Karin B. Michels *Editor*

Epigenetic Epidemiology

Second Edition

Springer

Epigenetic Epidemiology

Karin B. Michels
Editor

Epigenetic Epidemiology

Second Edition

Springer

Editor
Karin B. Michels
Department of Epidemiology
UCLA Fielding School of Public Health
Los Angeles, USA

ISBN 978-3-030-94474-2 ISBN 978-3-030-94475-9 (eBook)
https://doi.org/10.1007/978-3-030-94475-9

This Springer imprint is published by the registered company Springer Nature Switzerland AG
The registered company address is: Gewerbestrasse 11, 6330 Cham, Switzerland

Preface

This book joins two fields, epidemiology and epigenetics, to take advantage of their respective strengths in creating the science of epigenetic epidemiology. Epidemiology is the study of the frequency, distribution, and determinants of health and disease in humans. As a science fundamental to the study of public health, epidemiology is concerned with the prevention and effective control of disease. Epidemiology has early roots with the Greek physician Hippocrates, was and is essential in resolving infectious disease epidemics and pandemics ranging from cholera to COVID-19, and takes center stage in unveiling the causes of the chronic disease epidemics of our times including cardiovascular disease, diabetes, and cancer [1].

Epigenetics is the mitotically heritable state of the gene expression potential [2]. Gene expression is the response to cell-specific extracellular signals, and epigenetic mechanisms such as DNA methylation and histone modification govern the ability to respond appropriately to these signals. While the field of epigenetics has evolved over the past four decades, interest has been increasing exponentially. While initial work was conducted in plants and animal models, the focus has shifted on the role of epigenetics in human health and disease. Many intriguing and important observations have emerged, but numerous fundamental questions in epigenetic mechanisms in humans remain unanswered, providing abundant opportunities for discovery in the context of epidemiologic studies.

These two "epi" sciences (epidemiology: upon the people; epigenetics: above genetics) meet at the intersection of epigenetic variation and the distribution of disease [3]. Epigenetic epidemiology is defined as the study of the association between epigenetic variation and the risk of disease in humans [2]. Marrying a bench science and a population science creates both challenges and opportunities. The amalgamation of the two fields creates a science that supports the study of the role of epigenetic modifications in human disease etiology, the appreciation of epigenetics as a possible mechanistic link between environmental exposures and disease outcomes, and the discovery of new disease biomarkers [3]. Since the epigenetic signature is amenable to changes by environmental stressors, identifying factors that create or correct disease-specific patterns provides new possibilities for prevention and treatment.

Epidemiology and epigenetics share the elements of time and variability. Epidemics vary with time and infectious disease epidemics arise in regular intervals.

The epigenetic code—unlike the genetic code—is modifiable [4] and, while fairly robust [5] changes with age [6–8] and as a result of environmental influences [3, 8–10]. Epigenetic epidemiology relies on associations between those epigenetic marks with considerable interindividual variability and the incidence of disease.

This book is intended to be a resource for epidemiologists and epigeneticists alike. It provides insights into the mechanisms and methods in both fields to enable scientists to learn from each other, collaborate, and conduct qualitatively sound studies. Epidemiologists wishing to incorporate an epigenetic component into their epidemiologic study will find useful tools here such as guidance on the appropriate epigenetic methods and specifics about strengths and weaknesses of various laboratory assays. Epigeneticists will find relevant information on how to embed their research ideas into a population-based study, how to choose their study design and population, what pitfalls to watch out for, and the appropriate statistical analyses of their research findings.

In this second edition of "Epigenetic Epidemiology," we build and expand upon our successful previous collection and include chapters on the formation of the epigenome during development, the role of genomic imprinting, the role of epigenetics as a mechanistic underpinning of the developmental origins of health and disease, the potential impact of assisted reproductive technology on the epigenome, the influence of age and environmental factors on the epigenetic profile, and summaries of the state of the art in epigenetic epidemiologic research on a number of important diseases including cancer, infectious diseases, inflammation and rheumatoid arthritis, asthma, autism and other neurodevelopmental disorders, psychiatric disorders, metabolic disorders, and vascular disease. We also add significant new advances in epigenetics made since our first edition with particular emphasis on cutting-edge developments such as the epigenetic clock [11], epigenome editing [12], the Epigenome Roadmap [13], the Genotype-Tissue Expression (GTEx) project [14, 15], methylation patterns of circulating cell-free DNA [16], single-cell analysis [17, 18], and oncohistones [19].

We hope that this second edition of "Epigenetic Epidemiology" provides a useful tool in advancing this blossoming field encompassing the study of epigenetic variation in large heterogeneous populations as well as epigenome-wide association studies (EWAS).

Los Angeles, CA Karin B. Michels

References

1. Gordis L (2018) Epidemiology, 6th ed. Saunders, Elsevier, Philadelphia
2. Waterland RA, Michels KB (2007) Epigenetic epidemiology of the developmental origins hypothesis. Annu Rev Nutr 27:363–388. https://doi.org/10.1146/annurev.nutr.27.061406.093705
3. Michels KB (2010) The promises and challenges of epigenetic epidemiology. Exp Gerontol 45:297–301. https://doi.org/10.1016/j.exger.2009.12.011

4. Jaenisch R, Bird A (2003) Epigenetic regulation of gene expression: how the genome integrates intrinsic and environmental signals. Nat Genet 33 Suppl:245–254. https://doi.org/10.1038/ng1089
5. Cruz-Correa M, Zhao R, Oviedo M, et al (2009) Temporal stability and age-related prevalence of loss of imprinting of the insulin-like growth factor-2 gene. Epigenetics 4:114–118. https://doi.org/10.4161/epi.4.2.7954
6. Wilson VL, Jones PA (1983) DNA methylation decreases in aging but not in immortal cells. Science 220:1055–1057. https://doi.org/10.1126/science.6844925
7. Issa JP, Ottaviano YL, Celano P, et al (1994) Methylation of the oestrogen receptor CpG island links ageing and neoplasia in human colon. Nat Genet 7:536–540. https://doi.org/10.1038/ng0894-536
8. Bjornsson HT, Sigurdsson MI, Fallin MD, et al (2008) Intra-individual change over time in DNA methylation with familial clustering. JAMA 299:2877–2883. https://doi.org/10.1001/jama.299.24.2877
9. Feil R, Fraga MF (2012) Epigenetics and the environment: emerging patterns and implications. Nat Rev Genet 13:97–109. https://doi.org/10.1038/nrg3142
10. Cavalli G, Heard E (2019) Advances in epigenetics link genetics to the environment and disease. Nature 571:489–499. https://doi.org/10.1038/s41586-019-1411-0
11. Horvath S (2013) DNA methylation age of human tissues and cell types. Genome Biol 14:R115. https://doi.org/10.1186/gb-2013-14-10-r115
12. Nakamura M, Gao Y, Dominguez AA, Qi LS (2021) CRISPR technologies for precise epigenome editing. Nat Cell Biol 23:11–22. https://doi.org/10.1038/s41556-020-00620-7
13. Roadmap Epigenomics Consortium, Kundaje A, Meuleman W, et al (2015) Integrative analysis of 111 reference human epigenomes. Nature 518:317–330. https://doi.org/10.1038/nature14248
14. GTEx Consortium (2013) The Genotype-Tissue Expression (GTEx) project. Nat Genet 45:580–585. https://doi.org/10.1038/ng.2653
15. GTEx Portal. https://gtexportal.org/home. Accessed 25 Oct 2021
16. Moss J, Magenheim J, Neiman D, et al (2018) Comprehensive human cell-type methylation atlas reveals origins of circulating cell-free DNA in health and disease. Nat Commun 9:5068. https://doi.org/10.1038/s41467-018-07466-6
17. Argelaguet R, Clark SJ, Mohammed H, et al (2019) Multi-omics profiling of mouse gastrulation at single-cell resolution. Nature 576:487–491. https://doi.org/10.1038/s41586-019-1825-8
18. Stuart T, Satija R (2019) Integrative single-cell analysis. Nat Rev Genet 20:257–272. https://doi.org/10.1038/s41576-019-0093-7
19. Nacev BA, Feng L, Bagert JD, et al (2019) The expanding landscape of "oncohistone" mutations in human cancers. Nature 567:473–478. https://doi.org/10.1038/s41586-019-1038-1

Contents

List of Contributors

Nadya Ali Department of Psychiatry and Neuropsychology, School for Mental Health and Neuroscience (MHeNs), Maastricht University, Maastricht, The Netherlands

Karl Bacos Epigenetics and Diabetes Unit, Department of Clinical Sciences, Lund University Diabetes Centre, Lund University, Scania University Hospital, Malmö, Sweden

Ludovic Barault Department of Oncology, University of Torino, Torino, Italy
Candiolo Cancer Institute-FPO, IRCCS, Candiolo, Italy

Katherine Bassil Department of Psychiatry and Neuropsychology, School for Mental Health and Neuroscience (MHeNs), Maastricht University, Maastricht, The Netherlands

Jordana T. Bell Department of Twin Research and Genetic Epidemiology, Kings College London, London, UK

Alexandra M. Binder Population Sciences in the Pacific Program, University of Hawaii Cancer Center, Honolulu, HI, USA
Department of Epidemiology, Fielding School of Public Health, University of California, Los Angeles, CA, USA

James Chen Harvard College, Cambridge, MA, USA

Brock C. Christensen Department of Epidemiology, Department of Molecular and Systems Biology, Department of Community and Family Medicine, Geisel School of Medicine at Dartmouth, Hanover, NH, USA

Felicia Fei-Lei Chung Department of Medical Sciences, School of Medical and Life Sciences, Sunway University, Petaling Jaya, Malaysia

Montserrat Climent Humanitas Clinical and Research Center—IRCCS, Rozzano, MI, Italy

Christos Coutifaris Department of Obstetrics and Gynecology, University of Pennsylvania, Philadelphia, PA, USA

Leonardo Elia Humanitas Clinical and Research Center—IRCCS, Rozzano, MI, Italy
Department of Molecular and Translational Medicine, University of Brescia, Brescia, Italy

Todd M. Everson Department of Environmental Health, and Department of Epidemiology, Emory School of Public Health, Atlanta, GA, USA

Floriana Maria Farina Humanitas Clinical and Research Center—IRCCS, Rozzano, MI, Italy

Yumi Furuichi Division of Epigenomics, National Cancer Center Research Institute, Tokyo, Japan

Steffen Gay Department of Rheumatology, Center of Experimental Rheumatology, University Hospital Zurich, University of Zurich, Zurich, Switzerland

Akram Ghantous International Agency for Research on Cancer (IARC), Lyon, France

Jayashri Ghosh Fels Institute for Cancer Research and Molecular Biology, Lewis Katz School of Medicine, Temple University, Philadelphia, PA, USA

Chatura J. Gunasekara Department of Pediatrics, USDA/ARS Children's Nutrition Research Center, Baylor College of Medicine, Houston, TX, USA

Ignacio Fernando Hall Humanitas Clinical and Research Center—IRCCS, Rozzano, MI, Italy

Naoko Hattori Division of Epigenomics, National Cancer Center Research Institute, Tokyo, Japan

Zdenko Herceg International Agency for Research on Cancer (IARC), Lyon, France

Steve Horvath Department of Human Genetics, David Geffen School of Medicine, University of California, Los Angeles, CA, USA
Department of Biostatistics, Fielding School of Public Health, University of California, Los Angeles, CA, USA

Jean-Pierre J. Issa Coriell Institute for Medical Research, Camden, NJ, USA

Yong-Hui Jiang Department of Genetics, Yale University School of Medicine, New Haven, CT, USA
Department of Neuroscience and Pediatrics, Yale University School of Medicine, New Haven, CT, USA

R. M. John Cardiff School of Biosciences, Cardiff University, Cardiff, UK

Peter A. Jones Department of Epigenetics, Van Andel Institute, Grand Rapids, MI, USA

Kelsey Keith Coriell Institute for Medical Research, Camden, NJ, USA

Karl T. Kelsey Department of Epidemiology, and Department of Pathology and Laboratory Medicine, The Warren Alpert Medical School of Brown University, Providence, RI, USA

L. Lefebvre Department of Medical Genetics, University of British Columbia, Vancouver, BC, Canada

Gangning Liang Department of Epigenetics, Van Andel Institute, Grand Rapids, MI, USA

Charlotte Ling Epigenetics and Diabetes Unit, Department of Clinical Sciences, Lund University Diabetes Centre, Lund University, Scania University Hospital, Malmö, Sweden

Monica Mainigi Department of Obstetrics and Gynecology, University of Pennsylvania, Philadelphia, PA, USA

Sneha Mani Department of Obstetrics and Gynecology, University of Pennsylvania, Philadelphia, PA, USA

Carmen J. Marsit Department of Environmental Health, and Department of Epidemiology, Emory School of Public Health, Atlanta, GA, USA

Karin B. Michels Department of Epidemiology, Fielding School of Public Health, University of California, Los Angeles, CA, USA
Institute for Prevention and Cancer Epidemiology, Faculty of Medicine and Medical Center, University of Freiburg, Freiburg, Germany

Rachel L. Miller Division of Clinical Immunology, Department of Medicine, Icahn School of Medicine at Mount Sinai, New York, NY, USA

Caroline Ospelt Department of Rheumatology, Center of Experimental Rheumatology, University Hospital Zurich, University of Zurich, Zurich, Switzerland

Shoghag Panjarian Coriell Institute for Medical Research, Camden, NJ, USA

Ehsan Pishva Department of Psychiatry and Neuropsychology, School for Mental Health and Neuroscience (MHeNs), Maastricht University, Maastricht, The Netherlands

Rebecca C. Rancourt Division of 'Experimental Obstetrics', Clinic of Obstetrics, Charité—Universitätsmedizin Berlin, corporate member of Freie Universität Berlin, Humboldt-Universität zu Berlin, and Berlin Institute of Health, Berlin, Germany

Wendy P. Robinson BC Children's Hospital Research Institute, Vancouver, BC, Canada
Department of Medical Genetics, University of British Columbia, Vancouver, BC, Canada

Tina Rönn Epigenetics and Diabetes Unit, Department of Clinical Sciences, Lund University Diabetes Centre, Lund University, Scania University Hospital, Malmö, Sweden

Sabrina Ruhrmann Epigenetics and Diabetes Unit, Department of Clinical Sciences, Lund University Diabetes Centre, Lund University, Scania University Hospital, Malmö, Sweden

Richard Saffery Molecular Immunity, Murdoch Children's Research Institute (MCRI), Parkville, VIC, Australia
Department of Paediatrics, The University of Melbourne, Parkville, VIC, Australia

Johanna Säll Epigenetics and Diabetes Unit, Department of Clinical Sciences, Lund University Diabetes Centre, Lund University, Scania University Hospital, Malmö, Sweden

Carmen Sapienza Fels Institute for Cancer Research and Molecular Biology, Lewis Katz School of Medicine, Temple University, Philadelphia, PA, USA

Suneeta Senapati Department of Obstetrics and Gynecology, University of Pennsylvania, Philadelphia, PA, USA

M. A. Surani Wellcome Trust/Cancer Research UK Gurdon Institute, University of Cambridge, Cambridge, UK

Hideyuki Takeshima Division of Epigenomics, National Cancer Center Research Institute, Tokyo, Japan

Andrew E. Teschendorff CAS Key Laboratory of Computational Biology, CAS-MPG Partner Institute for Computational Biology, Shanghai, China
UCL Cancer Institute, University College London, London, UK

Rochelle L. Tiedemann Department of Epigenetics, Van Andel Institute, Grand Rapids, MI, USA

Toshikazu Ushijima Division of Epigenomics, National Cancer Center Research Institute, Tokyo, Japan

Daniel L. A. van den Hove Department of Psychiatry and Neuropsychology, School for Mental Health and Neuroscience (MHeNs), Maastricht University, Maastricht, The Netherlands
Laboratory of Translational Neuroscience, Department of Psychiatry, Psychosomatics and Psychotherapy, University of Würzburg, Würzburg, Germany

Sung Eun Wang Department of Genetics, Yale University School of Medicine, New Haven, CT, USA

Robert A. Waterland Department of Pediatrics, USDA/ARS Children's Nutrition Research Center, Baylor College of Medicine, Houston, TX, USA
Department of Molecular and Human Genetics, Baylor College of Medicine, Houston, TX, USA

Victor Yuan BC Children's Hospital Research Institute, Vancouver, BC, Canada
Department of Medical Genetics, University of British Columbia, Vancouver, BC, Canada

Part I

Basic Concepts and Methodologic Considerations

The Human Epigenome

1

Rochelle L. Tiedemann, Gangning Liang, and Peter A. Jones

Contents

Abstract

The output of the genome is controlled by the interaction of transcription factors with the epigenome. Epigenetic processes such as DNA methylation, histone modification, histone variants, noncoding RNAs, and nucleosomal remodeling machines interact with each other to ensure stable states of gene expression. These processes can become dysregulated during aging, exposure to environmental stressors, and the development of cancer and other diseases. DNA methylation patterns can be relatively easily read by high throughput techniques and provide information reflecting the influence of the environment and aging on the functionality of the epigenome. Analysis of DNA methylation patterns, therefore, provides an exciting new route to understanding how the environment interacts with the epigenome to cause disease. Despite the promise of DNA methylation patterns for epidemiologic studies, caution in interpreting data from surrogate tissues is necessary and cellular heterogeneity can also complicate interpretation of the data. In addition, DNA methylation within the

R. L. Tiedemann · G. Liang · P. A. Jones (✉)
Department of Epigenetics, Van Andel Institute, Grand Rapids, MI, USA
e-mail: rochelle.tiedemann@vai.org; gliang@usc.edu; peter.jones@vai.org

K. B. Michels (ed.), *Epigenetic Epidemiology*,
https://doi.org/10.1007/978-3-030-94475-9_1

body of genes can influence the response of the genome to the environment. Hypomethylation of repetitive elements can lead to genomic instability and ectopic gene expression. Methylation of coding regions can directly increase the rate of spontaneous hydrolytic mutations and increase the mutational frequency induced by carcinogens and radiation. Epigenetic processes can therefore contribute in multiple ways to the development of human diseases particularly cancer.

Abbreviations

5mC	5-methylcytosine
5hmC	5-hydroxymethylcytosine
5fC	5-formylcytosine
5caC	5-carboxylcytosine
CHD	Chromodomain Helicase DNA binding protein
ChIP	Chromatin immunoprecipitation
DNMT1	DNA methyltransferase 1
DNMT3A	DNA methyltransferase 3A
DNMT3B	DNA methyltransferase 3B
eRNA	enhancer RNA
ecRNA	extra-coding RNA
ISWI	imitation SWItch
KDM	lysine demethylase
lncRNA	long noncoding RNA
ncRNA	noncoding RNA
OxBS	oxidative bisulfite sequencing
PRC2	Polycomb Repressive Complex 2
SAM	S-adenosine methionine
SWI/SNF	SWItch/Sucrose Non-Fermentable
TAB-seq	Tet-assisted bisulfite sequencing
TDG	thymine DNA glycosylases
TETs	Ten-eleven translocation enzymes
UDG	uracil DNA glycosylases
UHRF1	Ubiquitin-like, containing PHD and RING finger domains, 1

1.1 Introduction

The genetic information encoded in the DNA of living organisms has to be read and interpreted in cells in such a way that its expression is highly controlled in response to developmental and environmental cues. Eukaryotic organisms, unlike prokaryotes, package their DNA into chromatin in which the fundamental building block is the nucleosome consisting of ~146 bp of DNA wrapped around an octamer of histones (Fig. 1.1). This packaging is essential to fit the DNA into the confines of the

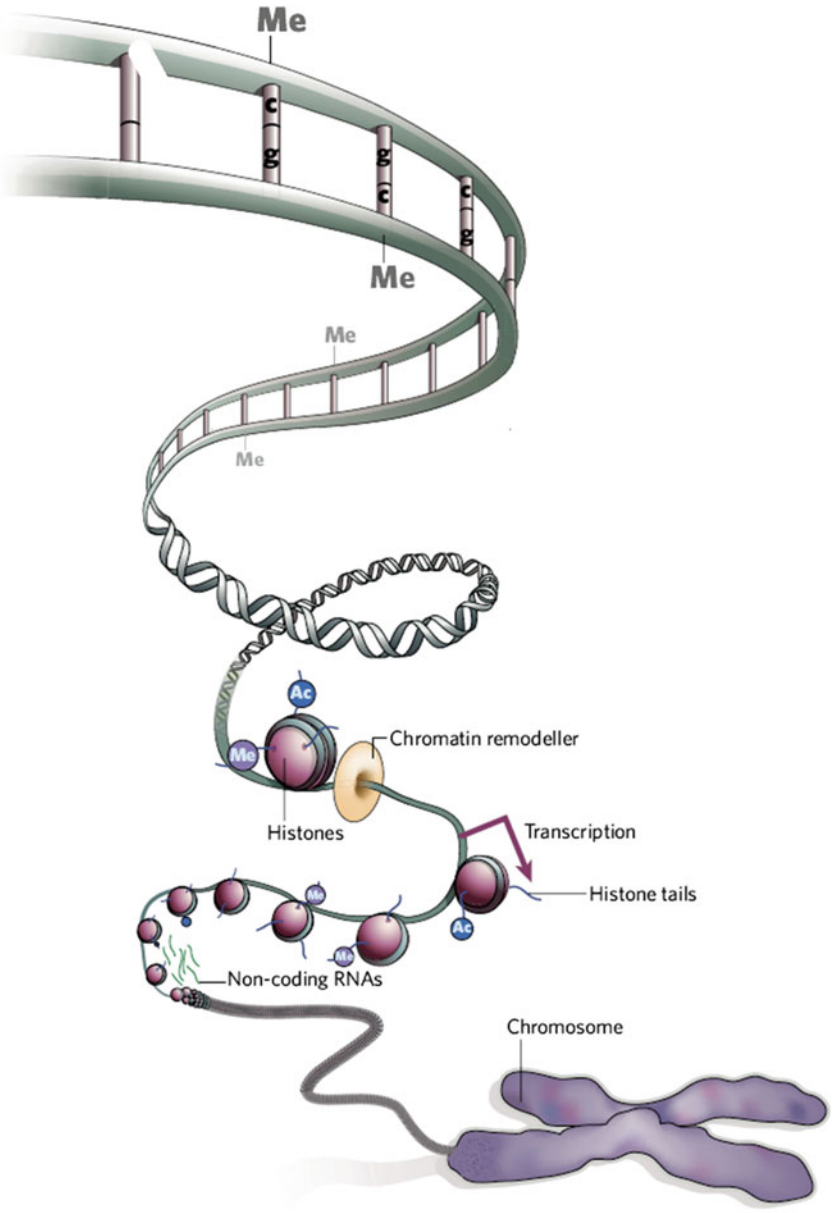

Fig. 1.1 Epigenetic processes. The DNA in living cells is complexed with proteins and RNA to fit into the structural confines of the mammalian nucleus. Most DNA is found in nucleosomes which contain about 146 base pairs of DNA wound around a histone octamer. The DNA can be modified by the application of methyl groups to cytosine residues in the simple palindromic sequence CpG. The tails of the histones and some internal amino acid residues are subject to posttranslational modifications which have significance in terms of dictating transcriptional competence. Nucleosome remodelers are necessary to expose regions of DNA so that it is accessible to the transcriptional and regulatory machinery. Noncoding RNAs also participate in the organization and functionality of chromatin. The various covalent marks communicate with each other and with

mammalian nucleus and also to provide functionality in different cell types. The combination of DNA and histones within the nucleosome is inherently refractory to transcription and nucleosomes have to be moved around or even evicted from particular places to allow gene expression to occur. This chromatin substrate, which is read by transcription factors in differentiated cell types, is what constitutes the epigenome. The accessibility or lack thereof of the genetic code is governed by chemical modifications which are applied to both the DNA and the protein components of chromatin and recent advances in high-throughput technologies now allow us to read these epigenetic modifications in their entireties in differentiated cell types.

Datei: 978-94-007-2495-2_2_Fig1_SB7.gif
Abb.-Typ: Strich-Abb.
Farbigkeit (IST): 1c
Farbigkeit (SOLL): 1c
Bildrechte: [Urheberrecht beim Autor]
Abdruckrechte: Nicht notwendig
Hinweise Verlag/Setzerei:

The roles of the different modifications during development and stability of differentiated states are now beginning to become apparent, as are the switching mechanisms which occur during the development of a human from a fertilized egg. Epigenetic information is heritable in somatic cells and can be copied after DNA synthesis and mitosis to ensure stability of cellular states. However, the epigenome is susceptible to alterations induced by the environment, nutrition, and other factors, so that potential changes in the packaging of genetic information can subsequently be copied in differentiated cells giving rise to both normal and abnormal cell states. Missteps in epigenetic processes can give rise to cancer and possibly to several other human diseases. We are just beginning to understand the multiple effects of the environment on epigenetic modifications and since these are potentially reversible, there is the possibility that several different diseases which have an epigenetic basis may be subject to pharmacological rectification [1].

1.2 Four Interacting Systems of Epigenetic Control

DNA in the nucleus of the cell is wrapped around an octamer containing 8 histone molecules in the fundamental structure of the nucleosome. The nucleosome contains about 146 bp of DNA and packages the DNA into the confines of the nucleus and also controls the output of the genome (Fig. 1.1). The nucleosome compactness is

Fig. 1.1 (continued) remodeling machines to define the structure of different epigenomes and different cell types (Reprinted with permission from *Nature* [1])

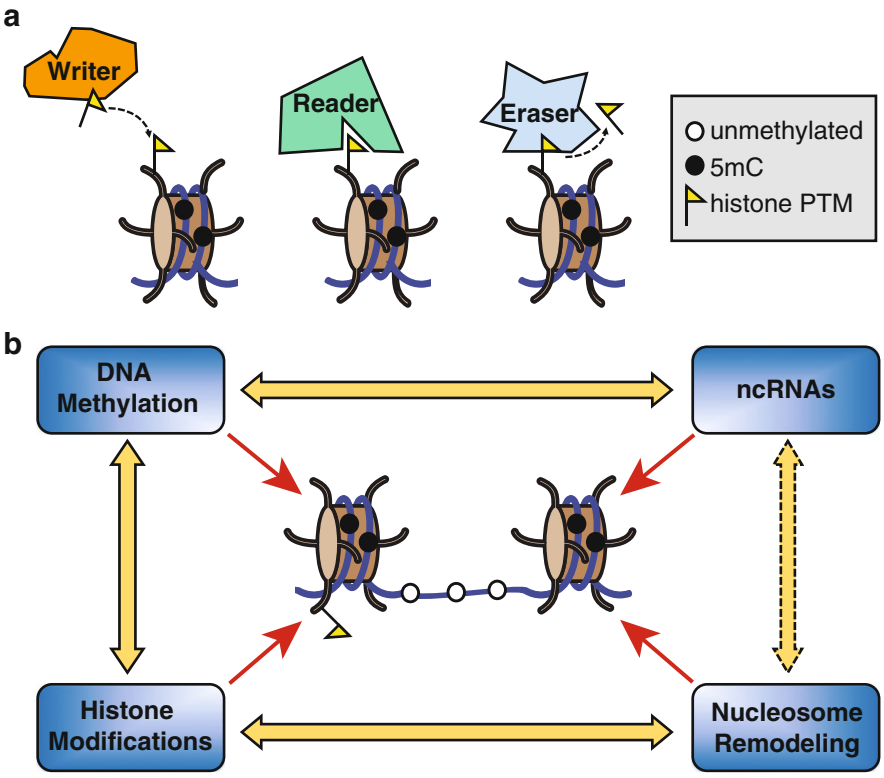

Fig. 1.2 Four interacting systems ensure epigenetic control. (**a**) Chromatin modifying enzymes act on chromatin in three primary ways. "Writers" deposit chemical modifications on the nucleosomal histones and DNA. "Readers" contain specific domains that bind certain chemical modifications on the nucleosome and DNA. "Erasers" remove chemical modifications from chromatin. (**b**) The four epigenetic processes outlined (among others) communicate to ensure somatically heritable states of gene expression in the context of the nucleosome. For example, DNA methylation and histone modification interact with each other and chromatin remodeling machines to position nucleosomes into active or repressive states. DNA methylation can also control the expression of noncoding RNAs which in turn can alter DNA methylation states in plants and human cells. These systems are mutually self-reinforcing and therefore can both initiate silencing and maintain previously silenced states. 5mC (5-methylcytosine); histone PTMs (histone post-translational modifications)

quite refractory to the initiation of transcription and nucleosomes need to be physically moved by nanomolecular machines to open up the DNA and allow transcription factors to initiate transcription.

The packaging and output are controlled by interaction between the various molecules that constitute chromatin and these systems interact with each other as depicted in Fig. 1.2. A variety of covalent marks and the presence of distinct histone variants, together with the involvement of noncoding RNA (ncRNA), are essential to the proper control of gene activity. DNA may become modified by the application of methyl groups to the 5 position of the cytosine ring and patterns of DNA methylation which are established during early development and differentiation can be copied

giving rise to somatically heritable states of gene expression which can be passed from one daughter cell to the next.

Datei: 978-94-007-2495-2_2_Fig2_SB7.gif
Abb.-Typ: Strich-Abb.
Farbigkeit (IST): 1c
Farbigkeit (SOLL): 1c
Bildrechte: [Urheberrecht beim Autor]
Abdruckrechte: Nicht notwendig
Hinweise Verlag/Setzerei:

The nucleosome contains two copies of each of the basic histone proteins, H2A, H2B, H3, and H4, and the tails of these histones are subject to a large number of covalent modifications which convey information regarding the stability of the nucleosome and its accessibility to transcription (Fig. 1.2). Some of these covalent modifications, such as acetylation and methylation of particular lysine residues are associated with active gene transcription. On the other hand, modifications such as methylation of other lysine residues can result in transcriptional silencing. In addition, certain histone variants such as histone H3.3 or histone H2A.Z are inserted into nucleosomes within specific places in the transcriptional unit and play important roles in gene activation or repression.

The covalent modifications of DNA and of histones communicate with each other at a biochemical level in ways that are now being unraveled. Together these processes can collaborate to ensure stable states of transcriptional competency. An enduring paradigm that has emerged over the past two decades recognizes that chromatin modifiers function in three predominant ways: (1) writers, (2) readers, and (3) erasers of covalent chromatin modifications [2] (Fig. 1.2a). Importantly, an individual chromatin modifier can perform a combination of these functions. For instance, the maintenance of DNA methylation is a coordinated effort between two different chromatin modifiers, UHRF1 (Ubiquitin-like, containing PHD, and RING finger domains, 1), and DNMT1 (DNA methyltransferase 1), which are both capable of reading and writing covalent chromatin modifications. UHRF1 can read both histone and DNA modifications in a nucleosomal context, and when it recognizes hemimethylated DNA, UHRF1 writes a ubiquitin mark onto the histone H3 tail [3–6]. DNMT1 reads the ubiquitin mark through its N-terminus, and then writes DNA methylation onto the daughter strand of DNA following replication [7]. Additionally, covalent modifications on chromatin can be removed (erased) through similar reader recognition and subsequent demethylase (eraser) activity. For example, Su and colleagues [8] identified that the lysine demethylase KDM4B reads H3K23me3, which subsequently stimulates the erasure of H3K36me3 by the enzyme. We have highlighted two examples of how chromatin modifiers coordinate writer, reader, and eraser functions to regulate chromatin structure and function; however, the predominance of these fundamental epigenetic processes is expansive and extends to nucleosome remodeling and ncRNA mechanisms as well.

Figure 1.2b also depicts the role of nucleosomal remodeling in epigenetic processes. As mentioned earlier, the presence of nucleosomes at transcriptional start sites is inherently refractory to transcriptional initiation and a whole series of multiprotein complexes use the energy of ATP to move nucleosomes around and expose different parts of the DNA thus allowing transcription to occur. Our understanding of how chromatin remodelers interact with the underlying epigenomic landscape to direct nucleosome positioning is becoming ever more clear and exceedingly complex. Four different chromatin remodeling subfamilies (ISWI (Imitation SWItch), CHD (Chromodomain Helicase DNA binding protein), SWI/SNF (SWItch/Sucrose Non-Fermentable), INO80) share the roles of nucleosome assembly and organization, chromatin accessibility, and exchange of histones (including variants) into nucleosome octamers [9]. As with other chromatin modifying enzymes, chromatin remodelers such as SWI/SNF and CHD complexes are targeted to particular genomic locations through reader domains that recognize the underlying covalent modifications on chromatin to mediate changes in nucleosome occupancy [9].

Recent work has cemented the important role of ncRNA species in the establishment and stability of epigenetic states. The role of RNA is quite well understood in organisms such as yeast and also in plants where it has been shown directly that ncRNAs can lead to DNA methylation and histone modifications which are important for keeping particular chromosomal regions silent by the formation of more densely packed configurations. In mammals, for example, the long ncRNA (lncRNA) Xist is expressed from the future inactive X-chromosome and initiates deposition of repressive histone modifications and DNA methylation to silence one of the X-chromosomes in females [10]. Additionally, certain microRNAs can downregulate chromatin modifiers such as DNA methyltransferases [11] or histone methyltransferases [12, 13]. The exact role of RNA in human epigenetic states has expanded over the past decade to encompass a large repertoire of RNA molecules with diverse functions for chromatin regulation. For example, we now appreciate that distal enhancers, when highly active, express enhancer RNAs (eRNAs) that bind to chromatin modifiers such as CBP and BRD4 to facilitate histone acetylation and transcriptional cofactor recruitment, respectively [14–16]. Additionally, extracoding RNAs (ecRNAs) transcribed from nearby protein-coding gene promoters can directly interact with DNMT1 to inhibit methyltransferase activity and permit mRNA transcription of target genes [17, 18]. The lncRNA HOTAIR binds to Polycomb Repressive Complex 2 (PRC2) and recruits the complex to specific genomic loci to deposit trimethylation on lysine 27 of histone H3 (H3K27me3) for gene silencing [19]. Collectively, these studies demonstrate that ncRNAs play an essential role in regulating chromatin structure to mediate transcriptional responses. As new technologies continue to be developed to dissect interactions between ncRNA and chromatin, our understanding of chromatin regulation by these mechanisms will become more clear.

Recent developments in chromatin immunoprecipitation (ChIP) procedures and the coupling of these approaches with high throughput sequencing now makes it relatively simple to map the distribution of the different epigenetic marks on a genome-wide basis. Over the past decade, major consortiums such as ENCODE,

the Epigenome Roadmap, and the International Human Epigenome Consortium (IHEC) have extensively mapped histone modifications, DNA methylation, gene expression, and chromatin structure across multiple tissue types, disease states, and cell lines providing an invaluable resource for researchers worldwide [20–22]. Significantly, the collection of these datasets has allowed for sophisticated computational methodologies, such as ChromHMM, to impute chromatin and transcriptional states across the genome [23]. Collectively, these efforts have provided a solid foundation for understanding the structure and function of the epigenomic landscape, and have allowed research groups to ask more pointed questions regarding the regulation of the epigenome.

Although it is clear that all of the processes outlined in Fig. 1.2b are contributing to epigenetic behavior and that all of them might be potentially altered by different environmental and nutritional influences. It is very likely that nutrition and the environment cause immediate and potentially reversible alterations to histone modifications, which could be the subject of detailed epidemiologic studies. This review will, however, focus on the promise and potential of DNA methylation analysis for epidemiologic studies because the 5-methylcytosine mark is inherently more stable than the chromatin structure and the code of DNA methylation can be more easily and quantitatively read so that its role in disease states can be better understood.

1.3 The Basics of DNA Methylation

About 1% of the cytosine residues in human DNA become methylated after the DNA is synthesized by the application of a methyl group from S-adenosine methionine (SAM) to the 5 position of the cytosine ring (Fig. 1.3a). The modification occurs very shortly after DNA has been synthesized although there clearly is some methylation that occurs hours after the DNA has left the replication fork [24]. Recent evidence supports this notion and suggests that high-density CpG dinucleotides are more rapidly maintained than low-density CpGs [25–27]. There are at least three enzymes that are responsible for setting up and maintaining DNA methylation patterns. DNA methyltransferase 3A (DNMT3A) and DNMT3B are thought to act early in development and can apply methyl groups to unmethylated and hemimethylated DNA (in which one strand has methylation and the other not). DNMT1 is thought to act primarily as a "maintenance enzyme" [28, 29] in that it has a preference for hemimethylated DNA and is the most active DNA methyltransferase in somatic cells [30]. All three of these enzymes have been shown, in gene knockout experiments in mice to be essential for mouse development demonstrating that DNA methylation is required for mammalian development.

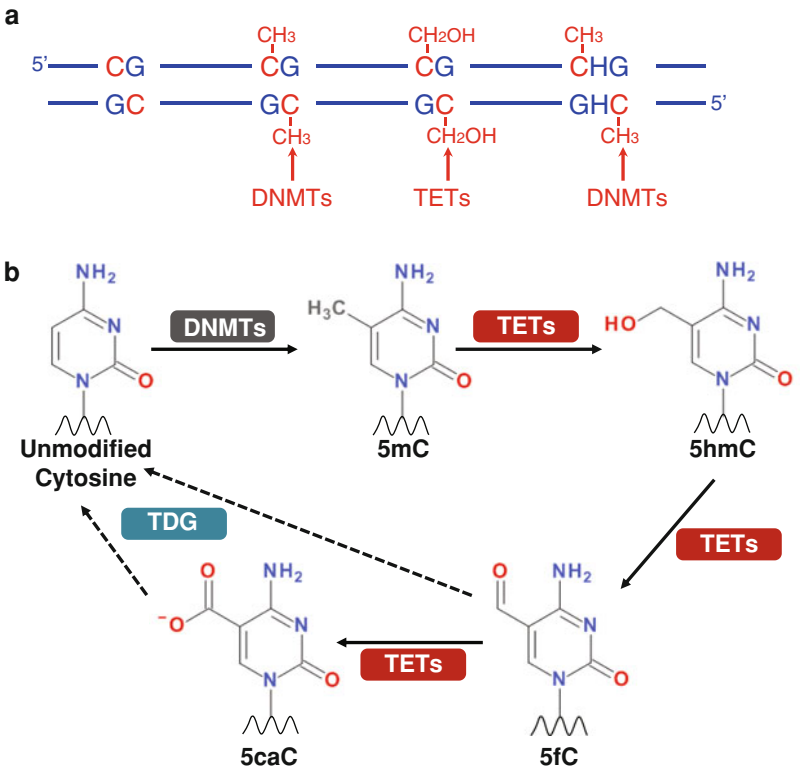

Fig. 1.3 Covalent modifications in DNA. (**a**) Almost all of the cytosine methylation in human DNA occurs in the simple palindrome CpG in which either both cytosines are methylated or neither are methylated as shown. The recent demonstration of 5-hydroxymethylcytosine (5hmC) in DNA suggests that a certain number of these sites are further modified by the TET proteins as indicated above. Recently non-CpG methylation in the sequence CHG (where *H* represents any base other than *G*) has been observed in human embryonic stem cells and brain tissue. These discoveries complicate the further dissection of the role of 5-methylcytosine (5mC) in gene control and need to be considered in epidemiologic studies. (**b**) Cytosine residues in DNA are modified by the application of a methyl group from S-adenosyl methionine to the 5 position of the cytosine ring. The TET proteins have been shown to be capable of further modification of 5-methylcytosine (5mC) by sequential oxidation of the methyl group substrate. 5hmC is the first oxidation product by the TET proteins; however, this modification is not recognized by thymine DNA glycosylase (TDG) for base excision repair. The TET proteins can further oxidize 5hmC to 5-formylcytosine (5fC) and 5-carboxylcytosine (5caC). Once the methyl group on cytosine is oxidized to 5fC and 5caC, TDG recognizes these modified cytosines for base excision repair to replace the oxidized bases with unmodified cytosines. Collectively, the process of sequential oxidation of 5mC by the TET proteins is widely accepted as the mechanism for active DNA demethylation

Datei: 978-94-007-2495-2_2_Fig3_SB7.gif
Abb.-Typ: Strich-Abb.
Farbigkeit (IST): 1c
Farbigkeit (SOLL): 1c
Bildrechte: [Urheberrecht beim Autor]
Abdruckrechte: Nicht notwendig
Hinweise Verlag/Setzerei:

These enzymes were thought to act largely in an autonomous manner with the DNMT3A and 3B "de novo methylases" required for early establishment of methylation patterns and DNMT1 then taking over to faithfully copy these patterns once they had been established. However, we have argued that this is unlikely to be the case and have proposed cooperativity between the enzymes in order to appropriately maintain DNA methylation patterns in normal and transformed cells [30]. The methylation activities of these enzymes are also regulated by accessory proteins which alter the methylating capability of the enzymes. For example, the catalytically inactive form of DNMT3 (DNMT3L) is strongly expressed in embryonic stem cells and is highly stimulatory to the DNA methylating activity of both DNMT3A and 3B [31]. Additionally, recent evidence from our group and others demonstrates that catalytically inactive DNMT3B isoforms can also act as accessory proteins for DNMT3A and 3B [32–34]. As described previously, DNMT1 also requires a cofactor, UHRF1, to effectively mediate maintenance DNA methylation in both embryonic and somatic cells [35–37].

The existence of tissue-specific patterns of DNA methylation has been known for a long time and these patterns are known to be strongly associated with gene expression. Methylation of gene promoters is commonly linked to silencing, and this mechanism may also exist at intragenic CpG islands to control tissue-specific expression of transcripts from alternative promoters [38]. Indeed, DNMT3B-mediated gene body methylation acts to deter spurious RNA-polymerase II entry and subsequent cryptic transcription [39]. The complexity of maintaining DNA methylation patterns is not completely understood; however, errors in the process can occur under normal conditions such as aging and in abnormal situations such as those which occur as a result of exposure to environmental insults. For this reason, the study of DNA methylation patterns in normal and diseased states has become of great importance. However, many of the observed alterations may have no direct role in the aging process or disease and it is still very difficult to distinguish between causative changes and alterations which have no functional consequence.

Almost all of the methylation of cytosine residues in somatic cell DNA occurs in the simple palindromic sequence, CpG (Fig. 1.3a). Most studies relating to DNA modification have focused on this covalent addition of the methyl group, however, it has recently become clear that other sequences and other modifications are also present in human DNA. For example, genome-wide studies in human embryonic stem cells have shown a high proportion of non-CpG methylation in particular

regions of human DNA in this early developmental state [40]. The physiologic significance of this is not understood, however, it may be due to the very high level of the DNMT3A and 3B enzymes in embryonic stem cells causing methylation at sites which subsequently lose their methylation at later stages of development possibly because the substrate generated after DNA synthesis would not be recognized and inherited by DNMT1. Notably, accumulation of non-CpG methylation (mediated by DNMT3A) is observed in an age-dependent manner in neurons of the brain; however, the functional role of non-CpG methylation in both of these tissue types remains poorly understood [41, 42].

Until recently, it was believed that 5-methylcytosine was the only modified base in human DNA but tremendous excitement has also been generated by the detection of 5-hydroxymethylcytosine in brain cell DNA and also in leukemia cells [43, 44]. The TET (Ten-Eleven Translocation) proteins are capable of oxidizing 5-methylcytosine in a stepwise manner to 5-hydroxymethylcytosine, 5-formylcytosine, and 5-carboxylcytosine (Fig. 1.3b). Importantly, thymine DNA glycosylase (TDG) can recognize 5-formylcytosine and 5-carboxylcytosine and excise these bases from the DNA strand to be replaced by unmodified cytosine [45, 46]. Collectively, oxidation of 5-methylcytosine is now widely accepted as a means of active DNA demethylation. The hydroxymethylation state is not distinguished from 5-methylcytosine by standard bisulfite sequencing technology which is often used to map DNA methylation in human cells; however, upstream modifications to standard bisulfite sequencing such as Tet-assisted bisulfite sequencing (TAB-seq) and Oxidative bisulfite sequencing (Ox-BS) now allow us to distinguish 5-methylcytosine from 5-hydroxymethylcytosine at base-pair resolution [47, 48]. By adopting these new approaches, active research continues to parse out the roles of these different DNA modifications in the genome to better understand how they contribute to chromatin accessibility and gene regulation.

A variety of methodologies have been used to determine DNA methylation levels and patterns, including restriction enzyme degradation, high pressure liquid chromatography, and bisulfite sequencing among others. The fact that the mark can be read in DNA extracted from formalin-fixed material and seems to be stable in specimens that have been kept for a long time has encouraged the use of DNA methylation as a marker for environmental exposures with the goal of determining the influence of these exposures on epigenetic processes.

1.4 Shaping of the Genome by DNA Methylation

DNA cytosine methylation has had a profound effect on the structure of the genome because the application of the methyl group to the 5 position on the cytosine ring creates mutational hotspots in DNA [49]. The methylation of DNA in germ cells has resulted in the depletion of the methylation acceptor site CpG during the course of evolution [50]. Recently, we have suggested that a major function of cytosine methylation is to suppress the activities of transposable elements in the genome. This has led not only to the huge expansion of the genomes of organisms that

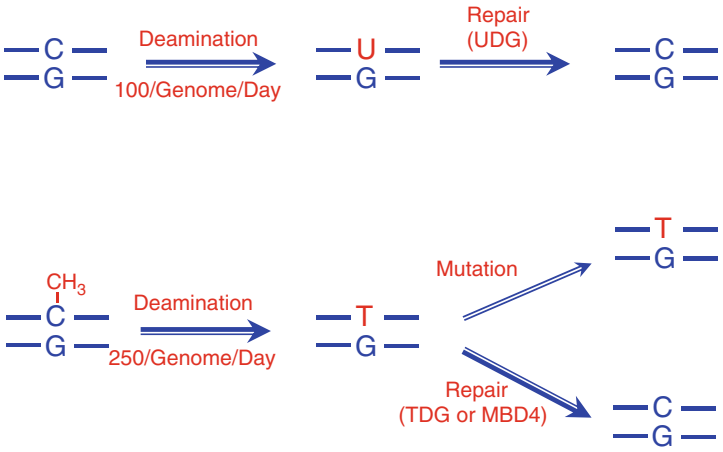

Fig. 1.4 5-Methylcytosine as a mutational hotspot. CpG sites in DNA are hotspots for C to T transition mutations in human DNA. Both cytosine and 5-methylcytosine can undergo spontaneous hydrolytic deamination to form uracil and thymine, respectively. Uracil is not a DNA base and is rapidly and accurately repaired by uracil DNA glycosylase (UDG). Thymine, being a natural component of DNA, is more difficult to accurately repair by thymine DNA glycosylase (TDG) or methyl binding protein domain 4 (MBD4). This has led to the loss of CpG sites which previously were methylated in the germline during evolution and has led to the generation of CpG islands which somehow have escaped methylation in the germline. CpG methylation in the germline contributes to about 30% of all new disease causing familial mutations in humans [50]. In somatic cells, it can lead to mutations in tumor suppressor genes thus causing cancer

methylate their DNAs but also to the creation of CpG islands [51]. Cytosine residues paired with guanines are known to undergo spontaneous hydrolytic deamination reactions of the order of 100 deaminations per genome per day (Fig. 1.4). The product of this deamination is an uracil residue which is not normally found in DNA and which can be rapidly and accurately repaired by ubiquitous and highly expressed uracil DNA glycosylase enzymes (UDG), which remove the uracil base and this results in the reinsertion of the cytosine residue so that no mutational events occur.

Datei: 978-94-007-2495-2_2_Fig4_SB7.gif
Abb.-Typ: Strich-Abb.
Farbigkeit (IST): 1c
Farbigkeit (SOLL): 1c
Bildrechte: [Urheberrecht beim Autor]
Abdruckrechte: Nicht notwendig
Hinweise Verlag/Setzerei:

The presence of a methyl group on the 5 position of the DNA increases the rate of spontaneous deamination by about 2.5-fold [52], but more importantly, results in the

generation of a thymine residue as a deamination product rather than uracil. Thymine, being a normal constituent of DNA, is more difficult to repair. Although thymine DNA glycosylases (TDG) or MBD4 are capable of repairing most deamination sites in the correct direction there is an increased possibility of the C to T transition mutation occurring following DNA methylation and deamination. As mentioned above, this process has led to the depletion of CpG sites in the bulk of human DNA because most of the CpG sites in germ cells are methylated and therefore subject to this increased mutability. Regions of DNA that are not methylated in germ cells have escaped this depletion of CpG sites and this has resulted in the presence of so-called "CpG islands" which are small regions of DNA about 1 kb in length, which occur in the promoters of a substantial portion of human genes [53]. These CpG islands usually remain unmethylated in all normal tissues and this is associated with transcriptional competency. On the other hand, the abnormal methylation of these CpG sites can cause gene silencing resulting in cancer and other diseases.

Figure 1.5 outlines how some cytosine methylations can be involved in normal gene control and can directly interact with the environment. About 50% of human genes contain unmethylated CpG islands in their promoters and first exons, whereas repetitive elements including Alus and LINES tend to be methylated, as do the coding portions of genes within the exons as indicated. Abnormal methylation of the CpG islands, which can be the result of copying errors associated with cell division, aging, diet, or exposure to carcinogens or other environmental stressors can result in the silencing of genes as indicated in Fig. 1.5a. This process has been well studied in cancer where between 1% and 10% of the CpG islands within genes have acquired abnormal methylation patterns during transformation [54]. The Fig. 1.5b also shows that demethylation of repetitive elements such as those within Alus and LINES can frequently occur [55] and this alteration is often related to disease outcomes in epidemiologic studies since these elements are abundant in DNA and their methylation status can be measured relatively easily using quantitative techniques such as pyrosequencing [56].

Datei: 978-94-007-2495-2_2_Fig5_SB7.gif
Abb.-Typ: Strich-Abb.
Farbigkeit (IST): 1c
Farbigkeit (SOLL): 1c
Bildrechte: [Urheberrecht beim Autor]
Abdruckrechte: Nicht notwendig
Hinweise Verlag/Setzerei:

Figure 1.5a also shows that the gene body methylation which occurs in exons can have profound effects on carcinogenesis. For example, the spontaneous deamination of these methylated sites can give rise to mutations in tumor suppressor genes [57]. The presence of 5-methylcytosine in the coding regions of genes increases

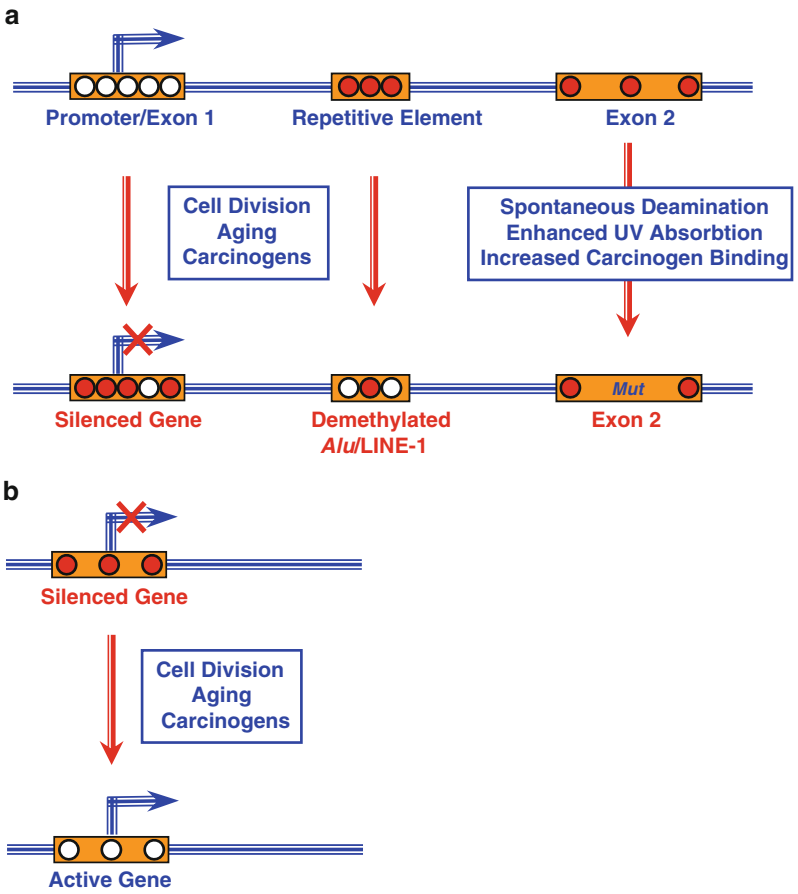

Fig. 1.5 The methylation status and effects of methylation on the carcinogenic process. (**a**) A promoter CpG island containing gene in which the gene is actively expressed due to a lack of methylation (shown as open circles) at the transcriptional start site. Oncogene and/or repetitive elements such as Alus or LINES located in the gene body are generally extensively methylated (closed circles) as are the CpG sites which are found within the coding regions of the gene. Promoter CpG islands can undergo inappropriate silencing and methylation of the CpG sites in response to aging, cell division, nutrition, or exposure to environmental carcinogens. Demethylation of repetitive elements has often been observed during carcinogenesis and can be easily measured because of the high copy number of these elements in human DNA. The methylation of CpG sites within the exon can increase the rate of somatic mutations directly by increasing the frequency of C to T transition mutations. It also can alter the interaction of the DNA with the environment. For example, it can increase UV absorption and increase the binding of carcinogens to DNA. (**b**) Often overlooked, is the demethylation of non-CpG island promoters which can result in inappropriate gene inactivation as opposed to silencing during carcinogenesis

the rate at which mutations are introduced by ultraviolet light during the development of skin cancers [58]. This is because 5-methylcytosine absorbs UV light at a wavelength more prevalent in sunlight than cytosine, thus increasing the chances of

mutations. Pfeifer et al. [58] have also shown that methylated CpG dinucleotides are the preferred targets for G to T transversions which are the most common mutations induced in mammalian cells by benzo(a)pyrene derived from tobacco smoke. Analysis of the mutational spectrum in human cancers can therefore provide evidence of value to the epidemiologist because the pattern of mutations can provide evidence of the most likely environmental cause [59]. For example, the high prevalence of C to T transition mutations at CpG dinucleotides in the p53 gene in colorectal cancer argues that these are most likely induced by endogenous processes potentially involving increased cell division [60] rather than being caused directly by exposure to carcinogens in the gut.

Figure 1.5b also shows that many tissue-specific genes which do not have CpG islands in their promoters can be ectopically activated as a function of aging and cell division. The potential role of non-CpG island methylation in gene control has been largely neglected in the field even though there is strong evidence that methylation of such regions can preclude gene expression [61]. Several studies have recently pointed to widespread hypomethylation of such regions in tumors and apparently normal cells adjacent to the tumor [55]. Since chemical carcinogens can inhibit DNA methylation, these processes can potentially result in the ectopic activation of genes which could play a significant role in the tumorigenic process.

1.5 Effects of the Environment on DNA Methylation

Soon after the discovery of the presence of 5-methylcytosine in DNA, work began to determine whether the levels of the modified base were altered in cancer and to determine whether chemical carcinogens could directly influence the methylation process. Lapeyre and Becker [62] showed that primary hepatocarcinoma and transplantable mouse liver tumors contained decreased levels of 5-methylcytosine relative to normal liver. Subsequently, human leukemias and other uncultured tumors were found to have altered levels of DNA methylation [63]. Many such studies showed alterations in the overall levels of DNA methylation in cancer cells; however, Feinberg and Vogelstein [64] were the first to show that the methylation of specific sites within individual gene bodies was decreased in uncultured tumors. These early studies summarized by Riggs and Jones [63] established clearly that DNA methylation was fundamentally altered in cell lines and cancers.

Given the emerging interest in the potential role of 5-methylcytosine in controlling gene expression [28, 29] there was increasing research activity in determining the potential role of DNA methylation in cancer. The thrust of this work was on the potential for carcinogens to heritably alter the regulation of genes rather than on their abilities to cause mutations as discussed earlier. Early pioneers such as Drahovsky and Morris [65] began work to determine whether chemical carcinogens could influence DNA methylation reactions in the test tube. These studies, also summarized in Riggs and Jones [63], pointed strongly to the possibility that chemicals in the environment including benzo(a)pyrene might be able to influence the DNA methylation machinery and that this could participate in the

oncogenic process. Evidence that this was indeed the case came from studies of Wilson and Jones [66] using cultured cells and Wilson et al. [67] using freshly explanted normal human bronchial epithelial cells.

In a recent landmark paper, Vaz et. al. [68] demonstrated that chronic cigarette smoke exposure of human bronchial epithelial cells in organoid cultures resulted in an accelerated change in the distribution of genomic methylation patterns. Importantly they observed de novo methylation of CpG islands which are normally silenced by the polycomb repressive complex. These studies along with many others which have been conducted over a long period of time, strongly support the idea that chemical and other carcinogens can directly impact genomic DNA methylation patterns. Indeed, analysis of genomic methylation patterns in many thousands of uncultured human cancers examined in the TCGA project, show that almost all of them harbor profoundly altered epigenomes compared to their normal counterparts. Also, the patterns can often be correlated with exposures to environmental factors or to somatically acquired or germline mutations [69–72].

1.6 The Role of Aging

A potential effect of aging on DNA methylation was suggested by Wilson and Jones [66] who showed that the lifespan of cells in culture was linked to the rate of overall loss of 5-methylcytosine levels. Subsequent experiments by Mays-Hoopes et al. [73] showed that alterations in DNA cytosine methylation also occurred in the intercisternal A particles (IAP genes) within mice. These studies which demonstrated hypomethylation as a function of age again emphasized that DNA methylation levels, while somatically heritable, were not completely invariant and could be used as markers for aging and exposure to chemicals.

The pioneering work of Ahuja and Issa and colleagues [74] in showing that hypermethylation of CpG islands in the colonic epithelium of people without cancer could be directly linked to the age, was also of great value in showing that the epigenome reacts to the increased cell division which accompanies aging. More recent work in mice has shown widespread and tissue-specific DNA methylation changes showing that epigenetic regulation is a common feature of aging in mammals [75]. Since aging is a major risk factor for cancer, these alterations might provide a biochemical basis for the subsequent development of tumors.

The development of bead-based arrays by Illumina now allows for the rapid interrogation of almost a million CpG sites from a given sample in a single experiment. Importantly the interrogated sites are widely distributed in the genome and include not only transcription start sites but also enhancers, gene bodies, and even some repetitive elements among others. An enormous amount of methylation data has now been deposited in databases leading to a much more comprehensive appreciation of methylation changes with aging. Several “epigenetic clocks” have been proposed. These clocks link developmental and maintenance processes to biological aging and are already showing themselves to have many practical and experimental uses [76]. Although some of the clocks were specifically developed to

use CpG sites which have no known functionality and are therefore useful across a wide range of tissues, others have concentrated on potentially functional changes such as the age-dependent methylation of genes that are suppressed in stem cells by the polycomb repressive complex [77]. It has also been proposed that DNA methylation loss in late-replicating domains is linked to the number of mitotic cell divisions providing alternative clocks [78].

Observations that DNA methylation patterns can be profoundly altered in aging in people without cancer show the plasticity of the epigenome. They also underline the importance of using age-matched controls in epidemiologic studies to investigate their alterations in this process and the relevance to development of cancer and other diseases. As mentioned previously, many DNA methylation changes may have no known significance in terms of genome function making it important for the important causative alterations to be determined in the future.

1.7 The Use of Surrogate Tissues

Epigenetic landscapes are tissue-specific and contribute to the phenotype of the cell. Unlike genetic studies, in which all differentiated cells in a subject have essentially the same markers such as SNPs, one cannot assume that surrogate tissues will necessarily have the same value in assessing the effect of the environment on a given marker such as DNA methylation. It is therefore not always feasible to take an easily available tissue such as peripheral blood and use this to examine DNA methylation changes that might be occurring in a different target tissue. Epigenomic epidemiologic studies are therefore more difficult to perform and evaluate than genetic epidemiologic studies. Another complicating factor, which limits the use of blood cells, is that they are a heterogeneous mixture whose composition can change dramatically in response to other cues such as infections. Because each specialized type of cell in the peripheral circulation would be expected to have a different epigenomic profile, a measured change might reflect a change in cellular composition rather than a change in the pattern in a given cell type. Since epidemiologic studies often demand a large number of subjects and sometimes repeat sampling, other cells to consider are buccal cells, urine sediments, sputum, and epidermal cells which can be relatively easily obtained.

Despite these reservations, useful information can be obtained from peripheral blood DNA methylation studies which might have value in determining the influence of nutrition or age, for example, on particular epigenomic marker. For example, the DNA methylation patterns of imprinted genes which are possibly methylated to similar extents in different tissues might be suitable as a surrogate although this remains to be shown. Another would be the methylation status of repetitive DNAs such as Alus and LINEs which do not show a great deal of inter tissue variations and which have been successfully used to measure changes in response to benzene exposure [79]. It is also important to consider the potential biological significance of relatively small changes which might be uncovered by these studies if there is an attempt being made to link the changes to a particular disease state. For example, it is

not known whether small changes in the methylation status of a given promoter necessarily translate an alteration in gene expression.

1.8 Appropriate Controls

As mentioned earlier, epigenetic analyses differ from genetic analysis because the epigenome is cell type-specific and is altered by environmental factors. This makes the appropriate selection of normal controls of great importance. For example, a comparison of the DNA methylation patterns in a set of tumors should be compared to age-matched controls because the epigenome is known to change with aging. Another complicating issue is the fact that epigenetic changes can often be observed in the cells surrounding the tumor. For example, we found that the entire urothelium of the bladders of patients with bladder cancer is altered with respect to DNA methylation patterns [55]. Thus, the surrounding normal tissue may already harbor DNA methylation changes which are either selected for in the tumor or more probably allow the tumor to grow by altering the integrity of the epithelium. Therefore, comparisons of normal appearing surrounding tissue to similar tissues from age-matched controls who do not have the particular disease of interest, is often necessary to fully appreciate the changes which occur during the process of transformation.

Despite this complication, the existence of DNA methylation changes in normal tissues surrounding a tumor may be of great value in understanding the mechanism of carcinogenesis. It is still not known whether these changes precede the formation of the tumor or are a response of the epithelium to the presence of a tumor in a bladder. It might be possible in the future to conduct these analyses on high-risk populations without cancer and predict cancer susceptibility. However, this will be limited to easily biopsied tissue and may not be applicable to all cancers, like brain cancer.

1.9 Summary

Interacting epigenetic processes ensure the somatic heritability of differentiated cell states and are set up early in development. These processes reinforce each other and can be influenced by environmental factors to alter gene expression in heritable ways which can cause disease. DNA methylation is a particularly attractive epigenetic process for epidemiologic studies since DNA methylation patterns can be quantitatively measured, are known to influence gene expression when located in controlling regions of genes and are subject to alterations associated with aging and exposure to environmental toxins. High-throughput approaches allow for the concomitant analysis of thousands of DNA methylation sites in large numbers of samples thus opening the door to future studies to link the influence of the environment to the epigenome. Caution, however, needs to be used when interpreting DNA methylation

data, particularly because many changes may have little functional significance and there is a need to examine the cell type of origin in order to gain useful information.

Although much neglected in the field, the methylation of cytosine residues within the coding regions of genes can directly contribute to carcinogenesis by increasing the frequency of both spontaneous and induced mutations. Analysis of the mutational spectrum in different disease states can give an indication of likely exogenous or endogenous causes [59, 80]. A large number of new epidemiologic studies including epigenetic analyses, such as those discussed in this book suggest that we are entering an age of epigenetic epidemiology and that much will be learned about the interaction of the epigenome and the environment and its relation to disease.

References

1. Jones PA, Archer TK, Baylin SB et al (2008) Moving AHEAD with an international human epigenome project. Nature 454:711–715. https://doi.org/10.1038/454711a
2. Rothbart SB, Strahl BD (2014) Interpreting the language of histone and DNA modifications. Biochimica et Biophysica Acta (BBA)—Gene Regulatory Mechanisms 1839:627–643. https://doi.org/10.1016/j.bbagrm.2014.03.001
3. Rothbart SB, Krajewski K, Nady N et al (2012) Association of UHRF1 with methylated H3K9 directs the maintenance of DNA methylation. Nat Struct Mol Biol 19:1155–1160. https://doi.org/10.1038/nsmb.2391
4. Nishiyama A, Yamaguchi L, Sharif J et al (2013) Uhrf1-dependent H3K23 ubiquitylation couples maintenance DNA methylation and replication. Nature 502:249–253. https://doi.org/10.1038/nature12488
5. Rothbart SB, Dickson BM, Ong MS et al (2013) Multivalent histone engagement by the linked tandem Tudor and PHD domains of UHRF1 is required for the epigenetic inheritance of DNA methylation. Genes Dev 27:1288–1298. https://doi.org/10.1101/gad.220467.113
6. Harrison JS, Cornett EM, Goldfarb D et al (2016) Hemi-methylated DNA regulates DNA methylation inheritance through allosteric activation of H3 ubiquitylation by UHRF1. eLife 5:e17101. https://doi.org/10.7554/eLife.17101
7. Qin W, Wolf P, Liu N et al (2015) DNA methylation requires a DNMT1 ubiquitin interacting motif (UIM) and histone ubiquitination. Cell Res 25:911–929. https://doi.org/10.1038/cr.2015.72
8. Su Z, Wang F, Lee J-H et al (2016) Reader domain specificity and lysine demethylase-4 family function. Nature Communications 7:13387. https://doi.org/10.1038/ncomms13387
9. Clapier CR, Iwasa J, Cairns BR, Peterson CL (2017) Mechanisms of action and regulation of ATP-dependent chromatin-remodelling complexes. Nature Reviews Molecular Cell Biology 18:407–422. https://doi.org/10.1038/nrm.2017.26
10. Cerase A, Pintacuda G, Tattermusch A, Avner P (2015) Xist localization and function: new insights from multiple levels. Genome Biology 16:166. https://doi.org/10.1186/s13059-015-0733-y
11. Iorio MV, Piovan C, Croce CM (2010) Interplay between microRNAs and the epigenetic machinery: An intricate network. Biochimica et Biophysica Acta (BBA)—Gene Regulatory Mechanisms 1799:694–701. https://doi.org/10.1016/j.bbagrm.2010.05.005
12. Friedman JM, Liang G, Liu C-C et al (2009) The Putative Tumor Suppressor microRNA-101 Modulates the Cancer Epigenome by Repressing the Polycomb Group Protein EZH2. Cancer Res 69:2623–2629. https://doi.org/10.1158/0008-5472.CAN-08-3114
13. Varambally S, Cao Q, Mani R-S et al (2008) Genomic loss of microRNA-101 leads to overexpression of histone methyltransferase EZH2 in cancer. Science 322:1695–1699. https://doi.org/10.1126/science.1165395

14. Bose DA, Donahue G, Reinberg D et al (2017) RNA Binding to CBP Stimulates Histone Acetylation and Transcription. Cell 168:135–149.e22. https://doi.org/10.1016/j.cell.2016.12.020
15. Rahnamoun H, Lee J, Sun Z et al (2018) RNAs interact with BRD4 to promote enhanced chromatin engagement and transcription activation. Nature Structural & Molecular Biology 25:687–697. https://doi.org/10.1038/s41594-018-0102-0
16. Sartorelli V, Lauberth SM (2020) Enhancer RNAs are an important regulatory layer of the epigenome. Nature Structural & Molecular Biology 27:521–528. https://doi.org/10.1038/s41594-020-0446-0
17. Di Ruscio A, Ebralidze AK, Benoukraf T et al (2013) DNMT1-interacting RNAs block gene-specific DNA methylation. Nature 503:371–376. https://doi.org/10.1038/nature12598
18. Savell KE, Gallus NVN, Simon RC et al (2016) Extra-coding RNAs regulate neuronal DNA methylation dynamics. Nature Communications 7:12091. https://doi.org/10.1038/ncomms12091
19. Gupta RA, Shah N, Wang KC et al (2010) Long non-coding RNA HOTAIR reprograms chromatin state to promote cancer metastasis. Nature 464:1071–1076. https://doi.org/10.1038/nature08975
20. ENCODE Project Consortium (2012) An integrated encyclopedia of DNA elements in the human genome. Nature 489:57–74. https://doi.org/10.1038/nature11247
21. Kundaje A, Meuleman W, Ernst J et al (2015) Integrative analysis of 111 reference human epigenomes. Nature 518:317–330. https://doi.org/10.1038/nature14248
22. Stunnenberg HG, Abrignani S, Adams D et al (2016) The International Human Epigenome Consortium: A Blueprint for Scientific Collaboration and Discovery. Cell 167:1145–1149. https://doi.org/10.1016/j.cell.2016.11.007
23. Ernst J, Kellis M (2010) Discovery and characterization of chromatin states for systematic annotation of the human genome. Nature Biotechnology 28:817–825. https://doi.org/10.1038/nbt.1662
24. Liang G, Chan MF, Tomigahara Y et al (2002) Cooperativity between DNA Methyltransferases in the Maintenance Methylation of Repetitive Elements. Mol Cell Biol 22:480–491. https://doi.org/10.1128/MCB.22.2.480-491.2002
25. Xu C, Corces VG (2018) Nascent DNA methylome mapping reveals inheritance of hemimethylation at CTCF/cohesin sites. Science 359:1166–1170. https://doi.org/10.1126/science.aan5480
26. Charlton J, Downing TL, Smith ZD et al (2018) Global delay in nascent strand DNA methylation. Nature Structural & Molecular Biology 25:327–332. https://doi.org/10.1038/s41594-018-0046-4
27. Ming X, Zhang Z, Zou Z et al (2020) Kinetics and mechanisms of mitotic inheritance of DNA methylation and their roles in aging-associated methylome deterioration. Cell Research 30:980–996. https://doi.org/10.1038/s41422-020-0359-9
28. Riggs AD (1975) X inactivation, differentiation, and DNA methylation. Cytogenet Cell Genet 14:9–25. https://doi.org/10.1159/000130315
29. Holliday R, Pugh JE (1975) DNA modification mechanisms and gene activity during development. Science 187:226–232. https://doi.org/10.1126/science.187.4173.226
30. Jones PA, Liang G (2009) Rethinking how DNA methylation patterns are maintained. Nature Reviews Genetics 10:805–811. https://doi.org/10.1038/nrg2651
31. Ooi SKT, Qiu C, Bernstein E et al (2007) DNMT3L connects unmethylated lysine 4 of histone H3 to de novo methylation of DNA. Nature 448:714–717. https://doi.org/10.1038/nature05987
32. Van Emburgh BO, Robertson KD (2011) Modulation of Dnmt3b function in vitro by interactions with Dnmt3L, Dnmt3a and Dnmt3b splice variants. Nucleic Acids Res 39:4984–5002. https://doi.org/10.1093/nar/gkr116
33. Duymich CE, Charlet J, Yang X et al (2016) DNMT3B isoforms without catalytic activity stimulate gene body methylation as accessory proteins in somatic cells. Nature Communications 7:11453. https://doi.org/10.1038/ncomms11453

34. Zeng Y, Ren R, Kaur G et al (2020) The inactive Dnmt3b3 isoform preferentially enhances Dnmt3b-mediated DNA methylation. Genes Dev. https://doi.org/10.1101/gad.341925.120
35. Sharif J, Muto M, Takebayashi S et al (2007) The SRA protein Np95 mediates epigenetic inheritance by recruiting Dnmt1 to methylated DNA. Nature 450:908–912. https://doi.org/10.1038/nature06397
36. Bostick M, Kim JK, Estève P-O et al (2007) UHRF1 Plays a Role in Maintaining DNA Methylation in Mammalian Cells. Science 317:1760–1764. https://doi.org/10.1126/science.1147939
37. Feng S, Cokus SJ, Zhang X et al (2010) Conservation and divergence of methylation patterning in plants and animals. PNAS 107:8689–8694. https://doi.org/10.1073/pnas.1002720107
38. Maunakea AK, Nagarajan RP, Bilenky M et al (2010) Conserved role of intragenic DNA methylation in regulating alternative promoters. Nature 466:253–257. https://doi.org/10.1038/nature09165
39. Neri F, Rapelli S, Krepelova A et al (2017) Intragenic DNA methylation prevents spurious transcription initiation. Nature 543:72–77. https://doi.org/10.1038/nature21373
40. Lister R, Pelizzola M, Dowen RH et al (2009) Human DNA methylomes at base resolution show widespread epigenomic differences. Nature 462:315–322. https://doi.org/10.1038/nature08514
41. Lister R, Mukamel EA, Nery JR et al (2013) Global Epigenomic Reconfiguration During Mammalian Brain Development. Science 341. https://doi.org/10.1126/science.1237905
42. Guo JU, Su Y, Shin JH et al (2014) Distribution, recognition and regulation of non-CpG methylation in the adult mammalian brain. Nature Neuroscience 17:215–222. https://doi.org/10.1038/nn.3607
43. Tahiliani M, Koh KP, Shen Y et al (2009) Conversion of 5-methylcytosine to 5-hydroxymethylcytosine in mammalian DNA by MLL partner TET1. Science 324:930–935. https://doi.org/10.1126/science.1170116
44. Kriaucionis S, Heintz N (2009) The nuclear DNA base 5-hydroxymethylcytosine is present in Purkinje neurons and the brain. Science 324:929–930. https://doi.org/10.1126/science.1169786
45. Ito S, Shen L, Dai Q et al (2011) Tet Proteins Can Convert 5-Methylcytosine to 5-Formylcytosine and 5-Carboxylcytosine. Science 333:1300–1303. https://doi.org/10.1126/science.1210597
46. He Y-F, Li B-Z, Li Z et al (2011) Tet-Mediated Formation of 5-Carboxylcytosine and Its Excision by TDG in Mammalian DNA. Science 333:1303–1307. https://doi.org/10.1126/science.1210944
47. Yu M, Hon GC, Szulwach KE et al (2012) Base-resolution analysis of 5-hydroxymethylcytosine in the mammalian genome. Cell 149:1368–1380. https://doi.org/10.1016/j.cell.2012.04.027
48. Booth MJ, Branco MR, Ficz G et al (2012) Quantitative Sequencing of 5-Methylcytosine and 5-Hydroxymethylcytosine at Single-Base Resolution. Science 336:934–937. https://doi.org/10.1126/science.1220671
49. Coulondre C, Miller JH, Farabaugh PJ, Gilbert W (1978) Molecular basis of base substitution hotspots in Escherichia coli. Nature 274:775–780. https://doi.org/10.1038/274775a0
50. Jones PA, Rideout WM, Shen J-C et al (1992) Methylation, mutation and cancer. BioEssays 14:33–36. https://doi.org/10.1002/bies.950140107
51. Zhou W, Liang G, Molloy PL, Jones PA (2020) DNA methylation enables transposable element-driven genome expansion. PNAS 117:19359–19366. https://doi.org/10.1073/pnas.1921719117
52. Shen J-C, Rideout WM, Jones PA (1994) The rate of hydrolytic deamination of 5-methylcytosine in double-stranded DNA. Nucleic Acids Res 22:972–976. https://doi.org/10.1093/nar/22.6.972
53. Takai D, Jones PA (2002) Comprehensive analysis of CpG islands in human chromosomes 21 and 22. PNAS 99:3740–3745. https://doi.org/10.1073/pnas.052410099

54. Jones PA, Baylin SB (2007) The Epigenomics of Cancer. Cell 128:683–692. https://doi.org/10.1016/j.cell.2007.01.029
55. Wolff EM, Byun H-M, Han HF et al (2010) Hypomethylation of a LINE-1 Promoter Activates an Alternate Transcript of the MET Oncogene in Bladders with Cancer. PLOS Genetics 6: e1000917. https://doi.org/10.1371/journal.pgen.1000917
56. Yang AS, Estécio MRH, Doshi K et al (2004) A simple method for estimating global DNA methylation using bisulfite PCR of repetitive DNA elements. Nucleic Acids Res 32:e38. https://doi.org/10.1093/nar/gnh032
57. Rideout WM, Coetzee GA, Olumi AF, Jones PA (1990) 5-Methylcytosine as an endogenous mutagen in the human LDL receptor and p53 genes. Science 249:1288–1290. https://doi.org/10.1126/science.1697983
58. Pfeifer GP, Tang M, Denissenko MF (2000) Mutation hotspots and DNA methylation. Curr Top Microbiol Immunol 249:1–19. https://doi.org/10.1007/978-3-642-59696-4_1
59. Jones PA, Buckley JD, Henderson BE et al (1991) From Gene to Carcinogen: A Rapidly Evolving Field in Molecular Epidemiology. Cancer Res 51:3617–3620
60. Preston-Martin S, Pike MC, Ross RK et al (1990) Increased Cell Division as a Cause of Human Cancer. Cancer Res 50:7415–7421
61. Simonsson S, Gurdon J (2004) DNA demethylation is necessary for the epigenetic reprogramming of somatic cell nuclei. Nature Cell Biology 6:984–990. https://doi.org/10.1038/ncb1176
62. Lapeyre J-N, Becker FF (1979) 5-Methylcytosine content of nuclear DNA during chemical hepatocarcinogenesis and in carcinomas which result. Biochemical and Biophysical Research Communications 87:698–705. https://doi.org/10.1016/0006-291X(79)92015-1
63. Riggs AD, Jones PA (1983) 5-Methylcytosine, Gene Regulation, and Cancer. In: Klein G, Weinhouse S (eds) Advances in Cancer Research. Academic Press, pp 1–30
64. Feinberg AP, Vogelstein B (1983) Hypomethylation distinguishes genes of some human cancers from their normal counterparts. Nature 301:89–92. https://doi.org/10.1038/301089a0
65. Drahovsky D, Morris NR (1972) The mechanism of action of rat liver DNA methylase: III. Nucleotide requirements for binding and methylation. Biochimica et Biophysica Acta (BBA)—Nucleic Acids and Protein Synthesis 277:245–250. https://doi.org/10.1016/0005-2787(72)90404-2
66. Wilson VL, Jones PA (1983) Inhibition of DNA methylation by chemical carcinogens in vitro. Cell 32:239–246. https://doi.org/10.1016/0092-8674(83)90514-7
67. Wilson VL, Smith RA, Longoria J et al (1987) Chemical carcinogen-induced decreases in genomic 5-methyldeoxycytidine content of normal human bronchial epithelial cells. Proc Natl Acad Sci U S A 84:3298–3301
68. Vaz M, Hwang SY, Kagiampakis I et al (2017) Chronic Cigarette Smoke-Induced Epigenomic Changes Precede Sensitization of Bronchial Epithelial Cells to Single-Step Transformation by KRAS Mutations. Cancer Cell 32:360–376.e6. https://doi.org/10.1016/j.ccell.2017.08.006
69. Noushmehr H, Weisenberger DJ, Diefes K et al (2010) Identification of a CpG Island Methylator Phenotype that Defines a Distinct Subgroup of Glioma. Cancer Cell 17:510–522. https://doi.org/10.1016/j.ccr.2010.03.017
70. Linehan WM et al (2016) Comprehensive Molecular Characterization of Papillary Renal-Cell Carcinoma. New England Journal of Medicine 374:135–145. https://doi.org/10.1056/NEJMoa1505917
71. Hoadley KA, Yau C, Hinoue T et al (2018) Cell-of-Origin Patterns Dominate the Molecular Classification of 10,000 Tumors from 33 Types of Cancer. Cell 173:291–304.e6. https://doi.org/10.1016/j.cell.2018.03.022
72. Liu Y, Sethi NS, Hinoue T et al (2018) Comparative Molecular Analysis of Gastrointestinal Adenocarcinomas. Cancer Cell 33:721–735.e8. https://doi.org/10.1016/j.ccell.2018.03.010
73. Mays-Hoopes LL, Brown A, Huang RC (1983) Methylation and rearrangement of mouse intracisternal a particle genes in development, aging, and myeloma. Molecular and Cellular Biology 3:1371–1380. https://doi.org/10.1128/MCB.3.8.1371

74. Ahuja N, Li Q, Mohan AL et al (1998) Aging and DNA Methylation in Colorectal Mucosa and Cancer. Cancer Res 58:5489–5494
75. Maegawa S, Hinkal G, Kim HS et al (2010) Widespread and tissue specific age-related DNA methylation changes in mice. Genome Res 20:332–340. https://doi.org/10.1101/gr.096826.109
76. Horvath S, Raj K (2018) DNA methylation-based biomarkers and the epigenetic clock theory of ageing. Nature Reviews Genetics 19:371–384. https://doi.org/10.1038/s41576-018-0004-3
77. Teschendorff AE, Menon U, Gentry-Maharaj A et al (2010) Age-dependent DNA methylation of genes that are suppressed in stem cells is a hallmark of cancer. Genome Res 20:440–446. https://doi.org/10.1101/gr.103606.109
78. Zhou W, Dinh HQ, Ramjan Z et al (2018) DNA methylation loss in late-replicating domains is linked to mitotic cell division. Nat Genet 50:591–602. https://doi.org/10.1038/s41588-018-0073-4
79. Bollati V, Baccarelli A, Hou L et al (2007) Changes in DNA Methylation Patterns in Subjects Exposed to Low-Dose Benzene. Cancer Res 67:876–880. https://doi.org/10.1158/0008-5472.CAN-06-2995
80. Alexandrov LB, Nik-Zainal S, Wedge DC et al (2013) Signatures of mutational processes in human cancer. Nature 500:415–421. https://doi.org/10.1038/nature12477

2 Considerations in the Design, Conduct, and Interpretation of Studies in Epigenetic Epidemiology

Karin B. Michels

Contents

K. B. Michels (✉)
Department of Epidemiology, Fielding School of Public Health, University of California, Los Angeles, CA, USA

Institute for Prevention and Cancer Epidemiology, Faculty of Medicine and Medical Center, University of Freiburg, Freiburg, Germany
e-mail: karin.michels@uniklinik-freiburg.de

K. B. Michels (ed.), *Epigenetic Epidemiology*,
https://doi.org/10.1007/978-3-030-94475-9_2

Abstract

Studies in epigenetic epidemiology may identify epigenetic aberrations associated with disease, link environmental and lifestyle factors to the epigenetic profile, or unveil epigenetic mechanisms underlying statistical relations between risk factors and disease outcomes. Epidemiologic studies provide the framework for identifying epigenetic biomarkers for disease risk or early detection of disease. Appropriate design considerations for studies in epigenetic epidemiology are imperative for their success. The tissue specificity of epigenetic marks represents a challenge in epigenetic epidemiology, and disease markers in easily accessible surrogate tissues are essential for large-scale population-based studies. Nested case-control studies using biospecimens collected prior to onset of disease provide appropriate data to identify epigenetic changes preceding disease. Selecting a representative study population with sufficiently large sample size and appropriate comparison group is crucial for the validity and reproducibility of the results. Challenges in epigenetic epidemiology studies include confounding and effect modification, and identifying epigenetic marks with sufficient systematic interindividual variation.

Abbreviations

EWAS	Epigenome-wide Association Studies
IGF 2	Insulin-like Growth Factor 2
LOI	Loss of Imprinting
NTD	Neural Tube Defect

2.1 The Objectives of Studies in Epigenetic Epidemiology

Epidemiology is primarily concerned with the frequency, distribution, and determinants of health and disease in humans [1]. Epidemiologic studies connect risk factors with disease outcomes based on distributions, often producing observations that inform basic research to identify mechanisms. A familiar example is the link between cigarette smoking and lung cancer [2], which guided basic research studies to classify nicotine as a carcinogen. Integrating epigenetics into an epidemiologic study recognizes the mechanistic link between a risk factor and disease risk [3] (Fig. 2.1). For instance, epidemiologic studies support an association between folate deficiency and neural tube defects (NTD) [4]. Including epigenetics in an epidemiologic study may shed light on the underlying mechanisms and identify mediators. Thus, in a case-control study including 48 induced abortions with NTD and 49 elective induced abortions without NTDs, the risk of NTDs increased with decreasing levels of LINE-1 methylation in brain tissue pointing toward epigenetic mechanisms underlying these associations [5]. Moreover, maternal serum folate

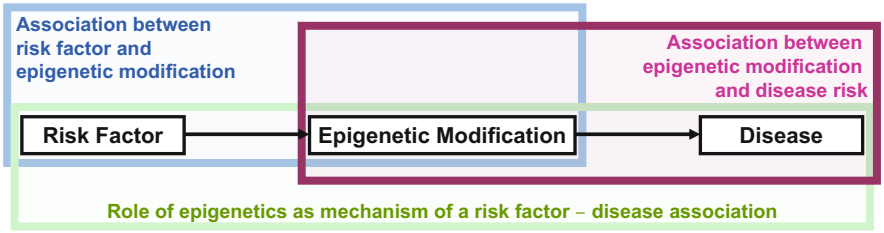

Fig. 2.1 Objectives of studies in epigenetic epidemiology

levels were lower in NTD cases than in controls and positively correlated with DNA methylation in fetal brain tissue, with a stronger correlation among cases [6].

The implications of epigenetic modifications in the risk of human disease can also be explored using the framework of an epidemiologic study (Fig. 2.1). Loss of imprinting (LOI) of insulin-like growth factor (*IGF*) 2, defined as biallelic expression of the normally monoallelically expressed gene in a parent-of-origin fashion, is an important etiologic factor in Beckwith-Wiedeman Syndrome [7] and Wilms' tumor [8] and was discovered by comparing the frequency of LOI of *IGF2* in children with and without these syndromes. We are only beginning to understand the role of epigenetic modifications in disease etiology; epigenetic epidemiology will provide the relevant methodologic underpinning to explore important correlations in human populations.

Epidemiologic studies can be used to explore whether aberrant DNA methylation, chromatin marks, or microRNAs may be candidates for biomarkers of disease risk or early detection of disease. To date, few epigenetic biomarkers with acceptable sensitivity and specificity for disease have been identified [9–11]. Tissue-specificity of epigenetic patterns, requirements for sophisticated lab methods (e.g., capturing cancer DNA from serum) and equipment, and heterogeneity in sample collection and laboratory routines are just some of the obstacles that have hindered progress in identifying such biomarkers and moving them toward clinical application [12]. Moreover, a disease is likely to be regulated by multiple epigenetic pathways.

Observational studies also allow investigation of the influence of environmental and lifestyle factors on the epigenome (Fig. 2.1). Several lifestyle factors including alcohol consumption and smoking affect DNA methylation, as discussed in more detail in Chap. 12. Insights into environmental influences on the epigenome may provide targets for disease prevention.

Understanding the role of epigenetic changes in the disease process will stimulate the development of targeted interventions to prevent and treat disease. Indeed, epigenetic drugs are used to target hematologic cancers. DNA methyltransferase inhibitors (which reverse acquired aberrant DNA methylation of tumor suppressor genes) have been approved in the USA and elsewhere on the basis of randomized clinical trials to treat patients with myelodysplastic syndrome [13] and acute myeloid leukemia, and histone deacetylase inhibitors are available for the treatment of cutaneous T-cell lymphoma [14]. More recent targets include activation of endogenously methylated sequences (e.g., cancer-testis antigens) [15]. To improve the

proportion of patients with therapeutics success with these therapies, treatments are increasingly combined [15]. Moreover, DNA methylation may also predict response to therapy and foster personalized medicine. For instance, promoter hypermethylation of the DNA repair protein MGMT is associated with a poor prognosis in various cancers because of the accumulation of mutations, however, these patients may respond well to alkylating drugs [16]. DNA methylation editing using the CRISPR technology opens new avenues in treatment, which can be evaluated in epidemiologic studies [17–19].

2.2 The Tissue Question

Due to its significant role in cell differentiation, one of the most profound characteristics of the epigenomic signature is its tissue specificity. Thus, the choice of the appropriate tissue is central to the success of an epigenetic epidemiologic study. Changes in DNA methylation or chromatin structure are often restricted to the target organ affected by disease. Thus, when studying the contribution of epigenetics to cancer, it is important to microdissect tumor tissue to ensure a uniform cancer cell population. Moreover, tissue specificity can make epidemiologic studies of aberrant epigenetic marks very difficult. Not only is it challenging to collect samples of difficult to obtain target tissue from a sufficient number of patients with illnesses, e.g., brain tissue from patients with Alzheimer's disease or liver tissue from patients with liver cancer, but it is nearly impossible to collect control tissues from healthy individuals to study the "normal" tissue-specific epigenome. When blood is used for epigenetic studies, cells should be optimally sorted. However, microdissection of tissue and sorting of blood cells are laborious and expensive; a practical biomarker should be detectable even in mixed cell populations.

These barriers (as well as the problem of interindividual confounding discussed below) have seduced cancer researchers into using tumor-adjacent tissue that has been histologically rated tumor free (from the same patient) for comparison. Of course, using control tissue samples from the same person avoids confounding by age or other interindividual differences. However, morphologically normal tissues adjacent to tumors harbor a number of genetic abnormalities [20–22], and epigenetic changes identified in cancer tissue have also been found in adjacent cancer-free tissue as far as 4 cm away from primary tumors [23–26]. This field-effect renders the use of adjacent tissues in the quest for epimutations obsolete, and potential DNA methylation markers of tumorigenesis may be missed due to the use of inappropriate control tissue [27]. However, when compared to tissue from healthy individuals, adjacent tissue may provide the opportunity for identification of early changes in DNA methylation which could serve as prediagnostic biomarkers of disease [28].

However, some tissue types are more accessible than others. For example, tissue harvested from invasive tumors in the intestinal wall of the colon displays characteristic epimutations when compared to colon mucosa from cancer-free patients [29]. Similarly, human skin biopsies show distinct methylation patterns according to sun exposure [30]. Conversely, collection of brain or heart muscle tissue may not be possible.

When obtaining tissues from a person with a particular disease, it is important to keep in mind that epigenetic changes either may have preceded or may be a result of the disease. Therefore, it is essential to precisely formulate the research question. If the focus is to identify an epigenetic mark as a biomarker for disease risk or to identify preclinical onset of disease, target tissues should be obtained prior to disease onset to preclude reverse causation. This approach is unfortunately prohibitive for most tissues, given the number of individuals from whom target tissues would have to be harvested while they are phenotypically healthy. Prospective sampling may be possible only for specimens that can be reasonably easily collected such as blood, saliva, buccal cells, skin cells, urine, and feces.

Aberrant DNA methylation or histone modifications in such tissues may serve as surrogate markers of disease risk for many illnesses. Surrogate tissues have been identified for some solid tumors [23, 31], and other chronic diseases including heart disease, asthma, and depression [32–34]. Statistical methods have been developed to improve prediction of locus-specific methylation in target tissue based on methylation marks in surrogate tissue [35]. Moreover, techniques to harvest circulating tumor DNA from the blood of cancer patients have been rapidly advanced over the past years [36–40] and bladder cancer DNA has been retrieved from urine sediments [41].

Importantly for studies in epigenetic epidemiology and epigenome-wide association studies (EWAS), DNA methylation displays exposure-dependent co-variability across different cell types [42].

2.3 Selecting the Epigenetic Mark to Study

The research question of interest generally determines the epigenetic mark to study. Generally, DNA methylation is a suitable marker for epidemiologic studies, because methyl groups are covalently bound to CpG dinucleotides and remain intact during routine DNA extraction, and cytosine methylation is fairly stable over the long term if samples are properly processed and stored [43]. Very few epidemiologic studies have incorporated histone modifications, since the chromatin structure requires an additional immunoprecipitation step in sample preparation, which generally rules out the use of biorepositories.

A profound understanding of the intraindividual variation of an epigenetic characteristic over time or the stochastic interindividual variations are essential for the successful design of epidemiologic studies. Epigenetic traits with large interindividual variation are the best candidates for epidemiologic studies, assuming that the majority of variation is systematic (and separates individuals with different characteristics such as differences in lifestyle or environmental factors or individuals with and without disease) rather than stochastic. Furthermore, for an epidemiologic study to detect interindividual variation in an epigenetic mark and link it to disease occurrence, interindividual variation has to exceed intraindividual variation. Some DNA methylation and genomic imprinting marks exhibit considerable interindividual variation and reasonable intraindividual stability over time [44], which makes them preferred markers for epidemiologic studies. However, a more detailed understanding of the DNA methylome is necessary to identify prime candidates.

In epidemiologic studies, the influence of environmental, nutritional, and lifestyle factors on the epigenome may be of interest. The focus will be on de novo methylation of unmethylated CpG islands in promoter regions. The stability of the methyl-cytosine bond makes active demethylation difficult to induce. While there is evidence of active demethylation in somatic cells [45], the mechanisms facilitating this process remain to be elucidated. Histone modifications are considered to be more volatile than CpG dinucleotide methylation and environmental factors may affect transcription also via this pathway. The world of histone modifications, however, is complex and includes methylation, acetylation, phosphorylation, and ubiquitination.

Transcription is likely governed by the interplay of DNA methylation, histone modifications, and miRNAs, hence capturing the state of all three markers simultaneously may be necessary for a more complete understanding of control mechanisms [46]. However, gene expression is not regulated by epigenetic features alone. Moreover, changes in DNA methylation or chromatin structure may not affect transcription levels. Unmethylated gene promoters (or enhancers) do not necessarily facilitate expression. The functional relevance of methylation of individual CpGs within or outside of a CpG island remains unclear. Similarly, whether a difference in methylation of a few percent has functional consequences is likely gene-specific.

The candidate-gene approach targets DNA methylation or histone modification at certain candidate genes suspected or known to play a role in a specific disease process. To identify epigenetic changes associated with initiation or promotion of a disease or resulting from a particular exposure, e.g., folic acid supplementation or exposure to a chemical like bisphenol A, a genome-wide approach such as DNA methylation microarrays or ChIP-on-chip may be preferable to the candidate-gene approach in order to allow new discoveries (discussed in more detail in Chap. 3). However, genome-wide approaches are generally less sensitive than the assays used for candidate-gene approaches and rely heavily on sophisticated bioinformatics methods (see Chap. 4 for further details), which complicates interpretation of results. Various genome-wide approaches are currently employed, and microarray-based methods, in particular, the Illumina Infinium arrays for DNA methylation are widely used [47]. With the rapid advance in sequencing technologies, sequencing the epigenome as part of an epidemiologic study may also be a realistic prospect. Reduced representation bisulfite sequencing (RRBS) provides an efficient approach to study genome-wide DNA methylation at single nucleotide resolution by enriching for areas with a high CpG content reducing the number of nucleotides sequenced to about 3% of the genome [48] [49]. For EWAS, samples from prospective cohorts are required to ensure the temporal sequence of epimutations and disease incidence.

2.4 Study Designs in Epigenetic Epidemiology

A number of study designs can be employed in epigenetic epidemiology; the choice of the appropriate design depends on the research question (Table 2.1).

Table 2.1 Study designs used in epigenetic epidemiology and their applications

Study design	Application
Cross-sectional study	Prevalence of an epigenetic mark in a well-defined population subgroup
Retrospective case-control study	Permanent epigenetic marks among individuals with and without disease
Cohort study	Epigenetic mechanisms underlying a risk factor-disease association Basis for a nested case-control study
Nested (prospective) case-control study	Epigenetic marks predisposing to disease Biomarkers for disease risk Biomarkers for early disease detection
Intervention study Crossover design Parallel group design	Effect of interventions on epigenetic pattern; effect of epigenetic therapies on disease
Epigenome-wide association studies	Epigenomic variation on a large-scale and biologic traits (e.g., disease)
Family-based study	Transgenerational inheritance of epigenetic traits
Birth cohort	Influence of preconceptional and prenatal factors on establishment of the epigenome

2.4.1 Cross-Sectional Study

In a cross-sectional study, all factors of interest are assessed at one time point. For example, the proportion of individuals with a methylated CpG in a specific gene promoter in a population defined by a special characteristic, such as female smokers age 20–25 years, or the prevalence of loss of imprinting of a particular gene, say IGF-2, in newborns, can be studied most easily with a cross-sectional design. Similarly, comparing the degree of global DNA methylation in two groups, e.g., among Caucasians age 60–65 years, males vs. females, can be accomplished effectively with a cross-sectional design.

2.4.2 Retrospective Case-Control Study

In a case-control study, individuals with a disease and appropriately selected individuals free of the disease are sampled from the same source population. To study epigenetic variation, relevant biospecimens would be obtained and DNA methylation or histone modification assessed among cases and controls. In the context of epigenetics, this type of design bears similarities to a cross-sectional study, except for the particular control selection, which must be independent of the epigenetic state. The purpose of the controls is to estimate the epigenetic state of cases had they not contracted the disease under study; if the controls are not properly chosen, selection bias results. In neither the cross-sectional study nor the retrospective case-control study can it be determined whether the particular epigenetic signature assessed at the time of case and control selection preceded (and possibly caused) the disease among cases or whether it may be a consequence of (and possibly have been caused by) the disease.

Hence, retrospective case-control studies have limited value for the study of epigenetic marks and have to be interpreted with caution.

2.4.3 Cohort Study

In a cohort, healthy individuals are recruited to participate in a longitudinal study over a certain period of follow-up time (weeks, months, years). At baseline, easy-to-obtain biospecimens are collected from all participants and stored. During follow-up, additional biospecimens may be collected. Such a biorepository provides the opportunity to study changes in DNA methylation or other epigenetic marks over time. Moreover, the unambiguous temporality allows identification of epigenetic marks that existed prior to clinical onset of disease. During follow-up, disease outcomes are recorded. Cohort studies are usually large and often include many thousands of participants. Because it is not cost-effective to analyze the samples from all participants obtained at baseline and during follow-up, a nested case-control study is usually embedded in the cohort to study epigenetic marks predisposing to disease. In addition, cohort studies provide the opportunity to study epigenetic mechanisms underlying a risk factor-disease association.

2.4.4 Nested Case-Control Study

A nested or prospective case-control study is embedded in a cohort study. All individuals who develop the disease of interest at any time during follow-up are selected, and appropriate controls (often 2 controls per case) are selected from those who remained free of the disease throughout follow-up. Since biospecimens from both cases and controls were obtained prior to the diagnosis of disease (and have been stored since then), this study design permits correlation of the prediagnostic epigenetic status with disease outcome, i.e., it is clear that the epigenetic state reflected in the biospecimens preceded the diagnosis of disease and is not influenced by (phenotypically manifest) disease. Though the nested case-control study is a cost-effective study design, of course, prospective cohorts are generally expensive to maintain.

2.4.5 Intervention Studies

The effect of interventions such as folate and other supplements that affect the one-carbon metabolism on the epigenome or the effect of demethylating agents on disease progression can be studied in (randomized) intervention studies (crossover or parallel group design). In a crossover study, the effect of supplements on the epigenetic pattern can be explored by comparing the profile before and after each supplement use in the same person, while the sequence of supplements can be randomized. Additionally, the effect of the different supplements can be compared. This approach reduces confounding. In a parallel group trial, a demethylating agent is randomly assigned to treat half of the patients in the study, while the other half

receives a different dose, a different drug, or even a placebo. Clinical outcomes are recorded in both groups.

2.4.6 EWAS

EWAS examine the association between interindividual epigenetic variation and human disease and other biologic traits on a large scale [50]. EWAS utilize array- and sequencing-based profiling technologies targeting large portions of the epigenome. While EWAS can be cohort, case-control studies, or cross-sectional studies, a prospective design is preferable to ensure that the epigenetic variation is a likely cause rather than a consequence of the disease. Appropriately stored samples may allow imbedding an EWAS into an ongoing cohort thus increasing efficiency. Numerous EWAS have been conducted and summarized in an EWAS Atlas [51, 52].

2.4.7 Family-Based Studies

Transgenerational inheritances of epigenetic traits have to be studied using triads of mother, father, and child. For an epigenetic mark to be inherited transgenerationally, the putatively inherited epigenetic change must be present in both the contributed gametes and the offspring soma: to demonstrate transgenerational inheritance along the male line two offspring generations are necessary, for demonstration along the female line three offspring generations are required [53]. To date, transgenerational inheritance of epigenetic marks in humans has not been established.

2.4.8 Birth Cohorts

In a birth cohort, preconceptional and prenatal exposures can be assessed and related to DNA methylation, imprinting profiles, and chromatin states of the offspring at birth in tissues that can be easily obtained such as cord blood, cord, placenta, and saliva. Additional follow-up of the birth cohort permits tracking of developmental stages, anthropometric variables, disease outcomes, and changes in epigenetic marks over time.

Thus, if the goal is to identify epigenetic changes that may predispose to disease, a nested case-control study would be the most appropriate design. Similarly, epigenetic biomarkers of susceptibility would be best studied using a nested case-control design. A retrospective case-control study may provide some useful information about an epigenetic mark of disease, but epimutations may actually be a consequence of the disease. The effect of a lifestyle factor such as nutrition or alcohol consumption on the epigenetic profile can be studied in either a cohort study or an intervention study. Epigenetics as a causal link connecting a risk factor and a disease requires samples from a cohort study, collected after the risk factor is manifest and before disease occurrence.

2.5 Other Methodologic Considerations in Epigenetic Epidemiology

Besides identifying the appropriate tissue, settling on the epigenetic marks to study, and choosing a suitable study design, other considerations are essential to ensure the success of an epigenetic epidemiology study (see Box 2.1). Among them are the study population and its characteristics.

Box 2.1 Some Important Considerations in Planning and Interpreting a Study in Epigenetic Epidemiology

- Cell-type heterogeneity
- Target tissue or surrogate tissue
- Restriction to a well-defined representative population subgroup
- Sufficient sample size
- Effect modification
- Adjustment for confounding
- Time sequence

2.5.1 Choice of the Study Population

As only a sample of the population of interest can be studied, the characteristics of this sample determine the extent to which the results can be generalized to the population of interest. A convenience sample for a birth cohort may focus on women with uncomplicated deliveries because of the logistical ease of obtaining the biospecimens. However, the epigenetic profile among the newborns may differ from those of infants born prematurely or those whose mothers suffered from preeclampsia. Similarly, if individuals with certain characteristics are over- or underrepresented in the sample drawn, the result, e.g., the prevalence of hypermethylation of a particular tumor suppressor gene promoter, may not be representative of the prevalence in the underlying population of interest from which the sample was drawn.

2.5.2 Sample Size

One of the most important—and common—limitations of basic science research is small sample size. A sufficiently large sample is a fundamental requirement of a high-quality study in epigenetic epidemiology: it increases the likelihood of a valid study result and is necessary to achieve adequate precision of the findings. Moderate to modest differences in epigenetic patterns between two groups may be important but cannot be detected in a small sample. Some factors may exert only modest changes in DNA methylation, e.g., diet or dietary supplements. A sufficiently large sample is necessary to detect such differences. Power calculations conducted during the planning phase of a study determine the number of participants necessary to detect a certain difference that may be of interest. Even if the expected difference is

large, the number of individuals studied has to be sufficient to ensure precision of the results reflected in a narrow confidence interval around the measure of interest.

2.5.3 Means Versus Outliers

In epidemiologic studies mostly means of groups are compared. Outliers or influential values might even be excluded from the analyses. Thus, we might compare the mean level of methylation in a group of individuals regularly taking folic acid supplements with the mean methylation level of a group of individuals who do not take folic acid supplements. However, individuals with extreme or unusual methylation values might be of particular interest and should not be excluded. On the contrary, it might be of interest to thoroughly explore their characteristics.

2.5.4 Effect Modification

Epigenetic marks are known not only to be tissue-specific but also to vary by race and ethnicity, sex, and age [54, 55]. If the prevalence of a particular epigenetic characteristic is of interest, e.g., the prevalence of LOI of IGF2 among healthy individuals, such frequency is best described in the context of a well-defined substratum of the population, e.g., Caucasian females age 30–40 years. Inferences about the prevalence in other population subgroups, e.g., Hispanic males age 60–70 years, cannot be made unless the prevalence is assessed specifically in this population.

When the goal is to study the relevance of an environmental factor for the epigenetic profile (e.g., the impact of smoking on DNA methylation) or the association between an epigenetic trait and a disease endpoint (e.g., a certain histone modification and the incidence of asthma), it is important to consider whether these associations may differ by sex, age, race or ethnicity, or other factors. Is there reason to suspect smoking may induce different methylation changes in Hispanics than in Asians? If the answer is yes or if there is sufficient uncertainty, the heterogeneity of this association ("effect modification") can be studied across ethnicity using a statistical test for interaction. If this test yields insignificant differences, the association between smoking and DNA methylation does not differ substantially between the different ethnic subgroups considered and can be assessed in a population with mixed ethnicity. Beware, however, that the strata of different ethnicities are not too small when applying the test for interaction. Similarly, we can test whether smoking affects DNA methylation differently in men and women. If significant effect modification by sex is identified, the association between smoking and DNA methylation should be reported separately for males and females. Each of the population subsets created by stratification needs to be sufficiently large to generate valid and reliable stratum-specific estimates.

2.5.5 Confounding

A distinctly different consideration from the above-described effect modification is whether a factor may influence the strength or even direction of an association. When evaluating whether LINE-1 hypomethylation in mammary tissue is more common among women with or without breast cancer, the distribution of age may distort the findings. Women with breast cancer may be older than women free of breast cancer, since risk increases with age. However, methylation of repetitive elements decreases with age. Unequal distribution of age in the subgroups may create a spurious or exaggerated association between LINE-1 hypomethylation and breast cancer that vanishes or at least diminishes after appropriate statistical adjustment for age differences. Confounding can be detected when proper statistical adjustment changes the association of interest, i.e., the effect estimates with and without statistical adjustment for the confounder differ. A factor is a confounder if it is associated with both factors of interest, the exposure or risk factor and the disease of interest; in the above example, age is associated with both LINE-1 methylation and breast cancer. Confounding is one of the most important threats to the validity of an epidemiologic study.

2.5.6 Misclassification

As the field of epigenetics matures, so do its methods. Methylation microarrays now allow us to assess the methylation status of an increasing number of individual CpGs within and outside of CpG islands. The assessment of loss of imprinting with allele-specific expression assays using pyrosequencing has become more quantitative; previous methods relied on gel electrophoresis and radioactive labeling. Each of these improvements in technology decreases random misclassification of the epigenetic state and improves precision. Misclassification will be further reduced by the next generation of genome-wide CpG dinucleotide methylation assessment. Correspondingly, increasingly sophisticated bioinformatics tools allow us to distinguish between true signals and noise.

2.6 Evaluating and Interpreting Results of Epigenetic Epidemiology Studies

The establishment of a benchmark of a “normal” or “disease-free” state is fundamental in genetics and epigenetics. Any departure from this state may be linked to diseases or phenotypes of interest. If an epigenetic change is already manifest prior to diagnosis of a disease, it may be a candidate for an early detection biomarker [3]. If it is present prior to disease development, it may be useful as a biomarker of disease risk. As part of the National Institutes of Health Roadmap Epigenomics Consortium and the ENDODE project, human reference epigenomes have been established for numerous tissues and cell types [56, 57]. Genome-wide, single-base resolution of methylated cytosines have also been presented for human

embryonic stem cells and fetal fibroblasts [58]. Furthermore, reference epigenomes have been derived for disease states [59]. Additional reference epigenomes are being decoded various consortia in the International Human Epigenome Consortium, including Blueprint, the Canadian Epigenetics, Environment and Health Consortium, the Japan Agency for Medical Research and Development Core Research for Evolutional Science and Technology, and the Hong Kong Epigenomics Project.

In epidemiology, comparing two groups (i.e., individuals with and without a certain characteristic of interest) is a substitute for the ideal (but impossible) set of information: data on the same individual measured at the same time twice, once with the characteristic of interest and once without. For example, two sets of data on the same woman, once with regular consumption of one glass of wine per day and once without alcohol consumption, would allow inference of whether her subsequent breast cancer would also have arisen if no alcohol was consumed, since all other factors are identical and no confounding was possible. Such counterfactual information would permit causal inference about the role of alcohol consumption in breast cancer etiology. Comparing the fate of two individuals, one with the characteristic and one without, even in the context of a randomized controlled trial, cannot exclude differences between these individuals besides the trait of interest.

Similarly, in epigenetic epidemiology information on the same person is not available twice: with and without a characteristic to see its bias-free effect on the epigenetic mark of interest, or with and without the epigenetic trait to evaluate its effect on disease risk. In order to reduce the risk of confounding, an intraindividual crossover study can be conducted to explore the effect of an intervention, such as supplementation with methyl group donors.

While DNA methylation changes with age (discussed in more detail in Chap. 10), age-related changes occur slowly, and the degree of change differs for individual genes and repetitive elements. Hence, intraindividual variation in cytosine methylation is limited. Moreover, genome-wide methylation studies have generally revealed statistically significant interindividual differences in only a fraction of genes studied, suggesting that DNA methylation is reasonably tightly controlled [60–63]. More profound differences are generally restricted to the comparison of cancerous and normal tissues [31, 64–66]. However, CpG clusters have been identified with high inter-individual epigenetic variation termed Variably Methylated Regions (VMRs) [67]. Environmentally responsive VMRs were enriched for imprinted regions suggesting their sensitivity to environmental conditions [67].

In the literature, most methylation differences between two groups of individuals with and without a certain characteristic are small and amount to less than 5%. This raises the question of the functional implication of theoretical differences in LINE-1 methylation of, let us say, 80% vs. 82% or in a gene promoter CpG island of 42% vs. 46%. While it is advisable to examine the expression profile of the gene studied, methylation and expression status of a gene have been found to be less closely correlated than previously assumed [68–70]. Nevertheless, downstream effects of methylation are possible but may be difficult to determine. Moreover, differences in methylation are often reported for individual CpGs, and the functional relevance of such differences is difficult to assess. Neighboring CpGs seem to

influence each other's DNA methylation states [60] and at least in cancer tissues, entire CpG islands rather than individual CpGs are aberrantly methylated [71].

The lack of precise definitions and standardization of epigenetic phenomena further complicates the interpretation of results of studies in epigenetic epidemiology. While hyper- and hypomethylation are terms commonly used in evaluating the results of global or locus-specific methylation experiments in populations, cut-off values applied in the literature vary considerably: above and below the median [72, 73], above and below the 25th and 75th percentile [5], statistically significantly higher or lower than the control [74], and percent methylated reference with a differently chosen cut-off for each gene studied [75]. Such a lack of standardization makes results from different studies difficult to compare. Similarly, there is no distinct definition of loss of imprinting, even in individuals. Often the normally silenced allele is expressed but at a considerably lower level than the normally expressed allele. Yet there is no set threshold level that defines loss of imprinting, leaving it to the individual investigator to make the call. A more standardized approach may aid comparability across studies.

There is increasing evidence that DNA methylation is correlated with DNA sequence and single nucleotide polymorphisms, suggesting that DNA composition predisposes CpG islands to DNA methylation [76, 77]. For a comprehensive understanding of the structural and functional implications of epimutations, it may be necessary to examine genetics, methylation status, histone modification, expression, and microRNAs in the loci studied.

2.7 Conclusion

The human epigenome in its complexity offers a broad target for innovative strategies in the prevention and treatment of disease. The malleable nature of the epigenetic signature, in particular, DNA methylation editing, provides a unique opportunity to design effective interventions. Identifying epigenetic marks of functional relevance is all the more important. Epigenetic epidemiology provides the framework to study the relevance of epigenetic variation in human health and disease on a population level. EWAS target the epigenome on a comprehensive level utilizing profiling technologies. Multiple challenges including tissue heterogeneity, surrogate tissue, uncertainties in time sequence, insufficient statistical power, and confounding may compromise interpretability of results. Careful considerations for the appropriate design and analysis of epigenetic epidemiology studies are essential to ensure validity. Recent advances in establishing reference epigenomes, in identifying epigenetic marks with reasonable interindividual variability, and in statistical methods will sharpen the approach.

References

1. Gordis L (2018) Epidemiology, 6th edn. Saunders, Elsevier, Philadelphia
2. Doll R, Hill AB (1950) Smoking and carcinoma of the lung; preliminary report. Br Med J 2: 739–748. https://doi.org/10.1136/bmj.2.4682.739
3. Michels KB (2010) The promises and challenges of epigenetic epidemiology. Exp Gerontol 45: 297–301. https://doi.org/10.1016/j.exger.2009.12.011
4. Daly LE, Kirke PN, Molloy A et al (1995) Folate levels and neural tube defects. Implications for prevention. JAMA 274:1698–1702. https://doi.org/10.1001/jama.1995.03530210052030
5. Wang L, Wang F, Guan J et al (2010) Relation between hypomethylation of long interspersed nucleotide elements and risk of neural tube defects. Am J Clin Nutr 91:1359–1367. https://doi.org/10.3945/ajcn.2009.28858
6. Chang H, Zhang T, Zhang Z et al (2011) Tissue-specific distribution of aberrant DNA methylation associated with maternal low-folate status in human neural tube defects. J Nutr Biochem 22:1172–1177. https://doi.org/10.1016/j.jnutbio.2010.10.003
7. Ohlsson R, Nyström A, Pfeifer-Ohlsson S et al (1993) IGF2 is parentally imprinted during human embryogenesis and in the Beckwith-Wiedemann syndrome. Nat Genet 4:94–97. https://doi.org/10.1038/ng0593-94
8. Ogawa O, Eccles MR, Szeto J et al (1993) Relaxation of insulin-like growth factor II gene imprinting implicated in Wilms' tumour. Nature 362:749–751. https://doi.org/10.1038/362749a0
9. Hegi ME, Diserens A-C, Gorlia T et al (2005) MGMT gene silencing and benefit from temozolomide in glioblastoma. N Engl J Med 352:997–1003. https://doi.org/10.1056/NEJMoa043331
10. Lofton-Day C, Model F, Devos T et al (2008) DNA methylation biomarkers for blood-based colorectal cancer screening. Clin Chem 54:414–423. https://doi.org/10.1373/clinchem.2007.095992
11. Chung W, Bondaruk J, Jelinek J et al (2011) Detection of bladder cancer using novel DNA methylation biomarkers in urine sediments. Cancer Epidemiol Biomark Prev 20:1483–1491. https://doi.org/10.1158/1055-9965.EPI-11-0067
12. Kagan J, Srivastava S, Barker PE et al (2007) Towards clinical application of methylated DNA sequences as cancer biomarkers: a joint NCI's EDRN and NIST workshop on standards, methods, assays, reagents and tools. Cancer Res 67:4545–4549. https://doi.org/10.1158/0008-5472.CAN-06-2888
13. Silverman LR, Demakos EP, Peterson BL et al (2002) Randomized controlled trial of azacitidine in patients with the myelodysplastic syndrome: a study of the cancer and leukemia group B. J Clin Oncol 20:2429–2440. https://doi.org/10.1200/JCO.2002.04.117
14. Piekarz RL, Frye R, Turner M et al (2009) Phase II multi-institutional trial of the histone deacetylase inhibitor romidepsin as monotherapy for patients with cutaneous T-cell lymphoma. J Clin Oncol 27:5410–5417. https://doi.org/10.1200/JCO.2008.21.6150
15. Jones PA, Ohtani H, Chakravarthy A, De Carvalho DD (2019) Epigenetic therapy in immune-oncology. Nat Rev Cancer 19:151–161. https://doi.org/10.1038/s41568-019-0109-9
16. Jacinto FV, Esteller M (2007) MGMT hypermethylation: a prognostic foe, a predictive friend. DNA Repair (Amst) 6:1155–1160. https://doi.org/10.1016/j.dnarep.2007.03.013
17. Kang JG, Park JS, Ko J-H, Kim Y-S (2019) Regulation of gene expression by altered promoter methylation using a CRISPR/Cas9-mediated epigenetic editing system. Sci Rep 9:11960. https://doi.org/10.1038/s41598-019-48130-3
18. Devesa-Guerra I, Morales-Ruiz T, Pérez-Roldán J et al (2020) DNA methylation editing by CRISPR-guided excision of 5-Methylcytosine. J Mol Biol 432:2204–2216. https://doi.org/10.1016/j.jmb.2020.02.007
19. Katayama S, Shiraishi K, Gorai N, Andou M (2021) A CRISPR/Cas9-based method for targeted DNA methylation enables cancer initiation in B lymphocytes. Adv Genet 2:e10040. https://doi.org/10.1002/ggn2.10040

20. Deng G, Lu Y, Zlotnikov G et al (1996) Loss of heterozygosity in normal tissue adjacent to breast carcinomas. Science 274:2057–2059. https://doi.org/10.1126/science.274.5295.2057
21. Ellsworth DL, Ellsworth RE, Love B et al (2004) Genomic patterns of allelic imbalance in disease free tissue adjacent to primary breast carcinomas. Breast Cancer Res Treat 88:131–139. https://doi.org/10.1007/s10549-004-1424-7
22. Kurose K, Hoshaw-Woodard S, Adeyinka A et al (2001) Genetic model of multi-step breast carcinogenesis involving the epithelium and stroma: clues to tumour-microenvironment interactions. Hum Mol Genet 10:1907–1913. https://doi.org/10.1093/hmg/10.18.1907
23. Cui H, Cruz-Correa M, Giardiello FM et al (2003) Loss of IGF2 imprinting: a potential marker of colorectal cancer risk. Science 299:1753–1755. https://doi.org/10.1126/science.1080902
24. Shen L, Kondo Y, Rosner GL et al (2005) MGMT promoter methylation and field defect in sporadic colorectal cancer. J Natl Cancer Inst 97:1330–1338. https://doi.org/10.1093/jnci/dji275
25. Yan PS, Venkataramu C, Ibrahim A et al (2006) Mapping geographic zones of cancer risk with epigenetic biomarkers in normal breast tissue. Clin Cancer Res 12:6626–6636. https://doi.org/10.1158/1078-0432.CCR-06-0467
26. Troester MA, Hoadley KA, D'Arcy M et al (2016) DNA defects, epigenetics, and gene expression in cancer-adjacent breast: a study from the cancer genome atlas. NPJ Breast Cancer 2:16007. https://doi.org/10.1038/npjbcancer.2016.7
27. Riazalhosseini Y, Hoheisel JD (2008) Do we use the appropriate controls for the identification of informative methylation markers for early cancer detection? Genome Biol 9:405. https://doi.org/10.1186/gb-2008-9-11-405
28. Gao Y, Widschwendter M, Teschendorff AE (2018) DNA methylation patterns in Normal tissue correlate more strongly with breast cancer status than copy-number variants. EBioMedicine 31:243–252. https://doi.org/10.1016/j.ebiom.2018.04.025
29. Ahuja N, Li Q, Mohan AL et al (1998) Aging and DNA methylation in colorectal mucosa and cancer. Cancer Res 58:5489–5494
30. Grönniger E, Weber B, Heil O et al (2010) Aging and chronic sun exposure cause distinct epigenetic changes in human skin. PLoS Genet 6:e1000971. https://doi.org/10.1371/journal.pgen.1000971
31. Marsit CJ, Koestler DC, Christensen BC et al (2011) DNA methylation array analysis identifies profiles of blood-derived DNA methylation associated with bladder cancer. J Clin Oncol 29:1133–1139. https://doi.org/10.1200/JCO.2010.31.3577
32. Kim M, Long TI, Arakawa K et al (2010) DNA methylation as a biomarker for cardiovascular disease risk. PLoS One 5:e9692. https://doi.org/10.1371/journal.pone.0009692
33. Lin P-I, Shu H, Mersha TB (2020) Comparing DNA methylation profiles across different tissues associated with the diagnosis of pediatric asthma. Sci Rep 10:151. https://doi.org/10.1038/s41598-019-56310-4
34. Uddin M, Koenen KC, Aiello AE et al (2011) Epigenetic and inflammatory marker profiles associated with depression in a community-based epidemiologic sample. Psychol Med 41:997–1007. https://doi.org/10.1017/S0033291710001674
35. Ma B, Wilker EH, Willis-Owen SAG et al (2014) Predicting DNA methylation level across human tissues. Nucleic Acids Res 42:3515–3528. https://doi.org/10.1093/nar/gkt1380
36. Bettegowda C, Sausen M, Leary RJ et al (2014) Detection of circulating tumor DNA in early- and late-stage human malignancies. Sci Transl Med 6:224ra24. https://doi.org/10.1126/scitranslmed.3007094
37. Francis G, Stein S (2015) Circulating cell-free tumour DNA in the Management of Cancer. Int J Mol Sci 16:14122–14142. https://doi.org/10.3390/ijms160614122
38. Wan JCM, Massie C, Garcia-Corbacho J et al (2017) Liquid biopsies come of age: towards implementation of circulating tumour DNA. Nat Rev Cancer 17:223–238. https://doi.org/10.1038/nrc.2017.7
39. Shohdy KS, West HJ (2020) Circulating tumor DNA testing-liquid biopsy of a cancer. JAMA Oncol 6:792. https://doi.org/10.1001/jamaoncol.2020.0346

40. Adashek JJ, Janku F, Kurzrock R (2021) Signed in blood: circulating tumor DNA in cancer diagnosis. Tx Screen Cancers (Basel) 13:3600. https://doi.org/10.3390/cancers13143600
41. Friedrich MG, Weisenberger DJ, Cheng JC et al (2004) Detection of methylated apoptosis-associated genes in urine sediments of bladder cancer patients. Clin Cancer Res 10:7457–7465. https://doi.org/10.1158/1078-0432.CCR-04-0930
42. Hannon E, Mansell G, Walker E et al (2021) Assessing the co-variability of DNA methylation across peripheral cells and tissues: implications for the interpretation of findings in epigenetic epidemiology. PLoS Genet 17:e1009443. https://doi.org/10.1371/journal.pgen.1009443
43. Waterland RA, Michels KB (2007) Epigenetic epidemiology of the developmental origins hypothesis. Annu Rev Nutr 27:363–388. https://doi.org/10.1146/annurev.nutr.27.061406.093705
44. Talens RP, Boomsma DI, Tobi EW et al (2010) Variation, patterns, and temporal stability of DNA methylation: considerations for epigenetic epidemiology. FASEB J 24:3135–3144. https://doi.org/10.1096/fj.09-150490
45. Ma DK, Jang M-H, Guo JU et al (2009) Neuronal activity-induced Gadd45b promotes epigenetic DNA demethylation and adult neurogenesis. Science 323:1074–1077. https://doi.org/10.1126/science.1166859
46. Bernstein BE, Meissner A, Lander ES (2007) The mammalian epigenome. Cell 128:669–681. https://doi.org/10.1016/j.cell.2007.01.033
47. Laird PW (2010) Principles and challenges of genomewide DNA methylation analysis. Nat Rev Genet 11:191–203. https://doi.org/10.1038/nrg2732
48. Carmona JJ, Accomando WP, Binder AM et al (2017) Empirical comparison of reduced representation bisulfite sequencing and Infinium BeadChip reproducibility and coverage of DNA methylation in humans. NPJ Genom Med 2:13. https://doi.org/10.1038/s41525-017-0012-9
49. Accomando WP, Michels KB (2018) Multiplexed reduced representation bisulfite sequencing with magnetic bead fragment size selection. Methods Mol Biol 1708:137–159. https://doi.org/10.1007/978-1-4939-7481-8_8
50. Rakyan VK, Down TA, Balding DJ, Beck S (2011) Epigenome-wide association studies for common human diseases. Nat Rev Genet 12:529–541. https://doi.org/10.1038/nrg3000
51. Michels KB, Binder AM (2018) Considerations for design and analysis of DNA methylation studies. Methods Mol Biol 1708:31–46. https://doi.org/10.1007/978-1-4939-7481-8_2
52. Li M, Zou D, Li Z et al (2019) EWAS atlas: a curated knowledgebase of epigenome-wide association studies. Nucleic Acids Res 47:D983–D988. https://doi.org/10.1093/nar/gky1027
53. van Otterdijk SD, Michels KB (2016) Transgenerational epigenetic inheritance in mammals: how good is the evidence? FASEB J 30:2457–2465. https://doi.org/10.1096/fj.201500083
54. Issa J-P (2003) Age-related epigenetic changes and the immune system. Clin Immunol 109:103–108. https://doi.org/10.1016/s1521-6616(03)00203-1
55. Zhang FF, Cardarelli R, Carroll J et al (2011) Significant differences in global genomic DNA methylation by gender and race/ethnicity in peripheral blood. Epigenetics 6:623–629. https://doi.org/10.4161/epi.6.5.15335
56. Roadmap Epigenomics Consortium, Kundaje A, Meuleman W et al (2015) Integrative analysis of 111 reference human epigenomes. Nature 518:317–330. https://doi.org/10.1038/nature14248
57. https://www.encodeproject.org/reference-epigenome-matrix/?type=Experiment&related_series.@type=ReferenceEpigenome&replicates.library.biosample.donor.organism.scientific_name=Mus+musculus
58. Lister R, Pelizzola M, Dowen RH et al (2009) Human DNA methylomes at base resolution show widespread epigenomic differences. Nature 462:315–322. https://doi.org/10.1038/nature08514
59. Beekman R, Chapaprieta V, Russiñol N et al (2018) The reference epigenome and regulatory chromatin landscape of chronic lymphocytic leukemia. Nat Med 24:868–880. https://doi.org/10.1038/s41591-018-0028-4
60. Bock C, Walter J, Paulsen M, Lengauer T (2008) Inter-individual variation of DNA methylation and its implications for large-scale epigenome mapping. Nucleic Acids Res 36:e55. https://doi.org/10.1093/nar/gkn122

61. Christensen BC, Houseman EA, Marsit CJ et al (2009) Aging and environmental exposures alter tissue-specific DNA methylation dependent upon CpG island context. PLoS Genet 5: e1000602. https://doi.org/10.1371/journal.pgen.1000602
62. Langevin SM, Houseman EA, Christensen BC et al (2011) The influence of aging, environmental exposures and local sequence features on the variation of DNA methylation in blood. Epigenetics 6:908–919. https://doi.org/10.4161/epi.6.7.16431
63. Dere E, Huse S, Hwang K et al (2016) Intra- and inter-individual differences in human sperm DNA methylation. Andrology 4:832–842. https://doi.org/10.1111/andr.12170
64. Marsit CJ, Christensen BC, Houseman EA et al (2009) Epigenetic profiling reveals etiologically distinct patterns of DNA methylation in head and neck squamous cell carcinoma. Carcinogenesis 30:416–422. https://doi.org/10.1093/carcin/bgp006
65. Teschendorff AE, Gao Y, Jones A et al (2016) DNA methylation outliers in normal breast tissue identify field defects that are enriched in cancer. Nat Commun 7:10478. https://doi.org/10.1038/ncomms10478
66. de Almeida BP, Apolónio JD, Binnie A, Castelo-Branco P (2019) Roadmap of DNA methylation in breast cancer identifies novel prognostic biomarkers. BMC Cancer 19:219. https://doi.org/10.1186/s12885-019-5403-0
67. Garg P, Joshi RS, Watson C, Sharp AJ (2018) A survey of inter-individual variation in DNA methylation identifies environmentally responsive co-regulated networks of epigenetic variation in the human genome. PLoS Genet 14:e1007707. https://doi.org/10.1371/journal.pgen.1007707
68. Gort EH, Suijkerbuijk KPM, Roothaan SM et al (2008) Methylation of the TWIST1 promoter, TWIST1 mRNA levels, and immunohistochemical expression of TWIST1 in breast cancer. Cancer Epidemiol Biomark Prev 17:3325–3330. https://doi.org/10.1158/1055-9965.EPI-08-0472
69. Diplas AI, Lambertini L, Lee M-J et al (2009) Differential expression of imprinted genes in normal and IUGR human placentas. Epigenetics 4:235–240. https://doi.org/10.4161/epi.9019
70. Fourkala E-O, Hauser-Kronberger C, Apostolidou S et al (2010) DNA methylation of polycomb group target genes in cores taken from breast cancer Centre and periphery. Breast Cancer Res Treat 120:345–355. https://doi.org/10.1007/s10549-009-0384-3
71. Jones PA, Baylin SB (2002) The fundamental role of epigenetic events in cancer. Nat Rev Genet 3:415–428. https://doi.org/10.1038/nrg816
72. Zhang FF, Morabia A, Carroll J et al (2011) Dietary patterns are associated with levels of global genomic DNA methylation in a cancer-free population. J Nutr 141:1165–1171. https://doi.org/10.3945/jn.110.134536
73. Zhu Z-Z, Sparrow D, Hou L et al (2011) Repetitive element hypomethylation in blood leukocyte DNA and cancer incidence, prevalence, and mortality in elderly individuals: the normative aging study. Cancer Causes Control 22:437–447. https://doi.org/10.1007/s10552-010-9715-2
74. Yoshida T, Yamashita S, Takamura-Enya T et al (2011) Alu and Satα hypomethylation in helicobacter pylori-infected gastric mucosae. Int J Cancer 128:33–39. https://doi.org/10.1002/ijc.25534
75. Saito K, Kawakami K, Matsumoto I et al (2010) Long interspersed nuclear element 1 hypomethylation is a marker of poor prognosis in stage IA non-small cell lung cancer. Clin Cancer Res 16:2418–2426. https://doi.org/10.1158/1078-0432.CCR-09-2819
76. Villicaña S, Bell JT (2021) Genetic impacts on DNA methylation: research findings and future perspectives. Genome Biol 22:127. https://doi.org/10.1186/s13059-021-02347-6
77. Bock C, Paulsen M, Tierling S et al (2006) CpG island methylation in human lymphocytes is highly correlated with DNA sequence, repeats, and predicted DNA structure. PLoS Genet 2: e26. https://doi.org/10.1371/journal.pgen.0020026

Laboratory Methods in Epigenetic Epidemiology

3

Ludovic Barault and Rebecca C. Rancourt

Contents

L. Barault
Department of Oncology, University of Torino, Torino, Italy

Candiolo Cancer Institute-FPO, IRCCS, Candiolo, Italy

R. C. Rancourt (✉)
Division of 'Experimental Obstetrics', Clinic of Obstetrics, Charité—Universitätsmedizin Berlin, corporate member of Freie Universität Berlin, Humboldt-Universität zu Berlin, and Berlin Institute of Health, Berlin, Germany
e-mail: Rebecca.rancourt@charite.de

K. B. Michels (ed.), *Epigenetic Epidemiology*,
https://doi.org/10.1007/978-3-030-94475-9_3

Abstract

The field of Epidemiology aims to pinpoint the risk factors for disease and health conditions and to quantify the association between lifestyle/environmental factors and disease frequency in populations. How and what is behind the epidemiological links at the cellular and molecular level are being researched, especially with the growing Epigenetic studies. Together these two fields as well as collecting the relevant biospecimens will aid in decoding the epigenome and its functional relevance. This chapter covers the importance of the thoughtful experimental design, collection, and preparation of biological samples in population-based studies, as well as providing a summary of the various techniques/technology used to analyze Epigenetic modifications at the DNA, RNA, and Chromatin level. Methodology involving both small- and large-scale approaches will be discussed (including their benefits and limitations), e.g., targeted gene sequencing and genome-wide arrays covering approaches such as DNA methylation, mRNA expression, and chromatin/histone modifications.

3.1 Introduction

Good laboratory practices are fundamental when investigating research questions. At the start, it is important to plan the study design and consider the choice of biological samples and study population, and the appropriate data analysis. This chapter covers a concise and clear introduction to the methods and techniques the most commonly for epigenetic research.

3.2 Study Design

3.2.1 Biospecimens and Materials

Deciding what samples are ideal/essential for addressing research aim/hypothesis. An individual has one genome but many different epigenomes across various tissue and cell types, and an observation may not be the same across tissues.

Epigenetic alterations can affect the different cellular components (DNA, methylation, RNA transcription levels, and chromatin modifications) that all require specific preparation. Planning out which biospecimen can be collected as well as how preparation will occur determines what epigenomic level can be investigated.

- *Cells from body fluids, fresh tissues, or relevant cell lines*: If the samples are freshly collected, nucleic acid isolation (DNA and RNA), and/or histone modification analyses (if the number of cells in the sample is large enough) can be performed. Once frozen, unless specific measures are taken (i.e., preservatives) cells will burst and will no longer be enabled for sorting or isolation. In the context of epigenetic epidemiology, cell lines can be useful for development and

optimization of experimental protocols, and to test the effect of certain drugs on the loci of interest.

- *Formalin fixed Paraffin embedded tissues* (FFPE): Until recently, FFPE tissues have mostly been used for DNA as these samples run the risk of being highly degraded. However, there are newer sample isolation protocols and kits available to help restore DNA, RNA, and even chromatin material for certain follow-up techniques. Recent advances in the preservation/recovery field are now allowing decent RNA extraction from FFPE blocks, opening experimental design to massive retrospective datasets. Although nucleic acids from FFPE are often fragmented, the ability to study large genomic fragments (more than 500 base pairs [bp]) may be impossible. New restoration kits have been recently developed enabling a better recovery of nucleic acids from FFPE tissue, and allowing omics experiments (e.g., microarrays and sequencing assays). Also, despite improvements in the field, the chromatin structure can be fragmented making methods such as chromosomal immunoprecipitation (ChIP) analyses limited. in the absence of good extraction methods [1].
- *Genomic DNA*: DNA can be isolated to assess the cytosine methylation and hydroxymethylation analyses, immunoprecipitation targeting DNA (methylcytosine), or DNA bound with supplemented transcription factors.
- *RNA*: RNA is isolated for gene expression/transcriptional analyses. RNA can be extracted from tissues and bodily fluids; however, optimal planning and timing must be considered. While RNA can be obtained from fresh frozen tissue, in the example of blood samples-it is essential to add an RNAse inhibitor buffer at the time of collection (e.g., RNA later), to prevent RNA degradation. Some experiments will, unfortunately, not be possible due to the initial cell lysis (e.g., microdissection, cell sorting, and phase separation for blood, and possibly expression microarrays).
- *cDNA*. For most RNA gene expression methods (see below), messenger RNA must be first converted to cDNA by reverse transcription. Further essential points for cDNA synthesis are outlined below in the gene expression section.
- Cell-free circulating DNA (cfDNA): The emerging field of liquid biopsy raised the importance of body fluid sample collection for analyses disregarding the blood cells. cfDNA is shed by surrounding tissue (e.g., fetus, dying cells, tumors…) and will circulate for a short period of time within the fluid (i.e., plasma, urine, saliva, and cerebrospinal fluid)[2]. This type of DNA bears the same genetic/epigenetic information as their tissue of origin enabling minimally invasive detection of diseases or conditions. Samples processed for this type of study will require special handlings in terms of processing time, preservative buffers to use, and pre-processing pipeline.
- Single-cell isolation: More molecular biological studies are striving to perform analysis at the single-cell level. Several proven single-cell isolation protocols/methods are available such as laser capture microdissection, Fluorescence-Activated Cell Sorting (FACS), dielectrophoretic digital sorting, and enzymatic digestion[3, 4].

3.2.2 Experimental Backgrounds for Epidemiologic Epigenetic Research Studies

The ability to have a multidisciplinary team or collaborations which greatly benefit Epidemiological-Epigenetic studies. The following are key aspects which out ought to be considered:

Sample collection: Collection of samples should be performed by someone with a background in molecular biology and in an ideal scenario by a pathologist or someone skilled in histology, ensuring the desired tissue- and cell type is collected. Microdissection of tissue may be necessary to ensure a homogenous cell type. Samples preservatives need to be selected with caution as some of them might disturb the epigenetic information (e.g., preservatives in cfDNA analyses [5]).

- *Experiments*: Various commercially available kits have made lab experiments more robust, however, a strong background in molecular biology allows for proper interpretation of the results and to avoid flaws in study design, experimental validity and reproducibility, and quality control which is required for epigenetic experiments. For example, more kit options have enabled simultaneous isolation of both RNA and DNA from the individual sample.
- *Analyses*: Statisticians and bioinformatics specialists are vital especially since most of the high-throughput techniques require complex analyses with large data sets. Furthermore, statisticians/bioinformatics involvement is important at the study design level to ensure that the hypothesis can be properly addressed on the potential cohort.
- Interpretation: Results should optimally be reviewed and interpreted by a multidisciplinary team such as molecular biologists, pathologists, epidemiologists, statisticians, and others.
- *Sample and data management*: Epidemiologic studies involve large sample sets/populations, creating vast amounts of data to manage. Large cohort could use bar-coded tubes which are read by a scanner generating a unique sample barcode number making sample labeling and tracking more efficient. Spreadsheets systems along are not an efficient solution for high-scale data management therefore, implementing such disciplines as bioinformatics or database management systems (SQL, Oracle, etc.) and/or using a LIMS (Laboratory Information Management Systems).

3.3 Epigenetic Methods Targeting DNA Modifications: DNA Methylation

3.3.1 DNA Modifications

DNA modifications encompass several chemical modifications of the nucleic acids which, despite affecting the affinity of DNA binding domains of enzymes and proteins, do not affect the pairing of the DNA strand. The DNA Methylation modifications have been increasingly studied since the 1980s, 5-Methyl Cytosine

(5mC) is referred to as the fifth DNA base. While 5-hydroxymethylcytosine (5hmC) which is referred to as DNA's sixth base was first discovered in bacteriophage DNA in the 1950s. 5hmC has been a fascinating discovery as it was found present in mammalian embryonic stem cells and in cerebellar Purkinje cells [6, 7]. 5hmC is known to be derived from 5mC by the action of TET family enzymes [7], leading to other DNA modifications such as 5-formylcytosine (5fC) and 5-carboxylcytosine (5caC).

The role of 5hmC/5fC/5caC remains not fully understood, but is thought to have the following effects:

- To affect the annealing of methyl-binding protein to the DNA.
- To induce demethylation through the DNA base pair repair mechanism.
- To initiate chromatin modification via the recognition of protein-specific interactions.

An important point regarding 5hmC/5fC/5caC is the absence of conversion into uracil, during bisulfite treatment, creating the possibility of some regions of the genome being mislabeled as methylated when they may contain other modifications.

The next paragraphs will give the readers an exhaustive view of past and present methods used for the most well-studied epigenetic DNA modifications. Table 3.1 will summarize all methods presented in the first version of this chapter and in the current; the following paragraph will emphasize on techniques currently used in epidemiologic studies based on the approach they are using.

3.3.1.1 Antibody-Based Methods

The last decade saw the emergence of new companies with an expertise in antibody production, improving the range of their applications. Advantages of these applications are their ability to discriminate the different types of modifications, and the possibility to identify their spatial location within a tissue; their caveats are related to the sensitivity of antibody, which usually requires a high amount of material and a high density of the DNA modification (e.g., number of CGs in the genomic locus).

- *ELISA-type methods*: In these assays, genomic DNA is coated on a plate in the presence of the DNA modification specific antibodies. The incorporation of the antibodies is quantified (like any standard ELISA) and corresponds to the global level of the modification. The development of recent assay kits has decreased the starting amount of DNA, and processing time. The disadvantage of this technique is possible underestimation of the global level of a modification since antibodies preferentially bind to areas of high CG density.
- Immunofluorescence (IF) methods: Usually performed on tissue section, this can also be performed on slides made from smear or cytocloats. The immunofluorescence assays targeting DNA modifications allow the spatial visualization of a modification, i.e., which cells are affected by a modification providing at the same time a semi-quantification. Antibodies are now available for all different types of DNA modifications enabling their individual discrimination in multiplex assay [9].

Table 3.1 Methods to study DNA methylation

	Requires bisulfite treatment	Discriminates DNA modifications	Quantitative	Global modification level	Locus specific level	Amenable for rare event detection	Amenable for single-cell analysis	Spatial detection	Requires specific/ proprietary equipment	Expensive	References
Antibody-based methods											
ELISA		Y	Y	Y							[8]
Immuno-fluorescence		Y		Y			Y	Y			[9]
In situ hybridization		Y			Y	Y	Y	Y	Y	Y	[10]
MeDIP		Y		Y	Y					Y	[11]
PCR based methods											
Bs-ddPCR	Y		Y	Y	Y	Y	Y		Y		[12]
BsPCR + MALDI-TOF	Y		Y	Y	Y				Y	Y	[13]
BsPyrosequencing	Y		Y	Y	Y				Y		[14]
COBRA	Y		Y		Y	Y					[15]
methylBEAMing	Y		Y		Y	Y				Y	[16]
Methylight	Y		Y		Y						[17]
MS-HRM	Y		Y		Y				Y		[18]
MSP	Y				Y						[19]
MSRE-PCR		Y			Y						[20]
Array based methods											
Infinium	Y		Y	Y	Y				Y	Y	
MCA		Y	Y	Y	Y						[21]
Sequencing based methods											
Bs-target sequencing	Y		Y		Y				Y	Y	[22]

CHIP-Seq		Y			Y				Y	Y	
EPIC-TruSeq	Y		Y		Y				Y	Y	
MeDIP-Seq		Y			Y				Y	Y	[23]
Nanopore		Y	Y	Y	Y				Y	Y	[24, 25]
RRBS	Y		Y		Y				Y	Y	[26]
WGBS	Y		Y	Y	Y				Y	Y	

- In situ hybridization: as for IF, In situ hybridization will be performed on slides. Combining two types of detection: DNA modification is recognized by the antibody while the genomic locus is hybridized with probes; proximity of the two compounds will allow amplification of a signal and its visualization. in situ hybridization should allow the detection of modified alleles in each individual cells. Such a method is dependent on the ploidy of the sample for the locus of interest and requires specific controls. While very interesting, this application remained very limited due to its difficulty in optimizing and reproducing the results. Development is ongoing to render such a method more reliable and amenable to screening.
- MeDIP and derivatives: DNA modification is targeted by the antibody and then subjected to immunoprecipitation. As with all antibody-based methods, MeDIP and derivatives usually require a large amount of material. Such caveats have recently been overcome using filler compounds to improve immunoprecipitation efficiency and highly sensitive methods such as NGS for output detection.

3.3.1.2 PCR-Based Methods

Bisulfite treatment is a key requirement for many of the current PCR-based techniques assessing DNA modification and particularly methylation. Various bisulfite treatment kits are now commercially available, enabling high recovery rate with a short incubation time, lower starting amounts of DNA and high conversion efficiency. Bisulfite treated DNA (BsDNA) results in single-stranded DNA of low complexity, since conversion of all unmodified cytosines into uracil will drastically decrease the percentage of GC in the sequence. Different techniques can be used to evaluate the difference between unmethylated and methylated DNA. Single-stranded DNA is very fragile and must be handled and stored properly (e.g., −20 °C for short-term use, −80 °C for long-term storage, and repeated freeze–thawing should be avoided). Bisulfite revolutionized the field of epigenetics, allowing for rapid identification of cytosine modifications with a decent cost expense. However, its huge caveats are the absence of discrimination between the modified base. In fact, without specific adjustment (addition of protective group, enzymatic conversion. . .) the chemical conversion by bisulfite of a 5mC or 5hmC template will result in the same output. Moreover, PCR-based methods working directly on previously extracted DNA will therefore lose the single-cell and spatial information of the sample.

The primers used for the PCR-based analyses will often differentiate the type of assay performed. Level of the modification is evaluated at a specific locus. Amplifying repetitive elements such as LINE1 or ALU allows for estimating of global level of methylation [27].

MSP (Methyl specific PCR): Developed by Herman and colleagues in 1996 [19], this technique remains very present in the literature. It consists of amplification specific to the methylated or unmethylated sequence. The targeted amplification is performed using primers with CG sites (or UG after treatment for the unmethylated DNA) in the 3′ sequence stabilizing the annealing specificity. This technique is sensitive and requires the amplification of unmethylated sequence to confirm that the absence of an amplicon for the methylated sequence is due to the lack of methylation and not a problem with the amplification protocol. *Methylight* is a real-time MSP

[17]. Primers and probes specific to the methylated or unmethylated sequences are used and the methylation ratio is calculated by the fluorescence ratio of the probes. The quantification of one amplicon is relative to the other making this technique semi-quantitative as it is not compared to a standard. Efficiencies of the two amplifications (methylated and unmethylated) must be equivalent. This technique is dependent on the quality and the careful positioning of the probes. Moreover, the methylation ratio resulting from the experiment is an average of the CpG sites on the probe and the primers. It is important to note that a mutation at one of the sites will completely alter the percentage of methylation and should be avoided by carefully confirming the sequence of interest. Compartmentalization of the template was proposed to allow the technique to detect rare events, though, new digital droplet platform outperformed this type of procedure.

As opposite to MSP, Methylation independent PCR (MIP) regroups all the methods that discriminate methylation not at the amplification level (i.e., primers), but at the output level (use of internal probes, restriction enzyme, primer extension…). In these cases, amplification is performed using primers which will anneal on regions possibly devoid of CG sites, avoiding amplification bias toward the modified or unmodified template.

Bisulfite sequencing (Bs-Sequencing) was the first technique developed to assess DNA methylation [28]. Amplification is first performed on BsDNA followed by sequencing. Setting aside its cost, this technique has the benefit of providing information at each CpG site in the sequence of interest and can be useful as an exploratory step. However, the methylation levels cannot be quantified limiting the ability to draw associations with any transcriptional consequences.

Among MIP assays, *MS-HRM* (Methylation-Specific High-Resolution Melting curve analysis) [18], *COBRA* (*COmBined Restriction enzyme Analyses*) [15], and (Bs-PCR w/MALDI-TOF) [13] might represent a good solution to routinely assess a biomarker, though, their recent use in the literature has decreased.

Nowadays, the most commonly used PCR-based methods remain *Bisulfite Pyrosequencing* (Bs-pyrosequencing) [14]. PCR amplification is performed on Bs-DNA using primers specific to the region of interest. One of the primers is biotinylated, allowing for the binding of avidin-coated sepharose beads. The amplicon is denatured, and the non-biotinylated strand is washed off. The remaining single strand is then bound to the bottom of the reaction plate, and buffers, sequencing primer, and DNA synthase enzyme are added to the reaction. Each nucleotide is then sequentially added to the reaction following a specific dispensation order. If the dispensed base is complementary to the upcoming base in the DNA strand, this nucleotide will be added to the sequence by the DNA synthase, which releases pyrophosphates that are converted into light. The emitted light is proportional to the amount of base incorporated into the DNA sequence. Therefore, the ratio of the signal following the dispensation of a C or a T at the CpG site will represent the level of methylation at this site. This technique requires a pyrosequencer, but is highly quantitative, very reproducible, and easy to develop. Bs-Pyrosequencing can also be used for allele-specific methylation, by the use of a sequencing primer with a single nucleotide polymorphism (SNP) in its 3′region [29].

Among PCR-based methods, MSRE-PCR should be mentioned, as this is one of the few that does not require prior bisulfite treatment of the DNA. In this type of protocol, methylation is detected by the presence of an amplicon after digestion of the locus by methylation sensitive restriction enzyme. Isoschizomers are used, coupled with primers specific to the locus of interest. For each sample, three reactions should be performed to ensure reliability of the assay (without enzyme with methylation insensitive enzyme, and with methylation sensitive enzyme). Coupled with quantitative PCR, this protocol allows for the pseudo quantification of methylation level. With the recent development of new enzymes as well as protective treatment [e.g., glucosyl or glucose] differences among the DNA modification might be achieved. However, this method is limited to the recognition site of the enzyme.

3.3.1.3 Array-Based Methods

Increasingly, more and more epigenetic studies have been performed at a genome-wide scale [30]. To date, these projects have mainly focused upon DNA methylation and histone modification.

- *Methylation microarray after bisulfite treatment* is currently the most used genome-wide application. In this protocol, BsDNA is amplified and put in contact with probes recognizing one CpG site each. A single base extension with labeled nucleotides will define the ratio of methylation of each CpG site. However, this method has several drawbacks: firstly, the microarray's targets are limited to the available probes defined by the company and are not always related to any functional annotations; secondly, the use of bisulfite treatment (as previously described) does not allow the discrimination of other DNA modifications. The most popular arrays have been the Infinium 450 K now replaced by the Infinium EPIC.
- Several genome-wide assays use restriction enzymes, and one of the first methods was *MCA* (Methylated CpG Island amplification) [21]. Enrichment for methylated DNA is performed by digestion with methylation sensitive and methylation-insensitive restriction enzymes, followed by ligation of adaptors and PCR amplification. The resulting amplicons, which are representative of the methylated fraction, are labeled and co-hybridized in a microarray platform or sequenced.

3.3.1.4 Sequencing Based Methods

- Whole Genome Bisulfite sequencing (WGBS) was the first next generation sequencing platform developed for assessing DNA methylation. However, due to high starting amount requirement and high cost for sequencing, use of WGBS has remained very limited. With reduction of cost, home-made custom panel target sequencing, as well as commercial ones (EPIC-TruSeq), came to life. In target sequencing DNA is generally, bisulfite treated prior to adapter ligation, library preparation and capture, since it requires several steps of amplification which would get rid of DNA modifications. By converting the DNA first target sequencing enable sequencing of template with relatively low input, though the library preparation will suffer from the low complexity of the converted DNA, possibly losing some of the regions of interest at the capture level. On the contrary, the EPIC-TruSeq pipeline will start with the capture, proceeded by bisulfite conversion and

library preparation. By starting with the capture, the EPIC-TruSeq bypasses the issue of complexity loss, though the enrichment observed in non-amplified template is very low and therefore requires large amount of material. EPIC-TruSeq assesses the same CpG sites as the microarray equivalent, its advantage being that it also gives information at other cytosines levels (i.e., CHG, CHH). Though the requirement for more complex bioinformatics analyses, the need for high depth and the higher cost compared to the microarray, lead to less frequent use of this platform.

- With *RRBS* (*Reduced Representation Bisulfite Sequencing*) [26], a restriction enzyme digest is followed by purification of the fragments (within a range of specific lengths), thus enabling the enrichment of sequences containing CpGs. DNA fragments are next bisulfite converted, amplified, and finally sequenced. This labor-intensive technique allows the widespread study of methylation across the genome (not exclusively in the promoter region) and works with a very low amount of DNA (less than 100 ng).
- *MeDIP-Seq and ChIP-Seq* are two methods using immunoprecipitation as the first step. In this protocol, DNA is precipitated using antibodies specific to 5mC (MeDIP), methyl-binding proteins, or histone modifications. The precipitated DNA is next used for massively parallel sequencing analysis. These protocols offer an in-depth screening assay. The disadvantages with antibodies are the need for a high density of the target, and the possibility of relatively low specificity creating potential false positives. Moreover, these methods do not allow for the direct measurement of a methylation ratio. The improvement of the MeDIP protocol by incorporation of a DNA filler, recently allowed its use for small input amounts such as circulating DNA, rendering this method highly interesting for individual screening.
- Nanopore sequencing. Nanopore sequencing uses a proprietary platform on which nucleic acid strands go through a protein nanopore. While going through this channel the electric current will change depending on the base (modified or not); such variation is monitored by the system and decrypted into the sequence. Nanopore sequencing is very interesting since it allows a real-time reading of the sequence, does not need specific equipment (the single unit machine is relatively inexpensive), and can therefore be used in a non-lab environment. Another advantage is the absence of bisulfite conversion requirement. However, the technique is still young and might require additional development prior to large population screening use.

3.3.2 Tips for Developing a Methylation Assay

- *Bisulfite conversion efficiency* is an important aspect in 5mC assessment. Incomplete conversion may lead to false estimation of the methylation (particularly true with homemade bisulfite treatment and methylation specific amplification methods such as MSP, and Methylight).
  - Avoid homemade conversion. Prefer commercial kits, which are much more reliable and reproducible. A large panel of kits is available which might fit your requirement in terms of cost or template to be assessed [31].

  - One assessment can involve the ratio of converted BsDNA by evaluating an amplification product bearing a C outside CpG sites and using primers independent of the conversion (without any cytosine in their sequence). The conversion of cytosine in the amplicon can be checked by probes, melting curve analysis, or even sequencing.
  - Another assessment is by investigating the complete conversion of an individual DNA strand for methods that look at individual CpG sites (Bs-sequencing, Bs-PCR w/MALDI-TOF, and Bs-Pyrosequencing). Failure of complete conversion will result in the presence of cytosine outside of CpG sites, visible in Bs-sequencing and Bs-PCR w/MALDI-TOF. This is available in Bs-Pyrosequencing by adding a C in the dispensation order (at the position of potential unconverted C).
  - *Primer design* is one of the most important features of methylation assay optimization [32]. There are several rules to follow, according to the experiment to perform.
  - If the assay is using a methylation-specific amplification (MSP, Methylight), primers should contain one or more CpG sites in their 3′region to ensure a specific amplification of the (un)methylated sequence. Freeware is available for primer design (e.g., Methprimer).
  - For an assay using a methylation independent amplification, primers should avoid having CG in their sequence or these dinucleotides must be limited to one or two in the 5′ region. When CpG sites in the primers cannot be avoided, the Cytosine of the dinucleotide (or its complementary depending on the primer direction) should be replaced by a mismatched base independent of the methylation status (A instead of C/T or T instead of G/A).
  - The salt adjusted melting temperature (Tm) should be higher than 60 °C, to enable a PCR Tm at 60 °C.
  - The 3′ region containing a T derived from the conversion of cytosines not included in a CpG site will ensure strict amplification of the BsDNA.
- Amplification biases: When we talk about DNA modification, we will often work with templates that are GC rich, and which are prone to amplification biases. As demonstrated by Aird and colleagues, temperature ramp can be a major source of biases [33]. When working on DNA modifications prefer a thermocycler with low-temperature ramp.
- *Linearity of detection*: Once the bisulfite conversion and the primer design are optimal, the ability to assess the different degree of methylation needs to be checked. A methylation scale should be created by pooling (in a dilution scale) unmethylated DNA with completely methylated DNA. Whole genome amplification can get rid of all methyl groups and complete methylation can be obtained with methylase treatment. The methylation scale provides a useful tool to confirm the accuracy of the primers and overall assay.
- *Reproducibility* must be confirmed by performing at least duplicates of the bisulfite treatment (not only the PCR!). Correlation between replications must be evaluated to ensure the best assay quality and accuracy of results (for example, a difference between the bisulfite replicates less than two times the standard deviation).

3.4 Epigenetic Studies Involving RNA modifications: Gene Expression

Analyzing gene expression is the pivotal methodology to investigate the biological function of specific genes along with any phenotypic differences that may be observed. Thorough planning of RNA isolation protocols (specific to tissue and/or bodily fluid samples) and which RNA expression experiments fit best should be considered due to the nature of cell type-specific expression of certain genes. Searching literature and genomic databases to assess any possible additional transcriptional features of the gene of interest such as: the presence of multiple splice variants and isoforms, or other transcripts that may span the region of interest (e.g., antisense RNAs) must be performed at the start.

3.4.1 Types of RNA

Translational RNA: Messenger RNA, transfer RNA, and ribosomal RNA work collectively the protein synthesis process by translating the genetic code into protein. Messenger RNA (mRNA) carries the protein sequence information to the ribosomes, while Transfer RNA (tRNA) is a small RNA chain (~80 nucleotides) that transfers specific amino acids for protein synthesis during translation. Ribosomal RNA (rRNA) combines with proteins in the cytoplasm to form the nucleoprotein ribosomes, examples of rRNA molecules are 18S, 5.8S, 28S, and 5S.

Noncoding RNAs (ncRNAs) are functional RNA molecules that do not result in a translated protein structure [34]. This group of RNAs can have regulatory functions acting to control gene expression which can be direct or through the targeting of other repressive modifications which will affect gene expression (e.g., targeting methylation). One example of an ncRNA type that has an increase in research interest is microRNA (miRNA, ~22 nucleotides) as the name indicates are small RNA molecules that direct posttranscriptional suppression of gene expression. Growing evidence shows ncRNAs such as miRNAs playing roles in the pathogenesis of human diseases such as cancers, metabolic diseases, neurological disorders, and infectious diseases [35]. Other ncRNAs are small interfering RNA (siRNA, ~21–22 nucleotides), small nucleolar RNAs (snoRNAs, ~80–300 nucleotides), and PIWI-interacting RNAs (piRNAs, ~26–30 nucleotides) and long ncRNA (greater than 200 nucleotides) [36, 37].

Antisense RNAs are transcribed in the opposite direction (antisense) of one or more endogenous mRNA transcripts which are in the sense orientation, both transcripts share the same sequence so prior investigating of sequence must be considered. Antisense RNA can act to either activate or downregulate gene expression by binding to the endogenous sense mRNA [38].

3.4.2 Methods for Overall Gene Expression

Expression analysis is used to examine the function of a gene(s) at the transcriptional level in a specific tissue or cell type and when possible, at a particular time-point (e.g., developmental, exposure, treatment). An overview of gene expression methods is in Table 3.2.

3.4.2.1 Classical Membrane-based Methods

The long studied, *classical methods* for gene expression such as Northern blots and ribonuclease protection assay (RPA) have a rich literary history and are well established. Both Northern blots and RPAs are generated via gel membrane/electrophoresis of RNA samples and usually involve visualization with the use of radiolabeled probes specific to the transcript of interest. A benefit to the RPA technique is the use of strand-specific probes which can differentiate sense versus antisense transcripts. However, these protocols are not ideal for large sample sizes, require a large starting amount of RNA material, and the need of a laboratory capable of working with radioactive isotopes, however, there are now various modifications to both protocols allowing for newer nonradioactive labeling reagents (e.g., Digoxigenin) for visualization.

Table 3.2 Methods to study gene expression

Technology	Membrane-based	Array-based	Quantitative PCR-based	Sequence-based
RNA assessment	Semi-quantitative and quantitative	Quantitative	Quantitative	Quantitative
Typical starting material (total RNA)	5–10 μg (minimum 1 μg)	As low as 100–200 ng	100–200 ng (minimum 50 ng)	2 μg (minimum 500 ng) 50,000 cells for single-cell analysis
Sample size/ scale	Small scale	High-throughput	Moderate throughput	High-throughput
Benefits	Cost-effective, not difficult to develop, capable of strand-specific analyses	Thousands of genes examined, reduced starting material	Cost-effective, easy to develop, capable of strand-specific analyses	Comprehensive profile of transcriptome and RNA biology, single-cell ability
Limitations	Labor intensive, higher RNA starting material	Costs, data analysis—labor intensive	Labor intensive	Costs, data analysis—labor intensive
Key requirements	Radioactivity, high starting material	Array core facility/service company, bioinformatics analysis	Real-time Thermocycler	Sequencing core facility/service company, bioinformatics analysis

3.4.2.2 Quantitative PCR-based Methods

Quantitative PCR (qPCR) is a very robust method for quantifying gene expression while requiring a small amount of starting material which can be ideal for precious human samples [39]. qPCR is performed on cDNA transcribed from RNA samples and uses a fluorescent reporter molecule which accumulates with the product amplification at each cycle. Real-time PCR (RT-PCR) provides analysis as the reaction progresses. There are two common detection methods when performing real-time qPCR: non-specific detection using DNA intercalating dyes (e.g., SYBR green technology) and specific detection with probes designed to a specific gene of interest/target region (e.g., hydrolysis probes aka TaqMan). Probe-based RT-PCR can additionally provide strand-specific discrimination, and validation of array experiments (see below). There are various algorithms used to quantify the levels of expression from the cycle threshold numbers/values (Ct) [40, 41].

3.4.2.3 RNA Expression Array-based Methods

Gene expression arrays allow for a high-throughput mRNA gene expression profile of thousands of genes in one sample using low amounts of starting RNA material. Microarrays are the most used in located RNA, these arrays are typically comprised of glass slides coated with transcript-specific probes. The RNA sample is fluorescently labeled and hybridized to the microarray slide, which is processed, and laser scanned. The normalization of the RNA sample for quality control is done using reference RNA (fluorescently labeled differently than the RNA sample). There are various commercially available platforms of expression arrays, and some companies offer custom made arrays, for the researcher to choose their favorite genes of interest (this provides a more moderate/mid-scale throughput and can provide a great approach to investigate numerous related genes in specific diseases or physiological pathways). Currently, there are several commercial options for expression arrays specific to identifying ncRNAs (e.g., miRNA, snoRNA). Important consideration is that array-based technology generates large amounts of data which requires bioinformatics analysis [42].

3.4.2.4 Expression Sequence-based Methods

During the past decade, an indispensable methodology for in-depth analysis of the transcriptome has been RNA sequencing (RNA-seq) technology. With over a hundred different methods rooted from the classic RNA-seq protocols, one must have careful consideration prior to planning. There are various technologies/service companies offering the many distinct RNA-seq approaches which have been comprehensively reviewed in Stark *et al.* [43]. Aside from differentially gene expression, RNA-seq based methods have provided key knowledge of gene regulation at ncRNA and RNA enhancer-elements, alternative mRNA splicing events, etc. As was mentioned in the array-based methods, sequencing technology generates large amounts of data which requires bioinformatics analysis.

3.4.3 Tips for RNA Expression Analyses

- Certain RNA isolation protocols and kits lose the ability to isolate small RNAs, so the type of RNA of interest needs to be clear when planning experiments.
- Stringent RNA quality assessment must be performed prior to experiments as possible degradation and contamination will impair the quality and validity of the experiments [44]. RNA quality can be assessed on a gel by measuring the ratio of 18S/28S or by using a more accurate technique like a bioanalyzer (which also provides important quantity values and requires less RNA).
- cDNA synthesis—pay attention to which reverse transcriptase primer(s) (e.g., OligoDT, random hexamers, or a mixture of both for priming oligos) is provided in the cDNA synthesis kits used.
  - OligoDT offers a more specific priming to poly(A) mRNA tail (as well as polyA rich regions) giving greater assurance that the mRNA is intact. [*Note: OligoDT cannot prime 18S rRNA eliminating the possibility of 18S being used as a housekeeping gene*.]
  - Random hexamer is recommended for total RNA especially if the transcription of interest does not have a polyA tail or is unknown.
- Controls: Without the proper controls, the analysis is impossible to interpret and trust.
  - Primer set choice: When possible, primers should be exon spanning to allow specific amplification of the RNA transcript instead of any possible contaminating gDNA.
  - Make RT minus negative control: In the cDNA synthesis, take half of the reaction as the negative control (RT minus) by adding water instead of the reverse transcriptase enzyme to control for gDNA contamination.
  - Housekeeping gene(s) (HKG) choice: Essential for normalization step to accurately measure the levels of cDNA samples. Multiple HKGs (2–4) must be tested and validated to confirm that HKG has ubiquitous expression in the tissue type studied and importantly does not show group differences across the study cohort [45]. Some examples of commonly used HKGs are GAPDH, β-actin, α-tubulin, HPRT, UBC, YWHAZ, and 18S.
- Replicates (biological and technical) are essential due to possible well-to-well variation, pipetting error, or poor efficiency of cDNA synthesis.
- Criteria for primer design and target amplicon
  - Primers are typically around 15–30 base pairs long with similar melting temperature (Tm), which is ideally around 58–62 °C (lower Tms can have non-specific annealing).
  - Using genome database software, confirm that the primers are specific to the targeted region and not aligning to other regions in the genome, and avoid repetitive regions.
  - The ideal amplicon length is 70–150 bp, and should not exceed 400 bp.

3.5 Epigenetic Studies Targeting Chromatin Modifications

The chromatin structure host epigenetic regulators, through proteins called histones that together with the DNA form the repeating nucleosomal units of chromatin [46]. Modifications to the histone N-terminal tails are known to be somatically heritable, have a role in regulating gene expression and occur post-translationally through methylation, acetylation, phosphorylation, ubiquitination, and other methods. Various histone modifications have been identified in many regions throughout the genome [47, 48].

3.5.1 Histones Modifications

Most histone modification analyses are through immunoprecipitation incubation of cells with antibodies targeting a specific modification after crosslinking the histones to the DNA (chemical fixation of the nucleosome). The chromatin immunoprecipitation (ChIP) protocol allows for the enrichment of DNA associated with histone modifications which can then be analyzed by sequencing (ChIP-Seq) [49], or microarray (ChIP-on-chip) [50]. As evident in a large number of ChIP-Seq studies in the literature, there are available datasets from previously performed ChIP assays found in genome databases (e.g., UCSC genome browser) across various tissue types, which provides a great resource when planning out a genomic region of interest or as to compliment DNA methylation and gene expression data.

3.5.2 Chromatin Conformation

The chromosome conformation capture assay is a technique used to identify and quantify possible physical interactions that are occurring between any two genomic loci enabling the study of in vivo genomic organization and interacting chromatin segments over vast regions of up to several hundreds of kilobases in size. The captured interactions could be specific interactions (e.g., between genes and regulatory elements), or random collisions between loci. In the chromatin conformation protocol, cells are first crosslinked which leads to stabilization or capture of the DNA–DNA, as well as DNA–protein interactions. The DNA is then digested with a restriction enzyme and the fragments are ligated, creating a mixture of interacting fragments. Sequence analysis of the area of ligation allows the identification of the interacting chromatin which enables the genome-wide analysis of the chromatin conformation [51].

3.6 Notes on Single-cell Epigenetic Methodology

With the advances in molecular genomic technology increasing so has the optimization of choosing the most optimal starting material. Across research involving DNA, RNA, and protein, cell-to-cell variation within individual tissue or blood samples has led to studies using single-cell epigenomic analysis to understand the biological functional mechanisms in a cell type-specific manner [52, 53]. Mehrmohamadi *et al*. provides a comprehensive review outlining current state of single-cell epigenomic methodology and assays [53].

3.6.1 Cell type-Specific Epigenomics

As mentioned above in the biospecimen section, generating/isolation of single-cell starting material depends on the available tissue/cells and research aims. With the adaptations of some of the above-mentioned assays, single-cell epigenomic analysis is possible for DNA methylation (e.g., single-cell bisulfite sequencing; scBS-Seq) [54], RNA expression (single-cell RNA-sequencing; scRNA-Seq) Dal Molin and Di Camillo proved an in-depth experimental design review [55], and chromatin modification (e.g., single-cell assay for transposase-accessible chromatin; ATAC-seq, scChIP-seq) [50]. Remarkably, current ground-breaking modifications to single-cell protocols/assays are being further advanced to simultaneously analyze the methylation and transcriptome or chromatin modifications (for a multi-omic approach) which is optimal for valuable hard-to-obtain material. [56].

3.6.2 Benefits and Limitations to Single-cell Epigenomics

The greatest benefit of single-cell approaches is the ability to study cellular heterogeneity, cell-to-cell interactions, and rare/sub-cell population, especially at critical time points in early stages of embryonic development and/or disease pathology etc. (possible biomarker applications). However, the generation/isolation of single-cell material is labor intensive/time consuming and requires expensive core facilities/equipment such as FACS machines, laser microdissection platforms, or cell culture units. In the instances where single-cell isolation can be clearly performed and would provide true biological functional relevant insight to the research aim/hypothesis this would be gold-standard material.

Single-cell epigenomics involves sequencing-based arrays/technologies which have been discussed in previous sections. These methods allow for medium to high throughput, great coverage, and resolution across the genome and transcriptome which is hugely beneficial when dealing with patient samples where starting material is limited. In addition to the high expense, large datasets are generated and require a great knowledge of computational analysis (bioinformatics, statistician). Most of these studies are performed on a smaller scale sample size, as it is cost- and time-prohibited to scale up to large population cohorts. With the

relatively new applications of single-cell epigenomic sequencing technologies/research tools, a great challenge and important consideration is what are best controls (e.g., technical, and biological controls for estimating and validating cellular variations, normalization, and spike—in standards. . .)[52].

3.7 Method Section Recommendations in Communications and Publications

Along with the epidemiological cohort/sample population information, there are essential molecular-genomic experimental details that should be provided in publications for validation, reproducibility, and data sharing within the research field (some examples include chromosomal locations of the targeted genes, primer sequences, tissue/sample selection and preparation, experimental controls) (Table 3.3).

Table 3.3 Epigenetic-epidemiology studies: Key information to be included in publications

Experimental analysis	Key requirements for publication	Examples
Cohort samples	Collection	Cohort details: e.g., BMI, age, treatment. . .
	Isolation protocol	Kits used, time until sample processing
	Samples	Sample/cell type (e.g., tissue, serum, buccal)
	Storage	Storage temperature and duration
Epigenetic studies (DNA, RNA, Chromatin)	Sample preparation	Tissue/cell type, isolation details, starting amounts, chromatin cross-linking info
	Target sequence	Accession number, probes filtering for arrays
	Primers and probe sequence	Sequence or reference number
	Bisulfite Primers sequence	Unconverted and converted sequence
	Reproducibility	Coefficient of variation
	Replication	Number and type of replicates
	Replication	Number and type of replicates
	Controls	Type of controls (e.g., negative control for RT-PCR, IP, antibody specificity control, chromatin conformation ligation template control)
	Normalization	Housekeeping genes
	Antibody information	Reference, dilution

3.8 Conclusion

Together Epidemiological-Epigenetic studies ought to have clearly defined experimental hypotheses and aims with thorough planning of cohort (especially for subgroup generation) and biospecimen collection as well as selection of the best/relevant methods. With clear epidemiological human/patient cohorts as an important starting material, researchers will strive to complete a more whole epigenome and transcriptome landscape/profile with the key combination of technologies on DNA, RNA, and Chromatin [53, 57–59].

References

1. Collas P (2010) The current state of chromatin immunoprecipitation. Mol Biotechnol 45(1): 87–100
2. Pos O, Biro O, Szemes T, Nagy B (2018) Circulating cell-free nucleic acids: characteristics and applications. Eur J Hum Genet 26(7):937–945
3. Gross A, Schoendube J, Zimmermann S, Steeb M, Zengerle R, Koltay P (2015) Technologies for single-cell isolation. Int J Mol Sci 16(8):16897–16919
4. Li X, Tao Y, Lee DH, Wickramasinghe HK, Lee AP (2017) In situ mRNA isolation from a microfluidic single-cell array using an external AFM nanoprobe. Lab Chip 17(9):1635–1644
5. Bartak BK, Kalmar A, Galamb O, Wichmann B, Nagy ZB, Tulassay Z et al (2019) Blood collection and cell-free DNA isolation methods influence the sensitivity of liquid biopsy analysis for colorectal cancer detection. Pathol Oncol Res 25(3):915–923
6. Kriaucionis S, Heintz N (2009) The nuclear DNA base 5-hydroxymethylcytosine is present in Purkinje neurons and the brain. Science 324(5929):929–930
7. Tahiliani M, Koh KP, Shen Y, Pastor WA, Bandukwala H, Brudno Y et al (2009) Conversion of 5-methylcytosine to 5-hydroxymethylcytosine in mammalian DNA by MLL partner TET1. Science 324(5929):930–935
8. Chowdhury B, Cho IH, Hahn N, Irudayaraj J (2014) Quantification of 5-methylcytosine, 5-hydroxymethylcytosine and 5-carboxylcytosine from the blood of cancer patients by an enzyme-based immunoassay. Anal Chim Acta 852:212–217
9. Singh RK, Diaz PE, Binette F, Nasonkin IO (2018) Immunohistochemical Detection of 5-Methylcytosine and 5-Hydroxymethylcytosine in Developing and Postmitotic Mouse Retina. J Vis Exp 138:58274
10. Nadiminti K, Silverman M, Bhagavathi S, Vikas P (2019) t(15; 17) associated with primary myelofibrosis: a case report of an unusual clinical presentation and diagnostic dilemma. Onco Targets Ther 12:5449–5455
11. Mohn F, Weber M, Schubeler D, Roloff TC (2009) Methylated DNA immunoprecipitation (MeDIP). Methods Mol Biol 507:55–64
12. Yu M, Heinzerling TJ, Grady WM (2018) DNA Methylation analysis using droplet digital PCR. Methods Mol Biol 1768:363–383
13. Suchiman HE, Slieker RC, Kremer D, Slagboom PE, Heijmans BT, Tobi EW (2015) Design, measurement and processing of region-specific DNA methylation assays: the mass spectrometry-based method EpiTYPER. Front Genet 6:287
14. Delaney C, Garg SK, Yung R (2015) Analysis of DNA methylation by pyrosequencing. Methods Mol Biol 1343:249–264
15. Xiong Z, Laird PW (1997) COBRA: a sensitive and quantitative DNA methylation assay. Nucleic Acids Res 25(12):2532–2534

16. Li M, Chen WD, Papadopoulos N, Goodman SN, Bjerregaard NC, Laurberg S et al (2009) Sensitive digital quantification of DNA methylation in clinical samples. Nat Biotechnol 27(9): 858–863
17. Campan M, Weisenberger DJ, Trinh B, Laird PW (2009) MethyLight. Methods Mol Biol 507: 325–337
18. Wojdacz TK, Dobrovic A, Hansen LL (2008) Methylation-sensitive high-resolution melting. Nat Protoc 3(12):1903–1908
19. Herman JG, Graff JR, Myohanen S, Nelkin BD, Baylin SB (1996) Methylation-specific PCR: a novel PCR assay for methylation status of CpG islands. Proc Natl Acad Sci U S A 93(18): 9821–9826
20. Melnikov AA, Gartenhaus RB, Levenson AS, Motchoulskaia NA, Levenson Chernokhvostov VV (2005) MSRE-PCR for analysis of gene-specific DNA methylation. Nucleic Acids Res 33(10):e93
21. Estecio MR, Yan PS, Ibrahim AE, Tellez CS, Shen L, Huang TH et al (2007) High-throughput methylation profiling by MCA coupled to CpG island microarray. Genome Res 17(10): 1529–1536
22. Masser DR, Stanford DR, Freeman WM (2015) Targeted DNA methylation analysis by next-generation sequencing. J Vis Exp 96:52488
23. Shen SY, Burgener JM, Bratman SV, De Carvalho DD (2019) Preparation of cfMeDIP-seq libraries for methylome profiling of plasma cell-free DNA. Nat Protoc 14(10):2749–2780
24. Simpson JT, Workman RE, Zuzarte PC, David M, Dursi LJ, Timp W (2017) Detecting DNA cytosine methylation using nanopore sequencing. Nat Methods 14(4):407–410
25. Ni P, Huang N, Zhang Z, Wang DP, Liang F, Miao Y et al (2019) DeepSignal: detecting DNA methylation state from Nanopore sequencing reads using deep-learning. Bioinformatics 35(22): 4586–4595
26. Smith ZD, Gu H, Bock C, Gnirke A, Meissner A (2009) High-throughput bisulfite sequencing in mammalian genomes. Methods 48(3):226–232
27. Yang AS, Estecio MR, Doshi K, Kondo Y, Tajara EH, Issa JP (2004) A simple method for estimating global DNA methylation using bisulfite PCR of repetitive DNA elements. Nucleic Acids Res 32(3):e38
28. Frommer M, McDonald LE, Millar DS, Collis CM, Watt F, Grigg GW et al (1992) A genomic sequencing protocol that yields a positive display of 5-methylcytosine residues in individual DNA strands. Proc Natl Acad Sci U S A 89(5):1827–1831
29. Wong HL, Byun HM, Kwan JM, Campan M, Ingles SA, Laird PW et al (2006) Rapid and quantitative method of allele-specific DNA methylation analysis. BioTechniques 41(6): 734–739
30. Laird PW (2010) Principles and challenges of genomewide DNA methylation analysis. Nat Rev Genet 11(3):191–203
31. Worm Orntoft MB, Jensen SO, Hansen TB, Bramsen JB, Andersen CL (2017) Comparative analysis of 12 different kits for bisulfite conversion of circulating cell-free DNA. Epigenetics 12(8):626–636
32. Wojdacz TK, Hansen LL, Dobrovic A (2008) A new approach to primer design for the control of PCR bias in methylation studies. BMC Res Notes 1:54
33. Aird D, Ross MG, Chen WS, Danielsson M, Fennell T, Russ C et al (2011) Analyzing and minimizing PCR amplification bias in Illumina sequencing libraries. Genome Biol 12(2):R18
34. Mattick JS, Makunin IV (2006) Non-coding RNA. Hum Mol Genet 15(1):R17–R29
35. Taft RJ, Pang KC, Mercer TR, Dinger M, Mattick JS (2010) Non-coding RNAs: regulators of disease. J Pathol 220(2):126–139
36. Klinge CM (2018) Non-coding RNAs: long non-coding RNAs and microRNAs in endocrine-related cancers. Endocr Relat Cancer 25(4):R259–R282
37. Liu B, Xiang W, Liu J, Tang J, Wang J, Liu B et al (2021) The regulatory role of antisense lncRNAs in cancer. Cancer Cell Int 21(1):459

38. Su WY, Xiong H, Fang JY (2010) Natural antisense transcripts regulate gene expression in an epigenetic manner. Biochem Biophys Res Commun 396(2):177–181
39. VanGuilder HD, Vrana KE, Freeman WM (2008) Twenty-five years of quantitative PCR for gene expression analysis. BioTechniques 44(5):619–626
40. Pfaffl MW (2001) A new mathematical model for relative quantification in real-time RT-PCR. Nucleic Acids Res 29(9):e45
41. Stahlberg A, Hakansson J, Xian X, Semb H, Kubista M (2004) Properties of the reverse transcription reaction in mRNA quantification. Clin Chem 50(3):509–515
42. Camacho L, Porter-Gill P, Silva CS (2019) Data on the effect of heat and other technical variables on the detection of microRNAs in human serum. Data Brief 24:103750
43. Stark R, Grzelak M, Hadfield J (2019) RNA sequencing: the teenage years. Nat Rev Genet 20(11):631–656
44. Bustin SA, Benes V, Garson JA, Hellemans J, Huggett J, Kubista M et al (2009) The MIQE guidelines: minimum information for publication of quantitative real-time PCR experiments. Clin Chem 55(4):611–622
45. Vandesompele J, De Preter K, Pattyn F, Poppe B, Van Roy N, De Paepe A et al (2002) Accurate normalization of real-time quantitative RT-PCR data by geometric averaging of multiple internal control genes. Genome Biol 3(7):0034
46. Consortium, EP (2012) An integrated encyclopedia of DNA elements in the human genome. Nature 489(7414):57–74
47. Kouzarides T (2007) Chromatin modifications and their function. Cell 128(4):693–705
48. Gelato KA, Fischle W (2008) Role of histone modifications in defining chromatin structure and function. Biol Chem 389(4):353–363
49. Kim TH, Dekker J (2018) ChIP-seq. Cold Spring Harb Protoc 2018(5):29610358
50. Buenrostro JD, Wu B, Litzenburger UM, Ruff D, Gonzales ML, Snyder MP et al (2015) Single-cell chromatin accessibility reveals principles of regulatory variation. Nature 523(7561): 486–490
51. Lieberman-Aiden E, van Berkum NL, Williams L, Imakaev M, Ragoczy T, Telling A et al (2009) Comprehensive mapping of long-range interactions reveals folding principles of the human genome. Science 326(5950):289–293
52. Kelsey G, Stegle O, Reik W (2017) Single-cell epigenomics: recording the past and predicting the future. Science 358(6359):69–75
53. Mehrmohamadi M, Sepehri MH, Nazer N, Norouzi MR (2021) A comparative overview of epigenomic profiling methods. Front Cell Dev Biol 9:714687
54. Liang J, Cai W, Sun Z (2014) Single-cell sequencing technologies: current and future. J Genet Genomics 41(10):513–528
55. Dal Molin A, Di Camillo B (2019) How to design a single-cell RNA-sequencing experiment: pitfalls, challenges and perspectives. Brief Bioinform 20(4):1384–1394
56. Hu Y, An Q, Guo Y, Zhong J, Fan S, Rao P et al (2019) Simultaneous profiling of mRNA transcriptome and DNA methylome from a single cell. Methods Mol Biol 1979:363–377
57. Rivera CM, Ren B (2013) Mapping human epigenomes. Cell 155(1):39–55
58. Lawrence M, Daujat S, Schneider R (2016) Lateral thinking: how histone modifications regulate gene expression. Trends Genet 32(1):42–56
59. Rothbart SB, Strahl BD (2014) Interpreting the language of histone and DNA modifications. Biochim Biophys Acta 1839(8):627–643

4 Cell-Type Heterogeneity in DNA Methylation Studies: Statistical Methods and Guidelines

Andrew E. Teschendorff

Contents

Abstract

Studies in epigenetic epidemiology have reported increasing numbers of epigenetic biomarkers associated with a wide range of exposures and outcomes. Due to cost and technical difficulties, these markers are usually derived from complex

A. E. Teschendorff (✉)
CAS Key Laboratory of Computational Biology, Shanghai Institute for Nutrition and Health, Chinese Academy of Sciences, Shanghai, China

UCL Cancer Institute, University College London, London, UK
e-mail: andrew@picb.ac.cn

K. B. Michels (ed.), *Epigenetic Epidemiology*,
https://doi.org/10.1007/978-3-030-94475-9_4

tissues that are composed of many different cell-types. This cell-type heterogeneity prevents the identification of cell-type specific epigenetic alterations, posing significant challenges to the interpretation and understanding of these markers. Consequently, there is a strong need to develop cost-effective computational solutions to tackle the cell-type heterogeneity problem. Here, I discuss some recently proposed cell-type deconvolution algorithms aimed at estimating cell-type fractions and identifying cell-type specific differential DNA methylation changes. I describe their successful application to epigenome studies. We also discuss their main limitations, providing general guidelines for their successful implementation and for correctly interpretating results derived from them.

Abbreviations

DMC	differentially methylated cytosine
DMCT	differentially methylated cell-type
DNAm	DNA methylation
EWAS	Epigenome-Wide Association Study
FDR	False Discovery Rate
FPR	False Positive Rate
LSR	least squares regression
mQTL	methylation quantitative trait loci
PR C2	Polycomb-Repressive-Complex-2
scRNA-Seq	single-cell RNA-Seq
SE	Sensitivity

4.1 Introduction

Over the last two decades, we have seen a rapid increase in the number of studies reporting associations of epigenetic marks, in particular DNA methylation (DNAm), with epidemiological and disease risk factors, as well as with disease itself [1–10]. Many of these associations have been derived by measuring DNA methylation in the tissue-of-origin, for example, in precursor cancer lesions [8, 11, 12], cancer-tissue [13] or post-mortem brain [14, 15]). However, by far most associations have been derived from easily accessible "surrogate" tissues like blood [1, 16], often under the assumption that DNAm changes in such tissues can be informative of disease or disease risk [17–20]. An ever-increasing number of epigenetic biomarker studies are also measuring DNAm of cell-free DNA fragments (cfDNAm) in serum, which offers great potential for noninvasive early detection of a wide range of diseases, including type-1 diabetes, multiple sclerosis, ischemic brain damage, pancreatitis, and cancer [21–26]. In most cases, however, the biological interpretation of the measured epigenetic alterations remains challenging [27]. One obvious

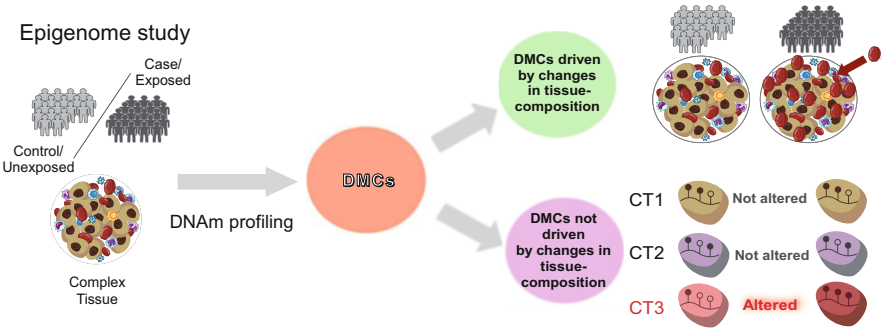

Fig. 4.1 Broad classes of DMCs. Epigenome studies performed in complex tissues (by definition these are composed of many different cell-types) only measure an average DNAm profile, and thus can only detect differentially methylated cytosines (DMCs) without knowledge of the underlying sources driving these DNAm changes. DMCs may result from shifts in cell-type proportions between cases and controls, or may result from DNAm changes in one or more cell-types (CT). All types of DMCs may be informative of say diagnosis or prognosis, but detection of cell-type specific DMCs is important to improve our understanding of the molecular pathways involved in disease development

reason for this is that, unlike gene or protein expression, an epigenetic change may not necessarily be functional. As a concrete example, age-associated DNAm alterations have been widely reported [6, 7, 28–31], yet a significant proportion of these changes do not appear to be functional or may only act to stabilize gene expression [32]. Another reason, which is still often overlooked, is cell-type heterogeneity [33]. This refers to the fact that most epigenome studies deriving biomarkers or molecular classifications of disease, do so by measuring DNAm in a complex tissue comprised of many different cell-types, each with its own characteristic DNAm profile. This is true not only for tissues like lung, liver, or blood, but also when measuring cfDNAm in serum, as most of the cfDNA derives from lymphocytes [21, 26]. Thus, by only measuring an average DNAm profile over many underlying cell-types, it is difficult to ascertain if DNAm changes associated with an exposure or outcome of interest is the result of DNAm changes in individual cell-types, and if so, in which cell-types, or whether the DNAm change is merely the result of underlying changes in cell-type proportions (Fig. 4.1).

Is it important, from a biological, clinical, or epidemiological perspective, to determine the source or nature of a DNAm change? In general, the answer to this question is yes: knowing in which cell-type (or cell-types) a DNAm change may be occurring in is critically important in order to understand how putative functional consequences of the DNAm changes may affect cell-function and disease development. For instance, in the context of asthma, an EWAS for immunoglobulin-E concentrations in blood revealed associations that were later validated in isolated eosinophils, highlighting molecular pathways in a relevant cell-type that mediate allergic inflammation [34]. Another example is that of HAND2, a transcription factor that mediates the tumor-suppressive effects of progesterone in the endometrium: here promoter hypermethylation and silencing of HAND2 is observed in endometrial fibroblasts from precursor cancer lesions, which results in increased

FGF paracrine signaling with endometrial epithelial cells, sensitizing these to oncogenic estrogen [35]. Another scenario where cell-type specificity may play an important role is in the context of mQTLs, i.e., SNP-CpG pairs where the methylation of the CpG is associated with genotype, with recent studies indicating that up to 20–30% of mQTLs derived in blood may be cell-type specific [36, 37]. Such cell-type-specific mQTLs could be informative of molecular pathways that contribute to disease predisposition in a cell-type-specific manner. Another example is aging, where the promoter of a given cell-specific transcription factor may become gradually hypermethylated with age, which could lead to irreversible silencing of the TF and to a skewed or impaired differentiation [38]. The same DNAm change occurring in a cell-type where the TF is already switched off, say by a repressive H3K27me3 mark, would not have a functional consequence.

One solution to the cell-type heterogeneity problem would be to perform the epigenome study in purified cell-types, but this remains labor intensive, costly and has mostly only been carried out in immune cells [36, 39–41], or to validate findings at a few selected loci [34]. Another potential solution is to generate DNAm data at the resolution of single cells [42–44]. However, this also remains costly, is unscalable to large numbers of individuals, and only generates very sparse data that cannot be used for building DNAm reference profiles for individual cell-types [44, 45]. Thus, there has been an ongoing attempt to address the cell-type heterogeneity problem computationally, by devising statistical algorithms (i) that can infer cell-type fractions and DMCs [46–48], (ii) that can infer latent components of variation and DMCs [49–51], and (iii) that can infer cell-type specific differential DNAm, i.e., differentially methylated cell-types (DMCTs) [52, 53] [54, 55], all in the context of epigenome studies performed in complex tissues.

The computational complexity posed by the cell-type heterogeneity problem can be quite substantial: assuming a study is performed in whole blood (a tissue with seven main blood cell subtypes: neutrophils, eosinophils, monocytes, natural-killer, B-cells, CD4+, and CD8+ T-cells), and that one finds a DMC associated with a factor of interest, there could be at least $2^7-1 = 127$ different combinatorial differential methylation (DM) patterns among the seven cell-types that could give rise to the observed DM at the whole tissue level. For instance, at one extreme, a DMC could be present in all cell-types of the tissue, while at another, it may only be present in one of the cell-types (Fig. 4.1). The above estimate does not even consider the potential combinatorial possibilities in terms of the directionality of DNAm change, which could involve increased (i.e., hyper) or decreased (i.e., hypo) methylation in different cell-types. Nor does the above estimate include scenarios where the DMC is driven purely by a change in cell-type composition. A well-known example of the latter is the increased myeloid to lymphocyte ratio that is observed in blood as a function of age [7, 56], cancer [57] or Rheumatoid Arthritis (RA) [58]. Shifts in cell-type composition are of course critically important in the context of cfDNAm in serum, since it is these shifts (e.g., increased circulating tumor DNA burden) that are informative of disease [26]. In other easily accessible tissues such as saliva or buccal swabs, the number of cell-types will be even bigger than in blood, because these tissues contain squamous epithelial cells besides immune cells [59–61]. The complexity only increases even further when we start to consider solid tissues, which in

addition to immune cells, may contain different types of epithelial, endothelial, and fibroblast cells. The full repertoire of cell-types within human tissues and organs is only now been elucidated thanks to major international efforts such as the Human Cell Atlas [62, 63]. For instance, some studies have estimated over 50 different cell-types in a tissue like lung [64, 65]. It should be noted though that given an observed magnitude of DNAm change, that this may impose substantial constraints on the allowed cell-type-specific DM patterns. For instance, if we observe a close to 100% change in DNAm between cases and controls, then this can only be realized if the change is happening unidirectionally in all the major cell-types of the issue. In general, it should be clear that the complexity of calling cell-type-specific DM can be at least 100-fold higher compared to calling DM [66].

Given this complexity, and given the inevitable limitations on the sample size of epigenome studies, it is understandable that statistical algorithms alone may not be able to fully address the above challenge. Nevertheless, as we shall see, statistical and computational methodology can help towards partial solutions or to solving the challenge in simpler scenarios, which can still be very informative and useful for disease diagnosis and early detection, for hypothesis generation or for devising validation experiments in purified cell-types. For instance, one way to simplify the problem is to consider a small number of "coarse" or "representative" cell-types, or only restrict to the main dominant cell-types within a tissue.

In the next sections, we shall describe some of the computational and statistical methods that have been proposed to estimate cell-type fractions and to detect cell-type-specific DNAm changes, as well as a number of applications where it has led to important novel insights.

4.2 General Considerations

In that follows, we shall refer to a cell-type specific differentially methylated cytosine as a differentially methylated cell-type, abbreviating this with "DMCT." Given a DNAm matrix defined over cytosines and samples, and given a factor of interest (e.g., an exposure, disease-status), the inference of DMCTs generally proceeds in two steps:

1. First, we need to estimate the fractions for all cell-types in the tissue. Without knowledge of the proportions of each cell-type in a given tissue, it is not possible to infer in which cell-types putative DNAm changes are happening. Some cell-types in a tissue may be present in such low numbers that their estimation is not possible, and therefore in this step one aims to infer fractions for as many cell-types so as to account for at least 95% of the tissue composition. This is because the error-rate of estimating cell-type fractions generally is about 5% [48].
2. Second, the estimated cell-type fractions are used as covariates in a linear model relating the observed DNAm profile of a cytosine to a factor/exposure of interest and other covariates representing potential confounding factors (e.g., batch effects). Importantly, the cell-type fractions enter the equations not only as

ordinary covariates, but also as part of interaction terms with the factor/exposure of interest. Later we shall explain why interaction terms capture DMCTs.

While in general, only these two steps are required, in practice, it is strongly advised to include two intermediate analyses in order to better interpret the overall results. Once we have estimated cell-type fractions, it is important to establish if these cell-type fractions vary significantly in relation to our factor/exposure of interest. Indeed, if tissue composition changes, then the underlying shifts in cellular proportions could play an important causal role in disease development. A concrete example is the infiltration of CD8+ T-cells in triple-negative breast cancer, which is a well-known predictor of good outcome in this usually aggressive type of cancer [67]. Cell-type compositional changes could potentially also be used for disease risk prediction or diagnosis [21, 22, 68–70]. For instance, in the context of cfDNAm in serum, an increased proportion of DNA fragments reflecting tissue-specific cell-death, as e.g., it happens with pancreatic beta-cell death in type-1 diabetes or with exocrine cell-death in pancreatic cancer, could be used for early detection or monitoring of therapy response [26]. A shift in tissue composition can also reflect a systemic effect of disease, as for instance, with the observed increased myeloid to lymphoid ratio in the blood of cancer patients [57]. If such shifts occur before the conventional diagnosis of disease, they could be used for early detection or for quantifying disease risk.

The other intermediate step one is advised to perform is the inference of differentially methylated cytosines (DMCs). This is accomplished using the same linear model as for DMCT inference, but without the inclusion of interaction terms. In this model, cell-type fractions only enter as ordinary covariates alongside other potential confounders, aim being to identify DMCs that are not driven by changes in cell-type composition or by any of the other potential confounders. It is important to stress that once we have identified such a DMC, that this does not tell us anything about which cell-types the DNAm change is happening in, although there could be constraints on this depending on the observed effect size, as mentioned earlier. We shall see one concrete example of this later. Now we turn to the specific task of estimating cell-type fractions in complex tissues.

4.3 Estimating Cell-Type Fractions in Complex Tissues

As mentioned earlier, this task is required before we can infer DMCTs. And the reliability of the inference of DMCTs in a given study hinges on our ability to accurately estimate the underlying cell-type fractions. Currently, the best way to estimate cell-type fractions in a given sample for which a genome-wide DNAm profile is available, is through the construction of a DNA methylation reference matrix (DMRM) [71]: the columns of this matrix represent the cell-types within the tissue, with the rows representing marker CpGs (or marker genes) whose DNAm levels vary substantially between cell-types. Here we shall discuss three different strategies that have been applied to build a DMRM (Fig. 4.2). The choice of strategy largely depends on the tissue-type being considered, which is why the ensuing discussion is structured based on tissue-type.

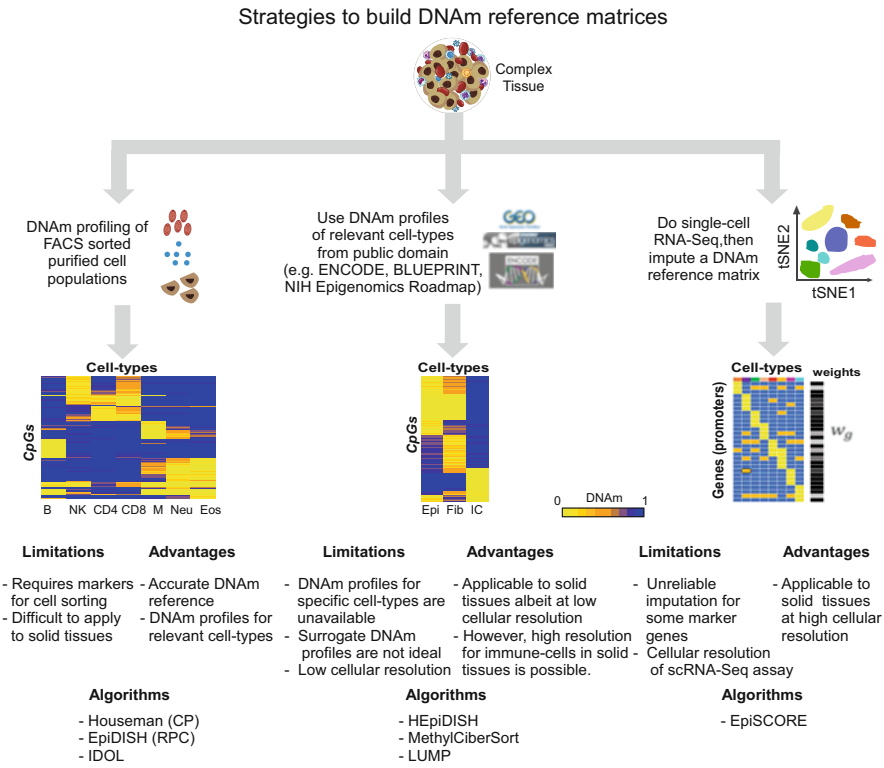

Fig. 4.2 Strategies to building a DNAm reference matrix. Depicted are three strategies to building a DNAm reference matrix (DMRM). One approach (left branch) is based on FACS sorting cells and subsequently generating genome-wide DNAm profiles for these purified cell populations. Subsequently, one identifies DMCs between these purified cell populations to build the DMRM. Another strategy (middle) is to use existing DNAm profiles of relevant purified cell samples in the public domain to build the DMRM, once again by identifying DMCs between these purified samples. This strategy can be applied more broadly to solid tissues, but compromising cellular resolution. Another strategy (right) is to leverage the high-resolution nature of a tissue-specific scRNA-Seq atlas to build an mRNA expression reference matrix and to subsequently impute a corresponding tissue-specific DMRM. The cellular resolution of this strategy is high (only limited by the cell-types the scRNA-Seq assay can measure), but the quality of the imputation may not suffice to distinguish similar cell-types

4.3.1 Estimation of Cell-Type Fractions in Blood

We first consider the case of whole blood (WB), peripheral blood mononuclear cells (PMBCs) and cord blood (CB), because for these tissues, it is possible to generate genome-wide DNAm reference profiles for all major underlying cell-types. This can be accomplished by the use of well-known markers for blood cell subtypes, whose specificity is relatively high in order to be able to generate purified samples (e.g., CD19 for B-cells) through, e.g., FACS sorting. Thus, in the case of WB, it has been possible to generate genome-wide DNAm profiles for all seven major cell subtypes,

which includes neutrophils, eosinophils, monocytes, B-cells, CD4+ T-cells, CD8+ T-cells, natural-killer cells, and B-cells [72]. PBMCs are generally devoid of granulocytes which includes neutrophils, eosinophils, and basophils. For CB, a tissue of increasing importance in epigenetic epidemiology [73–77], it has also been possible to build tailored DNAm reference matrices [78, 79], as indeed it is well recognized that both the composition and cell-type-specific DNAm profiles are different to those of whole blood. In all these cases, once genome-wide DNAm profiles for purified samples have been generated, it is then possible to build corresponding DMRMs.

It is important to briefly describe how the DMRM should be constructed and indeed why it is necessary to use this DMRM for estimating cell-type fractions. If we have a genome-wide DNAm profile for a WB/PBMC/CB sample, represented by a vector $\vec{\beta}$ defined over a large number of CpGs (around 450k or 850k, depending on the Illumina beadarray version), then it is sensible to express it as a linear mixture of corresponding genome-wide DNAm profiles for the constituent cell-types, as in:

$$\vec{\beta} = \sum_{k=1}^{K} f_k \vec{\beta}_k,$$

where k denotes the cell-type, K is the total number of different cell-types and f_k is the proportion/fraction of cells of type k in the mixture. These fractions are obviously unknown, and we would like to infer them given the measured DNAm profile of the mixture and that of the constituent cell-types. Mathematically, the inference of these fractions is indeed possible, by formulating the above equation as a linear least squares multivariate regression problem, with one key difference, however, which is that the fractions must all be bounded between 0 and 1, and that their sum must add to 1 (or to a number less than 1 since in practice we cannot possibly know or have DNAm profiles for all underlying cell-types in a tissue). These additional constraints on the regression problem turn the ordinary least squares regression (LSR) into a constrained LSR or constrained projection (CP) problem, which can be solved using techniques in quadratic programming (QP) [46]. While it might be tempting to solve this CP problem using all available CpGs, this is not advisable because a large chunk of the genome will not differ between blood cell subtypes. In other words, uninformative CpGs that show little variability between blood cell subtypes generally do not contribute, and therefore it is sensible to exclude these from the inference as they are not needed. Computationally, if the CP problem can be solved over a small number of informative CpGs this will also speed up the estimation of cell-type fractions very substantially, which is an important consideration. Thus, once the DNAm profiles for the purified cell-types have been generated, the next step is to identify the informative CpGs. There are several strategies to accomplish this [46, 48, 78, 79], but in general, they involve a process of identifying cell-type specific DMCs, i.e., CpGs that are highly methylated (or unmethylated) in one cell-type, with correspondingly low (or high) DNAm values in all other cell-types. Ideally, one desires a reasonable number of cell-type specific DMCs for each cell-type, which ensures robustness to potential "dropouts" in independent datasets where the fractions are to be estimated. By dropout we here mean a probe that does not pass QC in the

independent dataset and which therefore cannot be used in the inference procedure. Ideally, one would also want to ensure equal or approximately equal numbers of cell-type-specific markers for each cell-type, because can avoid biasing the inference to those cell-types with more markers. And thirdly, one would ideally only include cell-type-specific DMCs, where the difference in DNAm between the marker cell-type and all others is as large as possible, i.e., typically over 0.7 in the beta value [0,1] scale, i.e., at least a 70% methylation change. This number is motivated by the heavy use of Illumina DNAm beadarray technologies, where the two main modes (unmethylated and methylated peaks) are generally about 0.7 methylation units apart, with a value of 1 being the theoretical maximum. In practice, all three requirements above may not be met. For instance, for similar cell-types (e.g., CD4+ and CD8+ T-cells) it might be difficult to find many DMCs with over 70% DNAm difference between them, and indeed accurately estimating the relative fractions of highly similar cell-types still remains a challenge. Mathematically, the quality of a DNAm reference matrix can be expressed in terms of its conditioning number [79], which is an indicator of how robust the inference would be. Based on extensive experience estimating cell-type fractions, we advise on the following guidelines for selecting DMCs for a reference DNAm matrix. We note that these guidelines are not unique to DNAm but apply equally well to other data types (e.g., RNA-Seq):

1. Perform DM analysis between one cell-type (the "marker" cell-type) against all others (as one group) to identify DMCs using some sensible significance threshold (e.g., FDR < 0.05).
2. For these DMCs, compare the mean DNAm level in the marker cell-type to the highest (or lowest) mean DNAm level from among all other cell-types, to then subselect DMCs that have an effect size (i.e., difference in mean DNAm) larger than +/−0.7.
3. For each cell-type, rank these DMCs according to the absolute effect size and select a given top number of these. We recommend about 50 for each cell-type, if possible.
4. If in step-2 or step-3 there are not enough DMCs for a given cell-type, then the corresponding effect size threshold could be relaxed (for instance, one could go as low as +/−0.3 or +/−0.4), or alternatively, one may select a smaller number of top-ranked DMCs in step-3. However, as mentioned, due to QC-issues in independent data, we do not recommend that the final number of cell-type specific DMCs per cell-type should be less than 10.

Once the cell-type-specific DMCs have been identified, the DMRM is then built by taking the average or median DNAm of these CpGs in each cell-type separately. We note that although theoretically, one should take the average, in practice, taking the median is also justified as it may provide a more robust estimator. For K cell-types, the DMRM would then ideally be a matrix with $K*50$ rows labeling the cell-type-specific DMCs, and K columns labeling the cell-types. With this DMRM in place, we would then estimate cell-type fractions by solving the following CP problem:

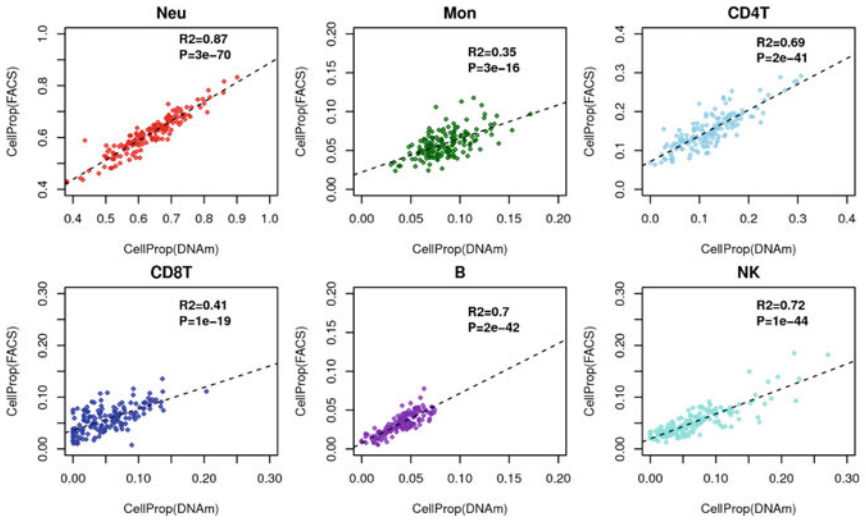

Fig. 4.3 Agreement between DNAm and flow cytometry-based cell fraction estimates in whole blood. For each of the six major blood cell subtypes, a scatterplot of the flow cytometric cell-type fraction estimate (y-axis) vs. the DNAm-based estimate (x-axis). The dashed lines represent the best fit regression line. R^2 and P-values are given. Dataset consists of 162 whole blood samples from healthy individuals

$$\vec{\beta} = \sum_{k=1}^{K} f_k \vec{\beta}_k^{(R)} + \vec{\epsilon} \text{ with}$$

$$0 \leq f_k \leq 1 \text{ and } \sum_{k=1}^{K} f_k \leq 1$$

DMRMs for blood and cord blood are available from various Bioconductor R-packages, including *minfi* [80], *EpiDISH* [48] and *FlowSorted.CordBlood.450 K* [81].

In the case of tissues like blood, it is also possible to obtain cell-type fractions using magnetic flow cytometric techniques (e.g., MACS), which thus allows objective testing of the DNAm-based estimates. This assumes of course that the sample taken for the DNAm-assay and the sample used to obtain MACS-based estimates are taken at the same timepoint, as blood cell fractions are known to vary with time [82–84]. In general, the agreement of estimated cell-type fractions obtained using a DMRM with those derived with MACS is very good (Fig. 4.3).

4.3.2 Estimation of Cell-Type Fractions in Saliva and Buccal Swabs

Like blood, saliva and buccal tissue can be obtained fairly easily and cheaply, allowing noninvasive measurement of DNA methylation in a tissue that contains squamous epithelial cells in addition to immune cells [59–61]. The reason for

treating saliva and buccal swabs separately from blood and solid tissue-types is that it is relatively straightforward to build DMRMs for these two tissue-types. Since they contain immune cells, the same DNAm profiles of immune cell subtypes in blood can be used for building the DMRM for saliva/buccal swabs. The only additional requirement is to generate a DNAm reference profile for squamous epithelial cells. However, since effectively most epithelial cells in saliva/buccal swabs are squamous in origin, and there is therefore no need to distinguish different epithelial cell-types from each other, a DNAm reference profile for any epithelial cell-type may suffice. For instance, ENCODE [85] and the NIH Epigenomics roadmap [86, 87] have generated genome-wide DNAm profiles for many different normal epithelial cell lines and primary epithelial cells. By using different types of epithelial samples one can thus hone in on the generic DNAm patterns of any epithelial sample, including the squamous epithelial cells from the oral cavity. Thus, for these tissues and given genome-wide DNAm profiles for epithelial and immune cell subtypes, one can apply the same DMRM-construction strategy outlined above, i.e., by first deriving cell-type-specific DMCs, from which the DMRM is then built. Estimation of cell-type fractions then proceeds exactly as described before for the case of blood.

A slight variation to the above procedure is obtained by recognizing that cell-types within a tissue exhibit markedly different levels of similarity, depending on their shared developmental trajectories. Thus, one expects substantially more DMCs between epithelial and immune cells than between CD4+ and CD8+ T-cells. In recognition of this, an algorithm called HEpiDISH (Hierarchical Epigenetic Dissection of Intra-Sample Heterogeneity) [59] uses a hierarchical 2-step strategy with two DNAm reference matrices, one aimed at obtaining a total epithelial and total immune cell fraction, and a second one to obtain fractions for all immune cell subtypes. Because we will revisit HEpiDISH in the context of solid tissue-types, we refer details to the next subsection.

4.3.3 Estimation of Cell-Type Fractions in Solid Tissues

The most challenging scenario is to estimate cell-type fractions in a solid tissue-type. This is because the number of cell-types is obviously much higher than in blood (there will be different kinds of fibroblasts, endothelial, and epithelial cell subtypes, besides immune cells), but also because this number is still generally unknown and generating DNAm profiles for even just a few of these cell-types is technically challenging. This means that for most cell-types in a solid tissue, there are no available DNAm reference profiles. This is despite efforts from the Epigenomics Roadmap [86] to generate a number of these (e.g., kidney podocytes, breast myoepithelial, breast luminal cells). Given these major challenges, how does one then aim to estimate cell-type fractions in a solid tissue like breast or lung where the number of cell-types will be substantially higher than 10 and for which reference DNAm profiles may not be available for specific cell-types?

In principle, one may be tempted to apply reference-free cell-type deconvolution algorithms [49, 50, 88], which aim to infer latent (i.e., hidden/unknown) sources of

variation in the data, some of which may correspond to variations in cell-type fractions. However, reference-free algorithms cannot yield direct cell-type fraction estimates and are therefore limited as far as DMCT inference is concerned, their application being primarily to infer DMCs not driven by changes in cell-type composition [49]. Given that biotechnology will eventually advance to a level that makes single-cell DNA methylomics more reliable, affordable, and scalable, thus allowing routine generation of DNAm reference matrices, it is sensible to focus on reference-based approaches as we are doing here.

So far, two different reference-based strategies have been developed. One approach is based on the HEpiDISH algorithm [59], mentioned earlier. This uses a two DMRM strategy, whereby in the first step, the algorithm uses one DMRM to estimate a total epithelial, a total fibroblast and a total immune cell fraction. While solid tissues like lung or breast contain other cell-types (e.g., endothelial cells), the lack of sufficient reference DNAm profiles for purified endothelial cells (to allow for training and validation) means that these may be challenging to include. In the case of breast, adipocytes form a major component in addition to epithelial, fibroblast, and immune cells, and for adipocytes, reference DNAm profiles are available in the public domain to allow construction and independent validation of a 4 cell-type DNAm reference matrix, defined over a generic epithelial, fibroblast, fat, and immune cell [59]. In the second optional step, HEpiDISH then estimates fractions for the immune cell subtypes. This is accomplished with a second DMRM defined over CpGs that discriminate the different immune cell subtypes from each other. Importantly, for each of these CpGs, it is also required that their baseline DNAm level, i.e., the level of DNAm seen across most of the immune cell subtypes, is similar to that of the epithelial, fibroblast, and fat cells, to ensure that variations in these other cell-type fractions would not bias the relative fractions of immune cell subtypes [59]. Thus, this second DMRM is distinct from the one used when inferring cell-type fractions in WB.

Another strategy is based on the EpiSCORE algorithm [89, 90]. EpiSCORE leverages the high-resolution nature of a tissue-specific scRNA-Seq atlas to first construct a mRNA expression reference matrix for all major cell-types in the tissue. This assumes that the scRNA-Seq assay has captured the most important cell-types in the tissue, which is not always the case: for instance, in the case of breast, fat cells are abundant but they are large cells that are often missed by current scRNA-Seq protocols [91]. Given the mRNA expression reference matrix, EpiSCORE then imputes a corresponding DMRM, defined over the promoters of a subset of the expression marker genes and the same number of cell-types. Of note, this imputation only works for 20–30% of the marker genes in the expression reference matrix, because for most marker genes there is no strong anti-correlative pattern between promoter DNAm and gene expression. The smaller number of "imputable" marker genes, for which there is such a strong anti-correlation, are identified using matched RNA-Seq and whole-genome bisulfite sequencing (WGBS) data from resources like the NIH Epigenomics Roadmap [86] and ENCODE [92, 93]. Unlike HEpiDISH, EpiSCORE can then yield cell-type fractions for all cell-types in the tissue in one step, by applying an analogous multivariate linear model as the one described earlier for blood.

4.3.4 Estimation of Cell-Type Fractions from cfDNAm in Serum

For completeness, we also describe briefly methods for estimating cell-type fractions from cfDNAm in serum, framing this in the context of cancer where such methods offer particular promise. Broadly speaking, the underlying strategy has been to build analogs of DMRMs by identifying informative CpGs (i.e., DMCs) from comparisons between DNAm measured in serum or blood from healthy individuals and DNAm measured in solid cancer-types. Since the cfDNA in plasma from healthy individuals stems predominantly from lymphocytes, such differential DNAm analyses naturally adjust for the immune cell contamination that is inevitable present in solid cancers [79, 94], thus identifying DNAm changes that are unique to the non-immune cells (e.g., epithelial, fibroblast cells) present in cancer-tissue. An algorithm called Cancer Locator [95] then builds a DMRM defined over a healthy and multiple cancer-type states, subsequently applying this DMRM in a beta-mixture modeling maximum likelihood framework to infer both tumor fraction and tumor-type. An important element of this approach has been the explicit modeling of variance in addition to the mean when building the DMRM [95]. Another strategy has been to perform targeted bisulfite sequencing of informative regions in case/control cohorts, using Machine Learning (ML) methods on a training set to build predictors of cancer-type, which are subsequently validated in independent plasma samples [21, 22]. Here, the ML method returns a probability measure that an individual has a particular cancer-type, which can be interpreted roughly as a relative measure of the corresponding tumor cell burden in the serum sample. Overall, these strategies have shown that sensitive detection, i.e., with sensitivities and specificities close to 90% or higher, is possible for a wide range of common tumor types (e.g., lung, liver, esophageal cancer) [21, 95], in some cases even up to 4 years before conventional diagnosis [22]. In this regard, it is worth noting that the cell-type specificity and meta-stability of DNAm offer significant advantages over other molecular data types in distinguishing the tissue-of-origin, as demonstrated by many studies (see e.g [96–99].). However, challenges remain in that informative features are not yet selected optimally. For instance, given that DNAm changes are widely altered and shared between cancer-types [100, 101], or that they could reflect alterations in other cellular compartments (e.g., fibroblasts or endothelial cells), it is plausible that these confounders could limit performance when validating in independent cohorts. Thus, cell-type deconvolution methods for solid tissues as described earlier should be the ideal starting point in which to identify an optimal set of informative DMCs that are unique to the cells of origin of each cancer-type.

4.4 Inferring Cell-Type-Specific Differential DNA Methylation (DMCTs)

Once we have inferred the cell-type fractions, we are now in a position to infer DMCTs. Here we shall describe one proposed DMCT-calling algorithm, the CellDMC algorithm [52]. Another very similar algorithm that subsumes CellDMC is TOAST [54]. Assuming for the time being that a CpG is altered in only one cell-

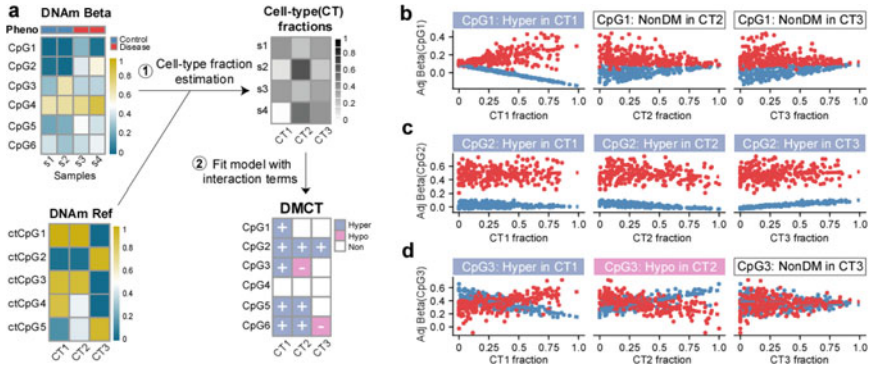

Fig. 4.4 Identification of DMCTs. (**a**) Overall strategy to infer DMCTs involves a step where we infer cell-type fractions for each sample, followed by fitting a linear model with interaction terms between phenotype and cell-type fractions to infer a map over CpGs and cell-types indicating which CpGs are altered in which cell-types. (**b–d**) Three examples of CpGs that define DMCTs in one cell-type (CT1) (**b**), all cell-types (CT1–3) (**c**) and two cell-types where the direction of DNAm change is different (CT1–2) (**d**). The scatterplots display the adjusted DNAm beta value against the corresponding cell-type fraction, with red datapoints indicating "cases," blue datapoints indicating "controls." Hyper = hypermethylated in cases, Hypo = hypomethylated in cases

type of the tissue, CellDMC/TOAST is based on the intuitive notion that the difference in DNAm between case/control status (for convenience we here assume a binary factor of interest, but similar arguments apply to a factor of interest that is continuously-valued, e.g., age or smoking exposure), would be bigger in those samples where the altered cell-type is more abundant. At the other extreme, i.e., in samples where the altered cell-type is only present in small numbers or not present at all, the difference in DNAm will correspondingly be much smaller. Extending this to more general scenarios, it is therefore plausible that by studying the patterns of DNAm change as a function of cell-type fractions, one can identify DMCTs (Fig. 3.4 to Fig. 4.4).

Statistically, the dependence of DNAm change with cell-type fractions can be captured using linear interaction terms, in this case, interactions between the factor of interest and cell-type fractions. From a modeling perspective, these interaction terms also emerge naturally: for a given cytosine c in a sample s, the methylation value β_{cs} will be given by the formula:

$$\beta_{cs} = \sum_{k=1}^{K} \widehat{f}_{ks}\beta_{cks},$$

where $\widehat{f}_{ks}$ are the estimated cell-type fractions and where β_{cks} denotes the DNAm value in cell-type k in sample s. It is natural to assume that this value follows a linear model in relation to the factor of interest (denoted here by y),

$$\beta_{cks} = \alpha_{ck} + \gamma_{ck} y_s + \varepsilon_{cs},$$

which means that the expression above becomes

$$\beta_{cs} = \sum_{k=1}^{K} \alpha_{ck} \widehat{f}_{ks} + \sum_{k=1}^{K} \gamma_{ck} \left(\widehat{f}_{ks} y_s \right) + \varepsilon_{cs},$$

where ε_{cs} is a Gaussian error term. The second term in brackets is the linear interaction term between cell-type fraction and our factor of interest. If there are covariates or confounders that affect DNAm independently of cell-type (e.g., batch effects), and which we represent as W_q, it is easy to show that the above formula becomes

$$\beta_{cs} = \sum_{k=1}^{K} \alpha_{ck} \widehat{f}_{ks} + \sum_{k=1}^{K} \gamma_{ck} \left(\widehat{f}_{ks} y_s \right) + \sum_{q=1}^{Q} \varphi_{cq} W_{qs} + \varepsilon_{cs}$$

We note that here the cell-type fractions have already been estimated and that their sum must add to 1, which means that implicit in the equation above, there is an intercept term, and a linear term in y (without interactions). The regression coefficients can be solved under a linear LSR, which returns t-statistics and P-values for the interaction terms, i.e., for assessing if there are DMCTs or not. It is important to note that the above model is a marginal conditional model, i.e., one aims to determine if a CpG is altered in a given cell-type conditioned on all other cell-types. This is to be contrasted with the marginal (unconditional) model for cell-type k

$$\beta_{cs} = \sum_{k=1}^{K} \alpha_{ck} \widehat{f}_{ks} + \gamma_{ck} \left(\widehat{f}_{ks} y_s \right) + \sum_{q=1}^{Q} \varphi_{cq} W_{qs} + \varepsilon_{cs},$$

where we estimate an interaction between the factor and the given cell-type fraction, ignoring

all other potential interaction terms. Incidentally, the model for inferring ordinary DMCs, would be

$$\beta_{cs} = \sum_{k=1}^{K} \alpha_{ck} \widehat{f}_{ks} + \gamma_c y_s + \sum_{q=1}^{Q} \varphi_{cq} W_{qs} + \varepsilon_{cs}$$

Later we present a power calculation to indicate that the marginal conditional model can display in certain scenarios limited sensitivity to detect DMCTs that occur in all cell-types. Thus, our guideline is the following 2-step strategy:

1. First, we infer ordinary DMCs adjusting for cell-type fractions, i.e., by adding cell-type fractions as covariates in the linear model. This ensures that DMCs are not driven by changes in cell-type proportion, while also keeping high sensitivity to detect most alterations, albeit without knowledge of which specific cell-types are altered.
2. Second, we run the marginal conditional model to identify DMCTs. Any DMCs called in step-2, which are not found to be DMCTs, most likely indicate CpGs that are altered in all cell-types.

4.5 Applications

We next describe a number of concrete examples where the CellDMC algorithm has been applied to real epigenome datasets, not only validating the algorithm, but also demonstrating how novel insights can be gained by inferring DMCTs as opposed to just DMCs.

4.5.1 Identification of Smoking-Associated DMCTs in Buccal Swabs

A compelling way to validate the CellDMC algorithm is by performing an EWAS in relation to smoking exposure in buccal swabs. To understand why, we first note that buccal swabs contain on average about 50% immune cells and 50% squamous epithelial cells [59], i.e., two main cell-types. Second, many smoking-EWAS performed in blood have shown that there is a highly reproducible signature of smoking-associated DNAm changes. For instance, a recent meta-analysis identified a gold-standard set of 62 smoking-associated CpGs, the great majority of which undergo hypomethylation in the blood of smokers [2]. This includes a well-known CpG mapping to the repressor of the aryl-hydrocarbon receptor (*AHRR*). A more recent study encompassing a much larger number of samples [102] further validated these 62 smoking-DMCs, while also identifying a larger number of alterations. Thus, it is reasonable to posit that in a smoking-EWAS performed in buccal swabs, that we should be able to retrieve this 62 CpG smoking hypomethylation signature in the immune cell compartment of the tissue. The result of applying CellDMC to an EWAS of 790 buccal swabs [18] is shown in Fig. 4.5a, which demonstrates that the algorithm indeed predicts most of the 62 CpGs to be hypomethylated in the immune cells from smokers. Interestingly, the algorithm also predicts that these alterations are not happening in the squamous epithelial compartment of buccal swabs, suggesting that these DNAm alterations do not occur in the cells of origin of smoking-related diseases like squamous cell carcinoma of the lung or head and neck. Instead, it has been shown that the DMCTs occurring in the epithelial compartment may be more relevant in mediating the risk of smoking in these cancer-types [18, 59, 103, 104].

4.5.2 HAND2 Hypermethylation in Endometrial Cancer

Another insightful application is to endometrial cancer. Endometrial tissue is composed mainly of epithelial, fibroblast, and immune cells. As mentioned in an earlier section, the promoter of HAND2, a transcription factor that mediates the tumor-suppressive effects of progesterone, is hypermethylated in the endometrial fibroblasts of precursor lesions, as well as in endometrial cancer itself, events that are associated with silencing of HAND2 [35]. HAND2's promoter also undergoes hypermethylation in blood as a function of age [32]. Thus, application of CellDMC to say the TCGA endometrial cancer study (403 cancers +46 normal-adjacent) [105]

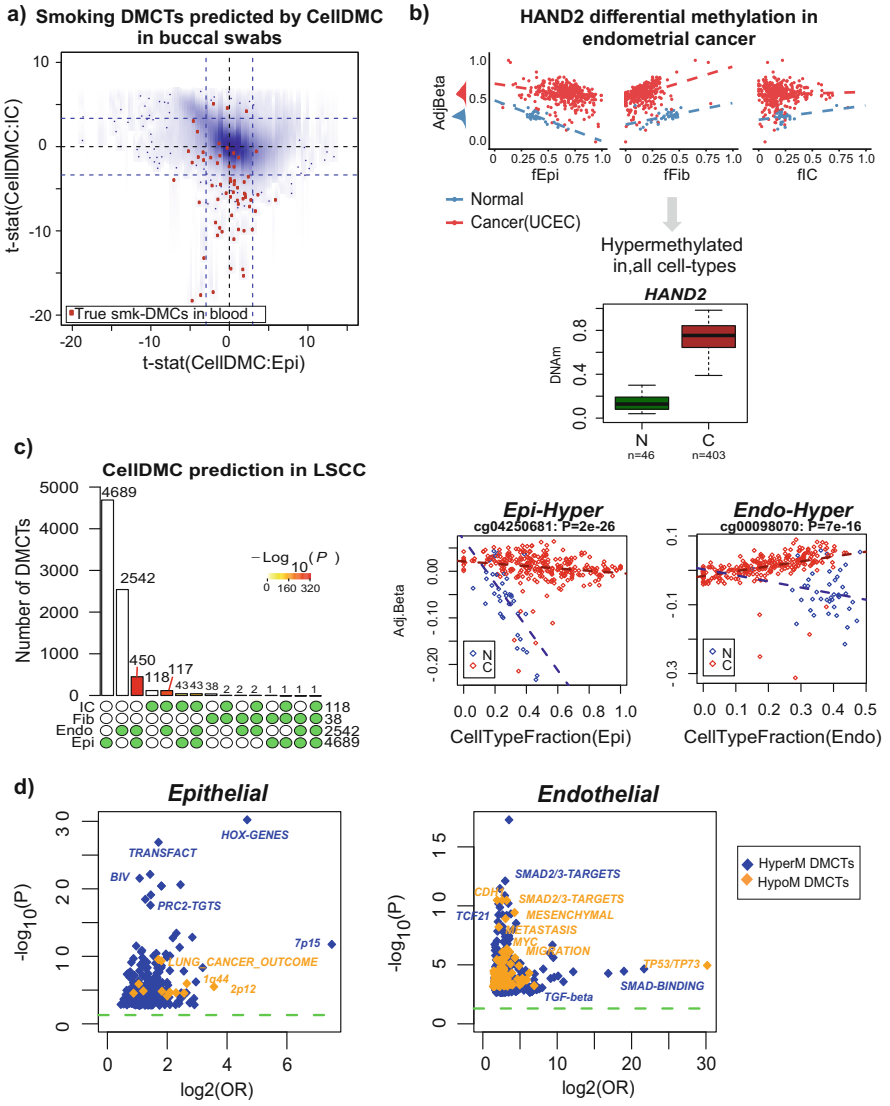

Fig. 4.5 Examples of DMCTs. (**a**) Scatterplot of *t*-statistics of association between smoking and DNA methylation (~480,000 CpGs) in immune cells (*y*-axis,IC) vs. the corresponding statistics in the epithelial cells (*x*-axis, Epi), as predicted by CellDMC in an EWAS of buccal swabs ($n = 790$ samples). CpGs belonging to a known 62 CpG hypomethylation smoking signature are highlighted in red. (**b**) Patterns of cell-type-specific differential DNAm for HAND2 (promoter & first exon region) in endometrial cancer, as predicted by CellDMC. Boxplot on the lower right corner is the differential DNAm pattern for whole tissue, i.e., averaged over all cell-types. (**c**) Distribution of DMCTs in an application of CellDMC to lung squamous cell carcinoma (LSCC) (TCGA dataset). Scatterplots at the bottom illustrate two DMCT examples: an epithelial and an endothelial-specific hypermethylated CpG. N = normal, C = cancer. (**d**) Scatterplots summarizing results of the GSEA of epithelial and endothelial-specific DMCTs. *y*-axis labels the statistical significance ($-\log_{10}$[*P*-value]), *x*-axis labels the $\log_2$[OddsRatio]. *P*-value and OR computed from a Fisher-test

should reveal that HAND2 is hypermethylated in fibroblasts and possibly also in the immune cells present within the endometrial tissue, because the average age of the 403 endometrial cancers is higher than that of the 46 normal-adjacent normals. Moreover, the average DNAm change between endometrial cancer and normal-adjacent samples is relatively large at around 0.6, a strong indication that HAND2 undergoes promoter hypermethylation in all main cell-types, including the epithelial compartment. Thus, it is not surprising that when applied to the TCGA dataset, at the resolution of three cell-types (epithelial, fibroblast, and immune cell) and using the corresponding DNAm reference matrix from HEpiDISH (see Fig. 4.2), that CellDMC predicts HAND2 promoter hypermethylation in all three cell-types (Fig. 4.5b). We note that in this instance, the marginal conditional model did have the sensitivity to detect HAND2 as changing in all three cell-types, despite being a scenario where an ordinary DMC-model is more appropriate. That the algorithm correctly predicts a DNAm change in the fibroblasts is critical, as it is the silencing of HAND2 in the fibroblasts that leads to an altered paracrine signaling with the epithelial cells, sensitizing these to oncogenic estrogen [35].

4.5.3 An Endothelial-to-Mesenchymal (EndoMT) Transformation Signature in Lung Cancer

An example of how novel insight can be gained through application of a tool like CellDMC is illustrated in the case of lung squamous cell carcinoma (LSCC). Here, the EpiSCORE algorithm (Fig. 4.2) was used to build a lung-specific DNAm reference matrix defined over epithelial, fibroblast, endothelial, and immune cells. While one could also consider different immune cell subtypes, we can refrain from doing so if our focus is on alterations occurring in the other cell-types. By simplifying the problem to four main cell-types also yields more power to detect changes in specific cellular compartments. The distribution of inferred DMCTs is displayed in Fig. 4.5c, together with a few examples, and a Gene Set Enrichment Analysis (GSEA) on the DMCTs appearing in the specific cellular compartments, which reveals patterns that are strongly consistent. For instance, the enrichment of bivalent and PRC2 marked domains among epithelial-specific hypermethylated DMCTs is a well-known universal cancer signature (Fig. 4.5d) [89]. Likewise, the concomitant enrichment of SMAD2/SMAD3 binding targets among endothelial-specific hypermethylated DMCTs, and of terms related to mesenchyme, migration and invasion among endothelial-specific hypomethylated DMCTs, is noteworthy given the role of SMAD2/SMAD3 in maintaining vascular integrity (Fig. 4.5d). Therefore, CellDMC makes the prediction that the observed hypermethylation in lung cancer endothelial cells is associated with disruption of SMAD2/SMAD3 binding, compromising vascular integrity and associated with an endothelial-to-mesenchymal (EndoMT) transformation. Such an EndoMT has been reported previously [106–108], but as revealed by CellDMC, epigenetic alterations could be critical in promoting such a transformation.

4.5.4 Myeloid and Lymphoid Specific Smoking-Associated DMCTs

As a final example, we revisit the case of smoking in blood. As mentioned earlier, smoking is associated with a highly reproducible DNAm signature in whole blood [2, 102]. However, to date, no large-scale EWAS in specific blood cell subtypes has been performed in relation to detecting smoking-associated DMCTs. Hence it is unknown whether the observed smoking signature is present in myeloid, lymphoid, or in both cell lineages together. Only smaller scale studies have converged on identifying a small number of CpGs that appear to be altered specifically in myeloid or lymphoid cells [109]. Specifically, Su et al. reported 7 CpGs that exhibit lineage-specific smoking-associated DNAm changes, with 5 of these being specific to myeloid cells, and 2 being specific to lymphocytes [109]. A recent study applied CellDMC to a number of independent EWAS with smoking information (2 in whole blood and 1 in buccal swabs), and at the resolution of 2 cell-types (myeloid and lymphoid), revealing good consistency with Su et al. [110]. A meta-analysis over a large number of EWAS in whole blood further revealed myeloid-specific hypermethylation and hypomethylation signatures, with the latter being strongly enriched for DNase Hypersensitive Sites (DHS) as defined in inflammatory macrophages, and with the former one only being enriched for DHSs as defined in acute myeloid leukemia (AML) [110]. In contrast, the same meta-analysis did not reveal an extended lymphocyte-specific smoking signature, suggesting that smoking imparts its effect on DNAm patterns in blood mostly via alterations in myeloid cells. This is a significant observation given that smoking is a moderate risk factor for AML but not for lymphocytic leukemias [111].

In summary, all the above examples illustrate successful applications of an algorithm such as CellDMC in identifying cell-type-specific DMCs, in some cases validating known DNAm signatures, and in other cases generating novel concrete hypotheses for further exploration and testing.

4.6 Limitations

Despite the successful applications described above, it is important to also emphasize the limitations associated with identifying DMCTs. We illustrate some of these limitations in the context of a power simulation on realistic data, focusing on the last example considered in the previous section, namely identifying smoking-associated DM in two cell lineages of blood (myeloid vs. lymphoid) [110]. This is a scenario where one cell-lineage (myeloid cells) accounts for most of the cells in the tissue (i.e., approximately 60–70% of cells in whole blood derive from the myeloid lineage), with variations in the myeloid fraction between individuals also accounting for most of the variation in DNAm. Details of the actual simulation model can be found in You et al. [110]. Very briefly, in this *in-silico* model, one simulates three classes of DMCTs, and subsequently assesses the sensitivity or power of the algorithm (in this case CellDMC) to detect these alterations. One class of DMCTs is specific to the lymphoid lineage, another class is specific to the myeloid lineage,

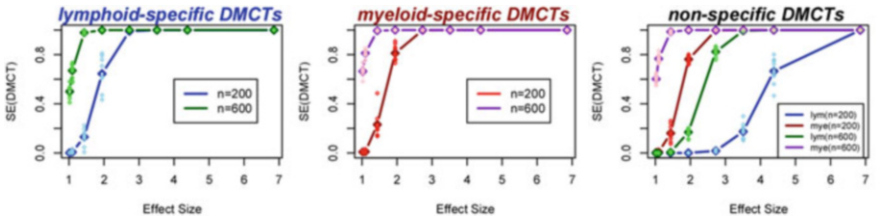

Fig. 4.6 Power calculation. Plots of the sensivity (*y*-axis) to detect each class of DMCT vs effect size (*x*-axis), and for two different sample sizes ($n = 200$ and $n = 600$). We note that n is the total sample size and number of cases and controls is assumed similar. The effect sizes are realistic and corresponding to average DNAm changes in individual cell-types that range from 0.05 to just over 0.6. In the last plot, we display the sensitivity to detect the non-specific DMCTs in each lineage separately

and the third class of DMCTs is present in both lineages, and therefore this class is non-specific. The sensitivity to detect each category of DMCT in a whole blood EWAS is shown in Fig. 4.6, for a range of different effect sizes and two different sample sizes.

Thus, for appropriately powered studies detecting lymphoid or myeloid-specific DMCTs is in general not a problem. For instance, for an epigenome study with 200 samples (100 cases and 100 controls), the sensitivity to detect myeloid-specific DMCTs is around 80% for an effect size of 2, which roughly corresponds to about a 10–20% average absolute DNAm change in myeloid cells. The corresponding sensitivity for lymphoid-specific DMCTs is just over 60%, i.e., lower than for myeloid-specific DMCTs, owing to the lower proportion and variance of lymphoid fractions in blood, but still reasonably high. For larger effect sizes, i.e., for average DNAm differences larger than say +/− 0.3, sensitivities are higher than 80 or 90% for both lineage-specific DMCTs. However, for non-specific DMCTs, the power calculation indicates markedly reduced sensitivity to detect the change in the lymphoid compartment. Thus, larger sample sizes would be needed to detect non-specific DMCTs as being non-specific. As far as the FPR and FDR are concerned, these are generally quite low, with the corresponding specificity and precision being close to 100% for all three classes of DMCTs (see You et al. [110]).

The above simple power calculation clearly illustrates the potential limitations associated with inferring DMCTs in real epigenome studies. As one wishes to infer DMCTs in the context of ever-increasing numbers of cell-types, one main limitation is on the sensitivity to detect non-specific DMCTs, and thus great caution needs to be exercised when interpreting the specificity of inferred DMCTs. Another major limitation is the potentially low variance in the fraction displayed by a given cell-type, as noted recently [112].

4.7 Useful Software

Here we give a brief overview of some of the important software that has been developed in the cell-type heterogeneity field. Most of this software is freely available as R-code from the websites as specified in the publications. For estimating cell-type fractions in whole blood or peripheral blood mononuclear cells, we can recommend the *minfi* Bioconductor package [80] which uses Houseman's constrained projection algorithm [46] as well as the original DMRM for blood used when validating the Houseman algorithm. Another option is the *EpiDISH* Bioconductor package [48], which offers robust partial correlations (RPCs) and Support Vector Regression as an alternative to constrained projection for estimating cell-type fractions. For those unfamiliar with R, the *EpiDISH* package has an associated webserver [113] and also incorporates its own DMRM for blood. Another important option for blood is the IDOL algorithm [78] and associated DMRM [114]. For cord blood, there are dedicated DMRMs available from the *FlowSorted.CordBlood.450 K* Bioconductor R-package [81] and from [115, 129]. For saliva, there is also a dedicated DMRM, and which has been specially designed for epidemiological studies in children [116]. For saliva and buccal swabs, one can also use a general three cell-type DNAm reference matrix defined for generic epithelial, fibroblast, and immune cells [59, 117], and which is also available within the *EpiDISH* package. For complex solid tissues, one can use the *HEpiDISH* algorithm [59] and associated 2-layer DMRMs for estimating cell-type fractions for epithelial, stromal, and all immune cell subtypes, and which is also part of the *EpiDISH* package. Another option for inferring epithelial and stromal fractions in complex tissues is *MethylCIBERSORT* [118]. The EpiSCORE algorithm can be used to build DMRMs for tissue-types for which there is a matching scRNA-Seq atlas, and DMRMs for lung and breast tissue have been provided [89]. An alternative statistical strategy for estimating cell-type fractions called *dtangle* that improves upon the constraint projection and the other frameworks has also been proposed [119]. A method called *MethylResolver* has recently been proposed to evaluate the quality of an inferred cell-type decomposition [120], which is an important step missing from all other software. Algorithms for identifying cell-type-specific DNA methylation changes include *TOAST* [54], *CellDMC* [52], *HIRE* [55], *Omicwas* [121] and *TCA* [122]. *Omicwas*, which implements a non-linear ridge regression framework is noteworthy for also dealing with the co-linearity problem in DMRMs of high cellular resolution. Reference-free and semi-reference-free methods remain useful for the purpose of identifying DMCs not confounded by cell-type heterogeneity [49–51, 123–125], but are generally speaking limited for the tasks of cell-type fraction estimation and inference of cell-type-specific DNAm changes.

4.8 Outlook

It is likely that over the next years, technologies for single-cell methylomics [126] will improve to allow the construction of DNAm reference profiles for most cell-types within a tissue. This would provide an alternative means to build a DMRM, which is more tailored to the tissue of interest, thus overcoming some of the limitations of existing strategies. However, in the foreseeable future, generating single-cell methylomics for large numbers of tissues (and individuals) will remain impractical. This means that for epigenome studies aiming to identify cell-type-specific biomarkers, and which will require genome-wide profiling of many 100s to 1000s of individuals, it will still be necessary to use cell-type deconvolution methods. Thus, single-cell methylomics may help towards the construction of tissue-specific DMRMs, and this in turn may improve the inference of cell-type fractions and DMCTs from large-scale epigenome studies.

Another important area of future development will be applications in personalized epigenetic epidemiology and medicine. One key area that is likely to grow even further is noninvasive detection of disease via measurement of cfDNAm in plasma. Cell-type deconvolution methods as well as more conventional Machine Learning methods need to be improved further, specially in relation to how informative features are selected. Currently, algorithms are also aimed at identifying biomarkers (cell-type specific or not) that display average differences in DNAm associated with an exposure or outcome. However, averaging over individuals also obscures in which particular individuals a DNAm change is happening. In the context of DMCTs, thus one would wish to infer in which cell-types and individual a particular DNAm change is happening in. Thus, given a DNAm data matrix and estimates of cell-type fractions, one would ideally want to infer an array of DNAm values, defined over CpGs, cell-types, and individuals. Mathematically, this is known as a tensor-object [37] and attempts to infer such a tensor in the context of EWAS have been made [122], albeit not yet successfully so [127]. Of note, such tensor-inference methods are also likely to overfit and to be computationally very intensive if applied in a genome-wide context, as the number of parameters to estimate is substantially higher.

An important biological question for the future is the relative fractions of non-specific vs. specific DMCTs. This is likely to depend on the tissue as well as phenotype. For instance, in the case of SNPs, the proportion of mQTLs that are shared between blood cell-types is likely to be large (approx. 70–80%) [36, 37]. Thus, in the case of DNAm variation associated with genetic variants, this variation appears to be largely cell-type independent, which however still allows for downstream functional effects to display cell-type specificity. In aging too, a recent study has shown that most age-associated DNAm changes appear to be independent of tissue and cell-type [128]. EWAS for body-mass index have also revealed shared DNAm changes between blood and fat cells, but also differences [3]. In the case of smoking in blood, there is evidence for both non-specific as well as specific DMCTs, whereas in buccal swabs the effects appear to be much more cell-type specific [18]. Other conditions such as type-1 diabetes [39], asthma [34],

Rheumatoid Arthritis [41] or cancer [89] appear to be associated with larger numbers of cell-type-specific DMCTs. Understanding why specific factors may affect DNAm in a cell-type independent manner and why others do not will be an important question for the epigenetics community that is also highly relevant for epigenetic epidemiological studies.

To conclude, this chapter provides a brief overview of the key concepts and tools needed to tackle the challenge posed by cell-type heterogeneity in the context of DNA methylation studies. We recommend that future epigenetic epidemiological studies make an attempt to tackle this challenge, using the software tools described herein, mainly as hypothesis generation tools, to be followed up with functional studies or validations in purified cell populations.

References

1. Petronis A (2010) Epigenetics as a unifying principle in the aetiology of complex traits and diseases. Nature 465:721–727
2. Gao X, Jia M, Zhang Y, Breitling LP, Brenner H (2015) DNA methylation changes of whole blood cells in response to active smoking exposure in adults: a systematic review of DNA methylation studies. Clin Epigenetics 7:113
3. Wahl S, Drong A, Lehne B, Loh M, Scott WR, Kunze S, Tsai PC, Ried JS, Zhang W, Yang Y et al (2017) Epigenome-wide association study of body mass index, and the adverse outcomes of adiposity. Nature 541:81–86
4. van Otterdijk SD, Binder AM, Michels KB (2017) Locus-specific DNA methylation in the placenta is associated with levels of pro-inflammatory proteins in cord blood and they are both independently affected by maternal smoking during pregnancy. Epigenetics 12:875–885
5. Non AL, Binder AM, Kubzansky LD, Michels KB (2014) Genome-wide DNA methylation in neonates exposed to maternal depression, anxiety, or SSRI medication during pregnancy. Epigenetics 9:964–972
6. Christensen BC, Houseman EA, Marsit CJ, Zheng S, Wrensch MR, Wiemels JL, Nelson HH, Karagas MR, Padbury JF, Bueno R et al (2009) Aging and environmental exposures alter tissue-specific DNA methylation dependent upon CpG island context. PLoS Genet 5: e1000602
7. Teschendorff AE, Menon U, Gentry-Maharaj A, Ramus SJ, Weisenberger DJ, Shen H, Campan M, Noushmehr H, Bell CG, Maxwell AP et al (2010) Age-dependent DNA methylation of genes that are suppressed in stem cells is a hallmark of cancer. Genome Res 20:440–446
8. Teschendorff AE, Jones A, Fiegl H, Sargent A, Zhuang JJ, Kitchener HC, Widschwendter M (2012) Epigenetic variability in cells of normal cytology is associated with the risk of future morphological transformation. Genome Med 4:24
9. Vandiver AR, Irizarry RA, Hansen KD, Garza LA, Runarsson A, Li X, Chien AL, Wang TS, Leung SG, Kang S, Feinberg AP (2015) Age and sun exposure-related widespread genomic blocks of hypomethylation in nonmalignant skin. Genome Biol 16:80
10. Laird PW (2003) The power and the promise of DNA methylation markers. Nat Rev Cancer 3: 253–266
11. Horvath S, Erhart W, Brosch M, Ammerpohl O, von Schonfels W, Ahrens M, Heits N, Bell JT, Tsai PC, Spector TD et al (2014) Obesity accelerates epigenetic aging of human liver. Proc Natl Acad Sci U S A 111:15538–15543
12. Issa JP (2014) Aging and epigenetic drift: a vicious cycle. J Clin Invest 124:24–29

13. Cancer Genome Atlas Research N, Weinstein JN, Collisson EA, Mills GB, Shaw KR, Ozenberger BA, Ellrott K, Shmulevich I, Sander C, Stuart JM (2013) The cancer genome atlas pan-cancer analysis project. Nat Genet 45:1113–1120
14. Montano C, Taub MA, Jaffe A, Briem E, Feinberg JI, Trygvadottir R, Idrizi A, Runarsson A, Berndsen B, Gur RC et al (2016) Association of DNA methylation differences with schizophrenia in an epigenome-wide association study. JAMA Psychiatry 73:506–514
15. Jaffe AE, Straub RE, Shin JH, Tao R, Gao Y, Collado-Torres L, Kam-Thong T, Xi HS, Quan J, Chen Q et al (2018) Developmental and genetic regulation of the human cortex transcriptome illuminate schizophrenia pathogenesis. Nat Neurosci 21:1117–1125
16. Rakyan VK, Down TA, Balding DJ, Beck S (2011) Epigenome-wide association studies for common human diseases. Nat Rev Genet 12:529–541
17. Jaffe AE, Kleinman JE (2016) Genetic and epigenetic analysis of schizophrenia in blood-a no-brainer? Genome Med 8:96
18. Teschendorff AE, Yang Z, Wong A, Pipinikas CP, Jiao Y, Jones A, Anjum S, Hardy R, Salvesen HB, Thirlwell C et al (2015) Correlation of smoking-associated DNA methylation changes in buccal cells with DNA methylation changes in epithelial cancer. JAMA Oncol 1: 476–485
19. Levine ME, Lu AT, Chen BH, Hernandez DG, Singleton AB, Ferrucci L, Bandinelli S, Salfati E, Manson JE, Quach A et al (2016) Menopause accelerates biological aging. Proc Natl Acad Sci U S A 113:9327–9332
20. Koestler DC, Marsit CJ, Christensen BC, Accomando W, Langevin SM, Houseman EA, Nelson HH, Karagas MR, Wiencke JK, Kelsey KT (2012) Peripheral blood immune cell methylation profiles are associated with nonhematopoietic cancers. Cancer Epidemiol Biomark Prev 21:1293–1302
21. Shen SY, Singhania R, Fehringer G, Chakravarthy A, Roehrl MHA, Chadwick D, Zuzarte PC, Borgida A, Wang TT, Li T et al (2018) Sensitive tumour detection and classification using plasma cell-free DNA methylomes. Nature 563:579–583
22. Chen X, Gole J, Gore A, He Q, Lu M, Min J, Yuan Z, Yang X, Jiang Y, Zhang T et al (2020) Non-invasive early detection of cancer four years before conventional diagnosis using a blood test. Nat Commun 11:3475
23. Widschwendter M, Zikan M, Wahl B, Lempiainen H, Paprotka T, Evans I, Jones A, Ghazali S, Reisel D, Eichner J et al (2017) The potential of circulating tumor DNA methylation analysis for the early detection and management of ovarian cancer. Genome Med 9:116
24. Liu L, Toung JM, Jassowicz AF, Vijayaraghavan R, Kang H, Zhang R, Kruglyak KM, Huang HJ, Hinoue T, Shen H et al (2018) Targeted methylation sequencing of plasma cell-free DNA for cancer detection and classification. Ann Oncol 29:1445–1453
25. Guo S, Diep D, Plongthongkum N, Fung HL, Zhang K, Zhang K (2017) Identification of methylation haplotype blocks aids in deconvolution of heterogeneous tissue samples and tumor tissue-of-origin mapping from plasma DNA. Nat Genet 49:635–642
26. Lehmann-Werman R, Neiman D, Zemmour H, Moss J, Magenheim J, Vaknin-Dembinsky A, Rubertsson S, Nellgard B, Blennow K, Zetterberg H et al (2016) Identification of tissue-specific cell death using methylation patterns of circulating DNA. Proc Natl Acad Sci U S A 113:E1826–E1834
27. Lappalainen T, Greally JM (2017) Associating cellular epigenetic models with human phenotypes. Nat Rev Genet 18:441–451
28. Fraga MF, Ballestar E, Paz MF, Ropero S, Setien F, Ballestar ML, Heine-Suner D, Cigudosa JC, Urioste M, Benitez J et al (2005) Epigenetic differences arise during the lifetime of monozygotic twins. Proc Natl Acad Sci U S A 102:10604–10609
29. Ahuja N, Issa JP (2000) Aging, methylation and cancer. Histol Histopathol 15:835–842
30. Ahuja N, Li Q, Mohan AL, Baylin SB, Issa JP (1998) Aging and DNA methylation in colorectal mucosa and cancer. Cancer Res 58:5489–5494

31. Maegawa S, Hinkal G, Kim HS, Shen L, Zhang L, Zhang J, Zhang N, Liang S, Donehower LA, Issa JP (2010) Widespread and tissue specific age-related DNA methylation changes in mice. Genome Res 20:332–340
32. Yuan T, Jiao Y, de Jong S, Ophoff RA, Beck S, Teschendorff AE (2015) An integrative multi-scale analysis of the dynamic DNA methylation landscape in aging. PLoS Genet 11:e1004996
33. Jaffe AE, Irizarry RA (2014) Accounting for cellular heterogeneity is critical in epigenome-wide association studies. Genome Biol 15:R31
34. Liang L, Willis-Owen SA, Laprise C, Wong KC, Davies GA, Hudson TJ, Binia A, Hopkin JM, Yang IV, Grundberg E et al (2015) An epigenome-wide association study of total serum immunoglobulin E concentration. Nature 520:670–674
35. Jones A, Teschendorff AE, Li Q, Hayward JD, Kannan A, Mould T, West J, Zikan M, Cibula D, Fiegl H et al (2013) Role of DNA methylation and epigenetic silencing of HAND2 in endometrial cancer development. PLoS Med 10:e1001551
36. Chen L, Ge B, Casale FP, Vasquez L, Kwan T, Garrido-Martin D, Watt S, Yan Y, Kundu K, Ecker S et al (2016) Genetic drivers of epigenetic and transcriptional variation in human immune cells. Cell 167:1398–1414.e24
37. Teschendorff AE, Jing H, Paul DS, Virta J, Nordhausen K (2018) Tensorial blind source separation for improved analysis of multi-omic data. Genome Biol 19:76
38. Baylin SB, Ohm JE (2006) Epigenetic gene silencing in cancer—a mechanism for early oncogenic pathway addiction? Nat Rev Cancer 6:107–116
39. Paul DS, Teschendorff AE, Dang MA, Lowe R, Hawa MI, Ecker S, Beyan H, Cunningham S, Fouts AR, Ramelius A et al (2016) Increased DNA methylation variability in type 1 diabetes across three immune effector cell types. Nat Commun 7:13555
40. Reynolds LM, Taylor JR, Ding J, Lohman K, Johnson C, Siscovick D, Burke G, Post W, Shea S, Jacobs DR Jr et al (2014) Age-related variations in the methylome associated with gene expression in human monocytes and T cells. Nat Commun 5:5366
41. Julia A, Absher D, Lopez-Lasanta M, Palau N, Pluma A, Waite Jones L, Glossop JR, Farrell WE, Myers RM, Marsal S (2017) Epigenome-wide association study of rheumatoid arthritis identifies differentially methylated loci in B cells. Hum Mol Genet 26:2803–2811
42. Angermueller C, Clark SJ, Lee HJ, Macaulay IC, Teng MJ, Hu TX, Krueger F, Smallwood S, Ponting CP, Voet T et al (2016) Parallel single-cell sequencing links transcriptional and epigenetic heterogeneity. Nat Methods 13:229–232
43. Smallwood SA, Lee HJ, Angermueller C, Krueger F, Saadeh H, Peat J, Andrews SR, Stegle O, Reik W, Kelsey G (2014) Single-cell genome-wide bisulfite sequencing for assessing epigenetic heterogeneity. Nat Methods 11:817–820
44. Lee DS, Luo C, Zhou J, Chandran S, Rivkin A, Bartlett A, Nery JR, Fitzpatrick C, O'Connor C, Dixon JR, Ecker JR (2019) Simultaneous profiling of 3D genome structure and DNA methylation in single human cells. Nat Methods 16:999–1006
45. Kelsey G, Stegle O, Reik W (2017) Single-cell epigenomics: recording the past and predicting the future. Science 358:69–75
46. Houseman EA, Accomando WP, Koestler DC, Christensen BC, Marsit CJ, Nelson HH, Wiencke JK, Kelsey KT (2012) DNA methylation arrays as surrogate measures of cell mixture distribution. BMC Bioinform 13:86
47. Accomando WP, Wiencke JK, Houseman EA, Butler RA, Zheng S, Nelson HH, Kelsey KT (2012) Decreased NK cells in patients with head and neck cancer determined in archival DNA. Clin Cancer Res 18:6147–6154
48. Teschendorff AE, Breeze CE, Zheng SC, Beck S (2017) A comparison of reference-based algorithms for correcting cell-type heterogeneity in epigenome-wide association studies. BMC Bioinform 18:105
49. Houseman EA, Molitor J, Marsit CJ (2014) Reference-free cell mixture adjustments in analysis of DNA methylation data. Bioinformatics 30:1431–1439

50. Lutsik P, Slawski M, Gasparoni G, Vedeneev N, Hein M, Walter J (2017) MeDeCom: discovery and quantification of latent components of heterogeneous methylomes. Genome Biol 18:55
51. Li Z, Wu H (2019) TOAST: improving reference-free cell composition estimation by cross-cell type differential analysis. Genome Biol 20:190
52. Zheng SC, Breeze CE, Beck S, Teschendorff AE (2018) Identification of differentially methylated cell types in epigenome-wide association studies. Nat Methods 15:1059–1066
53. Montano CM, Irizarry RA, Kaufmann WE, Talbot K, Gur RE, Feinberg AP, Taub MA (2013) Measuring cell-type specific differential methylation in human brain tissue. Genome Biol 14: R94
54. Li Z, Wu Z, Jin P, Wu H (2019) Dissecting differential signals in high-throughput data from complex tissues. Bioinformatics 35(20):3898–3905
55. Luo X, Yang C, Wei Y (2019) Detection of cell-type-specific risk-CpG sites in epigenome-wide association studies. Nat Commun 10:3113
56. Maegawa S, Gough SM, Watanabe-Okochi N, Lu Y, Zhang N, Castoro RJ, Estecio MR, Jelinek J, Liang S, Kitamura T et al (2014) Age-related epigenetic drift in the pathogenesis of MDS and AML. Genome Res 24:580–591
57. Teschendorff AE, Menon U, Gentry-Maharaj A, Ramus SJ, Gayther SA, Apostolidou S, Jones A, Lechner M, Beck S, Jacobs IJ, Widschwendter M (2009) An epigenetic signature in peripheral blood predicts active ovarian cancer. PLoS One 4:e8274
58. Liu Y, Aryee MJ, Padyukov L, Fallin MD, Hesselberg E, Runarsson A, Reinius L, Acevedo N, Taub M, Ronninger M et al (2013) Epigenome-wide association data implicate DNA methylation as an intermediary of genetic risk in rheumatoid arthritis. Nat Biotechnol 31:142–147
59. Zheng SC, Webster AP, Dong D, Feber A, Graham DG, Sullivan R, Jevons S, Lovat LB, Beck S, Widschwendter M, Teschendorff AE (2018) A novel cell-type deconvolution algorithm reveals substantial contamination by immune cells in saliva, buccal and cervix. Epigenomics 10:925–940
60. Theda C, Hwang SH, Czajko A, Loke YJ, Leong P, Craig JM (2018) Quantitation of the cellular content of saliva and buccal swab samples. Sci Rep 8:6944
61. Wong YT, Tayeb MA, Stone TC, Lovat LB, Teschendorff AE, Iwasiow R, Craig JM (2021) A comparison of epithelial cell content of oral samples estimated using cytology and DNA methylation. Epigenetics:1–8. https://doi.org/10.1080/15592294.2021.1950977
62. Regev A, Teichmann SA, Lander ES, Amit I, Benoist C, Birney E, Bodenmiller B, Campbell P, Carninci P, Clatworthy M et al (2017) The human cell atlas. elife 6:e27041
63. Rozenblatt-Rosen O, Stubbington MJT, Regev A, Teichmann SA (2017) The human cell atlas: from vision to reality. Nature 550:451–453
64. Vieira Braga FA, Kar G, Berg M, Carpaij OA, Polanski K, Simon LM, Brouwer S, Gomes T, Hesse L, Jiang J et al (2019) A cellular census of human lungs identifies novel cell states in health and in asthma. Nat Med 25:1153–1163
65. Schiller HB, Montoro DT, Simon LM, Rawlins EL, Meyer KB, Strunz M, Vieira Braga FA, Timens W, Koppelman GH, Budinger GRS et al (2019) The human lung cell atlas: a high-resolution reference map of the human lung in health and disease. Am J Respir Cell Mol Biol 61:31–41
66. Bock C (2012) Analysing and interpreting DNA methylation data. Nat Rev Genet 13:705–719
67. Ali HR, Provenzano E, Dawson SJ, Blows FM, Liu B, Shah M, Earl HM, Poole CJ, Hiller L, Dunn JA et al (2014) Association between CD8+ T-cell infiltration and breast cancer survival in 12,439 patients. Ann Oncol 25:1536–1543
68. Koestler DC, Usset J, Christensen BC, Marsit CJ, Karagas MR, Kelsey KT, Wiencke JK (2017) DNA methylation-derived neutrophil-to-lymphocyte ratio: an epigenetic tool to explore cancer inflammation and outcomes. Cancer Epidemiol Biomark Prev 26:328–338
69. Wiencke JK, Koestler DC, Salas LA, Wiemels JL, Roy RP, Hansen HM, Rice T, McCoy LS, Bracci PM, Molinaro AM et al (2017) Immunomethylomic approach to explore the blood neutrophil lymphocyte ratio (NLR) in glioma survival. Clin Epigenetics 9:10

70. Sun K, Jiang P, Chan KC, Wong J, Cheng YK, Liang RH, Chan WK, Ma ES, Chan SL, Cheng SH et al (2015) Plasma DNA tissue mapping by genome-wide methylation sequencing for noninvasive prenatal, cancer, and transplantation assessments. Proc Natl Acad Sci U S A 112: E5503–E5512
71. Zheng SC, Beck S, Jaffe AE, Koestler DC, Hansen KD, Houseman AE, Irizarry RA, Teschendorff AE (2017) Correcting for cell-type heterogeneity in epigenome-wide association studies: revisiting previous analyses. Nat Methods 14:216–217
72. Reinius LE, Acevedo N, Joerink M, Pershagen G, Dahlen SE, Greco D, Soderhall C, Scheynius A, Kere J (2012) Differential DNA methylation in purified human blood cells: implications for cell lineage and studies on disease susceptibility. PLoS One 7:e41361
73. den Dekker HT, Burrows K, Felix JF, Salas LA, Nedeljkovic I, Yao J, Rifas-Shiman SL, Ruiz-Arenas C, Amin N, Bustamante M et al (2019) Newborn DNA-methylation, childhood lung function, and the risks of asthma and COPD across the life course. Eur Respir J 53:1801795
74. Merid SK, Novoloaca A, Sharp GC, Kupers LK, Kho AT, Roy R, Gao L, Annesi-Maesano I, Jain P, Plusquin M et al (2020) Epigenome-wide meta-analysis of blood DNA methylation in newborns and children identifies numerous loci related to gestational age. Genome Med 12:25
75. Salas LA, Baker ER, Nieuwenhuijsen MJ, Marsit CJ, Christensen BC, Karagas MR (2019) Maternal swimming pool exposure during pregnancy in relation to birth outcomes and cord blood DNA methylation among private well users. Environ Int 123:459–466
76. Vehmeijer FOL, Kupers LK, Sharp GC, Salas LA, Lent S, Jima DD, Tindula G, Reese S, Qi C, Gruzieva O et al (2020) DNA methylation and body mass index from birth to adolescence: meta-analyses of epigenome-wide association studies. Genome Med 12:105
77. Yeung EH, Guan W, Zeng X, Salas LA, Mumford SL, de Prado BP, van Meel ER, Malmberg A, Sunyer J, Duijts L et al (2020) Cord blood DNA methylation reflects cord blood C-reactive protein levels but not maternal levels: a longitudinal study and meta-analysis. Clin Epigenetics 12:60
78. Koestler DC, Jones MJ, Usset J, Christensen BC, Butler RA, Kobor MS, Wiencke JK, Kelsey KT (2016) Improving cell mixture deconvolution by identifying optimal DNA methylation libraries (IDOL). BMC Bioinform 17:120
79. Newman AM, Liu CL, Green MR, Gentles AJ, Feng W, Xu Y, Hoang CD, Diehn M, Alizadeh AA (2015) Robust enumeration of cell subsets from tissue expression profiles. Nat Methods 12:453–457
80. Aryee MJ, Jaffe AE, Corrada-Bravo H, Ladd-Acosta C, Feinberg AP, Hansen KD, Irizarry RA (2014) Minfi: a flexible and comprehensive Bioconductor package for the analysis of Infinium DNA methylation microarrays. Bioinformatics 30:1363–1369
81. Gervin K, Salas LA, Bakulski KM, van Zelm MC, Koestler DC, Wiencke JK, Duijts L, Moll HA, Kelsey KT, Kobor MS et al (2019) Systematic evaluation and validation of reference and library selection methods for deconvolution of cord blood DNA methylation data. Clin Epigenetics 11:125
82. Oh G, Koncevicius K, Ebrahimi S, Carlucci M, Groot DE, Nair A, Zhang A, Krisciunas A, Oh ES, Labrie V et al (2019) Circadian oscillations of cytosine modification in humans contribute to epigenetic variability, aging, and complex disease. Genome Biol 20:2
83. Oh G, Ebrahimi S, Carlucci M, Zhang A, Nair A, Groot DE, Labrie V, Jia P, Oh ES, Jeremian RH et al (2018) Cytosine modifications exhibit circadian oscillations that are involved in epigenetic diversity and aging. Nat Commun 9:644
84. Lam LL, Emberly E, Fraser HB, Neumann SM, Chen E, Miller GE, Kobor MS (2012) Factors underlying variable DNA methylation in a human community cohort. Proc Natl Acad Sci U S A 109(Suppl 2):17253–17260
85. Dunham I, Kundaje A, Aldred SF, Collins PJ, Davis CA, Doyle F, Epstein CB, Frietze S, Harrow J, Kaul R et al (2012) An integrated encyclopedia of DNA elements in the human genome. Nature 489:57–74

86. Bernstein BE, Stamatoyannopoulos JA, Costello JF, Ren B, Milosavljevic A, Meissner A, Kellis M, Marra MA, Beaudet AL, Ecker JR et al (2010) The NIH Roadmap Epigenomics mapping consortium. Nat Biotechnol 28:1045–1048
87. Roadmap Epigenomics C, Kundaje A, Meuleman W, Ernst J, Bilenky M, Yen A, Heravi-Moussavi A, Kheradpour P, Zhang Z, Wang J et al (2015) Integrative analysis of 111 reference human epigenomes. Nature 518:317–330
88. Teschendorff AE, Relton CL (2018) Statistical and integrative system-level analysis of DNA methylation data. Nat Rev Genet 19:129–147
89. Teschendorff AE, Zhu T, Breeze CE, Beck S (2020) EPISCORE: cell type deconvolution of bulk tissue DNA methylomes from single-cell RNA-seq data. Genome Biol 21:221
90. Teschendorff AE (2020) EpiSCORE: cell-type deconvolution of bulk tissue DNA methylomes from single-cell RNA-seq data (version 0.9.1). Zenodo. https://doi.org/10.5281/zenodo.3893646
91. Tabula Muris C (2018) Overall c, logistical c, organ c, processing, library p, sequencing, computational data a, cell type a, writing g, et al: single-cell transcriptomics of 20 mouse organs creates a Tabula Muris. Nature 562:367–372
92. Gerstein M (2012) Genomics: ENCODE leads the way on big data. Nature 489:208
93. Consortium EP, Moore JE, Purcaro MJ, Pratt HE, Epstein CB, Shoresh N, Adrian J, Kawli T, Davis CA, Dobin A et al (2020) Expanded encyclopaedias of DNA elements in the human and mouse genomes. Nature 583:699–710
94. Gentles AJ, Newman AM, Liu CL, Bratman SV, Feng W, Kim D, Nair VS, Xu Y, Khuong A, Hoang CD et al (2015) The prognostic landscape of genes and infiltrating immune cells across human cancers. Nat Med 21:938–945
95. Kang S, Li Q, Chen Q, Zhou Y, Park S, Lee G, Grimes B, Krysan K, Yu M, Wang W et al (2017) CancerLocator: non-invasive cancer diagnosis and tissue-of-origin prediction using methylation profiles of cell-free DNA. Genome Biol 18:53
96. Hoadley KA, Yau C, Hinoue T, Wolf DM, Lazar AJ, Drill E, Shen R, Taylor AM, Cherniack AD, Thorsson V et al (2018) Cell-of-origin patterns dominate the molecular classification of 10,000 tumors from 33 types of cancer. Cell 173:291–304.e6
97. Moss J, Magenheim J, Neiman D, Zemmour H, Loyfer N, Korach A, Samet Y, Maoz M, Druid H, Arner P et al (2018) Comprehensive human cell-type methylation atlas reveals origins of circulating cell-free DNA in health and disease. Nat Commun 9:5068
98. Tang W, Wan S, Yang Z, Teschendorff AE, Zou Q (2018) Tumor origin detection with tissue-specific miRNA and DNA methylation markers. Bioinformatics 34:398–406
99. Moran S, Martinez-Cardus A, Sayols S, Musulen E, Balana C, Estival-Gonzalez A, Moutinho C, Heyn H, Diaz-Lagares A, de Moura MC et al (2016) Epigenetic profiling to classify cancer of unknown primary: a multicentre, retrospective analysis. Lancet Oncol 17:1386–1395
100. Nejman D, Straussman R, Steinfeld I, Ruvolo M, Roberts D, Yakhini Z, Cedar H (2014) Molecular rules governing de novo methylation in cancer. Cancer Res 74:1475–1483
101. Chen Y, Breeze CE, Zhen S, Beck S, Teschendorff AE (2016) Tissue-independent and tissue-specific patterns of DNA methylation alteration in cancer. Epigenetics Chromatin 9:10
102. Joehanes R, Just AC, Marioni RE, Pilling LC, Reynolds LM, Mandaviya PR, Guan W, Xu T, Elks CE, Aslibekyan S et al (2016) Epigenetic signatures of cigarette smoking. Circ Cardiovasc Genet 9:436–447
103. Chen Y, Widschwendter M, Teschendorff AE (2017) Systems-epigenomics inference of transcription factor activity implicates aryl-hydrocarbon-receptor inactivation as a key event in lung cancer development. Genome Biol 18:236
104. Zheng SC, Widschwendter M, Teschendorff AE (2016) Epigenetic drift, epigenetic clocks and cancer risk. Epigenomics 8:705–719
105. Cancer Genome Atlas Research N, Kandoth C, Schultz N, Cherniack AD, Akbani R, Liu Y, Shen H, Robertson AG, Pashtan I, Shen R et al (2013) Integrated genomic characterization of endometrial carcinoma. Nature 497:67–73

106. Platel V, Faure S, Corre I, Clere N (2019) Endothelial-to-mesenchymal transition (EndoMT): roles in tumorigenesis, metastatic extravasation and therapy resistance. J Oncol 2019:8361945
107. Tang YN, Ding WQ, Guo XJ, Yuan XW, Wang DM, Song JG (2015) Epigenetic regulation of Smad2 and Smad3 by profilin-2 promotes lung cancer growth and metastasis. Nat Commun 6: 8230
108. Potenta S, Zeisberg E, Kalluri R (2008) The role of endothelial-to-mesenchymal transition in cancer progression. Br J Cancer 99:1375–1379
109. Su D, Wang X, Campbell MR, Porter DK, Pittman GS, Bennett BD, Wan M, Englert NA, Crowl CL, Gimple RN et al (2016) Distinct epigenetic effects of tobacco smoking in whole blood and among leukocyte subtypes. PLoS One 11:e0166486
110. You C, Wu S, Zheng SC, Zhu T, Jing H, Flagg K, Wang G, Jin L, Wang S, Teschendorff AE (2020) A cell-type deconvolution meta-analysis of whole blood EWAS reveals lineage-specific smoking-associated DNA methylation changes. Nat Commun 11:4779
111. Fircanis S, Merriam P, Khan N, Castillo JJ (2014) The relation between cigarette smoking and risk of acute myeloid leukemia: an updated meta-analysis of epidemiological studies. Am J Hematol 89:E125–E132
112. Meier R, Nissen E, Koestler DC (2021) Low variability in the underlying cellular landscape adversely affects the performance of interaction-based approaches for conducting cell-specific analyses of DNA methylation in bulk samples. Stat Appl Genet Mol Biol 20(3):73–84
113. Zheng SC, Breeze CE, Beck S, Dong D, Zhu T, Ma L, Ye W, Zhang G, Teschendorff AE (2019) EpiDISH web server: epigenetic dissection of intra-sample-heterogeneity with online GUI. Bioinformatics 36(6):1950–1951
114. Salas LA, Koestler DC, Butler RA, Hansen HM, Wiencke JK, Kelsey KT, Christensen BC (2018) An optimized library for reference-based deconvolution of whole-blood biospecimens assayed using the Illumina HumanMethylationEPIC BeadArray. Genome Biol 19:64
115. Lin X, Tan JYL, Teh AL, Lim IY, Liew SJ, MacIsaac JL, Chong YS, Gluckman PD, Kobor MS, Cheong CY, Karnani N (2018) Cell type-specific DNA methylation in neonatal cord tissue and cord blood: a 850K-reference panel and comparison of cell types. Epigenetics 13: 941–958
116. Middleton LYM, Dou J, Fisher J, Heiss JA, Nguyen VK, Just AC, Faul J, Ware EB, Mitchell C, Colacino JA, Bakulski KM (2021) Saliva cell type DNA methylation reference panel for epidemiological studies in children. Epigenetics:1–17. https://doi.org/10.1080/15592294.2021.1890874
117. van Dongen J, Ehli EA, Jansen R, van Beijsterveldt CEM, Willemsen G, Hottenga JJ, Kallsen NA, Peyton SA, Breeze CE, Kluft C et al (2018) Genome-wide analysis of DNA methylation in buccal cells: a study of monozygotic twins and mQTLs. Epigenetics Chromatin 11:54
118. Chakravarthy A, Furness A, Joshi K, Ghorani E, Ford K, Ward MJ, King EV, Lechner M, Marafioti T, Quezada SA et al (2018) Pan-cancer deconvolution of tumour composition using DNA methylation. Nat Commun 9:3220
119. Hunt GJ, Freytag S, Bahlo M, Gagnon-Bartsch JA (2019) Dtangle: accurate and robust cell type deconvolution. Bioinformatics 35:2093–2099
120. Arneson D, Yang X, Wang K (2020) MethylResolver-a method for deconvoluting bulk DNA methylation profiles into known and unknown cell contents. Commun Biol 3:422
121. Takeuchi F, Kato N (2021) Nonlinear ridge regression improves cell-type-specific differential expression analysis. BMC Bioinform 22:141
122. Rahmani E, Schweiger R, Rhead B, Criswell LA, Barcellos LF, Eskin E, Rosset S, Sankararaman S, Halperin E (2019) Cell-type-specific resolution epigenetics without the need for cell sorting or single-cell biology. Nat Commun 10:3417
123. Onuchic V, Hartmaier RJ, Boone DN, Samuels ML, Patel RY, White WM, Garovic VD, Oesterreich S, Roth ME, Lee AV, Milosavljevic A (2016) Epigenomic deconvolution of breast tumors reveals metabolic coupling between constituent cell types. Cell Rep 17:2075–2086

124. Leek JT, Johnson WE, Parker HS, Jaffe AE, Storey JD (2012) The sva package for removing batch effects and other unwanted variation in high-throughput experiments. Bioinformatics 28(6):882–883
125. Teschendorff AE, Zhuang J, Widschwendter M (2011) Independent surrogate variable analysis to deconvolve confounding factors in large-scale microarray profiling studies. Bioinformatics 27:1496–1505
126. Karemaker ID, Vermeulen M (2018) Single-cell DNA methylation profiling: technologies and biological applications. Trends Biotechnol 36:952–965
127. Jing H, Zheng SC, Breeze CE, Beck S, Teschendorff AE (2019) Calling differential DNA methylation at cell-type resolution: an objective status-quo. bioRxiv. https://doi.org/10.1101/822940
128. Zhu T, Zheng SC, Paul DS, Horvath S, Teschendorff AE (2018) Cell and tissue type independent age-associated DNA methylation changes are not rare but common. Aging (Albany NY) 10:3541–3557
129. Cardenas A, Allard C, Doyon M, Houseman EA, Bakulski KM, Perron P, Bouchard L, Hivert MF (2016) Validation of a DNA methylation reference panel for the estimation of nucleated cells types in cord blood. Epigenetics 11:773–779

Epigenetics in Development

5

Victor Yuan and Wendy P. Robinson

Contents

Abstract

Epigenetic processes regulate cellular function at all stages of life. Epigenetic processes in their entirety are referred to as the epigenome, which include DNA methylation, non-coding RNAs, histone modifications, chromatin structure and accessibility. This multifaceted epigenome is highly dynamic across human

V. Yuan (✉) · W. P. Robinson
BC Children's Hospital Research Institute, Vancouver, BC, Canada

Department of Medical Genetics, University of British Columbia, Vancouver, BC, Canada
e-mail: vyuan@bcchr.ca; wrobinson@bcchr.ca

K. B. Michels (ed.), *Epigenetic Epidemiology*,
https://doi.org/10.1007/978-3-030-94475-9_5

development, requiring almost complete reprogramming at two developmental timepoints: during the development of germ cells, and also immediately after fertilization. Epigenetic modifications are also highly stable, for example, genomic imprinting must be protected from post-fertilization epigenetic reprogramming, and X-chromosome inactivation in females is crucial to balance gene dosage from the X chromosome. In this chapter, we describe the major epigenetic processes that occur throughout human development, from the DNA methylation erasures that occur in germ cells, to the epigenetic characteristics of differentiated cells that arise from previous lineage-specification events. Increasingly, advancing technologies, such as organoid systems and single-cell sequencing, are allowing the epigenome in development to be characterized in an unprecedented amount of detail, which has led to key insights into the epigenetics of not only normal human development, but also the developmental origins of disease.

Abbreviations

5hmC	5-hydroxymethylcytosine
5mC	5-methylcytosine
CGI	CpG island
CpG	Cytosine (phosphate) guanine site
DMR	Differentially methylated region
DNAm	DNA methylation
ERVs	Endogenous retroviral
EWAS	Epigenome-wide association study
gDMRs	germline differentially methylated regions
ICM	Inner cell mass
ICR	Imprinting control region
mQTL	methylated quantitative trait loci
PGC	Primordial germ cells
PMD	Partially methylated domain
TE	Trophectoderm
XCI	X-chromosome inactivation
ZGA	Zygotic genome activation

5.1 Introduction

In development, tissues and cell types acquire increasing specificity that allows them to fulfill their biological functions. Molecularly, much of this functional specificity is achieved through expression of highly cell-specific gene pathways, coordinated by a multifaceted epigenome. The epigenome here refers to several distinct molecular processes including DNA methylation (DNAm), histones and their post-translational

modifications, non-histone chromatin proteins, higher-order chromatin and chromosome organization, and non-coding RNAs. Establishment of these epigenetic features requires precise timing and specificity during the normal development of mammalian cells, throughout every generation. Epigenetic resetting rapidly occurs after fertilization, and then gradual acquisition of epigenetic features occurs in a cell-specific manner, resulting in the variety of cell types that fulfill various biological niches. Most of our knowledge of developmental epigenetics comes from mouse studies, due to the difficulties in collecting and studying human tissues early in development, and it is assumed that similar processes occur in humans. However, important studies verifying (or disputing) these processes in humans are noted. Developmentally significant epigenetic processes are described as "hard-wired" in that they occur without requiring extrinsic signals. However, an understanding of these pre-determined events is essential to also understanding how extrinsic signals, such as from environment, can alter the developmental trajectory of cells into disease-associated states. In this chapter, we provide a foundation for understanding the epigenome in development, and as a consequence, an understanding of the factors that can cause human disease.

5.2 The Epigenomes of Germ Cells

Germ cell formation involves a series of epigenetic changes including erasure of previous marks, suppression of many genes, activation of genes specific to gametogenesis, and changes in chromatin that facilitate appropriate chromosome interactions during the process of meiosis. As a consequence, genome-wide, sperm and eggs in mammals are highly epigenetically distinct compared to other differentiated somatic cells. Sperm and egg epigenomes are also highly distinct from each other, owing to differences in both the timing and pattern of the epigenomic marks that are established during spermatogenesis and oogenesis.

5.2.1 Primordial Germ Cells

Primordial germ cells (PGCs) are formed from the epiblast cell layer in the gastrulating embryo through a series of epigenetic remodeling events and expression of specific genes [1]. Chromatin-remodeling events occur as the PGCs differentiate from other epiblast cells, which include structural changes such as heterochromatin decondensation, loss of chromocenters, and increase in nuclear size [2]. Early changes in histone modifications include the loss of H3K9me2, a repressive mark, which is replaced by H3K27me3 from the core component of the polycomb repressive complex 2 [2–4]. Repressive methylation is also added onto histone tails of H2A and H4 by nuclear protein complexes, which include Blimp1 and Prtmt5 [5]. Chromatin remodeling continues as the PGCs migrate and take residence in the genital ridges.

Genome-wide erasure of DNAm in PGCs also begins during the migration to the genital ridge. DNA methylation is first lost by an active process where the actions of

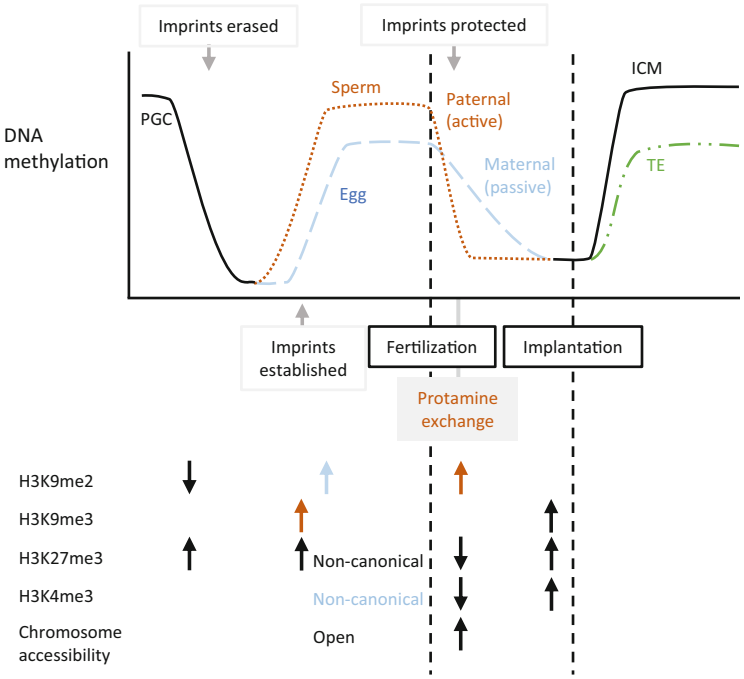

Fig. 5.1 Schematic of epigenetic reprogramming during development. The first major genome-wide DNA methylation erasure occurs in primordial germ cells (PGCs), and includes erasure of prior gametic imprints and replacement of H3K9me2 with H3K27me3. H3K27me3 undergoes further reconfiguration during migration to the genital ridge. H3K9me2 is reestablished in growing oocytes and H3K9me3 in growing spermatocytes. Concomitantly, DNA methylation is reestablished, but earlier and to a greater extent in sperm as compared to oocytes (egg). After fertilization, the asymmetric paternal and maternal epigenomes begin reprogramming. H3K4me3, which is distributed in a non-canonical pattern in the growing oocytes, is erased and a non-canonical distribution begins to be established. The paternal genome undergoes active demethylation, and protamines are replaced by histones. The maternal genome also undergoes demethylation, but more gradually through passive replication-dependent mechanisms. Gamete-specific differentially methylated regions (i.e., genomic imprints) and most repetitive sequences are protected during post-fertilization epigenetic reprogramming. At the blastocyst stage, there is the first lineage-specification event of inner cell mass (ICM) and trophectoderm (TE). By the blastocyst stage, most canonical patterns of histone modifications are established, such as bivalent promoters marked by H3K4me3/H3K27me3 at developmental genes. After implantation, DNA methylation is established in both ICM and TE, but in TE remains lower, a difference that is retained through development

TET enzymes convert 5'mC to 5'hmC, and then passively lost through successive replication events owing to the absence of normal DNAm maintenance activity (Fig. 5.1) [6, 7]. After migration to the genital ridge and completion of genome-wide demethylation, PGCs achieve a nearly completely demethylated genome [8, 9]. However, a subset of stable marks such as DNAm at imprinted loci and on the inactive X chromosome are removed only after migration to the genital ridge is complete, which may involve additional repression of *Dnmt3a*, *Dnmt3b*, and *Uhrgf1*

[9]. By the end of this process, genomic DNAm drops from around 71–80% to 7–14%, and PGCs enter either spermatogenesis or oogenesis depending on the sex of the embryo [9, 10].

5.2.2 Spermatogenesis

In male embryos, sperm progenitors (prospermatogonia) undergo alterations to DNAm starting before birth and continuing until completion of meiotic pachytene after puberty [11, 12]. Sperm DNA becomes highly remethylated owing largely to the activity of DNMT3A and DNMT3L [13]. DNMT3B is involved in methylating imprinted genes and repetitive sequences in sperm, but is inactive in oocytes [14, 15]. PIWI-interacting small RNAs are also responsible for establishing DNAm, which involve Dnmt3c, and H3K9me3 to suppress evolutionarily young copies of retrotransposons [16–18]. Another unique aspect of spermatogenesis is that initially, the sperm DNA is tightly wrapped around histones, but the majority are replaced first by non-canonical histone variants, then by transition proteins, and finally by protamines in the haploid stage after maturation, which allows DNA to be tightly packaged inside the sperm head [12, 19, 20]. However, approximately 1% of sperm histones are not replaced by protamines, and are instead retained in the mature sperm DNA specifically at gene promoters [21], suggesting a possible mechanism for transgenerational inheritance of paternal histone marks that can contribute to gene regulation in early development.

5.2.3 Oogenesis

In female embryos, after PGCs migrate to the genital ridge, there is short period of massive mitotic expansion, after which these oocyte precursors initiate meiosis, but then arrest at prophase 1 and remain dormant in the developing ovary until after puberty. Each month a small subset of this pool of oocytes will undergo a growth phase with only one fully maturing to ovulation. Over their development, oocytes acquire unique DNAm and histone modification profiles. In the late phase of oocyte growth, DNAm increases from 7–14% in PGCs up to 72% in mature oocytes, though remains less than the 86% in mature sperm (Table 5.1) [9, 10]. Unlike somatic cells or spermatocytes, de novo DNAm in oocytes is uncoupled from DNA replication and is acquired in a DNMT3A and DNMT3L dependent manner [22]. In contrast to sperm, where DNAm is more widespread and enriched at enhancers and repetitive short interspersed nuclear elements, DNAm in oocytes is uniquely distributed across actively transcribed gene bodies and CpG islands (CGI) [10], resulting in a oocyte-specific bimodal distribution of hypo- and hyper-methylated regions. Interestingly, DNMT1 and UHRF1, which are canonically responsible for the maintenance of DNAm, function in this context to complete *de-novo* methylation [15, 23]. Compared to other cell types, oocytes also show increased non-CpG methylation [15]. However, DNAm in oocytes is non-essential for the completion of oogenesis [24, 25]. Lower DNAm in oocytes may allow for subsets of transposable elements

Table 5.1 Major epigenetic differences between tissues and stages in development

	Gametes	Post-fertilization/ preimplantation	Somatic tissues (ICM-derived)	PLACENTAL TROPHOBLAST (TE-derived)
DNA methylation (DNAm)	Genome-wide erasure of DNAm, drops from 80% to 7–14% in PGCs	Loss of DNAm active (paternal) and passive (maternal)	Hypomethylation at CGI promoters of tissue-specific genes; Hypermethylation in gene bodies and intergenic regions	Lower relative genome-wide DNAm, owing to large PMDs in gene-poor regions
Partially methylated domains (PMDs)			Present in only a few tissues and with less genomic coverage relative to placenta; A key epigenomic feature when cells become malignant	Large regions that cover mostly intergenic regions; PMD-contained promoters are enriched for controlling placental genes; Placental-specific gene pathways co-opted in some cancers
Imprinting	Established in spermatogenesis and oogenesis	Imprints protected from DNA demethylation	Arises from a small number of gametic differences in DNAm that are resistant to reprogramming and act as imprinting control centers. Most imprinting is stable and pervasive across tissues; A small minority show tissue specificity in parent-of-origin expression	Most genes imprinted in somatic tissues are also imprinted in placenta. Some genes imprinted only in placenta show non-DNAm-dependent imprinting via maternal H3K27me3—Most of these are maternal and imprinting can be polymorphic
X-chromosome inactivation (XCI)	Erasure of inactive X-specific marks in female germ line	XIST and XACT expressed from both Xs in human female embryos initially. No initial imprinted XCI as there is in mouse	In females, the majority of CGI promoters are methylated on the inactive X and unmethylated on the active X. But chromosome-wide, the inactive X is less methylated	Less DNAm at CpG island promoters in female placentas, suggesting possible incomplete inactivation. Preferential paternal X inactivation in rodents, but not human

			than the active X. The inactive X also shows loss of promoter/enhancer histone acetylation; loss of promoter-associated H3K4me3 and H3K36me3; accumulation of H3K27me3, H4K20me1, H3K9me2, and macroH2A	
Histones	Low H3K9me2 and high H3K27me3 in PGCs; Sperm DNA packaged tightly around protamines; Oocytes have non-canonical distribution of H3K4me3	Paternal protamines are replaced by histones; Bivalent marks, marked by H3K27me3/H3K4me3, are established at promoters of developmental genes	Bivalent domains, marked by H3K4me3/H3K27me3, are prevalent in differentiation and cell type-specific	Histone deacetylation, increased H3K4me3, H3K9Ac, H3K27Ac, and H3K27me3 are associated with syncytialization. Less H3K27me3 in TE compared to ICM

DNAm, DNA methylation; PGC, primordial germ cell; XCI, X-chromosome inactivation; CGI, CpG Island; PMD, partially methylated domain; TE, trophectoderm; ICM, inner cell mass

to become active in late oocyte maturation, and serve as promoters or first exons for other genes, which can account for ~10% of oocyte-specific transcription [26].

Oocytes not only have a unique DNAm profile but also display non-canonical distribution of histone modifications. H3K4me3, which is associated with active promoters, is deposited across unusually broad intergenic domains, from the activity of the H3K4 methyltransferase, MLL2 [27–29]. H3K4me3 is required for genome-wide transcriptional silencing associated with oocyte maturation and for resumption of meiosis [27, 28, 30]. Although H3K4me3 is anti-correlated with DNAm in mouse oocytes, this inverse relation does not exist in humans [31]. Repressive H3K27me3 is found in large unmethylated regions with low transcriptional activity, but its functional role in oogenesis is unclear. However, a proportion of H3K27me3 in oocytes has been shown to regulate non-canonical DNAm-independent imprinting in the early embryo [32].

During maturation of the oocyte germinal vessicle, chromatin undergoes major conformational changes associated with transcriptional silencing. Upon resumption of meiosis, oocytes lose all higher-order chromatin structures, and interactions become uniform in strength across the chromosome [33, 34].

5.3 Post-fertilization

After fertilization, the paternal (sperm) and maternal (oocyte) genomes undergo extensive epigenetic reprogramming in order to equalize asymmetric epigenetic differences and prepare the totipotent zygote for cellular differentiation [35]. However, some genomic regions are resistant to this initial reprogramming, leading to the allele-specific expression of genes depending on parental origin, or genomic imprinting.

5.3.1 Post-fertilization Reprogramming of Gametic Genomes

Initially, the zygote is transcriptionally silent, relying on mostly maternally-inherited proteins and factors to carry out cellular processes during the first initial cell divisions. As maternally-inherited factors are degraded, the transcriptionally silent zygotic genome must become transcriptionally active, and this process, referred to as zygotic genome activation (ZGA), is associated with a variety of epigenetic changes in the paternal and maternal genomes [36]. In the paternal pronucleus, protamines are replaced with histones, and widespread erasure of DNAm is initiated before the first cell division through a rapid active demethylation process controlled by TET1. However, further demethylation occurs through replication-coupled passive dilution, which is aided by TET3 [8, 37]. TET3 converts 5'mC to 5'hmC which then impedes DNAm maintenance [38, 39], and TET3 activity itself can protect against DNMT3a-mediated de novo DNAm [40, 41]. Thus, TET proteins contribute to both active and passive modes of demethylation of the paternal genome after fertilization.

In contrast, the maternal pronucleus appears somewhat resistant to the initial TET demethylation steps, as evidenced by lower TET3 expression and lower TET3-

dependent 5'hmC [8]. The resistance of the maternal genome to this initial wave of DNA demethylation is due to H3K9me2-mediated recruitment of PGC7 (aka STELLA or DPPA3), which promotes maintenance of CpG DNAm and inhibits DNMT1-mediated de novo DNAm [42–45]. The maternal epigenome instead is demethylated by passive dilution through successive rounds of replication. By the blastocyst stage, about 20% of CpGs remain methylated, mostly at imprinted domains and repetitive elements [10, 22, 46].

Histone modifications are also dynamically reconfigured during preimplantation development. H3K4me3 is rapidly depleted after fertilization, but is then replaced by canonical H3K4me3 through H3K4 demethylases KDM5B and KDM1A [27, 28, 47]. In the zygote, there is also erasure of paternally-inherited H3K27me3, but maternally-inherited H3K27me3 at distal intergenic regions is maintained [48]. Despite promoters being devoid of H3K27me3, many still remain transcriptionally inactive, suggesting other silencing mechanisms are at play. H3K27me3 patterning is established by the blastocyst stage, where H3K27me3/H3K4me3 characterizes "bivalent" promoters of developmental genes, which are silent until rapid activation is needed for cell differentiation. Additionally, H3K27me3 maintains expression of some imprinted genes in early development until it is erased in the epiblast lineage at the blastocyst stage, but a subset of H3K27me3 imprinted genes are maintained in the extraembryonic lineage [32]. H3K9 di- and tri-methylation, which are associated to contribute to the silencing and protection of paternal DNAm at imprinted domains, are acquired by the 8-cell stage [43, 49].

From the zygote to blastocyst stages, higher-order chromatin structure such as long-range and local chromosome interactions are gradually established, including more enhancer–promoter interactions, increased number of DNase hypersensitivity sites, and increased chromatin accessibility [33, 34, 50].

5.3.2 Genomic Imprinting

Genomic imprinting is associated with human developmental disorders such as Prader–Willi Syndrome [51], Angelman Syndrome [52], and Beckwith–Wiedemann Syndrome [53, 54]. Many of the known imprinted genes are regulated by germline differentially methylated regions (gDMRs), where only one parental allele is methylated. As a consequence of differences in epigenetic reprogramming between oogenesis and spermatogenesis, over 1600 CGIs in mouse are gDMRs. Most of these gDMRs are subsequently lost either through removal of DNAm during preimplantation development [22, 55], or through acquisition of de novo DNAm on the alternate allele after implantation [56]. However, some gDMRs are resistant to early epigenetic reprogramming and stable throughout the lifespan, often acting as imprinting control regions (ICRs) that regulate the expression of nearby gene clusters. Mutations within ICRs can lead to loss of the ability to reprogram the appropriate parent-of-origin imprints and can result in unusual inheritance patterns in families. For example, mutations in the ICR regulating the *SNRPN* gene, implicated in Prader–Willi syndrome, can be passed from mother to child with no direct effect in the first generation. However, when a male inherits an ICR mutation from his mother, the maternal imprint

cannot be changed to a paternal imprint in his sperm, thus giving him a 50% risk of having a child affected by Prader–Willi syndrome [57].

In addition to canonical (DNAm-dependent) imprinting, there are a subset of non-canonical imprinted genes for which DNAm is absent, but instead imprinted gene expression is dependent on maternal H3K27me3 [32]. These sites of non-canonical imprinting may also lead to acquired DNAm differences (secondary imprinted DMRs) that are not derived from gDMRs. It is possible that even short-lived imprinting due to H3K27me3 on the maternal genome, leading to paternal-biased gene expression in human preimplantation embryos, could have lasting effects on embryonic development [58, 59]. Further, while the majority of maternal H3K27me3 is lost in embryonic lineages, it appears to be lost more gradually in the placenta and thus allelic differences in gene expression may play a greater role in placental function [12].

In mouse, the greatest numbers of imprinted loci were observed in early embryos, placenta, and brain [60]. A higher rate of imprinted genes in human placenta as compared to somatic tissues is also supported by the large number of placental-specific imprinted DMRs [61]. Placental-specific imprinted DMRs can arise from gDMRs (e.g., imprinted genes *Kcnq1* and *Igf2r*) or through acquisition of secondary DMRs via non-canonical imprinting (e.g., Gab1 and Sfmbt2), with loss of imprinting in the epiblast due to acquisition of DNAm, leading to bi-allelic silencing of the corresponding gene [62]. Interestingly, non-canonical imprinting was localized to endogenous retrovirus-K (ERVK) long terminal repeats, which can act as promoters in the placenta [63, 64]. It is possible that active ERVs in the placenta may interfere with epigenetic silencing in the placental lineages [62]. Placental-specific imprinting appears to be largely species specific and may have no clear function, but some knock-outs of placental-specific imprinted genes in mouse can lead to placental dysfunction [62].

5.4 Post-implantation

The first lineage-specification event occurs with the differentiation of the trophectoderm (TE) and inner cell mass (ICM) at the blastocyst stage of development. Trophoblast cells, which form both the outer layer of the placental chorionic villi, and the placental cells that invade into the maternal endometrium, are derived from the trophectoderm (TE). The remaining extraembryonic tissues (amnion, placental mesenchyme, and the inner layer of the chorion) derive from the ICM, as do all fetal tissues. At this critical developmental timepoint, there are already extensive genome-wide differences in epigenetic marks between ICM and TE that are essential for regulating the myriad of developmental events that follows.

5.4.1 Reprogramming of the Blastocyst

The blastocyst has the lowest levels of genome-wide DNAm compared to any other developmental stage, as a result of the genome-wide demethylation that occurs in the first few cell divisions after fertilization. At the blastocyst stage, DNAm begins to be

rapidly deposited across all ICM- and TE-derived tissues by DNMT3A and DNMT3B, which preferentially target broadly distributed H3K36me2/3-marked regions [9, 65, 66]. Although DNMT3A and -3B generally do not discriminate between genomic regions, the majority of CGI promoters remain unmethylated to avoid ectopic silencing of important genes. H3K4m3, which is associated with active promoters, protects promoter sequences from de novo DNAm by repelling the ADD domains of DNMT3A and -3B [67]. Developmental gene promoters, which are normally silent but must activate under specific developmental cues, are bivalently marked by PRC2-deposited H3K27me3 and H3K4me2 [68], and are devoid of CGI DNAm owing to TET1 activity [69]. Transcription factors, such as DPPA2, DPPA4, and others, are responsible for the targeted deposition of H3K4me3 and H3K27me3 at CGI promoters of bivalent genes in development [70–72].

5.4.2 Placental Trophoblasts

The embryo develops from the ICM but its growth also depends on the normal development of the placenta, which consists mostly of trophoblasts derived from TE. Due to early divergence of ICM- and TE-derived cell types, the resulting differentiated cell types are each epigenetically unique (Table 5.1). TE is resistant to the dramatic gain of methylation that ICM experiences starting at the blastocyst stage, resulting in a characteristic genome-wide hypomethylated profile of placental trophoblasts that make up the bulk of placenta [73, 74]. Fully developed somatic tissues have a bimodal distribution of genome-wide DNAm, where CGIs are mainly hypomethylated, and the rest of the genome is densely methylated. While unmethylated promoters are conserved in placenta, intergenic regions have lower DNAm compared to somatic tissues, and this lower methylation is organized into large partially methylated domains (PMDs) that are maintained throughout gestation [74, 75]. However, most CpG methylation increases in trophoblast over gestation, even in PMD regions [75, 76]. It is unclear if PMDs regulate transcription directly or are left from earlier transcription regulatory processes. However, even though these regions are relatively gene-poor overall, there is an enrichment for CGI promotors for genes involved in pathways related to the epithelial–mesenchymal transition, immune response, and inflammation [77]. Placental-specific genes and related pathways share epigenetic regulation that is similar to the epigenetic program in some cancers, where trophoblast-specific invasive and immune response pathways are co-opted [77, 78]. A proportion of placental hypomethylation is localized to various repetitive elements, such as LINE-1 [79] and human ERVs [80], which may regulate placental-specific functions [81]. Some repetitive elements serve as alternative promoters for trophoblast-related genes, such as for *KCNH5* and *IL2RB* [80, 82, 83]. Despite observations of PMD- and retrotransposon-driven gene expression in placenta, their contribution to placental function is relatively uncharacterized.

Histone modifications and chromatin dynamics are also under-characterized in placenta. Syncytiotrophoblasts, which line the outer surface of the placental chorionic villi are directly exposed to the maternal environment and govern molecular exchange, express hormones, and produce extracellular vesicles that are critical to

regulating interactions between mother and fetus. They are multi-nucleated, created by the fusion of underlying layer of cytotrophoblasts, and their DNA is heterogeneous; older nuclei are condensed into transcriptionally repressed heterochromatin-rich syncytial knots, whereas younger nuclei are euchromatic and transcriptionally active. Recent high-resolution studies indicate histone modifications are important to trophoblast differentiation. Differentiation of syncytiotrophoblast is associated with deacetylation of histone lysine residues, such as depletion of H3K27Ac and H3K9Ac at promoters by histone deacetylases HDAC1 and HDAC2 [84, 85]. Impairment of histone acetylation may be a feature of disease processes in placenta, as aberrant H3K27ac in placenta is associated with maternal preeclampsia and may be a more general feature of placental inflammation [76]. Other histone modifications, such as H3K27me3, H3K4me3, and H3K9me3, are dynamic in cytotrophoblast populations across gestation, but their functional impact on placental function is unclear [76]. H3K9me3, however, is associated with silencing of cytotrophoblast genes, and is correlated with placental DNA hypomethylation [76]. H3K27me3 might be important in establishing a trophoblast-specific transcription program in early placental development, as H3K27me3 has been shown to be uniquely distributed in trophoblast stem cells compared to other embryonic stem cells [86].

The dynamics of epigenetic regulation in the placenta throughout pregnancy is highly specific, especially in comparison to other cell types and tissues, and research in this exciting area is still in its infancy. However, growing interest in the early epigenetic regulation in the placenta promises that future research will lead to exciting new insights into human development.

5.4.3 X-Chromosome Inactivation

As female 46,XX embryos have two X chromosomes compared to one in 46,XY males, there is initially a sex difference in the expression of X-chromosome genes. As a mechanism for dosage compensation in female embryos, most genes on one of the two X chromosomes are transcriptionally silenced. X-chromosome inactivation (XCI) has been well studied and is a useful paradigm for understanding how monoallelic epigenetic silencing can occur during development. A critical trigger for XCI is the expression of XIST RNA, a 15–17 kb long non-coding RNA that accumulates along one X chromosome in *cis* and is required for the initiation and stable maintenance of XCI throughout development [87]. While XCI has been most well studied in mouse, there are some key differences with humans. For example, in mouse Xist expression is initiated at the 4–8 cell stage when only the paternal X is inactivated, followed by later reactivation of the paternal X and random XCI in the epiblast. In human embryos, XCI is not initiated until implantation or after, and there is no parental bias in XCI [88]. As human XCI is only completed after implantation and early tissue differentiation, there is a dosage imbalance of most X-linked genes up to this point [89]. Furthermore, as cells can differ by which parental X chromosome is inactivated in females, different cells can express different genetic variants. This not only has consequences for the presentation of genetic disorders but presents challenges to epigenome-wide association studies. Unfortunately, the

X-chromosome data is often discarded from epigenome-wide association studies (EWASs) as a consequence.

In human preimplantation embryos, XIST is expressed from both X's in females, as well as from the single X in males [89]. XACT, a primate-specific X-linked lncRNA, which eventually coats the active X chromosome, is also expressed biallelically at this time [90]. The mechanism by which XIST expression is subsequently dampened to keep only one X active is unclear, although it is hypothesized that XACT may antagonize XIST activity, and XIST may be able to trigger XCI on one randomly chosen X once XACT expression is lost [91]. It is also hypothesized that there is a dosage-sensitive repressor of XIST, encoded by an autosomal gene, that acts prior to the initiation of XCI [88]. In mouse, the RNA binding molecule SPEN has been shown to bind to Xist RNA, and once recruited to the X chromosome, targets active enhancers and promoters and induces repressive chromatin changes that shut down transcription [92]. Regardless of the mechanism, a series of sequential chromatin alterations are established as Xist RNA spreads along the X [93]. First, most euchromatic histone marks are lost, such as promoter/enhancer histone acetylation, and promoter-associated H3K4me3 and H3K36me3. Heterochromatic marks including H3K27me3, H2AK119Ub, and H4K20me1 accumulate later, and there is also accumulation of H3K9me2, and macroH2A. Lastly, CGI promoter methylation secures the stable nature of chromatin compaction characteristic of the inactive. Interestingly, even though H4K20me1 and H3K27me3 are early marks associated with the inactive X chromosome, they are not necessary for silencing and may accumulate as secondary effects of the inactive heterochromatin [94]. These epigenetic changes on the inactive X chromosome result in condensation of the inactive X into a transcriptionally silent structure called a Barr body that localizes to the nuclear periphery.

In addition to higher expression of X-linked genes in females prior to XCI, some genes escape from XCI altogether, and may show persistent higher expression in females compared to males. Roughly 15% of genes in human somatic tissues escape XCI, defined as having expression from the inactive X that is at least 10% of the active X [95]. Genes that escape XCI across different species tend to be enriched for CTCF-binding sites, ATAC-sequencing signal (indicative of open chromatin) and LTR repeats, compared to genes that are subject to silencing by XCI [96]. Thus, some sex differences may be due to persistent gene expression differences on the X, as well as to secondary effects on autosomal gene expression [97]. In addition, some genes show variable XCI, being variably inactivated in some cells, tissues, and individuals. Comparisons across tissues within individuals and in twin pairs show high concordance, indicating that a large portion of variability in XCI escape may be due to *cis*-acting genetic variation [95, 98].

The process of XCI in the placenta is less well studied and associated with distinct properties. In mouse, there is preferential inactivation of the paternal X, while in human placenta, the process is random but occurs in a patchy fashion across the placenta due to its clonal development [99, 100]. In comparison to males, placental DNAm on the X chromosome in females is greater at gene promoters and lower in gene bodies than expected for X-chromosome inactivation [101]. However, in

comparison to blood there is substantial hypomethylation of X-linked promoters in the female placenta, suggesting that there might be more escape from XCI in the placenta [79]. If this is true, then sex differences in fetal development may be driven in part by sex differences in placental function.

5.5 Factors That Contribute to Variability in Development

Epigenetic mechanisms ensure cellular identities are maintained throughout not only development but the entirety of human life, resulting in wide-ranging types of cellular function that fulfill highly tissue-specific niches. At the same time, extrinsic factors such as environment and aging can perturb the epigenetic program, increasing risk of aberrations that can result in disease. Intrinsic genetic variation is another factor that is increasingly becoming appreciated as a major factor contributing to epigenetic variation, highlighting the importance of considering human diversity.

5.5.1 Cell-Specific Epigenetics

Fundamentally, epigenetic marks are a cellular process. Despite this, most epigenetic studies rely on whole tissue samples, which are mixtures of distinct cell populations. Because every distinct cell population also has a distinct epigenetic profile, measurements of whole tissues are an average epigenetic profile of the cellular soup that consists of each sample. This results in challenges in interpreting epigenetic changes associated with disease, environment, or development; as observed epigenetic variation can be attributed to either changes that are present in a subset of cells, in all or most cells, or are simply reflecting cell composition variability without any changes in the epigenetic footprint [102]. For example, placental epigenetics, including canonical features like PMDs and placental-specific imprinting, are often features of trophoblast cells, rather than all component cell populations of the placenta [75, 76]. Fortunately, these challenges are surmountable. At the experimental design stage, cell populations most relevant to the research question should be as homogeneous as possible by using cell-specific isolation protocols. When this is not possible, cell composition can be estimated from whole tissue samples, which is commonly done in DNAm studies [103]. Single-cell technologies can address cell-specificity challenges, and have been useful in characterizing early developmental timepoints [104], but current costs and sample preparation requirements remain as major hurdles to employing these technologies at scale on a population level resolution to study disease and environment-related processes.

5.5.2 Genetic Influences on Epigenetic Variation

Genetic variation at the single nucleotide level is also a major factor contributing to epigenetic variation, especially in contexts where multiple human populations are

included. Methylation quantitative trait loci (mQTLs) are where a SNP affects the DNAm of a nearby CpG, which can account for up to 75% of the variance in DNAm [105, 106]. However, it is important to note that joint SNP plus environmental variation are larger contributors to DNAm variation than SNP variation alone [107]. The contribution of genetic and environmental factors can vary depending on tissue and cell type [108], and different human populations [109, 110]. The contribution of genotype-specific DNAm in human development is relatively uncharacterized. Even though the majority of mQTLs in adult brain are also found in fetal brain, a subset are fetal-specific mQTLs and enriched for variants associated with schizophrenia [111] and autism spectrum disorder [112]. Placental-specific mQTLs have also been identified, many of which are associated with expression changes, and a subset overlap known loci linked to fetal growth, diabetes, and inflammation [113]. In a study including blood samples from 4 timepoints, ranging from birth (cord blood) and adulthood (mean age = 47.5), a majority (>84%) of mQTLs were replicated across at all timepoints, suggesting that genetic effects are largely stable across life [114]. These studies highlight the importance of considering the contribution of human diversity and genetic influences on epigenetic processes in human development.

5.6 Transgenerational Epigenetics

It has been demonstrated in many organisms that a subset of environmentally induced epigenetic changes may be transmitted through the germline over multiple generations [115]. For example, exposures to endocrine disrupting chemicals, altered nutrition (e.g., high fat or low calorie), and stress have all been associated with transgenerational epigenetic alterations in rodent models [116]. In mammals, there is more evidence for male transgenerational epigenetics, as it is difficult to separate the in utero effects of maternal or grandmaternal exposures from true transgenerational inheritance in females [115, 116]. Epigenetic inheritance may potentially involve DNAm, histone modifications, short or long non-coding RNAs, and structural chromatin proteins [117, 118]. Although germline remodeling in spermatogenesis involves DNAm erasure and replacement of most histones with protamines, a small number of histones still remain, particularly in active CpG-rich gene regulatory regions [119, 120]. In a transgenic mouse model with overexpression of the histone demethylase KDM1A leading to depletion of H3K4me2, there was enrichment for H3K4me3 that escaped reprogramming in the embryo, supporting that some histone marks can serve as a potential mechanism for transgenerational inheritance [121]. Other studies have supported a role for small non-coding RNAs in transgenerational epigenetic inheritance. For example, male mice subjected to early life stress showed altered stress response pathways in their offspring, and by isolating microRNAs from the sperm of the stress-exposed males and injecting them into zygotes, similar behavioral effects were reproduced in offspring from control (non-stressed) parents [122, 123]. In another model whereby rats were transiently exposed to agricultural fungicide vinclozolin or pesticide DDT,

it was observed that changes in DNAm, histone localization, and non-coding RNAs colocalized to the same chromosomal regions, implying integrated effects that are mediated by non-coding RNA directed DNAm and DNAm-directed histone retention [124].

The possibility of epigenetic inheritance in humans is suggested by the Dutch famine, whereby the offspring of males, but not of females, prenatally exposed to caloric restriction showed increased adiposity [125]. Furthermore, small non-coding RNA expression and DNAm differed in the sperm from obese as compared to lean men [126]. Altered microRNA profiles were also reported in the sperm of men who smoked [127], while altered DNAm was reported in the sperm of infertile men taking high dose folate supplements [128]. However, direct evidence for epigenetic inheritance in humans is scarce, and in many cases where abnormal DNAm appeared to be transmitted in families, were actually due to genetic mutations that influence DNAm (e.g., imprinting mutations) [117, 129]. Likely, the same mechanisms shown in other mammals occur also in humans, but their significance in the context of the extensive genetic, environmental, and cultural variation in humans remains to be determined.

5.7 Conclusion

In conclusion, an important area of human epigenetics is understanding the epigenetic processes that occur during early development and gametogenesis. Histone modifications, chromosome structure, DNAm, and non-coding RNAs are highly dynamic in development, and are essential in setting up the epigenetic profiles that regulate gene expression in differentiated cells in later life. Recent and ongoing advancements in areas such as single-cell and low-input sequencing technologies are rapidly enabling characterization of the multifaceted epigenome at critical developmental timepoints. An important area for future research will be to not only continue measuring these epigenetic marks with advancing technologies, but also to understand the functional relevance of various epigenetic processes, how they contribute to development itself, and what factors disturb these processes that results in increased risk of disease in normal development. Recently developed ex vivo organoid technologies are promising, as more mechanistic studies and particularly human ones are needed. Although major epigenetic processes are often conserved across mammals, there is often evolutionary divergence in the timing and function of the proteins involved. Population studies measuring epigenetic marks in early life tissues, will also be important to understand how variability in normal development occurs under the context of genetic and environmental-related variation. Technology advancements and growing interest promise that the dynamics of the epigenome in development will be characterized at an unprecedented level of detail than ever before. This will lead to many exciting new insights that will aid our understanding of how genetic and environmental factors interact with the epigenome to create the vast diversity of human phenotypes.

Acknowledgements We thank Amy Inkster and Icíar Fernández Boyano for their invaluable help and feedback on the manuscript and figures.

References

1. Surani MA, Hayashi K, Hajkova P (2007) Genetic and epigenetic regulators of pluripotency. Cell 128:747–762. https://doi.org/10.1016/j.cell.2007.02.010
2. Hajkova P, Ancelin K, Waldmann T, Lacoste N, Lange UC, Cesari F, Lee C, Almouzni G, Schneider R, Surani MA (2008) Chromatin dynamics during epigenetic reprogramming in the mouse germ line. Nature 452:877–881. https://doi.org/10.1038/nature06714
3. Seki Y, Yamaji M, Yabuta Y, Sano M, Shigeta M, Matsui Y, Saga Y, Tachibana M, Shinkai Y, Saitou M (2007) Cellular dynamics associated with the genome-wide epigenetic reprogramming in migrating primordial germ cells in mice. Development 134:2627–2638. https://doi.org/10.1242/dev.005611
4. Seki Y, Hayashi K, Itoh K, Mizugaki M, Saitou M, Matsui Y (2005) Extensive and orderly reprogramming of genome-wide chromatin modifications associated with specification and early development of germ cells in mice. Dev Biol 278:440–458. https://doi.org/10.1016/j.ydbio.2004.11.025
5. Ancelin K, Lange UC, Hajkova P, Schneider R, Bannister AJ, Kouzarides T, Surani MA (2006) Blimp1 associates with Prmt5 and directs histone arginine methylation in mouse germ cells. Nat Cell Biol 8:623–630. https://doi.org/10.1038/ncb1413
6. Hackett JA, Sengupta R, Zylicz JJ, Murakami K, Lee C, Down TA, Surani MA (2013) Germline DNA demethylation dynamics and imprint erasure through 5-Hydroxymethylcytosine. Science 339:448–452. https://doi.org/10.1126/science.1229277
7. Hargan-Calvopina J, Taylor S, Cook H, Hu Z, Lee SA, Yen M-R, Chiang Y-S, Chen P-Y, Clark AT (2016) Stage-specific demethylation in primordial germ cells safeguards against precocious differentiation. Dev Cell 39:75–86. https://doi.org/10.1016/j.devcel.2016.07.019
8. Gu T-P, Guo F, Yang H, Wu H-P, Xu G-F, Liu W, Xie Z-G, Shi L, He X, Jin S, Iqbal K, Shi YG, Deng Z, Szabó PE, Pfeifer GP, Li J, Xu G-L (2011) The role of Tet3 DNA dioxygenase in epigenetic reprogramming by oocytes. Nature 477:606–610. https://doi.org/10.1038/nature10443
9. Seisenberger S, Andrews S, Krueger F, Arand J, Walter J, Santos F, Popp C, Thienpont B, Dean W, Reik W (2012) The dynamics of genome-wide DNA methylation reprogramming in mouse primordial germ cells. Mol Cell 48:849–862. https://doi.org/10.1016/j.molcel.2012.11.001
10. Zhu P, Guo H, Ren Y, Hou Y, Dong J, Li R, Lian Y, Fan X, Hu B, Gao Y, Wang X, Wei Y, Liu P, Yan J, Ren X, Yuan P, Yuan Y, Yan Z, Wen L, Yan L, Qiao J, Tang F (2018) Single-cell DNA methylome sequencing of human preimplantation embryos. Nat Genet 50:12–19. https://doi.org/10.1038/s41588-017-0007-6
11. Hajkova P, Erhardt S, Lane N, Haaf T, El-Maarri O, Reik W, Walter J, Surani MA (2002) Epigenetic reprogramming in mouse primordial germ cells. Mech Dev 117:15–23. https://doi.org/10.1016/S0925-4773(02)00181-8
12. Hanna CW, Demond H, Kelsey G (2018) Epigenetic regulation in development: is the mouse a good model for the human? Hum Reprod Update 24:556–576. https://doi.org/10.1093/humupd/dmy021
13. Okano M, Bell DW, Haber DA, Li E (1999) DNA methyltransferases Dnmt3a and Dnmt3b are essential for De novo methylation and mammalian development. Cell 99:247–257. https://doi.org/10.1016/S0092-8674(00)81656-6
14. Kato Y, Kaneda M, Hata K, Kumaki K, Hisano M, Kohara Y, Okano M, Li E, Nozaki M, Sasaki H (2007) Role of the Dnmt3 family in de novo methylation of imprinted and repetitive sequences during male germ cell development in the mouse. Hum Mol Genet 16:2272–2280. https://doi.org/10.1093/hmg/ddm179

15. Shirane K, Toh H, Kobayashi H, Miura F, Chiba H, Ito T, Kono T, Sasaki H (2013) Mouse oocyte Methylomes at base resolution reveal genome-wide accumulation of non-CpG methylation and role of DNA methyltransferases. PLoS Genet 9:e1003439. https://doi.org/10.1371/journal.pgen.1003439
16. Barau J, Teissandier A, Zamudio N, Roy S, Nalesso V, Hérault Y, Guillou F, Bourc'his D (2016) The DNA methyltransferase DNMT3C protects male germ cells from transposon activity. Science 354:909–912. https://doi.org/10.1126/science.aah5143
17. Pezic D, Manakov SA, Sachidanandam R, Aravin AA (2014) piRNA pathway targets active LINE1 elements to establish the repressive H3K9me3 mark in germ cells. Genes Dev 28: 1410–1428. https://doi.org/10.1101/gad.240895.114
18. Zamudio N, Barau J, Teissandier A, Walter M, Borsos M, Servant N, Bourc'his D (2015) DNA methylation restrains transposons from adopting a chromatin signature permissive for meiotic recombination. Genes Dev 29:1256–1270. https://doi.org/10.1101/gad.257840.114
19. Balhorn R, Brewer L, Corzett M (2000) DNA condensation by protamine and arginine-rich peptides: analysis of toroid stability using single DNA molecules. Mol Reprod Dev 56:230–234. https://doi.org/10.1002/(SICI)1098-2795(200006)56:2+<230::AID-MRD3>3.0.CO;2-V
20. Bao J, Bedford MT (2016) Epigenetic regulation of the histone-to-protamine transition during spermiogenesis. Reprod Camb Engl 151:R55–R70. https://doi.org/10.1530/REP-15-0562
21. Yoshida K, Muratani M, Araki H, Miura F, Suzuki T, Dohmae N, Katou Y, Shirahige K, Ito T, Ishii S (2018) Mapping of histone-binding sites in histone replacement-completed spermatozoa. Nat Commun 9:3885. https://doi.org/10.1038/s41467-018-06243-9
22. Smallwood SA, Tomizawa S, Krueger F, Ruf N, Carli N, Segonds-Pichon A, Sato S, Hata K, Andrews SR, Kelsey G (2011) Dynamic CpG island methylation landscape in oocytes and preimplantation embryos. Nat Genet 43:811–814. https://doi.org/10.1038/ng.864
23. Maenohara S, Unoki M, Toh H, Ohishi H, Sharif J, Koseki H, Sasaki H (2017) Role of UHRF1 in de novo DNA methylation in oocytes and maintenance methylation in preimplantation embryos. PLoS Genet 13:e1007042. https://doi.org/10.1371/journal.pgen.1007042
24. Bourc'his D, Xu G-L, Lin C-S, Bollman B, Bestor TH (2001) Dnmt3L and the establishment of maternal genomic imprints. Science 294:2536–2539. https://doi.org/10.1126/science.1065848
25. Kaneda M, Okano M, Hata K, Sado T, Tsujimoto N, Li E, Sasaki H (2004) Essential role for de novo DNA methyltransferase Dnmt3a in paternal and maternal imprinting. Nature 429:900–903. https://doi.org/10.1038/nature02633
26. Franke V, Ganesh S, Karlic R, Malik R, Pasulka J, Horvat F, Kuzman M, Fulka H, Cernohorska M, Urbanova J, Svobodova E, Ma J, Suzuki Y, Aoki F, Schultz RM, Vlahovicek K, Svoboda P (2017) Long terminal repeats power evolution of genes and gene expression programs in mammalian oocytes and zygotes. Genome Res 27:1384–1394. https://doi.org/10.1101/gr.216150.116
27. Dahl JA, Jung I, Aanes H, Greggains GD, Manaf A, Lerdrup M, Li G, Kuan S, Li B, Lee AY, Preissl S, Jermstad I, Haugen MH, Suganthan R, Bjørås M, Hansen K, Dalen KT, Fedorcsak P, Ren B, Klungland A (2016) Broad histone H3K4me3 domains in mouse oocytes modulate maternal-to-zygotic transition. Nature 537:548–552. https://doi.org/10.1038/nature19360
28. Zhang B, Zheng H, Huang B, Li W, Xiang Y, Peng X, Ming J, Wu X, Zhang Y, Xu Q, Liu W, Kou X, Zhao Y, He W, Li C, Chen B, Li Y, Wang Q, Ma J, Yin Q, Kee K, Meng A, Gao S, Xu F, Na J, Xie W (2016) Allelic reprogramming of the histone modification H3K4me3 in early mammalian development. Nature 537:553–557. https://doi.org/10.1038/nature19361
29. Hanna CW, Taudt A, Huang J, Gahurova L, Kranz A, Andrews S, Dean W, Stewart AF, Colomé-Tatché M, Kelsey G (2018) MLL2 conveys transcription-independent H3K4 trimethylation in oocytes. Nat Struct Mol Biol 25:73–82. https://doi.org/10.1038/s41594-017-0013-5

30. Andreu-Vieyra CV, Chen R, Agno JE, Glaser S, Anastassiadis K, Stewart AF, Matzuk MM (2010) MLL2 is required in oocytes for bulk histone 3 lysine 4 Trimethylation and transcriptional silencing. PLoS Biol 8:e1000453. https://doi.org/10.1371/journal.pbio.1000453
31. Xia W, Xu J, Yu G, Yao G, Xu K, Ma X, Zhang N, Liu B, Li T, Lin Z, Chen X, Li L, Wang Q, Shi D, Shi S, Zhang Y, Song W, Jin H, Hu L, Bu Z, Wang Y, Na J, Xie W, Sun Y-P (2019) Resetting histone modifications during human parental-to-zygotic transition. Science 365: 353–360. https://doi.org/10.1126/science.aaw5118
32. Inoue A, Jiang L, Lu F, Suzuki T, Zhang Y (2017) Maternal H3K27me3 controls DNA methylation-independent imprinting. Nature 547:419–424. https://doi.org/10.1038/nature23262
33. Ke Y, Xu Y, Chen X, Feng S, Liu Z, Sun Y, Yao X, Li F, Zhu W, Gao L, Chen H, Du Z, Xie W, Xu X, Huang X, Liu J (2017) 3D chromatin structures of mature gametes and structural reprogramming during mammalian embryogenesis. Cell 170:367–381.e20. https://doi.org/10.1016/j.cell.2017.06.029
34. Du Z, Zheng H, Huang B, Ma R, Wu J, Zhang X, He J, Xiang Y, Wang Q, Li Y, Ma J, Zhang X, Zhang K, Wang Y, Zhang MQ, Gao J, Dixon JR, Wang X, Zeng J, Xie W (2017) Allelic reprogramming of 3D chromatin architecture during early mammalian development. Nature 547:232–235. https://doi.org/10.1038/nature23263
35. Xu R, Li C, Liu X, Gao S (2021) Insights into epigenetic patterns in mammalian early embryos. Protein Cell 12:7–28. https://doi.org/10.1007/s13238-020-00757-z
36. Schulz KN, Harrison MM (2019) Mechanisms regulating zygotic genome activation. Nat Rev Genet 20:221–234. https://doi.org/10.1038/s41576-018-0087-x
37. Inoue A, Zhang Y (2011) Replication-dependent loss of 5-Hydroxymethylcytosine in mouse preimplantation embryos. Science 334:194–194. https://doi.org/10.1126/science.1212483
38. Hashimoto H, Liu Y, Upadhyay AK, Chang Y, Howerton SB, Vertino PM, Zhang X, Cheng X (2012) Recognition and potential mechanisms for replication and erasure of cytosine hydroxymethylation. Nucleic Acids Res 40:4841–4849. https://doi.org/10.1093/nar/gks155
39. Otani J, Kimura H, Sharif J, Endo TA, Mishima Y, Kawakami T, Koseki H, Shirakawa M, Suetake I, Tajima S (2013) Cell cycle-dependent turnover of 5-Hydroxymethyl cytosine in mouse embryonic stem cells. PLoS One 8:e82961. https://doi.org/10.1371/journal.pone.0082961
40. Amouroux R, Nashun B, Shirane K, Nakagawa S, Hill PWS, D'Souza Z, Nakayama M, Matsuda M, Turp A, Ndjetehe E, Encheva V, Kudo NR, Koseki H, Sasaki H, Hajkova P (2016) De novo DNA methylation drives 5hmC accumulation in mouse zygotes. Nat Cell Biol 18:225–233. https://doi.org/10.1038/ncb3296
41. Richard Albert J, Au Yeung WK, Toriyama K, Kobayashi H, Hirasawa R, Brind'Amour J, Bogutz A, Sasaki H, Lorincz M (2020) Maternal DNMT3A-dependent de novo methylation of the paternal genome inhibits gene expression in the early embryo. Nat Commun 11:5417. https://doi.org/10.1038/s41467-020-19279-7
42. Li Y, Zhang Z, Chen J, Liu W, Lai W, Liu B, Li X, Liu L, Xu S, Dong Q, Wang M, Duan X, Tan J, Zheng Y, Zhang P, Fan G, Wong J, Xu G-L, Wang Z, Wang H, Gao S, Zhu B (2018) Stella safeguards the oocyte methylome by preventing de novo methylation mediated by DNMT1. Nature 564:136–140. https://doi.org/10.1038/s41586-018-0751-5
43. Nakamura T, Arai Y, Umehara H, Masuhara M, Kimura T, Taniguchi H, Sekimoto T, Ikawa M, Yoneda Y, Okabe M, Tanaka S, Shiota K, Nakano T (2007) PGC7/Stella protects against DNA demethylation in early embryogenesis. Nat Cell Biol 9:64–71. https://doi.org/10.1038/ncb1519
44. Nakamura T, Liu Y-J, Nakashima H, Umehara H, Inoue K, Matoba S, Tachibana M, Ogura A, Shinkai Y, Nakano T (2012) PGC7 binds histone H3K9me2 to protect against conversion of 5mC to 5hmC in early embryos. Nature 486:415–419. https://doi.org/10.1038/nature11093
45. Zeng T-B, Han L, Pierce N, Pfeifer GP, Szabó PE (2019) EHMT2 and SETDB1 protect the maternal pronucleus from 5mC oxidation. Proc Natl Acad Sci U S A 116:10834–10841. https://doi.org/10.1073/pnas.1819946116

46. Wang L, Zhang J, Duan J, Gao X, Zhu W, Lu X, Yang L, Zhang J, Li G, Ci W, Li W, Zhou Q, Aluru N, Tang F, He C, Huang X, Liu J (2014) Programming and inheritance of parental DNA Methylomes in mammals. Cell 157:979–991. https://doi.org/10.1016/j.cell.2014.04.017
47. Ancelin K, Syx L, Borensztein M, Ranisavljevic N, Vassilev I, Briseño-Roa L, Liu T, Metzger E, Servant N, Barillot E, Chen C-J, Schüle R, Heard E (2016) Maternal LSD1/KDM1A is an essential regulator of chromatin and transcription landscapes during zygotic genome activation. eLife 5:e08851. https://doi.org/10.7554/eLife.08851
48. Zheng H, Huang B, Zhang B, Xiang Y, Du Z, Xu Q, Li Y, Wang Q, Ma J, Peng X, Xu F, Xie W (2016) Resetting epigenetic memory by reprogramming of histone modifications in mammals. Mol Cell 63:1066–1079. https://doi.org/10.1016/j.molcel.2016.08.032
49. Quenneville S, Verde G, Corsinotti A, Kapopoulou A, Jakobsson J, Offner S, Baglivo I, Pedone PV, Grimaldi G, Riccio A, Trono D (2011) In embryonic stem cells, ZFP57/KAP1 recognize a methylated Hexanucleotide to affect chromatin and DNA methylation of imprinting control regions. Mol Cell 44:361–372. https://doi.org/10.1016/j.molcel.2011.08.032
50. Lu F, Liu Y, Inoue A, Suzuki T, Zhao K, Zhang Y (2016) Establishing chromatin regulatory landscape during mouse preimplantation development. Cell 165:1375–1388. https://doi.org/10.1016/j.cell.2016.05.050
51. Nicholls RD, Knoll JHM, Butler MG, Karam S, Lalande M (1989) Genetic imprinting suggested by maternal heterodisomy in non-deletion Prader-Willi syndrome. Nature 342:281–285. https://doi.org/10.1038/342281a0
52. Knoll JH, Nicholls RD, Magenis RE, Graham JM, Lalande M, Latt SA (1989) Angelman and Prader-Willi syndromes share a common chromosome 15 deletion but differ in parental origin of the deletion. Am J Med Genet 32:285–290. https://doi.org/10.1002/ajmg.1320320235
53. Junien C (1992) Beckwith-Wiedemann syndrome, tumourigenesis and imprinting. Curr Opin Genet Dev 2:431–438. https://doi.org/10.1016/S0959-437X(05)80154-6
54. Ohlsson R, Nyström A, Pfeifer-Ohlsson S, Töhönen V, Hedborg F, Schofield P, Flam F, Ekström TJ (1993) IGF2 is parentally imprinted during human embryogenesis and in the Beckwith–Wiedemann syndrome. Nat Genet 4:94–97. https://doi.org/10.1038/ng0593-94
55. Kobayashi H, Sakurai T, Imai M, Takahashi N, Fukuda A, Yayoi O, Sato S, Nakabayashi K, Hata K, Sotomaru Y, Suzuki Y, Kono T (2012) Contribution of intragenic DNA methylation in mouse Gametic DNA Methylomes to establish oocyte-specific heritable Marks. PLoS Genet 8:e1002440. https://doi.org/10.1371/journal.pgen.1002440
56. Proudhon C, Duffié R, Ajjan S, Cowley M, Iranzo J, Carbajosa G, Saadeh H, Holland ML, Oakey RJ, Rakyan VK, Schulz R, Bourc'his D (2012) Protection against De novo methylation is instrumental in maintaining parent-of-origin methylation inherited from the gametes. Mol Cell 47:909–920. https://doi.org/10.1016/j.molcel.2012.07.010
57. Ohta T, Gray TA, Rogan PK, Buiting K, Gabriel JM, Saitoh S, Muralidhar B, Bilienska B, Krajewska-Walasek M, Driscoll DJ, Horsthemke B, Butler MG, Nicholls RD (1999) Imprinting-mutation mechanisms in Prader-Willi syndrome. Am J Hum Genet 64:397–413. https://doi.org/10.1086/302233
58. Duffié R, Ajjan S, Greenberg MV, Zamudio N, Escamilla del Arenal M, Iranzo J, Okamoto I, Barbaux S, Fauque P, Bourc'his D (2014) The Gpr1/Zdbf2 locus provides new paradigms for transient and dynamic genomic imprinting in mammals. Genes Dev 28:463–478. https://doi.org/10.1101/gad.232058.113
59. Zhang W, Chen Z, Yin Q, Zhang D, Racowsky C, Zhang Y (2019) Maternal-biased H3K27me3 correlates with paternal-specific gene expression in the human morula. Genes Dev 33:382–387. https://doi.org/10.1101/gad.323105.118
60. Babak T, DeVeale B, Tsang EK, Zhou Y, Li X, Smith KS, Kukurba KR, Zhang R, Li JB, van der Kooy D, Montgomery SB, Fraser HB (2015) Genetic conflict reflected in tissue-specific maps of genomic imprinting in human and mouse. Nat Genet 47:544–549. https://doi.org/10.1038/ng.3274

61. Hanna CW, Peñaherrera MS, Saadeh H, Andrews S, McFadden DE, Kelsey G, Robinson WP (2016) Pervasive polymorphic imprinted methylation in the human placenta. Genome Res 26: 756–767. https://doi.org/10.1101/gr.196139.115
62. Hanna CW (2020) Placental imprinting: emerging mechanisms and functions. PLoS Genet 16. https://doi.org/10.1371/journal.pgen.1008709
63. Bogutz AB, Brind'Amour J, Kobayashi H, Jensen KN, Nakabayashi K, Imai H, Lorincz MC, Lefebvre L (2019) Evolution of imprinting via lineage-specific insertion of retroviral promoters. Nat Commun 10:5674. https://doi.org/10.1038/s41467-019-13662-9
64. Hanna CW, Pérez-Palacios R, Gahurova L, Schubert M, Krueger F, Biggins L, Andrews S, Colomé-Tatché M, Bourc'his D, Dean W, Kelsey G (2019) Endogenous retroviral insertions drive non-canonical imprinting in extra-embryonic tissues. Genome Biol 20:225. https://doi.org/10.1186/s13059-019-1833-x
65. Borgel J, Guibert S, Li Y, Chiba H, Schübeler D, Sasaki H, Forné T, Weber M (2010) Targets and dynamics of promoter DNA methylation during early mouse development. Nat Genet 42: 1093–1100. https://doi.org/10.1038/ng.708
66. Dhayalan A, Rajavelu A, Rathert P, Tamas R, Jurkowska RZ, Ragozin S, Jeltsch A (2010) The Dnmt3a PWWP domain reads histone 3 lysine 36 Trimethylation and guides DNA methylation*. J Biol Chem 285:26114–26120. https://doi.org/10.1074/jbc.M109.089433
67. Zhang Y, Jurkowska R, Soeroes S, Rajavelu A, Dhayalan A, Bock I, Rathert P, Brandt O, Reinhardt R, Fischle W, Jeltsch A (2010) Chromatin methylation activity of Dnmt3a and Dnmt3a/3L is guided by interaction of the ADD domain with the histone H3 tail. Nucleic Acids Res 38:4246–4253. https://doi.org/10.1093/nar/gkq147
68. Bernstein BE, Mikkelsen TS, Xie X, Kamal M, Huebert DJ, Cuff J, Fry B, Meissner A, Wernig M, Plath K, Jaenisch R, Wagschal A, Feil R, Schreiber SL, Lander ES (2006) A bivalent chromatin structure Marks key developmental genes in embryonic stem cells. Cell 125:315–326. https://doi.org/10.1016/j.cell.2006.02.041
69. Verma N, Pan H, Doré LC, Shukla A, Li QV, Pelham-Webb B, Teijeiro V, González F, Krivtsov A, Chang C-J, Papapetrou EP, He C, Elemento O, Huangfu D (2018) TET proteins safeguard bivalent promoters from de novo methylation in human embryonic stem cells. Nat Genet 50:83–95. https://doi.org/10.1038/s41588-017-0002-y
70. Eckersley-Maslin MA, Parry A, Blotenburg M, Krueger C, Ito Y, Franklin VNR, Narita M, D'Santos CS, Reik W (2020) Epigenetic priming by Dppa2 and 4 in pluripotency facilitates multi-lineage commitment. Nat Struct Mol Biol 27:696–705. https://doi.org/10.1038/s41594-020-0443-3
71. Gretarsson KH, Hackett JA (2020) Dppa2 and Dppa4 counteract de novo methylation to establish a permissive epigenome for development. Nat Struct Mol Biol 27:706–716. https://doi.org/10.1038/s41594-020-0445-1
72. Kremsky I, Corces VG (2020) Protection from DNA re-methylation by transcription factors in primordial germ cells and pre-implantation embryos can explain trans-generational epigenetic inheritance. Genome Biol 21:118. https://doi.org/10.1186/s13059-020-02036-w
73. Ehrlich M, Gama-Sosa MA, Huang L-H, Midgett RM, Kuo KC, McCune RA, Gehrke C (1982) Amount and distribution of 5-methylcytosine in human DNA from different types of tissues or cells. Nucleic Acids Res 10:2709–2721. https://doi.org/10.1093/nar/10.8.2709
74. Schroeder DI, Blair JD, Lott P, Yu HOK, Hong D, Crary F, Ashwood P, Walker C, Korf I, Robinson WP, LaSalle JM (2013) The human placenta methylome. Proc Natl Acad Sci 110: 6037–6042. https://doi.org/10.1073/pnas.1215145110
75. Yuan V, Hui D, Yin Y, Peñaherrera MS, Beristain AG, Robinson WP (2021) Cell-specific characterization of the placental methylome. BMC Genomics 22:6. https://doi.org/10.1186/s12864-020-07186-6
76. Zhang B, Kim MY, Elliot G, Zhou Y, Zhao G, Li D, Lowdon RF, Gormley M, Kapidzic M, Robinson JF, McMaster MT, Hong C, Mazor T, Hamilton E, Sears RL, Pehrsson EC, Marra MA, Jones SJM, Bilenky M, Hirst M, Wang T, Costello JF, Fisher SJ (2021) Human placental

cytotrophoblast epigenome dynamics over gestation and alterations in placental disease. Dev Cell 56:1238–1252.e5. https://doi.org/10.1016/j.devcel.2021.04.001
77. Nordor AV, Nehar-Belaid D, Richon S, Klatzmann D, Bellet D, Dangles-Marie V, Fournier T, Aryee MJ (2017) The early pregnancy placenta foreshadows dna methylation alterations of solid tumors. Epigenetics 00–00. https://doi.org/10.1080/15592294.2017.1342912
78. Smith ZD, Shi J, Gu H, Donaghey J, Clement K, Cacchiarelli D, Gnirke A, Michor F, Meissner A (2017) Epigenetic restriction of extraembryonic lineages mirrors the somatic transition to cancer. Nat Publ Group 549:4–3. https://doi.org/10.1038/nature23891
79. Cotton AM, Avila L, Penaherrera MS, Affleck JG, Robinson WP, Brown CJ (2009) Inactive X chromosome-specific reduction in placental DNA methylation. Hum Mol Genet 18:3544–3552. https://doi.org/10.1093/hmg/ddp299
80. Reiss D, Zhang Y, Mager DL (2007) Widely variable endogenous retroviral methylation levels in human placenta. Nucleic Acids Res 35:4743–4754. https://doi.org/10.1093/nar/gkm455
81. Chatterjee A, Macaulay EC, Rodger EJ, Stockwell PA, Parry MF, Roberts HE, Slatter TL, Hung NA, Devenish CJ, Morison IM (2016) Placental Hypomethylation is more pronounced in genomic loci devoid of Retroelements. G3 GenesGenomesGenetics 6:1911–1921. https://doi.org/10.1534/g3.116.030379
82. Cohen CJ, Rebollo R, Babovic S, Dai EL, Robinson WP, Mager DL (2011) Placenta-specific expression of the Interleukin-2 (IL-2) receptor β subunit from an endogenous retroviral promoter. J Biol Chem 286:35543–35552. https://doi.org/10.1074/jbc.M111.227637
83. Macaulay EC, Weeks RJ, Andrews S, Morison IM (2011) Hypomethylation of functional retrotransposon-derived genes in the human placenta. Mamm Genome 22:722–735. https://doi.org/10.1007/s00335-011-9355-1
84. Jaju Bhattad G, Jeyarajah MJ, McGill MG, Dumeaux V, Okae H, Arima T, Lajoie P, Bérubé NG, Renaud SJ (2020) Histone deacetylase 1 and 2 drive differentiation and fusion of progenitor cells in human placental trophoblasts. Cell Death Dis 11:1–17. https://doi.org/10.1038/s41419-020-2500-6
85. Kwak Y-T, Muralimanoharan S, Gogate AA, Mendelson CR (2019) Human trophoblast differentiation is associated with profound gene regulatory and epigenetic changes. Endocrinology 160:2189–2203. https://doi.org/10.1210/en.2019-00144
86. Rugg-Gunn PJ, Cox BJ, Ralston A, Rossant J (2010) Distinct histone modifications in stem cell lines and tissue lineages from the early mouse embryo. Proc Natl Acad Sci 107:10783–10790. https://doi.org/10.1073/pnas.0914507107
87. Lee JT, Jaenisch R (1997) Long-range cis effects of ectopic X-inactivation centres on a mouse autosome. Nature 386:275–279. https://doi.org/10.1038/386275a0
88. Migeon BR (2017) Choosing the active X: the human version of X inactivation. Trends Genet 33:899–909. https://doi.org/10.1016/j.tig.2017.09.005
89. Petropoulos S, Edsgärd D, Reinius B, Deng Q, Panula SP, Codeluppi S, Plaza Reyes A, Linnarsson S, Sandberg R, Lanner F (2016) Single-cell RNA-Seq reveals lineage and X chromosome dynamics in human preimplantation embryos. Cell 165:1012–1026. https://doi.org/10.1016/j.cell.2016.03.023
90. Vallot C, Huret C, Lesecque Y, Resch A, Oudrhiri N, Bennaceur-Griscelli A, Duret L, Rougeulle C (2013) XACT, a long noncoding transcript coating the active X chromosome in human pluripotent cells. Nat Genet 45:239–241. https://doi.org/10.1038/ng.2530
91. Vallot C, Patrat C, Collier AJ, Huret C, Casanova M, Liyakat Ali TM, Tosolini M, Frydman N, Heard E, Rugg-Gunn PJ, Rougeulle C (2017) XACT noncoding RNA competes with XIST in the control of X chromosome activity during human early development. Cell Stem Cell 20:102–111. https://doi.org/10.1016/j.stem.2016.10.014
92. Dossin F, Pinheiro I, Żylicz JJ, Roensch J, Collombet S, Le Saux A, Chelmicki T, Attia M, Kapoor V, Zhan Y, Dingli F, Loew D, Mercher T, Dekker J, Heard E (2020) SPEN integrates transcriptional and epigenetic control of X-inactivation. Nature 578:455–460. https://doi.org/10.1038/s41586-020-1974-9

93. Brockdorff N, Bowness JS, Wei G (2020) Progress toward understanding chromosome silencing by Xist RNA. Genes Dev 34:733–744. https://doi.org/10.1101/gad.337196.120
94. Tjalsma SJD, Hori M, Sato Y, Bousard A, Ohi A, Raposo AC, Roensch J, Le Saux A, Nogami J, Maehara K, Kujirai T, Handa T, Bagés-Arnal S, Ohkawa Y, Kurumizaka H, da Rocha ST, Żylicz JJ, Kimura H, Heard E (2021) H4K20me1 and H3K27me3 are concurrently loaded onto the inactive X chromosome but dispensable for inducing gene silencing. EMBO Rep 22. https://doi.org/10.15252/embr.202051989
95. Tukiainen T, Villani A-C, Yen A, Rivas MA, Marshall JL, Satija R, Aguirre M, Gauthier L, Fleharty M, Kirby A, Cummings BB, Castel SE, Karczewski KJ, Aguet F, Byrnes A, Lappalainen T, Regev A, Ardlie KG, Hacohen N, MacArthur DG (2017) Landscape of X chromosome inactivation across human tissues. Nature 550:244–248. https://doi.org/10.1038/nature24265
96. Balaton BP, Fornes O, Wasserman WW, Brown CJ (2021) Cross-species examination of X-chromosome inactivation highlights domains of escape from silencing. Epigenetics Chromatin 14. https://doi.org/10.1186/s13072-021-00386-8
97. Raznahan A, Parikshak NN, Chandran V, Blumenthal JD, Clasen LS, Alexander-Bloch AF, Zinn AR, Wangsa D, Wise J, Murphy DGM, Bolton PF, Ried T, Ross J, Giedd JN, Geschwind DH (2018) Sex-chromosome dosage effects on gene expression in humans. Proc Natl Acad Sci 115:7398–7403. https://doi.org/10.1073/pnas.1802889115
98. Cotton AM, Price EM, Jones MJ, Balaton BP, Kobor MS, Brown CJ (2015) Landscape of DNA methylation on the X chromosome reflects CpG density, functional chromatin state and X-chromosome inactivation. Hum Mol Genet 24:1528–1539. https://doi.org/10.1093/hmg/ddu564
99. Moreira de Mello JC, de Araújo ESS, Stabellini R, Fraga AM, de Souza JES, Sumita DR, Camargo AA, Pereira LV (2010) Random X inactivation and extensive mosaicism in human placenta revealed by analysis of allele-specific gene expression along the X chromosome. PLoS One 5:e10947. https://doi.org/10.1371/journal.pone.0010947
100. Peñaherrera MS, Jiang R, Avila L, Yuen RKC, Brown CJ, Robinson WP (2012) Patterns of placental development evaluated by X chromosome inactivation profiling provide a basis to evaluate the origin of epigenetic variation. Hum Reprod Oxf Engl 27:1745–1753. https://doi.org/10.1093/humrep/des072
101. Gong S, Johnson MD, Dopierala J, Gaccioli F, Sovio U, Constância M, Smith GC, Charnock-Jones DS (2018) Genome-wide oxidative bisulfite sequencing identifies sex-specific methylation differences in the human placenta. Epigenetics 13:228–239. https://doi.org/10.1080/15592294.2018.1429857
102. Teschendorff AE, Relton CL (2017) Statistical and integrative system-level analysis of DNA methylation data. Nat Rev Genet 19:129–147. https://doi.org/10.1038/nrg.2017.86
103. Houseman EA, Accomando WP, Koestler DC, Christensen BC, Marsit CJ, Nelson HH, Wiencke JK, Kelsey KT (2012) DNA methylation arrays as surrogate measures of cell mixture distribution. BMC Bioinformatics 13:86–86. https://doi.org/10.1186/1471-2105-13-86
104. Wen L, Tang F (2019) Human germline cell development: from the perspective of single-cell sequencing. Mol Cell 76:320–328. https://doi.org/10.1016/j.molcel.2019.08.025
105. Bell JT, Pai AA, Pickrell JK, Gaffney DJ, Pique-Regi R, Degner JF, Gilad Y, Pritchard JK (2011) DNA methylation patterns associate with genetic and gene expression variation in HapMap cell lines. Genome Biol 12:R10–R10. https://doi.org/10.1186/gb-2011-12-1-r10
106. Teh AL, Pan H, Chen L, Ong M-L, Dogra S, Wong J, MacIsaac JL, Mah SM, McEwen LM, Saw S-M, Godfrey KM, Chong Y-S, Kwek K, Kwoh C-K, Soh S-E, Chong MFF, Barton S, Karnani N, Cheong CY, Buschdorf JP, Stünkel W, Kobor MS, Meaney MJ, Gluckman PD, Holbrook JD (2014) The effect of genotype and in utero environment on interindividual variation in neonate DNA methylomes. Genome Res 24:1064–1074. https://doi.org/10.1101/gr.171439.113
107. Czamara D, Eraslan G, Page CM, Lahti J, Lahti-Pulkkinen M, Hämäläinen E, Kajantie E, Laivuori H, Villa PM, Reynolds RM, Nystad W, Håberg SE, London SJ, O'Donnell KJ,

Garg E, Meaney MJ, Entringer S, Wadhwa PD, Buss C, Jones MJ, Lin DTS, MacIsaac JL, Kobor MS, Koen N, Zar HJ, Koenen KC, Dalvie S, Stein DJ, Kondofersky I, Müller NS, Theis FJ, Räikkönen K, Binder EB (2019) Integrated analysis of environmental and genetic influences on cord blood DNA methylation in new-borns. Nat Commun 10:2548. https://doi.org/10.1038/s41467-019-10461-0

108. Gutierrez-Arcelus M, Ongen H, Lappalainen T, Montgomery SB, Buil A, Yurovsky A, Bryois J, Padioleau I, Romano L, Planchon A, Falconnet E, Bielser D, Gagnebin M, Giger T, Borel C, Letourneau A, Makrythanasis P, Guipponi M, Gehrig C, Antonarakis SE, Dermitzakis ET (2015) Tissue-specific effects of genetic and epigenetic variation on gene regulation and splicing. PLoS Genet 11. https://doi.org/10.1371/journal.pgen.1004958
109. Fraser HB, Lam LL, Neumann SM, Kobor MS (2012) Population-specificity of human DNA methylation. Genome Biol 13:R8–R8. https://doi.org/10.1186/gb-2012-13-2-r8
110. Galanter JM, Gignoux CR, Oh SS, Torgerson D, Pino-Yanes M, Thakur N, Eng C, Hu D, Huntsman S, Farber HJ, Avila PC, Brigino-Buenaventura E, LeNoir MA, Meade K, Serebrisky D, Rodríguez-Cintrón W, Kumar R, Rodríguez-Santana JR, Seibold MA, Borrell LN, Burchard EG, Zaitlen N (2017) Differential methylation between ethnic sub-groups reflects the effect of genetic ancestry and environmental exposures. eLife 6: . https://doi.org/10.7554/eLife.20532
111. Hannon E, Spiers H, Viana J, Pidsley R, Burrage J, Murphy TM, Troakes C, Turecki G, O'Donovan MC, Schalkwyk LC, Bray NJ, Mill J (2016) Methylation quantitative trait loci in the developing brain and their enrichment in schizophrenia-associated genomic regions. Nat Neurosci 19:48–54. https://doi.org/10.1038/nn.4182
112. Andrews SV, Ellis SE, Bakulski KM, Sheppard B, Croen LA, Hertz-Picciotto I, Newschaffer CJ, Feinberg AP, Arking DE, Ladd-Acosta C, Fallin MD (2017) Cross-tissue integration of genetic and epigenetic data offers insight into autism spectrum disorder. Nat Commun 8:1011. https://doi.org/10.1038/s41467-017-00868-y
113. Delahaye F, Do C, Kong Y, Ashkar R, Salas M, Tycko B, Wapner R, Hughes F (2018) Genetic variants influence on the placenta regulatory landscape. PLoS Genet 14:e1007785–e1007785. https://doi.org/10.1371/journal.pgen.1007785
114. Gaunt TR, Shihab HA, Hemani G, Min JL, Woodward G, Lyttleton O, Zheng J, Duggirala A, McArdle WL, Ho K, Ring SM, Evans DM, Davey Smith G, Relton CL (2016) Systematic identification of genetic influences on methylation across the human life course. Genome Biol 17:61. https://doi.org/10.1186/s13059-016-0926-z
115. Rando OJ (2016) Intergenerational transfer of epigenetic information in sperm. Cold Spring Harb Perspect Med 6. https://doi.org/10.1101/cshperspect.a022988
116. Sharma U (2019) Paternal contributions to offspring health: role of sperm small RNAs in intergenerational transmission of epigenetic information. Front Cell Dev Biol 7. https://doi.org/10.3389/fcell.2019.00215
117. Heard E, Martienssen RA (2014) Transgenerational epigenetic inheritance: myths and mechanisms. Cell 157:95–109. https://doi.org/10.1016/j.cell.2014.02.045
118. Legoff L, D'Cruz SC, Tevosian S, Primig M, Smagulova F (2019) Transgenerational inheritance of environmentally induced epigenetic alterations during mammalian development. Cell 8. https://doi.org/10.3390/cells8121559
119. Erkek S, Hisano M, Liang C-Y, Gill M, Murr R, Dieker J, Schübeler D, van der Vlag J, Stadler MB, Peters AHFM (2013) Molecular determinants of nucleosome retention at CpG-rich sequences in mouse spermatozoa. Nat Struct Mol Biol 20:868–875. https://doi.org/10.1038/nsmb.2599
120. Jung YH, Sauria MEG, Lyu X, Cheema MS, Ausio J, Taylor J, Corces VG (2017) Chromatin states in mouse sperm correlate with embryonic and adult regulatory landscapes. Cell Rep 18:1366–1382. https://doi.org/10.1016/j.celrep.2017.01.034
121. Lismer A, Siklenka K, Lafleur C, Dumeaux V, Kimmins S (2020) Sperm histone H3 lysine 4 trimethylation is altered in a genetic mouse model of transgenerational epigenetic inheritance. Nucleic Acids Res 48:11380–11393. https://doi.org/10.1093/nar/gkaa712

122. Gapp K, Jawaid A, Sarkies P, Bohacek J, Pelczar P, Prados J, Farinelli L, Miska E, Mansuy IM (2014) Implication of sperm RNAs in transgenerational inheritance of the effects of early trauma in mice. Nat Neurosci 17:667–669. https://doi.org/10.1038/nn.3695
123. Rodgers AB, Morgan CP, Leu NA, Bale TL (2015) Transgenerational epigenetic programming via sperm microRNA recapitulates effects of paternal stress. Proc Natl Acad Sci U S A 112:13699–13704. https://doi.org/10.1073/pnas.1508347112
124. Beck D, Ben Maamar M, Skinner MK (2021) Integration of sperm ncRNA-directed DNA methylation and DNA methylation-directed histone retention in epigenetic transgenerational inheritance. Epigenetics Chromatin 14:6. https://doi.org/10.1186/s13072-020-00378-0
125. Veenendaal MVE, Painter RC, de Rooij SR, Bossuyt PMM, van der Post J a. M, Gluckman PD, Hanson MA, Roseboom TJ (2013) Transgenerational effects of prenatal exposure to the 1944-45 Dutch famine. BJOG Int J Obstet Gynaecol 120:548–553. https://doi.org/10.1111/1471-0528.12136
126. Donkin I, Versteyhe S, Ingerslev LR, Qian K, Mechta M, Nordkap L, Mortensen B, Appel EVR, Jørgensen N, Kristiansen VB, Hansen T, Workman CT, Zierath JR, Barrès R (2016) Obesity and bariatric surgery drive epigenetic variation of spermatozoa in humans. Cell Metab 23:369–378. https://doi.org/10.1016/j.cmet.2015.11.004
127. Marczylo EL, Amoako AA, Konje JC, Gant TW, Marczylo TH (2012) Smoking induces differential miRNA expression in human spermatozoa: a potential transgenerational epigenetic concern? Epigenetics 7:432–439. https://doi.org/10.4161/epi.19794
128. Aarabi M, San Gabriel MC, Chan D, Behan NA, Caron M, Pastinen T, Bourque G, MacFarlane AJ, Zini A, Trasler J (2015) High-dose folic acid supplementation alters the human sperm methylome and is influenced by the MTHFR C677T polymorphism. Hum Mol Genet 24:6301–6313. https://doi.org/10.1093/hmg/ddv338
129. Horsthemke B (2018) A critical view on transgenerational epigenetic inheritance in humans. Nat Commun 9:2973. https://doi.org/10.1038/s41467-018-05445-5

The Role of Epigenetics in the Developmental Origins of Health and Disease

6

Karin B. Michels, Chatura J. Gunasekara, and Robert A. Waterland

Contents

K. B. Michels (✉)
Department of Epidemiology, Fielding School of Public Health, University of California, Los Angeles, CA, USA

Institute for Prevention and Cancer Epidemiology, Faculty of Medicine and Medical Center, University of Freiburg, Freiburg, Germany
e-mail: karin.michels@uniklinik-freiburg.de

C. J. Gunasekara
Department of Pediatrics, USDA/ARS Children's Nutrition Research Center, Baylor College of Medicine, Houston, TX, USA

R. A. Waterland
Department of Pediatrics, USDA/ARS Children's Nutrition Research Center, Baylor College of Medicine, Houston, TX, USA

Department of Molecular & Human Genetics, Baylor College of Medicine, Houston, TX, USA

K. B. Michels (ed.), *Epigenetic Epidemiology*,
https://doi.org/10.1007/978-3-030-94475-9_6

Abstract

The Developmental Origins of Health and Disease (DOHaD) hypothesis posits that the prenatal and early postnatal environments shape the future probability of physical and mental well-being and risk of disease. A wealth of epidemiologic data document associations among maternal and infant nutrition, stress, and other exposures, and risk of chronic disease in later life including cardiovascular disease, hypertension, type 2 diabetes mellitus, obesity, neuropsychiatric disorders, and cancer. Extensive data from animal models support the biological plausibility of the DOHaD hypothesis. While the mechanisms underlying these observations remain unresolved, the DOHaD model assumes developmental plasticity, which allows adaptive regulation of embryonic, fetal, and infant development in response to nutritional and environmental perturbations. Establishment of epigenetic regulation during embryonic, fetal, and early postnatal life coincides with vulnerable ontogenic periods and provides a potential mechanism for long-lasting responses to transient environmental stimuli. In this chapter, we review recent progress in the epigenetic epidemiology of DOHaD and describe emerging approaches aimed at elucidating causal links between early environment, induced epigenetic alterations, and human disease.

Abbreviations

BMI	body mass index
CoRSIV	correlated region of systemic interindividual (epigenetic) variation
CpG	cytosine-guanine dinucleotide
DMR	differentially methylated region
DOHaD	Developmental Origins of Health and Disease
EWAS	epigenome-wide association study
GWAS	genome-wide association study
HM450	Illumina human methylation 450 microarray
IAP	intracisternal A particle
IGF2	insulin-like growth factor 2
IUGR	intrauterine growth retardation
PACE	Pregnancy and Childhood Epigenetics consortium
SIV	systemic interindividual (epigenetic) variation
SPLS-DA	sparse partial least squares discriminate analysis

6.1 The Developmental Origins of Health and Disease

The search for the origins of chronic disease has shifted the focus toward the earliest phases of the life course. While classic epidemiology has targeted lifestyle patterns of adults at various ages, in recent decades the importance of early life for determining lifelong health has been increasingly recognized. Following the seminal work of Rose [1], Forsdahl [2], and Barker [3, 4], the period from conception to birth and the first few years of life are considered critical in influencing disease susceptibility throughout life. This shift in thinking and research gave birth to the "Developmental Origins of Health and Disease" (DOHaD) hypothesis.

Epidemiologic studies support the hypothesis that chronic diseases have their roots in early life. Barker's work linked low birthweight to a number of cardiovascular diseases (including ischemic heart disease), hypertension, cholesterol levels, stroke, and impaired glucose tolerance [4–8]. His findings have been confirmed by other groups in different populations [9–11]. Data from the Dutch Famine in 1944/45, when food rations dropped below 1000 kcal/day for six months, suggest an increased risk of obesity among offspring of mothers exposed to the famine during the first and second trimester [12], glucose intolerance if exposure peaked during late gestation [13], and schizophrenia if conception occurred during the famine [14]. Other maternal characteristics such as maternal weight and malnutrition also increase the risk of coronary heart disease in the offspring [15].

However, there is trouble at both ends of the birthweight spectrum. Like low birthweight, high birthweight is also associated with adult obesity [16]. Similarly, women are more likely to become obese in adulthood if their mother was obese prior to pregnancy and/or had very high or very low gestational weight gain [17]. Furthermore, gestational diabetes (associated with fetal macrosomia) increases the risk of childhood and adult obesity in the offspring [18]. High birthweight is also associated with an elevated risk of several cancers. Numerous epidemiologic studies support the association between high birthweight and increased risk of premenopausal breast cancer [19, 20]. In addition, high birthweight has been linked to childhood leukemia [21], childhood brain tumors [22], and testicular cancer [23].

6.2 DOHaD Mechanisms

The fetal origins hypothesis suggests that perturbations at a critical period of development induce persistent alterations with potentially lifelong consequences. These epidemiologic observations led Hales and Barker to suggest the "thrifty phenotype" hypothesis, which proposes that poor fetal nutrition and growth lead to metabolic reprogramming of glycemic metabolism [24]. This adaptive developmental plasticity allows the fetus to adjust to and survive adverse environments. According to this model, a limited supply of transplacental nutrients compels the fetus to channel nutrients to the most vital organs, namely brain and heart, at the expense of other organs, which may remain underdeveloped and compromised in growth and function [25]. Moreover, permanent insulin resistance may be induced

during development and reduce basal metabolic requirements; this permits survival under suboptimal prenatal and predicted postnatal conditions [26]. Indeed, environmental perturbations may have a long-lasting impact at times of greatest plasticity during growth and development, while decreasing plasticity with increasing age allows less adaptation.

The potential benefits of such so-called predictive adaptive responses depend on the accuracy of the prediction; the cost of inaccurate predictions is high [26]. If developmental conditions that induce intrauterine growth retardation are followed by a resource-rich postnatal environment, high plasma glucose levels will coincide with insulin resistance, greatly increasing the risk for metabolic disease in later life [26]. This "mismatch" between predicted and actual postnatal environment may explain profound long-lasting implications for chronic disease among individuals prenatally exposed to the Dutch Hunger Winter, which lasted only nine months and was followed by normal nutritional availability [27]. Individual variation in sensitivity to mismatch and consequent disease susceptibility is likely due to a variety of factors including genetic variation and the degree of developmental plasticity [25].

6.3 Potential Critical Periods for Developmental Epigenetics

The biologic mechanisms underlying the long-term persistence of DOHaD phenomena are not well understood. Developmental plasticity allows a specific genotype to create alternative phenotypes depending on embryonic, intrauterine, and early postnatal conditions, which may induce lasting changes in chronic disease susceptibility. Among the various potential biologic mechanisms underlying developmental plasticity [28], environmental influences on developmental epigenetics are receiving increasing attention [29]. Epigenetics describes the study of mitotically heritable alterations in gene expression potential that are not mediated by DNA sequence alterations [30]. Essentially, epigenetic regulation involves a repertoire of cell-autonomous molecular modifications that govern selective access to the genetic information; because these are mitotically heritable, they are perpetuated in differentiated tissues. The specific molecular mechanisms that function interactively to heritably regulate chromatin conformation include DNA methylation (which occurs predominantly at cytosines within cytosine-guanine dinucleotides, i.e., CpG sites), various modifications of the histone proteins that package DNA in the nucleus, and autoregulatory DNA binding proteins [31]. The ontogenic periods during which these mechanisms undergo establishment and maturation suggest potential critical periods of environmental sensitivity (Fig. 6.1).

Many studies of epigenetics in DOHaD have focused on genomically imprinted genes. Genomic imprinting is the epigenetic silencing of either the maternal or paternal allele of specific genes by DNA methylation, leading to parent-of-origin-specific expression. Loss of imprinting results in the aberrant biallelic expression of an imprinted gene. Loss of imprinting of fetal growth genes, in particular that encoding insulin-like growth factor 2 (*IGF2*), has been associated with childhood disorders such as Beckwith–Wiedemann syndrome [32, 33], Silver–Russell

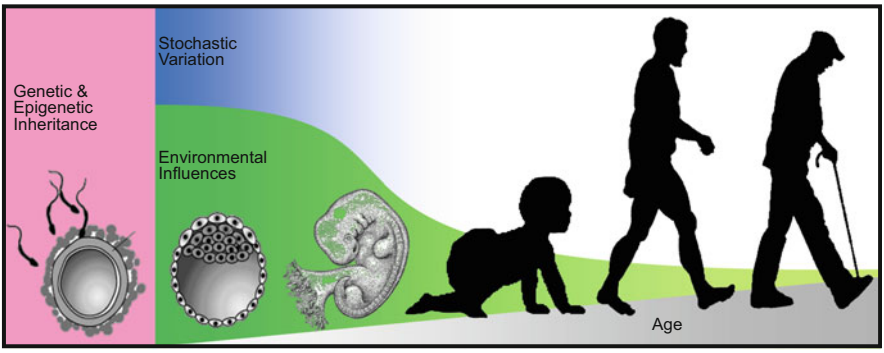

Fig. 6.1 Sources of interindividual variation in the epigenome. Environmental influences on the epigenome are likely most important during establishment of the epigenetic marks in prenatal and early postnatal development. [Reprinted with permission from R. A. Waterland and K. B. Michels: *Annu Rev Nutr* 27:363–388, 2007 [31]]

syndrome [34], and Wilms' tumor [35, 36], as well as with adult-onset diseases [37, 38]. In humans, approximately 100 imprinted genes have been identified. Since most imprinted genes play a role in intrauterine and early life growth, they have long been proposed as good candidates to translate early nutritional and environmental influences into fetal development [29]. Whether epigenetic regulation at imprinted genes is particularly susceptible to early developmental influences remains unresolved [39, 40].

Importantly, the epigenome is established at crucial developmental time points that coincide with vulnerable periods of adaptive plasticity. In the mouse model, each generation undergoes two waves of epigenomic reprogramming. As part of gametogenesis during mid-gestation development, primordial germ cells differentiate into oocyte and sperm [41], assuming distinct epigenomic profiles markedly different from those of somatic tissues. Then, after fertilization, the non-imprinted gene regions in the zygotic genome undergo another round of epigenetic reprogramming that restores totipotency. Genome-wide de novo methylation in the preimplantation embryo [42–45] permits cell fate commitment of the first cell lineages (discussed in more detail in Chap. 5). These dramatic waves of epigenetic reprogramming make mid-gestation and early embryonic development likely critical periods during which nutritional, environmental, and metabolic factors may affect the developmental establishment of epigenetic regulation in the gametes and somatic tissues, respectively.

As a first step toward understanding the role of epigenetic mechanisms in DOHaD, defining the window of susceptibility is crucial. In the mouse, for example, de novo methylation occurs at different times for imprinted and non-imprinted genes and in the developing female and male germline [43] (discussed in more detail in Chap. 5). The DNA methylation signature of non-imprinted genes may be most amenable to environmental stimuli just prior to implantation, when the totipotent blastocyst, largely stripped of genomic methylation, undergoes lineage-specific

remethylation during cellular differentiation. Nutritional and metabolic factors affecting the blastocyst during the early part of the first trimester therefore have great potential to augment or impair the introduction of cytosine methylation. The timing of remethylation of imprinted genes is less clear. By extrapolation from the mouse model, cytosine methylation of the differentially methylated regions (DMRs) of one of the two parental chromosomes is established at different time points for different imprinted genes [46–48]. In the mouse, maternal imprints are established at some point between oocyte development and ovulation [48], and paternal imprints are completed by the time spermatocytes enter meiosis [49]. Whether the establishment of imprinting marks is similar in humans remains to be established.

Periconceptional environmental stressors may yield downstream epigenetic effects in multiple tissues if induced epimutations are maintained during subsequent differentiation; perturbations during late gestation, on the other hand, are more likely to induce cell type-specific epigenetic changes [31]. Further, epigenetic development is not limited to prenatal life; for example, the early postnatal period appears to be a critical period for establishment of DNA methylation in the brain [50].

We have previously proposed two mechanisms to explain environmental influences on the developmental establishment of DNA methylation [31]. First, an imbalance in dietary methyl donors and/or activity of DNA methyltransferases may induce hyper- or hypomethylation. While most transposable elements in the mammalian genome are silenced by CpG methylation [51, 52], some are metastable and can also affect expression of neighboring genes [53]. Such metastable epialleles show large interindividual differences in DNA methylation and gene regulation—even among isogenic individuals—and appear particularly labile in response to environmental stimuli during developmental establishment of the epigenome.

Second, nutritional or environmental stimuli may alter transcriptional activity during periods of de novo DNA methylation, which may permanently alter epigenetic regulation and corresponding phenotypes. Genes actively transcribed during de novo methylation are protected from methylation and remain hypomethylated [54]. Interference with active transcription renders these promoters susceptible to de novo hypermethylation and alters their function [55].

The placenta's critical role in nutrient transfer from mother to fetus makes it particularly vulnerable to adverse intrauterine conditions. Whereas induced epigenetic changes in the soma persist to influence later phenotype, maternal nutrition may also induce epigenetic changes in the placenta, affecting nutrient transport and fetal growth [25]. Imprinted genes are highly expressed in the placenta, which may make them vulnerable to variation in maternal nutrition [56, 57].

6.4 First Clues from the Agouti Mouse

Seminal experiments in the agouti viable yellow (A^{vy}) mouse model support the idea that maternal nutrition can induce developmental programming via epigenetic mechanisms [58]. The *agouti* gene codes for yellow pigment in fur. Transposition of an IAP retrotransposon upstream of *agouti* resulted in the A^{vy} metastable epiallele.

DNA methylation of the retrotransposon exhibits spontaneous interindividual variation, controlling expression of the *agouti* gene and therefore the coat color of the animal. Moreover, supplementation of mouse dams during pregnancy with the dietary methyl donors and cofactors folic acid, vitamin B12, betaine, and choline shifts the coat color distribution of the offspring from yellow to brown [29, 59]. This was shown to occur by induced hypermethylation at the A^{vy} locus [58] systemically and permanently reducing expression of *agouti*.

Similarly, supplementation of the dams with the phytoestrogen genistein results in an analogous coat color shift also mediated through A^{vy} hypermethylation [60]. Maternal methyl donor supplementation studies in another murine metastable epiallele model, the *axin fused* mouse, corroborated the findings in the A^{vy} model [61], indicating that epigenetic regulation at metastable epialleles is generally susceptible to early environmental influences. Putative metastable epialleles are now being identified in humans [62]; as in the mouse models, these human loci show dramatic and systemic interindividual epigenetic variation that is influenced by maternal nutrition around the time of conception [63].

6.5 Epigenetic Epidemiology of DOHaD

Over the past two decades, numerous epidemiologic studies have been performed to explore the role of epigenetics in DOHaD. Within the epigenetic toolbox, DNA methylation is the most likely candidate to explain DOHaD observations due to its relative stability over time; furthermore, it is easiest to study due to its persistence within stored samples. Within the framework of DOHaD, epigenetic studies have addressed either the link between perinatal exposures and DNA methylation at various timepoints throughout life or the link between DNA methylation in early life and later health outcomes [64]. Studies exploring DNA methylation as a mediator connecting early life exposures and later life disease have been sparse [64]. In the following, we highlight some of the studies of particular interest.

Associations between prenatal and early life exposures and DNA methylation are being extensively examined by the Pregnancy and Childhood Epigenetics (PACE) consortium [65] as well as other groups. The best-established association is between maternal smoking during pregnancy and DNA methylation in the offspring cord blood, with a consistent change across most studies found in a CpG in the AHRR gene [66]. Four other CpGs that were changed in cord blood also showed changes in the placenta in a subset of these cohorts that had also collected placenta tissue [67]. In another PACE meta-analysis including 9340 mother–newborn dyads, both very high and very low maternal pre-pregnancy body mass index were linked to several small DNA methylation differences ($<0.2\%$ per BMI unit) in cord blood [68]. Interestingly, maternal alcohol consumption was not found to be associated with offspring cord blood methylation [69]. The PACE consortium and other cohorts have also linked preeclampsia and gestational diabetes to cord blood and placental DNA methylation [70–73].

Maternal self-reported high folic acid supplementation (defined as an average 1200 μg/day or more) was associated with 2.4% lower methylation at the *H19* differentially methylated region (DMR) in umbilical cord blood leukocytes if initiated before pregnancy and 3.7% less methylation at that locus if initiated during pregnancy, compared to mothers not reporting supplementation; no difference was found for the *IGF2* DMR0 [74]. A study from the Netherlands reported 4% higher methylation at one CpG of *IGF2* DMR0 in the blood of 17-month-old children whose mothers reported taking 400 μg folic acid during pregnancy compared to those whose mothers took no folic acid supplements; however, *IGF2* expression levels were not examined [75]. Maternal plasma folate levels during pregnancy were associated with DNA methylation in the cord blood of 1988 newborns [76].

In addition to smoking and folic acid supplementation, studies have linked other intrauterine exposures to DNA methylation in adulthood. Data on survivors of the Dutch Famine suggest that, compared to their unexposed siblings, individuals prenatally exposed to famine had somewhat lower DNA methylation at *IGF2* six decades later [77]. In this study of 60 same-sex sibling pairs, the authors examined five CpGs in the *IGF2* DMR0 and found, on average, 2.7% lower methylation among individuals exposed to famine in utero. Whether this small difference in methylation has any functional consequence remains unclear, in particular since the authors did not examine *IGF2* expression levels. Methylation differences of even smaller magnitude were observed for some other genes including *IL10, GNASAS, INSIGF, LEP*, and *MEG3* [78]. The association between an epigenetic difference assessed in adulthood and a prenatal exposure does not allow causal inference about the induction of that change by the prenatal factor, unless the change is already present directly after the exposure period [28, 31]. Of course, collecting appropriate samples in humans to test such causal pathways is logistically challenging.

Studies on DNA methylation in cord blood or placenta have considered several aspects of weight. A number of studies linked DNA methylation with birthweight with varying results [79–82]; in any event, changes observed did not persist to adulthood. The PACE meta-analyses of EWAS including 8825 neonates from 24 birth cohorts found birthweight associated with DNA methylation in neonatal blood at 914 sites, with a difference in birthweight ranging from −183 to +178 grams per 10% increase in methylation levels [83]. Some studies specifically explored the epigenetic profile of newborns with low birthweight or intrauterine growth retardation (IUGR). Einstein and colleagues compared cord blood samples from five IUGR and five normal pregnancies and identified methylation differences at a restricted number of loci [84]. A few small studies identified differences in methylation or expression of selected imprinted genes in the placenta and cord blood of IUGR or low birthweight compared to normal-weight infants [85–88] and in selected non-imprinted genes [89–91]. Conversely, high birthweight has been associated with increased promoter methylation of the glucocorticoid receptor gene in human placenta [84–88, 92]. Overall, differences in methylation in these studies were small, and it remains unclear whether DNA methylation changes are a cause or consequence of aberrant birthweight. Studies on childhood weight suggested an association between changes in newborn methylation of the *RXRA*

gene and the promoter or the long noncoding RNA *ANRIL* with childhood adiposity [93, 94]. CpG methylation of 68 CpGs in five candidate genes was assessed in umbilical cord tissue from healthy neonates in two prospective cohorts [93]; DNA methylation of one CpG was consistently associated with adiposity at the age of 9 in both cohorts.

In another childhood obesity study applying an array-based genome-scale screen to neonatal blood screening cards, although no statistically significant site emerged comparing the lowest and the highest BMI quartile at age 5, 13 CpG sites showed a $> 5\%$ difference in in DNA methylation levels [95]. All 13 were located in close proximity to the *nc886* gene. This gene, which encodes a small non-coding RNA, shows polymorphic imprinting in neonatal blood which appears to be modifiable by maternal age and nutrition status during pregnancy [40]. Methylation of the differentially methylated region nc886 may operate as a mediator between maternal characteristics and childhood outcomes, although a study demonstrating this link directly remains to be conducted.

Few studies have directly evaluated DNA methylation as a mediator between perinatal exposures and subsequent health outcomes. Cardenas et al. examined whether DNA methylation changes may mediate the association between intrauterine exposure to mercury and lower cognitive performance in childhood [96]. In newborn cord blood of 321 children, they found prenatal mercury levels were associated with lower DNA methylation at the paraoxonase 1 gene, which predicted lower regional cognitive test scores during early childhood. DNA methylation levels at this site, however, were attenuated in blood samples collected in mid-childhood, arguing against direct mediation.

Recently, focus has shifted to studies at the interface between epigenetics and the microbiome to explain DOHaD effect persistence [97]. Similar to epigenetic marks, the gut microbiome is established at birth, but remains malleable to a certain extent by lifestyle factors. The gut microbiome can influence DNA methylation and the activity of DNA methyltransferases and histone deacetylases, although the direction of this crosstalk is not always clear [98–101]. Bacterial metabolites, in particular short chain fatty acids, can function as HDAC inhibitors [102] and correlate with DNA methylation [103].

In summary, epigenetic mechanisms are likely candidates to explain at least some DOHaD phenomena. Nevertheless, despite a recent proliferation of studies in this area, it remains unclear whether the mostly small differences in DNA methylation at birth associated with intrauterine exposures have functional relevance and are maintained into adulthood. Whether the embryonic, intrauterine, and early postnatal environments affect adult disease susceptibility in humans via induced epigenetic alterations remains to be established. Although challenging, longitudinal cohorts assessing links between the periconceptional, pregnancy, and infant environment with adult health and disease status (including measurements of DNA methylation and potentially the gut microbiome at birth and throughout life) are needed to shed more light on these questions. Due to both their malleability by early nutrition and other lifestyle factors, and their noted long-term stability once established, DNA methylation and the gut microbiome hold promise for life-course prevention efforts.

6.6 Challenges for Epigenetic Epidemiology in DOHaD

In 2003 the International HapMap Project set out to identify common sequence variants in the human genome [104]. This "toolbox" enabled large-scale studies to test for associations between these variants and human diseases and phenotypes, heralding the dawn of the genome-wide association study (GWAS) era. In the past two decades, GWASs have identified an impressive and growing number of disease risk-associated genetic variants. Despite this success, however, the majority of individual variance in disease risk remains unexplained, contributing to increased interest in the idea that epigenetic variation could influence the etiology of disease [105–108] and leading to the development of so-called epigenome-wide association studies (EWAS) [109].

But the epigeneticists skipped a crucial step: no "epiHapMap" project was conducted. Rather, the overwhelming majority of the hundreds of "EWAS" studies in the literature employ DNA methylation arrays produced by Illumina (most notably the HM450 and more recently the EPIC850 array). Inexplicably, interindividual variation in DNA methylation was never considered in the design of the Illumina arrays [110, 111]. In fact, most of the probes on these arrays show negligible interindividual variation [112, 113]. A study evaluating the HM450 array in blood, using 256 technical replicates from 130 participants, showed that fewer than half of the CpG sites demonstrated greater interindividual variation than the variation due to technical errors [114]. Another study showed that the power of EWASs could be improved by focusing on the minority of CpG sites with substantial interindividual variation in DNA methylation [115]. A more recent study reported that in peripheral blood DNA, the greatest source of variation at most HM450 probes is intra-individual variability (most likely from variation in leukocyte composition) rather than interindividual variation [116]. The upgrade from the HM450 to the EPIC array in 2016 has not substantially improved the situation. Between HM450 and EPIC arrays, about 55% of the CpG sites show a correlation <0.20, due to low interindividual variability [117]. A recent study that used the EPIC array to examine test-retest reproducibility of peripheral blood DNA methylation of the same women over a one-year period [113] found extremely poor performance (average intraclass correlation coefficient of 0.22), and attributed this to the fact that "99.9% of CpG sites (covered by the array) in the non-sex chromosomes had similar methylation profiles between individuals." These data underscore the unfortunate fact that, over the last decade, over 1000 studies attempting to associate individual epigenetic variation with risk of disease have focused on genomic regions in which DNA methylation is largely invariant.

Another major factor overlooked by the HM450 and EPIC platforms is the cell type specificity of DNA methylation. Generally, we cannot "epigenotype" an individual using peripheral tissues such as blood; epigenetic variation detected in the blood may not be relevant for a disease involving the brain, for example. Reverse causality is another major confounding factor for epigenetic epidemiological studies. Even if the tissue of interest is obtained [118, 119], the disease process itself can

cause epigenetic differences, making it difficult to infer causality. Based on these observations, the designs of HM450 and EPIC arrays are far from ideal.

6.7 The Field Needs to Focus on Systemic Interindividual Epigenetic Variation

A logical approach to overcome these challenges is to focus on genomic regions displaying systemic interindividual variation (SIV) in DNA methylation [62, 120, 121]. A recent study, which could be viewed as a "mini-epiHapMap" project, conducted the largest screening of SIV regions in the human genome [122]. A computational algorithm was developed to analyze deep whole-genome bisulfite-sequencing data on tissues representing all three embryonic germ layers (thyroid, heart, and brain) from each of ten donors from the NIH Genotype-Tissue Expression project [123]. The authors identified 9926 correlated regions of systemic interindividual variation (CoRSIVs). Each CoRSIV is statistically significant ($P < 0.05$), includes at least 5 CpGs, and exhibits an interindividual methylation range of at least 20%. The multiple-tissue interindividual screening approach to identify SIV is similar to that previously used to identify candidate metastable epialleles [62, 120, 124], but unlike metastable epialleles, CoRSIVs are defined without regard to potential genetic influences on their interindividual variation.

Although only <1% of HM450 or EPIC probes are within CoRSIVs, these regions are often associated with a wide range of diseases. For example, the SIV region encompassing *nc886* (also known as *VTRNA2–1*) is a confirmed metastable epiallele; DNA methylation at this locus is influenced by maternal nutrition during periconceptional development [120]. More recently, evidence has emerged demonstrating this region is influenced by maternal alcohol use prior to pregnancy [40]. Additional studies found that methylation in this region is associated with risk of cancer [124, 125], type 2 diabetes [126], and preterm birth [127]. As mentioned above, a prospective study in infants found that *nc886* methylation in peripheral blood at birth predicts BMI at the age of 5 [95]. Consistent with the DOHaD hypothesis, hypermethylation at the *DUSP22* promoter (another CoRSIV) shows an association between in utero famine exposure and schizophrenia [128]. Methylation at a CoRSIV located in the promotor of the *PM20D1* gene has been linked with Alzheimer's disease [129]. More studies have found associations between CoRSIV gene methylation and Parkinson's disease [130], autism [131, 132], major depression and suicide [133], rheumatoid arthritis [134], multiple sclerosis [135], and obesity [136]. Methylation in SIV regions near the *OR2L13* promoter and gene body of *CYP2E1* [124] is associated with maternal gestational diabetes mellitus [71]. Hence, despite their under-representation on the Illumina arrays, CoRSIVs are often among top hits in HM450 and EPIC profiling studies screening for associations with disease and associated phenotypes, indicating immense potential for these regions to contribute to disease prediction, diagnosis, and prognosis. From a DOHaD perspective, a focus on CoRSIVs is particularly warranted, given their

well-documented plasticity to periconceptional environment [28, 62, 63, 121, 122, 137].

In addition to a focus on CoRSIVs, we believe the field will benefit from development of novel analytical approaches. Most studies of DNA methylation and disease have utilized univariate regression methodologies and focused on detecting associations rather than making predictions [138, 139]. It is increasingly recognized, however, that individual CpG sites do not provide as much information as coordinated interactions among multiple CpGs. Multivariate approaches can harness crucial synergistic biological effects [140], motivating increased interest in using machine learning to analyze DNA methylation. Target-capture approaches to study DNA methylation across the entire set of known CoRSIVs are under development. Meanwhile, there are many publicly available HM450 and EPIC datasets, in which ~10% of known CoRSIVs are covered by at least one probe. A recent study [141] took advantage of a publicly available HM450 data set on peripheral blood of schizophrenia (SZ) cases and controls [142] to develop a CoRSIV-focused machine learning classifier based on sparse partial least squares discriminant analysis (SPLS-DA). The model calculated an epigenetic risk score which was able to identify SZ cases with 80% positive predictive value, far surpassing the performance of an analogous SPLS-DA classifier based on polygenic risk score. Additional analyses indicated that these associations were not due to reverse causality, as might be caused by the tendency for SZ patients to smoke heavily and/or take psychotropic medications. Together these findings indicate that the systemic interindividual variants distinguishing SZ cases from controls were present prior to diagnosis; prospective studies will be required to confirm this. Nonetheless, this study provides compelling evidence that a focus on SIV, combined with sophisticated machine learning approaches, may ultimately enable blood-based disease risk prediction for a wide range of complex human diseases, with obvious implications for DOHaD.

6.8 Outlook

An epigenetic basis for DOHaD involves two steps: (1) early environmental influences during critical ontogenic periods can induce lasting epigenetic changes, and (2) these individual epigenetic differences must influence risk of disease later in life. There is now extensive evidence supporting the first step. Particularly in the context of human metastable epialleles and CoRSIVs, it is clear that periconceptional environment affects establishment of DNA methylation states that persist for years [62, 120, 122, 124]. The focus now must be on the second step of the pathway, i.e., establishing causal links between individual epigenetic variation and risk of disease. Despite the "failed start" due to the problems with the Illumina platforms, we believe that an increasing focus on CoRSIVs heralds great potential in the field of epigenetic epidemiology. The systemic nature of interindividual epigenetic variation means that CoRSIVs are essentially epigenetic polymorphisms, facilitating the use of DNA samples from blood, saliva, or buccal cells in large-scale epigenetic epidemiologic studies. Development of commercial

platforms focused on CoRSIVs, coupled with the establishment of prospective longitudinal cohorts, will allow epigenetic epidemiologists to probe causal links between early environment, DNA methylation, and disease.

6.9 Conclusions

The concept of DOHaD arose from epidemiologic studies. Developmental plasticity implies that fetal development adapts to transient nutritional and environmental experiences, resulting in lasting changes in chronic disease susceptibility. While our understanding of the underlying mechanisms is rudimentary, alterations in epigenetic regulation are likely contributors. Although CoRSIVs provide a promising avenue for future DOHaD-centered epigenetic studies, we emphasize that epigenetics is only one of several potential mechanisms explaining developmental plasticity. A better understanding of the mechanisms underlying DOHaD should someday make it possible to reduce individual risk of disease by both preventive strategies targeted to early life and corrective approaches designed to normalize malleable cellular and molecular mechanisms set askew by adverse early exposures.

References

1. Rose G (1964) Familial patterns in ischaemic heart disease. Br J Prev Soc Med 18:75–80. https://doi.org/10.1136/jech.18.2.75
2. Forsdahl A (1977) Are poor living conditions in childhood and adolescence an important risk factor for arteriosclerotic heart disease? Br J Prev Soc Med 31:91–95. https://doi.org/10.1136/jech.31.2.91
3. Barker DJ (1995) Fetal origins of coronary heart disease. BMJ 311:171–174. https://doi.org/10.1136/bmj.311.6998.171
4. Barker DJ, Winter PD, Osmond C et al (1989) Weight in infancy and death from ischaemic heart disease. Lancet 2:577–580. https://doi.org/10.1016/s0140-6736(89)90710-1
5. Barker DJ (1997) Intrauterine programming of coronary heart disease and stroke. Acta Paediatr Suppl 423:178–182.; discussion 183. https://doi.org/10.1111/j.1651-2227.1997.tb18408.x
6. Barker DJ, Bull AR, Osmond C, Simmonds SJ (1990) Fetal and placental size and risk of hypertension in adult life. BMJ 301:259–262. https://doi.org/10.1136/bmj.301.6746.259
7. Barker DJ, Osmond C (1986) Infant mortality, childhood nutrition, and ischaemic heart disease in England and Wales. Lancet 1:1077–1081. https://doi.org/10.1016/s0140-6736(86)91340-1
8. Hales CN, Barker DJ, Clark PM et al (1991) Fetal and infant growth and impaired glucose tolerance at age 64. BMJ 303:1019–1022. https://doi.org/10.1136/bmj.303.6809.1019
9. Curhan GC, Chertow GM, Willett WC et al (1996) Birth weight and adult hypertension and obesity in women. Circulation 94:1310–1315. https://doi.org/10.1161/01.cir.94.6.1310
10. Rich-Edwards JW, Stampfer MJ, Manson JE et al (1997) Birth weight and risk of cardiovascular disease in a cohort of women followed up since 1976. BMJ 315:396–400. https://doi.org/10.1136/bmj.315.7105.396

11. Andersson SW, Lapidus L, Niklasson A et al (2000) Blood pressure and hypertension in middle-aged women in relation to weight and length at birth: a follow-up study. J Hypertens 18:1753–1761. https://doi.org/10.1097/00004872-200018120-00008
12. Ravelli GP, Stein ZA, Susser MW (1976) Obesity in young men after famine exposure in utero and early infancy. N Engl J Med 295:349–353. https://doi.org/10.1056/NEJM197608122950701
13. Ravelli AC, van der Meulen JH, Michels RP et al (1998) Glucose tolerance in adults after prenatal exposure to famine. Lancet 351:173–177. https://doi.org/10.1016/s0140-6736(97)07244-9
14. Susser ES, Lin SP (1992) Schizophrenia after prenatal exposure to the Dutch hunger Winter of 1944-1945. Arch Gen Psychiatry 49:983–988. https://doi.org/10.1001/archpsyc.1992.01820120071010
15. Forsén T, Eriksson JG, Tuomilehto J et al (1997) Mother's weight in pregnancy and coronary heart disease in a cohort of Finnish men: follow up study. BMJ 315:837–840. https://doi.org/10.1136/bmj.315.7112.837
16. Sørensen HT, Sabroe S, Rothman KJ et al (1997) Relation between weight and length at birth and body mass index in young adulthood: cohort study. BMJ 315:1137. https://doi.org/10.1136/bmj.315.7116.1137
17. Stuebe AM, Forman MR, Michels KB (2009) Maternal-recalled gestational weight gain, pre-pregnancy body mass index, and obesity in the daughter. Int J Obes 33:743–752. https://doi.org/10.1038/ijo.2009.101
18. Lawlor DA, Lichtenstein P, Långström N (2011) Association of maternal diabetes mellitus in pregnancy with offspring adiposity into early adulthood: sibling study in a prospective cohort of 280,866 men from 248,293 families. Circulation 123:258–265. https://doi.org/10.1161/CIRCULATIONAHA.110.980169
19. Michels KB, Xue F (2006) Role of birthweight in the etiology of breast cancer. Int J Cancer 119:2007–2025. https://doi.org/10.1002/ijc.22004
20. Michels KB, Trichopoulos D, Robins JM et al (1996) Birthweight as a risk factor for breast cancer. Lancet 348:1542–1546. https://doi.org/10.1016/S0140-6736(96)03102-9
21. Caughey RW, Michels KB (2009) Birth weight and childhood leukemia: a meta-analysis and review of the current evidence. Int J Cancer 124:2658–2670. https://doi.org/10.1002/ijc.24225
22. Harder T, Plagemann A, Harder A (2008) Birth weight and subsequent risk of childhood primary brain tumors: a meta-analysis. Am J Epidemiol 168:366–373. https://doi.org/10.1093/aje/kwn144
23. Michos A, Xue F, Michels KB (2007) Birth weight and the risk of testicular cancer: a meta-analysis. Int J Cancer 121:1123–1131. https://doi.org/10.1002/ijc.22771
24. Hales CN, Barker DJ (2001) The thrifty phenotype hypothesis. Br Med Bull 60:5–20. https://doi.org/10.1093/bmb/60.1.5
25. Hochberg Z, Feil R, Constancia M et al (2011) Child health, developmental plasticity, and epigenetic programming. Endocr Rev 32:159–224. https://doi.org/10.1210/er.2009-0039
26. Rickard IJ, Lummaa V (2007) The predictive adaptive response and metabolic syndrome: challenges for the hypothesis. Trends Endocrinol Metab 18:94–99. https://doi.org/10.1016/j.tem.2007.02.004
27. Gluckman P, Hanson M (2006) Mismatch. Why our world no longer fits our bodies. Oxford University Press, Oxford
28. Waterland RA, Garza C (1999) Potential mechanisms of metabolic imprinting that lead to chronic disease. Am J Clin Nutr 69:179–197. https://doi.org/10.1093/ajcn/69.2.179
29. Waterland RA, Jirtle RL (2004) Early nutrition, epigenetic changes at transposons and imprinted genes, and enhanced susceptibility to adult chronic diseases. Nutrition 20:63–68. https://doi.org/10.1016/j.nut.2003.09.011
30. Jaenisch R, Bird A (2003) Epigenetic regulation of gene expression: how the genome integrates intrinsic and environmental signals. Nat Genet 33(Suppl):245–254. https://doi.org/10.1038/ng1089

31. Waterland RA, Michels KB (2007) Epigenetic epidemiology of the developmental origins hypothesis. Annu Rev Nutr 27:363–388. https://doi.org/10.1146/annurev.nutr.27.061406.093705
32. Ohlsson R, Nyström A, Pfeifer-Ohlsson S et al (1993) IGF2 is parentally imprinted during human embryogenesis and in the Beckwith-Wiedemann syndrome. Nat Genet 4:94–97. https://doi.org/10.1038/ng0593-94
33. Weksberg R, Shen DR, Fei YL et al (1993) Disruption of insulin-like growth factor 2 imprinting in Beckwith-Wiedemann syndrome. Nat Genet 5:143–150. https://doi.org/10.1038/ng1093-143
34. Gicquel C, Rossignol S, Cabrol S et al (2005) Epimutation of the telomeric imprinting center region on chromosome 11p15 in Silver-Russell syndrome. Nat Genet 37:1003–1007. https://doi.org/10.1038/ng1629
35. Ogawa O, Eccles MR, Szeto J et al (1993) Relaxation of insulin-like growth factor II gene imprinting implicated in Wilms' tumour. Nature 362:749–751. https://doi.org/10.1038/362749a0
36. Steenman MJ, Rainier S, Dobry CJ et al (1994) Loss of imprinting of IGF2 is linked to reduced expression and abnormal methylation of H19 in Wilms' tumour. Nat Genet 7:433–439. https://doi.org/10.1038/ng0794-433
37. Cui H, Cruz-Correa M, Giardiello FM et al (2003) Loss of IGF2 imprinting: a potential marker of colorectal cancer risk. Science 299:1753–1755. https://doi.org/10.1126/science.1080902
38. Jelinic P, Shaw P (2007) Loss of imprinting and cancer. J Pathol 211:261–268. https://doi.org/10.1002/path.2116
39. Waterland RA (2009) Early environmental effects on epigenetic regulation in humans. Epigenetics 4:523–525. https://doi.org/10.4161/epi.4.8.10155
40. Carpenter BL, Remba TK, Thomas SL et al (2021) Oocyte age and preconceptual alcohol use are highly correlated with epigenetic imprinting of a noncoding RNA (nc886). Proc Natl Acad Sci U S A 118:e2026580118. https://doi.org/10.1073/pnas.2026580118
41. Morgan HD, Santos F, Green K et al (2005) Epigenetic reprogramming in mammals. Hum Mol Genet 14(1):R47–R58. https://doi.org/10.1093/hmg/ddi114
42. Razin A, Riggs AD (1980) DNA methylation and gene function. Science 210:604–610. https://doi.org/10.1126/science.6254144
43. Reik W, Dean W, Walter J (2001) Epigenetic reprogramming in mammalian development. Science 293:1089–1093. https://doi.org/10.1126/science.1063443
44. Rideout WM, Eggan K, Jaenisch R (2001) Nuclear cloning and epigenetic reprogramming of the genome. Science 293:1093–1098. https://doi.org/10.1126/science.1063206
45. Santos F, Dean W (2004) Epigenetic reprogramming during early development in mammals. Reproduction 127:643–651. https://doi.org/10.1530/rep.1.00221
46. Obata Y, Kono T (2002) Maternal primary imprinting is established at a specific time for each gene throughout oocyte growth. J Biol Chem 277:5285–5289. https://doi.org/10.1074/jbc.M108586200
47. Lucifero D, Mann MRW, Bartolomei MS, Trasler JM (2004) Gene-specific timing and epigenetic memory in oocyte imprinting. Hum Mol Genet 13:839–849. https://doi.org/10.1093/hmg/ddh104
48. Tomizawa S, Kobayashi H, Watanabe T et al (2011) Dynamic stage-specific changes in imprinted differentially methylated regions during early mammalian development and prevalence of non-CpG methylation in oocytes. Development 138:811–820. https://doi.org/10.1242/dev.061416
49. Allegrucci C, Thurston A, Lucas E, Young L (2005) Epigenetics and the germline. Reproduction 129:137–149. https://doi.org/10.1530/rep.1.00360
50. Lister R, Mukamel EA, Nery JR et al (2013) Global epigenomic reconfiguration during mammalian brain development. Science 341:1237905. https://doi.org/10.1126/science.1237905

51. Yoder JA, Walsh CP, Bestor TH (1997) Cytosine methylation and the ecology of intragenomic parasites. Trends Genet 13:335–340. https://doi.org/10.1016/s0168-9525(97)01181-5
52. Yoder JA, Soman NS, Verdine GL, Bestor TH (1997) DNA (cytosine-5)-methyltransferases in mouse cells and tissues. Studies with a mechanism-based probe. J Mol Biol 270:385–395. https://doi.org/10.1006/jmbi.1997.1125
53. Rakyan VK, Blewitt ME, Druker R et al (2002) Metastable epialleles in mammals. Trends Genet 18:348–351. https://doi.org/10.1016/s0168-9525(02)02709-9
54. Bird A (2002) DNA methylation patterns and epigenetic memory. Genes Dev 16:6–21. https://doi.org/10.1101/gad.947102
55. Brandeis M, Frank D, Keshet I et al (1994) Sp1 elements protect a CpG island from de novo methylation. Nature 371:435–438. https://doi.org/10.1038/371435a0
56. Fowden AL, Sibley C, Reik W, Constancia M (2006) Imprinted genes, placental development and fetal growth. Horm Res 65(Suppl 3):50–58. https://doi.org/10.1159/000091506
57. Coan PM, Burton GJ, Ferguson-Smith AC (2005) Imprinted genes in the placenta--a review. Placenta 26(Suppl A):S10–S20. https://doi.org/10.1016/j.placenta.2004.12.009
58. Waterland RA, Jirtle RL (2003) Transposable elements: targets for early nutritional effects on epigenetic gene regulation. Mol Cell Biol 23:5293–5300. https://doi.org/10.1128/MCB.23.15.5293-5300.2003
59. Wolff GL, Kodell RL, Moore SR, Cooney CA (1998) Maternal epigenetics and methyl supplements affect agouti gene expression in Avy/a mice. FASEB J 12:949–957
60. Dolinoy DC, Weidman JR, Waterland RA, Jirtle RL (2006) Maternal genistein alters coat color and protects Avy mouse offspring from obesity by modifying the fetal epigenome. Environ Health Perspect 114:567–572. https://doi.org/10.1289/ehp.8700
61. Waterland RA, Dolinoy DC, Lin J-R et al (2006) Maternal methyl supplements increase offspring DNA methylation at Axin fused. Genesis 44:401–406. https://doi.org/10.1002/dvg.20230
62. Waterland RA, Kellermayer R, Laritsky E et al (2010) Season of conception in rural Gambia affects DNA methylation at putative human metastable epialleles. PLoS Genet 6:e1001252. https://doi.org/10.1371/journal.pgen.1001252
63. Dominguez-Salas P, Moore SE, Baker MS et al (2014) Maternal nutrition at conception modulates DNA methylation of human metastable epialleles. Nat Commun 5:3746. https://doi.org/10.1038/ncomms4746
64. Felix JF, Cecil C a M (2019) Population DNA methylation studies in the developmental origins of health and disease (DOHaD) framework. J Dev Orig Health Dis 10:306–313. https://doi.org/10.1017/S2040174418000442
65. https://www.niehs.nih.gov/research/atniehs/labs/epi/pi/genetics/pace/index.cfm
66. Joubert BR, Felix JF, Yousefi P et al (2016) DNA methylation in Newborns and maternal smoking in pregnancy: genome-wide consortium meta-analysis. Am J Hum Genet 98:680–696. https://doi.org/10.1016/j.ajhg.2016.02.019
67. Everson TM, Vives-Usano M, Seyve E et al (2021) Placental DNA methylation signatures of maternal smoking during pregnancy and potential impacts on fetal growth. Nat Commun 12:5095. https://doi.org/10.1038/s41467-021-24558-y
68. Sharp GC, Salas LA, Monnereau C et al (2017) Maternal BMI at the start of pregnancy and offspring epigenome-wide DNA methylation: findings from the pregnancy and childhood epigenetics (PACE) consortium. Hum Mol Genet 26:4067–4085. https://doi.org/10.1093/hmg/ddx290
69. Sharp GC, Arathimos R, Reese SE et al (2018) Maternal alcohol consumption and offspring DNA methylation: findings from six general population-based birth cohorts. Epigenomics 10:27–42. https://doi.org/10.2217/epi-2017-0095
70. Binder AM, LaRocca J, Lesseur C et al (2015) Epigenome-wide and transcriptome-wide analyses reveal gestational diabetes is associated with alterations in the human leukocyte antigen complex. Clin Epigenetics 7:79. https://doi.org/10.1186/s13148-015-0116-y

71. Howe CG, Cox B, Fore R et al (2020) Maternal gestational diabetes mellitus and Newborn DNA methylation: findings from the pregnancy and childhood epigenetics consortium. Diabetes Care 43:98–105. https://doi.org/10.2337/dc19-0524
72. Leseva MN, Binder AM, Ponsonby A-L et al (2020) Differential gene expression and limited epigenetic dysregulation at the materno-fetal interface in preeclampsia. Hum Mol Genet 29: 335–350. https://doi.org/10.1093/hmg/ddz287
73. Kazmi N, Sharp GC, Reese SE et al (2019) Hypertensive disorders of pregnancy and DNA methylation in Newborns. Hypertension 74:375–383. https://doi.org/10.1161/HYPERTENSIONAHA.119.12634
74. Hoyo C, Murtha AP, Schildkraut JM et al (2011) Methylation variation at IGF2 differentially methylated regions and maternal folic acid use before and during pregnancy. Epigenetics 6: 928–936. https://doi.org/10.4161/epi.6.7.16263
75. Steegers-Theunissen RP, Obermann-Borst SA, Kremer D et al (2009) Periconceptional maternal folic acid use of 400 microg per day is related to increased methylation of the IGF2 gene in the very young child. PLoS One 4:e7845. https://doi.org/10.1371/journal.pone.0007845
76. Joubert BR, den Dekker HT, Felix JF et al (2016) Maternal plasma folate impacts differential DNA methylation in an epigenome-wide meta-analysis of newborns. Nat Commun 7:10577. https://doi.org/10.1038/ncomms10577
77. Heijmans BT, Tobi EW, Stein AD et al (2008) Persistent epigenetic differences associated with prenatal exposure to famine in humans. Proc Natl Acad Sci U S A 105:17046–17049. https://doi.org/10.1073/pnas.0806560105
78. Tobi EW, Lumey LH, Talens RP et al (2009) DNA methylation differences after exposure to prenatal famine are common and timing- and sex-specific. Hum Mol Genet 18:4046–4053. https://doi.org/10.1093/hmg/ddp353
79. Michels KB, Harris HR, Barault L (2011) Birthweight, maternal weight trajectories and global DNA methylation of LINE-1 repetitive elements. PLoS One 6:e25254. https://doi.org/10.1371/journal.pone.0025254
80. Engel SM, Joubert BR, Wu MC et al (2014) Neonatal genome-wide methylation patterns in relation to birth weight in the Norwegian mother and child cohort. Am J Epidemiol 179:834–842. https://doi.org/10.1093/aje/kwt433
81. Haworth KE, Farrell WE, Emes RD et al (2014) Methylation of the FGFR2 gene is associated with high birth weight centile in humans. Epigenomics 6:477–491. https://doi.org/10.2217/epi.14.40
82. Simpkin AJ, Suderman M, Gaunt TR et al (2015) Longitudinal analysis of DNA methylation associated with birth weight and gestational age. Hum Mol Genet 24:3752–3763. https://doi.org/10.1093/hmg/ddv119
83. Küpers LK, Monnereau C, Sharp GC et al (2019) Meta-analysis of epigenome-wide association studies in neonates reveals widespread differential DNA methylation associated with birthweight. Nat Commun 10:1893. https://doi.org/10.1038/s41467-019-09671-3
84. Einstein F, Thompson RF, Bhagat TD et al (2010) Cytosine methylation dysregulation in neonates following intrauterine growth restriction. PLoS One 5:e8887. https://doi.org/10.1371/journal.pone.0008887
85. Apostolidou S, Abu-Amero S, O'Donoghue K et al (2007) Elevated placental expression of the imprinted PHLDA2 gene is associated with low birth weight. J Mol Med (Berl) 85:379–387. https://doi.org/10.1007/s00109-006-0131-8
86. Diplas AI, Lambertini L, Lee M-J et al (2009) Differential expression of imprinted genes in normal and IUGR human placentas. Epigenetics 4:235–240. https://doi.org/10.4161/epi.9019
87. Guo L, Choufani S, Ferreira J et al (2008) Altered gene expression and methylation of the human chromosome 11 imprinted region in small for gestational age (SGA) placentae. Dev Biol 320:79–91. https://doi.org/10.1016/j.ydbio.2008.04.025
88. McMinn J, Wei M, Schupf N et al (2006) Unbalanced placental expression of imprinted genes in human intrauterine growth restriction. Placenta 27:540–549. https://doi.org/10.1016/j.placenta.2005.07.004

89. Chelbi ST, Doridot L, Mondon F et al (2011) Combination of promoter hypomethylation and PDX1 overexpression leads to TBX15 decrease in vascular IUGR placentas. Epigenetics 6: 247–255. https://doi.org/10.4161/epi.6.2.13791
90. Fryer AA, Emes RD, Ismail KMK et al (2011) Quantitative, high-resolution epigenetic profiling of CpG loci identifies associations with cord blood plasma homocysteine and birth weight in humans. Epigenetics 6:86–94. https://doi.org/10.4161/epi.6.1.13392
91. Ferreira JC, Choufani S, Grafodatskaya D et al (2011) WNT2 promoter methylation in human placenta is associated with low birthweight percentile in the neonate. Epigenetics 6:440–449. https://doi.org/10.4161/epi.6.4.14554
92. Filiberto AC, Maccani MA, Koestler D et al (2011) Birthweight is associated with DNA promoter methylation of the glucocorticoid receptor in human placenta. Epigenetics 6:566–572. https://doi.org/10.4161/epi.6.5.15236
93. Godfrey KM, Sheppard A, Gluckman PD et al (2011) Epigenetic gene promoter methylation at birth is associated with child's later adiposity. Diabetes 60:1528–1534. https://doi.org/10.2337/db10-0979
94. Lillycrop K, Murray R, Cheong C et al (2017) ANRIL promoter DNA methylation: a perinatal marker for later adiposity. EBioMedicine 19:60–72. https://doi.org/10.1016/j.ebiom.2017.03.037
95. van Dijk SJ, Peters TJ, Buckley M et al (2018) DNA methylation in blood from neonatal screening cards and the association with BMI and insulin sensitivity in early childhood. Int J Obes 42:28–35. https://doi.org/10.1038/ijo.2017.228
96. Cardenas A, Rifas-Shiman SL, Agha G et al (2017) Persistent DNA methylation changes associated with prenatal mercury exposure and cognitive performance during childhood. Sci Rep 7:288. https://doi.org/10.1038/s41598-017-00384-5
97. Bianco-Miotto T, Craig JM, Gasser YP et al (2017) Epigenetics and DOHaD: from basics to birth and beyond. J Dev Orig Health Dis 8:513–519. https://doi.org/10.1017/S2040174417000733
98. Cortese R, Lu L, Yu Y et al (2016) Epigenome-microbiome crosstalk: a potential new paradigm influencing neonatal susceptibility to disease. Epigenetics 11:205–215. https://doi.org/10.1080/15592294.2016.1155011
99. Yu D-H, Gadkari M, Zhou Q et al (2015) Postnatal epigenetic regulation of intestinal stem cells requires DNA methylation and is guided by the microbiome. Genome Biol 16:211. https://doi.org/10.1186/s13059-015-0763-5
100. Kumar H, Lund R, Laiho A et al (2014) Gut microbiota as an epigenetic regulator: pilot study based on whole-genome methylation analysis. MBio 5:e02113–e02114. https://doi.org/10.1128/mBio.02113-14
101. Paschos K, Allday MJ (2010) Epigenetic reprogramming of host genes in viral and microbial pathogenesis. Trends Microbiol 18:439–447. https://doi.org/10.1016/j.tim.2010.07.003
102. Mischke M, Plösch T (2016) The gut microbiota and their metabolites: potential implications for the host epigenome. Adv Exp Med Biol 902:33–44. https://doi.org/10.1007/978-3-319-31248-4_3
103. Remely M, Aumueller E, Merold C et al (2014) Effects of short chain fatty acid producing bacteria on epigenetic regulation of FFAR3 in type 2 diabetes and obesity. Gene 537:85–92. https://doi.org/10.1016/j.gene.2013.11.081
104. International HapMap Consortium (2003) The international HapMap project. Nature 426:789–796. https://doi.org/10.1038/nature02168
105. Feinberg AP, Irizarry RA (2010) Evolution in health and medicine Sackler colloquium: stochastic epigenetic variation as a driving force of development, evolutionary adaptation, and disease. Proc Natl Acad Sci U S A 107(Suppl 1):1757–1764. https://doi.org/10.1073/pnas.0906183107
106. Petronis A (2010) Epigenetics as a unifying principle in the aetiology of complex traits and diseases. Nature 465:721–727. https://doi.org/10.1038/nature09230

107. Kulis M, Esteller M (2010) DNA methylation and cancer. Adv Genet 70:27–56. https://doi.org/10.1016/B978-0-12-380866-0.60002-2
108. Kulis M, Queirós AC, Beekman R, Martín-Subero JI (2013) Intragenic DNA methylation in transcriptional regulation, normal differentiation and cancer. Biochim Biophys Acta 1829: 1161–1174. https://doi.org/10.1016/j.bbagrm.2013.08.001
109. Rakyan VK, Down TA, Balding DJ, Beck S (2011) Epigenome-wide association studies for common human diseases. Nat Rev Genet 12:529–541. https://doi.org/10.1038/nrg3000
110. Bibikova M, Barnes B, Tsan C et al (2011) High density DNA methylation array with single CpG site resolution. Genomics 98:288–295. https://doi.org/10.1016/j.ygeno.2011.07.007
111. Pidsley R, Zotenko E, Peters TJ et al (2016) Critical evaluation of the Illumina MethylationEPIC BeadChip microarray for whole-genome DNA methylation profiling. Genome Biol 17:208. https://doi.org/10.1186/s13059-016-1066-1
112. Grundberg E, Meduri E, Sandling JK et al (2013) Global analysis of DNA methylation variation in adipose tissue from twins reveals links to disease-associated variants in distal regulatory elements. Am J Hum Genet 93:876–890. https://doi.org/10.1016/j.ajhg.2013.10.004
113. Zaimi I, Pei D, Koestler DC et al (2018) Variation in DNA methylation of human blood over a 1-year period using the Illumina MethylationEPIC array. Epigenetics 13:1056–1071. https://doi.org/10.1080/15592294.2018.1530008
114. Bose M, Wu C, Pankow JS et al (2014) Evaluation of microarray-based DNA methylation measurement using technical replicates: the atherosclerosis risk in communities (ARIC) study. BMC Bioinformatics 15:312. https://doi.org/10.1186/1471-2105-15-312
115. Hachiya T, Furukawa R, Shiwa Y et al (2017) Genome-wide identification of inter-individually variable DNA methylation sites improves the efficacy of epigenetic association studies. NPJ Genom Med 2:11. https://doi.org/10.1038/s41525-017-0016-5
116. Gallego-Paüls M, Hernández-Ferrer C, Bustamante M et al (2021) Variability of multi-omics profiles in a population-based child cohort. BMC Med 19:166. https://doi.org/10.1186/s12916-021-02027-z
117. Logue MW, Smith AK, Wolf EJ et al (2017) The correlation of methylation levels measured using Illumina 450K and EPIC BeadChips in blood samples. Epigenomics 9:1363–1371. https://doi.org/10.2217/epi-2017-0078
118. Wahl S, Drong A, Lehne B et al (2017) Epigenome-wide association study of body mass index, and the adverse outcomes of adiposity. Nature 541:81–86. https://doi.org/10.1038/nature20784
119. Viana J, Hannon E, Dempster E et al (2017) Schizophrenia-associated methylomic variation: molecular signatures of disease and polygenic risk burden across multiple brain regions. Hum Mol Genet 26:210–225. https://doi.org/10.1093/hmg/ddw373
120. Silver MJ, Kessler NJ, Hennig BJ et al (2015) Independent genomewide screens identify the tumor suppressor VTRNA2-1 as a human epiallele responsive to periconceptional environment. Genome Biol 16:118. https://doi.org/10.1186/s13059-015-0660-y
121. Kessler NJ, Waterland RA, Prentice AM, Silver MJ (2018) Establishment of environmentally sensitive DNA methylation states in the very early human embryo. Sci Adv 4:eaat2624. https://doi.org/10.1126/sciadv.aat2624
122. Gunasekara CJ, Scott CA, Laritsky E et al (2019) A genomic atlas of systemic interindividual epigenetic variation in humans. Genome Biol 20:105. https://doi.org/10.1186/s13059-019-1708-1
123. GTEx Consortium (2013) The genotype-tissue expression (GTEx) project. Nat Genet 45:580–585. https://doi.org/10.1038/ng.2653
124. Van Baak TE, Coarfa C, Dugué P-A et al (2018) Epigenetic supersimilarity of monozygotic twin pairs. Genome Biol 19:2. https://doi.org/10.1186/s13059-017-1374-0
125. Dugué P-A, Yu C, McKay T et al (2021) VTRNA2-1: genetic variation, heritable methylation and disease association. Int J Mol Sci 22:2535. https://doi.org/10.3390/ijms22052535

126. Lin C-H, Lee Y-S, Huang Y-Y, Tsai C-N (2021) Methylation status of vault RNA 2-1 promoter is a predictor of glycemic response to glucagon-like peptide-1 analog therapy in type 2 diabetes mellitus. BMJ Open Diabetes Res Care 9:e001416. https://doi.org/10.1136/bmjdrc-2020-001416
127. You Y-A, Kwon EJ, Hwang H-S et al (2021) Elevated methylation of the vault RNA2-1 promoter in maternal blood is associated with preterm birth. BMC Genomics 22:528. https://doi.org/10.1186/s12864-021-07865-y
128. Boks MP, Houtepen LC, Xu Z et al (2018) Genetic vulnerability to DUSP22 promoter hypermethylation is involved in the relation between in utero famine exposure and schizophrenia. NPJ Schizophr 4:16. https://doi.org/10.1038/s41537-018-0058-4
129. Sanchez-Mut JV, Heyn H, Silva BA et al (2018) PM20D1 is a quantitative trait locus associated with Alzheimer's disease. Nat Med 24:598–603. https://doi.org/10.1038/s41591-018-0013-y
130. Young JI, Sivasankaran SK, Wang L et al (2019) Genome-wide brain DNA methylation analysis suggests epigenetic reprogramming in Parkinson disease. Neurol Genet 5:e342. https://doi.org/10.1212/NXG.0000000000000342
131. Dall' Aglio L, Muka T, Cecil CAM et al (2018) The role of epigenetic modifications in neurodevelopmental disorders: a systematic review. Neurosci Biobehav Rev 94:17–30. https://doi.org/10.1016/j.neubiorev.2018.07.011
132. Zhao B-H, Jiang Y, Zhu H et al (2019) Placental Delta-like 1 gene DNA methylation levels are related to mothers' blood glucose concentration. J Diabetes Res 2019:9521510. https://doi.org/10.1155/2019/9521510
133. Murphy TM, Crawford B, Dempster EL et al (2017) Methylomic profiling of cortex samples from completed suicide cases implicates a role for PSORS1C3 in major depression and suicide. Transl Psychiatry 7:e989. https://doi.org/10.1038/tp.2016.249
134. Mok A, Rhead B, Holingue C et al (2018) Hypomethylation of CYP2E1 and DUSP22 promoters associated with disease activity and erosive disease among rheumatoid arthritis patients. Arthritis Rheumatol 70:528–536. https://doi.org/10.1002/art.40408
135. Maltby VE, Lea RA, Sanders KA et al (2017) Differential methylation at MHC in CD4+ T cells is associated with multiple sclerosis independently of HLA-DRB1. Clin Epigenetics 9:71. https://doi.org/10.1186/s13148-017-0371-1
136. Kühnen P, Handke D, Waterland RA et al (2016) Interindividual variation in DNA methylation at a putative POMC metastable Epiallele is associated with obesity. Cell Metab 24:502–509. https://doi.org/10.1016/j.cmet.2016.08.001
137. Estill MS, Bolnick JM, Waterland RA et al (2016) Assisted reproductive technology alters deoxyribonucleic acid methylation profiles in bloodspots of newborn infants. Fertil Steril 106:629–639.e10. https://doi.org/10.1016/j.fertnstert.2016.05.006
138. Bzdok D, Varoquaux G, Steyerberg EW (2021) Prediction, not association, paves the road to precision medicine. JAMA Psychiatry 78:127–128. https://doi.org/10.1001/jamapsychiatry.2020.2549
139. Rutherford S (2020) The promise of machine learning for psychiatry. Biol Psychiatry 88:e53–e55. https://doi.org/10.1016/j.biopsych.2020.08.024
140. Wiemken TL, Kelley RR (2020) Machine learning in epidemiology and health outcomes research. Annu Rev Public Health 41:21–36. https://doi.org/10.1146/annurev-publhealth-040119-094437
141. Gunasekara CJ, Hannon E, MacKay H et al (2021) A machine learning case-control classifier for schizophrenia based on DNA methylation in blood. Transl Psychiatry 11:412. https://doi.org/10.1038/s41398-021-01496-3
142. Hannon E, Dempster E, Viana J et al (2016) An integrated genetic-epigenetic analysis of schizophrenia: evidence for co-localization of genetic associations and differential DNA methylation. Genome Biol 17:176. https://doi.org/10.1186/s13059-016-1041-x

Epigenetics and Assisted Reproductive Technologies

7

Suneeta Senapati, Sneha Mani, Jayashri Ghosh, Christos Coutifaris, Carmen Sapienza, and Monica Mainigi

Contents

Abstract

The human genome undergoes several phases of epigenetic programming during gametogenesis and early embryo development. The myriad of exposures unique to assisted reproductive technologies (ART) including superovulation,

S. Senapati · S. Mani · C. Coutifaris · M. Mainigi
Department of Obstetrics & Gynecology, University of Pennsylvania, Philadelphia, PA, USA
e-mail: Suneeta.Senapati@pennmedicine.upenn.edu; srmani@pennmedicine.upenn.edu; Christos.Coutifaris@pennmedicine.upenn.edu; Monica.Mainigi@pennmedicine.upenn.edu

J. Ghosh · C. Sapienza (✉)
Fels Institute for Cancer Research and Molecular Biology, Lewis Katz School of Medicine, Temple University, Philadelphia, PA, USA
e-mail: jayashri.ghosh@temple.edu; sapienza@temple.edu

K. B. Michels (ed.), *Epigenetic Epidemiology*,
https://doi.org/10.1007/978-3-030-94475-9_7

fertilization procedures, embryo culture variations, embryo biopsy, and gamete/embryo cryopreservation all occur during this period of intense global reprogramming. While the vast majority of children born after assisted reproductive technologies are healthy, epidemiologic studies have demonstrated that ART is associated with a number of adverse perinatal outcomes including hypertensive disorders of pregnancy, preterm delivery, low birthweight, and an increased prevalence of imprinting disorders. Data from animal models support both global and gene-specific changes after ART exposures. Further studies including prospective longitudinal human cohort studies are needed to delineate the effects of these exposures on the epigenome, the long-term impacts on immediate offspring, and transgenerational effects.

Abbreviations

ART	assisted reproductive technologies
AS	Angelman syndrome
BWS	Beckwith–Wiedemann Syndrome
DNA	deoxyribonucleic acid
ICSI	intracytoplasmic sperm injection
IVF	in vitro fertilization
PGT	pre-implantation genetic testing
RNA	ribonucleic acid

7.1 Introduction

Since the delivery of the first child after in vitro fertilization in 1978, the utilization of assisted reproductive technologies (ART) for family building and fertility preservation has dramatically increased. ART children now comprise 1.8% of all births in the USA [1]; nearly 4% of all births in other countries [2] and more than eight million children have been born using ART [3–5]. While most children born after assisted reproductive technologies are healthy, ART has been associated with a number of adverse perinatal outcomes (independent of the higher incidence of multiple pregnancy), and an increased prevalence of imprinting disorders, including Beckwith–Wiedemann syndrome (BWS), Prader–Willi syndrome, Silver–Russel syndrome, and Angelman syndrome (AS) [6]. In the past two decades, we have also seen a parallel rise in fertility preservation (oocyte, sperm, and embryo cryopreservation), as well as third-party reproduction (donor oocyte, use of gestational carriers), which may bring new considerations as to the effects of extra vitrification/thaw steps, as well as fetal–maternal genomic compatibility. As such, the potential epigenetic impact of assisted reproductive technologies continues to evolve with changes in clinical care.

7.2 Biologic Plausibility and Methods of Epigenetic Investigations Peri-Conception

The biologic plausibility for epigenetic perturbations in assisted reproductive technologies is rooted in the dynamic changes observed in DNA methylation that occur during gametogenesis and early embryonic development. Oocytes and sperm are both highly methylated prior to fertilization. After fertilization occurs, the paternal genome is actively demethylated, while the maternal genome undergoes passive demethylation. In early embryogenesis, both the maternal and paternal contributions to the embryonic genome undergo remethylation [7, 8].

Assisted reproductive technologies, and specifically in vitro fertilization (IVF), involve multiple exposures including superovulation, fertilization procedures, embryo manipulation, embryo culture, and embryo transfer, exposing oocytes to an altered hormonal environment and embryos to changes in temperature, pH, and oxygen tension. These exposures transpire during the global reprogramming of the epigenome in early embryogenesis and implantation and the observed epigenetic changes persistent into adulthood, presumably providing the foundation to alter expression and clinical phenotypes observed throughout life [9]. The predominance of studies of epigenetic modification in ART have been performed via DNA methylation assessments of cord blood or placenta in human studies. Initial studies in human cohorts were restricted to candidate gene approaches focusing primarily on imprinted genes and imprinting control regions (ICRs) analyzed in cord blood, peripheral blood, buccal samples, and early/late gestation placental tissues [10–36]. While these findings seem plausible, there is conflicting data when human samples have been analyzed [2, 13, 20, 26, 27]. This demonstrates the inherent limitations of human studies. Notably, it is nearly impossible to completely control for potential known and unknown genetic and environmental confounders in human samples. Additionally, there are some concerns that epigenetic findings are highly cell type-specific, with some tissue types like the placenta demonstrating high levels of variability in DNA methylation. Extrapolating the impact on clinical phenotype (s) observed from accessible human tissue samples may be limited. Furthermore, some of these studies were cross-sectional in nature or otherwise limited in sample size; thus, study design considerations may further influence the conflicting findings observed. Finally, epigenetic investigations have primarily focused on changes in DNA methylation, due to the availability of cost-effective methods to profile genome-wide change. In the past 10 years, additional focus on other techniques of assessing epigenetic modifications including histone modifications and noncoding RNA (e.g., miRNA) expression has supplemented our understanding.

To address some of the limitations inherent to human studies, animal models are frequently used to examine the effect of ART. These models offer many advantages, foremost among them being the ability to use inbred strains and strictly controlled environments, which limit confounding effects and potential bias. However, animal models also come with challenges. These include basic physiological differences when it comes to the metabolomic requirement of embryos during development and overall developmental kinetics [37]. Additionally ART protocols differ between species. These factors, among others, must be kept in mind when extrapolating

results from animal models. Despite these differences, animal studies have mirrored phenotypic and epigenetic changes seen in humans and have offered a deeper understanding of the changes seen. The ability to use whole genome-wide techniques in controlled environments in animal studies to study tissue-specific effects has helped to solidify the associations noted from the clinical epidemiologic literature and further elucidate the mechanisms involved. A particularly powerful strategy to consider in future studies would be the use of parallel human and mouse studies when investigating the epigenetic impact of ART, which would offer unprecedented value in identifying translatable findings that can be examined in unbiased and manipulatable models.

7.3 Imprinting Disorders

Investigations regarding potential epigenetic perturbations in assisted reproductive technologies largely originate from clinical observations of an increased risk of imprinting disorders in children born after ART. Genomic imprinting is a method of epigenetic regulation in which the expression of a given gene or chromosomal region is dependent on differential allele expression based on the sex of the parent of origin. Importantly, this regulation does not alter the underlying DNA sequence, but alterations in these imprinting regions, particularly during the critical periods of gametogenesis and embryonic implantation, can disrupt normal growth and development leading to imprinting disorders. Interestingly, although there are over one hundred known imprinted genes, only a few have been associated with clinical syndromes.

The relationship between ART and imprinting disorders was first inferred by a case series of 2 children who were conceived by in vitro fertilization (IVF) with intracytoplasmic sperm injection (ICSI) who were diagnosed with Angelman syndrome—a disorder characterized by severe mental retardation, delayed motor development, poor balance, and absence of speech, with a pleasant disposition. DNA methylation and Southern blot analyses confirmed a sporadic imprinting defect in each child in the maternally expressed allele of the UBE3A gene on chromosome 15 [38]. Since that time, several cohorts have reported on the incidence of imprinting disorders with conflicting results. A recent meta-analysis of 23 studies reported positive associations between ART and four imprinting disorders: Angelman syndrome, summary odds ratio (sOR) = 4.7 (95% confidence interval (CI) 2.6–8.5, 4 studies); Beckwith–Wiedemann syndrome, sOR = 5.8 (95% CI 3.1–11.1, 8 studies); Prader–Willi Syndrome, sOR = 2.2 (95% CI 1.6–3.0, 6 studies); and Silver–Russell Syndrome, sOR = 11.3 (95% CI 4.5–28.5, 3 studies), with no differences in retinoblastoma [6]. A binational register-based cohort from the CoNARTaS group including over 74,000 children born over 24 years demonstrated an increased risk of imprinting disorders overall among children born after ART compared to unassisted conceptions (AOR 1.35 [95% CI: 0.80–2.29]); however, it was noted that this finding was mostly driven by an elevated incidence of Beckwith–Wiedemann syndrome specifically, with no significant differences noted in the incidence of Angelman syndrome, Silver–Russell syndrome, and Prader–Willi syndrome [39].

7.4 Potential Exposures in in Vitro Fertilization

In vitro fertilization (IVF) presents a number of opportunities for potential epigenetic perturbation. The process of IVF involves ovarian stimulation with exogenous gonadotropin hormones, followed by retrieval of oocytes under sedation, subsequent fertilization by sperm in the embryology laboratory (either by conventional or intracytoplasmic sperm injection), and development of embryos in culture prior to transfer into the uterus. Embryos can be transferred in a superovulated environment (fresh transfer), or can be cryopreserved for transfer in a more physiologic environment in the future (frozen transfer). Additionally, with the advent of fertility preservation techniques, oocytes can be cryopreserved (either at the MII stage or earlier with in vitro maturation techniques) and later warmed for fertilization and subsequent embryo transfer, introducing another potential exposure. Furthermore, with the advent of pre-implantation genetic testing, trophectoderm biopsy introduces yet another exposure that may impact the epigenome. In the following sections, we explore the impact of each of these steps or interventions on the epigenome and subsequent phenotypes observed (Fig. 7.1).

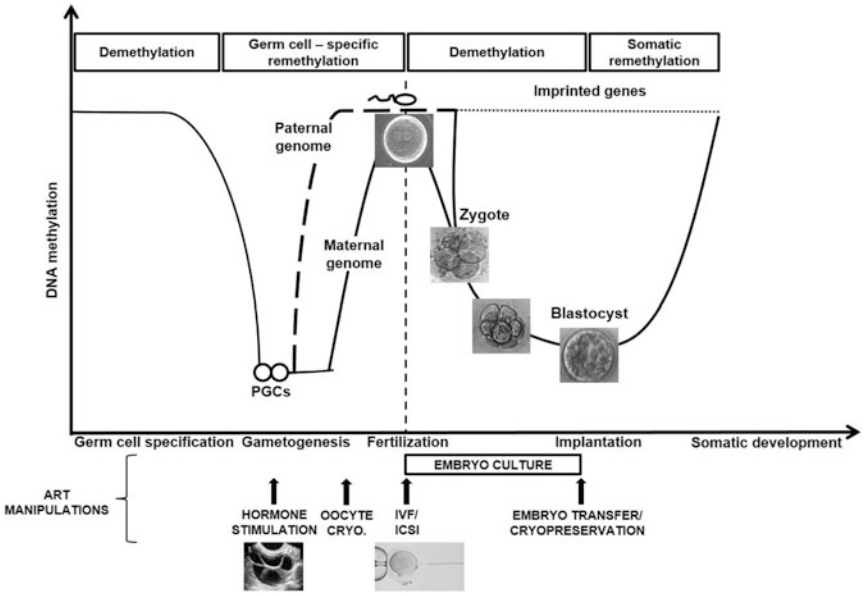

Fig. 7.1 *Schematic representation of DNA methylation dynamics showing concurrent timing of manipulations used in ARTs.* After global DNA demethylation takes place during primordial germ cell (PGC) specification, sex-specific methylation patterns are established. Initiation of remethylation occurs earlier on in the paternal genome (long dashed line) than the maternal genome. Superovulation occurs during the acquisition of maternal methylation. After fertilization (dashed line), genome-wide methylation, except within imprinted genes (round dotted line), occurs before somatic remethylation patterns are set up. This process coincides with hormone stimulation, oocyte cryopreservation, insemination techniques such as in vitro fertilization (IVF) and intracytoplasmic sperm injection (ICSI, embryo culture, and embryo transfer/embryo cryopreservation

7.5 Hormonal Stimulation

Superovulation, or controlled ovarian stimulation, involves the administration of exogenous gonadotropins to stimulate follicular growth and is a necessary step to allow for the retrieval of mature oocytes for IVF. Exposure of the gametes, embryos, and endometrium to these supraphysiologic hormonal levels has been noted to have an effect on the epigenome. Specifically, multiple animal studies have demonstrated the direct effects of superovulation on epigenetic programming during oogenesis, both with global methylation and at specific loci including *KCNQ1OT1* and *KvDMR1* [40–44]. A study examining both a murine model and adult human oocytes under the influence of superovulation demonstrated that superovulated oocytes from infertile women demonstrated a gain of *H19* methylation and a loss of *PEG1* methylation [42]. Changes in *H19* methylation, the *H19* promoter, and *H19* expression after superovulation have been reported by other groups as well [45]. Examinations of early cleavage stage embryos formed after superovulation in a murine model also demonstrated aberrant DNA methylation patterns (as determined by immunofluorescence patterns of an anti-5meC antibody) with greater prevalence in superovulated embryos compared to embryos from naturally cycling females (20 vs 10%, $p < 0.05$) [46]. A mouse model designed to isolate the effects of superovulation from embryo culture found that superovulation resulted in a loss of *Snrpn*, *Peg3*, and *Kcnq1ot1* imprinted methylation, and a gain of imprinted *H19* methylation in pre-implantation embryos. The perturbations noted were dose-dependent with higher gonadotropin dosing associated with more frequent dysregulation [40]. Of note, both maternal and paternal alleles were altered, indicating a potential effect on maintenance of imprinting marks post-fertilization. A study designed to isolate the effects of superovulation from the direct effects of superovulation on the uterine environment in a murine model demonstrated that *Peg3* DNA methylation levels were decreased in placentas derived from exposure to superovulation both pre- and peri-implantation compared with unexposed embryos and exposure of the pre-implantation embryo only, with RNA sequencing implicating genes involved in immune system regulation, specifically interferon signaling, as potential drivers of the differences in fetal and placental growth observed [47]. Furthermore, changes in DNA methylation may differentially impact embryonic and extraembryonic tissues. In a study designed to isolate the effect of superovulation and examine both embryonic and placental DNA methylation changes mid-gestation in a mouse model, superovulation was associated with biallelic expression of *Snrpn* and *H19* in 9.5 dpc placentas, while *Kcnq1ot1* was unaffected in placentas, and all three imprinted genes had normal monoallelic expression in the mouse embryos [48].

Histone modification studies in animal models have shown conflicting results regarding the impact of superovulation. Huffman et al. reported that superovulation doubled the acetylation abundance at H3K9 and H3K14 in embryos after superovulation compared to naturally ovulating controls [49], while another group specifically examining the level of H3K9me2 in mouse zygotes demonstrated no difference in histone modification after administration of exogenous hormones compared to controls [50]. Yet, another group specifically looking at histone acetylation in a

murine model of early vs late embryonic development demonstrated that repeated superovulation reduced H4K12 and H4K16 acetylation in pronuclei, while H3K9me2 and H3K27me3 were increased in four-cell embryos and blastocysts [51].

Few studies have examined potential embryonic epigenetic effects of hormonal stimulation captured by miRNA expression. Epigenetic effects on the endometrium, as assessed by DNA methylation and miRNA expression, similarly show conflicting results. In a murine model, ovarian stimulation with HMG followed by progesterone demonstrated an increase in endometrial miR-16-5p, VEGF protein expression, and angiogenesis. A study of human endometrium biopsied in the luteal phase after superovulation in oocyte donors compared to naturally cycling controls demonstrated 785 genes with differential methylation. Altered DNA methylation did not correlate with gene expression, although there were differences in expression noted in genes involved in endometrial remodeling including *PLAT*, *HSPE2*, *MMP2*, and *TIMP1* [52].

7.6 Culture Conditions (Media, Duration, Temp, pH)

In vitro embryo culture is an obligatory intervention performed in all ART-associated procedures and involves culture of fertilized embryos in a variety of commercially produced (or in-house proprietary) media for 3–5 days. Several laboratory-controlled variables, including media used for culture, pH, oxygen tension, and temperature, have the ability to impact embryo culture and act as stressors that influence embryo development and viability [53]. Of these, most studies have focused on the impact of different commercially available media used for culture, both in terms of composition (carbohydrate, amino acid, macromolecular, vitamin, growth factor content) and strategy used (sequential vs single-step or continuous) [54, 55]. In this section, we summarize existing data on how embryo culture itself, and culture media specifically, could change the epigenome. Due to ethical considerations regarding human embryo experimentation, and difficulties in isolating individual exposures utilized in ARTs, studies examining epigenetic perturbation due to embryo culture so far have primarily utilized animal models. Animal studies are also powerful as they allow for examining the effects of embryo culture in isolation, without the confounding factor of IVF. Most current studies examine DNA methylation changes as a proxy for epigenetic change, primarily due to the existence of affordable and widely available technologies that are able to investigate DNA methylation at high resolution even in a limited number of cells [56]. There is limited knowledge regarding the effects of embryo culture on histone modifications and small noncoding RNAs, though the increasing use of technologies such as CHIP-seq could change this. We will thus focus our summary on changes in DNA methylation due to in vitro culture, including information on alternate epigenetic features as available. Strategies used for examination of DNA methylation changes vary—due to the known association between ARTs and imprinting disorders [57, 58]; some studies approach methylation profiling in an unbiased manner, examining epigenetic change on a whole genome scale, while others choose a targeted approach, focusing on selected imprinted genes (Table 7.1). We separate out current knowledge based

Table 7.1 Summary of studies that examine DNA methylation changes due to in vitro culture

Type of analysis	Exposure	Species	Length of culture	Method	Summarized changes	Reference
Whole genome	-SO+no culture - SO + culture in M16 (mice), culture in M16 and mR1ECM (rat)	Mouse +rat	Zygote to 2C	IF 5-methylcytosine ab	Higher DNA methylation found in vitro vs in vivo in both mouse and rat	[59]
Whole genome	- in vivo - culture in B 2—Culture in B2 + 2.5% FCS	Rabbit	Zygote to 2-, 4-,8-, 16-cell stage, morula, or blastocyst	IF 5-methylcytosine ab	Demethylation occurred earlier in vitro than in vivo between the 2- and 8-cell stages in in vitro cultured vs. in vivo-developed embryos; kinetics were more normal without serum	[60]
Whole genome	- in vivo—Culture in G1+/G2+ media (sequential) vs culture in global media (one step)	Rabbit	Zygote to blastocyst	IF 5-methylcytosine ab; RT-qPCR	Neither culture medium showed similarity to in vivo condition. Global media showed overall hypermethylation, G1+/G2+ showed increase instead of decrease in methylation between 4- and 8-cell stages; aberrant expression of endogenous retroviral sequences with both media	[88]
Whole genome	In vivo vs SO vs SO + IVF + culture in M16, HH, M199M media	Mouse	2C to blastocyst	IF 5mC ab (i.e., immunofluorescence staining with an antibody against 5-methylcytosine—MeC)	Embryos after culture showed higher percentage of blastocysts with abnormal methylation patterns. Embryos cultured in HH showed the highest % of abnormal methylation patterns	[46]

Whole genome	SO+IVF + culture in HTF vs SO+ no culture	Mouse	Zygote to blastocyst	Methylated DNA immunoprecipitation sequencing, RNA-seq	400 genes hypo/ hypermethylated with corresponding changes in gene expression at E7.5 and E12.5, functionally involved in cytoskeleton maintenance, vasculogenesis, energy and amino acid metabolism	[62]
Whole genome	SO + in vivo vs SO + culture in Whitten's vs SO + culture in KSOM/AA	Mouse	Zygote to blastocyst	Affymetrix MOE430A Chip	Culture in KSOMaa led to gene expression that more closely resembled in vivo situation. Only 29 genes changed vs 114 when cultured in Whitten's	[64]
Whole genome	In vivo vs IVM + IVF + culture till zygote, 4C, 16C, or blastocyst	Cow	Zygote, 4C,16C, or blastocyst	EmbryoGENE DNA methylation array platform	Highest DNA methylation dysregulation found when cultured to blastocyst stage followed by 16C and 4C	[72]
Whole genome	IVM + IVF + culture in NCSU23 vs IVM + IVF + culture with natural reproductive fluids	Pig		Whole genome bisulfite sequencing	Embryos cultured in natural reproductive fluids have methylation and expression patterns closer to in vivo blastocysts	[74]
Targeted—Selected imprinted gene	In vivo vs culture	Sheep	5 days	Restriction enzyme digests	Hypomethylation and increased expression of IGF2R after in vitro culture	[78]

(continued)

Table 7.1 (continued)

Type of analysis	Exposure	Species	Length of culture	Method	Summarized changes	Reference
Targeted—Selected imprinted genes	SO + no culture vs SO + IVF + culture in HTF	Mouse	Zygote to morula or blastocyst	Allele-specific bisulfite PCR; chromatin immunoprecipitation (ChIP) assays	In ES cells derived from in vitro cultured blastocysts, hypermethylation of ICR controlling IGF2-H19; IGF2-H19 also showed increased H3 lysine-4 me (paternal) and H3K9me (maternal)	[84]
Targeted—Selected imprinted gene	SO + culture in Whitten's vs SO + culture in KSOM aa	Mouse	2C to blastocyst	Methylation status of HhaI site within H19 region by PCR	Loss of H19 methylation only when culture in Whitten's, Snrpn unchanged	[82]
Targeted—Selected imprinted genes	Culture in M16 vs SO + culture in M16 + 5% FCS vs SO + culture in M16 + 10% FCS	Mouse	1C/2C to blastocyst	Restriction enzyme digests	Presence of serum caused H19/IgF2 ICR hypermethylation and aberrant expression of H19, Igf2, and Grb10 imprinted genes	[83]
Targeted—Selected imprinted genes	Culture in Whitten's vs culture in KSOMaa	Mouse	2C to blastocyst	Bisulfite modification, PCR amplification, subcloning, and sequencing	H19 hypomethylation only in subset of blastocysts placental tissues showed loss of imprinting for H19, Ascl2, Snrpn, Peg3, and Xist, while in the embryo showed preserved imprinted expression	[86]

Targeted—Selected imprinted genes	In vivo vs SO + culture in Whitten's vs SO + culture in KSOMaa vs SO + culture in global (non-sequential) vs SO + HTF (non-sequential) vs SO + culture in preimplantation 1/multiblast (sequential) vs SO + culture in G1v5PLUS/G2v5PLUS (non-sequential)	Mouse	2C to blastocyst	Bisulfite modification, PCR amplification, and sequencing	H19 (paternal), Snrpn, and peg 3 (maternal) hypomethylation in all culture systems	[87]
Targeted—Selected imprinted gene	In vivo vs SO vs SO + culture in M16 vs SO + culture in G1.2/G2.2 vs SO + IVF + culture in M16 vs SO + IVF + culture in G1.2/G2.2	Mouse	Zygote to blastocyst	Bisulfite modification, PCR amplification, and sequencing	Methylation abnormalities of H19 were only found after IVF, but was higher in M16 medium than in the sequential medium	[45]
Targeted—Selected imprinted genes	In vivo vs SO + embryo transfer vs SO + culture in KSOMaa + embryo transfer	Mouse	2C to blastocyst	RT-PCR and allele-specific restriction enzyme digests Bisulfite modification, nested PCR amplification, and sequencing	Embryo culture led to loss of imprinting in extraembryonic and placental tissue that persisted up to E9.5 (83% of yolk sacs and 82% of placenta after ET and EC showed loss of imprinting in at least 1 gene vs. 27% of yolk sacs and 61% of placenta after ET only). EC also led to dysregulated imprinted gene expression (Igf2 and Ascl2)	[85]

(continued)

Table 7.1 (continued)

Type of analysis	Exposure	Species	Length of culture	Method	Summarized changes	Reference
Targeted—Chromosome 7	SO + no culture vs SO + KSOMaa + BSA	Mouse	Zygote to blastocyst	Microarray, NimbleGen custom array for mouse	Methylation/microarray results were consistent between in vivo embryos. Generalized hypermethylation and greater locus–locus variability in in vitro embryos	[61]
Targeted—Selected genes	IVM + IVF + culture in 1.25% Oviductal fluid vs IVM + IVF + culture in BSA	Cow	1C to blastocyst	Bifulfite sequencing	Hypermethylation of MTERF2, ABCA7, and OLFM1, LINE1 elements in oviductal fluid vs culture media with only BSA	[89]
Targeted—Avy gene	Natural mating vs SO + embryo transfer vs SO + embryo culture + embryo transfer	Mouse	Zygote to blastocyst	Bisulfite modification, sequencing of the Avy allele in tail DNA	Increased hypomethylation of agouti Avy allele resulting in increased agouti phenotype after embryo culture	[96]
Targeted—Avy gene	SO + culture in KSOMS vs SO +culture in KSOM and 10% FCS	Mouse	Zygote to blastocyst	Bisulfite modification, sequencing of the Avy allele in tail DNA	Axin 1 Fu epiallele activated through histone modifications, no changes in histone methylation seen	[95]

on these two strategies. Studies included in Table 7.1 were chosen from a literature search using the PubMed database and Google Scholar. Studies between 2000 and 2021 were chosen to make sure relevant information was surveyed. Keywords used included "embryo culture," "DNA methylation," "ART," "IVF," "imprinted genes," and "whole genome methylation."

7.6.1 Global Changes in DNA Methylation

Initial examination of DNA methylation changes on a whole genome scale was performed using 5-methylcytosine immunofluorescent staining. Mice, rabbits, and rats show aberrant DNA methylation kinetics during early embryo development when cultured in vitro, which varies with the type of culture medium used [46, 59, 60]. However, high-resolution quantification cannot be performed with immunofluorescence and locus-specific changes cannot be identified using this methodology. Microarrays and next-generation sequencing techniques allow for genome-wide examination of DNA with high resolution to identify locus/gene-specific changes. However, few studies have used high-throughput methods to analyze methylation differences after culture in animal models [37]. One group used methylation-sensitive PCR and microarray analysis of a single mouse chromosome to show that embryo culture in vitro led to overall hypermethylation and increased variation compared to in vivo conceived blastocysts [61]. Integrating methylated DNA immunoprecipitation sequencing (MeDIP-seq) and RNA-seq, Tan et al. examined changes in methylation and corresponding gene expression in mice through gestation. Examination of extraembryonic and placental tissue revealed 400 genes at each time point that showed hypo- or hypermethylation accompanied by corresponding changes in gene expression. These genes were functionally involved in cytoskeletal organization, vasculogenesis, and energy metabolism [62]. However embryo culture effects cannot be separated from the impact of IVF overall in this study. Compared to other ART manipulations, the Rinaudo group has found that in vitro culture of murine embryos has the greatest impact on global gene expression. They additionally found that using specific culture media can change the impact of embryo culture on gene expression—mouse embryos cultured in KSOMaa had a transcriptional profile that was more similar to that of in vivo cultured embryos when compared to those cultured in Whitten's media [63, 64].

Non-murine animal models also exhibit tissue-specific epigenetic and gene expression changes following ART [37, 65–71]. However, most of these studies do not separate out the effects of in vitro oocyte maturation, in vitro fertilization, and embryo culture on the abnormalities seen. A recent bovine study examined embryos cultured to the zygote, 4-cell, 16-cell, or blastocyst stage and found that length of culture proportionally affected genome-wide DNA methylation, suggesting that in vitro culture alone was able to alter DNA methylation profile of blastocysts [72]. An effect of culture media specifically on global gene expression was also seen in bovine embryos: the addition of serum to culture media led to sex-specific gene expression dysregulation, and embryos cultured serum-free more closely

resembled embryos developed in vivo [73]. Porcine embryos cultured in natural reproductive fluids additionally showed an epigenetic profile similar to embryos that develop in vitro. Changes in genome-wide methylation seen in embryos cultured in routinely used NCSU23 media led to epigenetic and transcriptional dysregulation in developmentally significant genes [74].

Genome-wide methylation effects of in vitro culture have not been examined in humans. However, genome-wide gene expression changes due to different culture media utilized in human IVF laboratories have been documented. Comparisons of G5 to HTF media showed changes in 951 genes, some of which are involved in DNA replication and oxidative phosphorylation pathways [75]., [76].

7.6.2 Gene-Specific Epigenetic Change

Global demethylation followed by gradual remethylation occurs during early embryo development to allow for lineage specification of early cell types. Imprinted genes are, however, protected from these dynamic changes, and epigenetic gene silencing restricts expression to either maternal or paternal alleles [77]. ART has been historically associated with imprinting disorders, leading many groups to focus efforts on the effects of in vitro embryo culture on imprinted genes. An early study in sheep that focused solely on the effect of embryo culture reported hypomethylation and corresponding increases in expression of the imprinted gene *Igf2R* [78]. Perturbations in imprint establishment in cultured mouse blastocysts have been comprehensively described underscoring the validity of this approach (see reviews by the Mann group [79, 80]). Epigenetic perturbations due to culture alone have been described in paternally imprinted (*H19*) and maternally imprinted (*Snrpn* and *Peg3*) loci and persist at least into mid-gestation [45, 81–85]. The presence of fetal calf serum in media and using Whitten's vs KSOM supplemented with amino acids (KSOMaa) additionally resulted in aberrant DNA methylation and gene expression of *H19*, *Snrpn*, *Ascl2*, and *Peg3,* suggesting that culture media alone could impact DNA methylation [82, 83, 86]. To examine these effects further, the Mann group evaluated five different culture media and found hypomethylation at imprinted genes including *H19*, *Snrpn*, and *Peg3* due to all five media, though some systems more closely mimicked in vivo imprinting patterns [87]. The use of both single-step and sequential culture media formulations shows evidence of epigenetic abnormalities [87, 88], though in one study, sequential media led to a lower proportion of blastocysts presenting epigenetic alterations [45]. Interestingly, tissue susceptibility to DNA methylation errors in imprinted genes is differential—compared to murine embryos, placental tissues show a higher propensity for loss of imprinting at the *Ascl2*, *Snrpn*, *Peg3*, and *Xist* loci [85, 86].

Changes due to embryo culture have also been demonstrated in non-imprinted genes. Using methylation-sensitive PCR and a custom microarray panel focused on mouse chromosome 7, one study showed global hypermethylation due to embryo culture. This result differs from findings in imprinted genes, which commonly show hypomethylation due to embryo culture, indicating that genome-wide effects could

differ from regulation at imprinted genes [61]. In line with these data, a recent bovine study also showed hypermethylation of non-imprinted genes and retrotransposons important for development [74, 89].

The above studies summarize the epigenetic impact of in vitro embryo culture. One common theme that emerges from studies on the effect of embryo culture is intergenic and inter-embryo variation in DNA methylation changes, suggesting that embryo response to in vitro culture is stochastic [79]. Variation in response to culture has also been found between mice of different genetic backgrounds [90]. Identifying culture conditions that reduce epigenetic perturbation is thus complicated by the fact that most molecular studies focus on the same 4–5 imprinted genes. Clearly, further studies that are standardized and adopt a whole genome approach are necessary.

Understanding how epigenetic change could occur due to embryo culture is especially vital due to findings that associate embryo culture conditions with the rate of embryo development, IVF success rates, and neonatal outcomes [91–94]. While few studies in mice additionally suggest transgenerational changes in select genes due to embryo culture [95, 96], and in mice and humans, evidence exists suggesting that culture media differences could lead to fetal birthweight changes that persist into childhood [97]. This indicates that epigenetic perturbation due to in vitro culture of embryos could lead to long-lasting developmental change. However, understanding the mechanism underlying these changes, and linking changes during embryo culture with epigenetic perturbations and adverse pregnancy outcomes, has proved challenging. This is primarily due to difficulties in correlating epigenetic differences to corresponding variation in gene expression, making it hard to identify candidate pathways and regulators that could contribute to the development of adverse outcomes. One reason for this might be that attempts to correlate epigenetic and gene expression differences are often conducted on tissue obtained at the end of gestation due to ease of availability. However, epigenetic change could in fact represent a kind of "fossil record" of perturbations that occur early on in pregnancy. Well-controlled animal studies isolating embryo culture use should therefore be performed through gestation to appropriately correlate epigenetic and gene expression changes. Conducting these studies in parallel to the use of physiologically accurate in vitro models using human tissue that enables investigation of early pregnancy would greatly improve our ability to identify factors changed due to embryo culture that could have diagnostic and therapeutic potential.

7.7 Insemination Techniques

Separate from the embryo culture techniques, the method of oocyte insemination may directly impact the epigenome. During IVF, oocytes can be conventionally inseminated or inseminated via intracytoplasmic sperm injection. Conventional insemination involves placing an oocyte with a droplet of processed semen containing roughly 50,000 sperm and observing for fertilization. Intracytoplasmic sperm injection involves isolating an individual sperm and injecting it into the cytoplasm of the oocyte under direct visualization with high powered magnification

[98]. An Illumina 450 K methylation array study on air-dried blood spots (Guthrie cards) from 76 children conceived by ICSI, 18 by intrauterine insemination, and 43 controls revealed numerous differences, in particular at metastable epialleles between ICSI and spontaneously conceived children including HERC3-NAP1L5, PEG10, and L3MBTL [99]. Another group evaluated global methylation via 450 k methylation array using cord blood samples of 48 ICSI and 46 control newborns and detected differences in 4730 CpG sites, albeit with small (beta<10%) effect sizes, and also noted a decreased "epigenetic age" at birth in children conceived with ICSI compared to controls [100]. There remains some debate as to whether the differences observed are related to the ICSI technology, or driven by abnormalities in the sperm (often the underlying indication for ICSI). Methylation errors have been shown to be present at imprinted loci in the sperm of oligospermic men [101, 102]. In a study of 78 paired paternal sperm and trophoblast samples from pregnancy terminations at 6–9 weeks gestation (IVF vs unassisted controls), 41% of cases with abnormal trophoblast DNA methylation, identical alterations were present in the parental sperm [11]. Interestingly, in a mouse model of IVF, mice conceived with ICSI that exhibited epimutations in *H19, Snrpn,* and *Peg3*, in somatic cells, normal epigenetic reprogramming in their germ cells was noted [41]. This observation suggests that while ICSI can lead to the formation of primary epimutations, and such epimutations are likely to be maintained indefinitely in somatic cells of those individuals, they are corrected in the germ line by epigenetic reprogramming and, thus, unlikely to be propagated to subsequent generations.

7.8 Embryo Cryopreservation

Embryo cryopreservation has gained increasing favor as a method of optimizing pregnancy and perinatal outcomes, allowing for pre-implantation genetic testing of embryos when indicated, decreasing maternal risk of ovarian hyperstimulation syndrome, and allowing for fertility preservation for future pregnancies [103]. Accumulating evidence suggests that adverse perinatal outcomes in singleton pregnancies are significantly reduced following frozen/thawed embryo transfer into a physiologic hormonal environment as opposed to immediate embryo transfer in the concurrent fresh/hormonally hyperstimulated IVF cycle [104–108]. In a murine model comparing pregnancies conceived after IVF with and without embryo cryopreservation and controls, embryo vitrification was noted to have differential effects on the methylation and expression of H19/Igf2 [109]. The H19/Igf2 DMD was hypomethylated in the vitrified group compared to both controls and fresh IVF. H19 expression was decreased in the vitrified group compared to the fresh IVF group, while Igf2 expression in the vitrified group was greater than in the fresh IVF group. Ma et al. examined the effect of embryo vitrification on changes in DNA methylation in E9.5 mouse fetuses and placentas and noted global methylation levels of fetuses were increased after vitrification compared with the in vitro cultured group without vitrification ($p < 0.05$), yet similar to unassisted conceptions ($p > 0.05$) [110]. In a study examining the effects of superovulation and vitrification on global DNA

methylation and histone modification among 4 exposures groups (control (C), superovulation (S), superovulation+in vitro culture (SI), and superovulation +vitrification+in vitro culture (SVI)) in a mouse model, vitrification decreased the fluorescence intensity of global DNA methylation in the inner cell mass (ICM), in SVI Group in comparison to C group ($P < 0.05$), and the fluorescence intensity of H4K12acetylation in the trophectoderm was higher in the SVI group than in C and S ($P < 0.05$) [111]. There are limited data on changes in miRNA expression associated with vitrification. In a murine model, while levels of miR-21 and let-7a were significantly decreased in vitrified 8-cell embryos and fresh blastocysts when compared with fresh 8-cell embryos, vitrification did not affect the expression level of these genes in preimplantation embryos [112].

While it has been hypothesized that the differences in clinical outcomes observed may be associated with epigenetic changes, few human studies have been done to date to corroborate these assertions—an opportunity for future research.

7.9 Embryo Biopsy

As IVF has evolved, so have the technologies to augment treatment outcomes, including pre-implantation genetic testing. Pre-implantation genetic testing (PGT) first gained traction in the realm of genetic testing for X-linked disorders, with the first live birth reported in 1990, involving blastomere biopsy followed by fluorescence in situ hybridization (FISH) and PCR amplification for detection of repetitive Y chromosome sequences [113]. As the technology developed from FISH to array comparative genomic hybridization (aCGH), single-nucleotide polymorphism (SNP) array, and most recently next-generation sequencing (NGS), so have the indications expanded to including testing for monogenic disorders (PGT-M), aneuploidy (PGT-A), structural re-arrangements (PGT-SR), and most recently, polygenic risk scores (PGT-P) [114, 115]. Pre-implantation genetic testing involves ovarian stimulation and oocyte retrieval, followed by in vitro fertilization methods and embryo culture to the blastocyst stage with a biopsy of 4–7 mural trophectoderm cells from an expanded blastocyst and, in most clinical scenarios, subsequent embryo vitrification. The impact of trophectoderm biopsy specifically on the epigenome is at this time poorly understood. A study of 72 human blastocysts donated from 33 couples after IVF for male factor infertility used Methyl Maxi-Seq (Zymo Research) for genome-wide DNA methylation followed by pyrosequencing and RT-PCR validation and small cell number-RNA-seq for transcriptome analyses [116]. Methylome and transcriptome analyses of individual blastocysts demonstrated perturbations in 1111 CpG sites and 469 transcripts, respectively ($P < 0.05$), specifically in genes involved in regulation of cellular metabolic processes. No studies regarding histone modification or miRNA expression specifically isolating trophectoderm biopsy as an exposure have been reported.

7.10 Oocyte Cryopreservation

Oocyte cryopreservation introduces yet another variable in IVF processes with an opportunity for changes to the gamete, and potentially by extension, the embryonic epigenome. The first birth after oocyte cryopreservation was described in 1986 [117], utilizing slow freezing methods. Challenges with oocyte post-thaw viability and subsequent embryo progression limited utilization until the transition to vitrification techniques. Now, oocyte cryopreservation is widely utilized for fertility preservation prior to gonadotoxic therapies, as well as for planned deferment of child-bearing [118, 119]. Animal models indicate a potential effect of vitrification on global DNA methylation and histone modification. Global DNA methylation (as reflected by 5-MeC fluorescence intensity) was lower in vitrified MII murine oocytes as well as the resulting 2–8 cell embryos, compared to control murine MII oocytes ($p < 0.001$) [120]. In a bovine model, levels of DNA methylation and H3K9me3 in previously vitrified/thawed oocytes and early cleavage embryos were lower ($P < 0.05$) than those in the control group, but the level of acH3K9 increased ($P < 0.05$) in the vitrification group during the early cleavage phases. No differences were observed for DNA methylation, H3K9me3, and acH3K9 in the inner cell mass of blastocysts, whereas decreased levels of DNA methylation and acH3K9 ($P < 0.05$) existed in TE cells after vitrification. The expression of imprinted genes *Peg10*, *Xist*, and *Kcnq1ot1* was upregulated in blastocysts formed from vitrified/thawed bovine oocytes [121]. A study comparing human MII oocytes without vitrification, MII oocytes that had been previously vitrified, and oocytes vitrified at the germinal vesicle stage and in vitro matured to the MII stage showed no difference in global DNA methylation between the three groups [122]. However, gene-specific studies in animal models have conflicting results. A study of murine oocytes demonstrated hypomethylation in promoters of *Oct4* and *Sox 2* genes in in vitro-matured MII oocytes compared to controls [123]. Yet, human oocytes previously vitrified at the germinal vesicle stage and in vitro matured demonstrate no differences in the DNA methylation at *H19* and *Kcnq1ot1* ICRs [124].

Analyses of histone modifications after oocyte vitrification to date are limited to animal studies. Lysine 12 of histone H4 (AcH4K12) acetylation levels increased significantly in vitrified murine oocytes compared to controls as assessed by immunofluorescence [125], while Yan et al. reported that H3K9 methylation and H4K5 acetylation were both increased in mouse oocytes subjected to vitrification [126].

The role of miRNA changes in oocytes after vitrification is similarly limited to animal studies to date. In a murine model of oocytes collected after superovulation and analyzed by Illumina Hiseq 2000/2500, twenty-two miRNAs were differentially expressed between fresh vs vitrified oocytes including miR-134-5p, miR-210-5p, and miR-21–3p (upregulated), and miR-465c-5p (downregulated). The target genes regulated by these miRNAs were predominantly metabolic pathway regulators and those involved in oxidative stress [127].

While the clinical standard of care for oocyte cryopreservation and utilization currently focuses on oocytes cryopreserved at the MII stage, the role of in vitro maturation of oocytes and their utilization remains to be optimized. As ovarian tissue

cryopreservation or retrieval of immature oocytes may be the only options for fertility preservation in pre-pubescent girls, understanding these potential effects on the epigenome is important. DNA methylation studies in murine models of fresh vs cryopreserved immature ovarian tissue grafts demonstrate that methylation status of *H19* and *Lit 1* ICR in kidney, muscle, and tongue tissues in the resulting offspring did not show significant changes in DNA methylation compared to controls [128]. Global methylation changes in human pregnancies resulting after oocyte cryopreservation, in vitro oocyte maturation, and ovarian tissue cryopreservation have not been reported.

7.11 Phenotypic Changes and Long-Term Outcomes

DNA methylation and gene expression differences between children conceived *in vitro* and children conceived *in vivo* have been noted in multiple studies [10, 11, 18, 20, 36, 129–133]; however, such differences have not been observed in all studies [13, 15–17, 20, 22, 25–28], raising the question as to whether there may be a particularly susceptible group and whether the changes observed have a lasting phenotypic effect.

Epidemiologic studies have demonstrated an increased risk of adverse obstetric and perinatal outcomes including hypertensive disorders of pregnancy, preterm delivery, and low birthweight (LBW) [134–137]. In a study of DNA methylation using human cord blood and placenta samples from 114 patients conceived after IVF or unassisted conceptions at the lowest and highest quintiles of the birthweight distribution, children from the lowest quintile of the birthweight distribution had a significantly greater number of disrupted CpGs than children from the highest quintile of the birthweight distribution particularly at genes involved in fetal and placental growth including *GRB10* [138]. As children with low birthweight can be predisposed to health concerns later in life, including obesity, hypertension, and diabetes, this phenomenon certainly warrants further investigation to identify those individuals who may be at particularly high risk.

While the vast majority of children born after IVF are healthy, what remains to be seen is whether these epigenetic perturbations observed with IVF result in changes in clinical phenotype in adulthood. Conception with in vitro fertilization (IVF) is associated with multiple altered outcomes in children including elevated blood pressure, fasting glucose, triglycerides, and increased total body fat composition, compared to unassisted conceptions, even after controlling for birthweight [139–143]. IVF is associated with increased epigenetic perturbations compared to unassisted conceptions, specifically, at loci involved in growth and development [15, 25, 129, 130, 133, 138]. The long-term persistence and health impact of these observed changes remains to be seen. A cohort study of 149 patients born after ART, using neonatal (Guthrie spot) peripheral whole blood, and 58 unassisted controls were followed to 18–28 years old. DNA methylation profiles from adult peripheral whole blood comparisons noted that epigenetic perturbations observed at birth had largely resolved in adulthood with no difference in health outcomes [144]. However, the

same group noted an increased rate of maternally reported hospital admissions, atopic respiratory conditions, and metabolic/endocrine/nutritional disease (ICD-10 coding category) in the ART-conceived group [145]. Thus, there may be a specific subset of pregnancies that may be at higher risk for persistence in these epigenetic perturbations observed, and they may be associated with long-term health outcomes that manifest beyond the fourth decade of life. Certainly, more long-term studies are needed to quantify and understand the impact of epigenetic perturbations in offspring conceived after ART.

7.12 Future Directions

With the increase in ART utilization, and evolving technologies in the ART space, questions remain as to the impact of added interventions. Specifically, a larger proportion of patients are now pursuing pre-implantation genetic testing for aneuploidy, monosomic disorders, or polygenic risk, introducing the added variable of trophectoderm biopsy. New technologies such as intravaginal IVF culture also introduce unique exposures. Additionally, with improvements in culture and vitrification techniques, gametes and embryos may be stored for years to decades, and the impact of storage duration on epigenetic perturbations has not been well studied. Both animal models and longitudinal human studies will be needed to delineate the effects of these exposures on the epigenome and any potential long-term impact on both immediate offspring and transgenerational effects.

Acknowledgments *Supported by NIH 2P50HD068157-06A1.*

References

1. Sunderam S, Kissin DM, Zhang Y, Folger SG, Boulet SL, Warner L et al (2019) Assisted reproductive technology surveillance—United States, 2016. MMWR Surveill Summ 68(4): 1–23
2. Barnhart KT (2013) Assisted reproductive technologies and perinatal morbidity: interrogating the association. Fertil Steril 99(2):299–302
3. Adamson GDTM, Macaluso M, deMouzon J (2013) The number of babies born globally after treatment with assisted reproductive technologies (ART). Fertil Steril 100(3):S42
4. Adamson GD, de Mouzon J, Chambers GM, Zegers-Hochschild F, Mansour R, Ishihara O et al (2018) International Committee for Monitoring Assisted Reproductive Technology: world report on assisted reproductive technology, 2011. Fertil Steril 110(6):1067–1080
5. Berntsen S, Soderstrom-Anttila V, Wennerholm UB, Laivuori H, Loft A, Oldereid NB et al (2019) The health of children conceived by ART: 'the chicken or the egg?'. Hum Reprod Update 25(2):137–158
6. Cortessis VK, Azadian M, Buxbaum J, Sanogo F, Song AY, Sriprasert I et al (2018) Comprehensive meta-analysis reveals association between multiple imprinting disorders and conception by assisted reproductive technology. J Assist Reprod Genet 35(6):943–952
7. Reik W, Dean W, Walter J (2001) Epigenetic reprogramming in mammalian development. Science 293(5532):1089–1093

8. Smith ZD, Chan MM, Humm KC, Karnik R, Mekhoubad S, Regev A et al (2014) DNA methylation dynamics of the human preimplantation embryo. Nature 511(7511):611–615
9. Gosden R, Trasler J, Lucifero D, Faddy M (2003) Rare congenital disorders, imprinted genes, and assisted reproductive technology. Lancet 361(9373):1975–1977
10. Gomes MV, Huber J, Ferriani RA, Amaral Neto AM, Ramos ES (2009) Abnormal methylation at the KvDMR1 imprinting control region in clinically normal children conceived by assisted reproductive technologies. Mol Hum Reprod 15(8):471–477
11. Kobayashi H, Hiura H, John RM, Sato A, Otsu E, Kobayashi N et al (2009) DNA methylation errors at imprinted loci after assisted conception originate in the parental sperm. European journal of human genetics : EJHG 17(12):1582–1591
12. Hiura H, Hattori H, Kobayashi N, Okae H, Chiba H, Miyauchi N et al (2017) Genome-wide microRNA expression profiling in placentae from frozen-thawed blastocyst transfer. Clin Epigenetics 9:79
13. Tierling S, Souren NY, Gries J, Loporto C, Groth M, Lutsik P et al (2010) Assisted reproductive technologies do not enhance the variability of DNA methylation imprints in human. J Med Genet 47(6):371–376
14. Kanber D, Buiting K, Zeschnigk M, Ludwig M, Horsthemke B (2009) Low frequency of imprinting defects in ICSI children born small for gestational age. European journal of human genetics : EJHG. 17(1):22–29
15. Li L, Wang L, Le F, Liu X, Yu P, Sheng J et al (2011) Evaluation of DNA methylation status at differentially methylated regions in IVF-conceived newborn twins. Fertil Steril 95(6): 1975–1979
16. Zechner U, Pliushch G, Schneider E, El Hajj N, Tresch A, Shufaro Y et al (2010) Quantitative methylation analysis of developmentally important genes in human pregnancy losses after ART and spontaneous conception. Mol Hum Reprod 16(9):704–713
17. Feng C, Tian S, Zhang Y, He J, Zhu XM, Zhang D et al (2011) General imprinting status is stable in assisted reproduction-conceived offspring. Fertil Steril 96(6):1417–23.e9
18. Shi X, Ni Y, Zheng H, Chen S, Zhong M, Wu F et al (2011) Abnormal methylation patterns at the IGF2/H19 imprinting control region in phenotypically normal babies conceived by assisted reproductive technologies. Eur J Obstet Gynecol Reprod Biol 158(1):52–55
19. Zheng HY, Shi XY, Wu FR, Wu YQ, Wang LL, Chen SL (2011) Assisted reproductive technologies do not increase risk of abnormal methylation of PEG1/MEST in human early pregnancy loss. Fertil Steril 96(1):84–9.e2
20. Zheng HY, Shi XY, Wang LL, Wu YQ, Chen SL, Zhang L (2011) Study of DNA methylation patterns of imprinted genes in children born after assisted reproductive technologies reveals no imprinting errors: a pilot study. Exp Ther Med 2(4):751–755
21. Zheng HY, Tang Y, Niu J, Li P, Ye DS, Chen X et al (2013) Aberrant DNA methylation of imprinted loci in human spontaneous abortions after assisted reproduction techniques and natural conception. Hum Reprod 28(1):265–273
22. Oliver VF, Miles HL, Cutfield WS, Hofman PL, Ludgate JL, Morison IM (2012) Defects in imprinting and genome-wide DNA methylation are not common in the in vitro fertilization population. Fertil Steril 97(1):147–53 e7
23. Sakian S, Louie K, Wong EC, Havelock J, Kashyap S, Rowe T et al (2015) Altered gene expression of H19 and IGF2 in placentas from ART pregnancies. Placenta 36(10):1100–1105
24. Vincent RN, Gooding LD, Louie K, Chan Wong E, Ma S (2016) Altered DNA methylation and expression of PLAGL1 in cord blood from assisted reproductive technology pregnancies compared with natural conceptions. Fertil Steril 106(3):739–48.e3
25. Wong EC, Hatakeyama C, Robinson WP, Ma S. DNA methylation at H19/IGF2 ICR1 in the placenta of pregnancies conceived by in vitro fertilization and intracytoplasmic sperm injection. Fertil Steril 2011;95(8):2524–6.e1–3
26. Rancourt RC, Harris HR, Michels KB (2012) Methylation levels at imprinting control regions are not altered with ovulation induction or in vitro fertilization in a birth cohort. Hum Reprod 27(7):2208–2216

27. Puumala SE, Nelson HH, Ross JA, Nguyen RH, Damario MA, Spector LG (2012) Similar DNA methylation levels in specific imprinting control regions in children conceived with and without assisted reproductive technology: a cross-sectional study. BMC Pediatr 12:33
28. Camprubí C, Iglesias-Platas I, Martin-Trujillo A, Salvador-Alarcon C, Rodriguez MA, Barredo DR et al (2013) Stability of genomic imprinting and gestational-age dynamic methylation in complicated pregnancies conceived following assisted reproductive technologies. Biol Reprod 89(3):50
29. Loke YJ, Galati JC, Saffery R, Craig JM (2015) Association of in vitro fertilization with global and IGF2/H19 methylation variation in newborn twins. J Dev Orig Health Dis 6(2):115–124
30. Whitelaw N, Bhattacharya S, Hoad G, Horgan GW, Hamilton M, Haggarty P (2014) Epigenetic status in the offspring of spontaneous and assisted conception. Hum Reprod 29(7): 1452–1458
31. Choux C, Binquet C, Carmignac V, Bruno C, Chapusot C, Barberet J et al (2018) The epigenetic control of transposable elements and imprinted genes in newborns is affected by the mode of conception: ART versus spontaneous conception without underlying infertility. Hum Reprod 33(2):331–340
32. Dimitriadou E, Noutsopoulos D, Markopoulos G, Vlaikou AM, Mantziou S, Traeger-Synodinos J et al (2013) Abnormal DLK1/MEG3 imprinting correlates with decreased HERV-K methylation after assisted reproduction and preimplantation genetic diagnosis. Stress 16(6):689–697
33. Liu Y, Tang Y, Ye D, Ma W, Feng S, Li X et al (2018) Impact of abnormal DNA methylation of imprinted loci on human spontaneous abortion. Reprod Sci 25(1):131–139
34. Chen XJ, Chen F, Lv PP, Zhang D, Ding GL, Hu XL et al (2018) Maternal high estradiol exposure alters CDKN1C and IGF2 expression in human placenta. Placenta 61:72–79
35. Tang L, Liu Z, Zhang R, Su C, Yang W, Yao Y et al (2017) Imprinting alterations in sperm may not significantly influence ART outcomes and imprinting patterns in the cord blood of offspring. PLoS One 12(11):e0187869
36. Turan N, Katari S, Gerson LF, Chalian R, Foster MW, Gaughan JP et al (2010) Inter- and intra-individual variation in allele-specific DNA methylation and gene expression in children conceived using assisted reproductive technology. PLoS Genet 6(7):e1001033
37. Duranthon V, Chavatte-Palmer P (2018) Long term effects of ART: what do animals tell us? Mol Reprod Dev 85(4):348–368
38. Cox GF, Burger J, Lip V, Mau UA, Sperling K, Wu BL et al (2002) Intracytoplasmic sperm injection may increase the risk of imprinting defects. Am J Hum Genet 71(1):162–164
39. Henningsen AA, Gissler M, Rasmussen S, Opdahl S, Wennerholm UB, Spangsmose AL et al (2020) Imprinting disorders in children born after ART: a Nordic study from the CoNARTaS group. Hum Reprod 35(5):1178–1184
40. Market-Velker BA, Zhang L, Magri LS, Bonvissuto AC, Mann MR (2010) Dual effects of superovulation: loss of maternal and paternal imprinted methylation in a dose-dependent manner. Hum Mol Genet 19(1):36–51
41. de Waal E, Yamazaki Y, Ingale P, Bartolomei MS, Yanagimachi R, McCarrey JR (2012) Gonadotropin stimulation contributes to an increased incidence of epimutations in ICSI-derived mice. Hum Mol Genet 21(20):4460–4472
42. Sato A, Otsu E, Negishi H, Utsunomiya T, Arima T (2007) Aberrant DNA methylation of imprinted loci in superovulated oocytes. Hum Reprod 22(1):26–35
43. Khoueiry R, Ibala-Rhomdane S, Mery L, Blachere T, Guerin JF, Lornage J et al (2008) Dynamic CpG methylation of the KCNQ1OT1 gene during maturation of human oocytes. J Med Genet 45(9):583–588
44. Geuns E, Hilven P, Van Steirteghem A, Liebaers I, De Rycke M (2007) Methylation analysis of KvDMR1 in human oocytes. J Med Genet 44(2):144–147
45. Fauque P, Jouannet P, Lesaffre C, Ripoche MA, Dandolo L, Vaiman D et al (2007) Assisted reproductive technology affects developmental kinetics, H19 imprinting control region methylation and H19 gene expression in individual mouse embryos. BMC Dev Biol 7:116

46. Shi W, Haaf T (2002) Aberrant methylation patterns at the two-cell stage as an indicator of early developmental failure. Mol Reprod Dev 63(3):329–334
47. Sullivan-Pyke C, Mani S, Rhon-Calderon EA, Ord T, Coutifaris C, Bartolomei MS et al (2020) Timing of exposure to gonadotropins has differential effects on the conceptus: evidence from a mouse modeldagger. Biol Reprod 103(4):854–865
48. Fortier AL, Lopes FL, Darricarrere N, Martel J, Trasler JM (2008) Superovulation alters the expression of imprinted genes in the midgestation mouse placenta. Hum Mol Genet 17(11): 1653–1665
49. Huffman SR, Pak Y, Rivera RM (2015) Superovulation induces alterations in the epigenome of zygotes, and results in differences in gene expression at the blastocyst stage in mice. Mol Reprod Dev 82(3):207–217
50. Liang XW, Cui XS, Sun SC, Jin YX, Heo YT, Namgoong S et al (2013) Superovulation induces defective methylation in line-1 retrotransposon elements in blastocyst. Reproductive biology and endocrinology : RB&E 11:69
51. Tang SB, Yang LL, Zhang TT, Wang Q, Yin S, Luo SM et al (2019) Multiple superovulations alter histone modifications in mouse early embryos. Reproduction 157(6):511–523
52. Senapati S, Wang F, Ord T, Coutifaris C, Feng R, Mainigi M (2018) Superovulation alters the expression of endometrial genes critical to tissue remodeling and placentation. J Assist Reprod Genet 35(10):1799–1808
53. Gardner DK, Kelley RL (2017) Impact of the IVF laboratory environment on human preimplantation embryo phenotype. J Dev Orig Health Dis 8(4):418–435
54. Lane M, Gardner DK (2007) Embryo culture medium: which is the best? Best Pract Res Clin Obstet Gynaecol 21(1):83–100
55. Summers MC, Biggers JD (2003) Chemically defined media and the culture of mammalian preimplantation embryos: historical perspective and current issues. Hum Reprod Update 9(6): 557–582
56. Kurdyukov S, Bullock M (2016) DNA methylation analysis: choosing the right method. Biology (Basel) 5(1)
57. Maher ER (2005) Imprinting and assisted reproductive technology. Hum Mol Genet 14(1): R133–R138
58. Owen CM, Segars JH Jr (2009) Imprinting disorders and assisted reproductive technology. Semin Reprod Med 27(5):417–428
59. Zaitseva I, Zaitsev S, Alenina N, Bader M, Krivokharchenko A (2007) Dynamics of DNA-demethylation in early mouse and rat embryos developed in vivo and in vitro. Mol Reprod Dev 74(10):1255–1261
60. Reis e Silva AR, Bruno C, Fleurot R, Daniel N, Archilla C, Peynot N et al (2012) Alteration of DNA demethylation dynamics by in vitro culture conditions in rabbit pre-implantation embryos. Epigenetics 7(5):440–446
61. Wright K, Brown L, Brown G, Casson P, Brown S (2011) Microarray assessment of methylation in individual mouse blastocyst stage embryos shows that in vitro culture may have widespread genomic effects. Hum Reprod 26(9):2576–2585
62. Tan K, Zhang Z, Miao K, Yu Y, Sui L, Tian J et al (2016) Dynamic integrated analysis of DNA methylation and gene expression profiles in in vivo and in vitro fertilized mouse post-implantation extraembryonic and placental tissues. Mol Hum Reprod 22(7):485–498
63. Feuer S, Liu X, Donjacour A, Simbulan R, Maltepe E, Rinaudo P (2016) Common and specific transcriptional signatures in mouse embryos and adult tissues induced by in vitro procedures. Reproduction
64. Rinaudo P, Schultz RM (2004) Effects of embryo culture on global pattern of gene expression in preimplantation mouse embryos. Reproduction 128(3):301–311
65. Hori N, Nagai M, Hirayama M, Hirai T, Matsuda K, Hayashi M et al (2010) Aberrant CpG methylation of the imprinting control region KvDMR1 detected in assisted reproductive technology-produced calves and pathogenesis of large offspring syndrome. Anim Reprod Sci 122(3–4):303–312

66. Suzuki J Jr, Therrien J, Filion F, Lefebvre R, Goff AK, Smith LC (2009) In vitro culture and somatic cell nuclear transfer affect imprinting of SNRPN gene in pre- and post-implantation stages of development in cattle. BMC Dev Biol 9:9
67. Hiendleder S, Wirtz M, Mund C, Klempt M, Reichenbach HD, Stojkovic M et al (2006) Tissue-specific effects of in vitro fertilization procedures on genomic cytosine methylation levels in overgrown and normal sized bovine fetuses. Biol Reprod 75(1):17–23
68. Wrenzycki C, Herrmann D, Lucas-Hahn A, Korsawe K, Lemme E, Niemann H (2005) Messenger RNA expression patterns in bovine embryos derived from in vitro procedures and their implications for development. Reprod Fertil Dev 17(1–2):23–35
69. Niemann H, Carnwath JW, Herrmann D, Wieczorek G, Lemme E, Lucas-Hahn A et al (2010) DNA methylation patterns reflect epigenetic reprogramming in bovine embryos. Cell Reprogram 12(1):33–42
70. El Hajj N, Haaf T (2013) Epigenetic disturbances in in vitro cultured gametes and embryos: implications for human assisted reproduction. Fertil Steril 99(3):632–641
71. Smith LC, Suzuki J Jr, Goff AK, Filion F, Therrien J, Murphy BD et al (2012) Developmental and epigenetic anomalies in cloned cattle. Reprod Domest Anim 47(Suppl 4):107–114
72. Salilew-Wondim D, Fournier E, Hoelker M, Saeed-Zidane M, Tholen E, Looft C et al (2015) Genome-wide DNA methylation patterns of bovine blastocysts developed in vivo from embryos completed different stages of development in vitro. PLoS One 10(11):e0140467
73. Heras S, De Coninck DI, Van Poucke M, Goossens K, Bogado Pascottini O, Van Nieuwerburgh F et al (2016) Suboptimal culture conditions induce more deviations in gene expression in male than female bovine blastocysts. BMC Genomics 17:72
74. Canovas S, Ivanova E, Romar R, Garcia-Martinez S, Soriano-Ubeda C, Garcia-Vazquez FA et al (2017) DNA methylation and gene expression changes derived from assisted reproductive technologies can be decreased by reproductive fluids. elife 6
75. Kleijkers SH, Eijssen LM, Coonen E, Derhaag JG, Mantikou E, Jonker MJ et al (2015) Differences in gene expression profiles between human preimplantation embryos cultured in two different IVF culture media. Hum Reprod 30(10):2303–2311
76. Mantikou E, Jonker MJ, Wong KM, van Montfoort AP, de Jong M, Breit TM et al (2016) Factors affecting the gene expression of in vitro cultured human preimplantation embryos. Hum Reprod 31(2):298–311
77. Barlow DP, Bartolomei MS (2014) Genomic imprinting in mammals. Cold Spring Harb Perspect Biol 6(2)
78. Young LE, Fernandes K, McEvoy TG, Butterwith SC, Gutierrez CG, Carolan C et al (2001) Epigenetic change in IGF2R is associated with fetal overgrowth after sheep embryo culture. Nat Genet 27(2):153–154
79. Velker BA, Denomme MM, Mann MR (2012) Embryo culture and epigenetics. Methods Mol Biol 912:399–421
80. Denomme MM, Mann MR (2012) Genomic imprints as a model for the analysis of epigenetic stability during assisted reproductive technologies. Reproduction 144(4):393–409
81. Sasaki H, Ferguson-Smith AC, Shum AS, Barton SC, Surani MA (1995) Temporal and spatial regulation of H19 imprinting in normal and uniparental mouse embryos. Development 121(12):4195–4202
82. Doherty AS, Mann MR, Tremblay KD, Bartolomei MS, Schultz RM (2000) Differential effects of culture on imprinted H19 expression in the preimplantation mouse embryo. Biol Reprod 62(6):1526–1535
83. Khosla S, Dean W, Brown D, Reik W, Feil R (2001) Culture of preimplantation mouse embryos affects fetal development and the expression of imprinted genes. Biol Reprod 64(3):918–926
84. Li T, Vu TH, Ulaner GA, Littman E, Ling JQ, Chen HL et al (2005) IVF results in de novo DNA methylation and histone methylation at an Igf2-H19 imprinting epigenetic switch. Mol Hum Reprod 11(9):631–640

85. Rivera RM, Stein P, Weaver JR, Mager J, Schultz RM, Bartolomei MS (2008) Manipulations of mouse embryos prior to implantation result in aberrant expression of imprinted genes on day 9.5 of development. Hum Mol Genet 17(1):1–14
86. Mann MR, Lee SS, Doherty AS, Verona RI, Nolen LD, Schultz RM et al (2004) Selective loss of imprinting in the placenta following preimplantation development in culture. Development 131(15):3727–3735
87. Market-Velker BA, Fernandes AD, Mann MR (2010) Side-by-side comparison of five commercial media systems in a mouse model: suboptimal in vitro culture interferes with imprint maintenance. Biol Reprod 83(6):938–950
88. Salvaing J, Peynot N, Bedhane MN, Veniel S, Pellier E, Boulesteix C et al (2016) Assessment of 'one-step' versus 'sequential' embryo culture conditions through embryonic genome methylation and hydroxymethylation changes. Hum Reprod 31(11):2471–2483
89. Barrera AD, Garcia EV, Hamdi M, Sanchez-Calabuig MJ, Lopez-Cardona AP, Balvis NF et al (2017) Embryo culture in presence of oviductal fluid induces DNA methylation changes in bovine blastocysts. Reproduction 154(1):1–12
90. de Waal E, Mak W, Calhoun S, Stein P, Ord T, Krapp C et al (2014) In vitro culture increases the frequency of stochastic epigenetic errors at imprinted genes in placental tissues from mouse concepti produced through assisted reproductive technologies. Biol Reprod 90(2):22
91. Quinn P, Kerin JF, Warnes GM (1985) Improved pregnancy rate in human in vitro fertilization with the use of a medium based on the composition of human tubal fluid. Fertil Steril 44(4): 493–498
92. Morbeck DE, Paczkowski M, Fredrickson JR, Krisher RL, Hoff HS, Baumann NA et al (2014) Composition of protein supplements used for human embryo culture. J Assist Reprod Genet 31(12):1703–1711
93. Biggers JD, Racowsky C (2002) The development of fertilized human ova to the blastocyst stage in KSOM(AA) medium: is a two-step protocol necessary? Reprod Biomed Online 5(2): 133–140
94. Mantikou E, Youssef MA, van Wely M, van der Veen F, Al-Inany HG, Repping S et al (2013) Embryo culture media and IVF/ICSI success rates: a systematic review. Hum Reprod Update 19(3):210–220
95. Fernandez-Gonzalez R, Ramirez MA, Pericuesta E, Calle A, Gutierrez-Adan A (2010) Histone modifications at the blastocyst Axin1(Fu) locus mark the heritability of in vitro culture-induced epigenetic alterations in mice. Biol Reprod 83(5):720–727
96. Morgan HD, Jin XL, Li A, Whitelaw E, O'Neill C (2008) The culture of zygotes to the blastocyst stage changes the postnatal expression of an epigentically labile allele, agouti viable yellow, in mice. Biol Reprod 79(4):618–623
97. Zandstra H, Brentjens L, Spauwen B, Touwslager RNH, Bons JAP, Mulder AL et al (2018) Association of culture medium with growth, weight and cardiovascular development of IVF children at the age of 9 years. Hum Reprod 33(9):1645–1656
98. Palermo G, Joris H, Devroey P, Van Steirteghem AC (1992) Pregnancies after intracytoplasmic injection of single spermatozoon into an oocyte. Lancet 340(8810):17–18
99. Estill MS, Bolnick JM, Waterland RA, Bolnick AD, Diamond MP, Krawetz SA (2016) Assisted reproductive technology alters deoxyribonucleic acid methylation profiles in bloodspots of newborn infants. Fertil Steril 106(3):629–39 e10
100. El Hajj N, Haertle L, Dittrich M, Denk S, Lehnen H, Hahn T et al (2017) DNA methylation signatures in cord blood of ICSI children. Hum Reprod 32(8):1761–1769
101. Marques CJ, Carvalho F, Sousa M, Barros A (2004) Genomic imprinting in disruptive spermatogenesis. Lancet 363(9422):1700–1702
102. Kobayashi H, Sato A, Otsu E, Hiura H, Tomatsu C, Utsunomiya T et al (2007) Aberrant DNA methylation of imprinted loci in sperm from oligospermic patients. Hum Mol Genet 16(21): 2542–2551
103. Shapiro BS, Daneshmand ST, Garner FC, Aguirre M, Hudson C, Thomas S (2011) Evidence of impaired endometrial receptivity after ovarian stimulation for in vitro fertilization: a

prospective randomized trial comparing fresh and frozen-thawed embryo transfer in normal responders. Fertil Steril 96(2):344–348
104. Chung K, Coutifaris C, Chalian R, Lin K, Ratcliffe SJ, Castelbaum AJ et al (2006) Factors influencing adverse perinatal outcomes in pregnancies achieved through use of in vitro fertilization. Fertil Steril 86(6):1634–1641
105. Kansal Kalra S, Ratcliffe SJ, Milman L, Gracia CR, Coutifaris C, Barnhart KT (2011) Perinatal morbidity after in vitro fertilization is lower with frozen embryo transfer. Fertil Steril 95(2):548–553
106. Kato O, Kawasaki N, Bodri D, Kuroda T, Kawachiya S, Kato K et al (2012) Neonatal outcome and birth defects in 6623 singletons born following minimal ovarian stimulation and vitrified versus fresh single embryo transfer. Eur J Obstet Gynecol Reprod Biol 161(1):46–50
107. Maheshwari A, Pandey S, Shetty A, Hamilton M, Bhattacharya S (2012) Obstetric and perinatal outcomes in singleton pregnancies resulting from the transfer of frozen thawed versus fresh embryos generated through in vitro fertilization treatment: a systematic review and meta-analysis. Fertil Steril 98(2):368–377
108. Roque M, Lattes K, Serra S, Sola I, Geber S, Carreras R et al (2013) Fresh embryo transfer versus frozen embryo transfer in in vitro fertilization cycles: a systematic review and meta-analysis. Fertil Steril 99(1):156–162
109. Wang Z, Xu L, He F (2010) Embryo vitrification affects the methylation of the H19/Igf2 differentially methylated domain and the expression of H19 and Igf2. Fertil Steril 93(8): 2729–2733
110. Ma Y, Ma Y, Wen L, Lei H, Chen S, Wang X (2019) Changes in DNA methylation and imprinting disorders in E9.5 mouse fetuses and placentas derived from vitrified eight-cell embryos. Mol Reprod Dev 86(4):404–415
111. Bakhtari A, Rahmani HR, Bonakdar E, Jafarpour F, Asgari V, Hosseini SM et al (2014) The interfering effects of superovulation and vitrification upon some important epigenetic biomarkers in mouse blastocyst. Cryobiology 69(3):419–427
112. Azizi E, Ghaffari Novin M, Naji M, Amidi F, Hosseinirad H, Shams MZ (2021) Effect of vitrification on biogenesis pathway and expression of development-related microRNAs in preimplantation mouse embryos. Cell Tissue Bank 22(1):103–114
113. Handyside AH, Kontogianni EH, Hardy K, Winston RM (1990) Pregnancies from biopsied human preimplantation embryos sexed by Y-specific DNA amplification. Nature 344(6268): 768–770
114. Practice Committees of the American Society for Reproductive M, the Society for Assisted Reproductive Technology. Electronic address Aao, Practice Committees of the American Society for Reproductive M, the Society for Assisted Reproductive T (2018) The use of preimplantation genetic testing for aneuploidy (PGT-A): a committee opinion. Fertil Steril 109(3):429–436
115. Treff NR, Marin D, Lello L, Hsu S, Tellier L (2020) PREIMPLANTATION GENETIC TESTING: Preimplantation genetic testing for polygenic disease risk. Reproduction 160(5): A13–AA7
116. Denomme MM, McCallie BR, Parks JC, Booher K, Schoolcraft WB, Katz-Jaffe MG (2018) Inheritance of epigenetic dysregulation from male factor infertility has a direct impact on reproductive potential. Fertil Steril 110(3):419–28 e1
117. Chen C (1986) Pregnancy after human oocyte cryopreservation. Lancet 1(8486):884–886
118. Ethics Committee of the American Society for Reproductive Medicine. Electronic address Aao (2018) Fertility preservation and reproduction in patients facing gonadotoxic therapies: an ethics committee opinion. Fertil Steril 110(3):380–386
119. Ethics Committee of the American Society for Reproductive Medicine. Electronic address aao, Ethics Committee of the American Society for Reproductive M (2018) Planned oocyte cryopreservation for women seeking to preserve future reproductive potential: an ethics committee opinion. Fertil Steril 110(6):1022–1028

120. Liang Y, Fu XW, Li JJ, Yuan DS, Zhu SE (2014) DNA methylation pattern in mouse oocytes and their in vitro fertilized early embryos: effect of oocyte vitrification. Zygote 22(2):138–145
121. Chen H, Zhang L, Deng T, Zou P, Wang Y, Quan F et al (2016) Effects of oocyte vitrification on epigenetic status in early bovine embryos. Theriogenology 86(3):868–878
122. Liu MH, Zhou WH, Chu DP, Fu L, Sha W, Li Y (2017) Ultrastructural changes and methylation of human oocytes vitrified at the germinal vesicle stage and matured in vitro after thawing. Gynecol Obstet Investig 82(3):252–261
123. Milroy C, Liu L, Hammoud S, Hammoud A, Peterson CM, Carrell DT (2011) Differential methylation of pluripotency gene promoters in in vitro matured and vitrified, in vivo-matured mouse oocytes. Fertil Steril 95(6):2094–2099
124. Al-Khtib M, Perret A, Khoueiry R, Ibala-Romdhane S, Blachere T, Greze C et al (2011) Vitrification at the germinal vesicle stage does not affect the methylation profile of H19 and KCNQ1OT1 imprinting centers in human oocytes subsequently matured in vitro. Fertil Steril 95(6):1955–1960
125. Suo L, Meng Q, Pei Y, Fu X, Wang Y, Bunch TD et al (2010) Effect of cryopreservation on acetylation patterns of lysine 12 of histone H4 (acH4K12) in mouse oocytes and zygotes. J Assist Reprod Genet 27(12):735–741
126. Yan LY, Yan J, Qiao J, Zhao PL, Liu P (2010) Effects of oocyte vitrification on histone modifications. Reprod Fertil Dev 22(6):920–925
127. Li J, Yang X, Liu F, Song Y, Liu Y (2019) Evaluation of differentially expressed microRNAs in vitrified oocytes by next generation sequencing. Int J Biochem Cell Biol 112:134–140
128. Sauvat F, Capito C, Sarnacki S, Poirot C, Bachelot A, Meduri G et al (2008) Immature cryopreserved ovary restores puberty and fertility in mice without alteration of epigenetic marks. PLoS One 3(4):e1972
129. Katari S, Turan N, Bibikova M, Erinle O, Chalian R, Foster M et al (2009) DNA methylation and gene expression differences in children conceived in vitro or in vivo. Hum Mol Genet 18(20):3769–3778
130. Song S, Ghosh J, Mainigi M, Turan N, Weinerman R, Truongcao M et al (2015) DNA methylation differences between in vitro- and in vivo-conceived children are associated with ART procedures rather than infertility. Clin Epigenetics 7:41
131. Katagiri Y, Aoki C, Tamaki-Ishihara Y, Fukuda Y, Kitamura M, Matsue Y et al (2010) Effects of assisted reproduction technology on placental imprinted gene expression. Obstet Gynecol Int 2010
132. Santos F, Hyslop L, Stojkovic P, Leary C, Murdoch A, Reik W et al (2010) Evaluation of epigenetic marks in human embryos derived from IVF and ICSI. Hum Reprod 25(9): 2387–2395
133. Nelissen EC, Dumoulin JC, Daunay A, Evers JL, Tost J, van Montfoort AP (2013) Placentas from pregnancies conceived by IVF/ICSI have a reduced DNA methylation level at the H19 and MEST differentially methylated regions. Hum Reprod 28(4):1117–1126
134. Jackson RA, Gibson KA, Wu YW, Croughan MS (2004) Perinatal outcomes in singletons following in vitro fertilization: a meta-analysis. Obstet Gynecol 103(3):551–563
135. Klemetti R, Gissler M, Hemminki E (2002) Comparison of perinatal health of children born from IVF in Finland in the early and late 1990s. Hum Reprod 17(8):2192–2198
136. Schieve LA, Ferre C, Peterson HB, Macaluso M, Reynolds MA, Wright VC (2004) Perinatal outcome among singleton infants conceived through assisted reproductive technology in the United States. Obstet Gynecol 103(6):1144–1153
137. Shevell T, Malone FD, Vidaver J, Porter TF, Luthy DA, Comstock CH et al (2005) Assisted reproductive technology and pregnancy outcome. Obstet Gynecol 106(5 Pt 1):1039–1045
138. Ghosh J, Mainigi M, Coutifaris C, Sapienza C (2016) Outlier DNA methylation levels as an indicator of environmental exposure and risk of undesirable birth outcome. Hum Mol Genet 25(1):123–129

139. Hart R, Norman RJ (2013) The longer-term health outcomes for children born as a result of IVF treatment. Part II--mental health and development outcomes. Hum Reprod Update 19(3): 244–250
140. Ceelen M, van Weissenbruch MM, Prein J, Smit JJ, Vermeiden JP, Spreeuwenberg M et al (2009) Growth during infancy and early childhood in relation to blood pressure and body fat measures at age 8-18 years of IVF children and spontaneously conceived controls born to subfertile parents. Hum Reprod 24(11):2788–2795
141. Ceelen M, van Weissenbruch MM, Vermeiden JP, van Leeuwen FE, Delemarre-van de Waal HA (2008) Cardiometabolic differences in children born after in vitro fertilization: follow-up study. J Clin Endocrinol Metab 93(5):1682–1688
142. Sakka SD, Loutradis D, Kanaka-Gantenbein C, Margeli A, Papastamataki M, Papassotiriou I et al (2010) Absence of insulin resistance and low-grade inflammation despite early metabolic syndrome manifestations in children born after in vitro fertilization. Fertil Steril 94(5): 1693–1699
143. Chen M, Wu L, Zhao J, Wu F, Davies MJ, Wittert GA et al (2014) Altered glucose metabolism in mouse and humans conceived by IVF. Diabetes 63(10):3189–3198
144. Novakovic B, Lewis S, Halliday J, Kennedy J, Burgner DP, Czajko A et al (2019) Assisted reproductive technologies are associated with limited epigenetic variation at birth that largely resolves by adulthood. Nat Commun 10(1):3922
145. Halliday J, Wilson C, Hammarberg K, Doyle LW, Bruinsma F, McLachlan R et al (2014) Comparing indicators of health and development of singleton young adults conceived with and without assisted reproductive technology. Fertil Steril 101(4):1055–1063

Genomic Imprinting: A Paradigm for Epigenetics of Human Diseases

8

R. M. John, L. Lefebvre, and M. A. Surani

Contents

R. M. John
Cardiff School of Biosciences, Cardiff University, Cardiff, UK

L. Lefebvre
Department of Medical Genetics, University of British Columbia, Vancouver, BC, Canada

M. A. Surani (✉)
Wellcome Trust/Cancer Research UK Gurdon Institute, University of Cambridge, Cambridge, UK
e-mail: a.surani@gurdon.cam.ac.uk

K. B. Michels (ed.), *Epigenetic Epidemiology*,
https://doi.org/10.1007/978-3-030-94475-9_8

Abstract

Genomic imprinting is a remarkable phenomenon through which certain genes show monoallelic expression depending on their parent of origin. While imprinting may have evolved for viviparity and potentially as a mechanism to balance resource allocation in mammals, functional haploidy presents a clear risk to human health. Both epigenetic and genetic and aberrations at imprinted loci contribute to genomic imprinting disorders, such as Beckwith–Wiedemann, Silver–Russell, Prader–Willi and Angelman syndromes. Beyond these well-documented disorders, changes in the tissue-specific expression levels of imprinted genes may contribute far more widely to human disease. The expression of imprinted genes can be disrupted at the level of a single gene, at the level of an imprinted domain or through changes in imprinted gene networks. Importantly, imprinted genes can respond to prenatal adversity leading to persistent changes in gene expression. Consequently, in addition to identifying the functions of individual imprinted genes, it is important to understand the mechanisms through which imprints are established, maintained and erased, with erasure critical to ensure comprehensive erasure of epimutations in the germline. We review the critical aspects of genomic imprinting and imprinted human diseases as a paradigm for future studies on epigenetics of human development and disease.

Abbreviations

5hmC	5-Hydroxymethylcytosine
5mC	5-Methylcytosine
Airn	*Antisense Of Igf2R non-protein coding RNA*
Ascl2	*Achaete-scute family bHLH transcription factor 2*
BWS	Beckwith–Wiedemann Syndrome
Cdkn1c	*Cyclin-dependent kinase inhibitor 1c*
CpG	Dinucleotide CG
CTCF	CCCTC-binding factor
Dio3	*Iodothyronine Deiodinase 3*
Dlk1	*Delta Like Non-Canonical Notch Ligand 1*
DNMT1	DNA methyltransferase 1
DNMT3A	DNA methyltransferase 3A
DNMT3B	DNA methyltransferase 3B
DNMT3L	DNA methyltransferase 3L
(E)	Embryonic
ERVK	Endogenous retrovirus-K
gDMR	Germline differentially methylated region
Gnas	*GNAS (guanine nucleotide binding protein, alpha stimulating) complex locus*
Grb10	*Growth factor receptor-bound protein 10*
H19	*H19* gene
H3K27me3	Histone H3 lysine 27 trimethylation

H3K36me2/3	Histone H3 di/trimethylated at lysine 36
H3K4me2/3	Histone H3 di/trimethylated at lysine 4
HELL2	Helicase, lymphoid specific
HELLP	Haemolysis, elevated liver enzymes and low platelet count
IC	Imprinting centre
Igf2	*Insulin-like growth factor 2*
Igf2r	*Insulin-like growth factor 2 receptor*
IGN	Imprinted gene network
IMAGe	Intrauterine growth retardation, metaphyseal dysplasia, adrenal hypoplasia congenita, and genital anomalies
Inpp5f-v2	*Inositol polyphosphate-5-phosphatase F variant 2*
Ins2	*Insulin 2*
Kcnq1ot1	*Kcnq1 opposite strand/antisense transcript 1 (non-protein coding) gene*
KvDMR	DMR in the *Kcnq1ot1* locus
LBW	Low birth weight
LTR	Long terminal repeat
Mcts2	*Malignant T cell amplified sequence*
Meg3 (aka *Gtl2*)	*Maternally expressed 3*
Mest (aka *Peg1*)	*Mesoderm-specific transcript*
MLID	Multilocus imprinting disorders
Nap1l5	*Nucleosome assembly protein 1-like 5*
NLRP5	NLR family pyrin domain containing 5
NSD1	Nuclear receptor binding SET domain protein 1
Peg1	*Paternally expressed gene-1*
Peg3	*Paternally expressed gene-3*
PGCs	Primordial germ cells
Phlda2	*Pleckstrin homology-like domain, family A, member 2 gene*
PIWI	P-element induced Wimpy testis
Plag1 (aka *Zac1*)	*Pleiomorphic adenoma gene-like 1*
Rasgrf1	*Ras protein-specific guanine nucleotide releasing factor 1*
sDMR	Somatic differentially methylated region
SETD2	SET domain containing 2, histone lysine methyltransferase
SETDB1	SET domain bifurcated histone lysine methyltransferase 1
Slc22a18	*Solute carrier family 22, member 18*
Slc22a2	*Solute carrier family 22 member 2*
Slc22a3	*Solute carrier family 22 member 3*
Slc38a4	*Solute carrier family 38 member 4*
Snrpn	*Small nuclear ribonucleoprotein polypeptide N*
SRS	Silver–Russell Syndrome
TAD	Topologically associating domain
Tet	Ten-eleven translocation protein
TNDM	Transient neonatal diabetes mellitus
TRIM28	Tripartite motif containing 28
UHRF1	Ubiquitin like with PHD and ring finger domains 1

UPD	Uniparental disomy
ZFP	Zinc-finger protein
Zrsr1 (aka *U2af1-rs1*)	*Zinc finger (CCCH type), RNA binding motif and serine/arginine rich 1*

8.1 Introduction

One of the remarkable discoveries of the twentieth century was the finding that some genes in mammals do not obey the rules of Mendelian inheritance. Accordingly, "imprinted" genes classically exhibit monoallelic expression in a parent-of-origin dependent manner [1–3]. Some imprinted genes are monoallelically expressed in all expressing cell types, while others show partial, tissue-specific, and/or temporal imprinting. Their most critical and fundamentally important characteristic is that parent-of-origin expression is dependent on passage through the parental germline [1]. Functional differences between the parental genomes were first experimentally demonstrated using pronuclear transfer experiments to artificially generate diploid mouse embryos whose genomes were exclusive of either maternal or paternal origin [4–9]. Development of these monoparental embryos was initially relatively normal. However, embryos consisting of two maternally-derived genomes became progressively growth-restricted and died around embryonic day (E) 10, with particularly limited development of extra-embryonic tissue. Conversely, embryos with two paternal copies were both developmentally delayed and growth restricted, dying around E8.5 with an abundance of extra-embryonic tissue. Similarly, monoparental embryonic stem cells in chimeras with wild-type cells allocate to essentially reciprocal regions in chimeric embryos [10, 11], highlighting the functional differences and requirement for both parental genomes. Studies of uniparental disomic (UPD) embryos delineated parent-of-origin functional differences to specific chromosomal regions [12], with these regions found to contain domains of maternally and paternally expressed genes [13–15].

The discovery of imprinted genes, along with their phylogenetic distribution with an increasing knowledge of their functions, has led to considerable speculation concerning the evolution of genomic imprinting with two prevalent explanations being genetic conflict over maternal resources [16, 17] and maternal-offspring adaptation [18]. These, and many other theories that attempt to ascribe the importance of genomic imprinting, account for the intimate and complex relationship between the mammalian mother and her offspring with increasing investment toward viviparity dependent on an elaborate placenta and considerable postnatal nurturing [19] (Fig. 8.1). An important aspect of imprinting is the control of gene dosage by epigenetic marks inherited from the parental germline. Epigenetic marks under some circumstances can undergo modifications in response to environmental cues such as low protein diet in pregnancy [20], with a potential to affect fetal growth and other phenotypic consequences later in life. The consequent flexibility conferred by the epigenetic mechanism may be advantageous for supporting adaptation to

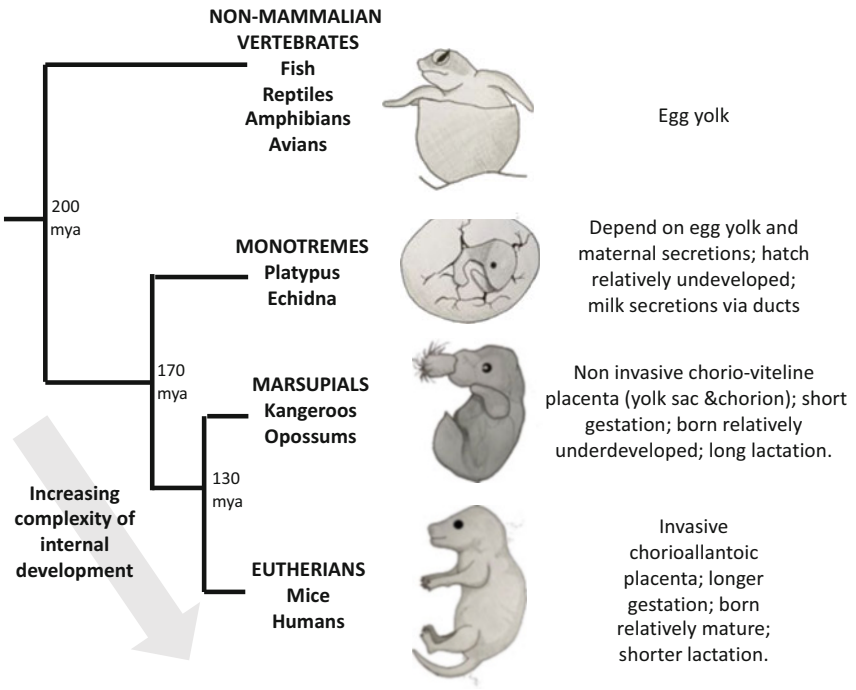

Fig. 8.1 Genomic imprinting and the increasing investment in viviparity in mammals. Within vertebrate lineages, genomic imprinting has only been observed in marsupials and, more prominently, in eutherians with prolonged fetal development in utero before birth. Mya—million years ago

different environments, but may also be responsible for diseases. Understanding the mechanisms that establish, maintain, and erase imprinting is therefore important for our understanding of some human diseases.

8.2 Establishment of Canonical Imprinting

Canonical imprinting resulting in monoallelic gene expression depends on the establishment of heritable DNA methylation marks in either the maternal or paternal germline, and these differentially methylated regions (gDMRs) are propagated in post-zygotic cells [1, 21, 22] (Fig. 8.2). In mice *de novo* DNA methylation at gDMRs is initiated in the male germline before birth [23–25] and detectable as DNA methylation more broadly at intergenic sequences and transposons [26]. For the two paternally imprinted loci, H19-DMR and IG-DMR, changes in histone modifications and high transcriptional read-through precede DNA methylation [27] by the *de novo* DNA methyltransferase DNMT3A and a DNMT-like protein DNMT3L; the latter lacks enzymatic activity [28–31]. *De novo* paternal methylation of the rodent-specific *Rasgrf1* gDMR requires components of the PIWI-interacting

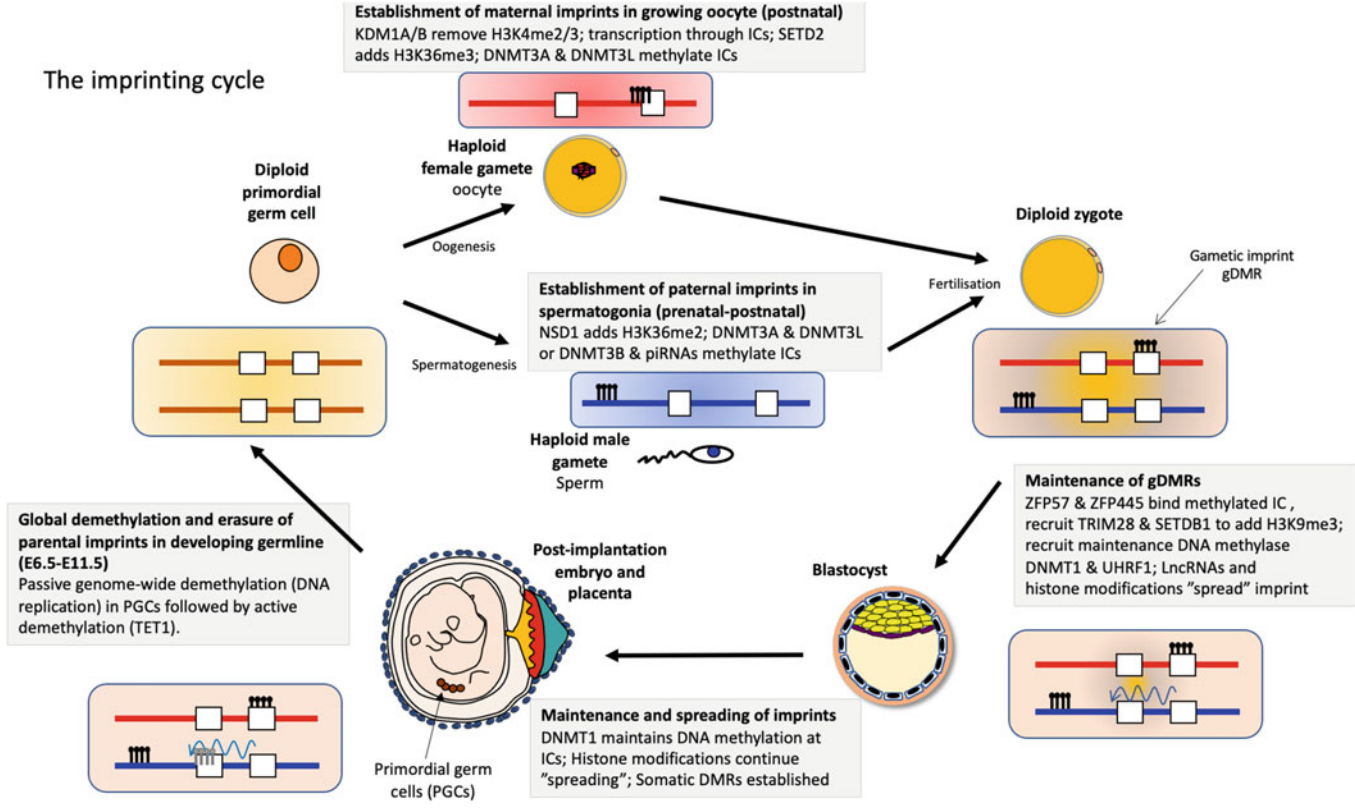

Fig. 8.2 The imprinting cycle of establishment, maintenance, spreading and erasure. DNA methylation imprints are initially established in the male and female germlines. After fertilization, imprints evade the waves of genome-wide reprogramming to become germline differentially methylated regions (gDMRs). These

gDMRs act as imprinting centers from which the imprints "spread" to neighboring genes, establishing domains of maternally and paternally expressed genes. These gDMRs are maintained for the lifetime of the animal along with the monoallelic expression of a subset of imprinted genes. For many genes, monoallelic gene expression does not persist and/or can be tissue-specific

RNA pathway which normally silences transposable elements in the male germline [32] and a second *de novo* DNA methyltransferase, DNMT3B [24]. For the H19-DMR and IG-DMR, these sequences acquire DNA methylation as part of a global mechanism methylating the mature sperm genome to an average level of ~90% [26]. A recent study showed recruitment of the DNMT3A/3L complex extensively to the genome of male germ cells by histone H3 lysine 36 dimethylation (H3K36me2) marks broadly deposited by the histone methyltransferase NSD1 [33].

De novo DNA methylation in the female germline occurs after birth, during the growth phase of oogenesis [34]. Between birth and adulthood, the average genomic DNA methylation level goes from ~2% in non-growing oocytes, to nearly 40% in fully grown, germinal vesicle oocytes [26, 35]. DNA methylation is preceded by the loss of histone H3 lysine 4 di/trimethylation (H3K4me2/3), and gain of trimethylation at lysine 36 (H3K36me3) [36]. These domains with the H3K36me3 marks are established over transcribed gene bodies by the histone methyltransferase SETD2, which associates with the elongating RNA polymerase II [37, 38]. The recruitment of DNMT3A/3L complex follows to the H3K36me3-marked regions to catalyze *de novo* DNA methylation [28–31, 34, 39–41], resulting in a characteristic oocyte DNA methylome with a strong correlation between DNA methylation and the levels of transcription [26, 42]. All maternal gDMRs discovered to date are CpG-rich promoter elements [43], and each of them is covered by a transcript initiating from nearby oocyte-specific promoters [42]. Transcription through gDMRs has been shown experimentally to be functionally required for *de novo* DNA methylation at 6 maternal germline DMRs: Nespas/Gnasxl and 1A, KvDMR1, PWS-IC, Zac1 igDMR, and Peg3DMR [42, 44–48]. The generality of this mechanism for all maternal gDMRs is furthermore supported by the loss of all maternal imprints in oocytes deficient for DNMT3A, DNMT3L, or SETD2 [29, 35, 37].

8.3 Maintenance

After fertilization, both parental genomes undergo extensive passive (maternal) and active (maternal and paternal) DNA demethylation [49]. However, DNA methylation at gDMRs survives this reprogramming, which is not seen for the rest of gametically methylated regions. The protection of DNA methylation at gDMRs is, in part, mediated by the zinc-finger proteins ZFP57 and ZFP445, which bind to a methylated TGCCGC consensus motif within gDMRs [50–54]. These KRAB zinc finger proteins recruit TRIM28 (also known as KAP1) and the histone methyltransferase SETDB1. Consequently, the recruitment of this protein complex at the methylated gDMRs introduces H3K9me3 marks in the region, which promotes preferential recruitment of the DNMT1 DNA methylation maintenance machinery. DNA methylation at gDMRs is subsequently propagated for the lifetime of the individual through the repeated action of the maintenance DNA methyltransferase DNMT1 [55–57] in conjunction with UHRF1 (also known as NP95) [58].

8.4 Spreading

gDMRs operate as imprinting centers (ICs; also known as imprinting control regions) to establish domains of maternally and paternally expressed genes [13–15]. In mice, there are estimated to be 21 gDMRs originating during oogenesis, while three of them are inherited from sperm. These epigenetic marks regulate the allelic expression of around 100 non-coding transcripts and protein-coding genes [15, 59, 60]. There is strong evidence that approximately fifteen of these domains are conserved in humans, while some other are unique to humans, notably with uniparental expression of genes restricted to the placenta [61]. While not all gDMRs have been functionally demonstrated to act as imprinting centers in mice, there are six that have been studied through targeted deletion and have been found to regulate imprinting of the domains within which they reside: *Airn/Igf2r*, *Igf2/H19*, *Snrpn*, *Kcnq1ot1/Cdkn1c*, *Dlk1/Gtl2*, and *Gnas* cluster [62–67]. Both spreading and maintenance of imprinted expression must be carried out in *cis* because the two alleles exist in the same nuclear environment. Some domains contain long non-coding RNAs (lncRNAs) [68] with the expression of the unmethylated parental allele. Some of these lncRNAs have been shown to have a role in the imprinting of a domain containing several imprinted loci, while others lncRNAs appear not to play such a direct epigenetic role [68–77].

8.5 Erasure

Erasure of imprints is prerequisite in each germline cycle in preparation for the establishment of new epigenetic marks according to the sex of the embryo. The erasure of DNA methylation imprints occurs in the precursors of the germline, the primordial germ cells (PGCs). Genome-wide analyses of DNA methylation in mouse PGCs showed that not all genomic regions follow the same demethylation kinetics and that near-complete erasure of the genome occurs in two distinct waves. In phase I of reprogramming, from E8.5 to E10, the genome of migrating PGCs is gradually demethylated by passive dilution of 5mC marks [25, 78]. Specific genomic sequences—called late-demethylating—are largely protected from this first wave of passive demethylation. These include germline-specific genes, X-linked CpG islands (CGIs) in female embryos, specific families of repetitive elements, and imprinted gDMRs [25]. The beginning of phase II of demethylation coincides with the colonization of the genital ridges by PGCs [79]. It is characterized by a rapid conversion of 5mC to 5-hydroxymethylcytosine (5hmC) by the ten-eleven translocation dioxygenase TET1, followed by a passive loss of this epigenetic mark, which is not maintained by DNMT1 [80–83]. From E10.5 to E13.5, the methylation levels of imprints fall from 70% to ~10% on average, while the genome reaches its lowest level of global DNA methylation known, at close to only 3% [25]. A strikingly similar dynamic of demethylation has also been observed in human PGCs, suggesting that epigenetic reprogramming of the PGC genome is a conserved process, but the temporal sequence and possibly the mechanism might differ from the observations in mice [84]. Failure to erase imprints in gonocytes leads to

abnormal expression of imprinted genes and embryonic lethality in the next generation because of the direct inheritance of abnormal grand-parental imprints [81, 85].

So far, TET1 is the only factor known to be implicated in phase II reprogramming in gonocytes in mice. A careful analysis of the progeny from *Tet1*-null mice revealed a broad spectrum of embryonic and postnatal phenotypes, suggesting abnormal expression of imprinted genes [85]. The analysis of DNA methylation patterns at imprinted gDMRs revealed that loss of TET1 in one of the parents is associated with abnormal biallelic DNA methylation at gDMRs in the offspring [85, 86]. These important results established the essential function of TET1 in imprint erasure since they are consistent with a failure to erase grand-parental imprints, which are then abnormally passed on to the progeny: perdurance of grand-paternal DNA methylation marks would be manifested when inherited maternally from *Tet1*-null females, and the converse for grand-maternal marks. The precise timing and mechanism of imprints erasure in the human germline merits further investigation.

8.6 Portrait of an Imprinted Domain: Mouse Distal Chromosome 7

The imprinted region on mouse distal chromosome 7/human chromosome 11p15 has been the focus of considerable study in both humans and mice due to links with the imprinting disorders Beckwith–Wiedemann Syndrome (OMIM 130650; BWS) and Silver-Russell Syndrome (OMIM 180860; SRS), which both show parent-of-origin features. Evidence for the presence of imprinted genes at this chromosomal location came from studies on UPD7 mice which suggested the presence of more than one imprinted gene [87, 88]. Maternal UPD of the distal chromosome 7 region resulted in significant impairment of fetal and placental growth, with fetal death by embryonic day (E) 17.5 [88]. Paternal UPD of the same region resulted in delayed embryonic development with lethality by E10.5, alongside significant developmental abnormalities in the placenta [88–90]. We now know that this region contains two adjacent but mechanistically distinct imprinted domains (Fig. 8.3A). Moreover, these domains best exemplify what is known about the mechanisms that establish domain-wide imprinting.

8.7 The IC1 Imprinted Domain

In mice, the IC1 domain spans over 100 kilobases and contains the paternally expressed protein-coding genes *Igf2* [87, 91] and *Ins2* [92], and the maternally expressed non-coding RNA *H19* [93]. Contained within *H19* is a functionally important imprinted microRNA, *miR-675* [94]. The gDMR within the IC1 domain, called H19-DMR, is located approximately 2 kb upstream of the *H19* promoter. This region acquires DNA methylation exclusively in the male germline, resulting in paternal-allele-specific DNA methylation in offspring [95]. Other regions of differential DNA methylation in the IC1 domain include the promoter of H19 on the silenced paternal allele, three other DMRs within the body of the *Igf2* gene also on

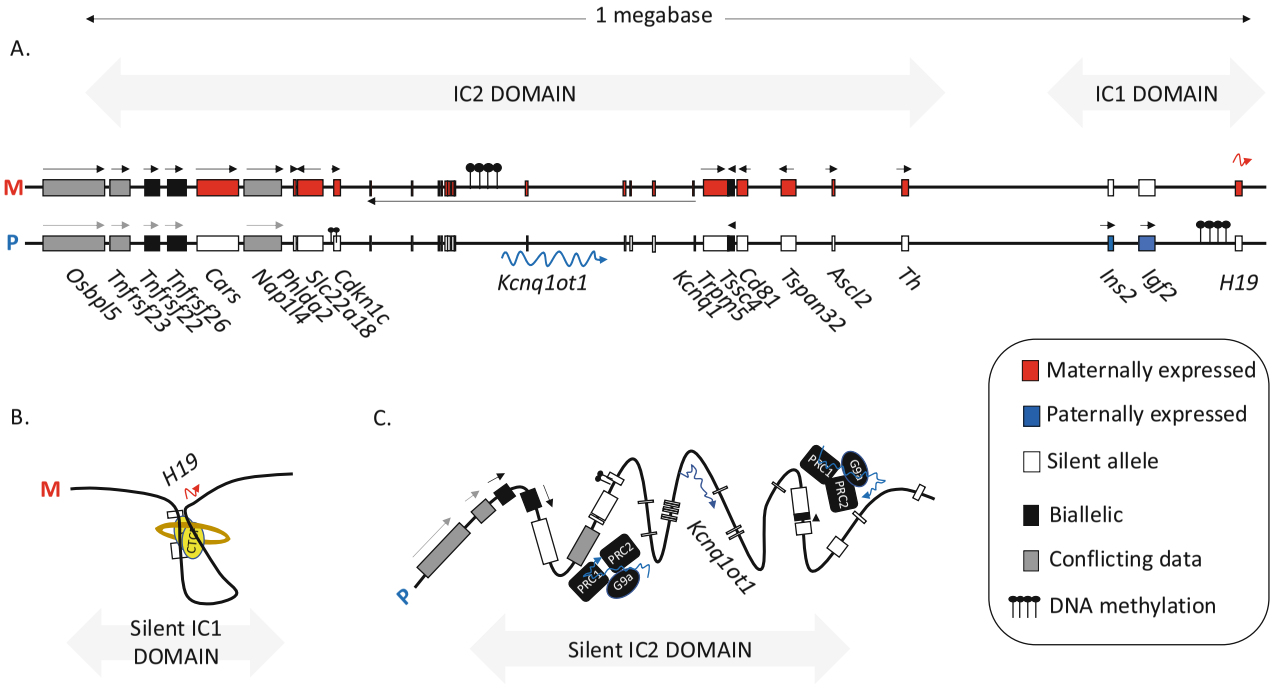

Fig. 8.3 The imprinted domains on mouse distal chromosome 7. (**A**) The imprinted regions on mouse distal chromosome 7 showing the organisation and direction of imprinting of genes within the IC1 and IC2 domains. (**B**) IC1 domain imprinting involves differential conformation of the paternal and maternal alleles dependent on methylation-sensitive binding of CTCF. (**C**) IC2 domain imprinting requires the transcription of the long, non-coding RNA *Kcnq1ot1* from the paternal allele, which is thought to attract the chromatin complexes PRC1 and PRC2 that add silencing histone marks along the domain, and may involve physical tethering to nucleoporins

the paternal allele which, in the case of *Igf2*, is the expressed allele. These sites in sperm are unmethylated but they acquire methylation after implantation and are referred to as somatic DMRs (sDMR) [96]. Maternal inheritance of targeted deletions of H19-DMR results in biallelic *Igf2* and *Ins2* expression alongside loss-of-expression of *H19*, even when the *H19* promoter is left intact [63, 97]. The physical arrangement of paternally and maternally expressed genes at this locus and the observation of similar expression patterns for *Igf2* and *H19* in the developing embryos led to the idea that these genes compete for the same enhancers [97]. H19-DMR contains binding sites for zinc finger protein CCCTC-binding factor (CTCF) that is conserved amongst mammals [98–100]. Binding of CTCF to H19-DMR only occurs when this region is unmethylated [98–100]. Accordingly, the binding of CTCF is thought to organize the domain into a specific chromatin conformation [101–104] which permits expression of only *H19* from the maternal allele [102, 105, 106]. In the absence of CTCF binding, a different conformation is adopted, resulting in the expression of *Igf2* and *Ins2* but not of *H19* (Fig. 8.3B). Similar mechanism involving a structural conformation dependent on methylation-sensitive CTCF binding regulates imprinted expression of the paternally methylated *Dlk1-Dio3* domain [107].

8.8 The IC2 Imprinted Domain

The IC2 domain spans approximately 800 kb and contains a number of protein-coding genes with preferential expression of the maternally inherited allele, including *Phlda2* [108], *Cdkn1c* [109] and *Ascl2* [110], and the paternally-expressed lncRNA, *Kcnq1ot1* (aka *Lit1*) [65] (Fig. 8.3A). The gDMR within the IC2 domain is known as KvDMR1 (Kv-differentially methylated region 1) [65, 111, 112]. KvDMR1 is methylated on the maternal allele and contains the promoter for *Kcnq1ot1* [111]. Paternal inheritance of a targeted deletion of this region in mice results in biallelic expression of the normally maternal-expressed genes upstream and downstream of the gDMR, and the loss of repressive histone modifications that normally coat the paternal allele [65, 75, 113–115]. Unlike in the IC1 domain, transcription of the lncRNA *Kcnq1ot1* rather than the DMR itself, is required to establish and maintain paternal silencing of the IC2 domain [75, 76]. All the protein-coding IC2 domain genes require DNA methylation in cis on the maternal KvDMR1 allele for their maternal expression [28–30, 116–119]. One of the consequences of *Kcnq1ot1* expression is the acquisition of a somatic DNA methylation mark on the silent paternal allele of *Cdkn1c* [120]. In mice deficient for the HELLS helicase (aka LSH), this somatic DMR is not appropriately established resulting in the biallelic expression of *Cdkn1c* [121]. Furthermore, the maintenance of this somatic DMR by DNMT1 is required for the continued silencing of the paternal allele of *Cdkn1c* during development [120, 122, 123]. This additional layer of epigenetic gene silencing suggests the importance of sustaining imprinting of *Cdkn1c* into adulthood.

The loss of domain-wide imprinting at the IC2 domain mirrors the earlier findings for the imprinted locus spanning the maternally-expressed *Igf2r* gene, whereby

truncation of the lncRNA *Airn* resulted in biallelic expression of *Igf2r* in the embryo and placenta, and of the downstream genes *Slc22a2* and *Slc22a3*, normally imprinted and maternally expressed in the placenta [69]. Various models have been proposed to explain how the expression of paternal lncRNAs mediates domain-wide imprinting. The presence of lncRNAs overlapping with and oppositely transcribed to protein-coding genes, as observed for *Kcnq1ot1/Kcnq1* and *Airn/Igf2r*, suggested the possibility of transcriptional interference [124], supported by the finding that *Airn* transcription through the *Igf2r* promoter was necessary to silence *Igf2r* [73]. However, while this could account for imprinted genes with promoters overlapping with lncRNAs, for the majority of loci there are imprinted genes that lie upstream of the lncRNA and ones that are too far downstream to be overlapping with the transcription of the lncRNA itself. An alternative idea was that lncRNAs function by disrupting the promoter-enhancer interactions required for active gene transcription, although there are no enhancers for *Slc22a2* or *Slc22a2* within the *Airn* transcribed region [125], and enhancers for *Cdkn1c* lie outside the region spanned by *Kcnq1ot1* [126]. LncRNAs may instead prevent the expression of nearby genes in cis by forming RNA "clouds" that coat the domain and recruit inactivating histone-modifying complexes [114, 127–129] (Fig. 8.3C). LncRNAs such as *Kcnq1ot1* may additionally function by physically localizing and tethering the paternal allele to the nuclear periphery [127, 130, 131]. There is some evidence that tethering involves complexing with nucleoporins which are the main components of nuclear pore complexes embedded in the nuclear membrane [132]. While posttranscriptional knockdown of *Kcnq1ot1* by RNAi does not impact imprinted gene expression in stem cells [133], conditional deletion of KvDMR1/IC2 in the early mouse embryo results in loss of both paternal gene silencing and acquisition of somatic DNA methylation [134], which suggests differences in the epigenetic mechanisms involved in short and longer-term silencing of the IC2 domain.

8.9 Non-canonical Imprinting

Recent genome-wide next-generation sequencing screens have identified genes exhibiting parent-of-origin behavior that lie outside imprinted domains and are not clearly associated with gDMRs [22, 135, 136]. These are termed non-canonical imprinted genes and their imprinted expression is mostly transient, generally restricted to the early embryo and extra-embryonic lineages. These genes rely on allelic histone modification, particularly histone H3 lysine 27 trimethylation (H3K27me3), following direct inheritance from the oocyte [137–139]. These observations revealed that inheritance of DNA methylation is not the only epigenetic mark for the direct inheritance from the mature gametes to regulate the expression of imprinted genes during development. The functional relevance of non-canonical imprinting in humans merits further investigation [140, 141].

8.10 Evolution of Genomic Imprinting

The evolution of genomic imprinting remains a fascinating subject for discussion since imprinting leads to monoallelic expression of specific genes with critical roles in development. Why would evolution favor a system that is vulnerable to recessive mutations [16, 17], and why is this specific phenomenon observed in mammals and flowering plants [142]. Clues may come from the observation that genomic imprinting leading to parent-of-origin-specific expression in not universal to all mammals and has not been reported in egg-laying mammalian species, such as monotremes, platypus and echidnas [143]. The emergence of imprinting in mammals predates the bifurcation between metatherians (marsupials) and eutherians (placentals), with some imprinted genes being shared amongst them. This suggests a link to the emergence of viviparity and fetal development *in utero*, leading to new energetic demands on the pregnant female [144] (Fig. 8.1). A few "classical" imprinted genes appear to be well conserved in mammals, but several species-specific differences have also been observed.

With respect to mechanisms, two different mechanisms that might have contributed to the emergence of species-specific imprinted genes, both related to transcription-coupled *de novo* DNA methylation in oocytes, have been described. The first mechanism implicates the generation of a new gene via retrotransposition within the intron of a host gene. Since this host gene is expressed in growing oocytes, the inserted retrogene acquires a maternal gDMR and becomes imprinted and paternally expressed in the progeny. Four different examples have been documented, one of which is rodent-specific (*Zrsr1/U2af1-rs1*), while the other three are more evolutionarily ancient, and also imprinted in human (*Mcts2*, *Nap1l5*, *Inpp5f_v2*) [145–148]. Although the retrotransposition event might have triggered the formation of a new imprinted gene, it is also possible that the necessary signals were already present at the ancestral host gene, although these are not imprinted themselves.

A second mechanism is based on the insertion of a new oocyte-specific promoter next to a host gene promoter. Here, the oocyte promoter is provided by a long terminal repeat (LTR) retrotransposon. Specific families of LTR elements evade silencing mechanisms operating in oocytes and act as promoters for diverse transcripts in growing oocytes. Since LTR elements are highly polymorphic and different families of elements have colonized different lineages during evolution, they have been shown to be responsible for species-specific differences in the DNA methylome of oocytes, which is guided by transcription [149]. Notably, specific LTR elements drive oocyte transcription through a number of species-specific gDMR. Indeed, a comparison of mouse and human maternal gDMRs has identified 4 mouse-specific and 17 human-specific examples of paternally-expressed imprinted genes acquiring DNA methylation in oocytes as a consequence of a nearby LTR promoter [61]. These results suggest that novel imprinted genes can be generated during evolution at least in part via the insertion of an active LTR promoter through retrotransposition. The targeted deletions of the upstream LTR at the imprinted genes *Slc38a4* and *Impact* lead to loss of imprinting and biallelic expression, providing functional support for this model [61].

Once established, further elaboration of canonical imprinted genes and arrangements in domains can be envisaged. Regulated by gDMRs, three essential requirements must be met for functional imprinting to evolve: (i) acquisition of a DNA methylation imprint in one germ line; (ii) maintenance of the gDMR through the preimplantation phase of epigenetic erasure; and (iii) prevention of acquisition of DNA methylation on the unmethylated allele during postimplantation waves of de novo methylation [150].

Endogenous retroviruses have also been implicated in the regulation of allelic expression at non-canonical imprinted genes, which are all paternally expressed and show imprinted expression only in preimplantation embryos and extra-embryonic lineages. More importantly, imprinting in this case is independent of gametic DNA methylation and relies instead on direct inheritance of repressive H3K27me3 histone marks from the oocyte [137], but these imprints are transient and do not persist after implantation. The expression of paternally inherited non-canonical imprinted genes in the extra-embryonic ectoderm showed that the maternal allele acquires DNA methylation as a somatic DMR, via DNMT3A/3B activity [138, 139]. So the germline imprint at those genes with a histone mark (H3K27me3) is eventually replaced by a DNA methylation mark for maintenance of imprinted expression in the extra-embryonic tissues. Furthermore, this epigenetic mechanism preferentially targets endogenous retroviruses, specifically ERVK LTRs [138]. These repetitive elements, which can act as promoters or enhancers in extra-embryonic tissues, appear to guide this unusual form of tissue-specific imprinting by protecting the paternal allele from DNA methylation-mediated silencing, which is the default pathway in embryonic lineages. Since endogenous retroviruses are implicated in non-canonical imprinting, we can expect this mechanism to regulate mostly species-specific imprinted genes.

8.11 The Function of Imprinted Genes in Mouse Development

Several imprinted genes function in a dosage-sensitive manner to regulate fetal growth and placental development early in life, and to influence both metabolic and behavioral processes later in life including those relating to mothering, as has been extensively and elegantly reviewed [22, 151–156]. As with imprinting mechanisms, the function of imprinted gene located within the mouse distal chromosome 7 epitomize the function of imprinted genes more generally.

8.12 IC1 Domain Genes

Igf2 was one of the first imprinted genes to be identified in mice [87, 91], along with *H19* [93] and *Igf2r* [157]. *Igf2* encodes a fetal growth factor structurally related to insulin that normally signals to promote cell proliferation, growth, differentiation, and survival via IGF1R. IGF2 also binds to IGF2R, which sequesters and degrades excess IGF2 [158]. The finding that IGF2R is a maternally-expressed imprinted gene

[157] was instrumental in supporting the idea that genomic imprinting evolved in response to a parental "tug-of-war" imposed by the development of mammalian offspring *in utero* [16, 17]. While deletion studies were important for understanding the normal function of IGF2, manipulating the dosage of *Igf2* can provide greater insight into the function of imprinting. In mice, loss of imprinting (LOI) of *Igf2* results from disruption of H19-DMR. Expression of both *Igf2* and *Ins2* occurs at approximately twice the normal level in this model with concomitant loss of *H19*, resulting in increased birth weight of between 8 and 30% together with placental overgrowth [97, 159]. Fetal growth is supported by the placenta which functions both in the transport of nutrients and in securing the availability of nutrients through the action of placental hormones on the mother [160]. IGF2 positively regulates the development of the region of the placenta involved in nutrient transport [161, 162]. Together, these experiments identify *Igf2* as a paternally-expressed imprinted gene that normally functions to promote fetal growth and instruct changes in the placenta required to enhance nutrient transport.

H19 was the first noncoding transcript to be identified as an imprinted gene, and one of the most abundant polyadenylated RNAs in the developing mouse embryo [93, 163]. The precise function of *H19* has been challenging to study because of the mechanistic and reciprocal link with *Igf2* imprinting. Accordingly, a deletion to facilitate the loss of *H19* resulted in upregulation of *Igf2* alongside an overgrowth phenotype despite not disrupting the gDMR [159]. Mice lacking the *H19* transcript appear to develop normally [164] but there is evidence that *H19* functions as a tumor suppressor with the loss of *H19* associated with faster progression in mouse tumorigenesis models [165]. *H19* function may, in part, be mediated by the microRNA *mir-675* which is processed from the first exon of the *H19* transcript in both mice and humans [94, 166, 167]. In mice, loss-of-function of *Ins2* alone does not lead to developmental or metabolic dysfunction [168, 169] but when combined with ablation of the related *Ins1* gene, pups are born small and die shortly thereafter from neonatal diabetes [168].

8.13 IC2 Domain Genes

Loss of imprinting of the IC2 domain, in which all paternally-silenced IC2 domain genes become biallelically expressed, results in both fetal and placental growth restriction [65]. Within this domain, the function of *Cdkn1c*, *Phlda2,* and *Ascl2* has been studied most intently. *Cdkn1c* encodes a cyclin-dependent kinase inhibitor, which normally suppresses proliferation and promotes differentiation in a wide range of tissues. Ablation of *Cdkn1c* in mice results in fetal overgrowth, placentomegaly with expansion of placental endocrine lineages, abdominal wall defects, cleft palate, renal dysplasia, adrenal cytomegaly, maternal preeclampsia, and premature birth [170–175]. Transgenic mice expressing *Cdkn1c* at twice the normal level are growth restricted identifying *Cdkn1c* as the major regulator of embryonic growth within the IC2 domain [176]. These experiments in mice identified *Cdkn1c* as a maternally-expressed imprinted gene that normally functions to restrain fetal growth and placental development. Imprinting of *Cdkn1c* (reduced

paternal expression) would therefore be predicted to enhance fetal and placental growth, and potentially support longer gestation.

Phlda2 encodes a PH domain-only protein normally maternally expressed in the mouse and human placenta where expression is highest [108, 177]. *Phlda2* functions to limit expansion of a major placental endocrine lineage in mice [178–181]. Precisely regulated expression of *Phlda2* is necessary for normal fetal growth with both loss- and gain-in-expression resulting in growth restriction [179, 181, 182]. Just two-fold expression of *Phlda2* driven by a transgene resulted in >10% reduction in birthweight with relative sparing of the head, neonatal hypoglycemia followed by catch-up growth [182]. Rather than playing a direct role in regulating fetal growth, fetal growth restriction occurs as a consequence of the placental endocrine insufficiency induced by the reduction in placental endocrine cells. In addition to regulating placental development and fetal growth, *Phlda2* plays a critically important role in instructing maternal caregiving behavior through controlling the production of placental hormones that act on the mother. Wild-type dams carrying and caring for offspring with different doses of *Phlda2*, behave atypically toward their offspring [183]. Dams carrying and caring for *Phlda2* knockout offspring with an expanded placental endocrine compartment engage in more pup-focused behaviors whereas dams carrying and caring for *Phlda2* transgenic offspring with a reduced placental endocrine compartment neglect their pups and prioritize nest building [183]. A direct role for imprinted genes in regulating maternal behavior has previously been highlighted by studies on the paternally expressed genes *Peg1* and *Peg3* but in both these examples, the dam was genetically altered [184, 185]. In the case of the *Phlda2* studies it is the genetically altered offspring that influence their mother's behavior before birth, consistent with the function of placental hormones in programming the maternal brain in pregnancy. High-quality maternal care is important for the offspring's later life behavior, and both transgenic *Phlda2* offspring and their wild-type siblings sharing the same abnormal environment were found to exhibit anxiety-like symptoms, mild depression, atypical social behavior, and reduced cognitive abilities as adults [186]. Together, these studies illustrate the far-reaching roles of imprinted genes both within and across generations.

Ascl2 encodes a bHLH transcription factor imprinted in the mouse placenta but not the fetus [110]. *Ascl2* plays a critically important role in the placenta supporting the development of the endocrine lineages with full ablation resulting in embryonic lethality at E9.5 [110, 187, 188]. More modest reductions in the expression level of *Ascl2* are also associated with placental defects alongside fetal growth restriction [189, 190]. Conversely, overexpression of *Ascl2* results in reduced development of several placental endocrine lineages and fetal growth restriction [191]. These experiments indicate that a single dose of *Ascl2* in the placenta is required for the proper allocation of progenitors to the different placental endocrine lineages required to support optimal fetal growth.

Cdkn1c, *Ascl2,* and *Phlda2*, through limiting the developing of the placental endocrine lineages, all function to limit the production of placental hormones. Consequently, alterations in their dosage can impact both fetal growth and maternal adaptations to pregnancy which in human populations may manifest as classic complications of pregnancy [192].

8.14 Imprinted Genes Influencing Adult Physiology and Behavior

In addition to these fundamentally important roles during gestation, imprinted genes also function to influence behavioral [22] and metabolic processes [193] critically important for the survival of mammals. Briefly, both *Igf2* and *Cdkn1c* are expressed in the developing and adult nervous system and both genes are important for adult neurogenesis, with IGF2 functioning both as a paracrine and autocrine factor [194] while *Cdkn1c* is required for neural stem cell quiescence and differentiation [195]. *Cdkn1c* is also required for the maturation of midbrain dopamine neuronal cells [196]. Consistent with important roles during brain development and into adulthood, manipulating the expression of these genes results in alterations in behavior. Increased dosage of *Cdkn1c* has been linked to altered reward behaviors and social functions in mice [197–199] and *Igf2* has been shown to be important for memory consolidation and cognitive function [200]. There is evidence for expression of the normally silenced allele of these two genes in the adult brain with the maternal expression of *Igf2* in the choroid plexus and leptomeninges [201], functionally important for neurogenesis [194], and paternal expression of *Cdkn1c* in the brain important for neocortical development [202]. In addition to the nervous system, both *Igf2* and *Cdkn1c* are expressed during development in tissues with important metabolic functions. For example, *Igf2* and *Cdkn1c* are both relatively highly expressed in the developing pituitary [203] where *Cdkn1c* regulates progenitor proliferation but not differentiation [204]. *Igf2* null mice possess more brown adipose tissue just prior to their birth [205] while *Cdkn1c* is both required for the proper differentiation of intrascapular brown adipose and stimulates "browning" within white adipose depots in mice [206]. *Igf2* functions in different lineages to regulate both the size and function of the pancreas [207] while *Cdkn1c* regulates cell cycle exit of progenitors during the early stages of pancreas formation [208]. These multi-fold, multi-tissue roles mean that precisely regulated expression of *Igf2* and *Cdkn1c* is key not only for early development but also for later life metabolism and behavior.

8.15 Variations in Genomic Imprinting and Human Disease

8.15.1 Genomic Imprinting Disorders

A number of classic congenital disorders are associated with alterations in imprinted domains, involving both genetic and epigenetic abnormalities in patients (Table 8.1) [209, 210]. These disorders classically exhibit parent-of-origin transmission, where transmission is possible, with clinical features commonly affecting growth, metabolism, and behavior consistent with the identified functions of many imprinted genes.

Table 8.1 Overview of eight different imprinting disorders in humans

Disorder	Frequency	MIM#	Chromosomal region(s)	Mutations
Beckwith–Wiedemann syndrome (BWS)	1 in 15,000	130650	11p15.5	KvDMR1 hypomethylation (40–50%) upd(11p15)pat (20%) H19-DMR hypermethylation (5–10%) *CDKN1C* point mutations (5%) Translocations (rare)
Silver–Russell syndrome (SRS)	1 in 75,000 to 1 in 100,000	180860	11p15.5	H19-DMR hypomethylation (30–60%) upd(11p15)mat (rare) *IGF2* point mutations (rare) *CDKN1C* point mutations (rare) Maternal *CDKN1C* duplication (rare)
			7p13q32	upd(7)mat (5–10%)
Prader–Willi syndrome (PWS)	1 in 25,000 to 1 in 10,000	176270	15q11.2	Paternal deletion (70%) upd(15)mat (<30%) Aberrant IC methylation (rare) chromosomal rearrangements/ translocations (rare)
Angelman syndrome (AS)	1 in 20, 000 to 1 in 12,000	105830	15q11.2	Maternal deletion (70%) *UBE3A* point mutations (10–15%) Aberrant IC methylation (4%) upd(15)pat (rare)
Transient neonatal diabetes mellitus (TNDM)	1 in 300,000	601410	6q24	Upd(6)pat (41%) Paternal duplications (29%) Methylation defects (30%)
Temple syndrome (TS14)	Unknown	616222	14q32	Upd(14)mat (29%) IGDMR Paternal deletion (10%) MEG3DMR

(continued)

Table 8.1 (continued)

Disorder	Frequency	MIM#	Chromosomal region(s)	Mutations
				Aberrant methylation (61%)
Kagami–Ogata syndrome (KOS14)	Unknown	608149	14q32	Upd(14)pat (65%) IGDMR maternal deletion (15%) MEG3DMR Aberrant methylation (20%)
Pseudohypoparathyroidism 1B (PHP1b)	Unknown	603233	20q13	*GNAS DMRs* hypomethylation (42.5%) Maternal deletion (8.5%) Point mutations 46.5% Upd(20)pat (rare)

Type and frequencies of alterations associated with disorders summarized from [209, 210]

8.16 Beckwith–Wiedemann Syndrome (BWS; OMIM #130650)

One of the most common imprinting disorders is BWS, which is associated with a variety of genetic and/or epigenetic alterations localized to human chromosome 11p15.5 [210] (Fig. 8.4). BWS is estimated to occur in one in every 15,000 births and primarily involves overgrowth features [211]. BWS is usually diagnosed based on the presence characteristics including macrosomia (birth weight >97th percentile), macroglossia (unusually large tongue), neonatal hypoglycemia, ear creases or pits and abdominal wall defects, hemihypertrophy (one side of the body or a part of one side of the body is larger than the other), visceromegaly, nervus flammeus (port-wine stain), cleft palate, cardiac abnormalities, advanced bone age, enlarged placenta and abnormalities in placental vasculature. There is a high incidence of premature birth for BWS infants, sometimes in combination with polyhydramnios (excessive amniotic fluid) and gestational hypertension [212], with some BWS mother's suffering the potentially life-threatening disorder of preeclampsia with HELLP (hemolysis, elevated liver enzymes, and low platelet count) [213]. Infants are at increased risk of developing congenital/childhood tumors such as Wilms' tumor, adrenocortical carcinoma, hepatoblastoma, and neuroblastoma, although this risk decreases with age. Five to ten percent of BWS patients exhibit gain of methylation at *H19-DMR* (IC1), predicted from mouse studies to result in overexpression of *IGF2*, while 40–50% of patients have loss of DNA methylation at *KvDMR1* (aka IC2) predicted to result in loss of expression of all the IC2 domain protein-coding genes. Although several BWS patients may sporadically lose DNA methylation at IC2 as an epimutation, recent evidence points to cis-acting genetic causes in some

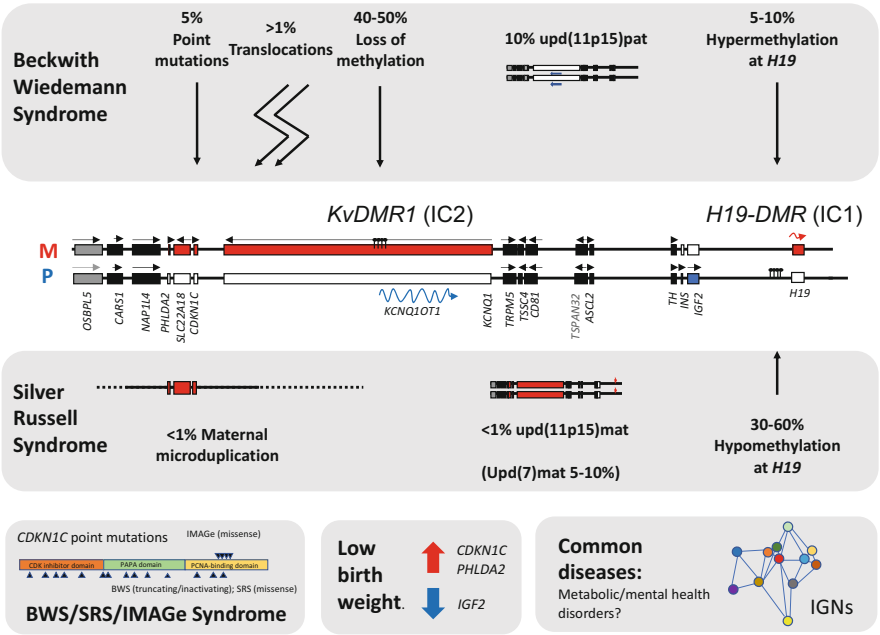

Fig. 8.4 Contribution of imprinted genes to both rare and highly common human diseases exemplified by human chromosome 11p15. Genetic and epigenetic disruptions to imprinted domains are associated with rare imprinting disorders while variations in the expression levels of individual genes or IGNs may underlie highly common conditions such as low birth weight, type 2 diabetes, obesity, neurodevelopmental and behavioral disorders

cases. Since IC2 is located in an intron of the *KCNQ1* gene, implicated in long QT syndrome-1 [111], *de novo* establishment of the maternal DNA methylation imprint at IC2 is predicted to be guided by transcription through the region during oocyte growth, possibly from a *KCNQ1* transcript initiating at its canonical start site. Such a model is in fact supported by studies of mouse mutants carrying the insertion of a transcriptional termination sequence adjacent to IC2, which prevents extension of a *Kcnq1* transcript across IC2 [46]. That such a mechanism is conserved in humans is supported by rare BWS patients, also affected by long QT syndrome, in which transcription from the *KCNQ1* promoter through IC2 is perturbed by maternally inherited translocations, promoter deletions, splice variants, or duplications within the *KCNQ1* locus [214–217]. Patients with familial BWS carry inactivating germ line mutations in the coding sequence of the maternally inherited *CDKN1C* allele [218, 219], further highlighting the pathological contribution of *CDKN1C* in BWS. Functional studies in mice have been highly informative for our understanding of the gene changes underlying BWS [220]. While overgrowth in BWS could result from either too much *IGF2* or too little *CDKN1C*, mice that overexpress *IGF2* do not exhibit defining features of BWS such as cleft palate and abdominal wall defects which are observed in response to loss of murine *Cdkn1c* [170, 171]. It seems likely that either or both alterations have potential to contribute to BWS.

8.17 Silver-Russell Syndrome (SRS; OMIM #180860)

SRS is a very rare imprinting disorder with approximately 70% of patients having alterations affecting human chromosome 7 or 11, with 30% of unknown origin [210] (Fig. 8.4). SRS is diagnosed with an approximate frequency of 1 in 300,000 but may be far more common. SRS is defined by fetal growth restriction and failure to thrive postnatally and some or all of the following: normal head circumference, triangular-shaped face with a large protruding forehead, clinodactyly, undergrowth of one side of the body (hemihypotrophy), fasting hypoglycaemia, night sweats and excessive thinness into adulthood. An international consensus statement summarizing recommendations for clinical diagnosis, investigation, and management was published in 2017 [221]. As with BWS, the cognitive and behavioral characteristics of SRS are less well established but there are reports of specific learning difficulties [222–225], hyperactivity [226], attention deficits [224], autistic regression [227], and eating difficulties [228–231]. Maternal uniparental disomy of chromosome 7 is present in 5–10% of SRS patients (mUPD7; two maternal copies). Within the duplicated region there are two imprinted domains one of which contains the maternally-expressed *GRB10* gene which has been implicated in both growth and behavior in mice [232, 233] and one of which spans *MEST* (aka *PEG1*) which has similar growth regulatory properties in mutant mice [184]. The larger proportion of SRS cases (30–60%) have hypomethylation of *H19-DMR* predicted, from mouse studies, to be associated with loss of expression of *IGF2*. A few SRS patients carry maternal microduplications of the 11p15 IC2 domain with the minimally duplicated region spanning *CDKN1C*, *KCNQ1*, *PHLDA2,* and *SLC22A18* [234–236]. These patients are predicted to have twice the normal level of *CDKN1C* expression. Rarely, patients diagnosed with SRS possess mutations within *CDKN1C* thought to increase the stability of the protein [237–239]. Maternally inherited dominant missense mutations primarily within the PCNA-binding domain of *CDKN1C* have been reported in IMAGe syndrome (Intrauterine growth retardation, metaphyseal dysplasia, adrenal hypoplasia congenita, and genital anomalies; # 614732) [240, 241] (Fig. 8.4). IMAGe syndrome is a very rare condition combining intrauterine growth restriction, metaphyseal dysplasia, adrenal hypoplasia congenita, and genital anomalies and is clinically distinguished from SRS by the presence of adrenal insufficiency. As with BWS, mouse studies provide support for too little *IGF2* or too much *CDKN1C* in SRS cases linked to chromosome 11p15. However, mice with loss-of-expression of *Igf2* do not exhibit the more defining features of SRS such as lack of body fat and altered behaviors, which are seen in mouse models with gain-in expression of *Cdkn1c* [197–199, 206, 242].

8.18 Multilocus Imprinting Disorders (MLID)

A number of patients with BWS have epigenetic alterations affecting additional imprinted regions with the first suggestion coming from the finding that some patients diagnosed with transient neonatal diabetes mellitus (TNDM) also have hypomethylation at *KvDMR1* [243]. It is now clear that a significant proportion of

patients with a diagnosed imprinting disorder have an MLID, whereby more than one imprinted domain is impacted [244, 245]. In BWS 20–50% of patients with loss of *KvDMR1* (IC2) methylation have an MLID, and MLID have been reported in approximately 15% of SRS cases [246]. One widely held explanation for simultaneous epigenetic defects at multiple ICs is a failure to maintain DNA methylation at these sites during the genome-wide reprogramming that occurs shortly after fertilization. Evidence for such a mechanism comes from studies on TNDM patients with DNA methylation defects at other imprinted loci, many of whom carry recessive mutations in ZFP57 [247], one of the proteins that maintain allelic DNA methylation and H3K9me3 at imprinted DMRs in mouse preimplantation embryos [52, 53]. Although mutations in ZFP57 have not, as yet, been reported in BWS or SRS patients with MLID [248, 249], a case of BWS with MLID has been described in the progeny of a mother homozygous for mutation in *NLRP2* [250]. Interestingly, mutations in the related *NLRP5* gene, coding for a component of the oocyte subcortical complex, are associated with MLID [251]. Given the multiple epigenetic modifiers involved in establishing and maintaining genomic imprinting, it is likely that other epigenetic regulators (such as ZNF445, DNMT1, UHRF1, SETDB1, or TRIM28) could be found to be involved in the DNA maintenance defects characteristic of MLID.

8.19 Imprinted Genes and More Common Complications of Pregnancy

Imprinting disorders are relatively rare (Table 8.1) but low birth weight (LBW), defined by The World Health Organisation as birth weight <2500 g at any gestational age (United Nations Children's Fund and World Health Organisation 2004), is one of the most common complications of pregnancy, affecting up to 19% of all births in the developing world and between 5 and 7.5% of births in developed countries [252]. Several genes within the 11p15.5 IC1/IC2 domain are known to play an important role in determining birth weight in mice, either by intrinsically regulating fetal growth potential or through regulating placental development to extrinsically impact fetal growth (Fig. 8.4). Alterations in the expression of IC1/IC2 domain genes have been reported in association with fetal growth disorders in human pregnancy. Placental expression of *IGF2* generally positively correlates with birthweight consistent with studies in mice [91, 253]. For example, elevated placental *IGF2* has been reported in large for gestational age infants [254, 255] and lower placental *IGF2* in small for gestational age infants [256, 257]. Conversely, placental expression of *CDKN1C* generally negatively correlates with birthweight [234, 235, 256, 258–260], consistent with the finding that birth weight decreases with increasing levels of *Cdkn1c* in mice [176, 206]. *IGF2* and *CDKN1C* are not the only fetal growth restriction genes located on human chromosome 11p15. In mice elevated *Phlda2* drives late fetal growth restriction resulting in LBW followed by catch-up growth [182]. Abnormally elevated placental *PHLDA2* is a highly common alteration linked to LBW [261] with a prevalence that may be as high as 25% in confirmed cases of fetal growth restriction [262]. Increased placental expression of

both *PHLDA2* and *CDKN1C* has been observed in small-for-gestational-age (SGA) infants, highlighting the potential for domain-wide loss of imprinting [256].

In addition to regulating fetal growth, imprinted genes within the IC1/IC2 domain are critically important for placental development in mice, involved in both nutrient transport and the regulation of placental hormone production. Consequently, aberrant expression of imprinted genes in the fetally-derived placenta may contribute to disorders impacting the mother during pregnancy. For example, loss of function of *Cdkn1c* in the offspring has been linked to preeclampsia-like symptoms in the mouse dams [174], consistent with the finding of HELLP syndrome in human mothers of BWS infants with maternally inherited *CDKN1C* mutations [213]. Similarly in mice increased expression of fetal/placental *Igf2* mediated by disrupted IC1 domain imprinting impacts maternal glucose management with potential relevance to gestational diabetes [263]. Changes in the expression of *Phlda2* in the mouse placenta have been linked to alterations in the behavior of mothers toward their newborns [154] with potential relevance to pregnancy-related mood disorders [264].

8.20 Imprinted Gene Networks

In addition to the domain-wide regulation of allelic expression, some imprinted genes have been observed to interact within imprinted gene networks (IGNs) with important roles in development [265]. A number of imprinted genes located in different domains possess a strikingly similar pattern of expression suggesting interactions, with *Cdkn1c* and *Plagl1* (aka *Zac1*) providing a key early example [243]. The interaction between imprinted genes in networks was further highlighted through the bioinformatic analysis of genes co-expressed with *Plagl1* [266]. Disruption of these networks can result from changes in the expression of individual imprinted genes as shown for *H19* [267] or disruption of epigenetic regulators, as shown for Trim28 [268]. To add further complexity, mutations disrupting networks can have more than one phenotypic outcome as a consequence of IGNs adjusting stochastically. Bimodal phenotypes have been reported for a number of mouse models in which individual imprinted genes have been genetically modified, further exemplified by studies on *Trim28* mutants, where haploinsufficiency disrupted IGNs in mice manifesting an obese phenotype but not those of normal weight [268]. Evidence for similar stratification of IGNs in human populations was also reported in this study. Disrupted expression of IGNs may therefore contribute to the considerable clinical overlap noted for several imprinting disorders [210]. Disruptions to IGNs may contribute to a wide range of common human diseases (Fig. 8.4).

8.21 Imprinted Genes and Early Life Adversity

Imprinted genes may respond to external exposures at the individual, domain-wide, or IGN level contributing to human disease. Early life adversity is a major threat to human health because of the harmful outcomes for offspring in the short and longer term, a phenomenon described as fetal programming or developmental origins of

health and disease [269, 270]. The clear role of imprinted genes in regulating fetal growth and postnatal phenotypes along with the predicted flexibility of epigenetic marks involved in the establishment, spreading, and maintenance of their expression has led to many researchers asking whether imprinted genes respond to adversity. Such studies are dependent on the demonstration of loss of monoallelic expression mediated by changes in epigenetics marks which can be challenging to demonstrate. Exposure to a maternal low-protein diet drives persistent loss-of-imprinting of at least one gene within the IC2 domain [20]. In this study, *Cdkn1c* luciferase reporter mice were used in which gene expression is visualized as a bioluminescent signal in living mice. Exposure to low-protein diet from conception resulted in loss of normal paternal silencing alongside loss of DNA methylation at the somatic Cdkn1c-sDMR but without impacting methylation of the IC2 domain gDMR, consistent with previous studies reporting that DNA methylation at gDMRs is relatively resistant to dietary adversities [271, 272]. Expression and methylation changes occurred before birth and persisted into adulthood even under conditions of a normal diet. Critically, supplementation with the methyl donor folate in pregnancy prevented loss of silencing. While this study provides definitive experimental evidence that imprinted genes can respond epigenetically to a dietary exposure, it remains to be determined how many genes are responsive, under what adversity conditions and to what extent the mechanisms underpinning responsiveness are conserved. Given that different mechanisms lead to the establishment and maintenance of imprinting, it seems likely that multiple pathways could disrupt the epigenetic regulation of imprinted loci. Outcomes of individual adversities may, however, be similar due to disruptions in IGNs.

8.22 Conclusion

Imprinted genes are associated with rare imprinting disorders such as BWS and SRS, but animal models suggest that imprinted genes have a greater potential to contribute more widely to human diseases including low birth weight, and chronic health conditions such as obesity, type 2 diabetes, and mental health disorders (Fig. 8.4). Imprinted genes may also, at least in part, be associated with the well-established relationship between early life adversity and human disease. A major challenge is how to translate functional outcomes from genetically modified mice to humans. Many studies have focused on comparing DNA methylation levels at gDMRs as a proxy for imprinted gene expression levels. However, the majority of imprinted genes are not directly spanned by gDMRs, and multistep epigenetic processes are required to establish and maintain their monoallelic expression. Developments in genome-wide approaches, single-cell technologies, combined with mathematical modeling are required to fully establish the extent to which variation in the expression of imprinted genes might contribute to wide-ranging human disease.

Acknowledgments M.A.S. is supported by Wellcome Trust grant 209475/Z/17/Z, L.L. is supported by CIHR grant PJT-165992 and R.M.J is supported by BBSRC grant BB/V008684/1.

The authors are grateful to the wider imprinting community—both those whose work is directly covered and the many studies which we have not been able to include due to space constraints. Help with Fig. 8.1 from F. M. John.

References

1. Surani MA (1998) Imprinting and the initiation of gene silencing in the germ line. Cell 93(3): 309–312
2. Reik W, Walter J (2001) Genomic imprinting: parental influence on the genome. Nat Rev Genet 2(1):21–32. https://doi.org/10.1038/35047554
3. Ferguson-Smith AC, Surani MA (2001) Imprinting and the epigenetic asymmetry between parental genomes. Science 293(5532):1086–1089
4. Surani MA, Barton SC (1983) Development of gynogenetic eggs in the mouse: implications for parthenogenetic embryos. Science 222(4627):1034–1036. https://doi.org/10.1126/science.6648518
5. McGrath J, Solter D (1983) Nuclear transplantation in the mouse embryo by microsurgery and cell fusion. Science 220(4603):1300–1302. https://doi.org/10.1126/science.6857250
6. McGrath J, Solter D (1984) Completion of mouse embryogenesis requires both the materal and paternal genomes. Cell 37:179–183
7. Surani MA, Barton SC, Norris ML (1984) Development of reconstituted mouse eggs suggests imprinting of the genome during gametogenesis. Nature 308(5959):548–550
8. Barton SC, Surani MA, Norris ML (1984) Role of paternal and maternal genomes in mouse development. Nature 311:374–376
9. Solter D, Aronson J, Gilbert SF, McGrath J (1985) Nuclear transfer in mouse embryos: activation of the embryonic genome. Cold Spring Harb Symp Quant Biol 50:45–50
10. Surani MA, Barton SC, Norris ML (1987) Influence of parental chromosomes on spatial specificity in androgenetic-parthenogenetic chimaeras in the mouse. Nature 326:395–397
11. Thomson JA, Solter D (1988) The developmental fate of androgenetic, parthenogenetic and gynogenetic cells in chimeric gastrulating mouse embryos. Genes Dev 2:1344–1351
12. Cattanach BM, Kirk M (1985) Differential activity of maternally and paternally derived chromosome regions in mice. Nature 315(6019):496–498
13. Ferguson-Smith AC (2011) Genomic imprinting: the emergence of an epigenetic paradigm. Nat Rev Genet 12(8):565–575. https://doi.org/10.1038/nrg3032
14. Barlow DP, Bartolomei MS (2014) Genomic imprinting in mammals. Cold Spring Harb Perspect Biol 6(2). https://doi.org/10.1101/cshperspect.a018382
15. Ferguson-Smith AC, Bourc'his D (2018) The discovery and importance of genomic imprinting. Elife 7. https://doi.org/10.7554/eLife.42368
16. Moore T, Haig D (1991) Genomic imprinting in mammalian development: a parental tug-of-war. TIG 7(2):45–49
17. Haig D, Graham C (1991) Genomic imprinting and the strange case of the insulin-like growth factor II receptor. Cell 64(6):1045–1046
18. Wolf JB, Hager R (2006) A maternal-offspring coadaptation theory for the evolution of genomic imprinting. PLoS Biol 4(12):e380. https://doi.org/10.1371/journal.pbio.0040380
19. Strahan R (1983) Complete book of Australian mammals. Angus and Robertson
20. Van de Pette M, Abbas A, Feytout A, McNamara G, Bruno L, To WK, Dimond A, Sardini A, Webster Z, McGinty J, Paul EJ, Ungless MA, French PMW, Withers DJ, Uren A, Ferguson-Smith AC, Merkenschlager M, John RM, Fisher AG (2017) Visualizing changes in cdkn1c expression links early-life adversity to imprint mis-regulation in adults. Cell Rep 18(5): 1090–1099. https://doi.org/10.1016/j.celrep.2017.01.010
21. SanMiguel JM, Bartolomei MS (2018) DNA methylation dynamics of genomic imprinting in mouse development. Biol Reprod 99(1):252–262. https://doi.org/10.1093/biolre/ioy036

22. Tucci V, Isles AR, Kelsey G, Ferguson-Smith AC, Erice Imprinting G (2019) Genomic imprinting and physiological processes in mammals. Cell 176(5):952–965. https://doi.org/10.1016/j.cell.2019.01.043
23. Li JY, Lees-Murdock DJ, Xu GL, Walsh CP (2004) Timing of establishment of paternal methylation imprints in the mouse. Genomics 84(6):952–960. https://doi.org/10.1016/j.ygeno.2004.08.012
24. Kato Y, Kaneda M, Hata K, Kumaki K, Hisano M, Kohara Y, Okano M, Li E, Nozaki M, Sasaki H (2007) Role of the Dnmt3 family in de novo methylation of imprinted and repetitive sequences during male germ cell development in the mouse. Hum Mol Genet 16(19):2272–2280. https://doi.org/10.1093/hmg/ddm179
25. Seisenberger S, Andrews S, Krueger F, Arand J, Walter J, Santos F, Popp C, Thienpont B, Dean W, Reik W (2012) The dynamics of genome-wide DNA methylation reprogramming in mouse primordial germ cells. Mol Cell 48(6):849–862. https://doi.org/10.1016/j.molcel.2012.11.001
26. Kobayashi H, Sakurai T, Imai M, Takahashi N, Fukuda A, Yayoi O, Sato S, Nakabayashi K, Hata K, Sotomaru Y, Suzuki Y, Kono T (2012) Contribution of intragenic DNA methylation in mouse gametic DNA methylomes to establish oocyte-specific heritable marks. PLoS Genet 8(1):e1002440. https://doi.org/10.1371/journal.pgen.1002440
27. Henckel A, Chebli K, Kota SK, Arnaud P, Feil R (2012) Transcription and histone methylation changes correlate with imprint acquisition in male germ cells. EMBO J 31(3):606–615. https://doi.org/10.1038/emboj.2011.425
28. Hata K, Okano M, Lei H, Li E (2002) Dnmt3L cooperates with the Dnmt3 family of de novo DNA methyltransferases to establish maternal imprints in mice. Development 129(8):1983–1993
29. Bourc'his D, Xu GL, Lin CS, Bollman B, Bestor TH (2001) Dnmt3L and the establishment of maternal genomic imprints. Science 294(5551):2536–2539. https://doi.org/10.1126/science.1065848
30. Kaneda M, Okano M, Hata K, Sado T, Tsujimoto N, Li E, Sasaki H (2004) Essential role for de novo DNA methyltransferase Dnmt3a in paternal and maternal imprinting. Nature 429(6994):900–903. https://doi.org/10.1038/nature02633
31. Okano M, Bell DW, Haber DA, Li E (1999) DNA methyltransferases Dnmt3a and Dnmt3b are essential for de novo methylation and mammalian development. Cell 99(3):247–257
32. Watanabe T, Tomizawa S, Mitsuya K, Totoki Y, Yamamoto Y, Kuramochi-Miyagawa S, Iida N, Hoki Y, Murphy PJ, Toyoda A, Gotoh K, Hiura H, Arima T, Fujiyama A, Sado T, Shibata T, Nakano T, Lin H, Ichiyanagi K, Soloway PD, Sasaki H (2011) Role for piRNAs and noncoding RNA in de novo DNA methylation of the imprinted mouse Rasgrf1 locus. Science 332(6031):848–852. https://doi.org/10.1126/science.1203919
33. Shirane K, Miura F, Ito T, Lorincz MC (2020) NSD1-deposited H3K36me2 directs de novo methylation in the mouse male germline and counteracts Polycomb-associated silencing. Nat Genet 52(10):1088–1098. https://doi.org/10.1038/s41588-020-0689-z
34. Smallwood SA, Tomizawa S, Krueger F, Ruf N, Carli N, Segonds-Pichon A, Sato S, Hata K, Andrews SR, Kelsey G (2011) Dynamic CpG island methylation landscape in oocytes and preimplantation embryos. Nat Genet 43(8):811–814. https://doi.org/10.1038/ng.864
35. Shirane K, Toh H, Kobayashi H, Miura F, Chiba H, Ito T, Kono T, Sasaki H (2013) Mouse oocyte methylomes at base resolution reveal genome-wide accumulation of non-CpG methylation and role of DNA methyltransferases. PLoS Genet 9(4):e1003439. https://doi.org/10.1371/journal.pgen.1003439
36. Stewart KR, Veselovska L, Kim J, Huang J, Saadeh H, Tomizawa S, Smallwood SA, Chen T, Kelsey G (2015) Dynamic changes in histone modifications precede de novo DNA methylation in oocytes. Genes Dev 29(23):2449–2462. https://doi.org/10.1101/gad.271353.115
37. Xu Q, Xiang Y, Wang Q, Wang L, Brind'Amour J, Bogutz AB, Zhang Y, Zhang B, Yu G, Xia W, Du Z, Huang C, Ma J, Zheng H, Li Y, Liu C, Walker CL, Jonasch E, Lefebvre L, Wu M, Lorincz MC, Li W, Li L, Xie W (2019) SETD2 regulates the maternal epigenome,

genomic imprinting and embryonic development. Nat Genet 51(5):844–856. https://doi.org/10.1038/s41588-019-0398-7
38. Sun XJ, Wei J, Wu XY, Hu M, Wang L, Wang HH, Zhang QH, Chen SJ, Huang QH, Chen Z (2005) Identification and characterization of a novel human histone H3 lysine 36-specific methyltransferase. J Biol Chem 280(42):35261–35271. https://doi.org/10.1074/jbc.M504012200
39. Chedin F, Lieber MR, Hsieh CL (2002) The DNA methyltransferase-like protein DNMT3L stimulates de novo methylation by Dnmt3a. Proc Natl Acad Sci U S A 99(26):16916–16921. https://doi.org/10.1073/pnas.262443999
40. Suetake I, Shinozaki F, Miyagawa J, Takeshima H, Tajima S (2004) DNMT3L stimulates the DNA methylation activity of Dnmt3a and Dnmt3b through a direct interaction. J Biol Chem 279(26):27816–27823. https://doi.org/10.1074/jbc.M400181200
41. Kaneda M, Hirasawa R, Chiba H, Okano M, Li E, Sasaki H (2010) Genetic evidence for Dnmt3a-dependent imprinting during oocyte growth obtained by conditional knockout with Zp3-Cre and complete exclusion of Dnmt3b by chimera formation. Genes Cells. https://doi.org/10.1111/j.1365-2443.2009.01374.x
42. Veselovska L, Smallwood SA, Saadeh H, Stewart KR, Krueger F, Maupetit-Mehouas S, Arnaud P, Tomizawa S, Andrews S, Kelsey G (2015) Deep sequencing and de novo assembly of the mouse oocyte transcriptome define the contribution of transcription to the DNA methylation landscape. Genome Biol 16(1):209. https://doi.org/10.1186/s13059-015-0769-z
43. Bourc'his D, Bestor TH (2006) Origins of extreme sexual dimorphism in genomic imprinting. Cytogenet Genome Res 113(1–4):36–40. https://doi.org/10.1159/000090813
44. Chotalia M, Smallwood SA, Ruf N, Dawson C, Lucifero D, Frontera M, James K, Dean W, Kelsey G (2009) Transcription is required for establishment of germline methylation marks at imprinted genes. Genes Dev 23(1):105–117. https://doi.org/10.1101/gad.495809
45. Smith EY, Futtner CR, Chamberlain SJ, Johnstone KA, Resnick JL (2011) Transcription is required to establish maternal imprinting at the Prader-Willi syndrome and Angelman syndrome locus. PLoS Genet 7(12):e1002422. https://doi.org/10.1371/journal.pgen.1002422
46. Singh VB, Sribenja S, Wilson KE, Attwood KM, Hillman JC, Pathak S, Higgins MJ (2017) Blocked transcription through KvDMR1 results in absence of methylation and gene silencing resembling Beckwith-Wiedemann syndrome. Development 144(10):1820–1830. https://doi.org/10.1242/dev.145136
47. Bretz CL, Kim J (2017) Transcription-driven DNA methylation setting on the mouse Peg3 locus. Epigenetics 12(11):945–952. https://doi.org/10.1080/15592294.2017.1377869
48. Joh K, Matsuhisa F, Kitajima S, Nishioka K, Higashimoto K, Yatsuki H, Kono T, Koseki H, Soejima H (2018) Growing oocyte-specific transcription-dependent de novo DNA methylation at the imprinted Zrsr1-DMR. Epigenetics Chromatin 11(1):28. https://doi.org/10.1186/s13072-018-0200-6
49. Wang L, Zhang J, Duan J, Gao X, Zhu W, Lu X, Yang L, Zhang J, Li G, Ci W, Li W, Zhou Q, Aluru N, Tang F, He C, Huang X, Liu J (2014) Programming and inheritance of parental DNA methylomes in mammals. Cell 157(4):979–991. https://doi.org/10.1016/j.cell.2014.04.017
50. Takahashi N, Coluccio A, Thorball CW, Planet E, Shi H, Offner S, Turelli P, Imbeault M, Ferguson-Smith AC, Trono D (2019) ZNF445 is a primary regulator of genomic imprinting. Genes Dev 33(1–2):49–54. https://doi.org/10.1101/gad.320069.118
51. Messerschmidt DM, de Vries W, Ito M, Solter D, Ferguson-Smith A, Knowles BB (2012) Trim28 is required for epigenetic stability during mouse oocyte to embryo transition. Science 335(6075):1499–1502. https://doi.org/10.1126/science.1216154
52. Li X, Ito M, Zhou F, Youngson N, Zuo X, Leder P, Ferguson-Smith AC (2008) A maternal-zygotic effect gene, Zfp57, maintains both maternal and paternal imprints. Dev Cell 15(4):547–557. https://doi.org/10.1016/j.devcel.2008.08.014
53. Quenneville S, Verde G, Corsinotti A, Kapopoulou A, Jakobsson J, Offner S, Baglivo I, Pedone PV, Grimaldi G, Riccio A, Trono D (2011) In embryonic stem cells, ZFP57/KAP1

recognize a methylated hexanucleotide to affect chromatin and DNA methylation of imprinting control regions. Mol Cell 44(3):361–372. https://doi.org/10.1016/j.molcel.2011.08.032
54. Shi H, Strogantsev R, Takahashi N, Kazachenka A, Lorincz MC, Hemberger M, Ferguson-Smith AC (2019) ZFP57 regulation of transposable elements and gene expression within and beyond imprinted domains. Epigenetics Chromatin 12(1):49. https://doi.org/10.1186/s13072-019-0295-4
55. Howell CY, Bestor TH, Ding F, Latham KE, Mertineit C, Trasler JM, Chaillet JR (2001) Genomic imprinting disrupted by a maternal effect mutation in the Dnmt1 gene. Cell 104(6): 829–838. https://doi.org/10.1016/s0092-8674(01)00280-x
56. Hirasawa R, Chiba H, Kaneda M, Tajima S, Li E, Jaenisch R, Sasaki H (2008) Maternal and zygotic Dnmt1 are necessary and sufficient for the maintenance of DNA methylation imprints during preimplantation development. Genes Dev 22(12):1607–1616. https://doi.org/10.1101/gad.1667008
57. Weaver JR, Sarkisian G, Krapp C, Mager J, Mann MR, Bartolomei MS (2010) Domain-specific response of imprinted genes to reduced DNMT1. Mol Cell Biol 30(16):3916–3928. https://doi.org/10.1128/MCB.01278-09
58. Sharif J, Muto M, Takebayashi S, Suetake I, Iwamatsu A, Endo TA, Shinga J, Mizutani-Koseki Y, Toyoda T, Okamura K, Tajima S, Mitsuya K, Okano M, Koseki H (2007) The SRA protein Np95 mediates epigenetic inheritance by recruiting Dnmt1 to methylated DNA. Nature 450(7171):908–912. https://doi.org/10.1038/nature06397
59. Hanna CW, Kelsey G (2014) The specification of imprints in mammals. Heredity 113(2): 176–183. https://doi.org/10.1038/hdy.2014.54
60. White CR, MacDonald WA, Mann MR (2016) Conservation of DNA methylation programming between mouse and human gametes and preimplantation embryos. Biol Reprod 95(3): 61. https://doi.org/10.1095/biolreprod.116.140319
61. Bogutz AB, Brind'Amour J, Kobayashi H, Jensen KN, Nakabayashi K, Imai H, Lorincz MC, Lefebvre L (2019) Evolution of imprinting via lineage-specific insertion of retroviral promoters. Nat Commun 10(1):5674. https://doi.org/10.1038/s41467-019-13662-9
62. Wutz A, Smrzka OW, Schweifer N, Schellander K, Wagner EF, Barlow DP (1997) Imprinted expression of the Igf2r gene depends on an intronic CpG island. Nature 389(6652):745–749
63. Thorvaldsen JL, Duran KL, Bartolomei MS (1998) Deletion of the H19 differentially methylated domain results in loss of imprinted expression of H19 and Igf2. Genes Dev 12(23):3693–3702
64. Bielinska B, Blaydes SM, Buiting K, Yang T, Krajewska-Walasek M, Horsthemke B, Brannan CI (2000) De novo deletions of SNRPN exon 1 in early human and mouse embryos result in a paternal to maternal imprint switch. Nat Genet 25(1):74–78. https://doi.org/10.1038/75629
65. Fitzpatrick GV, Soloway PD, Higgins MJ (2002) Regional loss of imprinting and growth deficiency in mice with a targeted deletion of KvDMR1. Nat Genet 32(3):426–431
66. Lin SP, Youngson N, Takada S, Seitz H, Reik W, Paulsen M, Cavaille J, Ferguson-Smith AC (2003) Asymmetric regulation of imprinting on the maternal and paternal chromosomes at the Dlk1-Gtl2 imprinted cluster on mouse chromosome 12. Nat Genet 35(1):97–102
67. Williamson CM, Turner MD, Ball ST, Nottingham WT, Glenister P, Fray M, Tymowska-Lalanne Z, Plagge A, Powles-Glover N, Kelsey G, Maconochie M, Peters J (2006) Identification of an imprinting control region affecting the expression of all transcripts in the Gnas cluster. Nat Genet 38(3):350–355
68. MacDonald WA, Mann MRW (2020) Long noncoding RNA functionality in imprinted domain regulation. PLoS Genet 16(8):e1008930. https://doi.org/10.1371/journal.pgen.1008930
69. Sleutels F, Zwart R, Barlow DP (2002) The non-coding Air RNA is required for silencing autosomal imprinted genes. Nature 415(6873):810–813. https://doi.org/10.1038/415810a
70. Santoro F, Mayer D, Klement RM, Warczok KE, Stukalov A, Barlow DP, Pauler FM (2013) Imprinted Igf2r silencing depends on continuous Airn lncRNA expression and is not restricted

to a developmental window. Development 140(6):1184–1195. https://doi.org/10.1242/dev.088849
71. Kota SK, Lleres D, Bouschet T, Hirasawa R, Marchand A, Begon-Pescia C, Sanli I, Arnaud P, Journot L, Girardot M, Feil R (2014) ICR noncoding RNA expression controls imprinting and DNA replication at the Dlk1-Dio3 domain. Dev Cell 31(1):19–33. https://doi.org/10.1016/j.devcel.2014.08.009
72. Tibbit CJ, Williamson CM, Mehta S, Ball ST, Chotalia M, Nottingham WT, Eaton SA, Quwailid MM, Teboul L, Kelsey G, Peters J (2015) Antisense activity across the nesp promoter is required for nespas-mediated silencing in the imprinted gnas cluster. Noncoding RNA 1(3):246–265. https://doi.org/10.3390/ncrna1030246
73. Latos PA, Pauler FM, Koerner MV, Senergin HB, Hudson QJ, Stocsits RR, Allhoff W, Stricker SH, Klement RM, Warczok KE, Aumayr K, Pasierbek P, Barlow DP (2012) Airn transcriptional overlap, but not its lncRNA products, induces imprinted Igf2r silencing. Science 338(6113):1469–1472. https://doi.org/10.1126/science.1228110
74. Meng L, Person RE, Huang W, Zhu PJ, Costa-Mattioli M, Beaudet AL (2013) Truncation of Ube3a-ATS unsilences paternal Ube3a and ameliorates behavioral defects in the Angelman syndrome mouse model. PLoS Genet 9(12):e1004039. https://doi.org/10.1371/journal.pgen.1004039
75. Mancini-Dinardo D, Steele SJ, Levorse JM, Ingram RS, Tilghman SM (2006) Elongation of the Kcnq1ot1 transcript is required for genomic imprinting of neighboring genes. Genes Dev 20(10):1268–1282
76. Shin JY, Fitzpatrick GV, Higgins MJ (2008) Two distinct mechanisms of silencing by the KvDMR1 imprinting control region. EMBO J 27(1):168–178
77. Keshavarz M, Tautz D (2021) The imprinted lncRNA Peg13 regulates sexual preference and the sex-specific brain transcriptome in mice. Proc Natl Acad Sci U S A 118(10). https://doi.org/10.1073/pnas.2022172118
78. Kobayashi H, Sakurai T, Miura F, Imai M, Mochiduki K, Yanagisawa E, Sakashita A, Wakai T, Suzuki Y, Ito T, Matsui Y, Kono T (2013) High-resolution DNA methylome analysis of primordial germ cells identifies gender-specific reprogramming in mice. Genome Res 23(4):616–627. https://doi.org/10.1101/gr.148023.112
79. Hajkova P, Erhardt S, Lane N, Haaf T, El-Maarri O, Reik W, Walter J, Surani MA (2002) Epigenetic reprogramming in mouse primordial germ cells. Mech Dev 117(1–2):15–23
80. Yamaguchi S, Hong K, Liu R, Inoue A, Shen L, Zhang K, Zhang Y (2013) Dynamics of 5-methylcytosine and 5-hydroxymethylcytosine during germ cell reprogramming. Cell Res 23(3):329–339. https://doi.org/10.1038/cr.2013.22
81. Dawlaty MM, Breiling A, Le T, Raddatz G, Barrasa MI, Cheng AW, Gao Q, Powell BE, Li Z, Xu M, Faull KF, Lyko F, Jaenisch R (2013) Combined deficiency of Tet1 and Tet2 causes epigenetic abnormalities but is compatible with postnatal development. Dev Cell 24(3):310–323. https://doi.org/10.1016/j.devcel.2012.12.015
82. Hackett JA, Sengupta R, Zylicz JJ, Murakami K, Lee C, Down TA, Surani MA (2013) Germline DNA demethylation dynamics and imprint erasure through 5-hydroxymethylcytosine. Science 339(6118):448–452. https://doi.org/10.1126/science.1229277
83. Tahiliani M, Koh KP, Shen Y, Pastor WA, Bandukwala H, Brudno Y, Agarwal S, Iyer LM, Liu DR, Aravind L, Rao A (2009) Conversion of 5-methylcytosine to 5-hydroxymethylcytosine in mammalian DNA by MLL partner TET1. Science 324(5929):930–935. https://doi.org/10.1126/science.1170116
84. Gkountela S, Zhang KX, Shafiq TA, Liao WW, Hargan-Calvopina J, Chen PY, Clark AT (2015) DNA demethylation dynamics in the human prenatal germline. Cell 161(6):1425–1436. https://doi.org/10.1016/j.cell.2015.05.012
85. Yamaguchi S, Shen L, Liu Y, Sendler D, Zhang Y (2013) Role of Tet1 in erasure of genomic imprinting. Nature 504(7480):460–464. https://doi.org/10.1038/nature12805

86. SanMiguel JM, Abramowitz LK, Bartolomei MS (2018) Imprinted gene dysregulation in a Tet1 null mouse model is stochastic and variable in the germline and offspring. Development 145(7). https://doi.org/10.1242/dev.160622
87. Ferguson-Smith AC, Cattanach BM, Barton SC, Beechey CV, Surani MA (1991) Molecular and embryological investigations of parental imprinting on mouse chromosome 7. Nature 351: 667–670
88. Searle AG, Beechey CV (1990) Genome imprinting phenomena on mouse chromosome 7. Genet Res (Cambridge) 56:237–244
89. McLaughlin KJ, Solter D, Mann J (1997) Developmental consequences of two paternal copies of imprinted chromosome region distal 7 in mice. J Cell Physiol 173(2):242–246. https://doi.org/10.1002/(SICI)1097-4652(199711)173:2<242::AID-JCP29>3.0.CO;2-G
90. McLaughlin KJ, Szabò P, Haegel H, Mann JR (1996) Mouse embryos with paternal duplication of an imprinted chromosome 7 region die around midgestation and lack placental spongiotrophoblast. Development 122:265–270
91. DeChiara TM, Robertson EJ, Efstratiadis A (1991) Parental imprinting of the mouse insulin-like growth factor II gene. Cell 64(4):849–859
92. Deltour L, Montagutelli X, Guenet J-L, Jami J, Paldi A (1995) Tissue and developmental stage-specific imprinting of the mouse proinsulin gene, *Ins-2*. Dev Biol 168:686–688
93. Bartolomei MS, Zemei S, Tilghman SM (1991) Parental imprinting of the mouse H19 gene. Nature 351:153–155
94. Cai X, Cullen BR (2007) The imprinted H19 noncoding RNA is a primary microRNA precursor. RNA (New York, NY) 13(3):313–316. https://doi.org/10.1261/rna.351707
95. Ferguson-Smith AC, Sasaki H, Cattanach BM, Surani MA (1993) Parental-origin-specific epigenetic modification of the mouse H19 gene. Nature 362(6422):751–755. https://doi.org/10.1038/362751a0
96. John RM, Lefebvre L (2011) Developmental regulation of somatic imprints. Differentiation 81(5):270–280. https://doi.org/10.1016/j.diff.2011.01.007
97. Leighton PA, Saam JR, Ingram RS, Stewart CL, Tilghman SM (1995) An enhancer deletion affects both H19 and Igf2 expression. Genes Dev 9(17):2079–2089
98. Bell AC, Felsenfeld G (2000) Methylation of a CTCF-dependent boundary controls imprinted expression of the Igf2 gene. Nature 405(6785):482–485. https://doi.org/10.1038/35013100
99. Hark AT, Schoenherr CJ, Katz DJ, Ingram RS, Levorse JM, Tilghman SM (2000) CTCF mediates methylation-sensitive enhancer-blocking activity at the H19/Igf2 locus. Nature 405(6785):486–489. https://doi.org/10.1038/35013106
100. Kanduri C, Pant V, Loukinov D, Pugacheva E, Qi CF, Wolffe A, Ohlsson R, Lobanenkov VV (2000) Functional association of CTCF with the insulator upstream of the H19 gene is parent of origin-specific and methylation-sensitive. Curr Biol 10(14):853–856. https://doi.org/10.1016/s0960-9822(00)00597-2
101. Engel N, Raval AK, Thorvaldsen JL, Bartolomei SM (2008) Three-dimensional conformation at the H19/Igf2 locus supports a model of enhancer tracking. Hum Mol Genet 17(19): 3021–3029. https://doi.org/10.1093/hmg/ddn200
102. Kurukuti S, Tiwari VK, Tavoosidana G, Pugacheva E, Murrell A, Zhao Z, Lobanenkov V, Reik W, Ohlsson R (2006) CTCF binding at the H19 imprinting control region mediates maternally inherited higher-order chromatin conformation to restrict enhancer access to Igf2. Proc Natl Acad Sci U S A 103(28):10684–10689. https://doi.org/10.1073/pnas.0600326103
103. Li T, Hu JF, Qiu X, Ling J, Chen H, Wang S, Hou A, Vu TH, Hoffman AR (2008) CTCF regulates allelic expression of Igf2 by orchestrating a promoter-polycomb repressive complex 2 intrachromosomal loop. Mol Cell Biol 28(20):6473–6482. https://doi.org/10.1128/MCB.00204-08
104. Qiu X, Vu TH, Lu Q, Ling JQ, Li T, Hou A, Wang SK, Chen HL, Hu JF, Hoffman AR (2008) A complex deoxyribonucleic acid looping configuration associated with the silencing of the maternal Igf2 allele. Mol Endocrinol 22(6):1476–1488. https://doi.org/10.1210/me.2007-0474

105. Murrell A, Heeson S, Reik W (2004) Interaction between differentially methylated regions partitions the imprinted genes Igf2 and H19 into parent-specific chromatin loops. Nat Genet 36(8):889–893. https://doi.org/10.1038/ng1402
106. Court F, Baniol M, Hagege H, Petit JS, Lelay-Taha MN, Carbonell F, Weber M, Cathala G, Forne T (2011) Long-range chromatin interactions at the mouse Igf2/H19 locus reveal a novel paternally expressed long non-coding RNA. Nucleic Acids Res 39(14):5893–5906. https://doi.org/10.1093/nar/gkr209
107. Lleres D, Moindrot B, Pathak R, Piras V, Matelot M, Pignard B, Marchand A, Poncelet M, Perrin A, Tellier V, Feil R, Noordermeer D (2019) CTCF modulates allele-specific sub-TAD organization and imprinted gene activity at the mouse Dlk1-Dio3 and Igf2-H19 domains. Genome Biol 20(1):272. https://doi.org/10.1186/s13059-019-1896-8
108. Qian N, Frank D, O'Keefe D, Dao D, Zhao L, Yuan L, Wang Q, Keating M, Walsh C, Tycko B (1997) The IPL gene on chromosome 11p15.5 is imprinted in humans and mice and is similar to TDAG51, implicated in Fas expression and apoptosis. Hum Mol Genet 6(12): 2021–2029
109. Hatada I, Mukai T (1995) Genomic imprinting of p57KIP2, a cyclin-dependent kinase inhibitor, in mouse. Nat Genet 11(2):204–206
110. Guillemot F, Caspary T, Tilghman SM, Copeland NG, Gilbert DJ, Jenkins NA, Anderson DJ, Joyner AL, Rossant J, Nagy A (1995) Genomic imprinting of *Mash2*, a mouse gene required for trophoblast development. Nat Genet 9:235–242
111. Smilinich NJ, Day CD, Fitzpatrick GV, Caldwell GM, Lossie AC, Cooper PR, Smallwood AC, Joyce JA, Schofield PN, Reik W, Nicholls RD, Weksberg R, Driscoll DJ, Maher ER, Shows TB, Higgins MJ (1999) A maternally methylated CpG island in KvLQT1 is associated with an antisense paternal transcript and loss of imprinting in Beckwith-Wiedemann syndrome. Proc Natl Acad Sci U S A 96(14):8064–8069. https://doi.org/10.1073/pnas.96.14.8064
112. Engemann S, Strodicke M, Paulsen M, Franck O, Reinhardt R, Lane N, Reik W, Walter J (2000) Sequence and functional comparison in the beckwith-wiedemann region: implications for a novel imprinting centre and extended imprinting. Hum Mol Genet 9(18):2691–2706
113. Lewis A, Mitsuya K, Umlauf D, Smith P, Dean W, Walter J, Higgins M, Feil R, Reik W (2004) Imprinting on distal chromosome 7 in the placenta involves repressive histone methylation independent of DNA methylation. Nat Genet 36(12):1291–1295
114. Terranova R, Yokobayashi S, Stadler MB, Otte AP, van Lohuizen M, Orkin SH, Peters AH (2008) Polycomb group proteins Ezh2 and Rnf2 direct genomic contraction and imprinted repression in early mouse embryos. Dev Cell 15(5):668–679
115. Umlauf D, Goto Y, Cao R, Cerqueira F, Wagschal A, Zhang Y, Feil R (2004) Imprinting along the Kcnq1 domain on mouse chromosome 7 involves repressive histone methylation and recruitment of Polycomb group complexes. Nat Genet 36(12):1296–1300
116. Kato Y, Iii WM, Hilton K, Barton SC, Tsunoda Y, Surani MA (1999) Developmental potential of mouse primordial germ cells. Development 126(9):1823–1832
117. Tada T, Tada M, Hilton K, Barton SC, Sado T, Takagi N, Surani MA (1998) Epigenotype switching of imprintable loci in embryonic germ cells. Dev Genes Evol 207(8):551–561
118. Kono T, Obata Y, Yoshimzu T, Nakahara T, Carroll J (1996) Epigenetic modifications during oocyte growth correlates with extended parthenogenetic development in the mouse. Nat Genet 13:91–94
119. Obata Y, Kaneko-Ishino T, Koide T, Takai Y, Ueda T, Domeki I, Shiroishi T, Ishino F, Kono T (1998) Disruption of primary imprinting during oocyte growth leads to the modified expression of imprinted genes during embryogenesis. Development 125(8):1553–1560
120. Bhogal B, Arnaudo A, Dymkowski A, Best A, Davis TL (2004) Methylation at mouse Cdkn1c is acquired during postimplantation development and functions to maintain imprinted expression. Genomics 84(6):961–970
121. Fan T, Hagan JP, Kozlov SV, Stewart CL, Muegge K (2005) Lsh controls silencing of the imprinted Cdkn1c gene. Development 132(4):635–644

122. Caspary T, Cleary MA, Baker CC, Guan XJ, Tilghman SM (1998) Multiple mechanisms regulate imprinting of the mouse distal chromosome 7 gene cluster. Mol Cell Biol 18(6): 3466–3474
123. Wood MD, Hiura H, Tunster S, Arima T, Shin JY, Higgins M, John RM (2010) Autonomous silencing of the imprinted Cdkn1c gene in stem cells. Epigenetics 5(3)
124. Pauler FM, Koerner MV, Barlow DP (2007) Silencing by imprinted noncoding RNAs: is transcription the answer? Trends Genet 23(6):284–292. https://doi.org/10.1016/j.tig.2007.03.018
125. Andergassen D, Muckenhuber M, Bammer PC, Kulinski TM, Theussl HC, Shimizu T, Penninger JM, Pauler FM, Hudson QJ (2019) The Airn lncRNA does not require any DNA elements within its locus to silence distant imprinted genes. PLoS Genet 15(7):e1008268. https://doi.org/10.1371/journal.pgen.1008268
126. John RM, Ainscough JF, Barton SC, Surani MA (2001) Distant cis-elements regulate imprinted expression of the mouse p57 (Kip2) (Cdkn1c) gene: implications for the human disorder, Beckwith—Wiedemann syndrome. Hum Mol Genet 10(15):1601–1609
127. Pandey RR, Mondal T, Mohammad F, Enroth S, Redrup L, Komorowski J, Nagano T, Mancini-Dinardo D, Kanduri C (2008) Kcnq1ot1 antisense noncoding RNA mediates lineage-specific transcriptional silencing through chromatin-level regulation. Mol Cell 32(2): 232–246. https://doi.org/10.1016/j.molcel.2008.08.022
128. Redrup L, Branco MR, Perdeaux ER, Krueger C, Lewis A, Santos F, Nagano T, Cobb BS, Fraser P, Reik W (2009) The long noncoding RNA Kcnq1ot1 organises a lineage-specific nuclear domain for epigenetic gene silencing. Development 136(4):525–530
129. Schertzer MD, Braceros KCA, Starmer J, Cherney RE, Lee DM, Salazar G, Justice M, Bischoff SR, Cowley DO, Ariel P, Zylka MJ, Dowen JM, Magnuson T, Calabrese JM (2019) lncRNA-induced spread of polycomb controlled by genome architecture, RNA abundance, and CpG Island DNA. Mol Cell 75(3):523–537. https://doi.org/10.1016/j.molcel.2019.05.028
130. Mohammad F, Pandey RR, Nagano T, Chakalova L, Mondal T, Fraser P, Kanduri C (2008) Kcnq1ot1/Lit1 noncoding RNA mediates transcriptional silencing by targeting to the perinucleolar region. Mol Cell Biol 28(11):3713–3728
131. Fedoriw AM, Calabrese JM, Mu W, Yee D, Magnuson T (2012) Differentiation-driven nucleolar association of the mouse imprinted Kcnq1 locus. G3 (Bethesda) 2(12):1521–1528. https://doi.org/10.1534/g3.112.004226
132. Sachani SS, Landschoot LS, Zhang L, White CR, MacDonald WA, Golding MC, Mann MRW (2018) Nucleoporin 107, 62 and 153 mediate Kcnq1ot1 imprinted domain regulation in extraembryonic endoderm stem cells. Nat Commun 9(1):2795. https://doi.org/10.1038/s41467-018-05208-2
133. Golding MC, Magri LS, Zhang L, Lalone SA, Higgins MJ, Mann MR (2011) Depletion of Kcnq1ot1 non-coding RNA does not affect imprinting maintenance in stem cells. Development 138(17):3667–3678. https://doi.org/10.1242/dev.057778
134. Mohammad F, Pandey GK, Mondal T, Enroth S, Redrup L, Gyllensten U, Kanduri C (2012) Long noncoding RNA-mediated maintenance of DNA methylation and transcriptional gene silencing. Development 139(15):2792–2803. https://doi.org/10.1242/dev.079566
135. DeVeale B, van der Kooy D, Babak T (2012) Critical evaluation of imprinted gene expression by RNA-Seq: a new perspective. PLoS Genet 8(3):e1002600. https://doi.org/10.1371/journal.pgen.1002600
136. Kelsey G, Bartolomei MS (2012) Imprinted genes . . . and the number is? PLoS Genet 8(3): e1002601. https://doi.org/10.1371/journal.pgen.1002601
137. Inoue A, Jiang L, Lu F, Suzuki T, Zhang Y (2017) Maternal H3K27me3 controls DNA methylation-independent imprinting. Nature 547(7664):419–424. https://doi.org/10.1038/nature23262
138. Hanna CW, Perez-Palacios R, Gahurova L, Schubert M, Krueger F, Biggins L, Andrews S, Colome-Tatche M, Bourc'his D, Dean W, Kelsey G (2019) Endogenous retroviral insertions

drive non-canonical imprinting in extra-embryonic tissues. Genome Biol 20(1):225. https://doi.org/10.1186/s13059-019-1833-x
139. Chen Z, Yin Q, Inoue A, Zhang C, Zhang Y (2019) Allelic H3K27me3 to allelic DNA methylation switch maintains noncanonical imprinting in extraembryonic cells. Sci Adv 5(12):eaay7246. https://doi.org/10.1126/sciadv.aay7246
140. Zhang W, Chen Z, Yin Q, Zhang D, Racowsky C, Zhang Y (2019) Maternal-biased H3K27me3 correlates with paternal-specific gene expression in the human morula. Genes Dev 33(7–8):382–387. https://doi.org/10.1101/gad.323105.118
141. Xia W, Xu J, Yu G, Yao G, Xu K, Ma X, Zhang N, Liu B, Li T, Lin Z, Chen X, Li L, Wang Q, Shi D, Shi S, Zhang Y, Song W, Jin H, Hu L, Bu Z, Wang Y, Na J, Xie W, Sun YP (2019) Resetting histone modifications during human parental-to-zygotic transition. Science 365(6451):353–360. https://doi.org/10.1126/science.aaw5118
142. Gehring M (2013) Genomic imprinting: insights from plants. Annu Rev Genet 47:187–208. https://doi.org/10.1146/annurev-genet-110711-155527
143. Renfree MB, Hore TA, Shaw G, Graves JA, Pask AJ (2009) Evolution of genomic imprinting: insights from marsupials and monotremes. Annu Rev Genomics Hum Genet 10:241–262. https://doi.org/10.1146/annurev-genom-082908-150026
144. Renfree MB, Ager EI, Shaw G, Pask AJ (2008) Genomic imprinting in marsupial placentation. Reproduction 136(5):523–531
145. Nabetani A, Hatada I, Morisaki H, Oshimura M, Mukai T (1997) Mouse U2af1-rs1 is a neomorphic imprinted gene. Mol Cell Biol 17(2):789–798
146. Choi JD, Underkoffler LA, Wood AJ, Collins JN, Williams PT, Golden JA, Schuster EF Jr, Loomes KM, Oakey RJ (2005) A novel variant of Inpp5f is imprinted in brain, and its expression is correlated with differential methylation of an internal CpG island. Mol Cell Biol 25(13):5514–5522. https://doi.org/10.1128/MCB.25.13.5514-5522.2005
147. Wood AJ, Roberts RG, Monk D, Moore GE, Schulz R, Oakey RJ (2007) A screen for retrotransposed imprinted genes reveals an association between X chromosome homology and maternal germ-line methylation. PLoS Genet 3(2):e20. https://doi.org/10.1371/journal.pgen.0030020
148. McCole RB, Loughran NB, Chahal M, Fernandes LP, Roberts RG, Fraternali F, O'Connell MJ, Oakey RJ (2011) A case-by-case evolutionary analysis of four imprinted retrogenes. Evolution 65(5):1413–1427. https://doi.org/10.1111/j.1558-5646.2010.01213.x
149. Brind'Amour J, Kobayashi H, Richard Albert J, Shirane K, Sakashita A, Kamio A, Bogutz A, Koike T, Karimi MM, Lefebvre L, Kono T, Lorincz MC (2018) LTR retrotransposons transcribed in oocytes drive species-specific and heritable changes in DNA methylation. Nat Commun 9(1):3331. https://doi.org/10.1038/s41467-018-05841-x
150. Proudhon C, Duffie R, Ajjan S, Cowley M, Iranzo J, Carbajosa G, Saadeh H, Holland ML, Oakey RJ, Rakyan VK, Schulz R, Bourc'his D (2012) Protection against de novo methylation is instrumental in maintaining parent-of-origin methylation inherited from the gametes. Mol Cell 47(6):909–920. https://doi.org/10.1016/j.molcel.2012.07.010
151. Smith FM, Garfield AS, Ward A (2006) Regulation of growth and metabolism by imprinted genes. Cytogenet Genome Res 113(1–4):279–291
152. Peters J (2014) The role of genomic imprinting in biology and disease: an expanding view. Nat Rev Genet 15(8):517–530. https://doi.org/10.1038/nrg3766
153. Cleaton MA, Edwards CA, Ferguson-Smith AC (2014) Phenotypic outcomes of imprinted gene models in mice: elucidation of pre- and postnatal functions of imprinted genes. Annu Rev Genomics Hum Genet 15:93–126. https://doi.org/10.1146/annurev-genom-091212-153441
154. Creeth HDJ, McNamara GI, Isles AR, John RM (2019) Imprinted genes influencing the quality of maternal care. Front Neuroendocrinol 53:100732. https://doi.org/10.1016/j.yfrne.2018.12.003
155. Kaneko-Ishino T, Ishino F (2019) Evolution of viviparity in mammals: what genomic imprinting tells us about mammalian placental evolution. Reprod Fertil Dev 31(7):1219–1227. https://doi.org/10.1071/RD18127

156. Hanin G, Ferguson-Smith AC (2020) The evolution of genomic imprinting: Epigenetic control of mammary gland development and postnatal resource control. Wiley Interdiscip Rev Syst Biol Med 12(3):e1476. https://doi.org/10.1002/wsbm.1476
157. Barlow DP, Stoger R, Herrmann BG, Saito K, Schweifer N (1991) The mouse insulin-like growth factor type-2 receptor is imprinted and closely linked to the Tme locus. Nature 349(6304):84–87. https://doi.org/10.1038/349084a0
158. Morgan DO, Edman JC, Standring DN, Fried VA, Smith MC, Roth R, Rutter WJ (1987) Insulin-like growth factor II receptor as a multifunctional binding protein. Nature 329:301–307
159. Ripoche M-A, Kress C, Poirier F, Dandolo L (1997) Deletion of the *H19* transcription unit reveals the existence of a putative imprinting control element. Genes Dev 11:1596–1604
160. John R, Hemberger M (2012) A placenta for life. Reprod Biomed Online 25(1):5–11. https://doi.org/10.1016/j.rbmo.2012.03.018
161. Coan PM, Fowden AL, Constancia M, Ferguson-Smith AC, Burton GJ, Sibley CP (2008) Disproportional effects of Igf2 knockout on placental morphology and diffusional exchange characteristics in the mouse. J Physiol 586(Pt 20):5023–5032. https://doi.org/10.1113/jphysiol.2008.157313
162. Esquiliano DR, Guo W, Liang L, Dikkes P, Lopez MF (2009) Placental glycogen stores are increased in mice with H19 null mutations but not in those with insulin or IGF type 1 receptor mutations. Placenta 30(8):693–699
163. Brannan CI, Dees EC, Ingram RS, Tilghman SM (1990) The product of the H19 gene may function as an RNA. Mol Cell Biol 10:28–36
164. Jones BK, Levorse JM, Tilghman SM (1998) Igf2 imprinting does not require its own DNA methylation or H19 RNA. Genes Dev 12(14):2200–2207. https://doi.org/10.1101/gad.12.14.2200
165. Yoshimizu T, Miroglio A, Ripoche MA, Gabory A, Vernucci M, Riccio A, Colnot S, Godard C, Terris B, Jammes H, Dandolo L (2008) The H19 locus acts in vivo as a tumor suppressor. Proc Natl Acad Sci U S A 105(34):12417–12422. https://doi.org/10.1073/pnas.0801540105
166. Mineno J, Okamoto S, Ando T, Sato M, Chono H, Izu H, Takayama M, Asada K, Mirochnitchenko O, Inouye M, Kato I (2006) The expression profile of microRNAs in mouse embryos. Nucleic Acids Res 34(6):1765–1771. https://doi.org/10.1093/nar/gkl096
167. Keniry A, Oxley D, Monnier P, Kyba M, Dandolo L, Smits G, Reik W (2012) The H19 lincRNA is a developmental reservoir of miR-675 that suppresses growth and Igf1r. Nat Cell Biol 14(7):659–665. https://doi.org/10.1038/ncb2521
168. Duvillie B, Cordonnier N, Deltour L, Dandoy-Dron F, Itier JM, Monthioux E, Jami J, Joshi RL, Bucchini D (1997) Phenotypic alterations in insulin-deficient mutant mice. Proc Natl Acad Sci U S A 94(10):5137–5140
169. Leroux L, Desbois P, Lamotte L, Duvillie B, Cordonnier N, Jackerott M, Jami J, Bucchini D, Joshi RL (2001) Compensatory responses in mice carrying a null mutation for Ins1 or Ins2. Diabetes 50(Suppl 1):S150–S153. https://doi.org/10.2337/diabetes.50.2007.s150
170. Zhang P, Liegeois NJ, Wong C, Finegold M, Hou H, Thompson JC, Silverman A, Harper JW, DePinho RA, Elledge SJ (1997) Altered cell differentiation and proliferation in mice lacking p57KIP2 indicates a role in Beckwith-Wiedemann syndrome. Nature 387(6629):151–158
171. Yan Y, Frisen J, Lee MH, Massague J, Barbacid M (1997) Ablation of the CDK inhibitor p57Kip2 results in increased apoptosis and delayed differentiation during mouse development. Genes Dev 11(8):973–983
172. Takahashi K, Kobayashi T, Kanayama N (2000) p57(Kip2) regulates the proper development of labyrinthine and spongiotrophoblasts. Mol Hum Reprod 6(11):1019–1025
173. Takahashi K, Nakayama K, Nakayama K (2000) Mice lacking a CDK inhibitor, p57Kip2, exhibit skeletal abnormalities and growth retardation. J Biochem (Tokyo) 127(1):73–83

174. Kanayama N, Takahashi K, Matsuura T, Sugimura M, Kobayashi T, Moniwa N, Tomita M, Nakayama K (2002) Deficiency in p57Kip2 expression induces preeclampsia-like symptoms in mice. Mol Hum Reprod 8(12):1129–1135
175. Tunster SJ, Van de Pette M, John RM (2011) Fetal overgrowth in the Cdkn1c mouse model of Beckwith-Wiedemann syndrome. Dis Model Mech 4(6):814–821. https://doi.org/10.1242/dmm.007328
176. Andrews SC, Wood MD, Tunster SJ, Barton SC, Surani MA, John RM (2007) Cdkn1c (p57Kip2) is the major regulator of embryonic growth within its imprinted domain on mouse distal chromosome 7. BMC Dev Biol 7:53. https://doi.org/10.1186/1471-213X-7-53
177. Frank D, Mendelsohn CL, Ciccone E, Svensson K, Ohlsson R, Tycko B (1999) A novel pleckstrin homology-related gene family defined by Ipl/Tssc3, TDAG51, and Tih1: tissue-specific expression, chromosomal location, and parental imprinting. Mamm Genome 10(12): 1150–1159
178. Frank D, Fortino W, Clark L, Musalo R, Wang W, Saxena A, Li CM, Reik W, Ludwig T, Tycko B (2002) Placental overgrowth in mice lacking the imprinted gene Ipl. Proc Natl Acad Sci U S A 99(11):7490–7495
179. Salas M, John R, Saxena A, Barton S, Frank D, Fitzpatrick G, Higgins MJ, Tycko B (2004) Placental growth retardation due to loss of imprinting of Phlda2. Mech Dev 121(10): 1199–1210
180. Tunster SJ, Tycko B, John RM (2010) The imprinted Phlda2 gene regulates extraembryonic energy stores. Mol Cell Biol 30(1):295–306. https://doi.org/10.1128/MCB.00662-09
181. Tunster SJ, Creeth HD, John RM (2016) The imprinted Phlda2 gene modulates a major endocrine compartment of the placenta to regulate placental demands for maternal resources. Dev Biol 409(1):251–260. https://doi.org/10.1016/j.ydbio.2015.10.015
182. Tunster SJ, Van De Pette M, John RM (2014) Isolating the role of elevated Phlda2 in asymmetric late fetal growth restriction in mice. Dis Model Mech 7(10):1185–1191. https://doi.org/10.1242/dmm.017079
183. Creeth HDJ, McNamara GI, Tunster SJ, Boque-Sastre R, Allen B, Sumption L, Eddy JB, Isles AR, John RM (2018) Maternal care boosted by paternal imprinting in mammals. PLoS Biol 16(7):e2006599. https://doi.org/10.1371/journal.pbio.2006599
184. Lefebvre L, Viville S, Barton SC, Ishino F, Keverne EB, Surani MA (1998) Abnormal maternal behaviour and growth retardation associated with loss of the imprinted gene Mest [see comments]. Nat Genet 20(2):163–169
185. Li L, Keverne EB, Aparicio SA, Ishino F, Barton SC, Surani MA (1999) Regulation of maternal behavior and offspring growth by paternally expressed Peg3. Science 284(5412): 330–333
186. Harrison DJ, Creeth HDJ, Tyson HR, Boque-Sastre R, Hunter S, Dwyer DM, Isles AR, John RM (2021) Placental endocrine insufficiency programs anxiety, deficits in cognition and atypical social behaviour in offspring. Hum Mol Genet. https://doi.org/10.1093/hmg/ddab154
187. Guillemot F, Nagy A, Auerbach A, Rossant J, Joyner AL (1994) Essential role of Mash-2 in extraembryonic development. Nature 371(6495):333–336
188. Tanaka M, Gertsenstein M, Rossant J, Nagy A (1997) Mash2 acts cell autonomously in mouse spongiotrophoblast development. Dev Biol 190(1):55–65
189. Oh-McGinnis R, Bogutz AB, Lefebvre L (2011) Partial loss of Ascl2 function affects all three layers of the mature placenta and causes intrauterine growth restriction. Dev Biol 351(2): 277–286. https://doi.org/10.1016/j.ydbio.2011.01.008
190. Bogutz AB, Oh-McGinnis R, Jacob KJ, Ho-Lau R, Gu T, Gertsenstein M, Nagy A, Lefebvre L (2018) Transcription factor ASCL2 is required for development of the glycogen trophoblast cell lineage. PLoS Genet 14(8):e1007587. https://doi.org/10.1371/journal.pgen.1007587
191. Tunster SJ, McNamara GI, Creeth HD, John RM (2016) Increased dosage of the imprinted Ascl2 gene restrains two key endocrine lineages of the mouse Placenta. Dev Biol 418(1): 55–65. https://doi.org/10.1016/j.ydbio.2016.08.014

192. John RM (2017) Imprinted genes and the regulation of placental endocrine function: Pregnancy and beyond. Placenta 56:86–90. https://doi.org/10.1016/j.placenta.2017.01.099
193. Millership SJ, Van de Pette M, Withers DJ (2019) Genomic imprinting and its effects on postnatal growth and adult metabolism. Cell Mol Life Sci 76(20):4009–4021. https://doi.org/10.1007/s00018-019-03197-z
194. Ferron SR, Radford EJ, Domingo-Muelas A, Kleine I, Ramme A, Gray D, Sandovici I, Constancia M, Ward A, Menheniott TR, Ferguson-Smith AC (2015) Differential genomic imprinting regulates paracrine and autocrine roles of IGF2 in mouse adult neurogenesis. Nat Commun 6:8265. https://doi.org/10.1038/ncomms9265
195. Furutachi S, Matsumoto A, Nakayama KI, Gotoh Y (2013) p57 controls adult neural stem cell quiescence and modulates the pace of lifelong neurogenesis. EMBO J. https://doi.org/10.1038/emboj.2013.50
196. Joseph B, Wallen-Mackenzie A, Benoit G, Murata T, Joodmardi E, Okret S, Perlmann T (2003) p57(Kip2) cooperates with Nurr1 in developing dopamine cells. Proc Natl Acad Sci U S A 100(26):15619–15624
197. McNamara GI, Davis BA, Dwyer DM, John RM, Isles AR (2016) Behavioural abnormalities in a novel mouse model for Silver Russell Syndrome. Hum Mol Genet. https://doi.org/10.1093/hmg/ddw357
198. McNamara GI, John RM, Isles AR (2018) Territorial behavior and social stability in the mouse require correct expression of imprinted Cdkn1c. Front Behav Neurosci 12:28. https://doi.org/10.3389/fnbeh.2018.00028
199. McNamara GI, Davis BA, Browne M, Humby T, Dalley JW, Xia J, John RM, Isles AR (2018) Dopaminergic and behavioural changes in a loss-of-imprinting model of Cdkn1c. Genes Brain Behav 17(2):149–157. https://doi.org/10.1111/gbb.12422
200. Chen DY, Stern SA, Garcia-Osta A, Saunier-Rebori B, Pollonini G, Bambah-Mukku D, Blitzer RD, Alberini CM (2011) A critical role for IGF-II in memory consolidation and enhancement. Nature 469(7331):491–497. https://doi.org/10.1038/nature09667
201. Charalambous M, Menheniott TR, Bennett WR, Kelly SM, Dell G, Dandolo L, Ward A (2004) An enhancer element at the Igf2/H19 locus drives gene expression in both imprinted and non-imprinted tissues. Dev Biol 271(2):488–497. https://doi.org/10.1016/j.ydbio.2004.04.022
202. Imaizumi Y, Furutachi S, Watanabe T, Miya H, Kawaguchi D, Gotoh Y (2020) Role of the imprinted allele of the Cdkn1c gene in mouse neocortical development. Sci Rep 10(1):1884. https://doi.org/10.1038/s41598-020-58629-9
203. Scagliotti V, Costa Fernandes Esse R, Willis TL, Howard M, Carrus I, Lodge E, Andoniadou CL, Charalambous M (2021) Dynamic expression of imprinted genes in the developing and postnatal pituitary gland. Genes (Basel) 12(4). https://doi.org/10.3390/genes12040509
204. Bilodeau S, Roussel-Gervais A, Drouin J (2009) Distinct developmental roles of cell cycle inhibitors p57Kip2 and p27Kip1 distinguish pituitary progenitor cell cycle exit from cell cycle reentry of differentiated cells. Mol Cell Biol 29(7):1895–1908. https://doi.org/10.1128/MCB.01885-08
205. Borensztein M, Viengchareun S, Montarras D, Journot L, Binart N, Lombes M, Dandolo L (2012) Double Myod and Igf2 inactivation promotes brown adipose tissue development by increasing Prdm16 expression. FASEB J 26(11):4584–4591. https://doi.org/10.1096/fj.12-208496
206. Van De Pette M, Tunster SJ, McNamara GI, Shelkovnikova T, Millership S, Benson L, Peirson S, Christian M, Vidal-Puig A, John RM (2016) Cdkn1c boosts the development of brown adipose tissue in a murine model of silver russell syndrome. PLoS Genet 12(3):e1005916. https://doi.org/10.1371/journal.pgen.1005916
207. Hammerle CM, Sandovici I, Brierley GV, Smith NM, Zimmer WE, Zvetkova I, Prosser HM, Sekita Y, Lam BYH, Ma M, Cooper WN, Vidal-Puig A, Ozanne SE, Medina-Gomez G, Constancia M (2020) Mesenchyme-derived IGF2 is a major paracrine regulator of pancreatic growth and function. PLoS Genet 16(10):e1009069. https://doi.org/10.1371/journal.pgen.1009069

208. Georgia S, Soliz R, Li M, Zhang P, Bhushan A (2006) p57 and Hes1 coordinate cell cycle exit with self-renewal of pancreatic progenitors. Dev Biol 298(1):22–31. https://doi.org/10.1016/j.ydbio.2006.05.036
209. Eggermann T, Perez de Nanclares G, Maher ER, Temple IK, Tumer Z, Monk D, Mackay DJ, Gronskov K, Riccio A, Linglart A, Netchine I (2015) Imprinting disorders: a group of congenital disorders with overlapping patterns of molecular changes affecting imprinted loci. Clin Epigenetics 7:123. https://doi.org/10.1186/s13148-015-0143-8
210. Eggermann T, Davies JH, Tauber M, van den Akker E, Hokken-Koelega A, Johansson G, Netchine I (2021) Growth restriction and genomic imprinting-overlapping phenotypes support the concept of an imprinting network. Genes (Basel) 12(4). https://doi.org/10.3390/genes12040585
211. Weksberg R, Shuman C, Beckwith JB (2010) Beckwith-Wiedemann syndrome. Eur J Hum Genet 18(1):8–14. https://doi.org/10.1038/ejhg.2009.106
212. Wangler MF, Chang AS, Moley KH, Feinberg AP, Debaun MR (2005) Factors associated with preterm delivery in mothers of children with Beckwith-Wiedemann syndrome: a case cohort study from the BWS registry. Am J Med Genet A 134A(2):187–191. https://doi.org/10.1002/ajmg.a.30595
213. Romanelli V, Belinchon A, Campos-Barros A, Heath KE, Garcia-Minaur S, Martinez-Glez V, Palomo R, Mercado G, Gracia R, Lapunzina P (2009) CDKN1C mutations in HELLP/preeclamptic mothers of Beckwith-Wiedemann Syndrome (BWS) patients. Placenta 30(6):551–554. https://doi.org/10.1016/j.placenta.2009.03.013
214. Valente FM, Sparago A, Freschi A, Hill-Harfe K, Maas SM, Frints SGM, Alders M, Pignata L, Franzese M, Angelini C, Carli D, Mussa A, Gazzin A, Gabbarini F, Acurzio B, Ferrero GB, Bliek J, Williams CA, Riccio A, Cerrato F (2019) Transcription alterations of KCNQ1 associated with imprinted methylation defects in the Beckwith-Wiedemann locus. Genet Med 21(8):1808–1820. https://doi.org/10.1038/s41436-018-0416-7
215. Beygo J, Burger J, Strom TM, Kaya S, Buiting K (2019) Disruption of KCNQ1 prevents methylation of the ICR2 and supports the hypothesis that its transcription is necessary for imprint establishment. Eur J Hum Genet 27(6):903–908. https://doi.org/10.1038/s41431-019-0365-x
216. Beygo J, Joksic I, Strom TM, Ludecke HJ, Kolarova J, Siebert R, Mikovic Z, Horsthemke B, Buiting K (2016) A maternal deletion upstream of the imprint control region 2 in 11p15 causes loss of methylation and familial Beckwith-Wiedemann syndrome. Eur J Hum Genet 24(9):1280–1286. https://doi.org/10.1038/ejhg.2016.3
217. Essinger C, Karch S, Moog U, Fekete G, Lengyel A, Pinti E, Eggermann T, Begemann M (2020) Frequency of KCNQ1 variants causing loss of methylation of Imprinting Centre 2 in Beckwith-Wiedemann syndrome. Clin Epigenetics 12(1):63. https://doi.org/10.1186/s13148-020-00856-y
218. Hatada I, Ohashi H, Fukushima Y, Kaneko Y, Inoue M, Komoto Y, Okada A, Ohishi S, Nabetani A, Morisaki H, Nakayama M, Niikawa N, Mukai T (1996) An imprinted gene p57KIP2 is mutated in Beckwith-Wiedemann syndrome. Nat Genet 14(2):171–173. https://doi.org/10.1038/ng1096-171
219. O'Keefe D, Dao D, Zhao L, Sanderson R, Warburton D, Weiss L, Anyane-Yeboa K, Tycko B (1997) Coding mutations in p57KIP2 are present in some cases of Beckwith-Wiedemann syndrome but are rare or absent in Wilms tumors. Am J Hum Genet 61(2):295–303
220. Chang S, Bartolomei MS (2020) Modeling human epigenetic disorders in mice: Beckwith-Wiedemann Syndrome and Silver-Russell Syndrome. Dis Model Mech. https://doi.org/10.1242/dmm.044123
221. Wakeling EL, Brioude F, Lokulo-Sodipe O, O'Connell SM, Salem J, Bliek J, Canton AP, Chrzanowska KH, Davies JH, Dias RP, Dubern B, Elbracht M, Giabicani E, Grimberg A, Gronskov K, Hokken-Koelega AC, Jorge AA, Kagami M, Linglart A, Maghnie M, Mohnike K, Monk D, Moore GE, Murray PG, Ogata T, Petit IO, Russo S, Said E, Toumba M, Tumer Z, Binder G, Eggermann T, Harbison MD, Temple IK, Mackay DJ,

Netchine I (2017) Diagnosis and management of Silver-Russell syndrome: first international consensus statement. Nat Rev 13(2):105–124. https://doi.org/10.1038/nrendo.2016.138
222. Lai KY, Skuse D, Stanhope R, Hindmarsh P (1994) Cognitive abilities associated with the Silver-Russell syndrome. Arch Dis Child 71(6):490–496
223. Noeker M, Wollmann HA (2004) Cognitive development in Silver-Russell syndrome: a sibling-controlled study. Dev Med Child Neurol 46(5):340–346
224. Plotts CA, Livermore CL (2007) Russell-Silver syndrome and nonverbal learning disability: a case study. Appl Neuropsychol 14(2):124–134. https://doi.org/10.1080/09084280701322684
225. Saal HM, Pagon RA, Pepin MG (1985) Reevaluation of Russell-Silver syndrome. J Pediatr 107(5):733–737
226. Russell A (1954) A syndrome of intra-uterine dwarfism recognizable at birth with cranio-facial dysostosis, disproportionately short arms, and other anomalies (5 examples). Proc R Soc Med 47(12):1040–1044
227. Vardi O, Davidovitch M, Vinkler C, Michelson M, Lerman-Sagie T, Lev D (2012) Autistic regression in a child with Silver-Russell syndrome and maternal UPD 7. Eur J Paediatr Neurol 16(1):95–98. https://doi.org/10.1016/j.ejpn.2011.05.009
228. Azcona C, Stanhope R (2005) Hypoglycaemia and Russell-Silver syndrome. J Pediatr Endocrinol Metab 18(7):663–670
229. Blissett J, Harris G, Kirk J (2001) Feeding problems in Silver-Russell syndrome. Dev Med Child Neurol 43(1):39–44
230. Marsaud C, Rossignol S, Tounian P, Netchine I, Dubern B (2015) Prevalence and management of gastrointestinal manifestations in Silver-Russell syndrome. Arch Dis Child 100(4): 353–358. https://doi.org/10.1136/archdischild-2013-305864
231. Netchine I, Rossignol S, Dufourg MN, Azzi S, Rousseau A, Perin L, Houang M, Steunou V, Esteva B, Thibaud N, Demay MC, Danton F, Petriczko E, Bertrand AM, Heinrichs C, Carel JC, Loeuille GA, Pinto G, Jacquemont ML, Gicquel C, Cabrol S, Le Bouc Y (2007) 11p15 imprinting center region 1 loss of methylation is a common and specific cause of typical Russell-Silver syndrome: clinical scoring system and epigenetic-phenotypic correlations. J Clin Endocrinol Metab 92(8):3148–3154. https://doi.org/10.1210/jc.2007-0354
232. Charalambous M, Smith FM, Bennett WR, Crew TE, Mackenzie F, Ward A (2003) Disruption of the imprinted Grb10 gene leads to disproportionate overgrowth by an Igf2-independent mechanism. Proc Natl Acad Sci U S A 100(14):8292–8297
233. Garfield AS, Cowley M, Smith FM, Moorwood K, Stewart-Cox JE, Gilroy K, Baker S, Xia J, Dalley JW, Hurst LD, Wilkinson LS, Isles AR, Ward A (2011) Distinct physiological and behavioural functions for parental alleles of imprinted Grb10. Nature 469(7331):534–538. https://doi.org/10.1038/nature09651
234. Schonherr N, Meyer E, Roos A, Schmidt A, Wollmann HA, Eggermann T (2007) The centromeric 11p15 imprinting centre is also involved in Silver-Russell syndrome. J Med Genet 44(1):59–63. https://doi.org/10.1136/jmg.2006.044370
235. Bonaldi A, Mazzeu JF, Costa SS, Honjo RS, Bertola DR, Albano LM, Furquim IM, Kim CA, Vianna-Morgante AM (2011) Microduplication of the ICR2 domain at chromosome 11p15 and familial Silver-Russell syndrome. Am J Med Genet A 155A(10):2479–2483. https://doi.org/10.1002/ajmg.a.34023
236. Nakashima S, Kato F, Kosho T, Nagasaki K, Kikuchi T, Kagami M, Fukami M, Ogata T (2015) Silver-Russell syndrome without body asymmetry in three patients with duplications of maternally derived chromosome 11p15 involving CDKN1C. J Hum Genet 60(2):91–95. https://doi.org/10.1038/jhg.2014.100
237. Brioude F, Oliver-Petit I, Blaise A, Praz F, Rossignol S, Le Jule M, Thibaud N, Faussat AM, Tauber M, Le Bouc Y, Netchine I (2013) CDKN1C mutation affecting the PCNA-binding domain as a cause of familial Russell Silver syndrome. J Med Genet 50(12):823–830. https://doi.org/10.1136/jmedgenet-2013-101691
238. Kerns SL, Guevara-Aguirre J, Andrew S, Geng J, Guevara C, Guevara-Aguirre M, Guo M, Oddoux C, Shen Y, Zurita A, Rosenfeld RG, Ostrer H, Hwa V, Dauber A (2014) A novel

variant in CDKN1C is associated with intrauterine growth restriction, short stature, and early-adulthood-onset diabetes. J Clin Endocrinol Metab 99(10):E2117–E2122. https://doi.org/10.1210/jc.2014-1949
239. Sabir AH, Ryan G, Mohammed Z, Kirk J, Kiely N, Thyagarajan M, Cole T (2019) Familial Russell-Silver syndrome like phenotype in the PCNA domain of the CDKN1C gene, a further case. Case Rep Genet 2019:1398250. https://doi.org/10.1155/2019/1398250
240. Arboleda VA, Lee H, Parnaik R, Fleming A, Banerjee A, Ferraz-de-Souza B, Delot EC, Rodriguez-Fernandez IA, Braslavsky D, Bergada I, Dell'angelica EC, Nelson SF, Martinez-Agosto JA, Achermann JC, Vilain E (2012) Mutations in the PCNA-binding domain of CDKN1C cause IMAGe syndrome. Nat Genet. https://doi.org/10.1038/ng.2275
241. Cabrera-Salcedo C, Kumar P, Hwa V, Dauber A (2017) IMAGe and related undergrowth syndromes: the complex spectrum of gain-of-function CDKN1C mutations. Pediatr Endocrinol Rev 14(3):289–297. https://doi.org/10.17458/per.vol14.2017.SKHD.imageandrelatedundergrowth
242. Van de Pette M, Tunster SJ, John RM (2018) Loss of imprinting of Cdkn1c protects against age and diet-induced obesity. Int J Mol Sci 19(9). https://doi.org/10.3390/ijms19092734
243. Arima T, Kamikihara T, Hayashida T, Kato K, Inoue T, Shirayoshi Y, Oshimura M, Soejima H, Mukai T, Wake N (2005) ZAC, LIT1 (KCNQ1OT1) and p57KIP2 (CDKN1C) are in an imprinted gene network that may play a role in Beckwith-Wiedemann syndrome. Nucleic Acids Res 33(8):2650–2660
244. Mackay DJ, Eggermann T, Buiting K, Garin I, Netchine I, Linglart A, de Nanclares GP (2015) Multilocus methylation defects in imprinting disorders. Biomol Concepts 6(1):47–57. https://doi.org/10.1515/bmc-2014-0037
245. Eggermann T, Heilsberg AK, Bens S, Siebert R, Beygo J, Buiting K, Begemann M, Soellner L (2014) Additional molecular findings in 11p15-associated imprinting disorders: an urgent need for multi-locus testing. J Mol Med (Berl) 92(7):769–777. https://doi.org/10.1007/s00109-014-1141-6
246. Sanchez-Delgado M, Riccio A, Eggermann T, Maher ER, Lapunzina P, Mackay D, Monk D (2016) Causes and consequences of multi-locus imprinting disturbances in humans. Trends Genet 32(7):444–455. https://doi.org/10.1016/j.tig.2016.05.001
247. Mackay DJ, Callaway JL, Marks SM, White HE, Acerini CL, Boonen SE, Dayanikli P, Firth HV, Goodship JA, Haemers AP, Hahnemann JM, Kordonouri O, Masoud AF, Oestergaard E, Storr J, Ellard S, Hattersley AT, Robinson DO, Temple IK (2008) Hypomethylation of multiple imprinted loci in individuals with transient neonatal diabetes is associated with mutations in ZFP57. Nat Genet 40(8):949–951. https://doi.org/10.1038/ng.187
248. Boonen SE, Hahnemann JM, Mackay D, Tommerup N, Brondum-Nielsen K, Tumer Z, Gronskov K (2012) No evidence for pathogenic variants or maternal effect of ZFP57 as the cause of Beckwith-Wiedemann Syndrome. Eur J Hum Genet 20(1):119–121. https://doi.org/10.1038/ejhg.2011.140
249. Spengler S, Gogiel M, Schonherr N, Binder G, Eggermann T (2009) Screening for genomic variants in ZFP57 in Silver-Russell syndrome patients with 11p15 epimutations. Eur J Med Genet 52(6):415–416. https://doi.org/10.1016/j.ejmg.2009.07.005
250. Meyer E, Lim D, Pasha S, Tee LJ, Rahman F, Yates JR, Woods CG, Reik W, Maher ER (2009) Germline mutation in NLRP2 (NALP2) in a familial imprinting disorder (Beckwith-Wiedemann Syndrome). PLoS Genet 5(3):e1000423. https://doi.org/10.1371/journal.pgen.1000423
251. Docherty LE, Rezwan FI, Poole RL, Turner CL, Kivuva E, Maher ER, Smithson SF, Hamilton-Shield JP, Patalan M, Gizewska M, Peregud-Pogorzelski J, Beygo J, Buiting K, Horsthemke B, Soellner L, Begemann M, Eggermann T, Baple E, Mansour S, Temple IK, Mackay DJ (2015) Mutations in NLRP5 are associated with reproductive wastage and multilocus imprinting disorders in humans. Nat Commun 6:8086. https://doi.org/10.1038/ncomms9086

252. Valero De Bernabe J, Soriano T, Albaladejo R, Juarranz M, Calle ME, Martinez D, Dominguez-Rojas V (2004) Risk factors for low birth weight: a review. Eur J Obstet Gynecol Reprod Biol 116(1):3–15. https://doi.org/10.1016/j.ejogrb.2004.03.007
253. DeChiara TM, Efstratiadis A, Robertson EJ (1990) A growth-deficiency phenotype in heterozygous mice carrying an insulin-like growth factor II gene disrupted by targeting. Nature 345: 78–80
254. Kappil MA, Green BB, Armstrong DA, Sharp AJ, Lambertini L, Marsit CJ, Chen J (2015) Placental expression profile of imprinted genes impacts birth weight. Epigenetics 10(9): 842–849. https://doi.org/10.1080/15592294.2015.1073881
255. Su R, Wang C, Feng H, Lin L, Liu X, Wei Y, Yang H (2016) Alteration in expression and methylation of IGF2/H19 in placenta and umbilical cord blood are associated with macrosomia exposed to intrauterine hyperglycemia. PLoS One 11(2):e0148399. https://doi.org/10.1371/journal.pone.0148399
256. Xing Y, Liu H, Cui Y, Wang X, Tong X (2019) Abundances of placental imprinted genes CDKN1C, PHLDA2 and IGF-2 are related to low birth weight and early catch-up growth in full-term infants born small for gestational age. PLoS One 14(6):e0218278. https://doi.org/10.1371/journal.pone.0218278
257. Demetriou C, Abu-Amero S, Thomas AC, Ishida M, Aggarwal R, Al-Olabi L, Leon LJ, Stafford JL, Syngelaki A, Peebles D, Nicolaides KH, Regan L, Stanier P, Moore GE (2014) Paternally expressed, imprinted insulin-like growth factor-2 in chorionic villi correlates significantly with birth weight. PLoS One 9(1):e85454. https://doi.org/10.1371/journal.pone.0085454
258. Boonen SE, Freschi A, Christensen R, Valente FM, Lildballe DL, Perone L, Palumbo O, Carella M, Uldbjerg N, Sparago A, Riccio A, Cerrato F (2016) Two maternal duplications involving the CDKN1C gene are associated with contrasting growth phenotypes. Clin Epigenetics 8:69. https://doi.org/10.1186/s13148-016-0236-z
259. Xue Y, Shankar S, Cornell K, Dai Z, Wang CE, Rudd MK, Coffee B (2015) Paternal duplication of the 11p15 centromeric imprinting control region is associated with increased expression of CDKN1C in a child with Russell-Silver syndrome. Am J Med Genet A 167A (12):3229–3233. https://doi.org/10.1002/ajmg.a.37371
260. López-Abad M, Iglesias-Platas I, Monk D (2016) Epigenetic characterization of CDKN1C in placenta samples from non-syndromic intrauterine growth restriction. Front Genet 7:62. https://doi.org/10.3389/fgene.2016.00062
261. Jensen AB, Tunster SJ, John RM (2014) The significance of elevated placental PHLDA2 in human growth restricted pregnancies. Placenta 35(8):528–532. https://doi.org/10.1016/j.placenta.2014.04.018
262. McMinn J, Wei M, Schupf N, Cusmai J, Johnson EB, Smith AC, Weksberg R, Thaker HM, Tycko B (2006) Unbalanced placental expression of imprinted genes in human intrauterine growth restriction. Placenta 27(6–7):540–549
263. Petry CJ, Evans ML, Wingate DL, Ong KK, Reik W, Constancia M, Dunger DB (2010) Raised late pregnancy glucose concentrations in mice carrying pups with targeted disruption of H19delta13. Diabetes 59(1):282–286. https://doi.org/10.2337/db09-0757
264. Janssen AB, Capron LE, O'Donnell K, Tunster SJ, Ramchandani PG, Heazell AE, Glover V, John RM (2016) Maternal prenatal depression is associated with decreased placental expression of the imprinted gene PEG3. Psychol Med 46(14):2999–3011. https://doi.org/10.1017/S0033291716001598
265. Patten MM, Cowley M, Oakey RJ, Feil R (2016) Regulatory links between imprinted genes: evolutionary predictions and consequences. Proc Biol Sci 283(1824). https://doi.org/10.1098/rspb.2015.2760
266. Varrault A, Gueydan C, Delalbre A, Bellmann A, Houssami S, Aknin C, Severac D, Chotard L, Kahli M, Le Digarcher A, Pavlidis P, Journot L (2006) Zac1 regulates an imprinted gene network critically involved in the control of embryonic growth. Dev Cell 11(5):711–722. https://doi.org/10.1016/j.devcel.2006.09.003

267. Gabory A, Ripoche MA, Le Digarcher A, Watrin F, Ziyyat A, Forne T, Jammes H, Ainscough JF, Surani MA, Journot L, Dandolo L (2009) H19 acts as a trans regulator of the imprinted gene network controlling growth in mice. Development 136(20):3413–3421. https://doi.org/10.1242/dev.036061
268. Dalgaard K, Landgraf K, Heyne S, Lempradl A, Longinotto J, Gossens K, Ruf M, Orthofer M, Strogantsev R, Selvaraj M, Lu TT, Casas E, Teperino R, Surani MA, Zvetkova I, Rimmington D, Tung YC, Lam B, Larder R, Yeo GS, O'Rahilly S, Vavouri T, Whitelaw E, Penninger JM, Jenuwein T, Cheung CL, Ferguson-Smith AC, Coll AP, Korner A, Pospisilik JA (2016) Trim28 haploinsufficiency triggers bi-stable epigenetic obesity. Cell 164(3):353–364. https://doi.org/10.1016/j.cell.2015.12.025
269. Barker DJ (1990) The fetal and infant origins of adult disease. BMJ 301(6761):1111
270. Gluckman PDaH M (2006) Developmental origins of health and disease, 1st edn. Cambridge University Press, Cambridge
271. Ivanova E, Chen JH, Segonds-Pichon A, Ozanne SE, Kelsey G (2012) DNA methylation at differentially methylated regions of imprinted genes is resistant to developmental programming by maternal nutrition. Epigenetics 7(10):1200–1210. https://doi.org/10.4161/epi.22141
272. Radford EJ, Isganaitis E, Jimenez-Chillaron J, Schroeder J, Molla M, Andrews S, Didier N, Charalambous M, McEwen K, Marazzi G, Sassoon D, Patti ME, Ferguson-Smith AC (2012) An unbiased assessment of the role of imprinted genes in an intergenerational model of developmental programming. PLoS Genet 8(4):e1002605. https://doi.org/10.1371/journal.pgen.1002605

The Utility of Twins for Epigenetic Analysis

9

Richard Saffery and Jordana T. Bell

Contents

Abstract

Twin studies have played an important role in our understanding of individual variation for over a century. The strength of these lies in the capacity to almost perfectly control for inter-individual genetic variation through the study of

R. Saffery (✉)
Molecular Immunity, Murdoch Children's Research Institute (MCRI), Parkville, VIC, Australia

Department of Paediatrics, The University of Melbourne, Parkville, VIC, Australia
e-mail: richard.saffery@mcri.edu.au

J. T. Bell
Department of Twin Research and Genetic Epidemiology, Kings College London, London, UK
e-mail: jordana.bell@kcl.ac.uk

K. B. Michels (ed.), *Epigenetic Epidemiology*,
https://doi.org/10.1007/978-3-030-94475-9_9

monozygotic (MZ) twin pairs. Despite their genetic identity, MZ twins usually show phenotypic variability, often solely ascribed to non-shared environmental exposures. Given that epigenetic mechanisms are widely believed to be the mediators of the influence of environmental factors on the underlying genome, it is not surprising that study of twins in epigenetic research is an important approach to help unravel the complexities associated with gene: environment interactions in human development and disease. In addition, the strategic use of twins in epigenetic studies has revealed the importance of genetic factors and both in utero and postnatal environments to the establishment and maintenance of the human epigenome. However, a note of caution is warranted given emerging evidence for epigenetic variation as an inherent feature of MZ twinning and the potential for MZ twins to show genetic variability. Irrespective of this, twin studies are beginning to reveal evidence linking epigenetic disruption to disease-associated risk in humans.

Abbreviations

2D	Two dimensional
3′UTR	Non-protein coding 3′ untranslated region of RNA transcript
5MeC	5-methylcytosine
AIMS	Amplification of inter-methylated sites
ANOVA	Analysis of variance
CBMC	Cord blood mononuclear cells
CGI	CpG island region (enriched for CpG dinucleotides)
COMT	Catechol-O-methyltransferase
DC	Dichorionic (two placentas)
DM	Dermatomyositis
DMR	Differentially methylated regions
DNA	Deoxyribonucleic acid
DRD2	Dopaminergic receptor D2
DZ	Dizygotic
EEA	Equal environments assumption
HUVEC	Human umbilical vein endothelial cells
ICC	Intraclass correlation
MC	Monochorionic
MSRDA	Methylation-sensitive representational difference analysis
MZ	Monozygotic
PPIEL	Peptidylprolyl isomerase E-like
RA	Rheumatoid arthritis
RNA	Ribonucleic acid
RRBS	Reduced representation bisulphite sequencing
SLE	Systemic lupus erythematosus
SMS	Spermine synthase

9.1 Why Study Epigenetics in Twins?

Twin studies have played an important role in understanding the determinants of phenotypic variation for over a century. The classical definition of phenotype as the 'physical manifestation of genotype' has expanded over time to include not only phenotype at the macro level (e.g appearance, health status), but also the micro/molecular level, as encompassed by a suite of cellular and molecular variation. This includes variation in gene expression, itself subject to regulation by genetic and environmental influence and ultimately universally regulated by epigenetic factors.

The similarity of genetically 'identical' Monozygotic (MZ) twins as a group relative to non-identical dizygotic (DZ) twins has traditionally allowed an estimation of heritability (the proportion of the total variance in a trait attributable to genetic factors). MZ twins are generally considered to be genetically identical, while DZ twins share on average 50% of their genetic (DNA sequence) variation (0–100% theoretical range). MZ twins exhibit a wide range of concordance rates for any phenotype, including disease. A MZ phenotypic concordance approaching 100% likely indicates the coinheritance of highly dominant/penetrant genetic variants. However, most diseases or traits show MZ concordance rates far lower than this, indicating a modifying role of environmental and/or stochastic factors in regulating phenotype in association with underlying genotypes.

Given that epigenetic mechanisms are widely believed to be the mediators of the influence of environmental factors on the underlying genome, it is not surprising that the study of twins in epigenetic research is becoming increasingly popular. However, studies to date in this field have relied on the assumption of genetic identity of MZ twin pairs and the equivalent responsiveness of embryonic epigenetic variation between MZ and DZ twin pairs. Each of these assumptions has been directly challenged by observational data and thus caution is warranted when applying classical twin approaches to understand the determinants of epigenetic variation and/or its link to phenotype in humans.

9.2 Twin Studies and the Regulation of Epigenetic Profile

9.2.1 Gene Expression Differences in Twins

Previous studies examining gene expression in MZ twins have identified many hundreds of genes whose within-pair expression discordance [the absolute value of the ratio of gene expression of twin 1: twin 2] is greater than an arbitrarily defined threshold (set above the level of experimental 'noise') [1, 2]. This supports divergence in gene expression between MZ twins in response to environmental and stochastic factors, as the driver of subsequent phenotypic divergence. In accordance with expectations, studies have also demonstrated a higher degree of similarity between specific gene expression patterns in MZ than in DZ pairs, highlighting the contribution of genetic factors to overall expression profile [1, 3]. This relation equally applies to newborn twin pairs [4]. However, MZ pairs also show clear discordance in gene expression profiles at birth, highlighting the importance of the

non-shared in utero environment in determining gene expression profile [5]. Given the unequivocal role of epigenetic variation in regulating gene expression, it is clear that such discordance in MZ twins is likely to have an epigenetic origin.

9.2.2 Contributors to Epigenetic Variation

Elegant studies in mice have demonstrated that almost any phenotype measured in genetically identical (inbred) animals reared in essentially identical environments shows a normal distribution, highlighting the role of probabilistic (stochastic) factors in phenotype [6]. This is also true to DNA methylation levels at specific loci. A similar pattern of distribution is also apparent in relation to epigenetic marks assessed in any reasonably sized groups of humans. An exception to this is the highly skewed distribution seen in many cancers, relatively rare imprinting disorders, or so-called 'metastable epialleles'. However, unlike inbred mice, the relative contributions of genetic, environmental and stochastic factors to the establishment and maintenance of epigenetic profile in outbred humans, and its relative stability over time, remain poorly understood. What is becoming increasingly clear is that all four factors (genes, environment, stochastic and age/time) play some role in both the levels of epigenetic variation and resulting phenotypes more generally. The strategic study of twins provides a unique opportunity to circumvent some of the complexities associated with dissecting the relative contributions of these factors to a range of complex health and disease phenotypes in humans or to molecular variation such as epigenetic profile.

9.2.3 The Importance of Genetic (Heritable) Factors in Regulating the Epigenetic Profile

MZ twins share maternal, obstetric and genetic factors at birth, but differences in epigenetic profile between MZ twins within a pair may still accumulate in response to differences in environmental exposures and stochastic factors, accumulated both in utero and postnatally. In contrast DZ pairs also share maternal and obstetric factors but the likelihood of sharing specific genetic variation is the same as that of non-twin siblings. Given the large number of genes and/or proteins implicated in the establishment and maintenance of the epigenetic profile, and the demonstrated link between genetic variation and epigenetic profile [7, 8], it is reasonable to speculate that DZ twins as a group will be more epigenetically divergent than MZ twins due to inherent genetic differences.

A powerful method for estimating the relative contribution of heritable and environmental/stochastic influences on variation in any quantitative trait such as DNA methylation is to compare the degree of discordance in MZ pairs as a group relative to that of DZ twin pairs. Both MZ and DZ twins are required to estimate the relative contributions of genetic vs. common environmental contributors. Age differences between groups of MZ and DZ twin pairs may be a potential confounding factor specific to epigenetic analyses [9, 10] and, as such, care must

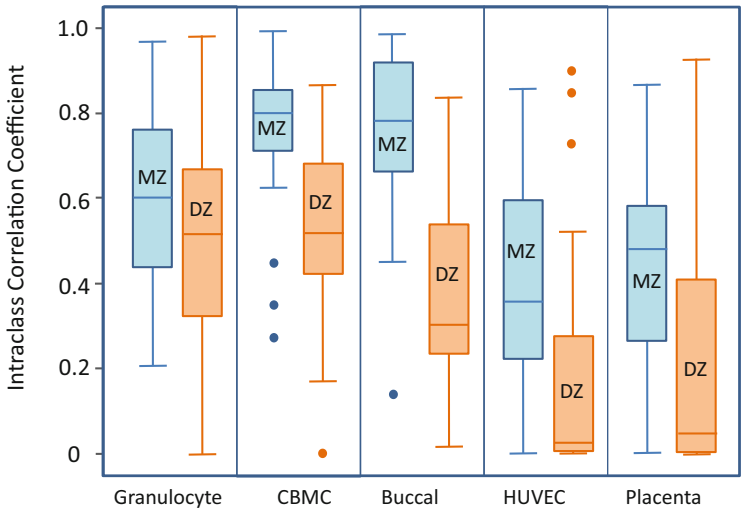

Fig. 9.1 Distribution of intraclass correlations reveals a higher median correlation coefficient for DNA methylation at the IGF2/H19 locus in MZ than DZ twin pairs in five tissues from newborns. The ICC measures the proportion of total variance attributable to within pair variation within MZ and DZ groups of twins. ICC analysis. CBMC—cord blood mononuclear cells; HUVEC—human umbilical vein endothelial cells. Figure adapted from [11]

be taken to match as closely as possible in this regard. Additionally, the contribution of stochastic factors to epigenetic divergence is equally assumed to be constant across populations in these analyses. Nevertheless, by treating epigenetic variation, such as DNA methylation, as a quantitative phenotype, the classical twin model can be applied to estimate the relative genetic and environmental components to variation in epigenetic profile. For example the variance in DNA methylation within twins at a specific site is due to a combination of genetic make-up (A), shared environmental influences (C) and residual (including stochastic and non-shared) environmental factors (E), that in combination form the basis of the ACE model (Fig. 9.1) that can be used to calculate A, C and E components while also allowing adjusting for a range of covariates such as age. Various derivative models that account for only some of the three components can also be tested (AC, E, CE) to identify the model which best fits the observed data. Alternatives to ACE include the ADE model, whereby D stands for a dominant genetic effect [12]. Such studies can provide an estimate of the proportion of epigenetic variance at a locus that can be attributed to genetic variation and also the proportion of epigenetic variance that can be attributed to shared environment [13].

Heritability is an estimate of the proportion of the total phenotypic (outcome) variance in a population that is attributable to genetic effects. In twin studies, heritability is estimated by comparing the degree of phenotypic similarities between groups of MZ and DZ twins, either as the concordance rate or intraclass correlation. In the context of twins, the additive heritability (h^2) is determined as twice the difference between MZ and DZ concordance rates (correlation; $h^2 = 2(r_{MZ} - r_{DZ})$,

where r is the concordance or intraclass correlation between each type of twin [14]. Heritability estimates are population specific unless the environment is constant.

Early studies involving twins estimated the heritability of DNA methylation profiles both at specific genes [11, 15] and across the genome [4, 16]. The average heritability of methylation across all sites yet examined appears in the range of 0.1–0.3, supported by non-twin family studies [17].

A strength of the twin study design is the lack of confounding by age effects between individuals within a twin pair. Despite this, epigenetic heritability estimates will invariably be cell, tissue, time and locus specific. For example whereas certain regions of the genome show strong evidence of being under genetic control, others do not. Heritability estimates of DNA methylation at specific sites vary widely from 0 to 0.99. Such wide variation in the influence of genetic variation in DNA methylation is now fully supported by a host of singleton studies assessing the widespread genetic influence on DNA methylation profile [18, 19].

Classic studies in twins are predicated on the equal environments assumption (EEA) that no interaction exists between genes and environment [i.e. no twin-type (DZ vs. MZ) specific environmental differences]. If this is not the case, then the classical model may lack utility in that genetic factors may not be the sole driver of cumulative MZ vs. DZ disparity. Furthermore, MZ and DZ twins as groups may be inherently epigenetically different from early embryogenesis, either through the very early influence of genetic variation and environment in DZ relative to MZ twins or in association with the process of twinning itself. This is supported by recent findings of inherent epigenetic variation as a hallmark of MZ twinning [20]. The implications of such observations challenge the assumption that classical twin studies as suitable models for dissecting the variance components of epigenetic variation.

9.2.4 Twin Studies Reveal Cumulative Environmental Contributors to 'Epigenetic Drift' Over Time

The cumulative effects of environmental and stochastic variation on changing epigenetic profile were first illustrated by a widely cited study that examined both genome wide and locus-specific DNA methylation variation in a small number of young and middle-aged MZ twins [10]. Whereas 3-year-old MZ twins showed relatively few epigenetic differences within pairs, those aged 50 years showed considerable variability within pairs, and this was greater if the twins had divergent lifestyles. A multi-level statistical analysis was performed that generated a single descriptive value for each type of epigenetic measure. Using this approach, the general conclusion was that epigenetic profile is in constant 'drift' from early in life, although very few young twins were studied and statistical precision was low [10]. In addition, this was a cross-sectional study and did not evaluate methylation in the same individuals over time. Finally, the specific epigenetic analyses employed were all low resolution in that they did not examine the distribution of epigenetic differences within the genome, but rather provided a global 'snapshot' of different classes of epigenetic disparity within twin pairs.

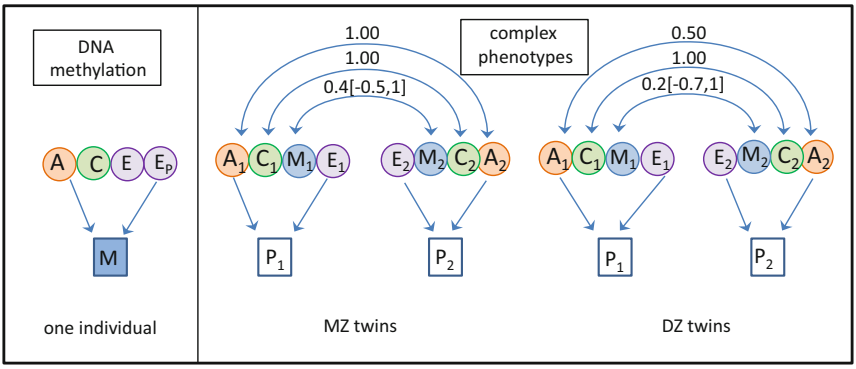

Fig. 9.2 The relation between epigenetic and phenotypic heritability. Proposed contribution of latent variables to the methylation status of an individual at a genomic region (M) and to their phenotype (P). The left pane shows variables contributing to DNA methylation status at one genomic region in one individual; effects will be specific to age, sex, population (genetic factors) and tissue sampled, and will also include stochastic factors. Methylation latent factors include additive genetic factors (A), common environmental factors (C), unique environment (E), and heritable and stable epigenetic factors that are not DNA sequence dependent (Ep). The right pane represents the path model in twins, depicting the contribution of DNA methylation and other factors to the phenotype (P) in twin *i* with correlation estimates in MZ (left) and DZ (right) twins of latent variables including additive genetic effects (Ai), common environment (Ci), DNA methylation (Mi), and unique environment (Ei). Correlation estimates were obtained from previous genetic [22] and epigenetic studies [16] in twins. In siblings, the correlation in M will be lower than that observed in DZ twins due to age differences and associated increased level of cumulative stochastic change. Figure adapted from [14]

More recent studies have also examined epigenetic drift longitudinally in twins, for example in young twins [21]. Analysis focused on buccal cell DNA methylation in 3 genes in 46 MZ and 45 DZ at 5 and 10 years of age. Longitudinal change within individuals was calculated by assessing the correlation in methylation at each age. To assess the relative contributions of heritable and environmental/stochastic components to methylation levels, correlations within MZ pairs were compared to correlations within DZ pairs. Finally, to assess the relative contributions of heritable and environmental/stochastic components to changes in DNA methylation over time, intra-individual change scores were calculated and correlations within MZ twin pairs were compared to correlations within DZ pairs. This study was the first to reveal the extent of epigenetic discordance in MZ twins in early life and highlighted the ongoing instability of methylation levels over time. Importantly different genomic regions were found to show varying levels of epigenetic divergence over time.

Another examination of DNA methylation levels at several sites in multiple tissues from newborn twins confirmed that epigenetic drift between genetically identical individuals (MZ twins) begins in utero and in a tissue-specific manner. In this study, intraclass correlation coefficients (ICC) were higher in MZ than in DZ twins [11] (Fig. 9.2), supporting previous findings of a role for genetic/heritable factors in the establishment of epigenetic profile. Interestingly, a subsequent study of a different tissue (buccal epithelial cells) from some of the same MZ twins found

evidence of both epigenetic drift and epigenetic convergence in different sets of twins from birth to 18 months of age [23]. Similar differences in the level of within pair methylation variation over a relatively short period of time have similarly been described in MZ adolescent twins [24].

Many non-twin studies have now described epigenetic variation (drift) in association with ageing. Often described as the 'epigenetic clocks', it has become clear that a subset of epigenetic variation is directly attributable to ageing and has potential as a marker of 'biological' (as opposed to chronological) age [25]. Epigenetic clocks have now been generated across a variety of tissue types and developmental stages. For example gestational age-associated epigenetic clocks [26, 27] may be sensitive to specific pregnancy exposures/risk factors or predict subsequent offspring characteristics [28]. The relative contribution of genetic, environmental and stochastic factors to epigenetic variation associated with 'drift' and/or 'epigenetic ageing' remains to be fully elaborated, but is clearly itself subject to genetic influence. A combined genetic/DNA methylation analysis of more than 40,000 individuals identified 137 loci implicated in the biological ageing captured by four epigenetic clocks [29].

Numerous twin studies have attempted to disentangle the relative contribution of cumulative genetic and/or environmental influence to age-associated epigenetic variation (as assessed by various DNA methylation 'clocks'). For example study of ~100 female twins identified a large number of sites showing heritable DNA methylation ageing across two different timepoints, with decreasing heritability apparent over 10 years (23.8% to 18.0%). Previously identified age-sensitive sites showed strong genetic contribution, while other sites of variation were more environmentally sensitive [30]. Another study of 104 Swedish and Danish twin pairs estimated the genetic and environmental influence on two epigenetic 'clocks' at the mean age of 70 (baseline) and 79 years (follow-up). Individual specific environmental influences were found to influence epigenetic age over the two timepoints examined, even at this advanced age [31]. Further, a study of 413 pairs of Finnish twins (age 63–76) tested the influence of genetic and lifestyle factors to accelerated ageing (assessed by GrimAge clock) and mortality. The findings confirmed the strong predictive value of the GrimAge with mortality and further that this predictive value was independent of genetic influence [32]. In one of the largest study of its kind studying 4217 individuals, including twins (0–92 years of age), variation in DNAm age is was mostly attributed to environmental factors, including those shared within twins/families [33]. This suggested the equal environment assumption of the classic twin study may not hold in studies exploring epigenetic ageing.

9.3 Epigenetics and the Discordant Twin Model

9.3.1 The Assumption of Genetic Identity

Many studies have reported phenotypic differences between MZ co-twins (reviewed in [34]). Discordant MZ pairs have proved very valuable in revealing the contribution of non-genetic variation to disease penetrance, etiology or effect of therapies. The underlying assumption in all such studies is that MZ twins are genetically

identical. This has been directly disproved in many instances, where either specific point mutations [35–37], uniparental disomy [38], triplet repeat expansion [39, 40], chromosomal mosaicism [41], heteroplasmy for mitochondrial-encoded mutations [42, 43] or chromosomal aneuploidies (reviewed in [34]), have been linked to specific phenotypic differences in pairs of MZ twins. Importantly, differences in both copy number [44] and telomere length [45, 46] have been also described in phenotypically discordant and concordant MZ twin pairs.

In the most sensitive analysis of its kind, an in-depth examination of the DNA sequence of 381 MZ twin pairs identified on average 5.2 early developmental mutations within pairs. Of these, mutations were specific to one individual in approximately 15% of pairs. Overall, this revealed the potential different cellular origins of some MZ co-twins within the very early embryo, highlighting the need to consider the potential for genomic discordance in MZ pairs when analyzing complex traits [47].

9.3.2 Inferring Causation: Epigenetic Vs. Genetic Analysis

Since every cell in an individual derives from a single zygote with one genome, genetic studies generally require only a single DNA sample per individual, taken from any tissue at any age. Cause and effect can be reliably predicted or excluded, as genomic variation is not considered to be variable over the life course (except in the case of relatively 'rare' somatic mutation). Such rules may not apply to most epigenetic studies where repeated samples from more than one tissue may be desirable. Unless biospecimens are collected very early in life (prior to phenotypic manifestation), unravelling cause and effect in human epigenetic studies is problematic. Despite this and the many other caveats associated with inferring causation in any epigenetic association study (covered in other chapters of this book), several investigators have 'taken the plunge' and attempted to link specific epigenetic changes to disease phenotypes.

9.3.3 Localized Epigenetic Variation in Phenotypically Discordant Twins

Several studies have attempted to characterize the epigenetic contribution to a range of complex human phenotypes through a variety of variable resolution approaches applied to a wide range of, generally low to modest, sample sizes. Given its inherent stability as a covalent modification, and relative ease of simultaneously mapping DNA methylation profile across large regions of the genome relative to other order epigenetic marks, it is not surprising that twin studies carried out in this space have generally focused on DNA methylation variation in phenotypically discordant MZ twins (examples in Table 9.1). The advent of the Illumina Infinium Human Methylation arrays (27 K, 450 k and now 850 K EPIC) has revolutionized this analysis by allowing robust measurements of up to over 850,000 individual sites of DNA methylation within any human samples for which genomic DNA can be obtained.

Table 9.1 Examples of epigenetic analysis in discordant MZ twin pairs

Focus	Number of twin pairs	Biological samples	Epigenetic target/Methodology	Major finding	Ref.
Schizophrenia (SCZ)	1 discordant, 1 concordant SCZ twin pair	Blood	DNA methylation of *DRD2* gene by Bisulphite sequencing	Discordant twin with schizophrenia shows more similar *DRD2* methylation profile to affected concordant twins than to unaffected co-twin	[48]
Birth weight	12 highly discordant MZ pairs	Buccal	DNA methylation at 2 CpG sites of *COMT* gene by bisulphite pyrosequencing	Highly variable methylation concordance rates between pairs. Average within pair methylation discordance was 10.3% at site 1 and 16.1% at site 2. Strong correlation in differences between both sites ($r = 0.87$; $p < 0.001$). No correlation between birth weight and methylation level	[49]
Primary biliary cirrhosis (PBC)	4 discordant MZ, 1 concordant affected	Blood	DNA methylation by bisulphite sequencing	Decreased expression of *CLIC2* and *PIN4* in ¾ PBC affected individuals in discordant pairs. No evidence for DNA methylation changes driving expression.	[50]
Beckwith Wiedemann syndrome (BWS)	10 discordant MZ 5 control MZ	Skin fibroblasts and/or peripheral blood	DNA methylation by Southern blotting	BWS affected individuals show loss of imprinting at *KVDMR1* with biallelic expression of *KCNQ1OT1*	[51]
Silver-Russell syndrome (SRS)	1 discordant MZ pair	Blood	COBRA and bisulfite sequencing for two H19 DMRs	Loss of *H19*-DMR in roughly half of cells in the affected twin only with associated decreased expression of *IGF2*	[52]

Caudal duplication Anomaly (CDA)	1 discordant MZ pair 9 unaffected MZ pairs	PBMC	DNA methylation of *AXIN-1* by Bisulphite sequencing	Higher methylation at *AXIN-1* promoter in affected vs. unaffected co-twin ($p < 0.0001$). Higher methylation in both twins than in controls ($p = 0.02$). ICC correlation of 0.76 for all 10 MZ pairs.	[53]
Alzheimer's disease (AD)	1 discordant MZ	Post mortem temporal neocortex	Various epigenetic markers by immunohistochemistry	Significantly reduced levels of DNA methylation in temporal neocortex neuronal nuclei in AD	[54]
Body mass index (BMI)	16 discordant MZ pairs	Saliva	DNA methylation at 9 regions implicated in growth, by bisulphite based, amplification and primer extension and HPLC.	Only small intrapair differences in methylation observed. No significant correlations between intrapair BMI differences and methylation levels.	[55]
1. Systemic lupus Erythematosus (SLE) **2**. Rheumatoid arthritis (RA) **3**. Dermatomyositis (DM)	5 discordant pairs for each disorder	Blood	Golden Gate arrays Validation by Bisulphite sequencing Global 5MeC	No methylation differences associated with DM and RA. Consistent methylation changes in SLE twins relative to unaffected co-twins in immune functioning genes. Global decrease in 5-methylcytosine in SLE and hypomethylation of 28S and 18S *rDNA* genes	[56]

(continued)

Table 9.1 (continued)

Focus	Number of twin pairs	Biological samples	Epigenetic target/Methodology	Major finding	Ref.
Multiple sclerosis (MS)	3 discordant MZ pairs	CD4+ T cells	RRBS (50–90 million reads covering >2 million CpG sites)	Between 2 and 176 significant methylation differences (of 2 million CpGs tested) between MZ co-twins (> 800 methylation differences in unrelated individuals). No evidence for genetic, epigenetic or transcriptome differences underlying MS discordance.	[57]
Bipolar disorder (BPD)	1 discordant MZ pair, 16 unrelated singletons with BPD. Independent set of 14 unrelated BPD singletons	Lymphoblastoid cell lines	MSRDA with validation by bisulphite DNA sequencing and pyrosequencing	10 genomic regions showing differential methylation between co-twins by MSRDA. 4 of these confirmed by bisulphite sequencing of co-twin DNA. 2 of these (*PPIEL*, *SMS*) confirmed as significant by pyrosequencing in case-control study design (25 BPD cases and 18 controls) with $p < 0.05$.	[58]
Risk taking behaviour	1 discordant plus 9 MZ control pairs	Blood	>12,000 CpG island fragments	No technical comparisons survived FDR correction. Differential methylation of *DLX1* gene specifically in discordant MZ pair	[59]
T2 diabetes	17 T2D discordant, 3 T2D concordant, 7 healthy MZ twin pairs	Blood	MeDIP-seq 450 K Infinium array	Single DMR following FDR correction (<5%), 31 DMRs (<10%). 12/20 DMRs tested validated by 450 K, considerable replication in an independent T2D sample	[60]

Idiopathic scoliosis	6 discordant MZ pairs, 1 concordant MZ	Blood	850 K Infinium EPIC array	Four DMPs (FDR <0.1) and many DMRs associated with curve severity	[61]
Depression	58 score-dependent discordant MZ twin pairs	Blood	RRBS.	66 DMPs associated with depression score reached the level of $P < 1 \times 10^{-4}$. Evidence of common biological functions.	[62]
T1 diabetes	52 discordant MZ pairs	Purified blood CD4+ T, CD19+ B cells, CD14 + CD16- monocytes		Single T1D associated DMP in T cells only in DDIT4 gene. 10,548 DVPs in B cells, 4314 in T cells and 6508 in monocytes at a stringent FDR of <0.001 enriched in T1D twins	[63]
T2 diabetes	14 discordant MZ twin pairs	Adipose tissue	450 K Infinium array	Many thousands of DMPs ($p < 0.05$).Significant associations between 20 DMPs, and 11 DEGs and T2D. Some replication of previous findings	[64]
IgA nephropathy	2 discordant MZ pairs	PBMC	850 K Infinium EPIC array	521 DMRs were detected for 2 IgAN-discordant MZ twins. A total of 521 DMRs detected across 2 pairs. 9 DMRs mapped to DEGs in MZ twins	[65]
Non-syndromic cleft lip/palate	6 discordant MZ twin pairs	Saliva	30X coverage WGBS	151 DMPs in total spanning 147 genes from all 6 discordant pairs combined. Smaller number of DMRs with little overlap between twin pairs.	[66]

(continued)

Table 9.1 (continued)

Focus	Number of twin pairs	Biological samples	Epigenetic target/Methodology	Major finding	Ref.
Teratology of Fallot	2 discordant MZ pairs	Blood	21X coverage WGBS	>300,000 DMPs with >25% difference within each pair. Differential methylation of *NFATC1* promoter in both pairs	[67]
Major depression (MD)	27 discordant, 42 concordant −ve, 6 concordant +ve MZ twin pairs, aged 15–20 years	Blood	450 K Infinium array	17 DMRs associated with MD (all hypermethylated). 10 VMRs more variable in MD cases. Enriched for neurodevelopment genes, including those previously implicated in mood and psychiatric disorders	[68]
Major depressive disorder (MDD)	79 discordant MZ twin pairs	Purified blood monoctyes	850 K Infinium EPIC array	39 DMRs associated with MDD (39 hyper and 6 hypomethylated)	[69]
Autism spectrum disorder	5 discordant MZ twin pairs	Blood	450 K Infinium array (n = 3); RRBS (n = 2)	2,397 differentially methylated genes in ASD enriched for 27 gene pathways including neurotrophin signaling	[70]
Focal/generalised epilepsy	15 discordant MZ twin pairs total	Blood Buccal	850 K Infinium EPIC array	No significant associations at FDR correction. Some evidence of DMRs in both buccal and blood for both focal and generalised epilepsy but no overlap between epilepsy types or tissues	[71]

Multiple sclerosis	45 discordant MZ twin pairs	PBMC	850 K Infinium EPIC array, validation by TDBS	No significant associations at FDR correction. Some evidence for 39 MS-associated DMPs ($p < 5 \times 10E\text{-}6$). Validation of TMEM232 and ZBTB16 DMPs by TDBS	[72]
Rheumatoid arthritis	79 discordant MZ twin pairs	Blood	450 K Infinium array	No significant associations at FDR correction. 763 hypervariable DVP associated with RA with FDR correction ($p < 0.001$)	[73]
Childhood persistent asthma	37 discordant MZ twin pairs	Buccal	450 K Infinium array	No significant associations following Bonferroni correction. Some evidence for consistent DMPs across the majority of discordant twin pairs. Top nominally associated asthma DMP in *HGSNAT* gene	[74]
BMI	30 discordant MZ twin pairs	Blood	450 K Infinium array	No significant associations with BMI at FDR correction. However, >1000 DMPs apparent if BMI linked to high liver fat and metabolic changes.	[75]

Note: * transcription factor involved in the formation of GABAergic interneurons [76]

Abbreviations: 3'UTR—non-protein coding 3′ untranslated region of RNA transcript; DMP—differentially methylate probe; DMR—differentially methylated regions; VMR—variably methylated region; DVP—differentially variable positions; DEG—differentially expressed gene; ICC—intraclass correlation; DRD2—dopaminergic receptor D2; COMT—catechol-O-methyltransferase; COBRA—Combined bisulphite restriction analysis; RRBS—reduced representation bisulphite sequencing; MSRDA—methylation-sensitive representational difference analysis; MeDIP-seq—methylation dependent immunoprecipitation and sequencing; GSNAP [77]; PPIEL—Peptidylprolyl isomerase E-like; PBMC—peripheral blood mononuclear cells; SMS—Spermine synthase; 850 K, 450 K—Infinium Human Methylation Beadchips; WGBS—whole genome bisulphite sequencing; TDBS—targeted deep bisulphite sequencing

The most commonly approached twin-based study design involves associating DNA methylation variation to phenotypic discordance in MZ twin pairs. Many studies have now explored the link in a wide range of phenotypes (Table 9.1). However to date, utilizing the phenotype discordant MZ twin pair model has played a relatively limited role in identifying epigenetic modifications associated with specific human phenotypes, likely due to limited sample sizes, despite the advantages afforded by largely controlling for underlying genetic variation. Compelling findings are few with limited biological plausibility and a lack of independent replication. As such, the biological and clinical relevance of many such findings remains largely unclear.

9.3.4 Skewed X-Chromosome Inactivation in Phenotypically Discordant Twins

One of the most widely studied epigenetic phenomena in mammals involves the inactivation of the majority of genes on one X-chromosome in females. This 'dosage compensation' equalizes the expression of most X-chromosome genes in males (XY) and females (XX) and usually occurs in a random manner, with roughly 50% of cells in a female showing inactivation of the maternally-derived X, whereas the other 50% show inactivation of the paternally-derived X-chromosome. However, in some individuals, a skewed pattern of X-inactivation is apparent, with one parental X over represented in a particular tissue or cell type. In such instances, an otherwise recessive mutation can have profound adverse phenotypic effects. Thus, skewed X-inactivation, or more precisely the chance inactivation of a single functional allele in a physiologically-relevant tissue, has been associated with discordance for several disorders including hemophilia [78], Fragile-X syndrome [40, 79] and Duchenne Muscular Dystrophy [80].

9.3.5 Ultra-High Resolution of Genetic and Epigenetic Variation in Twins

The future of both genomic and epigenomic analysis of twins lies in the use of high throughput nucleotide sequencing. Baranzini et al. [57] were amongst the first to adopt this approach, searching for genetic, expression and DNA methylation differences in purified CD4+ T cells in a small number of MZ twin pairs discordant for multiple sclerosis (MS). Between 50 and 68 million different sequencing reads of messenger RNAs were counted for each pair, broadly speaking as the expression 'output' in each sample. A diagnosis of MS accounted for only 9.4% of the total variance in gene expression, with no robust differences ascribed to the MS phenotype in isolation. An examination of 50–90 million high-quality reads of reduced representation bisulphite sequencing (RRBS) methylation data from CD4+ T cells of the three discordant twins similarly showed little evidence of DNA methylation changes that could be specifically associated with the MS phenotype. Unfortunately, there was no independent validation of limited observed within pair differences

using an alternative methodology, so the robustness of the few observed DNA methylation changes within MZ pairs remains unclear [57].

9.4 Future Directions

There may be epigenetic effects specific to twins that limit their utility as a model to reveal contributors to disease more broadly applicable in the wider singleton population. For example recent data suggest the processes of MZ twinning itself involves a distinct subset of epigenetic variation, stable throughout life [20], that may potentially underpin the embryo splitting process [81]. If specific to MZ twins, this would have the effect of confounding any MZ vs. DZ comparisons aimed at identifying components of epigenetic variation as several key assumptions of equivalence would be violated. Further, it remains to be determined if this also results in differential epigenomes of two newly created MZ embryos independently of later onset of phenotypic discordance.

Nearly all considerations associated with epigenetic analyses generally also apply to studies examining twins. This includes issues of tissue/cell specificity, variation over time and the potential for confounding or reverse causation. The primary advantage afforded by twin studies, namely the additional power and thereby smaller sample size requirements afforded through largely controlling for genetic and some environmental contributors to variation, makes the ongoing longitudinal study of twin cohorts of paramount importance in this ever-expanding research area.

It is hard to believe that just over 20 years ago, the human genome had not been characterized fully, micro RNAs and other non-coding RNAs were considered largely artefactual, and epigenetics was often considered to be of 'limited' clinical importance, beyond rare imprinted disorders and potentially some cancers. Today, reference epigenomes, mapping many different epigenetic marks, exist for almost every human tissue and cell type, across several different ages, and health states. It is now unequivocal that epigenetic variation plays a role in healthy development and ageing throughout the life course and is also key in the progression of all human neoplasia, and likely a gamut of other adverse health outcomes. Distinct subsets of DNA methylation variation not only act as a robust biomarker of tobacco smoke exposure, but also define accelerated 'epigenetic ageing' one of the best population predictors of a range of adult-onset non-communicable diseases and all-cause mortality. As a new generation of human cohorts begins to mature and techniques for measuring epigenetic markers increase in sensitivity and decline in cost, it is likely that a full characterization of both genomic and epigenomic data in humans will soon be within the reach of most researchers. It is also likely that the level of complexity and variation revealed within individual cells/tissues/organs of any individual, *and* between individuals over time will be immense, potentially hindering the identification of clinically relevant variation in epigenetic profile. Despite this, the inherent capacity to control for genetic variation in MZ twin pairs (and to a lesser extent environmental factors) will prove incredibly valuable in helping to unravel the mind boggling complexity of gene:environment: epigenetic interactions that underpin human health and disease.

References

1. Hasler R et al (2009) Genetic control of global gene expression levels in the intestinal mucosa: a human twin study. Physiol Genomics 38(1):73–79
2. Sharma K (2005) Genetic epidemiology of epilepsy: a twin study. Neurol India 53(1):93–98
3. Tan Q et al (2005) Genetic dissection of gene expression observed in whole blood samples of elderly Danish twins. Hum Genet 117(2–3):267–274
4. Gordon L et al (2012) Neonatal DNA methylation profile in human twins is specified by a complex interplay between intrauterine environmental and genetic factors, subject to tissue-specific influence. Genome Res 22(8):1395–1406
5. Gordon L et al (2011) Clues to the effect of intrauterine milieu on gene expression from a study of two tissues from neonatal monozygotic twins. Epigenetics 6(5):579–592
6. Gartner K (1990) A third component causing random variability beside environment and genotype. A reason for the limited success of a 30 year long effort to standardize laboratory animals? Lab Anim 24(1):71–77
7. Kerkel K et al (2008) Genomic surveys by methylation-sensitive SNP analysis identify sequence-dependent allele-specific DNA methylation. Nat Genet 40(7):904–908
8. Schalkwyk LC et al (2010) Allelic skewing of DNA methylation is widespread across the genome. Am J Hum Genet 86(2):196–212
9. Fuke C et al (2004) Age related changes in 5-methylcytosine content in human peripheral leukocytes and placentas: an HPLC-based study. Ann Hum Genet 68(Pt 3):196–204
10. Fraga MF et al (2005) Epigenetic differences arise during the lifetime of monozygotic twins. Proc Natl Acad Sci U S A 102(30):10604–10609
11. Ollikainen M et al (2010) DNA methylation analysis of multiple tissues from newborn twins reveals both genetic and intrauterine components to variation in the human neonatal epigenome. Hum Mol Genet 19(21):4176–4188
12. Tan Q et al (2015) Twin methodology in epigenetic studies. J Exp Biol 218(Pt 1):134–139
13. Boomsma D, Busjahn A, Peltonen L (2002) Classical twin studies and beyond. Nat Rev Genet 3(11):872–882
14. Bell JT, Spector TD (2011) A twin approach to unraveling epigenetics. Trends Genet 27:116–125
15. Heijmans BT et al (2007) Heritable rather than age-related environmental and stochastic factors dominate variation in DNA methylation of the human IGF2/H19 locus. Hum Mol Genet 16(5): 547–554
16. Kaminsky ZA et al (2009) DNA methylation profiles in monozygotic and dizygotic twins. Nat Genet 41(2):240–245
17. McRae AF et al (2014) Contribution of genetic variation to transgenerational inheritance of DNA methylation. Genome Biol 15(5):R73
18. Villicana S, Bell JT (2021) Genetic impacts on DNA methylation: research findings and future perspectives. Genome Biol 22(1):127
19. Min JL et al (2021) Genomic and phenotypic insights from an atlas of genetic effects on DNA methylation. Nat Genet 53(9):1311–1321
20. van Dongen J et al (2021) Identical twins carry a persistent epigenetic signature of early genome programming. Nat Commun 12(1):5618
21. Wong CC et al (2010) A longitudinal study of epigenetic variation in twins. Epigenetics 5(6): 516–526
22. Spector TD (2000) Advances in twin and sib-pair analysis. Greenwich Medical Media
23. Martino D et al (2013) Longitudinal, genome-scale analysis of DNA methylation in twins from birth to 18 months of age reveals rapid epigenetic change in early life and pair-specific effects of discordance. Genome Biol 14(5):R42
24. Levesque ML et al (2014) Genome-wide DNA methylation variability in adolescent monozygotic twins followed since birth. Epigenetics 9(10):1410–1421

25. Horvath S (2013) DNA methylation age of human tissues and cell types. Genome Biol 14(10): R115
26. Knight AK et al (2016) An epigenetic clock for gestational age at birth based on blood methylation data. Genome Biol 17(1):206
27. Lee Y et al (2019) Placental epigenetic clocks: estimating gestational age using placental DNA methylation levels. Aging (Albany NY) 11(12):4238–4253
28. Girchenko P et al (2017) Associations between maternal risk factors of adverse pregnancy and birth outcomes and the offspring epigenetic clock of gestational age at birth. Clin Epigenetics 9: 49
29. McCartney DL et al (2021) Genome-wide association studies identify 137 genetic loci for DNA methylation biomarkers of aging. Genome Biol 22(1):194
30. Reynolds CA et al (2020) A decade of epigenetic change in aging twins: genetic and environmental contributions to longitudinal DNA methylation. Aging Cell 19(8):e13197
31. Jylhava J et al (2019) Longitudinal changes in the genetic and environmental influences on the epigenetic clocks across old age: evidence from two twin cohorts. EBioMedicine 40:710–716
32. Fohr T et al (2021) Does the epigenetic clock GrimAge predict mortality independent of genetic influences: an 18 year follow-up study in older female twin pairs. Clin Epigenetics 13(1):128
33. Li S et al (2020) Genetic and environmental causes of variation in epigenetic aging across the lifespan. Clin Epigenetics 12(1):158
34. Zwijnenburg PJ, Meijers-Heijboer H, Boomsma DI (2010) Identical but not the same: the value of discordant monozygotic twins in genetic research. Am J Med Genet B Neuropsychiatr Genet 153B(6):1134–1149
35. Sakuntabhai A et al (1999) Mutations in ATP2A2, encoding a Ca2+ pump, cause Darier disease. Nat Genet 21(3):271–277
36. Robertson SP et al (2006) Postzygotic mutation and germline mosaicism in the otopalatodigital syndrome spectrum disorders. Eur J Hum Genet 14(5):549–554
37. Kondo S et al (2002) Mutations in IRF6 cause Van der Woude and popliteal pterygium syndromes. Nat Genet 32(2):285–289
38. Smith AC et al (2006) New chromosome 11p15 epigenotypes identified in male monozygotic twins with Beckwith-Wiedemann syndrome. Cytogenet Genome Res 113(1–4):313–317
39. Helderman-van den Enden AT et al (1999) Monozygotic twin brothers with the fragile X syndrome: different CGG repeats and different mental capacities. J Med Genet 36(3):253–257
40. Kruyer H et al (1994) Fragile X syndrome and the (CGG)n mutation: two families with discordant MZ twins. Am J Hum Genet 54(3):437–442
41. Bonilla E et al (1990) Partial dystrophin deficiency in monozygous twin carriers of the Duchenne gene discordant for clinical myopathy. Neurology 40(8):1267–1270
42. Biousse V et al (1997) De novo 14484 mitochondrial DNA mutation in monozygotic twins discordant for Leber's hereditary optic neuropathy. Neurology 49(4):1136–1138
43. Blakely EL et al (2004) Mitochondrial DNA deletion in "identical" twin brothers. J Med Genet 41(2):e19
44. Bruder CE et al (2008) Phenotypically concordant and discordant monozygotic twins display different DNA copy-number-variation profiles. Am J Hum Genet 82(3):763–771
45. Bakaysa SL et al (2007) Telomere length predicts survival independent of genetic influences. Aging Cell 6(6):769–774
46. Slagboom PE, Droog S, Boomsma DI (1994) Genetic determination of telomere size in humans: a twin study of three age groups. Am J Hum Genet 55(5):876–882
47. Jonsson H et al (2021) Differences between germline genomes of monozygotic twins. Nat Genet 53(1):27–34
48. Petronis A et al (2003) Monozygotic twins exhibit numerous epigenetic differences: clues to twin discordance? Schizophr Bull 29(1):169–178
49. Mill J et al (2006) Evidence for monozygotic twin (MZ) discordance in methylation level at two CpG sites in the promoter region of the catechol-O-methyltransferase (COMT) gene. Am J Med Genet B Neuropsychiatr Genet 141B(4):421–425

50. Mitchell MM et al (2011) Epigenetic investigation of variably X chromosome inactivated genes in monozygotic female twins discordant for primary biliary cirrhosis. Epigenetics 6(1):95–102
51. Weksberg R et al (2002) Discordant KCNQ1OT1 imprinting in sets of monozygotic twins discordant for Beckwith-Wiedemann syndrome. Hum Mol Genet 11(11):1317–1325
52. Yamazawa K et al (2008) Monozygotic female twins discordant for silver-Russell syndrome and hypomethylation of the H19-DMR. J Hum Genet 53(10):950–955
53. Oates NA et al (2006) Increased DNA methylation at the AXIN1 gene in a monozygotic twin from a pair discordant for a caudal duplication anomaly. Am J Hum Genet 79(1):155–162
54. Mastroeni D et al (2009) Epigenetic differences in cortical neurons from a pair of monozygotic twins discordant for Alzheimer's disease. PLoS One 4(8):e6617
55. Souren NY et al (2011) DNA methylation variability at growth-related imprints does not contribute to overweight in monozygotic twins discordant for BMI. Obesity (Silver Spring) 19:1519–1522
56. Javierre BM et al (2010) Changes in the pattern of DNA methylation associate with twin discordance in systemic lupus erythematosus. Genome Res 20(2):170–179
57. Baranzini SE et al (2010) Genome, epigenome and RNA sequences of monozygotic twins discordant for multiple sclerosis. Nature 464(7293):1351–1356
58. Kuratomi G et al (2008) Aberrant DNA methylation associated with bipolar disorder identified from discordant monozygotic twins. Mol Psychiatry 13(4):429–441
59. Kaminsky Z et al (2008) Epigenetics of personality traits: an illustrative study of identical twins discordant for risk-taking behavior. Twin Res Hum Genet 11(1):1–11
60. Yuan W et al (2014) An integrated epigenomic analysis for type 2 diabetes susceptibility loci in monozygotic twins. Nat Commun 5:5719
61. Carry PM et al (2021) Severity of idiopathic scoliosis is associated with differential methylation: an epigenome-wide association study of monozygotic twins with idiopathic scoliosis. Genes (Basel) 12(8):1191
62. Wang W et al (2021) Genome-wide DNA methylation and gene expression analyses in monozygotic twins identify potential biomarkers of depression. Transl Psychiatry 11(1):416
63. Paul DS et al (2016) Increased DNA methylation variability in type 1 diabetes across three immune effector cell types. Nat Commun 7:13555
64. Nilsson E et al (2014) Altered DNA methylation and differential expression of genes influencing metabolism and inflammation in adipose tissue from subjects with type 2 diabetes. Diabetes 63(9):2962–2976
65. Wei M et al (2021) Monozygotic twins discordant for immunoglobulin a nephropathy display differences in DNA methylation and gene expression. Kidney Dis (Basel) 7(3):200–209
66. Young JI et al (2021) DNA methylation variation is identified in monozygotic twins discordant for non-syndromic cleft lip and palate. Front Cell Dev Biol 9:656865
67. Grunert M et al (2020) The needle in the haystack-searching for genetic and epigenetic differences in monozygotic twins discordant for tetralogy of Fallot. J Cardiovasc Dev Dis 7(4):55
68. Roberson-Nay R et al (2020) An epigenome-wide association study of early-onset major depression in monozygotic twins. Transl Psychiatry 10(1):301
69. Zhu Y et al (2019) Genome-wide profiling of DNA methylome and transcriptome in peripheral blood monocytes for major depression: a monozygotic discordant twin study. Transl Psychiatry 9(1):215
70. Liang S et al (2019) Genome-wide DNA methylation analysis reveals epigenetic pattern of SH2B1 in Chinese monozygotic twins discordant for autism Spectrum disorder. Front Neurosci 13:712
71. Mohandas N et al (2019) Evidence for type-specific DNA methylation patterns in epilepsy: a discordant monozygotic twin approach. Epigenomics 11(8):951–968
72. Souren NY et al (2019) DNA methylation signatures of monozygotic twins clinically discordant for multiple sclerosis. Nat Commun 10(1):2094

73. Webster AP et al (2018) Increased DNA methylation variability in rheumatoid arthritis-discordant monozygotic twins. Genome Med 10(1):64
74. Murphy TM et al (2015) Methylomic markers of persistent childhood asthma: a longitudinal study of asthma-discordant monozygotic twins. Clin Epigenetics 7:130
75. Ollikainen M et al (2015) Genome-wide blood DNA methylation alterations at regulatory elements and heterochromatic regions in monozygotic twins discordant for obesity and liver fat. Clin Epigenetics 7:39
76. Cobos I et al (2005) Mice lacking Dlx1 show subtype-specific loss of interneurons, reduced inhibition and epilepsy. Nat Neurosci 8(8):1059–1068
77. Wu TD, Nacu S (2010) Fast and SNP-tolerant detection of complex variants and splicing in short reads. Bioinformatics 26(7):873–881
78. Bennett CM, Boye E, Neufeld EJ (2008) Female monozygotic twins discordant for hemophilia a due to nonrandom X-chromosome inactivation. Am J Hematol 83(10):778–780
79. Willemsen R et al (2000) Twin sisters, monozygotic with the fragile X mutation, but with a different phenotype. J Med Genet 37(8):603–604
80. Zneimer SM, Schneider NR, Richards CS (1993) In situ hybridization shows direct evidence of skewed X inactivation in one of monozygotic twin females manifesting Duchenne muscular dystrophy. Am J Med Genet 45(5):601–605
81. Shur N (2009) The genetics of twinning: from splitting eggs to breaking paradigms. Am J Med Genet C Semin Med Genet 151C(2):105–109

Age-Related Variation in DNA Methylation 10

Kelsey Keith, Jean-Pierre J. Issa, and Shoghag Panjarian

Contents

Abstract

Methylation is a ubiquitous, naturally occurring biochemical modification of DNA in mammalian cells that regulates gene expression. DNA methylation patterns are generally stable in the short term but show prominent changes in aging cells. DNA methylation changes are tissue specific and bidirectional, including gains of methylation at previously protected promoter regions and losses of methylation genome wide. In fact, DNA methylation changes linearly with age and could be used to estimate the biological age (methylation age). To this end, in the past decade, several epigenetic clocks have been developed to measure biological rather than chronological ages. The differences between the estimated and the chronological ages could reflect biological variability and correlate with life expectancy. These

K. Keith · J.-P. J. Issa · S. Panjarian (✉)
Coriell Institute for Medical Research, Camden, NJ, USA
e-mail: spanjarian@coriell.org

K. B. Michels (ed.), *Epigenetic Epidemiology*,
https://doi.org/10.1007/978-3-030-94475-9_10

age-related methylation changes remain mechanistically mysterious and complex but are conserved across species and are likely caused by infidelity in the replication of the epigenome over time. Because of the link between DNA methylation and gene expression, these changes result in a mosaic epigenome in aged cells that could underlie diseases of aging such as cancer.

Abbreviations

5hmC	5-hydroxymethylcytosine
5mC	5-methylcytosine
BER	base excision repair
CGI	CpG island
DHEA	dehydroepiandrosterone
DNMT	DNA methyl transferase
NGS	next generation sequencing
PRC2	polycomb repressor complex 2
rhGH	recombinant human growth hormone
TDG	thymine DNA glycosylase
TET	Ten-eleven translocation

10.1 DNA Methylation, Demethylation, and Gene Expression

Methylation of human DNA consists of the transfer of a methyl (CH_3) group from S-adenosylmethionine to the C5 position of cytosine when followed by guanosine (the CpG dinucleotide) to form 5-methylcytosine (5mC) [1, 2]. This covalent biochemical modification takes place after DNA synthesis and is catalyzed by DNA methyltransferases (DNMTs). There are three DNMT enzymes: DNMT1, DNMT3A, and DNMT3B. DNMT3A and DNMT3B are *de novo* methyltransferases that are essential in initiating and establishing the methylation pattern during embryogenesis and in stem cells. The methyl mark is then faithfully copied from the parental strand to the daughter strand and propagated by the maintenance methyltransferase DNMT1 enzyme [3–5]. There are two isoforms of DNMT3A and over thirty isoforms of DNMT3B that were suggested to play important roles as accessory proteins to restore DNA methylation and in aberrant methylation in tumorigenesis [6, 7]. Other related proteins include DNMT3L (DNMT 3-like) that lacks catalytic activity and DNMT2 which has RNA methyltransferase activity [8, 9].

Until the discovery of TET enzymes in 2009, DNA demethylation was thought to take place only because of a passive dilution process during DNA replication in the absence of DNMTs. TET enzymes (TET1, TET2, TET3) are Fe2+ and 2-oxoglutarate-dependent dioxygenases that oxidize 5mC into 5-hydroxymethylcytosine (5hmC) which then can be further oxidized into 5-formylcytosine (5fC) and

5-carboxylcytosine (5caC) [10, 11]. These oxidized derivatives are excised and repaired by the thymine DNA glycosylase (TDG) and base excision repair (BER) pathways, respectively, leading to demethylation of 5mC [12]. All TET enzymes have a common core catalytic domain with key residues that interact with substrates and cofactors such as 2-oxoglutarate, Fe (II), and molecular oxygen to carry the oxidation reaction. At the N-terminal region, TET1 and TET3 contain a CXXC domain that targets these enzymes to unmethylated CpG islands while TET2 lacks CXXC domain and pairs up with CXXC containing IDAX protein. TET enzymes are primarily expressed during embryogenesis but are also detected in different adult tissues [13–15].

DNA methylation is a highly regulated mechanism that plays important roles in gene expression. This effect on gene expression is dependent on the CpG location within the genomic compartment (gene body, promoter, intergenic regions) as well as whether the CpG is in CpG islands (CGI, short stretches of CpG rich DNA regions) or non-CGIs [1, 16, 17]. The best-known role of DNA methylation and its concomitant effect on gene expression is at promoter CGIs. Methylation at these loci, though rare, is associated with stable gene silencing, such as on imprinted genes, on the inactive X-chromosome in women, and on germ cell specific genes [1, 18]. The effect of methylation on CpG sites at other loci has variable effects on gene expression. On the other hand, unlike methylation, which is maintained and propagated through cell division, TET-mediated demethylation could also induce gene expression by the demethylation of promoter non-CGIs [19].

10.2 DNA Methylation Changes with Age: Location, Tissue Variability, and Rate of Change

The understanding of DNA methylation changes with age has been facilitated by and evolved with the advancement of genome-wide quantitative technologies. The earliest studies used global measurement of 5-methylcytosine, which primarily reflects non-promoter, non-CpG island DNA [20]. These initial studies showed progressive depletion associated with passaging of fibroblasts in vitro and subsequently similar changes were observed in aging mouse and human tissues [21, 22].

Later, quantitative analysis of DNA methylation at specific sites indicated that human aging was accompanied by small, but measurable, changes in many genes. Initial studies focused on a handful of genes with unmethylated promoter CpG islands in normal tissues that hypermethylated in tumors. For example, *ERα* was one of the first genes where it was shown that DNA methylation in normal tissue increased linearly with age at a rate of 1% every 3 years in human colon [23]. This was also true for many other genes and genome-wide studies showed that promoter CpG islands that gain DNA methylation in cancer also gain methylation with age [24, 25]. Additionally, DNA methylation changes as well as the degree of the changes vary in a tissue-specific manner [26, 27]. For example, in mouse tissues the most prominent changes are observed in the most proliferative organ, the GI tract, followed by spleen and other tissues [27]. Studies in human tissues also showed similar changes with age. Thus it has been proposed that, although the changes are tissue specific, one can estimate an

individual's age through the DNA methylation in peripheral blood through calculating the rate of change with age [28–32].

10.3 Methylation Changes and Correlation with Lifespan

DNA methylation changes with age are characterized by both losses and gains of methylation at different loci [27, 33]. This phenomenon is referred to as methylation drift. Methylation drift takes place in different species such as mice, rhesus monkeys, dogs, whales, and humans. Age-related methylation drift is evolutionary conserved [34], allowing one to compare rates of change across species. In one study, a multilevel linear mixed effect model was applied to 10 age-related hypermethylated genes with high sequence conservation between the species, yielding drift rates (mean ± SEM) for mice (4.1 ± 1.2% per year), monkeys (0.34 ± 0.14% per year), and humans (0.1 ± 0.02% per year) that show that methylation drift is inversely proportional to longevity [34]. Consistent with this, age-related DNA methylation changes in 26 different bat species showed that in long-lived bat species age-related sites had slower rate of change while short-lived species had a faster rate of change [35]. Additionally, it has been shown that a methylation age can be estimated using different sets of CpG sites with age-related DNA methylation changes [31, 36, 37]. The difference between the estimated age and the chronological age could reflect biological variability in aging and correlate with life expectancy within a species. In fact, it was demonstrated that when estimated age is much higher than chronological age, this accelerated methylation aging is associated with higher all-cause mortality [38].

10.4 Epigenetic Clocks

As age and aging-related diseases are a large burden on individuals and on healthcare systems, there is interest in developing a measure of biological, rather than chronological age. DNA methylation patterns, which are generally stable but also change with age, have been used as an important tool towards this end. Using CpG sites that are strongly correlated with chronological age, researchers have constructed DNA methylation clocks that accurately predict an individual's age [31, 36, 39, 40]. Most commonly used are the Hannum clock for human blood and the human Horvath clock, which claims to be a pan-tissue master clock [31, 36]. While these clocks have been shown to be accurate at predicting chronological age, which could be useful for forensic applications, clinicians and researchers are interested in the biological age of the subjects. There is wide variation in the visible aging of people and in their burden of age-related frailty and disease. This is reflected in many biological measures, such as mutation rate, telomere length, and, the subject of this chapter, DNA methylation [41]. By calculating individual's ages using these clocks, clinicians could theoretically identify individuals with aging faster than average (i.e., accelerated aging), and target them for interventions, such as screening them more carefully for diseases of age. On the other hand, screening could be relaxed for individuals with decelerated

aging. Predicting chronological age directly, as the Horvath and Hannum clocks do, might not be the best way to detect accelerated biological aging and therefore disease risk. Other clocks have been designed to specifically examine disease and mortality risk, incorporating additional measurement such as tissue mitoses [42, 43], known physiological changes [44–46], and environmental factors [47]. In addition to clocks developed in humans, there have been several clocks developed in other organisms, including whales [48], canines [49], and mice [50–52]. These clocks are useful to test interventions that may alter the rate of aging in model organisms. See Table 10.1 for a list of the clocks and Chap. 11: Epigenetic Clocks for a more detailed description of epigenetic clock construction and applications.

Since the inception of these clocks, they have been applied to study an array of aging-related changes and disorders. Looking at healthy individuals, age-related methylation changes begin early in life [46, 57–59] and continue over an individual's lifetime, although theoretically the rate decelerates the older an individual is [60]. Many diseases are associated with accelerated aging, both environmental [42, 43, 61–70] and non-environmental genetic diseases [71–73] as well as physiological markers of aging [56, 74–76], and environmental and lifestyle factors [31, 77–82]. However, with the exceptions of cancer and HIV, the association between accelerated aging and these factors have not been replicated, so more work needs to be done to confirm those findings. Especially for lifestyle factors, the association could be correlation, not causation, where both the variable of interest and aging could be correlated with some third factor that is causing the accelerated aging. Finally, multiple studies have found that increased epigenetic age is associated with all-cause mortality [38, 45, 54, 66, 68, 83, 84]. On the other hand, as you would expect, very long-lived super centenarian humans show decelerated aging [85] and positive lifestyle factors are associated with modest decelerated aging in the clocks [79]. Then, having established that the clocks are associated with disease and mortality risk, but also that some interventions can slow the clocks, researchers have tested whether the rate of the clock can be actively perturbed. In model organisms, researchers have been able to test interventions to slow down or speed up aging. Interventions in mice have showed the ability to accelerate, ovariectomy or high-fat diet [51], or decelerate, dwarfism, calorie restriction, or rapamycin treatment [50–52], aging. Additionally, a very early and small safety study in 10 human males treated with recombinant human growth hormone (rhGH), dehydroepiandrosterone (DHEA), and metformin over 9 months has tentatively showed some age deceleration effects. Using a blood sample and calculating biological age using the Horvath, Levine, Hannum, and GrimAge/Lu clocks, the biological age of subjects was on average 1.5 years younger after treatment [86]. While the sample size is wholly inadequate to draw any conclusions from the study, it does demonstrate that trials testing aging interventions can be safely carried out in humans and is a first step down the road to further work.

Despite the progress that has been made using these clocks, there are structural issues with the construction of the clocks that affect their accuracy. First, clocks have been developed primarily with 450k arrays, ignoring 99% of the possible CpG sites in the human genome which could carry interesting biology as well. Second, many of the clocks have been developed using "normal" tissue sampled during cancer

Table 10.1 Overview of existing epigenetic clocks

Clock	Significance	Number of CpG sites	Tissue	Species	Year	References
Bocklandt	One of the first clocks	88	Saliva	*Homo sapiens*	2011	[39]
Koch	One of the first clocks	5	Multi-tissue	*Homo sapiens*	2011	[40]
Hannum	One of the first clocks	71	Blood	*Homo sapiens*	2013	[31]
Horvath	Most used clock	353	Multi-tissue	*Homo sapiens*	2013	[36]
Weidner	Hypothetically, only a few [3] CpG sites are necessary for accurate age prediction	3	Blood	*Homo sapiens*	2014	[37]
Humpback Epigenetic Age Assay (HEAA)	Only whale clock. Useful to estimate whale age since the only other way to determine whale age is through morphological features	3	Skin	*Megaptera novaeangliae*	2014	[48]
Lin	Clock results are associated with mortality	99	Blood	*Homo sapiens*	2016	[38]
Vidal-Bralo	Again, only a few CpG sites need and this clock can be run in a single multiplex assay	10	Blood	*Homo sapiens*	2016	[53]
Epigenetic Timer of Cancer (epiToc)	Incorporates number of mitoses in tissue and baseline methylation at birth, making it more accurate	385	Multi-tissue	*Homo sapiens*	2016	[42]
Intrinsic Epigenetic Age Acceleration (IEAA)	Attempts to eliminate the effect of changing blood cell composition with age in the model	71 or 353	Multi-tissue	*Homo sapiens*	2016	[45]
Extrinsic Epigenetic Age Acceleration (EEAA)	Incorporates changing blood cell composition with age into the model along with DNA methylation changes	71 or 353	Multi-tissue	*Homo sapiens*	2016	[45]
Zhang	Provides a mortality risk score rather than predicting biological age	58	Blood	*Homo sapiens*	2017	[54]
Thompson	Only dog clock. Useful for aging studies that use dogs as model organisms and to estimate wolf ages, which like whales, is difficult	115	Blood	*Canis familiaris* and *Canis lupus*	2017	[49]
Petkovich	Published the same month as Stubbs, one of the first two mouse clocks	90	Blood	*Mus musculus*	2017	[50]

Stubbs	Published the same month as Petkovish, one of the first two mouse clocks. Also shows that mouse and human aging is similar	329	Pan-tissue	*Mus musculus*	2017	[51]
Wang	For liver aging specifically rather than blood or a universal multi-tissue clock	148	Liver	*Mus musculus*	2017	[52]
MiAge	Estimates the number of cell divisions a sample has undergone rather than biological age	268	Multi-tissue	*Homo sapiens*	2018	[55]
DNAm PhenoAge	Uses phenotypic age based on the subject's health rather than chronological age to build the model	513	Blood	*Homo sapiens*	2018	[44]
DunedinPoAm	Includes 18 physiological markers associated with aging in additional to biological age in a longitudinal cohort	46	Blood	*Homo sapiens*	2020	[46]
Epigenetic Timer of Cancer 2 (epiTOC2)	Improvement to original epiTOC, expanding on number and types of sites including with more training data	163	Multi-tissue	*Homo sapiens*	2020	[43]
HypoClock	Only clock looking solely about hypomethylating sites	678	Multi-tissue	*Homo sapiens*	2020	[56]

biopsies. This is routinely collected in order to compare the cancer to the patient's own tissue, but these normal adjacent samples have an intermediate phenotype between truly normal tissue and cancer [87, 88] and also show accelerated aging compared to normal tissues [32, 89]. The normal adjacent samples do not represent normal human aging and skew the clocks including them, such as the Horvath clock. Third, these clocks have been developed, either partially or totally, using elastic net regularized regression. The rule of thumb for model building is that one needs 10 times as many samples as there are features to select the most informative features; in the case of the 450k array, 4.5 million samples would be needed, a number of samples that simply does not exist. Using elastic net solves this problem since, out of all the CpG sites measured, the technique finds a small subset of CpG sites that are strongly associated with age and eliminates the rest of the sites, removing the need for an exorbitantly large sample size. The small subset of CpGs elastic net selects also allow the clocks to be easily applied to new datasets if those same CpGs are measured [90]. Measuring the same CpGs is not an issue if one continues to use the same 450K arrays from the original papers, but with various NGS sequencing technologies, measuring all the same CpGs is unlikely, with as little as 24% overlap between libraries [91]. Also, many sites are equivalently predictive of age and there elastic net will select the "best" ones based on the underlying data [51]. Since most of these clocks have been developed with different data in different tissues, the underlying data will affect which CpG sites are "best," causing tissue-specific effects, and certainly making it difficult to compare clocks since few CpG sites overlap between any of them. Fourth, there are accuracy issues with the clocks themselves, both with their age predictions and especially in applying the same age prediction across multiple tissues. As mentioned earlier, even in the original Horvath paper, correlation with age was only 0.55 in some tissues [36]. Similarly, the DNAm PhenoAge clock, developed using whole blood, had a correlation of 0.35 in breast tissue [44] and in the cerebellum, in an independent study, the Horvath clock only has a correlation of 0.66 with age [92]. Recently, a study found that blood cell composition does affect predictions and that compensating for it can improve the accuracy of clocks, demonstrating that cell composition does confound the clocks [93]. Additionally, in an independent study of 7 tissues, the majority of CpG sites (85% of hypermethylated CpGs, 97% of hypomethylated CpGs) were tissue specific [94]. More troublingly than the poor correlations, the Horvath clock, and clocks constructed using elastic net, systematically overestimate the age of young individuals and underestimate the age of old individuals [92]. In longitudinal study, the Horvath and Hannum clocks predict that aging slows over an individual's lifetime [60]. Some of this may be, as the study authors suggest, driven by survivor bias since only healthy individuals live to be included in a longitudinal study, but it could also be a consequence of the math behind elastic net optimizing for the mean of the data. Because it optimizes for the mean, elastic net will predict the ages of people close to the mean, middle-aged people from about their 30s–50s, the most accurately. Young people aged under 30 will be overestimated because that is closer to the middle-aged mean of the data, while the opposite will happen with older individuals and their ages will be underestimated, again because that is closer to the mean of the data. The underestimation of the age of older individuals is a huge

confounding error with epigenetic clocks overall as their main purpose is to detect individuals with accelerated aging and the systematic underestimation undermines that. Finally, a recent systematic review and meta-analysis of epigenetic clocks found that there was a positive publication bias [95], offering evidence that negative data is not being published, so more work needs to be done to acknowledge and address the issues with the existing clocks.

In addition to the technical issues behind clock construction, there are many unanswered questions about the biology underlying the phenomenon. First, the biological and chronological components of clocks need to be teased apart. While a chronological clock could be useful for forensics, the greater the accuracy of the chronological age prediction, the less it is associated with disease and mortality [54]. To construct better biological clocks, we need to understand the mechanisms underlying the changes in DNA methylation with age and construct clocks guided by the underlying biology. The current clocks also suffer from the type and quantity of the data available. Currently, at best, clocks can be constructed using a few thousand samples that have 450k array data, which measures only 1% of the genome with only a few hundred longitudinal samples existing. Thousands more samples measuring DNA methylation genome wide and with longitudinal data are necessary to start teasing out the biological underpinnings of age. In addition to bulk data, single cell data is vitally necessary to identify how different cell types and changing proportions of cell types within a tissue contribute to the clocks and aging generally. These samples need to be acquired not only for humans but also for important model organisms like mice and rats, so that pre-clinical intervention testing can be done [41].

10.5 Methylation Outliers

In recent years, interest in DNA methylation outliers has been the focus of many publications within the context of identifying these phenotypes, understanding the causes of this variation, and the biological implications in terms of disease susceptibility and clinical outcomes [89, 96–98]. To understand the causes and the consequences of the outlier phenotype, it is necessary to understand what the outlier phenotype is. Individuals are considered outliers for DNA methylation when DNA methylation levels in their normal tissues vary significantly from the population mean. Different studies have used different statistical algorithms to identify these individuals with the outlier phenotype, which is a rare event, making their identification a non-trivial task. Additionally, identifying these individuals is also dependent on important technical and biological considerations, such as the number of CpG sites, the nature of the CpG sites, and the direction of the change—gain or loss of methylation.

In a very recent study using whole genome DNA methylation analyses, we found that individuals with outlier DNA methylation phenotype could be identified by about 150 CpG sites that show tissue-specific age-dependent DNA methylation changes [32]. These rare outliers could be best identified using tissue-specific changes rather than tissue-agnostic clocks and could not be identified by Horvath's multi-tissue estimator CpG sites nor random CpG sites. Outlier individuals were predicted to be older than their chronological age based on the methylation levels of

these CpG sites. Thus, accelerated aging is one phenotype that characterizes a group of outlier individuals based on alterations in their epigenomes. However, there are many unanswered questions concerning the outliers, such as if these individuals with an accelerated aging phenotype are at higher risk of cancer development sooner since cancer is an aging disease? Alternatively, are outlier individuals with decelerated aging protected from developing cancer? Would it be possible to predict who is at higher risk of phenotypic variation based on DNA methylation sites that change with age?

10.6 Causes of Age-Related DNA Methylation Alterations

10.6.1 Replication Error and the Mitotic Clock Hypothesis

There is no simple explanation for age-related changes in methylation. As it is bidirectional, increasing at some sites and decreasing at others, this eliminates simple explanations such as changes in DNMTs or TETs alone [27]. Changes are also not completely random; certain genomic compartments and features are more prone to age-related errors [99]. Age-related changes are also conserved over evolution and tissue specific, with the most proliferative tissues, like intestinal epithelium, showing the most change with age [27]. This gives rise to the mitotic clock hypothesis, which postulates the error accumulation is the result of errors in the maintenance of epigenetic states during replication. Emerging evidence is beginning to show how small errors accumulate in stem cells over time and are passed on to their differentiated daughters, leading to dysfunction and pathogenesis with age.

Some mechanisms of aging DNA methylation alterations have been identified. While, again, it does not explain all age-related changes, alterations in readers and writers of methylation like mutations [100], altered gene expression [101–103], altered enzymatic activity through changes in the levels of their co-enzymes [104], and altered chromatin recruitment [105, 106] do result in age-related changes. Another cause of age-related changes, primarily hypomethylation, is mistakes in DNA re-methylation during replication. Not totally elucidated, the kinetics of re-methylation after DNA replication are complicated; most methylation is re-established quickly after replication, but a large portion of DNA takes several hours after replication to re-methylate, especially in heterochromatin [99, 107, 108]. This later methylating DNA has been found to be more prone to errors and loss of methylation [57, 109]. DNMT1 is also less efficient at re-methylation where CpGs marked for demethylation are present, so the activity of TETs can also lead to passive demethylation in cycling cells [110, 111]. These errors can accumulate over a lifetime because, unlike with DNA mutations, there seems to be no strict checkpoint or error checking for methylation after replication. Even in extreme cases, such as treatment with strong demethylating agent 5-aza-2′-deoxycytidine, it does not prevent cells from entering the cell cycle and completing replication [112]. Another known mechanism of age-related hypomethylation is transcription-factor induced hypomethylation. Transcription factors can recruit TET to demethylate DNA when a cell needs to activate a previously silent gene, but this could cause errors in two

ways, either through erroneous recruitment itself and/or through re-methylation never occurring [113–117].

Considering mechanisms that cause gain of methylation, inactivation or reduction in expressions of TETs leads to gains in DNA methylation since methylation is not removed [29, 118, 119]. Going back to replication, a recent study using hairpin-DNA sequencing, which distinguishes the mother and daughter strands, showed up to 5% new methylation after replication [99]. One mechanism for gain of methylation during replication could be through aberrant recruitment of DNMT1 (through misplaced histone marks or errors from recruitment partners) during replication [120]. Another could be loss of DNMT1 regulatory partners that manage the levels of DNMT1 in the cell, which has been shown to occur *in vitro* [121]. For example, polycomb repressor targets, which are unmethylated during development, are known features that gain methylation with age, [42, 43, 122] and polycomb repressor complex 2 (PRC2) is a known recruitment partner of DNMTs [123], so likely the methylation gain is through aberrant recruitment. The frequency of de novo DNA methylation is correlated with the density of CpGs [124], where if there are many methylated CpGs close together, neighboring CpGs are more likely to gain methylation, [99, 125, 126], while solo CpG sites are likely to lose methylation [99]. It is proposed that *de novo* DNA methylation can also happen through DNMT1 sliding on DNA, where it is appropriately recruited to hemi-methylated DNA and then erroneously continues down the strand [127]; there is biochemical evidence that DNMT1 can travel down a piece of DNA, methylating each CpG [128, 129], although there is currently no *in vitro* or *in vivo* confirmation.

10.6.2 Inflammation and Environmental Exposures

As mentioned earlier, the causes of DNA methylation changes are multifactorial, with chronic inflammation and environmental exposures playing important direct and indirect roles. Understanding how these contribute to age-related methylation changes is important for interventions to modulate the impact of those changes and reduce the associated disease incidence.

Chronic inflammation has been strongly linked to age-related DNA methylation changes and the observed changes are again bidirectional (gains and losses). In the gastrointestinal system, chronic inflammation in colon, esophagus, stomach, and liver is associated with increases in DNA methylation in apparently normal tissues [130–133]. Additionally, individuals with inflammatory bowel diseases have more DNA methylation changes in the colon compared with individuals without the chronic inflammatory diseases [130, 134, 135]. Increase in methylation is also associated with *Helicobacter pylori* infection-related inflammation in the stomach [136].

On the other hand, the associations between environmental factors such as diet, nutrition, stress, physical activity, air pollutants, and smoking, and age-related DNA methylation changes are less well studied [31, 137, 138]. This is partly because it is hard to design a precise study about the effect of, for example, dietary factors on DNA methylation changes, but it is plausible that dietary factors can induce inflammation and hence cause epigenetic changes, or they can directly impact the

epigenome through their metabolites. One-carbon metabolism provides a direct link between nutrients such as folate and B-vitamins, which influence the availability of the substrate of DNMTs, S-adenosyl methionine, and drives the methylation reaction. In mice, it was shown that gestational exposure to folate and vitamin B12 could impact methylation patterns in neonates. In another study in humans, it was reported that low dietary intake of vitamin B12 can lead to increased DNA methylation of genes involved in one-carbon metabolism which could confer cardiovascular disease [139]. Additionally, an epigenome-wide association study revealed that supplementation of vitamin B12 and folate modulates DNA methylation patterns in older adults, highlighting key targets of age-related DNA methylation sites for further investigation [140, 141]. In another cross-sectional study of postmenopausal women from the Women's Health Initiative and the Italian cohort study InCHIANTI, that included both male and female participants, association of environmental factors on epigenetic age acceleration in blood was studied. The study reported that there was significant association with a younger epigenetic age and potential health benefits associated with higher intake of fish, poultry, fruits and vegetables, BMI [79].

Indeed, DNA methylation drift is also modulated by not just the type of dietary exposure but also by calorie count. Twenty-two to 30 year old rhesus monkeys exposed to 30% caloric restriction and 2.7–3.2 year old mice exposed to 40% caloric restriction, showed attenuated age-related DNA methylation drift [34]. In fact, based on age-related methylation changes early life dietary and lifestyle interventions were recently proposed to prevent colorectal cancer, because the premalignant cells form decades prior to giving rise to detectable cancer [142].

10.6.3 Epigenetic Mosaicism

As discussed in the previous section, errors in DNA methylation accumulate with age through intrinsic and extrinsic factors, and these erroneous DNA methylation states in somatic cells are faithfully inherited [109]. The variation in these errors increases with age as well. Considering an individual CpG site, at young ages methylation is highly similar, but as age increases, methylation at the same CpG site in different individuals can vary by up to 20% [143]. The number of individuals with outlier methylation increases with age as well [144]. These changes are also highly heterogeneous, as one longitudinal study found that only 0.32% of CpGs had age-related changes in 5% or more of samples in the study [145]. These intrinsic heterogeneous age-related changes must be a stochastic process innate to our biology, as several studies examining twins have found that the increase in methylation errors and disorder is independent of both genetics and environment, because even monozygotic twins living in the same environment do not accumulate the same errors [83, 145, 146].

Epigenetically repressed genes can be reactivated by age-related changes [147]. Changes in promoter methylation with age can affect the expression of genes, turning on previously silenced genes and turning off currently active genes. Promoters do not have to be completely methylated to affect gene expression; even

small amounts of stochastic methylation at the promoter have been associated with a reduction in gene expression [148, 149]. Examining single cell data, clonal populations were found which had activated previously repressed genes through demethylation of their promoters and were stably expressing them. Importantly, this was expression of 98 different genes in 97 different clones, genes that were totally uncorrelated with each other, demonstrating how stochastic methylation changes can lead to transcriptionally diverse subpopulations [148]. In addition to abnormal promoter methylation, aberrant targeting of transcription factors due to differentially methylated DNA can alter the gene expression program. Approximately 30% of transcription factors preferentially bind methylated motifs [150], so disruption to the methylation regulatory sequences can alter the expression program of the cell as well. Finally, methylation also silences transposable elements, which tend to lose methylation with age and become active, moving about the genome [151–153]. Reactivation of transposable elements causes DNA damage [154–156] and mutagenesis through breaking the DNA to insert new copies, and can also lead to inflammation through the innate immune system [153, 157–159].

We have postulated that errors in methylation accumulate in stem cells, and as the stem cells divide those errors accumulate and cause functional errors in their differentiated daughter cells. It is known that a high number of mutations and epigenetic modifications correlates with the number of cell divisions, and that higher mitotic age is associated with worse survival [55]. Also, considering disease for a moment, cancers and their associated adjacent tissues with the highest degree of aberrant methylation had the highest cell division rates [42, 43]. Considering stem cells directly, bulk analyses of stem cells in the intestine [160], germ line [161], hematopoiesis [162], and skeletal muscle [163] have shown that, with age, these cells become mosaics of different DNA methylation changes. Using unique molecular identifiers and single-cell cloning, it was found that embryonic stem cells maintain disordered methylation patterns [109]. However, to tease out what epigenetic subpopulations exist, we need to look at single cell data, which has become possible with the rise of high-throughput technologies in recent years. Looking at single cell data over many cell types in the liver methylome, 3.3% of locations show methylation variation with age, an order of magnitude higher than the number of somatic mutations, and heterogeneity increased with age [164]. Examining single cells of immortalized fibroblasts, lung, and colon cancer cells, sequencing both their founding populations and expanded clonal populations, most hypomethylating CpG sites were found in late-replicating domains and were putatively caused by DNA replication. Additionally, looking at hypermethylation, the changes were associated with cell proliferation and poor maintenance of methylation [148]. Looking directly at muscle stem cells, where they simultaneously measured methylation and transcription in the same cells, aged stem cells have increased cell-to-cell variability and genes with promoters with increased methylation heterogeneity have increased expression heterogeneity. Importantly while global heterogeneity was increased, individual cells expressed fewer genes [165]. This reduced gene expression program indicates competition and natural selection between the stem cells which has caused focal stem cell proliferation.

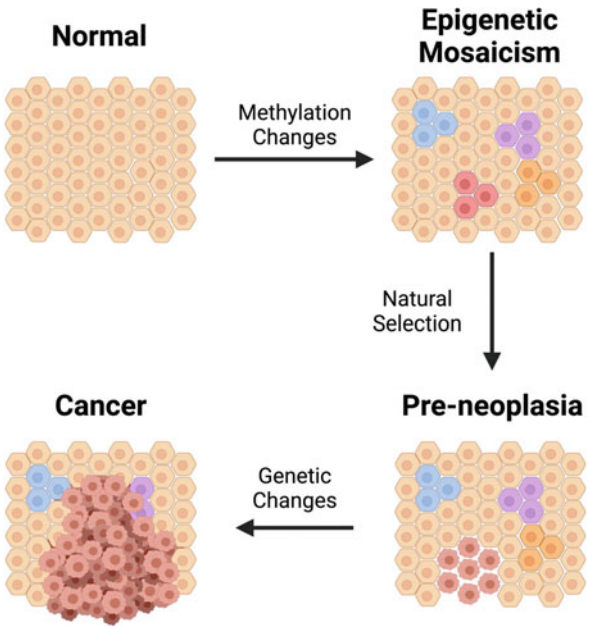

Fig. 10.1 Model of how age-related DNA methylation changes contribute to focal diseases, specifically cancer. Normal stem cells (upper left) start with uniform patterns of epigenetic regulation. Age-related DNA methylation changes result in epigenetic mosaicism (upper right) as represented by the differently colored cells and creating cell-to-cell expression variation. Gene expression variation drives natural selection and results in the overgrowth of the variant that is randomly most fit (bottom right). Because of this overgrowth, those cells are more susceptible to further changes, either random further epigenetic changes or DNA mutations or carcinogen-induced changes. The combination of all the damage acquired allows the cells to transform into full-fledged malignancy (bottom left)

Epigenetic mosaicism leading to changes in gene expression is an engine for natural selection. Evolutionary theory predicts that epigenetic variation would lead to competition among cells in a tissue and lead to the emergence of subpopulations with a selective growth advantage. From single cell data, we know this epigenetic variation takes place with age and leads to clonal subpopulations [148, 163, 164] and we also know that cancers have high rates of epigenetic heterogeneity [42, 43]. Therefore, the simplest model predicts that variation-driven natural selection results in hyperproliferative lesions and eventually cancer. As many oncogenic events lead to the death of normal cells, the convergence of selection of a lifetime likely explains the enormous rise in cancer incidence as humans age (Fig. 10.1). Even outside of the extreme of cancer, age-related methylation changes are known to happen in many disease of age like cardiovascular disease [66], where methylation-driven alterations can result in gene expression changes [147, 148] or inflammation [153, 157–159].

10.7 Emerging Quantitative Approaches to Mosaicism

Quantifying epigenetic mosaicism is an ongoing challenge. Current approaches draw from information theory and attempt to quantify the disorder in the system. As discussed previously, the variability in the system is important because even small amounts of methylation can alter gene expression [148, 149]. A new metric, Cell Heterogeneity-Adjusted cLonal Methylation (CHALM), is a variation on standard percent methylation, where if any CpG on a read is methylated, that location counts as methylated. CHALM, which treats any amount of local methylation the same as if the region was fully methylated, correlates much better with gene expression than standard percent methylation. This is particularly true at lowly methylated promoters suggesting that a small amount of methylation in regulatory regions is sufficient to disrupt the expression program [149]. Instead of compressing the effects of methylation disorder as CHALM does, entropy measures attempt to directly quantify the disorder present in a region (Table 10.2). They work by measuring regions of bisulfite sequencing that are covered by a single read and quantifying how many of the reads have the same methylation patterns (Fig. 10.2). Entropy is associated with gene expression [167] and transcriptional heterogeneity [171]. It has been used successfully to detect regions and genes with accelerated aging both in bulk and single-cell bisulfite data [143, 165, 172]. As entropy measures and methylation are calculated from the same source, they detect many of the same regions, but also each find unique patterns that have relevant biological information [169]. Of course, there are issues to be aware of with these scores. Disordered methylation can be confounded by cell-type heterogeneity and sample

Table 10.2 Published entropy scores. Some scores have more intuitive ranges where 0 is no entropy and 1 is maximum entropy. Others have maximum entropy equal to the number of CpGs analyzed; for example, if there are 4 CpGs in a region, a score of 4 would indicate maximum entropy. Differently from the rest, combinatorial entropy reaches maximum entropy at a score of −144

Entropy score	Score range	Publication year	References
Epipolymorphism	0 to 1	2012	[166]
Shannon entropy	0 to number of CpGs analyzed	2012	[166]
Proportion of Discordant Reads (PDR)	0 to 1	2014	[167]
Methclone	0 to −144	2014	[168]
Hamming distance	0 to number of CpGs analyzed	2014	[168]
Methylation haplotype load (MHL)	0 to 1	2017	[124]
Normalized Shannon entropy	0 to 1	2018	[169]
Jensen–Shannon Distance	0 to 1	2018	[169]
Fraction of Discordant Read Pairs (FDRP)	0 to number of CpGs analyzed	2020	[170]
Quantitative Fraction of Discordant Read Pairs (qFDRP)	0 to number of CpGs analyzed	2020	[170]

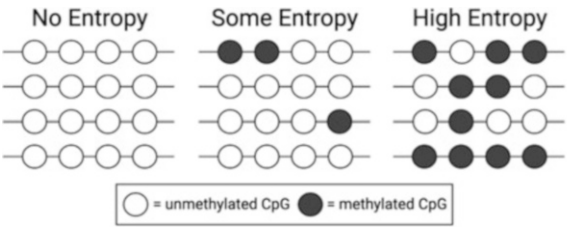

Fig. 10.2 Simple Entropy Example. Each line in the figure represents a singled read as measured by NGS sequencing and each circle is a CpG. The left-most image has no entropy because all reads are totally unmethylated. The middle image has some entropy since some of the reads have randomly gained methylation, but its entropy score would be low because half of the reads are identically unmethylated. The right-most image demonstrates maximum entropy, because every read has a different methylation pattern and number of CpGs methylated

contamination [170], although as single-cell technologies improve and become cheaper the cell-type issues can be resolved. Entropy measures can also detect allele-specific methylation [170], which does occur with age [148], but can also come from imprinted genes, which must be controlled for. It is also important to consider the underlying biology when looking at entropy scores since they can have the same values for a baseline methylated and a baseline unmethylated state, but that would lead to opposite effects in the cell. For this reason, entropy scores are best used in conjunction with other measures, usually percent methylation, as combinations of percent methylation and multiple entropy scores can yield more information than any one metric alone [169].

10.8 Conclusions

Aging is accompanied by progressive, proliferation-dependent DNA methylation changes that regulate gene expression and create epigenetic mosaicism in older tissues. These changes may compromise stem cell function and contribute to diseases of aging such as cancer. Epigenetic clocks and quantitative measurement of epigenetic mosaicism, despite their limitations, could serve in developing a measure of biological variation and estimating methylation age rather than chronological age. Interventions aimed at reducing age-related methylation deregulation have the potential to attenuate the occurrence and severity of aging diseases.

References

1. Bird A (2002) DNA methylation patterns and epigenetic memory. Genes Dev 16(1):6–21
2. Breiling A, Lyko F (2015) Epigenetic regulatory functions of DNA modifications: 5–methylcytosine and beyond. Epigenetics Chromatin 8:24

3. Hermann A, Goyal R, Jeltsch A (2004) The Dnmt1 DNA-(cytosine-C5)-methyltransferase methylates DNA processively with high preference for hemimethylated target sites. J Biol Chem 279(46):48350–48359
4. Okano M, Bell DW, Haber DA, Li E (1999) DNA methyltransferases Dnmt3a and Dnmt3b are essential for de novo methylation and mammalian development. Cell 99(3):247–257
5. Jones PA, Liang G (2009) Rethinking how DNA methylation patterns are maintained. Nat Rev Genet 10(11):805–811
6. Ostler KR, Davis EM, Payne SL, Gosalia BB, Expósito-Céspedes J, Le Beau MM et al (2007) Cancer cells express aberrant DNMT3B transcripts encoding truncated proteins. Oncogene 26(38):5553–5563
7. Gopalakrishnan S, Van Emburgh BO, Shan J, Su Z, Fields CR, Vieweg J et al (2009) A novel DNMT3B splice variant expressed in tumor and pluripotent cells modulates genomic DNA methylation patterns and displays altered DNA binding. Mol Cancer Res 7(10):1622–1634
8. Wienholz BL, Kareta MS, Moarefi AH, Gordon CA, Ginno PA, Chédin F (2010) DNMT3L modulates significant and distinct flanking sequence preference for DNA methylation by DNMT3A and DNMT3B in vivo. PLoS Genet 6(9):e1001106
9. Goll MG, Kirpekar F, Maggert KA, Yoder JA, Hsieh CL, Zhang X et al (2006) Methylation of tRNAAsp by the DNA methyltransferase homolog Dnmt2. Science 311(5759):395–398
10. Tahiliani M, Koh KP, Shen Y, Pastor WA, Bandukwala H, Brudno Y et al (2009) Conversion of 5-methylcytosine to 5-hydroxymethylcytosine in mammalian DNA by MLL partner TET1. Science 324(5929):930–935
11. Ito S, Shen L, Dai Q, Wu SC, Collins LB, Swenberg JA et al (2011) Tet proteins can convert 5-methylcytosine to 5-formylcytosine and 5-carboxylcytosine. Science 333(6047):1300–1303
12. He YF, Li BZ, Li Z, Liu P, Wang Y, Tang Q et al (2011) Tet-mediated formation of 5-carboxylcytosine and its excision by TDG in mammalian DNA. Science 333(6047): 1303–1307
13. Ko M, An J, Bandukwala HS, Chavez L, Aijö T, Pastor WA et al (2013) Modulation of TET2 expression and 5-methylcytosine oxidation by the CXXC domain protein IDAX. Nature 497(7447):122–126
14. Xu Y, Xu C, Kato A, Tempel W, Abreu JG, Bian C et al (2012) Tet3 CXXC domain and dioxygenase activity cooperatively regulate key genes for Xenopus eye and neural development. Cell 151(6):1200–1213
15. Wu X, Zhang Y (2017) TET-mediated active DNA demethylation: mechanism, function and beyond. Nat Rev Genet 18(9):517–534
16. Deaton AM, Bird A (2011) CpG islands and the regulation of transcription. Genes Dev 25(10): 1010–1022
17. Estécio MR, Issa JP (2011) Dissecting DNA hypermethylation in cancer. FEBS Lett 585(13): 2078–2086
18. Shen L, Kondo Y, Guo Y, Zhang J, Zhang L, Ahmed S et al (2007) Genome-wide profiling of DNA methylation reveals a class of normally methylated CpG island promoters. PLoS Genet 3(10):2023–2036
19. Bruniquel D, Schwartz RH (2003) Selective, stable demethylation of the interleukin-2 gene enhances transcription by an active process. Nat Immunol 4(3):235–240
20. Wilson VL, Jones PA (1983) DNA methylation decreases in aging but not in immortal cells. Science 220(4601):1055–1057
21. Wilson VL, Smith RA, Ma S, Cutler RG (1987) Genomic 5-methyldeoxycytidine decreases with age. J Biol Chem 262(21):9948–9951
22. Fuke C, Shimabukuro M, Petronis A, Sugimoto J, Oda T, Miura K et al (2004) Age related changes in 5-methylcytosine content in human peripheral leukocytes and placentas: an HPLC-based study. Ann Hum Genet 68(Pt 3):196–204
23. Issa JP, Ottaviano YL, Celano P, Hamilton SR, Davidson NE, Baylin SB (1994) Methylation of the oestrogen receptor CpG island links ageing and neoplasia in human colon. Nat Genet 7(4):536–540

24. Ahuja N, Li Q, Mohan AL, Baylin SB, Issa JP (1998) Aging and DNA methylation in colorectal mucosa and cancer. Cancer Res 58(23):5489–5494
25. Toyota M, Ahuja N, Ohe-Toyota M, Herman JG, Baylin SB, Issa JP (1999) CpG island methylator phenotype in colorectal cancer. Proc Natl Acad Sci U S A 96(15):8681–8686
26. Waki T, Tamura G, Sato M, Motoyama T (2003) Age-related methylation of tumor suppressor and tumor-related genes: an analysis of autopsy samples. Oncogene 22(26):4128–4133
27. Maegawa S, Hinkal G, Kim HS, Shen L, Zhang L, Zhang J et al (2010) Widespread and tissue specific age-related DNA methylation changes in mice. Genome Res 20(3):332–340
28. Hernandez DG, Nalls MA, Gibbs JR, Arepalli S, van der Brug M, Chong S et al (2011) Distinct DNA methylation changes highly correlated with chronological age in the human brain. Hum Mol Genet 20(6):1164–1172
29. Rakyan VK, Down TA, Maslau S, Andrew T, Yang TP, Beyan H et al (2010) Human aging-associated DNA hypermethylation occurs preferentially at bivalent chromatin domains. Genome Res 20(4):434–439
30. Grönniger E, Weber B, Heil O, Peters N, Stäb F, Wenck H et al (2010) Aging and chronic sun exposure cause distinct epigenetic changes in human skin. PLoS Genet 6(5):e1000971
31. Hannum G, Guinney J, Zhao L, Zhang L, Hughes G, Sadda S et al (2013) Genome-wide methylation profiles reveal quantitative views of human aging rates. Mol Cell 49(2):359–367
32. Panjarian S, Madzo J, Keith K, Slater CM, Sapienza C, Jelinek J et al (2021) Accelerated aging in normal breast tissue of women with breast cancer. Breast Cancer Res 23(1):58
33. Issa J-P (2014) Aging and epigenetic drift: a vicious cycle. J Clin Invest 124(1):24–29
34. Maegawa S, Lu Y, Tahara T, Lee JT, Madzo J, Liang S et al (2017) Caloric restriction delays age-related methylation drift. Nat Commun 8(1):539
35. Wilkinson GS, Adams DM, Haghani A, Lu AT, Zoller J, Breeze CE et al (2021) DNA methylation predicts age and provides insight into exceptional longevity of bats. Nat Commun 12(1):1615
36. Horvath S (2013) DNA methylation age of human tissues and cell types. Genome Biol 14(10): R115
37. Weidner C, Lin Q, Koch C, Eisele L, Beier F, Ziegler P et al (2014) Aging of blood can be tracked by DNA methylation changes at just three CpG sites. Genome Biol 15(2):R24
38. Lin Q, Weidner CI, Costa IG, Marioni RE, Ferreira MRP, Deary IJ et al (2016) DNA methylation levels at individual age-associated CpG sites can be indicative for life expectancy. Aging 8(2):394–401
39. Bocklandt S, Lin W, Sehl ME, Sánchez FJ, Sinsheimer JS, Horvath S et al (2011) Epigenetic Predictor of Age. PLoS One 6(6):e14821
40. Koch CM, Wagner W (2011) Epigenetic-aging-signature to determine age in different tissues. Aging 3(10):1018–1027
41. Bell CG, Lowe R, Adams PD, Baccarelli AA, Beck S, Bell JT et al (2019) DNA methylation aging clocks: challenges and recommendations. Genome Biol 20(1):249
42. Yang Z, Wong A, Kuh D, Paul DS, Rakyan VK, Leslie RD et al (2016) Correlation of an epigenetic mitotic clock with cancer risk. Genome Biol 17(1):205
43. Teschendorff AE (2020) A comparison of epigenetic mitotic-like clocks for cancer risk prediction. Genome Med 12(1):56
44. Levine ME, Lu AT, Quach A, Chen BH, Assimes TL, Bandinelli S et al (2018) An epigenetic biomarker of aging for lifespan and healthspan. Aging 10(4):573–591
45. Chen BH, Marioni RE, Colicino E, Peters MJ, Ward-Caviness CK, Tsai P-C et al (2016) DNA methylation-based measures of biological age: meta-analysis predicting time to death. Aging 8(9):1844–1865
46. Belsky DW, Caspi A, Arseneault L, Baccarelli A, Corcoran DL, Gao X et al (2020) Quantification of the pace of biological aging in humans through a blood test, the DunedinPoAm DNA methylation algorithm. elife 9:e54870
47. Lu AT, Quach A, Wilson JG, Reiner AP, Aviv A, Raj K et al (2019) DNA methylation GrimAge strongly predicts lifespan and healthspan. Aging 11(2):303–327

48. Polanowski AM, Robbins J, Chandler D, Jarman SN (2014) Epigenetic estimation of age in humpback whales. Mol Ecol Resour 14(5):976–987
49. Thompson MJ, Vonholdt B, Horvath S, Pellegrini M (2017) An epigenetic aging clock for dogs and wolves. Aging 9(3):1055–1068
50. Petkovich DA, Podolskiy DI, Lobanov AV, Lee S-G, Miller RA, Gladyshev VN (2017) Using DNA methylation profiling to evaluate biological age and longevity interventions. Cell Metab 25(4):954–960
51. Stubbs TM, Bonder MJ, Stark A-K, Krueger F, Von Meyenn F, Stegle O et al (2017) Multi-tissue DNA methylation age predictor in mouse. Genome Biol 18(1):68
52. Wang T, Tsui B, Kreisberg JF, Robertson NA, Gross AM, Yu MK et al (2017) Epigenetic aging signatures in mice livers are slowed by dwarfism, calorie restriction and rapamycin treatment. Genome Biol 18(1):57
53. Vidal-Bralo L, Lopez-Golan Y, Gonzalez A (2016) Simplified assay for epigenetic age estimation in whole blood of adults. Front Genet 7:126
54. Zhang Y, Wilson R, Heiss J, Breitling LP, Saum K-U, Schöttker B et al (2017) DNA methylation signatures in peripheral blood strongly predict all-cause mortality. Nat Commun 8(1):14617
55. Youn A, Wang S (2018) The MiAge Calculator: a DNA methylation-based mitotic age calculator of human tissue types. Epigenetics 13(2):192–206
56. Levine ME, Lu AT, Chen BH, Hernandez DG, Singleton AB, Ferrucci L et al (2016) Menopause accelerates biological aging. PNAS 113(33):9327–9332
57. Zhou W, Dinh HQ, Ramjan Z, Weisenberger DJ, Nicolet CM, Shen H et al (2018) DNA methylation loss in late-replicating domains is linked to mitotic cell division. Nat Genet 50(4): 591–602
58. Martino D, Loke Y, Gordon L, Ollikainen M, Cruickshank MN, Saffery R et al (2013) Longitudinal, genome-scale analysis of DNA methylation in twins from birth to 18 months of age reveals rapid epigenetic change in early life and pair-specific effects of discordance. Genome Biol 14(5):R42
59. Alisch RS, Barwick BG, Chopra P, Myrick LK, Satten GA, Conneely KN et al (2012) Age-associated DNA methylation in pediatric populations. Genome Res 22(4):623–632
60. Marioni RE, Suderman M, Chen BH, Horvath S, Bandinelli S, Morris T et al (2019) Tracking the epigenetic clock across the human life course: a meta-analysis of longitudinal cohort data. J Gerontol: Series A 74(1):57–61
61. Gross M, Andrew JA, Philipp KF, Jason LK, Jepsen L, Kristen KM et al (2016) Methylome-wide analysis of chronic HIV infection reveals five-year increase in biological age and epigenetic targeting of HLA. Mol Cell 62(2):157–168
62. Horvath S, Levine AJ (2015) HIV-1 infection accelerates age according to the epigenetic clock. J Infect Dis 212(10):1563–1573
63. Levine AJ, Quach A, Moore DJ, Achim CL, Soontornniyomkij V, Masliah E et al (2016) Accelerated epigenetic aging in brain is associated with pre-mortem HIV-associated neurocognitive disorders. J Neuro-Oncol 22(3):366–375
64. Loomba R, Gindin Y, Jiang Z, Lawitz E, Caldwell S, Djedjos CS et al (2018) DNA methylation signatures reflect aging in patients with nonalcoholic steatohepatitis. JCI Insight 3(2): e96685
65. Horvath S, Ritz BR (2015) Increased epigenetic age and granulocyte counts in the blood of Parkinson's disease patients. Aging 7(12):1130–1142
66. Perna L, Zhang Y, Mons U, Holleczek B, Saum K-U, Brenner H (2016) Epigenetic age acceleration predicts cancer, cardiovascular, and all-cause mortality in a German case cohort. Clin Epigenetics 8(1):64
67. Levine ME, Hosgood HD, Chen B, Absher D, Assimes T, Horvath S (2015) DNA methylation age of blood predicts future onset of lung cancer in the women's health initiative. Aging 7(9): 690–700

68. Zheng Y, Joyce BT, Colicino E, Liu L, Zhang W, Dai Q et al (2016) Blood epigenetic age may predict cancer incidence and mortality. EBioMedicine 5:68–73
69. Ambatipudi S, Horvath S, Perrier F, Cuenin C, Hernandez-Vargas H, Le Calvez-Kelm F et al (2017) DNA methylome analysis identifies accelerated epigenetic ageing associated with postmenopausal breast cancer susceptibility. Eur J Cancer 75:299–307
70. Durso DF, Bacalini MG, Sala C, Pirazzini C, Marasco E, Bonafé M et al (2017) Acceleration of leukocytes' epigenetic age as an early tumor and sex-specific marker of breast and colorectal cancer. Oncotarget 8(14):23237–23245
71. Maierhofer A, Flunkert J, Oshima J, Martin GM, Haaf T, Horvath S (2017) Accelerated epigenetic aging in Werner syndrome. Aging 9(4):1143–1152
72. Horvath S, Garagnani P, Bacalini MG, Pirazzini C, Salvioli S, Gentilini D et al (2015) Accelerated epigenetic aging in Down syndrome. Aging Cell 14(3):491–495
73. Horvath S, Langfelder P, Kwak S, Aaronson J, Rosinski J, Vogt TF et al (2016) Huntington's disease accelerates epigenetic aging of human brain and disrupts DNA methylation levels. Aging 8(7):1485–1512
74. Marioni RE, Shah S, Mcrae AF, Ritchie SJ, Muniz-Terrera G, Harris SE et al (2015) The epigenetic clock is correlated with physical and cognitive fitness in the Lothian Birth Cohort 1936. Int J Epidemiol 44(4):1388–1396
75. Raina A, Zhao X, Grove ML, Bressler J, Gottesman RF, Guan W et al (2017) Cerebral white matter hyperintensities on MRI and acceleration of epigenetic aging: the atherosclerosis risk in communities study. Clin Epigenetics 9(1):21
76. Binder AM, Corvalan C, Mericq V, Pereira A, Santos JL, Horvath S et al (2018) Faster ticking rate of the epigenetic clock is associated with faster pubertal development in girls. Epigenetics 13(1):85–94
77. Horvath S, Erhart W, Brosch M, Ammerpohl O, Von Schonfels W, Ahrens M et al (2014) Obesity accelerates epigenetic aging of human liver. PNAS 111(43):15538–15543
78. Nevalainen T, Kananen L, Marttila S, Jylhävä J, Mononen N, Kähönen M et al (2017) Obesity accelerates epigenetic aging in middle-aged but not in elderly individuals. Clin Epigenetics 9(1):20
79. Quach A, Levine ME, Tanaka T, Lu AT, Chen BH, Ferrucci L et al (2017) Epigenetic clock analysis of diet, exercise, education, and lifestyle factors. Aging 9(2):419–446
80. Carroll JE, Irwin MR, Levine M, Seeman TE, Absher D, Assimes T et al (2017) Epigenetic aging and immune senescence in women with insomnia symptoms: findings from the women's health initiative study. Biol Psychiatry 81(2):136–144
81. Zannas AS, Arloth J, Carrillo-Roa T, Iurato S, Röh S, Ressler KJ et al (2015) Lifetime stress accelerates epigenetic aging in an urban, African American cohort: relevance of glucocorticoid signaling. Genome Biol 16(1):61
82. Jovanovic T, Vance LA, Cross D, Knight AK, Kilaru V, Michopoulos V et al (2017) Exposure to violence accelerates epigenetic aging in children. Sci Rep 7(1):8962
83. Christiansen L, Lenart A, Tan Q, Vaupel JW, Aviv A, Mcgue M et al (2016) DNA methylation age is associated with mortality in a longitudinal Danish twin study. Aging Cell 15(1):149–154
84. Marioni RE, Shah S, Mcrae AF, Chen BH, Colicino E, Harris SE et al (2015) DNA methylation age of blood predicts all-cause mortality in later life. Genome Biol 16(1):25
85. Horvath S, Pirazzini C, Bacalini MG, Gentilini D, Di Blasio AM, Delledonne M et al (2015) Decreased epigenetic age of PBMCs from Italian semi-supercentenarians and their offspring. Aging 7(12):1159–1170
86. Fahy GM, Brooke RT, Watson JP, Good Z, Vasanawala SS, Maecker H et al (2019) Reversal of epigenetic aging and immunosenescent trends in humans. Aging Cell 18(6):e13028
87. Aran D, Camarda R, Odegaard J, Paik H, Oskotsky B, Krings G et al (2017) Comprehensive analysis of normal adjacent to tumor transcriptomes. Nat Commun 8(1):1077
88. Russi C, Ruggieri L, Rocca A et al (2019) Gastric normal adjacent mucosa versus healthy and cancer tissues: distinctive transcriptomic profiles and biological features. Cancer 11(9):1248

89. Teschendorff AE, Gao Y, Jones A, Ruebner M, Beckmann MW, Wachter DL et al (2016) DNA methylation outliers in normal breast tissue identify field defects that are enriched in cancer. Nat Commun 7:10478
90. Zou H, Hastie T (2005) Regularization and variable selection via the elastic net. J R Stat Soc Ser B Stat Methodol 67(2):301–320
91. Kacmarczyk TJ, Fall MP, Zhang X, Xin Y, Li Y, Alonso A et al (2018) "Same difference": comprehensive evaluation of four DNA methylation measurement platforms. Epigenetics Chromatin 11(1):21
92. El Khoury LY, Gorrie-Stone T, Smart M, Hughes A, Bao Y, Andrayas A et al (2019) Systematic underestimation of the epigenetic clock and age acceleration in older subjects. Genome Biol 20(1):283
93. Zhang Q, Vallerga CL, Walker RM, Lin T, Henders AK, Montgomery GW et al (2019) Improved precision of epigenetic clock estimates across tissues and its implication for biological ageing. Genome Med 11(1):54
94. Slieker RC, Relton CL, Gaunt TR, Slagboom PE, Heijmans BT (2018) Age-related DNA methylation changes are tissue-specific with ELOVL2 promoter methylation as exception. Epigenetics Chromatin 11(1):25
95. Fransquet PD, Wrigglesworth J, Woods RL, Ernst ME, Ryan J (2019) The epigenetic clock as a predictor of disease and mortality risk: a systematic review and meta-analysis. Clin Epigenetics 11(1):62
96. Teschendorff AE, Jones A, Widschwendter M (2016) Stochastic epigenetic outliers can define field defects in cancer. BMC Bioinform 17:178
97. Ghosh J, Schultz B, Coutifaris C, Sapienza C (2019) Highly variant DNA methylation in normal tissues identifies a distinct subclass of cancer patients. Adv Cancer Res 142:1–22
98. Ghosh J, Mainigi M, Coutifaris C, Sapienza C (2016) Outlier DNA methylation levels as an indicator of environmental exposure and risk of undesirable birth outcome. Hum Mol Genet 25(1):123–129
99. Xing M, Zhugiang Z, Zhuoning Z, Cong L, Qiang D, Qixiang H et al (2020) Kinetics and mechanisms of mitotic inheritance of DNA methylation and their roles in aging-associated methylome deterioration. Cell Res 30:980–996
100. Shih AH, Abdel-Wahab O, Patel JP, Levine RL (2012) The role of mutations in epigenetic regulators in myeloid malignancies. Nat Rev Cancer 12:599–612
101. Jia Y, Li P, Fang L, Zhu H, Xu L, Cheng H et al (2016) Negative regulation of DNMT3A de novo DNA methylation by frequently overexpressed UHRF family proteins as a mechanism for widespread DNA hypomethylation in cancer. Cell Discov 2(1):16007
102. Ronn T, Volkov P, Gillberg L, Kokosar M, Perfilyev A, Jacobsen AL et al (2015) Impact of age, BMI and HbA1c levels on the genome-wide DNA methylation and mRNA expression patterns in human adipose tissue and identification of epigenetic biomarkers in blood. Hum Mol Genet 24(13):3792–3813
103. Casillas MA, Lopatina N, Andrews LG, Tollefsbol TO (2003) Transcriptional control of the DNA methyltransferases is altered in aging and neoplastically-transformed human fibroblasts. Mol Cell Biochem 252:33–43
104. Berger SL, Sassone-Corsi P (2016) Metabolic signaling to chromatin. Cold Spring Harb Perspect Biol 8(11):a019463
105. Veland N, Hardikar S, Zhong Y, Gayatri S, Dan J, Strahl BD et al (2017) The arginine Methyltransferase PRMT6 regulates DNA methylation and contributes to global DNA hypomethylation in cancer. Cell Rep 21(12):3390–3397
106. Cruickshanks HA, Mcbryan T, Nelson DM, Vanderkraats ND, Shah PP, Van Tuyn J et al (2013) Senescent cells harbour features of the cancer epigenome. Nat Cell Biol 15(12): 1495–1506
107. Charlton J, Downing TL, Smith ZD, Gu H, Clement K, Pop R et al (2018) Global delay in nascent strand DNA methylation. Nat Struct Mol Biol 25(4):327–332

108. Liang G, Chan MF, Tomigahara Y, Tsai YC, Gonzales FA, Li E et al (2002) Cooperativity between DNA methyltransferases in the maintenance methylation of repetitive elements. Mol Cell Biol 22:480–491
109. Zohar S, Zohar M, Mendelson CN, Gilad L, Elad C, Reich ZS et al (2014) Dynamic and static maintenance of epigenetic memory in pluripotent and somatic cells. Nature 513(7516): 115–119
110. Valinluck V, Sowers LC (2007) Endogenous cytosine damage products alter the site selectivity of Human DNA maintenance methyltransferase DNMT1. Cancer Res 67(3):946–950
111. Hashimoto H, Liu Y, Upadhyay AK, Chang Y, Howerton SB, Vertino PM et al (2012) Recognition and potential mechanisms for replication and erasure of cytosine hydroxymethylation. Nucleic Acids Res 40(11):4841–4849
112. Bender CM, Pao MM, Jones PA (1998) Inhibition of DNA Methylation by 5-Aza-2-′-deoxycytidine suppresses the growth of human tumor cell lines. Cancer Res 58:95–101
113. Schubeler D (2015) Function and information content of DNA methylation. Nature 517:321–326
114. Feldmann A, Ivanek R, Murr R, Gaidatzis D, Burger L, Schübeler D (2013) Transcription Factor occupancy can mediate active turnover of DNA methylation at regulatory regions. PLoS Genet 9(12):e1003994
115. Suzuki T, Maeda S, Furuhata E, Shimizu Y, Nishimura H, Kishima M et al (2017) A screening system to identify transcription factors that induce binding site-directed DNA demethylation. Epigenetics Chromatin 10(1):60
116. Mayran A, Drouin J (2018) Pioneer transcription factors shape the epigenetic landscape. J Biol Chem 293(36):13795–13804
117. Stadler MB, Murr R, Burger L, Ivanek R, Lienert F, Scholer A et al (2011) DNA-binding factors shape the mouse methylome at distal regulatory regions. Nature 480:490–495
118. Reimer M, Pulakanti K, Shi L, Abel A, Liang M, Malarkannan S et al (2019) Deletion of Tet proteins results in quantitative disparities during ESC differentiation partially attributable to alterations in gene expression. BMC Dev Biol 19(1):16
119. Dawlaty M, Meelad BA, Le T, Barrasa IM, Raddatz G, Gao Q et al (2014) Loss of tet enzymes compromises proper differentiation of embryonic stem cells. Dev Cell 29(1):102–111
120. Li T, Wang L, Du Y, Xie S, Yang X, Lian F et al (2018) Structural and mechanistic insights into UHRF1-mediated DNMT1 activation in the maintenance DNA methylation. Nucleic Acids Res 46(6):3218–3231
121. Zhang H, Gao Q, Tan S, You J, Lyu C, Zhang Y et al (2019) SET8 prevents excessive DNA methylation by methylation-mediated degradation of UHRF1 and DNMT1. Nucleic Acids Res 47(17):9053–9068
122. Dozmorov MG (2015) Polycomb repressive complex 2 epigenomic signature defines age-associated hypermethylation and gene expression changes. Epigenetics 10(6):484–495
123. Viré E, Brenner C, Deplus R, Blanchon L, Fraga M, Didelot C et al (2006) The Polycomb group protein EZH2 directly controls DNA methylation. Nature 439(7078):871–874
124. Guo S, Diep D, Plongthongkum N, Fung H-L, Zhang K, Zhang K (2017) Identification of methylation haplotype blocks aids in deconvolution of heterogeneous tissue samples and tumor tissue-of-origin mapping from plasma DNA. Nat Genet 49(4):635–642
125. Busto-Moner L, Morival J, Ren H, Fahim A, Reitz Z, Downing TL et al (2020) Stochastic modeling reveals kinetic heterogeneity in post-replication DNA methylation. PLoS Comput Biol 16(4):e1007195
126. Wang Q, Yu G, Ming X, Xia W, Xu X, Zhang Y et al (2020) Imprecise DNMT1 activity coupled with neighbor-guided correction enables robust yet flexible epigenetic inheritance. Nat Genet 52:828–839
127. Ming X, Zhu B, Li Y (2021) Mitotic inheritance of DNA methylation: more than just copy and paste. J Genet Genomics 48(1):1–13
128. Tollefsbol TO, Hutchison CA III (1997) Control of methylation spread in synthetic DNA sequences by the murine DNA methyltransferase. J Mol Biol 269(4):494–504

129. Vilkaitis G, Suetake I, Klimašauskas S, Tajima S (2005) Processive methylation of hemimethylated CpG sites by mouse Dnmt1 DNA methyltransferase. J Biol Chem 280(1): 64–72
130. Issa JP, Ahuja N, Toyota M, Bronner MP, Brentnall TA (2001) Accelerated age-related CpG island methylation in ulcerative colitis. Cancer Res 61(9):3573–3577
131. Eads CA, Lord RV, Wickramasinghe K, Long TI, Kurumboor SK, Bernstein L et al (2001) Epigenetic patterns in the progression of esophageal adenocarcinoma. Cancer Res 61(8): 3410–3418
132. Maekita T, Nakazawa K, Mihara M, Nakajima T, Yanaoka K, Iguchi M et al (2006) High levels of aberrant DNA methylation in Helicobacter pylori-infected gastric mucosae and its possible association with gastric cancer risk. Clin Cancer Res 12(3 Pt 1):989–995
133. Shen L, Ahuja N, Shen Y, Habib NA, Toyota M, Rashid A et al (2002) DNA methylation and environmental exposures in human hepatocellular carcinoma. J Natl Cancer Inst 94(10): 755–761
134. Cooke J, Zhang H, Greger L, Silva AL, Massey D, Dawson C et al (2012) Mucosal genome-wide methylation changes in inflammatory bowel disease. Inflamm Bowel Dis 18(11): 2128–2137
135. Nimmo ER, Prendergast JG, Aldhous MC, Kennedy NA, Henderson P, Drummond HE et al (2012) Genome-wide methylation profiling in Crohn's disease identifies altered epigenetic regulation of key host defense mechanisms including the Th17 pathway. Inflamm Bowel Dis 18(5):889–899
136. Niwa T, Tsukamoto T, Toyoda T, Mori A, Tanaka H, Maekita T et al (2010) Inflammatory processes triggered by Helicobacter pylori infection cause aberrant DNA methylation in gastric epithelial cells. Cancer Res 70(4):1430–1440
137. Christensen BC, Houseman EA, Marsit CJ, Zheng S, Wrensch MR, Wiemels JL et al (2009) Aging and environmental exposures alter tissue-specific DNA methylation dependent upon CpG island context. PLoS Genet 5(8):e1000602
138. Fiorito G, McCrory C, Robinson O, Carmeli C, Rosales CO, Zhang Y et al (2019) Socioeconomic position, lifestyle habits and biomarkers of epigenetic aging: a multi-cohort analysis. Aging (Albany NY) 11(7):2045–2070
139. Fiorito G, Guarrera S, Valle C, Ricceri F, Russo A, Grioni S et al (2014) B-vitamins intake, DNA-methylation of One Carbon Metabolism and homocysteine pathway genes and myocardial infarction risk: the EPICOR study. Nutr Metab Cardiovasc Dis 24(5):483–488
140. Irwin RE, Thursby SJ, Ondičová M, Pentieva K, McNulty H, Richmond RC et al (2019) A randomized controlled trial of folic acid intervention in pregnancy highlights a putative methylation-regulated control element at ZFP57. Clin Epigenetics 11(1):31
141. Caffrey A, Irwin RE, McNulty H, Strain JJ, Lees-Murdock DJ, McNulty BA et al (2018) Gene-specific DNA methylation in newborns in response to folic acid supplementation during the second and third trimesters of pregnancy: epigenetic analysis from a randomized controlled trial. Am J Clin Nutr 107(4):566–575
142. Luebeck GE, Hazelton WD, Curtius K, Maden SK, Yu M, Carter KT et al (2019) Implications of epigenetic drift in colorectal neoplasia. Cancer Res 79(3):495–504
143. Slieker RC, Van Iterson M, Luijk R, Beekman M, Zhernakova DV, Moed MH et al (2016) Age-related accrual of methylomic variability is linked to fundamental ageing mechanisms. Genome Biol 17(1):191
144. Gentilini D, Garagnani P, Pisoni S, Bacalini MG, Calzari L, Mari D et al (2015) Stochastic epigenetic mutations (DNA methylation) increase exponentially in human aging and correlate with X chromosome inactivation skewing in females. Aging 7(8):568–578
145. Wang Y, Karlsson R, Jylhävä J, Hedman ÅK, Almqvist C, Karlsson IK et al (2019) Comprehensive longitudinal study of epigenetic mutations in aging. Clin Epigenetics 11(1):187
146. Pirazzini C, Giuliani C, Bacalini MG, Boattini A, Capri M, Fontanesi E et al (2012) Space/Population and Time/Age in DNA methylation variability in humans: a study on IGF2/H19

locus in different Italian populations and in mono- and di-zygotic twins of different age. Aging 4(7):509–520
147. Bennett-Baker PE, Wilkowski J, Burke DT (2003) Age-associated activation of epigenetically repressed genes in the mouse. Genetics 165(4):2055–2062
148. Meir Z, Mukamel Z, Chomsky E, Lifshitz A, Tanay A (2020) Single-cell analysis of clonal maintenance of transcriptional and epigenetic states in cancer cells. Nat Genet 52(7):709–718
149. Xu J, Shi J, Cui X, Cui Y, Li JJ, Goel A et al (2021) Cellular Heterogeneity–Adjusted cLonal Methylation (CHALM) improves prediction of gene expression. Nat Commun 12(1):40
150. Yin Y, Morgunova E, Jolma A, Kaasinen E, Sahu B, Khund-Sayeed S et al (2017) Impact of cytosine methylation on DNA binding specificities of human transcription factors. Science 356(6337):eaaj2239
151. De Cecco M, Criscione SW, Peckham EJ, Hillenmeyer S, Hamm EA, Manivannan J et al (2013) Genomes of replicatively senescent cells undergo global epigenetic changes leading to gene silencing and activation of transposable elements. Aging Cell 12(2):247–256
152. De Cecco M, Criscione SW, Peterson AL, Neretti N, Sedivy JM, Kreiling JA (2013) Transposable elements become active and mobile in the genomes of aging mammalian somatic tissues. Aging 5(12):867–883
153. Van Meter M, Kashyap M, Rezazadeh S, Geneva AJ, Morello TD, Seluanov A et al (2014) SIRT6 represses LINE1 retrotransposons by ribosylating KAP1 but this repression fails with stress and age. Nat Commun 5(1):5011
154. Gasior SL, Wakeman TP, Xu B, Deininger PL (2006) The Human LINE-1 Retrotransposon creates DNA double-strand breaks. J Mol Biol 357(5):1383–1393
155. Gilbert N, Lutz-Prigge S, Moran JV (2002) Genomic Deletions Created upon LINE-1 Retrotransposition. Cell 110(3):315–325
156. Iskow RC, Mccabe MT, Mills RE, Torene S, Pittard WS, Neuwald AF et al (2010) Natural mutagenesis of human genomes by endogenous retrotransposons. Cell 141(7):1253–1261
157. Ablasser A, Hemmerling I, Schmid-Burgk JL, Behrendt R, Roers A, Hornung V (2014) TREX1 Deficiency Triggers cell-autonomous immunity in a cGAS-dependent manner. J Immunol 192(12):5993–5997
158. Oberdoerffer P, Michan S, Mcvay M, Mostoslavsky R, Vann J, Park S-K et al (2008) SIRT1 Redistribution on chromatin promotes genomic stability but alters gene expression during aging. Cell 135(5):907–918
159. Thomas CA, Tejwani L, Trujillo CA, Negraes PD, Herai RH, Mesci P et al (2017) Modeling of TREX1-dependent autoimmune disease using human stem cells highlights l1 accumulation as a source of neuroinflammation. Cell Stem Cell 21(3):319–331
160. Snippert HJ, van der Flier LG, Sato T, van Es JH, van den Born M, Kroon-Veenboer C et al (2010) Intestinal crypt homeostasis results from neutral competition between symmetrically dividing Lgr5 stem cells. Cell 143(1):134–144
161. Klein AM, Nakagawa T, Ichikawa R, Yoshida S, Simons BD (2010) Mouse germ line stem cells undergo rapid and stochastic turnover. Cell Stem Cell 7(2):214–224
162. Sun D, Luo M, Jeong M, Rodriguez B, Xia Z, Hannah R et al (2014) Epigenomic profiling of young and age HSCs reveals concerted changes during aging that reinforce self-renewal. Cell Stem Cell 14(5):673–688
163. Tierney MT, Stec MJ, Rulands S, Simons BD, Sacco A (2018) Muscle stem cells exhibit distinct clonal dynamics in response to tissue repair and homeostatic aging. Cell Stem Cell 22(1):119–127
164. Gravina S, Dong X, Yu B, Vijg J (2016) Single-cell genome-wide bisulfite sequencing uncovers extensive heterogeneity in the mouse liver methylome. Genome Biol 17(1):150
165. Hernando-Herraez I, Evano B, Stubbs T, Commere P-H, Jan Bonder M, Clark S et al (2019) Ageing affects DNA methylation drift and transcriptional cell-to-cell variability in mouse muscle stem cells. Nat Commun 10(1):4361

166. Landan G, Cohen NM, Mukamel Z, Bar A, Molchadsky A, Brosh R et al (2012) Epigenetic polymorphism and the stochastic formation of differentially methylated regions in normal and cancerous tissues. Nat Genet 44(11):1207–1214
167. Landau A, Dan CK, Ziller J, Michael BP, Fan J, Gu H et al (2014) Locally disordered methylation forms the basis of intratumor methylome variation in chronic lymphocytic leukemia. Cancer Cell 26(6):813–825
168. Li S, Garrett-Bakelman F, Perl AE, Luger SM, Zhang C, To BL et al (2014) Dynamic evolution of clonal epialleles revealed by methclone. Genome Biol 15(9):472
169. Jenkinson G, Abante J, Feinberg AP, Goutsias J (2018) An information-theoretic approach to the modeling and analysis of whole-genome bisulfite sequencing data. BMC Bioinformatics 19(1):87
170. Scherer M, Nebel A, Franke A, Walter J, Lengauer T, Bock C et al (2020) Quantitative comparison of within-sample heterogeneity scores for DNA methylation data. Nucleic Acids Res 48(8):e46
171. Gaiti F, Chaligne R, Gu H, Brand RM, Kothen-Hill S, Schulman RC et al (2019) Epigenetic evolution and lineage histories of chronic lymphocytic leukaemia. Nature 569(7757):576–580
172. Martin-Herranz DE, Aref-Eshghi E, Bonder MJ, Stubbs TM, Choufani S, Weksberg R et al (2019) Screening for genes that accelerate the epigenetic aging clock in humans reveals a role for the H3K36 methyltransferase NSD1. Genome Biol 20(1):146

Epigenetic Clocks

11

Alexandra M. Binder and Steve Horvath

Contents

Abstract

Epigenetic clocks are DNA methylation-based estimators of chronological age or mortality risk that have been integrated into a wide array of epidemiologic studies. In this chapter, we review fundamental considerations in the development

A. M. Binder (✉)
Population Sciences in the Pacific Program, University of Hawaii Cancer Center, Honolulu, HI, USA

Department of Epidemiology, Fielding School of Public Health, University of California, Los Angeles, CA, USA
e-mail: abinder@cc.hawaii.edu

S. Horvath
Department of Human Genetics, David Geffen School of Medicine, University of California, Los Angeles, CA, USA

Department of Biostatistics, Fielding School of Public Health, University of California, Los Angeles, CA, USA
e-mail: shorvath@mednet.ucla.edu

K. B. Michels (ed.), *Epigenetic Epidemiology*,
https://doi.org/10.1007/978-3-030-94475-9_11

of epigenetic clocks, including the features of aging upon which they were trained, and the characteristics of the training dataset. We discuss the implications of these decisions on the relation of epigenetic clocks with morbidity and mortality across populations. We review the assumptions underlying the incorporation of epigenetic clocks into human studies of disease etiology, as well as distinct analytic considerations when the study objective is to appraise the value of epigenetic clocks as risk predictors. We expand on a few key characteristics expected of candidate biomarkers proposed for integration into a clinical setting. Throughout this chapter, we highlight the extensive epidemiologic literature on epigenetic clocks to date, and opportunities to expand on the current research.

Abbreviations

DNAm	DNA methylation
epiTOC	Epigenetic Timer of Cancer clocks
MiAge	Mitotic Age clock
IEAA	Intrinsic epigenetic age acceleration
EEAA	Extrinsic epigenetic age acceleration
PedBE	Pediatric-Buccal-Epigenetic clock
MR	Mendelian Randomization

11.1 Developing an Epigenetic Estimator of Biologic Aging

The term "epigenetic clock" has been used to refer to multivariate regression models that estimate age or mortality risk based on specific DNA methylation (DNAm) signatures. As such, a clock is defined by a set of CpG loci and coefficient values. Many epigenetic clock algorithms have been published to date, and the number continues to expand as new clocks are developed for distinct purposes. Some clocks apply to a specific source of DNA (e.g., saliva), while others apply to all sources of DNA (e.g., pan-tissue clocks). Epigenetic clocks vary based on the features of aging upon which they were trained (e.g., telomere length), as well as the characteristics of the training dataset (e.g., pediatric samples). In this section, we describe some of the considerations in epigenetic clock development, which influence the relation between each measure and aging-related conditions, as well as mortality rate.

11.1.1 Choice of Outcome

A defining aspect of an epigenetic clock is the characteristic upon which it was trained. The first generation of epigenetic clocks was built to predict chronologic age based on DNAm levels at specific CpG loci. The residual variation in DNAm-

estimated age (i.e., "epigenetic age"), independent of chronologic age, was proposed to serve as an indicator of biologic aging. In this framework, those with a DNAm-estimated age higher than their chronologic age are described to have an older ("accelerated") epigenetic age, whereas those with a relatively lower DNAm-estimated age have a younger ("decelerated") epigenetic age. Since this first generation of clocks, additional variants have been developed to predict mortality or morbidity risk (e.g., GrimAge). These clocks also give rise to estimates of epigenetic age acceleration that measure the discrepancy between predicted mortality risk relative to that expected on the basis of chronologic age.

11.1.1.1 Chronologic Age

Training epigenetic clocks on chronologic age is a logical choice given its strong correlation with mortality/morbidity risk and biologic age (Fig. 11.1). While several clocks have been developed based on chronologic age, they vary in how CpG loci were prioritized for inclusion into the clock algorithm, and the types of datasets in which the algorithms were trained [1–10]. Regularized regression models, such as elastic net or lasso, are commonly used to select CpG loci for inclusion in these prediction algorithms. Given the relatively high degree of shared variance across the methylome, a major benefit of this approach is the capacity to accommodate a large amount of collinearity in the prediction model. With the goal to minimize prediction

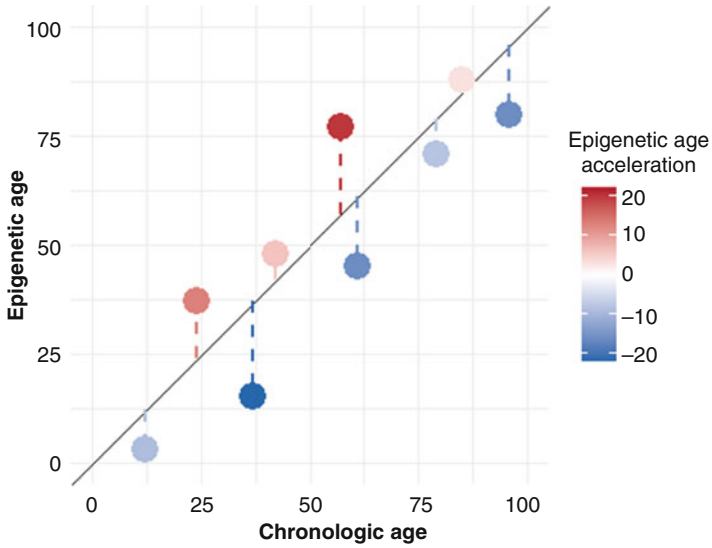

Fig. 11.1 Evaluating relative biologic age based on an epigenetic clock. DNAm-estimated age (i.e., "epigenetic age") is plotted against chronologic age, with a line of best fit between the two age measures. The dashed line represents the difference between observed epigenetic age and predicted epigenetic age among individuals of the same chronologic age; in other words, the residual variation in epigenetic age, independent of chronologic age. Epigenetic age is "accelerated" if it is higher than expected (red), and considered "decelerated" if it is lower than expected (blue)

error, regularized regression introduces bias into the traditional linear regression to decrease the high variance induced by the incorporation of highly correlated variables [11]. This is accomplished by imposing a size constraint on the model coefficients that is controlled by a penalty, the strength of which is modified by a tuning parameter selected to optimize the cross-validated penalized likelihood [11]. The lasso penalty optimizes a sparse solution that tends to select a predictor among highly correlated variables, and is somewhat indifferent to predictor choice [11]. The ridge penalty optimizes a model that shrinks the coefficients of correlated predictors towards each other [11]. The elastic net penalty is a compromise between the lasso and ridge, and encourages balancing the contributions of correlated variables and a sparse solution [11]. Both the Horvath pan-tissue (PanTissue) clock [3] and the Hannum clock [2] were developed by modeling chronologic age as a function of CpG loci interrogated on DNAm arrays using elastic net regression. For other clocks, the CpG loci considered for inclusion into the prediction model were first reduced to a subset based on their site-specific association with chronologic age [4, 10]. The skin and blood clock is an example of this approach, which first estimated the correlation between site-specific DNAm and chronologic age across human fibroblasts, keratinocytes, buccal cells, endothelial cells, lymphoblastoid cells, skin, blood, and saliva samples [4]. An elastic net regression for calibrated chronologic age was then estimated, including the CpG loci with the strongest correlation with chronological age in different cell types, and 500 CpGs with the least significant correlation with age [4]. The Weidner clock similarly first applied a filtering approach to develop a clock that prioritized minimizing the number of CpG loci necessary to predict age [7]. Restricting to CpG loci with an absolute correlation with chronologic age greater than 0.85 in blood, recursive feature elimination was used to inform the selection of three CpG loci to predict age [7]. Several authors have aimed to develop sparse clocks on the basis of a single CpG (e.g., a single CpG in the *ELOVL2* gene [12]) or a handful of CpGs. Sparse clocks are attractive for two main reasons: (1) reduced assay costs and (2) enhanced interpretability with respect to neighboring genes. However, sparse clocks tend to be less robust and accurate than clocks based on larger numbers of CpGs. For all chronologic age-trained clocks, the accuracy of the predictor has some important implications for associations with health outcomes. As demonstrated by Zhang et al. [9], it is feasible to develop a near-perfect age predictor when the training sample size is sufficiently large. However, when the prediction of chronologic age is very accurate, the variation in epigenetic age independent of chronologic age is no longer associated with mortality [9]. Optimally, a chronologic age predictor should characterize average trends in the methylome with age, while accommodating individual variation in that trend.

11.1.1.2 Mitotic Age

An alternative approach to describe the aging process is on the cellular level. A number of epigenetic clocks have been developed to estimate "mitotic age," which is a term used to describe the cumulative number of divisions within the stem cell pool [13–16]. This age is influenced by the intrinsic cell turnover rate of a given tissue, as well as factors that contribute to cellular injury. To estimate mitotic age,

investigators have taken advantage of trends in DNAm maintenance over successive cell divisions. Clocks applying this approach include the HypoClock, which predicts mitotic age based on DNAm loss in late-replicating regions due to incomplete methylation maintenance [16]. The Epigenetic Timer of Cancer (epiTOC) clocks were alternatively developed based CpG-rich regions marked by the polycomb repressive complex-2, which gradually gain methylation over cell divisions [13, 14]. Assuming mitotic age is a predictor of neoplastic potential, the Mitotic Age (MiAge) clock used tumor and adjacent normal tissue samples to develop an estimator based on a model of stochastic errors in DNAm maintenance and de novo DNAm [15]. As expected, each of these mitotic clocks estimate greater cell proliferation among tumor tissue relative to normal tissue [13–16]. Given telomeres undergo progressive degradation with DNA replication, a recently developed DNAm-based estimator of telomere length also falls within the category of mitotic clocks [17]. Leukocyte telomere length measured using Southern blots was predicted based on site-specific DNAm using elastic net regression [17]. This DNAm-based estimator of telomere length was shown to have a stronger association with time-to-death than measured leukocyte telomere length [17]. Analysis of cultured cells with and without telomerase suggested this DNAm-based estimator of telomere length is correlated with cellular proliferation, even independent of telomere attrition [17]. Given the cell-type specificity of cell turnover rates, the interpretation of associations with mitotic clocks should consider the potential influence of cellular heterogeneity among the samples assayed.

11.1.1.3 Time-to-Death and Characteristics Associated with Mortality Rates

Estimation of biologic age can inform an appraisal of an individual's relative risk of an accelerated aging trajectory, characterized by a greater incidence of aging-related chronic conditions, and increased hazard of death. Given this potential application of biologic age estimators, a few epigenetic clocks have been specifically developed to capture variation in the methylome associated with time-to-death and clinical characteristics associated with aging and mortality rates. One of these clocks, Mortality Score, was created by first identifying CpG loci with a genome-wide significant association with mortality rate that replicated in a validation set [18]. Among this subset of sites, a lasso Cox regression was used to select 10 CpG loci that were the most predictive of time-to-death to estimate the DNAm-based Mortality Score [18]. For the PhenoAge clock, Levine et al. [19] began by first creating a composite measure of "phenotypic age" based on characteristics associated with time-to-death. To develop this measure, a penalized Cox regression was used to identify a subset of clinical markers that were the most predictive of all-cause mortality in the nationally representative National Health and Nutrition Examination Survey (NHANES) [19]. Among the 42 clinical markers considered for inclusion, the final subgroup included: albumin, creatinine, glucose, c-reactive protein, lymphocyte percent, mean red cell volume, red cell distribution width, alkaline phosphatase, and white blood cell count, as well as chronologic age [19]. A DNAm-based estimator of this phenotypic age measure was then trained in an independent dataset using elastic net regression [19]. To construct the

DunedinPoAm clock, an estimated "Pace of Aging" was initially developed based on a panel of 18 blood-chemistry and organ-system-function biomarkers measured at ages 26, 32, and 38 years old in the Dunedin Study birth cohort [20]. The Pace of Aging was calculated as an individual's rate of change in these 18 biomarkers relative to the average change across all study participants over time [20]. Accelerated aging based on this measure was associated with less physical ability, greater cognitive decline, and worse self-reported health among Dunedin Study participants [20]. Among the subset of individuals with DNAm assayed at age 38, a DNAm-based estimator of this Pace of Aging was derived using elastic net regression [21]. The GrimAge clock was created using a distinct two-step approach, beginning with the development of DNAm-based estimators of plasma protein levels and smoking pack-years [22]. A penalized Cox regression model was then used to identify the subset of DNAm-based estimators that were the most predictive of time-to-death [22]. GrimAge is calculated as the transformed linear combination of covariates resulting from this model, which included age, sex, DNAm pack-years, and seven DNAm-based surrogate markers of plasma proteins (adrenomedullin, beta-2 microglobulin, cystatin C, growth differentiation factor 15, leptin, plasminogen activator inhibitor-1, tissue inhibitor metalloproteinase 1) [22].

11.1.2 Choice of Tissue

The choice of tissue used to train an epigenetic clock is shaped by three primary factors: study feasibility, the intended ease of assay for the target population, and the tissue-specificity of the aging process the clock is designed to evaluate. For clocks developed based on a composite measure of aging processes (e.g., GrimAge, PhenoAge, and DunedinPoAm), a large, well-defined cohort is required for the training set. Clocks partially trained on time-to-death, such as GrimAge and Mortality Score, require many years of participant follow-up to build a relatively precise predictor. Using samples of convenience, the GrimAge, PhenoAge, Mortality Score, and DunedInPoAm clocks were all trained using peripheral blood, which is the most commonly available stored biospecimen for existing cohorts. Blood is a heterogeneous tissue that can capture distinct patterns of aging-related DNAm variation across several cell types. For biomarkers intended to measure multiple facets of aging, this can be a benefit. To minimize the degree to which blood composition drives associations with epigenetic clocks, investigators can adjust for DNAm-based estimators of blood cell counts and relative abundance [23, 24]. Intrinsic epigenetic age acceleration (IEAA) is an example of this approach, as it estimates PanTissue age adjusting for chronologic age and DNAm-estimated counts of naïve CD8+ T cells, exhausted CD8+ T cells, plasma B cells, CD4+ T cells, natural killer cells, monocytes, and granulocytes [23]. This residual variation is assumed to measure age-related changes in DNAm that are consistent across cell types. When epigenetic clocks are intended to serve as an indicator of risk, removing this variation may have important implications for predictive accuracy. Aging is characterized by changes in the distribution of immune cells, including a decrease in naïve cytotoxic T cells and an increase in memory or exhausted cytotoxic T cells [25, 26]. Estimated extrinsic

epigenetic age acceleration (EEAA) was designed to capture these age-related changes by calculating an age-adjusted weighted average of Hannum epigenetic age and DNAm-estimated blood cell counts of cytotoxic T cell sub-populations and plasma B cells [23]. Highlighting the predictive value of these age-related changes in blood composition, EEAA was found to have a stronger association with all-cause mortality than the PanTissue clock and IEAA in a meta-analysis of 13 cohorts [27]. However, EEAA is inferior to second-generation clocks, such as GrimAge, when it comes to predicting mortality risk.

Epigenetic clocks that perform well in blood samples have high clinical utility given the relative ease of collecting peripheral blood. The drawback to clocks trained exclusively using blood samples is potential poor transportability to other tissues. For pediatric studies and investigations that require multiple assessments of epigenetic age over time, estimation of epigenetic age from buccal swabs and saliva collections is likely more practical. Blood may also not be a strong surrogate for more tissue-specific mechanisms of age-related complications. The PanTissue clock was explicitly built to have broad external validity, incorporating samples from 51 healthy tissues and cell types into its development [3]. The trade-off is that clocks trained for specific target tissues tend to have higher accuracy. For example, the skin and blood clock has a much stronger correlation with chronologic age than the PanTissue clock in buccal swabs [4]. Generally, most tissues exhibit a similar PanTissue age, but there are some exceptions. Notably, average PanTissue age is significantly younger in the cerebellum [28], while female breast tissue is estimated to be much older relative to chronologic age [3, 29]. The PanTissue clock was trained using wide-spectrum tissue types, but it only involved few skeletal muscle samples. A predictor of chronologic age in skeletal muscle was recently developed to address this gap, and found to have a stronger correlation with age than the PanTissue clock in muscle samples [6]. Tissue-specific epigenetic clocks, like the skeletal muscle clock, may provide more nuanced insight into the etiology of age-related conditions, such as frailty. However, clocks for less accessible tissues reduce the feasibility of integrating epigenetic clock estimation into a broad public health screening paradigm. Therefore, the intent to measure tissue-specific mechanisms should be balanced by consideration of whether the clock is intended to be used as a clinical risk predictor.

11.1.3 Choice of Age Range

The factors influencing the rate of aging change across the life course. For this reason, epigenetic clocks trained among individuals of a relatively narrow age range may be stronger predictors of health for target populations of a similar age range. In contrast, clocks trained across a broad range of ages, like the Hannum (ages 19 to 101 years) and PanTissue (ages 0 to 101 years) clocks, are more likely to be generalizable across studies. Epigenetic clocks created for specific critical periods of development include predictors of gestational age. Gestational age is frequently used as a proxy for developmental maturity and is strongly related to both short-term and long-term morbidity and mortality [30]. Epigenetic clocks for gestational age

have been developed for both neonatal cord blood [31–33] and placenta [34, 35]. Comparison of these clocks across chorionic villus, placenta, and cord blood tissues found epigenetic age acceleration in the cord and placental samples was not correlated, suggesting these clocks capture distinct variation in gestational age-related DNAm [36]. The training sets for these predictors should be carefully considered when evaluating associations of these gestational age clocks with pregnancy complications, as well as maternal characteristics and exposures. Early delivery is indicated in the case of certain maternal and fetal conditions, including pre-eclampsia or eclampsia, diabetes, and intrauterine growth restriction [37]. Some of the risk factors for spontaneous preterm birth include inflammation, stress, smoking, low pre-pregnancy body mass index, and periodontal disease [37]. Training sets that include an overrepresentation of preterm births are partially built to capture these characteristics. Potentially due to the influence of multiple factors on gestational age, the predictive accuracy of cord blood gestational age clocks has been shown to be particularly sensitive to the size of training set [38]. For the cord blood clock trained on a larger sample size [31], epigenetic age acceleration has been associated with higher birthweight and greater birth length [39], but lower weight in childhood [40]. Childhood and adolescence also represent critical windows of development with implications for adult health. The timing and rate of pubertal development have been associated with both early life adversity and future cardiometabolic health and cancer risk [41]. Pediatric epigenetic clocks trained on chronologic age have been developed to more accurately characterize aging rates during this period [5, 8]. The first of these predictors was developed using peripheral blood from individuals ages 9 months to 18 years [8], while the Pediatric-Buccal-Epigenetic (PedBE) clock was derived using DNAm assayed in buccal swabs collected from individuals ages 0 to 20 years [5]. Comparison of epigenetic clocks across pediatric tissues highlights the added value of clocks specific to pediatric samples [42]. Although the correlation between chronologic age and the PanTissue, Hannum, PhenoAge, skin and blood, and pediatric clocks was similar in pediatric blood samples, the correlation was much stronger for the PedBE and skin and blood clocks in pediatric buccal and saliva samples [42]. Across tissues, the deviation between chronologic age and age estimated by the PedBE and blood pediatric clocks was lower among young individuals [42].

11.1.4 Validation

To appraise the potential value of a predictor, it must first be compared to the gold standard upon which it was developed. For epigenetic clocks trained on chronologic age, we expect the DNAm-based estimate to be highly correlated with chronologic age. Epigenetic clocks derived using penalized regression models tend to use cross-validation to reduce the likelihood of overfitting the model to the training dataset. Overfitting creates a predictor that is highly correlated with the gold standard among the training data, but tends to perform poorly among external samples. While the cross-validation procedure can inform selection of a tuning parameter that minimizes the estimated prediction error, the characteristics of the training dataset can influence

the external validity of the predictor. For example, an epigenetic clock developed among older women may not perform well among a cohort of young men. If assuming broad external validity, it is useful to assess the concordance between the predictor and the gold standard in an external dataset. As an example of the ideal approach, 13 data sets were used to construct the skin and blood clock age predictor, and 13 independent data sets were used to validate it [3]. When this ideal approach is not feasible, the prediction error can be appraised by separating a cohort of samples into distinct training and testing subgroups. Validation of the DNAm-based estimators of plasma proteins integrated into the GrimAge clock applied this alternative approach. A random sample of the Framingham Heart Study Offspring Cohort was used to develop the prediction models for each plasma protein, which were then tested in a distinct subgroup of cohort participants [22].

The characteristics of the training dataset require careful consideration when the surrogate measure of aging used to train the epigenetic clock must also be developed. To train the DunedinPoAm clock, the Pace of Aging estimate was first derived among Dunedin Study birth cohort participants, who were then used to create a DNAm-based estimator of this measure [21]. Estimation of this Pace of Aging measure was only feasible due to the rich, longitudinal biomarker data available for this birth cohort. However, the uniqueness of this data impeded validation of this composite rate of aging estimator in an external dataset. Given participants were primarily white, and all were young adults, the transportability of the DunedinPoAm clock to ethnic minorities and older individuals requires additional appraisal. Although the phenotypic age measure used to derive the PhenoAge clock was not built to capture longitudinal trends in clinical markers like the DunedinPoAm clock, it was trained to be broadly generalizable to US demographics. One cross-sectional wave of the nationally representative NHANES participants was used to create the phenotypic age measure, which was validated among a subsequent, independent wave of NHANES participants [19]. A separate cohort was then used to train a DNAm-based estimator of phenotypic age [19].

As indicators of biologic aging, validation of epigenetic clocks is ultimately dependent on well-replicated relations with aging-related conditions and mortality rate. To support validity claims, publications introducing a new epigenetic clock tend to include evaluation of associations with morbidity and/or mortality in cohorts that were not used to train the clock [4, 18, 19, 21, 22]. While the original publications for the PanTissue and Hannum clocks did not assess relations with aging outcomes, subsequent meta-analyses demonstrate that these clocks are consistent predictors of morbidity and mortality across cohorts [27, 43, 44]. With the expansion of new epigenetic clocks, investigators must not only validate associations with aging outcomes, but also appraise the strength of a new predictor relative to established epigenetic clocks. If a new clock does not improve the predictive accuracy relative to other clocks in some target population, other considerations could be used to argue utility, e.g., sparsity (few CpGs) or biological interpretability (e.g., clocks that are based on CpGs in specific genomic locations).

11.2 Integration into Epidemiologic Studies

Epigenetic clocks have been integrated into a wide array of epidemiologic investigations [45, 46]. The intent of these studies can be broadly categorized into two primary goals: to understand disease etiology and improve risk prediction. In practice, it is common to combine these interests by minimizing all forms of bias in the analysis and making assumptions about the underlying mechanism in the interpretation. The characteristics how an epigenetic clock was trained may inform its inclusion into a specific research question. However, the ease of estimating multiple clocks from a single DNAm assay enables a more hypothesis-generating approach. Many studies evaluate multiple epigenetic clocks to inform discussion of the aging mechanisms driving observed associations. In this section, we review the ways epidemiology studies of epigenetic clocks can inform our understanding of aging and risk of aging-related conditions.

11.2.1 To Understand Disease Etiology

When the interest is causality, a fundamental question is whether the shifts in DNAm underlying estimated epigenetic age have a direct effect on disease risk, or whether these changes in DNAm are surrogates for a disease process. The distinction has important implications for creating targeted therapies intended to reverse epigenetic age. As described in the first section, several epigenetic clocks have been developed to serve as surrogates for specific facets of aging (e.g., the mitotic clocks) or composite measures of aging (e.g., the PhenoAge, GrimAge, and DunedinPoAm clocks). In these cases, it is possible to more confidently infer the mechanisms underlying the associations after adjusting for putative confounders. If the goal is to provide insight into disease etiology, a logical question is whether it would be more sensible to directly assay the facets of aging used to train these clocks. This viewpoint assumes that any discrepancies between the characteristic the clock was trained on and its DNAm-based estimate is uninformative error. However, there is compelling evidence that this error is partially what characterizes the risk of accelerated aging trajectories. This is well illustrated in case of epigenetic clocks trained on chronologic age, where the residual variation in the clock, independent of age, is associated with mortality and morbidity [27, 43, 44]. The stronger association between DNAm-estimated leukocyte telomere length and mortality rate, relative to Southern blot measured leukocyte telomere length, also supports this premise [17]. Similarly, the DunedinPoAm clock was more strongly associated with several physical and cognitive function characteristics than the original Pace of Aging measure on which it was trained [21]. The assumption that shifts in DNAm used to estimate epigenetic clocks are directly associated with outcomes is further complicated by the relatively high degree of shared variation across the methylome. Given prediction models that feature variable selection are commonly used to train epigenetic clocks, the specific CpG loci included in the epigenetic clock model may not be causal, but highly correlated with causal sites. The specific sites in the clock

may therefore be considered representative examples of a specific pattern of DNAm variation. To estimate a causal effect of this variation on conditions of aging, we must assume the capacity to control for all confounders of the relation. This presumes we know all possible confounders, which is a strong supposition given evidence for a quickly expanding list of epigenetic age predictors. An alternative approach that is more robust to residual unmeasured confounding is Mendelian Randomization (MR) [47, 48]. This approach uses a genetic proxy for epigenetic age to estimate the causal relation between epigenetic age and aging-related outcomes. MR also requires some strong assumptions, notably that the polymorphisms associated with epigenetic age do not have a direct effect on the outcome of interest (pleiotropy) [48]. However, an adaption of Egger regression can be used to detect and correct for this potential bias [49]. Most critically, this approach requires a genetic instrument that describes sufficient variation in epigenetic age. To date, there have been a few genome-wide association studies of different epigenetic clocks [50, 51]. Estimated SNP-based heritability has been relatively low, ranging from 0.15 to 0.19, suggesting an MR analysis would require a very large sample size to be adequately powered [50, 51]. Beyond causal modeling techniques, the biologic mechanism of epigenetic clocks can be informed by multi-omic analyses, as well as in vivo and in vitro studies [52]. These complementary investigations will be important validation of findings in human studies and will inform the interpretation of subsequent analyses.

11.2.2 Epigenetic Clocks as Risk Predictors

Our analytic considerations are distinct when appraising the potential value of epigenetic clocks as risk predictors. When the objective is to use epigenetic clocks as a general screening tool to inform risk management, causality is irrelevant if epigenetic clocks provide an easy, cost-effective means to measure facets of aging that strongly predict risk of morbidity and mortality. For example, a single blood-based estimate of the DunedinPoAm clock may be more feasible to integrate into a clinical setting than measuring the longitudinal change in the 18 blood-chemistry and organ-system-function biomarkers used train this clock. The relative ease of measurement is highly dependent on the tissue proposed for assay, incentivizing risk prediction studies to focus on easily accessible samples like peripheral blood or buccal swabs. For more targeted screening proposals, the use of less accessible tissues may be justifiable if the predictive accuracy can be improved. For example, the estimation of mitotic age in tumor tissue to predict prognosis among cancer patients [15].

Another important factor informing the use of epigenetic clocks in a clinical setting is the external validity of these algorithms across populations, which is characterized by both their generalizability and transportability. Generalizability refers to the consistency of study findings from a potentially biased sampling of the target population to the full target population. For example, if sex is an effect modifier of the relation between chronologic age and DNAm, and the study

population included an overrepresentation of females, the model may not be generalizable to a sampling of the source population with a more equal distribution of males and females. Distinctly, transportability refers to the consistency of study findings to the target population when the source population and target population only partially overlap or do not overlap. Modifying our example, a model trained only among females may not be transportable to a population of males. Ideally, the external validity of the epigenetic clock algorithms should be appraised by evaluating the consistency between the trait used to train the epigenetic clock and its corresponding DNAm-based estimate across groups of interest within our target population. As discussed in Sect. 11.1.4, new clocks frequently use a test set that is distinct from the training set to validate the epigenetic age algorithm. Alternatively, we can evaluate whether specific characteristics modify the relation between epigenetic age and conditions of aging. This second approach captures both the external validity of the epigenetic clock algorithms and potential effect measure modification of associations with epigenetic clocks, both of which are informative in the context of prediction modeling. Notably, studies evaluating epigenetic clocks as predictors of morbidity and mortality should consider heterogeneity in associations across sex and race/ethnicity. An extreme test of transportability arises when generalizing findings from one species (e.g., mice) to another (e.g., primates). This challenge is met by third-generation epigenetic clocks that use a single model to estimate age in all mammalian species (e.g., universal mammalian clocks [53]).

To track longitudinal variation in risk profile, epigenetic clocks must have a high reliability and be responsive to interventions. High reliability is indicated by consistent epigenetic age estimates across technical replicates and over short timeframes. Extraneous variation that contributes to low reliability reduces our capacity to track changes in epigenetic age over time. To date, specific epigenetic age estimates have been shown to have relatively high reliability across DNAm preprocessing pipelines and DNAm arrays [54]. However, additional characterization of ways to maximize the reliability of these estimates is needed to optimize the calculation of epigenetic age in a clinical setting. There is growing evidence from longitudinal studies that the rate of epigenetic aging is responsive to specific interventions [55–58]. In a phase I clinical trial of 9 healthy men, GrimAge was reversed by 2.5 years after one year of treatment with a protocol intended to regenerate the thymus and revert immunosenescent trends [56]. Recently, PanTissue age was shown to be significantly reversed after 16 weeks of vitamin D_3 supplementation in a clinical trial of overweight/obese African Americans [55]. In spite of their relatively small sample sizes, these studies provide compelling initial evidence that specific interventions can rejuvenate epigenetic age, even within a relatively short timeframe. To serve as an indicator of shifts in in risk profile with time, intervention-induced epigenetic age rejuvenation needs to correspond to a reduced risk of conditions of aging. With the expansion of longitudinal studies of epigenetic age, we will be able to better appraise whether the rate of epigenetic aging improves risk prediction above a point in time estimate of epigenetic age.

11.3 Conclusions

Epigenetic clocks have been and will be widely used in epidemiological studies. Investigators need to carefully think about how the clock was trained (e.g., as a mortality predictor) and the source of DNA (blood, buccal swabs, saliva, urine, adipose tissue), since the results often depend on it. The new frontier is to adapt epigenetic clocks for specific applications and to qualify these molecular biomarkers of aging as surrogate endpoints for human clinical trials.

References

1. Bocklandt S, Lin W, Sehl ME et al (2011) Epigenetic predictor of age. PLoS One 6:e14821. https://doi.org/10.1371/journal.pone.0014821
2. Hannum G, Guinney J, Zhao L et al (2013) Genome-wide methylation profiles reveal quantitative views of human aging rates. Mol Cell 49:359–367. https://doi.org/10.1016/j.molcel.2012.10.016
3. Horvath S (2013) DNA methylation age of human tissues and cell types. Genome Biol 14:3156. https://doi.org/10.1186/gb-2013-14-10-r115
4. Horvath S, Oshima J, Martin GM et al (2018) Epigenetic clock for skin and blood cells applied to Hutchinson Gilford progeria syndrome and ex vivo studies. Aging (Albany NY) 10:1758–1775. https://doi.org/10.18632/aging.101508
5. McEwen LM, O'Donnell KJ, McGill MG et al (2020) The PedBE clock accurately estimates DNA methylation age in pediatric buccal cells. Proc Natl Acad Sci U S A 117:23329–23335. https://doi.org/10.1073/pnas.1820843116
6. Voisin S, Harvey NR, Haupt LM et al (2020) An epigenetic clock for human skeletal muscle. J Cachexia Sarcopenia Muscle 11:887–898. https://doi.org/10.1002/jcsm.12556
7. Weidner CI, Lin Q, Koch CM et al (2014) Aging of blood can be tracked by DNA methylation changes at just three CpG sites. Genome Biol 15:R24. https://doi.org/10.1186/gb-2014-15-2-r24
8. Wu X, Chen W, Lin F et al (2019) DNA methylation profile is a quantitative measure of biological aging in children. Aging (Albany NY) 11:10031–10051. https://doi.org/10.18632/aging.102399
9. Zhang Q, Vallerga CL, Walker RM et al (2019) Improved precision of epigenetic clock estimates across tissues and its implication for biological ageing. Genome Med 11:54. https://doi.org/10.1186/s13073-019-0667-1
10. Zhu T, Gao Y, Wang J et al (2019) CancerClock: a DNA methylation age predictor to identify and characterize aging clock in pan-cancer. Front Bioeng Biotechnol 7. https://doi.org/10.3389/fbioe.2019.00388
11. Hastie T, Tibshirani R, Friedman J (2009) The elements of statistical learning: data mining, inference, and prediction, 2nd edn. Springer, New York
12. Garagnani P, Bacalini MG, Pirazzini C et al (2012) Methylation of ELOVL2 gene as a new epigenetic marker of age. Aging Cell 11:1132–1134. https://doi.org/10.1111/acel.12005
13. Teschendorff AE (2020) A comparison of epigenetic mitotic-like clocks for cancer risk prediction. Genome Med 12:56. https://doi.org/10.1186/s13073-020-00752-3
14. Yang Z, Wong A, Kuh D et al (2016) Correlation of an epigenetic mitotic clock with cancer risk. Genome Biol 17:205. https://doi.org/10.1186/s13059-016-1064-3
15. Youn A, Wang S (2018) The MiAge calculator: a DNA methylation-based mitotic age calculator of human tissue types. Epigenetics 13:192–206. https://doi.org/10.1080/15592294.2017.1389361

16. Zhou W, Dinh HQ, Ramjan Z et al (2018) DNA methylation loss in late-replicating domains is linked to mitotic cell division. Nat Genet 50:591–602. https://doi.org/10.1038/s41588-018-0073-4
17. Lu AT, Seeboth A, Tsai P-C et al (2019b) DNA methylation-based estimator of telomere length. Aging (Albany NY) 11:5895–5923. https://doi.org/10.18632/aging.102173
18. Zhang Y, Wilson R, Heiss J et al (2017) DNA methylation signatures in peripheral blood strongly predict all-cause mortality. Nat Commun 8:14617. https://doi.org/10.1038/ncomms14617
19. Levine ME, Lu AT, Quach A et al (2018) An epigenetic biomarker of aging for lifespan and healthspan. Aging (Albany NY) 10:573–591. https://doi.org/10.18632/aging.101414
20. Belsky DW, Caspi A, Houts R et al (2015) Quantification of biological aging in young adults. Proc Natl Acad Sci U S A 112:E4104–E4110. https://doi.org/10.1073/pnas.1506264112
21. Belsky DW, Caspi A, Arseneault L et al (2020) Quantification of the pace of biological aging in humans through a blood test, the DunedinPoAm DNA methylation algorithm. elife 9. https://doi.org/10.7554/eLife.54870
22. Lu AT, Quach A, Wilson JG et al (2019a) DNA methylation GrimAge strongly predicts lifespan and healthspan. Aging (Albany NY) 11:303–327. https://doi.org/10.18632/aging.101684
23. Horvath S, Gurven M, Levine ME et al (2016) An epigenetic clock analysis of race/ethnicity, sex, and coronary heart disease. Genome Biol 17:171. https://doi.org/10.1186/s13059-016-1030-0
24. Houseman EA, Accomando WP, Koestler DC et al (2012) DNA methylation arrays as surrogate measures of cell mixture distribution. BMC Bioinformatics 13:86. https://doi.org/10.1186/1471-2105-13-86
25. Fagnoni FF, Vescovini R, Passeri G et al (2000) Shortage of circulating naive CD8(+) T cells provides new insights on immunodeficiency in aging. Blood 95:2860–2868
26. Gruver AL, Hudson LL, Sempowski GD (2007) Immunosenescence of ageing. J Pathol 211:144–156. https://doi.org/10.1002/path.2104
27. Chen BH, Marioni RE, Colicino E et al (2016) DNA methylation-based measures of biological age: meta-analysis predicting time to death. Aging (Albany NY) 8:1844–1865. https://doi.org/10.18632/aging.101020
28. Horvath S, Mah V, Lu AT et al (2015) The cerebellum ages slowly according to the epigenetic clock. Aging (Albany NY) 7:294–306. https://doi.org/10.18632/aging.100742
29. Sehl ME, Henry JE, Storniolo AM et al (2017) DNA methylation age is elevated in breast tissue of healthy women. Breast Cancer Res Treat. https://doi.org/10.1007/s10549-017-4218-4
30. Engle WA (2011) Morbidity and mortality in late preterm and early term newborns: a continuum. Clin Perinatol 38:493–516. https://doi.org/10.1016/j.clp.2011.06.009
31. Bohlin J, Håberg SE, Magnus P et al (2016) Prediction of gestational age based on genome-wide differentially methylated regions. Genome Biol 17:207. https://doi.org/10.1186/s13059-016-1063-4
32. Haftorn KL, Lee Y, Denault WRP et al (2021) An EPIC predictor of gestational age and its application to newborns conceived by assisted reproductive technologies. Clin Epigenetics 13:82. https://doi.org/10.1186/s13148-021-01055-z
33. Knight AK, Craig JM, Theda C et al (2016) An epigenetic clock for gestational age at birth based on blood methylation data. Genome Biol 17:206. https://doi.org/10.1186/s13059-016-1068-z
34. Lee Y, Choufani S, Weksberg R et al (2019) Placental epigenetic clocks: estimating gestational age using placental DNA methylation levels. Aging (Albany NY) 11:4238–4253. https://doi.org/10.18632/aging.102049
35. Mayne BT, Leemaqz SY, Smith AK et al (2017) Accelerated placental aging in early onset preeclampsia pregnancies identified by DNA methylation. Epigenomics 9:279–289. https://doi.org/10.2217/epi-2016-0103
36. Dieckmann L, Lahti-Pulkkinen M, Kvist T et al (2021) Characteristics of epigenetic aging across gestational and perinatal tissues. Clin Epigenetics 13:97. https://doi.org/10.1186/s13148-021-01080-y

37. Goldenberg RL, Culhane JF, Iams JD, Romero R (2008) Epidemiology and causes of preterm birth. Lancet 371:75–84. https://doi.org/10.1016/S0140-6736(08)60074-4
38. Simpkin AJ, Suderman M, Howe LD (2017) Epigenetic clocks for gestational age: statistical and study design considerations. Clin Epigenetics 9:100. https://doi.org/10.1186/s13148-017-0402-y
39. Khouja JN, Simpkin AJ, O'Keeffe LM et al (2018) Epigenetic gestational age acceleration: a prospective cohort study investigating associations with familial, sociodemographic and birth characteristics. Clin Epigenetics 10:86. https://doi.org/10.1186/s13148-018-0520-1
40. Bright HD, Howe LD, Khouja JN et al (2019) Epigenetic gestational age and trajectories of weight and height during childhood: a prospective cohort study. Clin Epigenetics 11:194. https://doi.org/10.1186/s13148-019-0761-7
41. Golub MS, Collman GW, Foster PMD et al (2008) Public health implications of altered puberty timing. Pediatrics 121(Suppl 3):S218–S230. https://doi.org/10.1542/peds.2007-1813G
42. Kling T, Wenger A, Carén H (2020) DNA methylation-based age estimation in pediatric healthy tissues and brain tumors. Aging (Albany NY) 12:21037–21056. https://doi.org/10.18632/aging.202145
43. Fransquet PD, Wrigglesworth J, Woods RL et al (2019) The epigenetic clock as a predictor of disease and mortality risk: a systematic review and meta-analysis. Clin Epigenetics 11:62. https://doi.org/10.1186/s13148-019-0656-7
44. Marioni RE, Shah S, McRae AF et al (2015) DNA methylation age of blood predicts all-cause mortality in later life. Genome Biol 16. https://doi.org/10.1186/s13059-015-0584-6
45. Oblak L, van der Zaag J, Higgins-Chen AT et al (2021) A systematic review of biological, social and environmental factors associated with epigenetic clock acceleration. Ageing Res Rev 69:101348. https://doi.org/10.1016/j.arr.2021.101348
46. Ryan J, Wrigglesworth J, Loong J et al (2020) A systematic review and meta-analysis of environmental, lifestyle, and health factors associated with DNA methylation age. J Gerontol A Biol Sci Med Sci 75:481–494. https://doi.org/10.1093/gerona/glz099
47. Relton CL, Davey Smith G (2012) Two-step epigenetic Mendelian randomization: a strategy for establishing the causal role of epigenetic processes in pathways to disease. Int J Epidemiol 41:161–176. https://doi.org/10.1093/ije/dyr233
48. Relton CL, Davey Smith G (2015) Mendelian randomization: applications and limitations in epigenetic studies. Epigenomics 7:1239–1243. https://doi.org/10.2217/epi.15.88
49. Bowden J, Davey Smith G, Burgess S (2015) Mendelian randomization with invalid instruments: effect estimation and bias detection through egger regression. Int J Epidemiol 44:512–525. https://doi.org/10.1093/ije/dyv080
50. Gibson J, Russ TC, Clarke T-K et al (2019) A meta-analysis of genome-wide association studies of epigenetic age acceleration. PLoS Genet 15:e1008104. https://doi.org/10.1371/journal.pgen.1008104
51. Lu AT, Xue L, Salfati EL et al (2018) GWAS of epigenetic aging rates in blood reveals a critical role for TERT. Nat Commun 9:387. https://doi.org/10.1038/s41467-017-02697-5
52. Liu Z, Leung D, Thrush K et al (2020) Underlying features of epigenetic aging clocks in vivo and in vitro. Aging Cell 19:e13229. https://doi.org/10.1111/acel.13229
53. Consortium MM, Lu AT, Fei Z et al (2021) Universal DNA methylation age across mammalian tissues. bioRxiv. https://doi.org/10.1101/2021.01.18.426733
54. McEwen LM, Jones MJ, Lin DTS et al (2018) Systematic evaluation of DNA methylation age estimation with common preprocessing methods and the Infinium MethylationEPIC BeadChip array. Clin Epigenetics 10:123. https://doi.org/10.1186/s13148-018-0556-2

55. Chen L, Dong Y, Bhagatwala J et al (2019) Effects of vitamin D3 supplementation on epigenetic aging in overweight and obese African Americans with suboptimal vitamin D status: a randomized clinical trial. J Gerontol A Biol Sci Med Sci 74:91–98. https://doi.org/10.1093/gerona/gly223
56. Fahy GM, Brooke RT, Watson JP et al (2019) Reversal of epigenetic aging and immunosenescent trends in humans. Aging Cell 18:e13028. https://doi.org/10.1111/acel.13028
57. Fitzgerald KN, Hodges R, Hanes D et al (2021) Potential reversal of epigenetic age using a diet and lifestyle intervention: a pilot randomized clinical trial. Aging (Albany NY) 13:9419–9432. https://doi.org/10.18632/aging.202913
58. Gensous N, Garagnani P, Santoro A et al (2020) One-year Mediterranean diet promotes epigenetic rejuvenation with country- and sex-specific effects: a pilot study from the NU-AGE project. Geroscience 42:687–701. https://doi.org/10.1007/s11357-019-00149-0

Influence of Environmental Factors on the Epigenome

12

Brock C. Christensen, Todd M. Everson, Carmen J. Marsit, and Karl T. Kelsey

Contents

B. C. Christensen (✉)
Department of Epidemiology, Department of Molecular and Systems Biology, Department of Community and Family Medicine, Geisel School of Medicine at Dartmouth, Hanover, NH, USA
e-mail: Brock.Christensen@Dartmouth.edu

T. M. Everson · C. J. Marsit
Department of Environmental Health, and Department of Epidemiology, Emory School of Public Health, Atlanta, GA, USA
e-mail: Todd.M.Everson@emory.edu; carmen.j.marsit@emory.edu

K. T. Kelsey
Department of Epidemiology, and Department of Pathology and Laboratory Medicine, The Warren Alpert Medical School of Brown University, Providence, RI, USA
e-mail: Karl_Kelsey@brown.edu

K. B. Michels (ed.), *Epigenetic Epidemiology*,
https://doi.org/10.1007/978-3-030-94475-9_12

Abstract

In this chapter we consider the role of environmental factors on the epigenome. The importance of research into exposures that may alter epigenetic marks is now well recognized. Relations of exposures such as tobacco, alcohol, diet, endocrine disruptors, metals, and environmental contaminants with epigenetic states have been investigated and are reviewed here. We will briefly cover environmental exposures and imprinting and development, as well as discuss potential mechanisms for exposures to modify epigenetic states. Appropriate epidemiologic studies are crucial to understanding the true effect of environmental exposures on the human epigenome and this work is urgently needed to better understand the biology of epigenetic alterations which may constitute biology underlying risk for pathogenesis of disease. With a more comprehensive understanding of the effects of exposures on the epigenome (including consideration of genetic background), not only will the prediction of the toxic potential of new compounds be more readily achieved, but precision prevention and intervention strategies also may be developed.

Abbreviations

ADHD	attention deficit hyperactivity disorder
ART	assisted reproductive technology
AUC	area under the curve
BMI	body mass index
BPA	bisphenol A
BPF	bisphenol F
BPS	bisphenol S
DES	diethylstilbestrol
DMRs	differentially methylated regions
EWAS	Epigenome wide association study
GST	glutathione transferase
H3K4	Histone 3 lysine 4
HNSCC	head and neck squamous cell carcinoma
ICR	imprinting control regions
IQR	inter-quartile range
LINE	long interspersed nucleotide elements
mQTL	methylation quantitative trait loci
NTD	neural tube defects
PBMC	peripheral blood mononuclear cells
PCBs	polychlorinated biphenyls
PFASs	per-and polyfluoroalkyl substances
PM_{10}	inhalable particulate matter
$PM_{2.5}$	fine particulate matter
RRBS	reduced-representation bisulfite sequencing
SAM	S-adenosyl methionine
UV	ultraviolet

12.1 Introduction

The epigenome is the landscape of mitotically heritable changes in gene expression and gene expression potential that are mediated without altering genetic sequence. A role for environmental exposures in epigenetic alteration is increasingly recognized and investigated, and much of the initial evidence implicating a role of environmental factors on the epigenome came from studies of disease outcomes such as cancer and adverse reproductive/developmental events. Initially, epigenetic alterations were identified in various human tumors and consequently, environmental exposures known to have an etiologic role in cancer, such as tobacco, air pollutants, and metals have been studied and implicated in the modification of epigenetic marks. Of course, this has highlighted the need to consider how environmental factors may influence the epigenome in pathologically normal tissues, perhaps representing alterations necessary to initiate or predispose to disease phenotypes. Since the first edition of this work, evidence of relations between environmental factors and the epigenome also has emerged from cohort studies. The majority of epigenetic alteration work in human biospecimens focused on DNA methylation, though emerging work is beginning to test potential associations of environmental exposures on chromatin conformation. This chapter highlights evidence that epigenetic alterations are associated with environmental exposures, focusing on human studies, including epigenome-wide association studies (EWASs). The potential mechanisms behind environmentally related epigenetic alterations and methods for studying the relation between the epigenome and the environment that are amenable to epidemiologic research are also presented. As the data from human biospecimens has, for the most part, centered on alterations to DNA methylation, this chapter will focus on the examination of DNA methylation and environment.

12.2 Environmental Exposures of Diet and Lifestyle and Epigenetics

12.2.1 Folate, One-Carbon Metabolism, Diet, and Body Mass Index

One-carbon metabolism is the network of biochemical reactions essential to both DNA synthesis and all cellular biomolecule (nucleic acids, proteins, lipids) methylation reactions that involve the transfer of one-carbon groups. A critical nutrient central to one-carbon metabolism is folate (vitamin B_9), which donates its methyl group for homocysteine remethylation to methionine. Subsequently, methionine is the methyl donor for all cellular methylation reactions, most notable for this discussion being DNA and histone methylation via S-adenosyl methionine (SAM), and other B vitamins (B_2, B_6, B_{12}) that act as enzymatic co-factors in the network. Collectively, B vitamins, homocysteine, and methionine are important contributors to the maintenance of DNA integrity and DNA methylation. ⊙ Figure 12.1 represents the links between diet, one-carbon metabolism, DNA synthesis, DNA methylation, and toxicant metabolism.

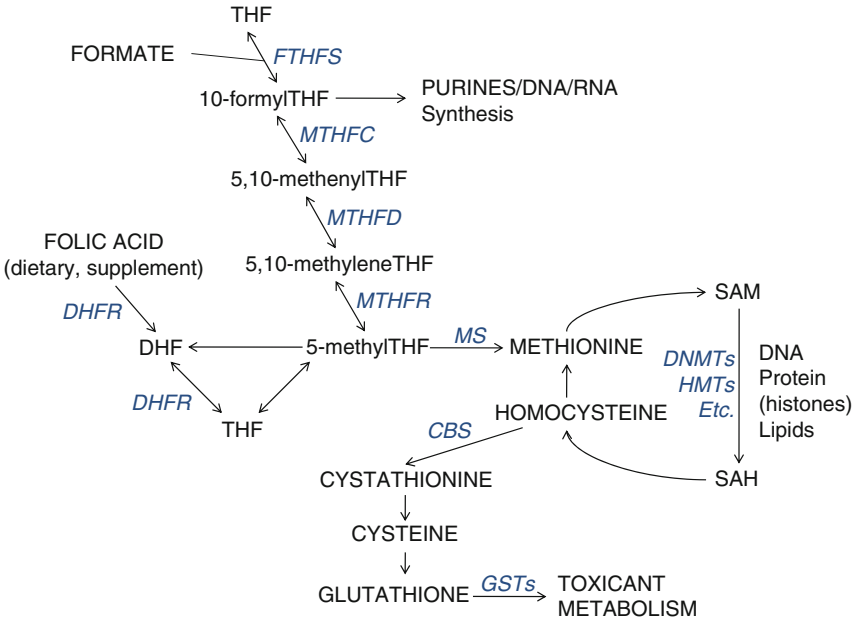

Fig. 12.1 Diet, one-carbon metabolism, DNA synthesis, DNA methylation, and toxicant metabolism

Datei: 978–94–007-2495-2_12_Fig 1_SB7.gif
Abb.-Typ: Strich-Abb.
Farbigkeit (IST): 1c
Farbigkeit (SOLL): 1c
Bildrechte: [Urheberrecht beim Autor]
Abdruckrechte: Nicht notwendig
Hinweise Verlag/Setzerei:

In mammals, establishment of somatic cell epigenetic patterns occurs early in fetal development subsequent to genome-wide reprogramming of epigenetic patterns at the multiple developmental stages (fertilization, implantation, blastocyst) when totipotency and pluripotency requirements are dynamic [1]. In addition, DNA methylation is necessary for the mitotic inheritance process of genomic imprinting (mono-allelic gene expression which can also be tissue-specific). Embryonic and fetal nutrient availability have great potential to affect the epigenetic reprogramming and patterning phenomena, which has implications for proper development and perhaps even life-long conditioning and health [2, 3]. Transmission of both nutrients and environmental exposures to developing mammals in utero proceeds through the placenta, and maternal folate status is recognizably associated with fetal development. For instance, it is well-known that folate deficiency is associated with neural

tube defects (NTD) [4] and animal studies have indicated the necessity of sufficient methyl group availability for proper neural tube development [5]. Further, hypomethylation of long interspersed nucleotide elements (LINE) and genomic DNA has been associated with increased risk of NTDs in humans [6]. At the same time, genetics is at the intersection of methyl group availability and NTD risk. In particular, perhaps not surprisingly, there has been a focus on genetic variation in genes that code for enzymes in the one-carbon metabolism network. Yan et al. led a meta-analysis of 25 studies to better understand the association of *MTHFR* genotype, one-carbon metabolites, and NTD risk which identified TT genotype to be associated with lower folate, lower enzymatic activity, and increased NTD risk [7]. Heterozygous genotype also was associated with lower folate and lower enzymatic activity compared with wild type.

Of course, Waterland and Michels [8] proposed that epigenetic mechanisms may contribute to the developmental origins of health and disease hypothesis, which states that nutrition and environmental stimuli at critical developmental periods (pre- and postnatal) can induce permanent alterations in susceptibility to metabolic and chronic diseases [9]. Additional details on epigenetic changes during development are available in Chap. 5, and the developmental origins of health and disease hypothesis is covered in Chap. 6.

There have been several animal studies that directly link folate intake, diet, and epigenetic alterations and evidence from human studies has been slower to emerge. Initially, human studies described associations between folate, B vitamins, or diet and disease risk and suggested the potential for epigenetic mechanisms to account for these observations. For instance, elevated serum vitamin B_6 and methionine have been associated with reduced risk of lung cancer [10], folate intake has been associated with reduced risk of breast cancer (especially among women with high alcohol consumption) [11, 12] and a protective role for folate in colon cancer [13] has been consistently reported. Investigating some of the mechanisms of protective effects, breast and lung cancer cell lines treated with methyl donors observed reduced cell proliferation (through MAPK/ERK and AKT pathways) as well as increased apoptosis [14]. In a pooled analysis of 13 prospective cohort studies on folate intake and colon cancer, Kim and colleagues reinforced the significant, though modest protective role of folate for reducing risk of colon cancer by approximately 2% per 100 μg/day of folate [15]. Similarly, a meta-analysis of 19 studies identified that 100 μg/day of folate supplementation resulted in a 12% reduced risk of esophageal cancer [16]. Extending that work, vitamin B_6 intake of 1 mg/day decreased esophageal cancer risk, though results for B_{12} intake indicated a slight increase in cancer risk [17]. Work from the Netherlands Cohort Study on diet and cancer investigated risk of colorectal cancer according to methyl donor (folate, methionine, and vitamins B_2 and B_6) intake and considering genotypes of enzymes in the one-carbon metabolism pathway [18]. These authors did not find significant diet–gene interactions, but did observe a significant inverse association between methionine intake and risk of colon cancer among individuals with common *DNMT3B* (de novo DNA methyltransferase) genotypes as well between B_2 intake and risk of colon cancer in individuals with one or less rare allele among the folate enzyme genes *MTHFR, MTRR*, and *MTR* [18]. More recently, in a case–control study with over 5000 subjects in a Chinese

population, increased intake of each of folate, vitamin B2, B6, and B12 was associated with reduced colorectal cancer risk [19].

In addition, a study of serum concentrations of folate and plasma SAM levels in over 330 women of reproductive age showed that body mass index (BMI) was the strongest determinant of SAM concentration, suggesting that nutrient intake and nutrient availability are distinguishable, and that BMI is a potential confounder of folate–methylation associations [20]. However, in work measuring folate and vitamin B12 in early pregnancy, among nearly 500 women, high BMI was associated with decreased serum and plasma folate levels for women with equivalent intake of folate (O'Malley 2018). Thus, although BMI may be positively associated with one-carbon factors and co-factors, current guidelines recommend that women with high BMI ($> = 30$) supplement with high-dose folic acid in the periconception period. Although small, one study that also investigated high BMI and compared with normal BMI on folic acid supplementation included measures of neutrophil DNA methylation at genome scale. Obese women had lower blood folate concentrations and changes in DNA methylation among genes related with NTD and folate were different between normal weight and obese women [21].

Associations of folic acid supplementation with epigenetic states also have been investigated in early life and mother–infant pairs. Candidate gene methylation measures for growth and metabolism genes in a study of mother–infant pairs identified lower *LEP* and *IGF2* methylation in infant buccal cell DNA at six months associated with preconception folate and folic acid intake [22]. Measures of cord blood DNA methylation at candidate genes and *LINE-1* from infants whose mothers had folic acid supplementation in second and third trimesters compared with placebo showed moderately reduced *LINE-1, IFG2,* and *BDNF* methylation though results did not adjust for potential variation in cell type proportions [23]. Another study of cord blood that stratified analysis of genome-scale methylation on *MTHFR* genotype identified global methylation dependency on maternal genotype with higher methylation in wild type mothers associated with multivitamin supplementation, whereas variant allele presence indicated no association of methylation with supplementation [24]. Additional work in pregnant women and their offspring that incorporates one-carbon metabolism enzyme genotypes, genome-scale methylation measures, and adjustment for cell type proportions in mixture biospecimens is warranted to understand the relation of one-carbon donors with DNA methylation and infant growth and early life health.

Although relatively abundant work exists examining one-carbon pathway factors and co-factors in association with demographic factors, inclusion of epigenetic measures has emerged more recently, and most studies have been in the context of cancer risk. Early, innovative work in uremia patients with hyperhomocyteinemia used cytosine extension and Southern blot identified global DNA hypomethylation and successfully intervened using folate treatment, restoring DNA methylation levels [25]. Among disease-free controls, colorectal tissue global methylation was higher compared with nontumor adjacent tissue from patients with adenomas. In addition, folate status also was lower in cases, and DNA hypomethylation was associated with increased risk of adenoma and weakly associated with colorectal cancer risk [26]. Folate status has been inversely correlated with colon tissue DNA

methylation extent [27]. In head and neck squamous cell carcinoma (HNSCC) repetitive element methylation extent was associated with folate intake, was reduced among individuals with a variant *MTHFR* genotype, and was a significant risk factor for disease [28]. Using breast tissue biopsy specimens from disease-free premenopausal women, Frederick et al. [29] measured tissue folate and repeat element gene methylation and identified higher tissue folate in obese women and higher *LINE-1* methylation. In addition, contrasting much of what is observed for associations of folate with blood DNA methylation, increased breast tissue folate was associated with increased LINE-1 methylation. Another study of breast tissue DNA methylation from healthy women used a genome-scale approach and tested the relation of DNA methylation with eight SNPs in one-carbon metabolism pathway genes finding 57 differentially methylated CpGs associated with various genotypes in models adjusting for age, race, and BMI [30].

Outside of cancer, although there is a large body of literature examining the relation of diet and disease risk, particularly examining nutrients involved in one-carbon metabolism. However, approaches that include measures of epigenetic states such as DNA methylation remain limited, with some literature presenting the hypothesis that altered methylation status is associated with cardiovascular disease through folate status [31–33]. As part of the Singapore Chinese Health Study, higher rank (rather than measured value) of Alu and Satellite methylation was associated with elevated cardiovascular disease risk and conditions associated with its predisposure, but only in males, and not females [23]. It is unclear whether large cohort studies that could test associations of dietary factors with methylation alterations in normal and diseased states such as cardiovascular disease while accounting for important potential confounders such as BMI and alcohol remain unpublished or if additional investigation is necessary.

Other areas of interest for future study include the potential for differential contribution of nutrient factors dependent upon disease context (healthy vs. diseased), the potential risks of high-dose nutrient supplementation practices, and interactions between one-carbon network participants. Folate supplementation is associated with a reduced risk of many forms of cancer, though once disease is present it is possible that a relative increase in one-carbon nutrient levels may speed disease progression or invasiveness dependent upon the epigenetic character of the disease state. Increased availability of one-carbon groups could enable nucleotide synthesis for neoplastic cell division. Evidence of the relation between disease severity and one-carbon nutrient availability was provided in a study of breast tumor methylation from over 160 patients where increased folate intake (controlling for potential confounders such as alcohol intake) was associated with a methylation profile that was independently associated with increased tumor size [34]. Further, in a long-term study of elderly individuals supplementing with folic acid over a dozen years, high folate was associated with increased (albeit still low) levels of tumor suppressor DNA methylation (*CDKN2A*, *MLH1*, and *MGMT*) in blood [35]. However, in patients with liver cancer, Yeh et al. [36] observed that low plasma folate and low plasma LINE-1 methylation were both associated with worse prognosis. Certainly, relations of one-carbon pathway nutrients and DNA

methylation in carcinogenesis and disease progression are complex, and much of the little existing work includes one-carbon metabolism enzyme genotypes. An approach that included both primary human lesions along the continuum of cervical cancer progression with folate level data and cervical cancer cell lines treated with folate indicated that adequate folate maintained appropriate regulation of *FHIT* tumor suppressor methylation and gene expression [37]. Excessive folate supplementation may have potential adverse effects including masking of B_{12} deficiency, disruption of zinc function, or interfering with one-carbon homeostasis and additional study is required to elucidate these potential effects on epigenetic states in health and disease [38, 39].

The one-carbon pathway links nutrient availability with not only DNA methylation, but also with toxicant metabolism through glutathione and the glutathione transferase (GST) enzyme family (⊙ Fig. 12.1). Therefore, interactions between one-carbon pathway participants may modify associations between exposures, pathway participants, and methylation alterations. In fact, a study of reduced GST enzyme activity in a mouse model of Alzheimer's disease and neuronal health has shown that SAM can mediate the activity of GST enzymes [40]. This suggests that reactions dependent upon SAM are necessary for GST activity and that SAM may be a critical mediator of neuronal health [40]. Furthermore, it could be predicted that GST enzyme family genotypes (among other genotypes) will be shown to modify associations between certain exposures and DNA methylation alterations. Hence, integrative studies that incorporate genetics (genotype and/or alterations) with measures of epigenetic alterations will allow a more comprehensive understanding of the relation among exposures, epigenetic alterations, and genetic states. Such approaches are increasingly important as the extent to which one-carbon metabolism pathway enzyme genotypes interact with one-carbon levels and DNA methylation is now better established. More research is needed to ascertain best practices for assessing folate/methyl availability, perhaps comparing food-frequency-questionnaire data to homocysteine and/or SAM/SAH levels in a prospective manner. In addition, the recently released Illumina mouse DNA methylation array may offer opportunities for more controlled studies of the relation of folate intake, one-carbon metabolism, diet, and/or BMI with DNA methylation in mouse models, and such studies may inform approaches or target genes/regions for investigations in large cohort studies. Another line of future investigation that has the potential for value added as in the candidate gene study above from Li et al. [37] would be in vitro cell line or organoid approaches that complement epidemiologic investigation at genome scale.

12.2.2 Alcohol

Alcohol is known to interfere with folate absorption in the intestine and hepatic release of folate, and hence, supply to tissues [41]. Although it is a weak mutagen [42], by inhibiting the one-carbon metabolism network, alcohol may contribute to increased risk of disease in an epigenetic fashion. More specifically, ethanol can

interfere with several steps of methionine metabolism and can lead to activation of betaine homocysteine methyltransferase to activate a compensatory pathway for maintenance of SAM levels [43]. However, with continued exposure, the compensatory pathway cannot be maintained, and liver injury will result. In addition to liver injury, an etiologic role for alcohol in several cancers is well established [44]. For example, alcohol consumption increases colon cancer risk [45], though the association may be modified by dietary folate, and may be specific to LINE-1 hypomethylated tumors [46]. More specifically, subjects with high folate intake were less likely to develop LINE-1 hypomethylated colon cancers but subjects with high alcohol consumption had a significantly increased risk of developing LINE-1 hypomethylated colon cancers with no association for more highly methylated LINE-1 tumors [46]. Strong evidence of a general trend for decreased methylation associated with alcohol intake is available from a very large meta-analysis of 13 population-based cohorts with over 13,000 subjects investigating heavy drinking [47]. Not only do Liu et al. observe decreased methylation for 29 of the 30 CpGs whose methylation is most strongly associated with alcohol intake, but they also show a dramatic skew to negative regression coefficients on the volcano plot from the meta-EWAS for alcohol intake.

Beyond its potential to interfere with folate metabolism, alcohol has a well-recognized etiologic role in multiple cancers and may contribute to cancer risk through epigenetic alterations. For example, in head and neck squamous cell carcinoma (HNSCC), alcohol consumption has been significantly associated with reduced LINE-1 methylation [48], and HNSCC tumor methylation profiles are significantly and independently associated with alcohol intake [49]. An excess risk for each alcoholic drink per day of about 10% has been reported in multiple meta-analyses of prospective and case–control studies of breast cancer [50, 51], and a study of breast tumor DNA methylation demonstrated a significant, independent association (controlling for age, dietary folate, and other variables) between alcohol intake and tumor DNA methylation profile [34]. In nontumor breast tissue, a candidate gene approach that analyzed tissue folate and *CDKN2A* methylation observed alcohol consumption to be associated with lower breast folate, higher *CDKN2A* methylation, and lower gene expression [52]. More recently, a large study of alcohol and DNA methylation in both blood and breast tissue leveraged almost 3000 samples from the Sister Study prospective cohort and nearly 200 samples from the Normal Breast Study at the University of North Carolina and employed robust EWAS methods including adjustment for cell type proportions [53]. Thousands of CpGs were identified to have significant methylation changes associated with alcoholic drinks per week in blood. In addition to including an independent replication set, the authors also identified overlap between blood and breast tissue methylation associations with alcohol at nearly 100 CpGs. Similar to prior work, mean methylation across all measured CpGs was significantly lower for subjects with high alcohol consumption. Collectively, these studies suggest that a major carcinogenic mechanism of action of alcohol is interference with epigenetic regulation, in part through disruption of one-carbon metabolism.

Studies of alcohol and epigenetic states include comparisons of subjects discordant for alcohol use, as well as those focused on alcohol use disorders and efforts to develop DNA methylation biomarkers of alcohol exposures. A promising data set from a study of genome-scale blood DNA methylation in monozygotic twins discordant for alcohol consumption identified some evidence of differential methylation by alcohol consumption in adjusted models and included an independent replication cohort [54]. Although the work has generally strong study design that accounts for genetics, lack of adjustment for potential confounders such as cell type limits interpretability of the findings. Compared with healthy subjects, Alu repeat element methylation from blood DNA in Korean men with alcohol use disorder was significantly increased [55]. An investigation of blood methylation alterations in alcohol dependence in both European American and African American subjects that measured hundreds of candidate genes identified some limited evidence of population specific alterations associated with alcohol dependence [56]. In a pilot study comparing genome-scale blood DNA methylation in alcohol abstainers to heavy users identified over 8000 CpGs with differential methylation below after FDR correction that were enriched for pathways related to cell death and apoptosis though without adjustment for age, sex, ethnicity, or cell type [57]. More recently, a large meta-analysis investigated heavy drinking and identified strong evidence of alcohol related methylation changes in blood DNA from European ancestry subjects and African ancestry subjects [47]. In addition, monocyte-specific alcohol related methylation changes were identified. Perhaps most importantly, these authors offer a preliminary DNA methylation biomarker for heavy drinking with an area under the curve (AUC) range of 0.90–0.99 with 144 CpG sites that has potential utility for supplementing questionnaire data in epidemiologic studies and assist clinicians with a more accurate and objective approach to discern heavy drinking behavior in patients.

12.2.3 Aging

The aging process itself and differences in environment have been hypothesized to influence clinically significant changes in methylation profiles as individuals accumulate varying exposures with age. Early studies hypothesized that epigenetic variation is a cause of underlying differences in disease susceptibility among monozygotic twins, and in one study young twin pairs seemed more epigenetically similar than older monozygotic twins, though there were a limited number of twin pairs studied [58]. After the introduction of the Illumina methylation array, a study of ten different normal tissue types observed consistent associations between methylation and age at previously reported candidate genes, and indicated the context dependent nature of age-related methylation changes. Though it is a simplification, densely grouped CpG sites in CpG islands tend to be unmethylated, and sparsely grouped CpGs tend to be unmethylated. Thus, the direction of age-associated changes in methylation is dependent on the reference or baseline state of methylation for any given locus and this context-dependency of age-related changes became more

clearly established [59]. The observed pattern of age-associated methylation in [59] was also irrespective of tissue type, suggesting a common mechanism or dysregulation to explain these alterations. Similarly, important work from Teschendorff et al. in 2010 suggested that age-dependent methylation of polycomb-group target genes (genes suppressed in stem cells allowing differentiation) is a hallmark of cancer and is independent of gender, tissue type, or disease state, suggesting a mechanism for aging to predispose to carcinogenesis [60]. Since these and other early reports of age-related methylation in normal human tissues, major developments emerged that allow use of DNA methylation to predict chronological age, introducing a new sub-field of aging epigenetics that centers around age clocks. Initial DNA methylation age clocks from Horvath [61] and Hannum [62] uncovered new opportunities for aging research. Briefly, the Horvath DNA methylation age clock developed a pan-tissue predictor of chronological age using elastic net regression to select a few hundred CpG sites associated with age across many tissue types and over 7800 samples from 80 data sets [61]. One potential limitation of the Horvath clock was restriction of the input data to ~21,000 CpG sites available on both the Illumina 27k and 450k array (~27,000 and ~ 450,000 CpG sites respectively). Other major DNA methylation clocks include the Hannum methylation age predictor which is focused on peripheral blood and was built using input data from the 450k Illumina methylation array [62], and the Levine clock called DNAm PhenoAge [63]. One exciting opportunity for the use of DNA methylation age clocks is to test the relation of environmental exposures with biological aging. For instance, using a DNA methylation clock approach, long-term exposure to fine particulate matter ($PM_{2.5}$) has been shown to associate with increased epigenetic age acceleration in the KORA F4 cohort [64]. Similarly, participants in the Normative Aging Study were observed to have age acceleration in association with increased $PM_{2.5}$ exposure [65, 66]. Reports focusing on relations of epigenetic aging with environmental exposures are beginning to emerge [67], and tools to define epigenetic aging variables are poised for combination with methylation quantitative trait loci (mQTL) approaches [68] to address important questions and uncover biology at the intersection of aging and environmental exposures. The development and use of DNA methylation age clocks has been reviewed elsewhere [69], and additional details on age-related changes in DNA methylation are available in Chap. 10, and additional details on epigenetic age are available in Chap. 11.

12.2.4 Tobacco

Among exposures, the epigenetic epidemiology of tobacco smoking is relatively well-developed. Much of the initial work on DNA methylation in smoking-related tumors identified associations of tumor suppressor gene methylation with smoking variables using candidate gene approaches [70–73]. For example, methylation of the *RASSF1A* gene was associated with an earlier age at starting smoking in lung cancer patients [74, 75], suggesting that still growing adolescent lungs may be particularly susceptible to tobacco-related epigenetic events [75, 76]. Similarly, oral cancer

patients who started smoking at a younger age or who had higher pack-years smoked were more likely to have hypermethylated tumor suppressor genes [77]. An early approach to dense methylation profiling found that head and neck cancer methylation profiles were significantly associated with smoking intensity [49]. Then, using sputum and pre-malignant lung epithelium, work from several studies turned the focus from tumors to more specifically investigate the role of tobacco exposure in carcinogenesis, again identifying associations of tumor suppressor gene promoter hypermethylation with tobacco exposures [78–85]. However, as array-based methods gained visibility from studies in tumor, peripheral blood DNA methylation measures became more common.

An early genome-scale approach using the Illumina 27k array measured DNA methylation in blood from fewer than 200 current, former, and never smokers identified decreased *F2RL3* methylation associated with smoking and replicated the finding using a targeted approach [86]. The coagulation factor II receptor-like 3 gene (*F2RL3*) codes for protease-activated receptor 4 which functions in platelet activation. To follow-up on this discovery, the authors performed a follow-up study in over 1200 subjects in the KAROLA prospective cohort who had a history of cardiovascular pathology and found that low *F2RL3* methylation (consistent with smoking exposure), was associated with increased hazard of death, both cardiovascular-specific and all-cause [87]. Around the same time, the aryl hydrocarbon receptor repressor (*AHRR*) was identified as another gene of focus for tobacco-exposure-related DNA methylation. Lymphoblast cell lines (derived from blood) from participants in the Iowa Adoption Studies were profiled for genome-scale DNA methylation using the Illumina 450k array and identified *AHRR* hypomethylation associated with current smoking [88]. This finding was validated in alveolar macrophages from the same study. The aryl hydrocarbon receptor functions to metabolize byproducts of tobacco smoke and at the time, *AHRR* was already a known tumor suppressor gene. With measures of genome-scale blood DNA methylation in the NOWAC prospective cohort *AHRR* and *F2RL3* hypomethylation were associated with prospective risk of lung cancer. These findings were replicated in case–control pairs from three additional prospective cohorts with adjustment for cell type proportion providing further evidence linking tobacco-exposure-related DNA methylation alterations to carcinogenesis [89]. Further, in the Multi-Ethnic Study of Atherosclerosis, a novel association of monocyte *AHRR* methylation and carotid plaque scores was identified and validated in liver DNA [90].

Epigenome-wide association studies (EWASs) of DNA methylation and smoking repeatedly identified decreased *AHRR* and *F2RL3* methylation. Additional CpG loci associated with smoking were reported as sample sizes grew. The European Prospective Investigation into Cancer and Nutrition (EPIC) measured nearly 400 matched pairs of subjects nested in its prospective cohort with the Illumina 450k array and identified methylation of two intergenic regions associated with tobacco smoking status [91]. Further, this group not only provided evidence that effects of smoking on *AHRR* hypomethylation extended from blood to lung tissue, but that *AHRR* gene expression increased. At nearly the same time, peripheral blood DNA from over 2000 subjects in the KORA study F3 and F4 groups was measured

with the 450k array and significant associations with smoking were observed on all autosomes [92]. Importantly, methods to adjust for potential variation in peripheral blood cell type proportions [93, 94] were increasingly being adopted in EWAS studies. A very large meta-analysis of smoking that included 16 cohorts and nearly 16,000 subjects with blood-derived 450k methylation data with current, former, never smoking status and adjusted for cell type composition identified over 2600 CpGs with altered smoking-related methylation using a strict Bonferroni threshold, and nearly 18,000 CpGs at FDR < 0.05 [95]. Depending on the significance cutoff, only approximately 10–15% of smoking-associated differentially methylated CpGs were differentially methylated between current and former smokers indicating persistent effects of smoking on methylation. With the relation of smoking and DNA methylation solidifying in peripheral blood samples from healthy subjects, attention somewhat turned to newer questions raised about cell-specific methylation. Although decreased *AHRR* and *F2RL3* methylation related with smoking appeared somewhat robust to measured cell or tissue type, many associations of blood DNA methylation with smoking had small effect sizes suggesting that lower prevalence cell types in blood may specifically be affected. Using data sets from whole blood and peripheral blood mononuclear cells (PBMC, no granulocytes) data sets, Bauer et al. observed consistent *AHRR* hypomethylation between granulocytes and T-cells related with smoking, but cell-type dependent patterns of smoking-related methylation change at other CpGs [96]. Cell-dependent patterns of methylation associated with smoking were then more comprehensively studied using 450k arrays and reduced-representation bisulfite sequencing (RRBS), in both whole blood and purified cell subsets, uncovering a clearer picture of tobacco smoking-associated DNA methylation alterations [97].

With implications for life course epidemiology, prenatal smoking exposure also has been shown to alter epigenetic states in offspring. In some early work that used a birth cohort sample from the late 1950s and early 1960s, granulocyte repeat element methylation (Sat2, Alu, LINE), was measured in adult offspring and observed decreased Sat2 repeat methylation associated with prenatal tobacco smoke exposure [98]. Less than a year later, a genome-scale study with the 450k array measured methylation in over 1000 cord blood samples from the Norwegian Mother and Child Cohort Study and identified *AHRR* and *CYP1A1* methylation alterations among 10 genes and 26 CpGs that met a strict EWAS p-value threshold (Joubert 2012). In work measuring blood DNA methylation in over 500 children aged 5–12 from the Asthma BRIDGE study methylation of 19 CpGs was associated with prenatal smoke exposure and associations for CpGs in *FRMD4A* and C11orf52 were independently replicated [99]. Extending tests of the hypothesis that prenatal smoking exposure has long-lasting postnatal effects, Lee et al. measured blood methylation during adolescence to test associations with prenatal smoking using the 450k array and including adjustment for cell type. CpG sites in *MYO1G* and *CNTNAP2* associated with prenatal smoking exposure were validated in independent samples with data from birth, childhood, and adolescence [100]. An arguably even more detailed to track the persistence of prenatal smoking exposure on methylation started with 800 cord blood samples and described dose-dependency of altered methylation and uncovered

evidence using serial samples that prenatal smoking exposure effects on methylation are reversible for some genes, but show persistence for others [101]. Although not a study of persistent effects, a comparative meta-analysis identified extensive overlap of gene pathways for results from newborn and adult smoking-related DNA methylation alterations [102]. Additional evidence for persistent epigenetic effects of prenatal smoking exposure was shown by Rauschert et al., who measured effects in 17-year-old adolescents from the Raine Study [103]. These authors also showed that smoking-related DNA methylation alterations were associated with cardiometabolic risk factors. In follow-up work, they also present a DNA methylation score for prenatal smoking exposure identified with an elastic net approach [104]. Additional opportunities exist to understand potential mediating effects of prenatal tobacco exposure on health outcomes over the life course, develop quantitative biomarkers of direct and indirect exposures.

12.3 Environmental Toxicants and Epigenetic Alterations

12.3.1 Endocrine Disruptors

Exposures to endocrine disrupting chemicals such as diethylstilbestrol (DES), bisphenol A (BPA), and per-and polyfluoroalkyl substances (PFASs) are of particular concern in the context of development. DES is a non-genotoxic carcinogen with developmental stage-specificity, that in the past was administered to women during pregnancy to prevent miscarriages [105]. In utero exposure of mice to DES has been shown to result in the hypermethylation of the developmentally critical (specifically to uterine organogenesis) *Hoxa10* gene [106]. Epidemiologic evidence from individuals exposed to DES during the first 3 months in utero indicates an increase in vaginal clear cell carcinoma incidence and reproductive disorders [107]. In addition, grandchildren of DES exposed women reported higher incidences of rare reproductive disorders; whether this reflects detection bias or possibly implicates a role for epigenetic transgenerational inheritance remains to be clarified [107]. Though the mechanism through which DES establishes altered epigenetic marks capable of transgenerational inheritance remains unclear, a model for the epigenetic effects of DES has been proposed by Ruden et al. drawing similarities between DES and Hsp90, which acts to play a role in modifying H3K4 methylation by increasing the activity of the H3K4 methyltransferase *SMYD3*, thereby altering epigenetic control of various genes [108]. Since the first edition of this chapter was published, a few additional investigations of DES have examined the multigenerational consequences of this potent EDC. A study of women whose mothers were exposed to DES in-utero (i.e., third generation of DES exposure during pregnancy) were at heightened risk of abnormal menstruation [109]. While a multi-generational study (mothers, their children, and their grandchildren) of the Nurses' Health Study II found that the grandchildren of grandmothers that used DES during pregnancy had increased odds of attention deficit hyperactivity disorder (ADHD), with even greater odds when DES was taken during the first trimester

[110]. However, further investigation of the potential transgenerational and epigenetic effects of DES still requires longer-term follow-up in multi-generational cohorts.

BPA is a monomer used in the production of polycarbonate plastic, a flame retardant, a fungicide, and a surface-coating for everyday objects [111]. BPA has been shown to readily cross the placental barrier and accumulate both in the placenta and the fetus [112]. Animal studies have demonstrated that developmental exposure to BPA can alter epigenetic profiles. Briefly, Dolinoy et al. showed that in utero BPA exposure decreases CpG methylation in agouti mice and that methyl donor supplementation negated BPA related hypomethylation [113]. In human placental cell lines BPA exposure has been shown to alter miRNA expression levels [114]. More specifically, miR-146a was strongly induced by BPA treatment and resulted in both slower proliferation rate and higher sensitivity to the DNA damaging agent bleomycin [114]. While BPA is still ubiquitous, replacement bisphenols with alternative but similar chemical structures have been increasingly produced, such as bisphenol F (BPF) and bisphenol S [115]. Recent publications have demonstrated that all three of BPA, BPS, and BPF are associated with differentially methylated CpGs in breast cancer cells, and the perturbations to the methylome were strongest for BPA, followed by BPS then BPF, and these were enriched for genes involved in cancer-pathways [116]. Another recent study of MCF7 breast cancer cells showed that BPA exposure resulted in hypermethylation of multiple tumor suppressors, and similarly, a separate study of prostate cancer cells also observed perturbed methylation in the promoter regions of tumor suppressors [117]. In the context of prenatal exposure, maternal BPA concentrations have been associated with differential methylation in cord blood samples, while BPS and BPF were not, although this may be due, in part, to the small sample size and lower exposure concentrations of BPF and BPS in this study [118]. Additionally, a large epidemiologic study found that children with higher gestational exposure to BPF had increased DNA methylation at *GRIN2B* during childhood, and that higher methylation at this CpG was associated with cognitive performance [119]. This builds upon prior findings from this group, which demonstrated that *GRIN2B* methylation and expression were perturbed in rats, and the DNA methylation was higher in *GRIN2B* in a sub-sample of their larger cohort [120]. These recent studies further demonstrate that bisphenols are important endocrine disrupting chemicals that can perturb epigenetic regulation, while the emergence of new compounds warrants continued investigation.

PFASs are a family of chemical compounds with carbon chains and carbon-fluorine bonds. This family includes thousands of fluorinated substances that are used in numerous consumer products [121] to repel water and oil, many of which are highly pervasive and persistent in the environment [122]. These compounds can act as EDCs and have been linked to increased cholesterol and dyslipidemia, higher blood pressure, liver disease [123], and immunotoxicity [124], while prenatal exposures are associated with decreased birth weight, postnatal growth [125], and other children's health outcomes. Epidemiologic studies have begun to characterize some of the epigenetic effects of PFAS in human populations. In adults, PFOS has

been associated with increased global DNA methylation levels in blood leukocytes in one study [126], while another study of sperm DNA methylation identified no association with global methylation levels [127]; the different tissue types in these two studies may explain their divergent findings. Another small study of adults observed PFOS-associated differential methylation at genes involved in carcinogenesis, immune activity, and metabolic functions [128], although these findings need to be confirmed in larger independent study populations. Three studies of prenatal PFAS exposures have examined effects on global DNA methylation in cord blood. One observed global hypomethylation associated with higher PFOA [129], another observed hypomethylation with PFOS [130], while the third observed no relation with global DNA methylation levels, but instead found that PFOA was linked to lower methylation levels at *IGF2* gene [131]. Four studies have utilized DNA methylation microarrays to perform epigenome-wide association studies of prenatal PFAS exposure. Two of these were small studies of cord blood that reported differential methylation with PFAO and with PFOS [132, 133], while one larger study identified then replicated differential methylation at CpGs within *ZBTB7A* (PFOS), *USP2-AS1* (PFOA), *TCP11L2* (PFOA), and *NTN1* (PFOA) [134]. Additionally, one study of placental epigenomic and transcriptomic responses to PFAS identified 39 differentially methylated CpGs, and 4 differentially expressed genes, that were associated with five different PFAS from maternal serum and were in turn associated with neonatal anthropometrics [135]. Of note, PFDA and PFUnDA yielded the most statistically significant finding (*SPG20*) and the largest magnitude of effect (*ILF3*), rather than the more commonly studied PFAS compounds [135]. In vitro studies also provide some insights into the molecular cascades and epigenetic responses that are induced by PFAS exposures. A study of cultured trophoblasts demonstrated increased microRNA expression (miR29-b) and DNA hypomethylation with PFOS exposure, suggesting an oxidative stress response [136], while another study showed that PFOS, PFOA, and GenX affected the expression of genes involved in migration and inflammatory signaling [137]; these findings indicate that PFAS may induce a stress response in placental tissues. While a study of human mesenchymal stem cells demonstrated that environmentally relevant levels of PFOS induce alterations to DNA methylation during adipogenesis [138]. These studies provide supporting evidence that epigenetic regulation and genetic activity are responsive to PFAS exposures, and that these effects are induced in developmentally relevant tissues. However, this is still an emerging field of research and these findings are preliminary given the small sample sizes in the majority of currently published epidemiologic studies. Additionally, most of the epidemiologic and experimental studies to date have focused on PFOS and PFOA, which are no longer produced in the United States, but have been replaced with alternative PFAS in many products. Thus, a major gap remains for less common and more recently emergent PFAS compounds.

12.3.2 Ionizing and Ultraviolet Radiation

Exposures to ionizing radiation have long been linked to cancer, through DNA damage in the form of large deletions and in some cases point mutation [139–141]. Limited epidemiological data are available on the association of ionizing radiation with DNA methylation alterations, though rat lung tumors induced by ^{239}plutonium exposure exhibited tumor suppressor gene silencing [142]. This is consistent with human lung tumor *CDKN2A* methylation in workers from the Russian Mayak weapons-grade plutonium plant, where the prevalence of altered methylation exhibited a dose–response with radiation internal exposure dose [143]. Methylation of *CDKN2A* has also been linked to reactive oxygen species produced by radiation exposures [144], and murine models of radiation-induced lymphoma have also demonstrated hypermethylation of *Cdkn2b* (encoding p15ink4b [145–147]). The relation of childhood cancer treatment with adverse health outcomes is established. A childhood cancer survival cohort, the St. Jude Lifetime Cohort Study, measured genome-scale blood DNA methylation with the 450k array in over 2500 subjects to study potential relations of chemotherapy and radiation therapy with DNA methylation and cardiometabolic outcomes. Over 300 CpG sites were significantly associated with each of chest and abdominal radiation therapy, and nearly 250 CpGs sites were associated with pelvic radiation therapy [148]. Importantly, Song et al. also showed that CpG methylation alterations associated with radiation therapy mediated associations of treatment with hypercholesterolemia and hypertriglyceridemia.

The genetic effects of UV radiation are well described, and epigenetic effects of UV radiation include spontaneous deamination of methylcytosine, though there is limited epidemiologic data on UV exposure and DNA methylation alterations. Reduced LINE-1 repeat element methylation in blood DNA was observed in a study of over 200 participants in Australia, who used personal exposure measures over a six week period [149]. Work measuring blood methylation and assessing lifetime UV exposure with questionnaire data in subjects from the Norwegian Women in Cancer cohort identified a CpG with altered methylation at a genome-wide significance threshold but it did not replicate, and two other sites that were weakly associated with other UV exposure variables [150]. A more complete understanding of methylation alterations associated with UV radiation in relation with skin carcinogenesis is needed. Opportunities to measure methylation in skin directly, leverage dermoscopic photo-aging related with UV exposure assessment, and deploy emerging cell deconvolution approaches for skin all hold promise for future work.

12.3.3 Arsenic

Arsenic exposure has been associated with DNA methylation alterations in non-pathologic as well as tumor tissues and there is some data suggesting that changes in miRNA expression and histone tail modifications are also associated

with exposure to arsenic. Although the mode(s) of arsenic's carcinogenicity is not completely clear, there is some speculation about arsenite-generated free radicals and reactive oxygen species lead to genotoxic damage [151]. However, in vitro exposures to inorganic arsenic species have demonstrated dose-dependent increases in promoter region hypermethylation of CpG sites, although not to those resulting in altered gene expression, as well as to the occurrence of genome-wide hypomethylation [152–154]. In human bladder cancer, relatively low levels of inorganic arsenic exposure have been associated with methylation of *RASSF1A* and *PRSS3* (but not *CDKN2A*) [155], and in mouse models of methyl- or folate-deficient diets, arsenic exposure through water supply led to hypomethylation in hepatic-derived DNA [156], as well as to increases in chromosomal aberrations in blood lymphocytes [157]. Thus, one potential mechanism for arsenic-related carcinogenicity is via the depletion of S-adenosyl methionine (SAM, the universal methyl donor) due to the metabolism of inorganic arsenic to its methylated forms resulting in altered DNA methylation.

As more studies are conducted a complex picture of dose-dependent DNA methylation alterations with arsenic exposure is beginning to emerge. Chanda et al. have shown that arsenic exposure can result in *CDKN2A* hypermethylation in human blood DNA, but that a subgroup of cases had hypomethylation with high arsenic exposure [158]. More recently, a group of some of the same authors has better quantified the relation showing that exposures to 250–500 ug/L of arsenic in drinking water results in global hypermethylation (^{3}H methyl group uptake), but that >500 ug/L of arsenic results in global hypomethylation [159]. In another study of peripheral blood DNA methylation by ^{3}H methyl group uptake, Pilsner et al. showed increased DNA methylation to be associated with urinary and plasma arsenic and plasma folate, and that the association between arsenic and methylation was modified by folate in that it was restricted to individuals with high plasma folate [160]. This same group has shown that folic acid supplementation lowers blood arsenic [161], and folate deficiency, hyperhomocysteinemia, and leukocyte hypomethylation are associated with arsenic-induced skin lesions [162]. A number of studies have also demonstrated prenatal exposures to arsenic can be linked with observed variation in DNA methylation in newborn cord blood, and importantly, identified similarities in the estimates of reductions of the proportions of CD8+ T-cells that likely underlies some of the observed differences in DNA methylation observed [163, 164]. More recent evidence from this work also suggests that these observed changes may be mediating arsenic's impacts on gestational age [165]. Importantly, a recent analysis using Mendelian Randomization provides greater evidence for a causal mechanism underlying the observed relations between arsenic exposure and peripheral blood DNA methylation alterations [166].

Beyond interfering with one-carbon metabolism and affecting DNA methylation, treatment of human lymphoblastoid cells with sodium arsenite led to global increases in miRNA expression [167]. A number of human population studies have also begun to identify relations between arsenic exposure and altered miRNA expression in multiple tissue types, suggesting this may be an additional model of genotoxicity for arsenic [168]. Finally, altered chromatin modifications have been

linked to arsenic exposures in vitro [169] and human studies have observed global alterations in histone post-translational modifications, although more gene-specific findings have not been clearly identified due to technical limitations of their assessment in population studies [170]. Nonetheless, additional studies are necessary to further elucidate the associations between arsenic and all major forms of epigenetic alteration with particular attention being given to dose and modification by one-carbon metabolism pathway participants.

12.3.4 Cadmium

Cadmium is an established human carcinogen [171] and is ranked seventh on the Agency for Toxic Substances and Disease Registry substance priority list [172]. Recent cell line, animal, and epidemiologic studies have characterized potential roles for cadmium-induced responses in DNA methylation, histone modifications, and miRNA and mRNA expression, in its carcinogenic and toxic effects. A comparison of cadmium-transformed prostate epithelial cells versus control RWPE-1 cells showed that cadmium exposure resulted in increased DNA methyltransferase activity (*DNMT3b*), increased global hypermethylation, and decreased expression of two tumor suppressor genes (*RASSF1A* and *p16*), likely due to increased methylation of the promoter regions for these genes [173]. A mouse model of Cd-induced malignant transformation observed increased global DNA methylation which was associated with the overexpression of *DNMT1* and *DNMT3a*. These authors also found that DNA repair genes were downregulated (*hMSH2*, *ERCC1*, *XRCC1*, and *hOGG1*), and DNA damage increased [174]. While a study of human B lymphoblast cells demonstrated that Cd exposure resulted in increased cellular proliferation and increased *DNMT1* and *DNMT3b* expression, with repression of the *p16* tumor repressor [175]. A study of Cd-induced transformation in immortalized human bronchial epithelial cells observed increased methylation of histone tails, specifically H3K4me3 and H3K9me2, with corresponding inhibition of histone demethylases and that these changes persisted even after cadmium was removed [176]. While Cd-exposed MCF-7 breast cancer cells exhibited altered DNA methylation, as well as differential expression of regulatory RNA (lncRNA and miRNA) and mRNA expression, within biological pathways involved in Wnt-signaling, metabolism, and HPV infection [177]. These experimental studies demonstrate that cadmium impacts the activities of the epigenetic machinery involved in establishing and maintaining methylation patterns, and that transcriptional activities of multiple genes, including tumor suppressors, may be affected by these responses.

Cadmium is also a developmental toxicant and exposure during the early developmental period can have adverse effects on pregnancy outcomes, as well as children's health [178]. Additionally, epigenetic responses to prenatal cadmium exposure appear to be involved in these effects [179]. Three epigenome-wide association studies of prenatal cadmium exposure have been performed in humans. One examined the relations between maternal blood cadmium concentrations with

DNA methylation levels in cord blood and in peripheral blood during early childhood, identifying potential sex-specific effects [180]. Another focused on placental tissues and found that placental Cd levels were associated with differential DNA methylation and expression of genes (*TNFAIP2*, *ACOT7*, and *RORA*) involved in inflammatory activities; differential expression of these genes was also associated with differences in birth weight [181]. Both of these EWASs measured DNA methylation with microarrays. While a third EWAS performed whole genome bisulfite sequencing in maternal blood and cord blood, and observed differentially methylated regions (DMRs) that were associated with Cd, and these DMRs were enriched for imprinting control regions (ICR) [182]. Imprinted genes play critical roles in early growth and development, and their dysregulation can have long-term consequences on growth and metabolism [183], and on neurodevelopment [184]. An additional study of the 74 putative imprinted genes in human placenta found that increasing Cd exposure was associated with higher expression of *DLX5*, reduction of *h19,* and reduction of *NDN*, which were consistent across two independent studies [185]. Similar to the above EWAS, this study of imprinted genes also observed sex-specific effects, where *CPA4*, *GRB10*, and *ILK* were differentially expressed only among female placenta [185]. Another study of several imprinted genes in cord blood found that maternal blood Cd levels were associated with increased DNA methylation in a DMR that regulates *PEG3* expression [186].

In addition to the above epidemiologic studies, experimental in vivo and in vitro models also provide evidence that altered epigenetic regulation is involved in the developmental toxicity of prenatal Cd. A mouse model of maternal Cd exposure showed that *PEG10* expression was repressed and *CDK1NC* expression was up-regulated, with corresponding changes in promoter DNA methylation, in the placental tissues of Cd-exposed mice [187]. Another mouse study showed that maternal Cd impairs embryonic development, increases the activity of histone deacetylase 1 with corresponding decreases in histone acetylation (H4K8ac and H4K12ac), and significantly decreases DNA methylation of the *h19* gene [188]. Additionally, cultured placental trophoblast cells that were exposed to Cd have inhibited trophoblast migration and altered TGF-B signaling [189], which was associated with perturbed expression of miR-26a which is also responsive to cadmium exposure [190]. Another study linked increased Cd exposure to reduced 11 beta-HSD2 expression (*HSD11B2*), a gene that involved in fetal growth restriction, in cultured human trophoblast cells [191]. These studies add to the mounting evidence that cadmium exposure perturbs epigenetic activities, and may be involved in the carcinogenic and developmental toxicity of this metal. However, the above studies employed a variety of study designs, some with small sample sizes, and measured Cd and epigenetic features in multiple different biological matrices. Additional research is needed to validate the above findings.

12.3.5 Other Metals

Evidence for epigenetic alterations due to exposure to other metals in humans also has been observed for lead, mercury, and chromium. A study of metal-rich particulate matter exposure among electric-furnace steel plant workers has shown significant alterations of microRNA expression in peripheral blood leukocytes [192]. In the Normative Aging Study, patella lead levels were associated with reduced global DNA methylation (LINE-1 elements) though not Alu repeat regions [193]. Neonatal lead exposure also has been shown to result in reduced infant cord blood repeat element methylation [194]. In project Viva, an EWAS observed a small number of CpGs with differential methylation in cord blood associated with prenatal lead exposure including some evidence suggesting brain-specific gene alterations that was independently validated [195].

Certainly mercury is a recognized neurotoxic metal and emerging work has described impacts of prenatal exposures on DNA methylation as well as cognitive outcomes. For instance, EWAS approaches identified *PON1* methylation alterations in males that persisted into childhood [196]. Another EWAS of prenatal mercury exposure identified enrichment for CpG island shore methylation which could poise those regions for spreading altered methylation into the island itself [197]. Altered 5-hydroxymethylation levels also have been observed in association with prenatal mercury exposure [198]. Accumulating evidence for prenatal mercury exposure related with DNA methylation alterations has led to a meta-analysis including up to seven studies with over 1400 subjects total and measures of methylation during childhood to examine persistence. Though the effect sizes were relatively modest, CpG sites in three genes *MED31*, *GRK1*, and *GGH* were identified [199].

Chromium has industrial applications including chrome plating and stainless steel welding and is implicated in occupationally related lung cancers [200]. Mechanisms of chromium carcinogenicity include adduct formation [201] and DNA strand breakage [202], and some limited evidence suggests the potential for chromium to have epigenetic effects. In cases of chromate-induced lung cancer, Kondo et al. showed that although prevalence of *CDKN2A* methylation was similar between chromate and non-chromate lung cancers, chromate lung cancers were more likely to have methylated *CDKN2A* with increasing duration of occupational exposure [203]. An in vitro exposure of mouse hepatoma cells to chromium showed transcriptional repression of *Cyp1a1* by local cross-linking of Hdac and Dnmt1 and altered histone marks [204]. However, much additional study of this diminishing exposure has not emerged.

12.3.6 Particulates/Air Pollution

Environmental particulate matter (PM) and air pollutant exposures are harmful and have been linked to increased risk of lung cancer [205], metabolic disorders, and increased morbidity and mortality from cardiovascular and respiratory illnesses [206]. In some of the early work from human subjects, exposure to particulate matter

of <10 μm (PM_{10}) in steel plant workers was associated with significantly lower peripheral blood *NOS2* promoter methylation [207]. In addition, this same study reported significantly reduced methylation of LINE-1 and Alu repeat elements in blood DNA associated with long-term PM_{10} exposure [207]. Consistent with this, a recent report examining the effect of ambient particulate pollutants on repeat element methylation in subjects from the Boston area Normative Aging Study found significantly decreased LINE-1 methylation following recent exposure to higher black carbon [208]. Several studies in adults have followed testing associations of $PM_{2.5}$, PM_{10}, nitric oxides, polychlorinated biphenyls (PCBs), and black carbon with repeat element methylation leading to emergence of systematic reviews and at least one meta-analysis. Results from some of the largest studies of $PM_{2.5}$ and repeat element methylation in selected participants (cross-sectional) from the Sister Study [209], the Multi-Ethnic Study of Atherosclerosis [210], and the Normative Aging Study [211] did not identify significant associations. However, across these studies and in a recent meta-analysis including these and other study results indicate a trend of decreased Alu and LINE-1 repeat element methylation levels with $PM_{2.5}$ exposure [212]. Although less work has been done in children, early prenatal exposures to air pollution indicate similar trends of decreased repeat element methylation in cord blood [213, 214], newborn blood [215], and placenta [216]. Some insight into a potential mechanism for air pollution associated epigenetic alterations, and genetic interactions that modify effects of exposure has been provided from results showing that increases in $PM_{2.5}$ were associated with increases in plasma homocysteine (part of one-carbon metabolism), and glutathione transferase theta genotype (an antioxidant dependent upon homocysteine levels) [217]. This same group of authors also reported that increases in mean air pollution PM_{10} concentrations did not significantly alter fasting or postmethionine-load total homocysteine in non-smokers, but was associated with significantly increased homocysteine levels in smokers, suggesting interactions between exposures may contribute to epigenetic alterations [218].

Genome-wide measures of DNA methylation have provided evidence of epigenetic alterations related with air pollution. Several cross-sectional studies in adults leveraging cohorts [219–222], some with sample sizes over a thousand and into the thousands of subjects have identified specific CpG sites with significantly altered methylation in association with $PM_{2.5}$, PM_{10}, NO_2, and PCB exposures. Although specific CpG loci and genes identified vary, and appear lower in scope than tobacco-related alterations, multiple studies identified increased methylation of *NXN* which is a negative regulator of the Wnt pathway. A randomized controlled trial of short-term exposure to $PM_{2.5}$ (in natural Shanghai conditions representing high-exposure), with genome-scale blood DNA methylation data that adjusted for cell type proportions identified dozens of CpGs with significantly altered methylation [223]. Although studies in newborn cord blood and placenta tissues are fewer, meta-analyses have begun to emerge and include replication that leverage data from childhood blood samples. For example, among several CpG sites with reasonably consistent findings among newborns for relation with $PM_{2.5}$, or PM_{10}, independent replication in at least one of two follow-up cohorts of 7–9 year old children was observed for CpGs in

FAM13A, NOTCH4, and *P4HA2* [224]. Similarly, a meta-analysis of prenatal NO_2 exposure identified CpGs with altered methylation in mitochondria related genes and the site in *SLC25A28* replicated in an independent set of samples from older children [225]. Although these studies seek in part to identify persistent effects of air pollution effects on DNA methylation into childhood, potential negative health implications of effects observed at birth that do not necessarily persist should not be discounted. In addition, expectedly, limited tissue types have been studied and the patterns and scope of exposure-related alterations incurred during development on other tissues are unclear and likely underappreciated. Multiple reviews of air pollution and DNA methylation also have emerged and provided utility [226–228, 212]. Air pollution exposure studies also have leveraged genome-scale DNA methylation data to study biological age acceleration, finding generally increased biological age associated with exposure, more detail is provided above in section (12.2.3 Aging). One such study that focused on age acceleration related with $PM_{2.5,}$ PM_{10}, and NO_2 exposures clustered the analysis on profiles of $PM_{2.5}$ components indicating that the makeup of $PM_{2.5}$ resulted in varying associations with biological age acceleration [229]. Specifically, in over 2700 female subjects a $PM_{2.5}$ cluster with higher earth crust elements had over a 6 year age acceleration associated with an IQR increase in $PM_{2.5}$, whereas a $PM_{2.5}$ cluster with low sulfur had two year age acceleration, and a $PM_{2.5}$ cluster with low nitrate was modestly inversely associated with age acceleration. Varying associations of $PM_{2.5}$ biological age from DNA methylation data based on its profile provide an opportunity to further investigate prior age acceleration findings, as well as EWAS findings. In addition, future prospective studies are warranted, as is the use of leading-edge approaches to identify cell type proportions that can be applied to directly test associations of immune profile with air pollution and more comprehensively adjust for potential confounding in EWAS approaches.

12.3.7 Asbestos

Exposure to asbestos is the main risk factor for malignant pleural mesothelioma with approximately 80% of cases reporting a known exposure to asbestos [230]. Importantly, in contrast to tobacco smoke and radiation, asbestos is known to be a weak mutagen [231], and there have been reports of altered tumor methylation in mesothelioma using candidate gene approaches [232–237]. Following the introduction of array-based methylation measures, initial work identified that a quantitative asbestos exposure measure was associated with altered methylation at over 100 discrete CpG loci, and that in almost all cases (94%) there was increased methylation associated with increased exposure [238]. Further, overall methylation profiles for mesotheliomas were significantly associated with quantitative asbestos exposure burden [238]. One of the challenges with assessing the relation of methylation alterations with asbestos exposure is that most studies of asbestos-related cancers like mesothelioma and lung cancer do not include quantitative measures of exposure. Using the 450k Illumina array, Kettunen et al. studied lung tumor DNA methylation in asbestos exposed and non-exposed patients and identified significant DNA

methylation alterations that were validated in an independent set of samples [239]. In addition to study of tumor methylation alterations, comparisons of mesothelioma patient blood DNA methylation to controls subjects identified and validated differential methylation including for *FOXK1* in one study with a training and testing set approach, which is known to biologically interact with *BAP1* an often-mutated gene in mesothelioma [240]. Another case–control EWAS in blood identified *FKBP5* hypomethylation and *MLLT1* hypermethylation in cases [241]. Interest in the continued development of screening approaches for mesothelioma among asbestos exposed individuals may benefit from multitarget approaches that incorporate epigenetic alterations associated with asbestos exposures and disease. Identifying methylation alterations associated with asbestos exposure [242], and single-walled carbon nanotubes [243, 244] in vitro may benefit epidemiologic studies.

12.4 Environmental Exposures and Imprinting and Development

Genomic imprinting involves epigenetic silencing of a single allele distinguishing the maternal or paternal inheritance pattern of a specific locus [245]. Imprinting and its importance in development as well as alterations to imprinting and its effect on disease, including cancer, have been thoroughly reviewed by Feinberg and colleagues [246, 247, 245]. Allele-specific expression associated with imprinting is based upon promoter CpG island methylation which to repress transcription and insulate neighboring repressive elements [248, 249]. Alterations to the imprinted status observed after prolonged culture of mouse embryonic stem cells, lead to a number of phenotypic abnormalities, through both gains and losses of appropriate methylation of these regions [250]. In humans, it has been suggested that assisted reproductive technology (ART) may affect the epigenetics of development thus altering imprinting status. In sheep and cattle, large offspring syndrome has been linked to culture of pre-implantation embryos [251] and in humans, associations between Angelman syndrome [252] and Beckwith–Wiedemann [253] syndrome and in-vitro fertilization have been reported, and are related to changes in imprinting status of specific genes. More on epigenome changes during development can be found in Chap. 5, details of imprinting disorders are available in Chap. 8, and assisted reproductive technologies are covered in Chap. 7.

12.5 Mechanisms

Different mechanisms likely can contribute to epigenetic alterations dependent upon context and exposure and ascribing a mode by which environmental exposures drive or select epigenetic alteration is difficult. Although exposure-related EWAS has begun to leverage gene pathway and ontology approaches to detail implications of exposures on gene deregulation, to date there has been less consistent use of genomic context enrichment and transcription factor enrichment analyses. As distinct

enzymes are thought responsible for maintenance DNA methylation dependent on CpG density [254], contextual elements may differentially affect the likelihood of exposure-related alterations. Incorporating contextual details (e.g., repeat element, promoter, enhancer, existing methylation state, nucleosome position, and transcription factor binding sites), in exposure-related EWAS is one way to add value to existing and future work. Even in early age-related methylation analysis, Rakyan et al. showed that hypermethylation was more likely to occur at promoters associated with bivalent chromatin domains [255]. Also, contextual elements may differentially affect the likelihood of exposure-related alterations dependent upon cell type and the blood samples used in most environmental epigenetic epidemiology constitute a complex mixture of cell types.

One mechanism by which an exposure may result in altered DNA methylation is increased reactive oxygen species as a result of inflammatory response. More specifically, it has been reported that 5-hydroxymethylcytosine can be generated by oxidation of 5-methylcytosine [256], and both 5-methylcytosine adjacent to 8-oxoguanine, and 5-hydroxymethylcytosine have been shown to inhibit binding of methyl-CpG binding protein 2, a critical epigenetic regulator that recruits cytosine methyltransferases and histone deacetylases [257]. It is also known that 5-hydroxymethylcytosine is not recognized as 5-methylcytosine by the maintenance methyltransferase DNMT1, and hence, may lead to aberrant loss of methylation during cell replication [258]. Additional base alterations occur via neutrophil and eosinophil peroxidase-derived HOCl and HOBr which can react with DNA to form 5-chlorocytosine and 5-bromocytosine, respectively [259]. These halogenated cytosines can be mistaken by DNMT1 as 5-methylcytosine during replication, thus providing a potential mechanism for inflammation-induced aberrant hypermethylation [258].

Another mechanism that has been hypothesized is a spreading process, whereby CpG methylation from within the gene migrates into the promoter due to the loss of protective boundary elements which normally protect promoters from CpG island methylation [260–262]. Alternatively, for example, in aging, a spreading process subsequent to a stochastic methylation alteration (in pathologically normal tissue) without functional consequences for gene expression, may allow the initiation of a spreading process that eventually confers an altered expression phenotype. Age or exposure-related drift of normal epigenomes (without prominent changes in gene expression) may confer significantly increased risk of conversion to a pathologic phenotype by enhancing both the likelihood and frequency of methylation events that ultimately result in altered expression or genomic stability. For example, in the context of acquired "non-functional" CpG methylation in the promoter region of an aged individual, continued stochastic methylation events (e.g., "methylation spreading") increase the chance of methylation induced silencing at that promoter (or silencing of another locus through action at a distance via silencing of other important regions such as enhancers), and hence, progression to a pathologic phenotype. Certainly, this hypothesis is especially plausible for the many diseases of aging such as cancer. One demonstrated example in breast cancer observed CpG island shore methylation in normal breast tissue adjacent to tumor compared with

distant normal breast tissue as a seed for spreading methylation to the CpG island in tumors from matched subjects [263]. Alternatively, aberrant CpG methylation that silences a gene on a single allele may not appear to have a functional consequence if the complementary allele can provide compensatory expression. As a result, for example, clusters of cellular clones with mono-allelic gene expression could contribute to an increased risk of progression to a pathologic phenotype. Mutagenic and non-mutagenic compounds can impart selective pressures on cellular clones, and selection pressure may be more important than direct epi/genotoxicity. The modes by which exposures drive the clonal selection of genetically or epigenetically altered cells or modulate the epigenetic landscape of individual cells require further exploration.

12.6 Development of Environmental Epigenomics

The fields of environmental genomics and toxicogenomics are bringing the wealth of knowledge gained through the human genome project to the fields of environmental health and toxicology. This will stimulate novel approaches to understanding the modes of action of various chemical and physical agents. It has also been suggested that collaborative efforts between the various disciplines of environmental science, genomics, systems biology, medicine, and basic biology may hold great promise to elucidate more thoroughly the effect of exposures on human health [264]. In addition to studying the effects at the genotoxic and transcriptional levels, it is important to also consider the role of carcinogenic, teratogenic, and developmentally toxic exposures on the epigenome, and on clonal selection, again requiring multidisciplinary work from in-vitro and model systems to epidemiologic studies of human disease. Established genome-scale approaches allow high resolution analysis of DNA methylation alterations with the latest generation of CpG methylation arrays interrogating nearly to a million CpG loci at costs that permit population-level studies. In addition, the availability of methylation array technology for genome-scale measures in mouse enables important new opportunities to compare the effects of controlled experimental exposures on epigenetics with epidemiological data. In the future, it is possible that third wave sequencing approaches will allow accurate direct measurement of cytosine modifications without the need for bisulfite modification. Mammalian stem cells may also prove extremely useful in bettering the understanding of exposure-related epigenetics in development and lineage commitment, providing insightful models for demonstrating how these pathways become dysregulated in disease pathophysiology. Incorporation of inferred cell type information in EWAS models and considering potential direct relation of exposures on cell type proportion profiles also is expected to improve our understanding of environmental epigenomics. Finally, appropriate epidemiologic studies, including both case-series and prospective designs, requiring defined and consistent methodologies and data collection will be crucial to understanding the effect of environmental exposures on the human epigenome. This work is urgently needed to better understand the biology of epigenetic alterations and the effects of toxic

exposures on these disease-associated somatic alterations. Detailed understanding of existing exposures may inform better prediction of the toxic potential of new compounds and mixtures allowing for more targeted, and appropriate, disease prevention strategies.

References

1. Morgan HD, Santos F, Green K, Dean W, Reik W (2005) Epigenetic reprogramming in mammals. Hum Mol Genet 14 Spec No 1:R47–58. https://doi.org/10.1093/hmg/ddi114
2. Gluckman PD, Hanson MA, Cooper C, Thornburg KL (2008) Effect of in utero and early-life conditions on adult health and disease. N Engl J Med 359(1):61–73. https://doi.org/10.1056/NEJMra0708473
3. Li E (2002) Chromatin modification and epigenetic reprogramming in mammalian development. Nat Rev Genet 3(9):662–673. https://doi.org/10.1038/nrg887
4. Daly LE, Kirke PN, Molloy A, Weir DG, Scott JM (1995) Folate levels and neural tube defects. Implications for prevention JAMA 274(21):1698–1702. https://doi.org/10.1001/jama.1995.03530210052030
5. Dunlevy LP, Burren KA, Mills K, Chitty LS, Copp AJ, Greene ND (2006) Integrity of the methylation cycle is essential for mammalian neural tube closure. Birth Defects Res A Clin Mol Teratol 76(7):544–552. https://doi.org/10.1002/bdra.20286
6. Wang L, Wang F, Guan J, Le J, Wu L, Zou J, Zhao H, Pei L, Zheng X, Zhang T (2010) Relation between hypomethylation of long interspersed nucleotide elements and risk of neural tube defects. Am J Clin Nutr 91(5):1359–1367. https://doi.org/10.3945/ajcn.2009.28858
7. Yan L, Zhao L, Long Y, Zou P, Ji G, Gu A, Zhao P (2012) Association of the maternal MTHFR C677T polymorphism with susceptibility to neural tube defects in offsprings: evidence from 25 case-control studies. PLoS One 7(10):e41689. https://doi.org/10.1371/journal.pone.0041689
8. Waterland RA, Michels KB (2007) Epigenetic epidemiology of the developmental origins hypothesis. Annu Rev Nutr 27:363–388. https://doi.org/10.1146/annurev.nutr.27.061406.093705
9. Gluckman PD, Hanson MA (2004) Developmental origins of disease paradigm: a mechanistic and evolutionary perspective. Pediatr Res 56(3):311–317. https://doi.org/10.1203/01.PDR.0000135998.08025.FB
10. Johansson M, Relton C, Ueland PM, Vollset SE, Midttun O, Nygard O, Slimani N, Boffetta P, Jenab M, Clavel-Chapelon F, Boutron-Ruault MC, Fagherazzi G, Kaaks R, Rohrmann S, Boeing H, Weikert C, Bueno-de-Mesquita HB, Ros MM, van Gils CH, Peeters PH, Agudo A, Barricarte A, Navarro C, Rodriguez L, Sanchez MJ, Larranaga N, Khaw KT, Wareham N, Allen NE, Crowe F, Gallo V, Norat T, Krogh V, Masala G, Panico S, Sacerdote C, Tumino R, Trichopoulou A, Lagiou P, Trichopoulos D, Rasmuson T, Hallmans G, Riboli E, Vineis P, Brennan P (2010) Serum B vitamin levels and risk of lung cancer. JAMA 303(23):2377–2385. https://doi.org/10.1001/jama.2010.808
11. Freudenheim JL, Marshall JR, Vena JE, Laughlin R, Brasure JR, Swanson MK, Nemoto T, Graham S (1996) Premenopausal breast cancer risk and intake of vegetables, fruits, and related nutrients. J Natl Cancer Inst 88(6):340–348. https://doi.org/10.1093/jnci/88.6.340
12. Graham S, Hellmann R, Marshall J, Freudenheim J, Vena J, Swanson M, Zielezny M, Nemoto T, Stubbe N, Raimondo T (1991) Nutritional epidemiology of postmenopausal breast cancer in western New York. Am J Epidemiol 134(6):552–566. https://doi.org/10.1093/oxfordjournals.aje.a116129
13. Martinez ME, Marshall JR, Giovannucci E (2008) Diet and cancer prevention: the roles of observation and experimentation. Nat Rev Cancer 8(9):694–703. https://doi.org/10.1038/nrc2441

14. Kiss E, Forika G, Mohacsi R, Nemeth Z, Krenacs T, Dank M (2021) Methyl-donors can induce apoptosis and attenuate both the Akt and the Erk1/2 mediated proliferation pathways in breast and lung cancer cell lines. Int J Mol Sci 22(7). https://doi.org/10.3390/ijms22073598
15. Kim DH, Smith-Warner SA, Spiegelman D, Yaun SS, Colditz GA, Freudenheim JL, Giovannucci E, Goldbohm RA, Graham S, Harnack L, Jacobs EJ, Leitzmann M, Mannisto S, Miller AB, Potter JD, Rohan TE, Schatzkin A, Speizer FE, Stevens VL, Stolzenberg-Solomon R, Terry P, Toniolo P, Weijenberg MP, Willett WC, Wolk A, Zeleniuch-Jacquotte A, Hunter DJ (2010) Pooled analyses of 13 prospective cohort studies on folate intake and colon cancer. Cancer Causes Control 21(11):1919–1930. https://doi.org/10.1007/s10552-010-9620-8
16. Zhao Y, Guo C, Hu H, Zheng L, Ma J, Jiang L, Zhao E, Li H (2017) Folate intake, serum folate levels and esophageal cancer risk: an overall and dose-response meta-analysis. Oncotarget 8(6):10458–10469. https://doi.org/10.18632/oncotarget.14432
17. Qiang Y, Li Q, Xin Y, Fang X, Tian Y, Ma J, Wang J, Wang Q, Zhang R, Wang J, Wang F (2018) Intake of dietary one-carbon metabolism-related B vitamins and the risk of Esophageal cancer: a dose-response meta-analysis. Nutrients 10(7). https://doi.org/10.3390/nu10070835
18. de Vogel S, Wouters KA, Gottschalk RW, van Schooten FJ, de Goeij AF, de Bruine AP, Goldbohm RA, van den Brandt PA, van Engeland M, Weijenberg MP (2011) Dietary methyl donors, methyl metabolizing enzymes, and epigenetic regulators: diet-gene interactions and promoter CpG island hypermethylation in colorectal cancer. Cancer Causes Control 22(1):1–12. https://doi.org/10.1007/s10552-010-9659-6
19. Huang CY, Abulimiti A, Zhang X, Feng XL, Luo H, Chen YM, Fang YJ, Zhang CX (2020) Dietary B vitamin and methionine intakes and risk for colorectal cancer: a case-control study in China. Br J Nutr 123(11):1277–1289. https://doi.org/10.1017/S0007114520000501
20. van Driel LM, Eijkemans MJ, de Jonge R, de Vries JH, van Meurs JB, Steegers EA, Steegers-Theunissen RP (2009) Body mass index is an important determinant of methylation biomarkers in women of reproductive ages. J Nutr 139(12):2315–2321. https://doi.org/10.3945/jn.109.109710
21. Park HJ, Bailey LB, Shade DC, Hausman DB, Hohos NM, Meagher RB, Kauwell GP, Lewis RD, Smith AK (2017) Distinctions in gene-specific changes in DNA methylation in response to folic acid supplementation between women with normal weight and obesity. Obes Res Clin Pract 11(6):665–676. https://doi.org/10.1016/j.orcp.2017.06.004
22. Pauwels S, Ghosh M, Duca RC, Bekaert B, Freson K, Huybrechts I, Langie SAS, Koppen G, Devlieger R, Godderis L (2017) Maternal intake of methyl-group donors affects DNA methylation of metabolic genes in infants. Clin Epigenetics 9:16. https://doi.org/10.1186/s13148-017-0321-y
23. Caffrey A, Irwin RE, McNulty H, Strain JJ, Lees-Murdock DJ, McNulty BA, Ward M, Walsh CP, Pentieva K (2018) Gene-specific DNA methylation in newborns in response to folic acid supplementation during the second and third trimesters of pregnancy: epigenetic analysis from a randomized controlled trial. Am J Clin Nutr 107(4):566–575. https://doi.org/10.1093/ajcn/nqx069
24. Bakulski KM, Dou JF, Feinberg JI, Brieger KK, Croen LA, Hertz-Picciotto I, Newschaffer CJ, Schmidt RJ, Fallin MD (2020) Prenatal multivitamin use and MTHFR genotype are associated with Newborn cord blood DNA methylation. Int J Environ Res Public Health 17(24). https://doi.org/10.3390/ijerph17249190
25. Ingrosso D, Cimmino A, Perna AF, Masella L, De Santo NG, De Bonis ML, Vacca M, D'Esposito M, D'Urso M, Galletti P, Zappia V (2003) Folate treatment and unbalanced methylation and changes of allelic expression induced by hyperhomocysteinaemia in patients with uraemia. Lancet 361(9370):1693–1699. https://doi.org/10.1016/S0140-6736(03)13372-7
26. Pufulete M, Al-Ghnaniem R, Leather AJ, Appleby P, Gout S, Terry C, Emery PW, Sanders TA (2003) Folate status, genomic DNA hypomethylation, and risk of colorectal adenoma and cancer: a case control study. Gastroenterology 124(5):1240–1248. https://doi.org/10.1016/s0016-5085(03)00279-8

27. Pufulete M, Al-Ghnaniem R, Rennie JA, Appleby P, Harris N, Gout S, Emery PW, Sanders TA (2005) Influence of folate status on genomic DNA methylation in colonic mucosa of subjects without colorectal adenoma or cancer. Br J Cancer 92(5):838–842. https://doi.org/10.1038/sj.bjc.6602439
28. Hsiung DT, Marsit CJ, Houseman EA, Eddy K, Furniss CS, McClean MD, Kelsey KT (2007) Global DNA methylation level in whole blood as a biomarker in head and neck squamous cell carcinoma. Cancer Epidemiol Biomark Prev 16(1):108–114. https://doi.org/10.1158/1055-9965.EPI-06-0636
29. Frederick AM, Guo C, Meyer A, Yan L, Schneider SS, Liu Z (2021) The influence of obesity on folate status, DNA methylation and cancer-related gene expression in normal breast tissues from premenopausal women. Epigenetics 16(4):458–467. https://doi.org/10.1080/15592294.2020.1805687
30. Song MA, Brasky TM, Marian C, Weng DY, Taslim C, Llanos AA, Dumitrescu RG, Liu Z, Mason JB, Spear SL, Kallakury BV, Freudenheim JL, Shields PG (2016) Genetic variation in one-carbon metabolism in relation to genome-wide DNA methylation in breast tissue from heathy women. Carcinogenesis 37(5):471–480. https://doi.org/10.1093/carcin/bgw030
31. Gluckman PD, Hanson MA, Buklijas T, Low FM, Beedle AS (2009) Epigenetic mechanisms that underpin metabolic and cardiovascular diseases. Nat Rev Endocrinol 5(7):401–408. https://doi.org/10.1038/nrendo.2009.102
32. Kim M, Long TI, Arakawa K, Wang R, Yu MC, Laird PW (2010) DNA methylation as a biomarker for cardiovascular disease risk. PLoS One 5(3):e9692. https://doi.org/10.1371/journal.pone.0009692
33. Papait R, Condorelli G (2010) Epigenetics in heart failure. Ann N Y Acad Sci 1188:159–164. https://doi.org/10.1111/j.1749-6632.2009.05096.x
34. Christensen BC, Kelsey KT, Zheng S, Houseman EA, Marsit CJ, Wrensch MR, Wiemels JL, Nelson HH, Karagas MR, Kushi LH, Kwan ML, Wiencke JK (2010) Breast cancer DNA methylation profiles are associated with tumor size and alcohol and folate intake. PLoS Genet 6(7):e1001043. https://doi.org/10.1371/journal.pgen.1001043
35. Sanchez H, Hossain MB, Lera L, Hirsch S, Albala C, Uauy R, Broberg K, Ronco AM (2017) High levels of circulating folate concentrations are associated with DNA methylation of tumor suppressor and repair genes p16, MLH1, and MGMT in elderly Chileans. Clin Epigenetics 9:74. https://doi.org/10.1186/s13148-017-0374-y
36. Yeh CC, Goyal A, Shen J, Wu HC, Strauss JA, Wang Q, Gurvich I, Safyan RA, Manji GA, Gamble MV, Siegel AB, Santella RM (2017) Global level of plasma DNA methylation is associated with overall survival in patients with hepatocellular carcinoma. Ann Surg Oncol 24(12):3788–3795. https://doi.org/10.1245/s10434-017-5913-4
37. Li Q, Ding L, Jing N, Liu C, Yang Z, Chen F, Hou L, Wang J (2018) Folate deficiency and aberrant DNA methylation and expression of FHIT gene were associated with cervical pathogenesis. Oncol Lett 15(2):1963–1972. https://doi.org/10.3892/ol.2017.7471
38. Hathcock JN (1997) Vitamins and minerals: efficacy and safety. Am J Clin Nutr 66(2):427–437. https://doi.org/10.1093/ajcn/66.2.427
39. Sauer J, Mason JB, Choi SW (2009) Too much folate: a risk factor for cancer and cardiovascular disease? Curr Opin Clin Nutr Metab Care 12(1):30–36. https://doi.org/10.1097/MCO.0b013e32831cec62
40. Tchantchou F, Graves M, Falcone D, Shea TB (2008) S-adenosylmethionine mediates glutathione efficacy by increasing glutathione S-transferase activity: implications for S-adenosyl methionine as a neuroprotective dietary supplement. J Alzheimers Dis 14(3):323–328. https://doi.org/10.3233/jad-2008-14306
41. Hillman RS, Steinberg SE (1982) The effects of alcohol on folate metabolism. Annu Rev Med 33:345–354. https://doi.org/10.1146/annurev.me.33.020182.002021
42. Dumitrescu RG, Shields PG (2005) The etiology of alcohol-induced breast cancer. Alcohol 35(3):213–225. https://doi.org/10.1016/j.alcohol.2005.04.005

43. Kharbanda KK (2009) Alcoholic liver disease and methionine metabolism. Semin Liver Dis 29(2):155–165. https://doi.org/10.1055/s-0029-1214371
44. Alcohol drinking. IARC Working Group, Lyon, 13–20 October 1987 (1988) IARC Monogr Eval Carcinog Risks Hum 44:1–378
45. Giovannucci E, Rimm EB, Ascherio A, Stampfer MJ, Colditz GA, Willett WC (1995) Alcohol, low-methionine – low-folate diets, and risk of colon cancer in men. J Natl Cancer Inst 87 (4):265–273. https://doi.org/10.1093/jnci/87.4.265
46. Schernhammer ES, Giovannucci E, Kawasaki T, Rosner B, Fuchs CS, Ogino S (2010) Dietary folate, alcohol and B vitamins in relation to LINE-1 hypomethylation in colon cancer. Gut 59(6):794–799. https://doi.org/10.1136/gut.2009.183707
47. Liu C, Marioni RE, Hedman AK, Pfeiffer L, Tsai PC, Reynolds LM, Just AC, Duan Q, Boer CG, Tanaka T, Elks CE, Aslibekyan S, Brody JA, Kuhnel B, Herder C, Almli LM, Zhi D, Wang Y, Huan T, Yao C, Mendelson MM, Joehanes R, Liang L, Love SA, Guan W, Shah S, McRae AF, Kretschmer A, Prokisch H, Strauch K, Peters A, Visscher PM, Wray NR, Guo X, Wiggins KL, Smith AK, Binder EB, Ressler KJ, Irvin MR, Absher DM, Hernandez D, Ferrucci L, Bandinelli S, Lohman K, Ding J, Trevisi L, Gustafsson S, Sandling JH, Stolk L, Uitterlinden AG, Yet I, Castillo-Fernandez JE, Spector TD, Schwartz JD, Vokonas P, Lind L, Li Y, Fornage M, Arnett DK, Wareham NJ, Sotoodehnia N, Ong KK, van Meurs JBJ, Conneely KN, Baccarelli AA, Deary IJ, Bell JT, North KE, Liu Y, Waldenberger M, London SJ, Ingelsson E, Levy D (2018) A DNA methylation biomarker of alcohol consumption. Mol Psychiatry 23(2):422–433. https://doi.org/10.1038/mp.2016.192
48. Smith IM, Mydlarz WK, Mithani SK, Califano JA (2007) DNA global hypomethylation in squamous cell head and neck cancer associated with smoking, alcohol consumption and stage. Int J Cancer 121(8):1724–1728. https://doi.org/10.1002/ijc.22889
49. Marsit CJ, Christensen BC, Houseman EA, Karagas MR, Wrensch MR, Yeh RF, Nelson HH, Wiemels JL, Zheng S, Posner MR, McClean MD, Wiencke JK, Kelsey KT (2009) Epigenetic profiling reveals etiologically distinct patterns of DNA methylation in head and neck squamous cell carcinoma. Carcinogenesis 30(3):416–422. https://doi.org/10.1093/carcin/bgp006
50. Hamajima N, Hirose K, Tajima K, Rohan T, Calle EE, Heath CW, Jr., Coates RJ, Liff JM, Talamini R, Chantarakul N, Koetsawang S, Rachawat D, Morabia A, Schuman L, Stewart W, Szklo M, Bain C, Schofield F, Siskind V, Band P, Coldman AJ, Gallagher RP, Hislop TG, Yang P, Kolonel LM, Nomura AM, Hu J, Johnson KC, Mao Y, De Sanjose S, Lee N, Marchbanks P, Ory HW, Peterson HB, Wilson HG, Wingo PA, Ebeling K, Kunde D, Nishan P, Hopper JL, Colditz G, Gajalanski V, Martin N, Pardthaisong T, Silpisornkosol S, Theetranont C, Boosiri B, Chutivongse S, Jimakorn P, Virutamasen P, Wongsrichanalai C, Ewertz M, Adami HO, Bergkvist L, Magnusson C, Persson I, Chang-Claude J, Paul C, Skegg DC, Spears GF, Boyle P, Evstifeeva T, Daling JR, Hutchinson WB, Malone K, Noonan EA, Stanford JL, Thomas DB, Weiss NS, White E, Andrieu N, Bremond A, Clavel F, Gairard B, Lansac J, Piana L, Renaud R, Izquierdo A, Viladiu P, Cuevas HR, Ontiveros P, Palet A, Salazar SB, Aristizabel N, Cuadros A, Tryggvadottir L, Tulinius H, Bachelot A, Le MG, Peto J, Franceschi S, Lubin F, Modan B, Ron E, Wax Y, Friedman GD, Hiatt RA, Levi F, Bishop T, Kosmelj K, Primic-Zakelj M, Ravnihar B, Stare J, Beeson WL, Fraser G, Bullbrook RD, Cuzick J, Duffy SW, Fentiman IS, Hayward JL, Wang DY, McMichael AJ, McPherson K, Hanson RL, Leske MC, Mahoney MC, Nasca PC, Varma AO, Weinstein AL, Moller TR, Olsson H, Ranstam J, Goldbohm RA, van den Brandt PA, Apelo RA, Baens J, de la Cruz JR, Javier B, Lacaya LB, Ngelangel CA, La Vecchia C, Negri E, Marubini E, Ferraroni M, Gerber M, Richardson S, Segala C, Gatei D, Kenya P, Kungu A, Mati JG, Brinton LA, Hoover R, Schairer C, Spirtas R, Lee HP, Rookus MA, van Leeuwen FE, Schoenberg JA, McCredie M, Gammon MD, Clarke EA, Jones L, Neil A, Vessey M, Yeates D, Appleby P, Banks E, Beral V, Bull D, Crossley B, Goodill A, Green J, Hermon C, Key T, Langston N, Lewis C, Reeves G, Collins R, Doll R, Peto R, Mabuchi K, Preston D, Hannaford P, Kay C, Rosero-Bixby L, Gao YT, Jin F, Yuan JM, Wei HY, Yun T, Zhiheng C, Berry G, Cooper Booth J, Jelihovsky T, MacLennan R, Shearman R, Wang QS,

Baines CJ, Miller AB, Wall C, Lund E, Stalsberg H, Shu XO, Zheng W, Katsouyanni K, Trichopoulou A, Trichopoulos D, Dabancens A, Martinez L, Molina R, Salas O, Alexander FE, Anderson K, Folsom AR, Hulka BS, Bernstein L, Enger S, Haile RW, Paganini-Hill A, Pike MC, Ross RK, Ursin G, Yu MC, Longnecker MP, Newcomb P, Bergkvist L, Kalache A, Farley TM, Holck S, Meirik O, Collaborative Group on Hormonal Factors in Breast C (2002) Alcohol, tobacco and breast cancer – collaborative reanalysis of individual data from 53 epidemiological studies, including 58,515 women with breast cancer and 95,067 women without the disease. Br J Cancer 87(11):1234–1245. https://doi.org/10.1038/sj.bjc.6600596

51. Key J, Hodgson S, Omar RZ, Jensen TK, Thompson SG, Boobis AR, Davies DS, Elliott P (2006) Meta-analysis of studies of alcohol and breast cancer with consideration of the methodological issues. Cancer Causes Control 17(6):759–770. https://doi.org/10.1007/s10552-006-0011-0
52. Llanos AA, Dumitrescu RG, Brasky TM, Liu Z, Mason JB, Marian C, Makambi KH, Spear SL, Kallakury BV, Freudenheim JL, Shields PG (2015) Relationships among folate, alcohol consumption, gene variants in one-carbon metabolism and p16INK4a methylation and expression in healthy breast tissues. Carcinogenesis 36(1):60–67. https://doi.org/10.1093/carcin/bgu219
53. Wilson LE, Xu Z, Harlid S, White AJ, Troester MA, Sandler DP, Taylor JA (2019) Alcohol and DNA methylation: an epigenome-wide association study in blood and Normal breast tissue. Am J Epidemiol 188(6):1055–1065. https://doi.org/10.1093/aje/kwz032
54. Lu M, Xueying Q, Hexiang P, Wenjing G, Hagg S, Weihua C, Chunxiao L, Canqing Y, Jun L, Zengchang P, Liming C, Hua W, Xianping W, Yunzhang W, Liming L (2021) Genome-wide associations between alcohol consumption and blood DNA methylation: evidence from twin study. Epigenomics 13(12):939–951. https://doi.org/10.2217/epi-2021-0039
55. Kim DS, Kim YH, Lee WK, Na YK, Hong HS (2016) Effect of alcohol consumption on peripheral blood Alu methylation in Korean men. Biomarkers 21(3):243–248. https://doi.org/10.3109/1354750X.2015.1134661
56. Zhang H, Herman AI, Kranzler HR, Anton RF, Zhao H, Zheng W, Gelernter J (2013) Array-based profiling of DNA methylation changes associated with alcohol dependence. Alcohol Clin Exp Res 37(Suppl 1):E108–E115. https://doi.org/10.1111/j.1530-0277.2012.01928.x
57. Philibert RA, Penaluna B, White T, Shires S, Gunter T, Liesveld J, Erwin C, Hollenbeck N, Osborn T (2014) A pilot examination of the genome-wide DNA methylation signatures of subjects entering and exiting short-term alcohol dependence treatment programs. Epigenetics 9(9):1212–1219. https://doi.org/10.4161/epi.32252
58. Fraga MF, Ballestar E, Paz MF, Ropero S, Setien F, Ballestar ML, Heine-Suner D, Cigudosa JC, Urioste M, Benitez J, Boix-Chornet M, Sanchez-Aguilera A, Ling C, Carlsson E, Poulsen P, Vaag A, Stephan Z, Spector TD, Wu YZ, Plass C, Esteller M (2005) Epigenetic differences arise during the lifetime of monozygotic twins. Proc Natl Acad Sci U S A 102(30):10604–10609. https://doi.org/10.1073/pnas.0500398102
59. Christensen BC, Houseman EA, Marsit CJ, Zheng S, Wrensch MR, Wiemels JL, Nelson HH, Karagas MR, Padbury JF, Bueno R, Sugarbaker DJ, Yeh RF, Wiencke JK, Kelsey KT (2009) Aging and environmental exposures alter tissue-specific DNA methylation dependent upon CpG island context. PLoS Genet 5(8):e1000602. https://doi.org/10.1371/journal.pgen.1000602
60. Teschendorff AE, Menon U, Gentry-Maharaj A, Ramus SJ, Weisenberger DJ, Shen H, Campan M, Noushmehr H, Bell CG, Maxwell AP, Savage DA, Mueller-Holzner E, Marth C, Kocjan G, Gayther SA, Jones A, Beck S, Wagner W, Laird PW, Jacobs IJ, Widschwendter M (2010) Age-dependent DNA methylation of genes that are suppressed in stem cells is a hallmark of cancer. Genome Res 20(4):440–446. https://doi.org/10.1101/gr.103606.109
61. Horvath S (2013) DNA methylation age of human tissues and cell types. Genome Biol 14(10):R115. https://doi.org/10.1186/gb-2013-14-10-r115

62. Hannum G, Guinney J, Zhao L, Zhang L, Hughes G, Sadda S, Klotzle B, Bibikova M, Fan JB, Gao Y, Deconde R, Chen M, Rajapakse I, Friend S, Ideker T, Zhang K (2013) Genome-wide methylation profiles reveal quantitative views of human aging rates. Mol Cell 49(2):359–367. https://doi.org/10.1016/j.molcel.2012.10.016
63. Levine ME, Lu AT, Quach A, Chen BH, Assimes TL, Bandinelli S, Hou L, Baccarelli AA, Stewart JD, Li Y, Whitsel EA, Wilson JG, Reiner AP, Aviv A, Lohman K, Liu Y, Ferrucci L, Horvath S (2018) An epigenetic biomarker of aging for lifespan and healthspan. Aging (Albany NY) 10(4):573–591. https://doi.org/10.18632/aging.101414
64. Ward-Caviness CK, Nwanaji-Enwerem JC, Wolf K, Wahl S, Colicino E, Trevisi L, Kloog I, Just AC, Vokonas P, Cyrys J, Gieger C, Schwartz J, Baccarelli AA, Schneider A, Peters A (2016) Long-term exposure to air pollution is associated with biological aging. Oncotarget 7(46):74510–74525. https://doi.org/10.18632/oncotarget.12903
65. Nwanaji-Enwerem JC, Colicino E, Trevisi L, Kloog I, Just AC, Shen J, Brennan K, Dereix A, Hou L, Vokonas P, Schwartz J, Baccarelli AA (2016) Long-term ambient particle exposures and blood DNA methylation age: findings from the VA normative aging study. Environ Epigenet 2(2). https://doi.org/10.1093/eep/dvw006
66. Nwanaji-Enwerem JC, Dai L, Colicino E, Oulhote Y, Di Q, Kloog I, Just AC, Hou L, Vokonas P, Baccarelli AA, Weisskopf MG, Schwartz JD (2017) Associations between long-term exposure to PM2.5 component species and blood DNA methylation age in the elderly: the VA normative aging study. Environ Int 102:57–65. https://doi.org/10.1016/j.envint.2016.12.024
67. Dhingra R, Nwanaji-Enwerem JC, Samet M, Ward-Caviness CK (2018) DNA methylation age-environmental influences, health impacts, and its role in environmental epidemiology. Curr Environ Health Rep 5(3):317–327. https://doi.org/10.1007/s40572-018-0203-2
68. Min JL, Hemani G, Hannon E, Dekkers KF, Castillo-Fernandez J, Luijk R, Carnero-Montoro-E, Lawson DJ, Burrows K, Suderman M, Bretherick AD, Richardson TG, Klughammer J, Iotchkova V, Sharp G, Al Khleifat A, Shatunov A, Iacoangeli A, McArdle WL, Ho KM, Kumar A, Soderhall C, Soriano-Tarraga C, Giralt-Steinhauer E, Kazmi N, Mason D, McRae AF, Corcoran DL, Sugden K, Kasela S, Cardona A, Day FR, Cugliari G, Viberti C, Guarrera S, Lerro M, Gupta R, Bollepalli S, Mandaviya P, Zeng Y, Clarke TK, Walker RM, Schmoll V, Czamara D, Ruiz-Arenas C, Rezwan FI, Marioni RE, Lin T, Awaloff Y, Germain M, Aissi D, Zwamborn R, van Eijk K, Dekker A, van Dongen J, Hottenga JJ, Willemsen G, Xu CJ, Barturen G, Catala-Moll F, Kerick M, Wang C, Melton P, Elliott HR, Shin J, Bernard M, Yet I, Smart M, Gorrie-Stone T, Consortium B, Shaw C, Al Chalabi A, Ring SM, Pershagen G, Melen E, Jimenez-Conde J, Roquer J, Lawlor DA, Wright J, Martin NG, Montgomery GW, Moffitt TE, Poulton R, Esko T, Milani L, Metspalu A, Perry JRB, Ong KK, Wareham NJ, Matullo G, Sacerdote C, Panico S, Caspi A, Arseneault L, Gagnon F, Ollikainen M, Kaprio J, Felix JF, Rivadeneira F, Tiemeier H, van IMH, Uitterlinden AG, Jaddoe VWV, Haley C, McIntosh AM, Evans KL, Murray A, Raikkonen K, Lahti J, Nohr EA, Sorensen TIA, Hansen T, Morgen CS, Binder EB, Lucae S, Gonzalez JR, Bustamante M, Sunyer J, Holloway JW, Karmaus W, Zhang H, Deary IJ, Wray NR, Starr JM, Beekman M, van Heemst D, Slagboom PE, Morange PE, Tregouet DA, Veldink JH, Davies GE, de Geus EJC, Boomsma DI, Vonk JM, Brunekreef B, Koppelman GH, Alarcon-Riquelme ME, Huang RC, Pennell CE, van Meurs J, Ikram MA, Hughes AD, Tillin T, Chaturvedi N, Pausova Z, Paus T, Spector TD, Kumari M, Schalkwyk LC, Visscher PM, Davey Smith G, Bock C, Gaunt TR, Bell JT, Heijmans BT, Mill J, Relton CL (2021) Genomic and phenotypic insights from an atlas of genetic effects on DNA methylation. Nat Genet 53(9):1311–1321. https://doi.org/10.1038/s41588-021-00923-x
69. Bell CG, Lowe R, Adams PD, Baccarelli AA, Beck S, Bell JT, Christensen BC, Gladyshev VN, Heijmans BT, Horvath S, Ideker T, Issa JJ, Kelsey KT, Marioni RE, Reik W, Relton CL, Schalkwyk LC, Teschendorff AE, Wagner W, Zhang K, Rakyan VK (2019) DNA methylation aging clocks: challenges and recommendations. Genome Biol 20(1):249. https://doi.org/10.1186/s13059-019-1824-y

70. Divine KK, Pulling LC, Marron-Terada PG, Liechty KC, Kang T, Schwartz AG, Bocklage TJ, Coons TA, Gilliland FD, Belinsky SA (2005) Multiplicity of abnormal promoter methylation in lung adenocarcinomas from smokers and never smokers. Int J Cancer 114(3):400–405. https://doi.org/10.1002/ijc.20761
71. Hou M, Morishita Y, Iljima T, Inadome Y, Mase K, Dai Y, Noguchi M (1999) DNA methylation and expression of p16(INK4A) gene in pulmonary adenocarcinoma and anthracosis in background lung. Int J Cancer 84(6):609–613. https://doi.org/10.1002/(sici)1097-0215(19991222)84:6<609::aid-ijc12>3.0.co;2-q
72. Kim DH, Nelson HH, Wiencke JK, Zheng S, Christiani DC, Wain JC, Mark EJ, Kelsey KT (2001) p16(INK4a) and histology-specific methylation of CpG islands by exposure to tobacco smoke in non-small cell lung cancer. Cancer Res 61(8):3419–3424
73. Toyooka S, Maruyama R, Toyooka KO, McLerran D, Feng Z, Fukuyama Y, Virmani AK, Zochbauer-Muller S, Tsukuda K, Sugio K, Shimizu N, Shimizu K, Lee H, Chen CY, Fong KM, Gilcrease M, Roth JA, Minna JD, Gazdar AF (2003) Smoke exposure, histologic type and geography-related differences in the methylation profiles of non-small cell lung cancer. Int J Cancer 103(2):153–160. https://doi.org/10.1002/ijc.10787
74. Kim DH, Kim JS, Ji YI, Shim YM, Kim H, Han J, Park J (2003) Hypermethylation of RASSF1A promoter is associated with the age at starting smoking and a poor prognosis in primary non-small cell lung cancer. Cancer Res 63(13):3743–3746
75. Marsit CJ, Kim DH, Liu M, Hinds PW, Wiencke JK, Nelson HH, Kelsey KT (2005) Hypermethylation of RASSF1A and BLU tumor suppressor genes in non-small cell lung cancer: implications for tobacco smoking during adolescence. Int J Cancer 114(2):219–223. https://doi.org/10.1002/ijc.20714
76. Wiencke JK, Kelsey KT (2002) Teen smoking, field cancerization, and a "critical period" hypothesis for lung cancer susceptibility. Environ Health Perspect 110(6):555–558. https://doi.org/10.1289/ehp.02110555
77. Hasegawa M, Nelson HH, Peters E, Ringstrom E, Posner M, Kelsey KT (2002) Patterns of gene promoter methylation in squamous cell cancer of the head and neck. Oncogene 21(27):4231–4236. https://doi.org/10.1038/sj.onc.1205528
78. Belinsky SA, Nikula KJ, Palmisano WA, Michels R, Saccomanno G, Gabrielson E, Baylin SB, Herman JG (1998) Aberrant methylation of p16(INK4a) is an early event in lung cancer and a potential biomarker for early diagnosis. Proc Natl Acad Sci U S A 95(20):11891–11896. https://doi.org/10.1073/pnas.95.20.11891
79. Kersting M, Friedl C, Kraus A, Behn M, Pankow W, Schuermann M (2000) Differential frequencies of p16(INK4a) promoter hypermethylation, p53 mutation, and K-ras mutation in exfoliative material mark the development of lung cancer in symptomatic chronic smokers. J Clin Oncol 18(18):3221–3229. https://doi.org/10.1200/JCO.2000.18.18.3221
80. Palmisano WA, Divine KK, Saccomanno G, Gilliland FD, Baylin SB, Herman JG, Belinsky SA (2000) Predicting lung cancer by detecting aberrant promoter methylation in sputum. Cancer Res 60(21):5954–5958
81. Sozzi G, Pastorino U, Moiraghi L, Tagliabue E, Pezzella F, Ghirelli C, Tornielli S, Sard L, Huebner K, Pierotti MA, Croce CM, Pilotti S (1998) Loss of FHIT function in lung cancer and preinvasive bronchial lesions. Cancer Res 58(22):5032–5037
82. von Zeidler SV, Miracca EC, Nagai MA, Birman EG (2004) Hypermethylation of the p16 gene in normal oral mucosa of smokers. Int J Mol Med 14(5):807–811. https://doi.org/10.3892/ijmm.14.5.807
83. Zochbauer-Muller S, Fong KM, Maitra A, Lam S, Geradts J, Ashfaq R, Virmani AK, Milchgrub S, Gazdar AF, Minna JD (2001) 5' CpG island methylation of the FHIT gene is correlated with loss of gene expression in lung and breast cancer. Cancer Res 61(9):3581–3585
84. Zochbauer-Muller S, Lam S, Toyooka S, Virmani AK, Toyooka KO, Seidl S, Minna JD, Gazdar AF (2003) Aberrant methylation of multiple genes in the upper aerodigestive tract epithelium of heavy smokers. Int J Cancer 107(4):612–616. https://doi.org/10.1002/ijc.11458

85. Kulkarni V, Saranath D (2004) Concurrent hypermethylation of multiple regulatory genes in chewing tobacco associated oral squamous cell carcinomas and adjacent normal tissues. Oral Oncol 40(2):145–153. https://doi.org/10.1016/s1368-8375(03)00143-x
86. Breitling LP, Yang R, Korn B, Burwinkel B, Brenner H (2011) Tobacco-smoking-related differential DNA methylation: 27K discovery and replication. Am J Hum Genet 88(4): 450–457. https://doi.org/10.1016/j.ajhg.2011.03.003
87. Breitling LP, Salzmann K, Rothenbacher D, Burwinkel B, Brenner H (2012) Smoking, F2RL3 methylation, and prognosis in stable coronary heart disease. Eur Heart J 33(22):2841–2848. https://doi.org/10.1093/eurheartj/ehs091
88. Monick MM, Beach SR, Plume J, Sears R, Gerrard M, Brody GH, Philibert RA (2012) Coordinated changes in AHRR methylation in lymphoblasts and pulmonary macrophages from smokers. Am J Med Genet B Neuropsychiatr Genet 159B(2):141–151. https://doi.org/10.1002/ajmg.b.32021
89. Fasanelli F, Baglietto L, Ponzi E, Guida F, Campanella G, Johansson M, Grankvist K, Johansson M, Assumma MB, Naccarati A, Chadeau-Hyam M, Ala U, Faltus C, Kaaks R, Risch A, De Stavola B, Hodge A, Giles GG, Southey MC, Relton CL, Haycock PC, Lund E, Polidoro S, Sandanger TM, Severi G, Vineis P (2015) Hypomethylation of smoking-related genes is associated with future lung cancer in four prospective cohorts. Nat Commun 6:10192. https://doi.org/10.1038/ncomms10192
90. Reynolds LM, Wan M, Ding J, Taylor JR, Lohman K, Su D, Bennett BD, Porter DK, Gimple R, Pittman GS, Wang X, Howard TD, Siscovick D, Psaty BM, Shea S, Burke GL, Jacobs DR Jr, Rich SS, Hixson JE, Stein JH, Stunnenberg H, Barr RG, Kaufman JD, Post WS, Hoeschele I, Herrington DM, Bell DA, Liu Y (2015) DNA methylation of the aryl hydrocarbon receptor repressor associations with cigarette smoking and subclinical atherosclerosis. Circ Cardiovasc Genet 8(5):707–716. https://doi.org/10.1161/CIRCGENETICS.115.001097
91. Shenker NS, Polidoro S, van Veldhoven K, Sacerdote C, Ricceri F, Birrell MA, Belvisi MG, Brown R, Vineis P, Flanagan JM (2013) Epigenome-wide association study in the European prospective investigation into cancer and nutrition (EPIC-Turin) identifies novel genetic loci associated with smoking. Hum Mol Genet 22(5):843–851. https://doi.org/10.1093/hmg/dds488
92. Zeilinger S, Kuhnel B, Klopp N, Baurecht H, Kleinschmidt A, Gieger C, Weidinger S, Lattka E, Adamski J, Peters A, Strauch K, Waldenberger M, Illig T (2013) Tobacco smoking leads to extensive genome-wide changes in DNA methylation. PLoS One 8(5):e63812. https://doi.org/10.1371/journal.pone.0063812
93. Houseman EA, Accomando WP, Koestler DC, Christensen BC, Marsit CJ, Nelson HH, Wiencke JK, Kelsey KT (2012) DNA methylation arrays as surrogate measures of cell mixture distribution. BMC Bioinformatics 13:86. https://doi.org/10.1186/1471-2105-13-86
94. Reinius LE, Acevedo N, Joerink M, Pershagen G, Dahlen SE, Greco D, Soderhall C, Scheynius A, Kere J (2012) Differential DNA methylation in purified human blood cells: implications for cell lineage and studies on disease susceptibility. PLoS One 7(7):e41361. https://doi.org/10.1371/journal.pone.0041361
95. Joehanes R, Just AC, Marioni RE, Pilling LC, Reynolds LM, Mandaviya PR, Guan W, Xu T, Elks CE, Aslibekyan S, Moreno-Macias H, Smith JA, Brody JA, Dhingra R, Yousefi P, Pankow JS, Kunze S, Shah SH, McRae AF, Lohman K, Sha J, Absher DM, Ferrucci L, Zhao W, Demerath EW, Bressler J, Grove ML, Huan T, Liu C, Mendelson MM, Yao C, Kiel DP, Peters A, Wang-Sattler R, Visscher PM, Wray NR, Starr JM, Ding J, Rodriguez CJ, Wareham NJ, Irvin MR, Zhi D, Barrdahl M, Vineis P, Ambatipudi S, Uitterlinden AG, Hofman A, Schwartz J, Colicino E, Hou L, Vokonas PS, Hernandez DG, Singleton AB, Bandinelli S, Turner ST, Ware EB, Smith AK, Klengel T, Binder EB, Psaty BM, Taylor KD, Gharib SA, Swenson BR, Liang L, DeMeo DL, O'Connor GT, Herceg Z, Ressler KJ, Conneely KN, Sotoodehnia N, Kardia SL, Melzer D, Baccarelli AA, van Meurs JB, Romieu I, Arnett DK, Ong KK, Liu Y, Waldenberger M, Deary IJ, Fornage M, Levy D,

London SJ (2016) Epigenetic signatures of cigarette smoking. Circ Cardiovasc Genet 9(5): 436–447. https://doi.org/10.1161/CIRCGENETICS.116.001506
96. Bauer M, Fink B, Thurmann L, Eszlinger M, Herberth G, Lehmann I (2015) Tobacco smoking differently influences cell types of the innate and adaptive immune system-indications from CpG site methylation. Clin Epigenetics 7:83. https://doi.org/10.1186/s13148-016-0249-7
97. Su D, Wang X, Campbell MR, Porter DK, Pittman GS, Bennett BD, Wan M, Englert NA, Crowl CL, Gimple RN, Adamski KN, Huang Z, Murphy SK, Bell DA (2016) Distinct epigenetic effects of tobacco smoking in whole blood and among leukocyte subtypes. PLoS One 11(12):e0166486. https://doi.org/10.1371/journal.pone.0166486
98. Flom JD, Ferris JS, Liao Y, Tehranifar P, Richards CB, Cho YH, Gonzalez K, Santella RM, Terry MB (2011) Prenatal smoke exposure and genomic DNA methylation in a multiethnic birth cohort. Cancer Epidemiol Biomark Prev 20(12):2518–2523. https://doi.org/10.1158/1055-9965.EPI-11-0553
99. Breton CV, Siegmund KD, Joubert BR, Wang X, Qui W, Carey V, Nystad W, Haberg SE, Ober C, Nicolae D, Barnes KC, Martinez F, Liu A, Lemanske R, Strunk R, Weiss S, London S, Gilliland F, Raby B, Bc A (2014) Prenatal tobacco smoke exposure is associated with childhood DNA CpG methylation. PLoS One 9(6):e99716. https://doi.org/10.1371/journal.pone.0099716
100. Lee KW, Richmond R, Hu P, French L, Shin J, Bourdon C, Reischl E, Waldenberger M, Zeilinger S, Gaunt T, McArdle W, Ring S, Woodward G, Bouchard L, Gaudet D, Smith GD, Relton C, Paus T, Pausova Z (2015) Prenatal exposure to maternal cigarette smoking and DNA methylation: epigenome-wide association in a discovery sample of adolescents and replication in an independent cohort at birth through 17 years of age. Environ Health Perspect 123(2): 193–199. https://doi.org/10.1289/ehp.1408614
101. Richmond RC, Simpkin AJ, Woodward G, Gaunt TR, Lyttleton O, McArdle WL, Ring SM, Smith AD, Timpson NJ, Tilling K, Davey Smith G, Relton CL (2015) Prenatal exposure to maternal smoking and offspring DNA methylation across the lifecourse: findings from the Avon longitudinal study of parents and children (ALSPAC). Hum Mol Genet 24(8): 2201–2217. https://doi.org/10.1093/hmg/ddu739
102. Sikdar S, Joehanes R, Joubert BR, Xu CJ, Vives-Usano M, Rezwan FI, Felix JF, Ward JM, Guan W, Richmond RC, Brody JA, Kupers LK, Baiz N, Haberg SE, Smith JA, Reese SE, Aslibekyan S, Hoyo C, Dhingra R, Markunas CA, Xu T, Reynolds LM, Just AC, Mandaviya PR, Ghantous A, Bennett BD, Wang T, Consortium TB, Bakulski KM, Melen E, Zhao S, Jin J, Herceg Z, Meurs JV, Taylor JA, Baccarelli AA, Murphy SK, Liu Y, Munthe-Kaas MC, Deary IJ, Nystad W, Waldenberger M, Annesi-Maesano I, Conneely K, Jaddoe VW, Arnett D, Snieder H, Kardia SL, Relton CL, Ong KK, Ewart S, Moreno-Macias H, Romieu I, Sotoodehnia N, Fornage M, Motsinger-Reif A, Koppelman GH, Bustamante M, Levy D, London SJ (2019) Comparison of smoking-related DNA methylation between newborns from prenatal exposure and adults from personal smoking. Epigenomics 11(13):1487–1500. https://doi.org/10.2217/epi-2019-0066
103. Rauschert S, Melton PE, Burdge G, Craig JM, Godfrey KM, Holbrook JD, Lillycrop K, Mori TA, Beilin LJ, Oddy WH, Pennell C, Huang RC (2019) Maternal smoking during pregnancy induces persistent epigenetic changes into adolescence, independent of postnatal smoke exposure and is associated with Cardiometabolic risk. Front Genet 10:770. https://doi.org/10.3389/fgene.2019.00770
104. Rauschert S, Melton PE, Heiskala A, Karhunen V, Burdge G, Craig JM, Godfrey KM, Lillycrop K, Mori TA, Beilin LJ, Oddy WH, Pennell C, Jarvelin MR, Sebert S, Huang RC (2020) Machine learning-based DNA methylation score for Fetal exposure to maternal smoking: development and validation in samples collected from adolescents and adults. Environ Health Perspect 128(9):97003. https://doi.org/10.1289/EHP6076
105. Herbst AL, Ulfelder H, Poskanzer DC (1971) Adenocarcinoma of the vagina. Association of maternal stilbestrol therapy with tumor appearance in young women. N Engl J Med:284 (15): 878-881. https://doi.org/10.1056/NEJM197104222841604

106. Bromer JG, Wu J, Zhou Y, Taylor HS (2009) Hypermethylation of homeobox A10 by in utero diethylstilbestrol exposure: an epigenetic mechanism for altered developmental programming. Endocrinology 150(7):3376–3382. https://doi.org/10.1210/en.2009-0071
107. Newbold RR (2004) Lessons learned from perinatal exposure to diethylstilbestrol. Toxicol Appl Pharmacol 199(2):142–150. https://doi.org/10.1016/j.taap.2003.11.033
108. Ruden DM, Xiao L, Garfinkel MD, Lu X (2005) Hsp90 and environmental impacts on epigenetic states: a model for the trans-generational effects of diethylstilbestrol on uterine development and cancer. Hum Mol Genet 14 Spec No 1:R149–155. https://doi.org/10.1093/hmg/ddi103
109. Titus L, Hatch EE, Drake KM, Parker SE, Hyer M, Palmer JR, Strohsnitter WC, Adam E, Herbst AL, Huo D, Hoover RN, Troisi R (2019) Reproductive and hormone-related outcomes in women whose mothers were exposed in utero to diethylstilbestrol (DES): a report from the US National Cancer Institute DES third generation study. Reprod Toxicol 84:32–38. https://doi.org/10.1016/j.reprotox.2018.12.008
110. Kioumourtzoglou MA, Coull BA, O'Reilly EJ, Ascherio A, Weisskopf MG (2018) Association of Exposure to Diethylstilbestrol during pregnancy with multigenerational neurodevelopmental deficits. JAMA Pediatr 172(7):670–677. https://doi.org/10.1001/jamapediatrics.2018.0727
111. vom Saal FS, Hughes C (2005) An extensive new literature concerning low-dose effects of bisphenol a shows the need for a new risk assessment. Environ Health Perspect 113(8):926–933. https://doi.org/10.1289/ehp.7713
112. Schonfelder G, Wittfoht W, Hopp H, Talsness CE, Paul M, Chahoud I (2002) Parent bisphenol a accumulation in the human maternal-fetal-placental unit. Environ Health Perspect 110(11):A703–A707. https://doi.org/10.1289/ehp.110-1241091
113. Dolinoy DC, Huang D, Jirtle RL (2007) Maternal nutrient supplementation counteracts bisphenol A-induced DNA hypomethylation in early development. Proc Natl Acad Sci U S A 104(32):13056–13061. https://doi.org/10.1073/pnas.0703739104
114. Avissar-Whiting M, Veiga KR, Uhl KM, Maccani MA, Gagne LA, Moen EL, Marsit CJ (2010) Bisphenol a exposure leads to specific microRNA alterations in placental cells. Reprod Toxicol 29(4):401–406. https://doi.org/10.1016/j.reprotox.2010.04.004
115. Lehmler HJ, Liu B, Gadogbe M, Bao W (2018) Exposure to bisphenol a, bisphenol F, and bisphenol S in U.S. adults and children: the National Health and nutrition examination survey 2013–2014. ACS Omega 3(6):6523–6532. https://doi.org/10.1021/acsomega.8b00824
116. Nair VA, Valo S, Peltomaki P, Bajbouj K, Abdel-Rahman WM (2020) Oncogenic potential of bisphenol a and common environmental contaminants in human mammary epithelial cells. Int J Mol Sci 21(10). https://doi.org/10.3390/ijms21103735
117. Fatma Karaman E, Caglayan M, Sancar-Bas S, Ozal-Coskun C, Arda-Pirincci P, Ozden S (2019) Global and region-specific post-transcriptional and post-translational modifications of bisphenol a in human prostate cancer cells. Environ Pollut 255(Pt 2):113318. https://doi.org/10.1016/j.envpol.2019.113318
118. McCabe CF, Padmanabhan V, Dolinoy DC, Domino SE, Jones TR, Bakulski KM, Goodrich JM (2020) Maternal environmental exposure to bisphenols and epigenome-wide DNA methylation in infant cord blood. Environ Epigenet 6(1):dvaa021. https://doi.org/10.1093/eep/dvaa021
119. Engdahl E, Svensson K, Lin PD, Alavian-Ghavanini A, Lindh C, Ruegg J, Bornehag CG (2021) DNA methylation at GRIN2B partially mediates the association between prenatal bisphenol F exposure and cognitive functions in 7-year-old children in the SELMA study. Environ Int 156:106617. https://doi.org/10.1016/j.envint.2021.106617
120. Alavian-Ghavanini A, Lin PI, Lind PM, Risen Rimfors S, Halin Lejonklou M, Dunder L, Tang M, Lindh C, Bornehag CG, Ruegg J (2018) Prenatal bisphenol a exposure is linked to epigenetic changes in glutamate receptor subunit gene Grin2b in female rats and humans. Sci Rep 8(1):11315. https://doi.org/10.1038/s41598-018-29732-9

121. Wang Z, DeWitt JC, Higgins CP, Cousins IT (2017) A never-ending story of per- and Polyfluoroalkyl substances (PFASs)? Environ Sci Technol 51(5):2508–2518. https://doi.org/10.1021/acs.est.6b04806
122. Blum A, Balan SA, Scheringer M, Trier X, Goldenman G, Cousins IT, Diamond M, Fletcher T, Higgins C, Lindeman AE, Peaslee G, de Voogt P, Wang Z, Weber R (2015) The Madrid statement on poly- and Perfluoroalkyl substances (PFASs). Environ Health Perspect 123(5):A107–A111. https://doi.org/10.1289/ehp.1509934
123. Fenton SE, Ducatman A, Boobis A, DeWitt JC, Lau C, Ng C, Smith JS, Roberts SM (2021) Per- and Polyfluoroalkyl substance toxicity and human health review: current state of knowledge and strategies for informing future research. Environ Toxicol Chem 40(3):606–630. https://doi.org/10.1002/etc.4890
124. Program NT (2016) NTP monograph on immunotoxicity associated with exposure to perfluorooctanoic acid (PFOA) or perfluorooctane sulfonate (PFOS)
125. Liew Z, Goudarzi H, Oulhote Y (2018) Developmental exposures to Perfluoroalkyl substances (PFASs): an update of associated health outcomes. Curr Environ Health Rep 5(1):1–19. https://doi.org/10.1007/s40572-018-0173-4
126. Watkins DJ, Wellenius GA, Butler RA, Bartell SM, Fletcher T, Kelsey KT (2014) Associations between serum perfluoroalkyl acids and LINE-1 DNA methylation. Environ Int 63:71–76. https://doi.org/10.1016/j.envint.2013.10.018
127. Leter G, Consales C, Eleuteri P, Uccelli R, Specht IO, Toft G, Moccia T, Budillon A, Jonsson BA, Lindh CH, Giwercman A, Pedersen HS, Ludwicki JK, Zviezdai V, Heederik D, Bonde JP, Spano M (2014) Exposure to perfluoroalkyl substances and sperm DNA global methylation in Arctic and European populations. Environ Mol Mutagen 55(7):591–600. https://doi.org/10.1002/em.21874
128. van den Dungen MW, Murk AJ, Kampman E, Steegenga WT, Kok DE (2017) Association between DNA methylation profiles in leukocytes and serum levels of persistent organic pollutants in Dutch men. Environ Epigenet 3(1):dvx001. https://doi.org/10.1093/eep/dvx001
129. Guerrero-Preston R, Goldman LR, Brebi-Mieville P, Ili-Gangas C, Lebron C, Witter FR, Apelberg BJ, Hernandez-Roystacher M, Jaffe A, Halden RU, Sidransky D (2010) Global DNA hypomethylation is associated with in utero exposure to cotinine and perfluorinated alkyl compounds. Epigenetics 5(6):539–546. https://doi.org/10.4161/epi.5.6.12378
130. Liu CY, Chen PC, Lien PC, Liao YP (2018) Prenatal Perfluorooctyl sulfonate exposure and Alu DNA Hypomethylation in cord blood. Int J Environ Res Public Health 15(6). https://doi.org/10.3390/ijerph15061066
131. Kobayashi S, Azumi K, Goudarzi H, Araki A, Miyashita C, Kobayashi S, Itoh S, Sasaki S, Ishizuka M, Nakazawa H, Ikeno T, Kishi R (2017) Effects of prenatal perfluoroalkyl acid exposure on cord blood IGF2/H19 methylation and ponderal index: the Hokkaido study. J Expo Sci Environ Epidemiol 27(3):251–259. https://doi.org/10.1038/jes.2016.50
132. Kingsley SL, Kelsey KT, Butler R, Chen A, Eliot MN, Romano ME, Houseman A, Koestler DC, Lanphear BP, Yolton K, Braun JM (2017) Maternal serum PFOA concentration and DNA methylation in cord blood: a pilot study. Environ Res 158:174–178. https://doi.org/10.1016/j.envres.2017.06.013
133. Leung YK, Ouyang B, Niu L, Xie C, Ying J, Medvedovic M, Chen A, Weihe P, Valvi D, Grandjean P, Ho SM (2018) Identification of sex-specific DNA methylation changes driven by specific chemicals in cord blood in a Faroese birth cohort. Epigenetics 13(3):290–300. https://doi.org/10.1080/15592294.2018.1445901
134. Miura R, Araki A, Miyashita C, Kobayashi S, Kobayashi S, Wang SL, Chen CH, Miyake K, Ishizuka M, Iwasaki Y, Ito YM, Kubota T, Kishi R (2018) An epigenome-wide study of cord blood DNA methylations in relation to prenatal perfluoroalkyl substance exposure: the Hokkaido study. Environ Int 115:21–28. https://doi.org/10.1016/j.envint.2018.03.004
135. Ouidir M, Mendola P, Buck Louis GM, Kannan K, Zhang C, Tekola-Ayele F (2020) Concentrations of persistent organic pollutants in maternal plasma and epigenome-wide

placental DNA methylation. Clin Epigenetics 12(1):103. https://doi.org/10.1186/s13148-020-00894-6
136. Sonkar R, Kay MK, Choudhury M (2019) PFOS modulates interactive epigenetic regulation in first-trimester human trophoblast cell line HTR-8/SVneo. Chem Res Toxicol 32(10): 2016–2027. https://doi.org/10.1021/acs.chemrestox.9b00198
137. Szilagyi JT, Freedman AN, Kepper SL, Keshava AM, Bangma JT, Fry RC (2020) Per- and Polyfluoroalkyl substances differentially inhibit placental trophoblast migration and invasion in vitro. Toxicol Sci 175(2):210–219. https://doi.org/10.1093/toxsci/kfaa043
138. van den Dungen MW, Murk AJ, Kok DE, Steegenga WT (2017) Persistent organic pollutants alter DNA methylation during human adipocyte differentiation. Toxicol In Vitro 40:79–87. https://doi.org/10.1016/j.tiv.2016.12.011
139. Little JB (2000) Radiation carcinogenesis. Carcinogenesis 21(3):397–404. https://doi.org/10.1093/carcin/21.3.397
140. McDonald JW, Taylor JA, Watson MA, Saccomanno G, Devereux TR (1995) p53 and K-ras in radon-associated lung adenocarcinoma. Cancer Epidemiol Biomark Prev 4(7):791–793
141. Prise KM, Pinto M, Newman HC, Michael BD (2001) A review of studies of ionizing radiation-induced double-strand break clustering. Radiat Res 156(5 Pt 2):572–576. https://doi.org/10.1667/0033-7587(2001)156[0572:arosoi]2.0.co;2
142. Swafford DS, Middleton SK, Palmisano WA, Nikula KJ, Tesfaigzi J, Baylin SB, Herman JG, Belinsky SA (1997) Frequent aberrant methylation of p16INK4a in primary rat lung tumors. Mol Cell Biol 17(3):1366–1374. https://doi.org/10.1128/MCB.17.3.1366
143. Belinsky SA, Klinge DM, Liechty KC, March TH, Kang T, Gilliland FD, Sotnic N, Adamova G, Rusinova G, Telnov V (2004) Plutonium targets the p16 gene for inactivation by promoter hypermethylation in human lung adenocarcinoma. Carcinogenesis 25(6): 1063–1067. https://doi.org/10.1093/carcin/bgh096
144. Romanenko A, Morell-Quadreny L, Lopez-Guerrero JA, Pellin A, Nepomnyaschy V, Vozianov A, Llombart-Bosch A (2002) P16INK4A and p15INK4B gene alteration associated with oxidative stress in renal cell carcinomas after the Chernobyl accident (pilot study). Diagn Mol Pathol 11(3):163–169. https://doi.org/10.1097/00019606-200209000-00007
145. Cleary HJ, Boulton E, Plumb M (1999) Allelic loss and promoter hypermethylation of the p15INK4b gene features in mouse radiation-induced lymphoid – but not myeloid – leukaemias. Leukemia 13(12):2049–2052. https://doi.org/10.1038/sj.leu.2401616
146. Malumbres M, Perez de Castro I, Santos J, Fernandez Piqueras J, Pellicer A (1999) Hypermethylation of the cell cycle inhibitor p15INK4b 3′-untranslated region interferes with its transcriptional regulation in primary lymphomas. Oncogene 18(2):385–396. https://doi.org/10.1038/sj.onc.1202299
147. Malumbres M, Perez de Castro I, Santos J, Melendez B, Mangues R, Serrano M, Pellicer A, Fernandez-Piqueras J (1997) Inactivation of the cyclin-dependent kinase inhibitor p15INK4b by deletion and de novo methylation with independence of p16INK4a alterations in murine primary T-cell lymphomas. Oncogene 14(11):1361–1370. https://doi.org/10.1038/sj.onc.1200969
148. Song N, Hsu CW, Pan H, Zheng Y, Hou L, Sim JA, Li Z, Mulder H, Easton J, Walker E, Neale G, Wilson CL, Ness KK, Krull KR, Srivastava DK, Yasui Y, Zhang J, Hudson MM, Robison LL, Huang IC, Wang Z (2021) Persistent variations of blood DNA methylation associated with treatment exposures and risk for cardiometabolic outcomes in long-term survivors of childhood cancer in the St. Jude Lifetime Cohort Genome Med 13(1):53. https://doi.org/10.1186/s13073-021-00875-1
149. Nair-Shalliker V, Dhillon V, Clements M, Armstrong BK, Fenech M (2014) The association between personal sun exposure, serum vitamin D and global methylation in human lymphocytes in a population of healthy adults in South Australia. Mutat Res 765:6–10. https://doi.org/10.1016/j.mrfmmm.2014.04.001
150. Page CM, Djordjilovic V, Nost TH, Ghiasvand R, Sandanger TM, Frigessi A, Thoresen M, Veierod MB (2020) Lifetime ultraviolet radiation exposure and DNA methylation in blood

leukocytes: the Norwegian women and cancer study. Sci Rep 10(1):4521. https://doi.org/10.1038/s41598-020-61430-3
151. Rossman TG (2003) Mechanism of arsenic carcinogenesis: an integrated approach. Mutat Res 533(1–2):37–65. https://doi.org/10.1016/j.mrfmmm.2003.07.009
152. Mass MJ, Wang L (1997) Arsenic alters cytosine methylation patterns of the promoter of the tumor suppressor gene p53 in human lung cells: a model for a mechanism of carcinogenesis. Mutat Res 386(3):263–277. https://doi.org/10.1016/s1383-5742(97)00008-2
153. Zhao CQ, Young MR, Diwan BA, Coogan TP, Waalkes MP (1997) Association of arsenic-induced malignant transformation with DNA hypomethylation and aberrant gene expression. Proc Natl Acad Sci U S A 94(20):10907–10912. https://doi.org/10.1073/pnas.94.20.10907
154. Zhong CX, Mass MJ (2001) Both hypomethylation and hypermethylation of DNA associated with arsenite exposure in cultures of human cells identified by methylation-sensitive arbitrarily-primed PCR. Toxicol Lett 122(3):223–234. https://doi.org/10.1016/s0378-4274(01)00365-4
155. Marsit CJ, Karagas MR, Danaee H, Liu M, Andrew A, Schned A, Nelson HH, Kelsey KT (2006) Carcinogen exposure and gene promoter hypermethylation in bladder cancer. Carcinogenesis 27(1):112–116. https://doi.org/10.1093/carcin/bgi172
156. Okoji RS, Yu RC, Maronpot RR, Froines JR (2002) Sodium arsenite administration via drinking water increases genome-wide and ha-ras DNA hypomethylation in methyl-deficient C57BL/6J mice. Carcinogenesis 23(5):777–785. https://doi.org/10.1093/carcin/23.5.777
157. McDorman EW, Collins BW, Allen JW (2002) Dietary folate deficiency enhances induction of micronuclei by arsenic in mice. Environ Mol Mutagen 40(1):71–77. https://doi.org/10.1002/em.10085
158. Chanda S, Dasgupta UB, Guhamazumder D, Gupta M, Chaudhuri U, Lahiri S, Das S, Ghosh N, Chatterjee D (2006) DNA hypermethylation of promoter of gene p53 and p16 in arsenic-exposed people with and without malignancy. Toxicol Sci 89(2):431–437. https://doi.org/10.1093/toxsci/kfj030
159. Majumdar S, Chanda S, Ganguli B, Mazumder DN, Lahiri S, Dasgupta UB (2010) Arsenic exposure induces genomic hypermethylation. Environ Toxicol 25(3):315–318. https://doi.org/10.1002/tox.20497
160. Pilsner JR, Liu X, Ahsan H, Ilievski V, Slavkovich V, Levy D, Factor-Litvak P, Graziano JH, Gamble MV (2007) Genomic methylation of peripheral blood leukocyte DNA: influences of arsenic and folate in Bangladeshi adults. Am J Clin Nutr 86(4):1179–1186. https://doi.org/10.1093/ajcn/86.4.1179
161. Gamble MV, Liu X, Slavkovich V, Pilsner JR, Ilievski V, Factor-Litvak P, Levy D, Alam S, Islam M, Parvez F, Ahsan H, Graziano JH (2007) Folic acid supplementation lowers blood arsenic. Am J Clin Nutr 86(4):1202–1209. https://doi.org/10.1093/ajcn/86.4.1202
162. Pilsner JR, Liu X, Ahsan H, Ilievski V, Slavkovich V, Levy D, Factor-Litvak P, Graziano JH, Gamble MV (2009) Folate deficiency, hyperhomocysteinemia, low urinary creatinine, and hypomethylation of leukocyte DNA are risk factors for arsenic-induced skin lesions. Environ Health Perspect 117(2):254–260. https://doi.org/10.1289/ehp.11872
163. Kile ML, Houseman EA, Baccarelli AA, Quamruzzaman Q, Rahman M, Mostofa G, Cardenas A, Wright RO, Christiani DC (2014) Effect of prenatal arsenic exposure on DNA methylation and leukocyte subpopulations in cord blood. Epigenetics 9(5):774–782. https://doi.org/10.4161/epi.28153
164. Koestler DC, Avissar-Whiting M, Houseman EA, Karagas MR, Marsit CJ (2013) Differential DNA methylation in umbilical cord blood of infants exposed to low levels of arsenic in utero. Environ Health Perspect 121(8):971–977. https://doi.org/10.1289/ehp.1205925
165. Bozack AK, Cardenas A, Quamruzzaman Q, Rahman M, Mostofa G, Christiani DC, Kile ML (2018) DNA methylation in cord blood as mediator of the association between prenatal arsenic exposure and gestational age. Epigenetics 13(9):923–940. https://doi.org/10.1080/15592294.2018.1516453

166. DiGiovanni A, Demanelis K, Tong L, Argos M, Shinkle J, Jasmine F, Sabarinathan M, Rakibuz-Zaman M, Sarwar G, Islam MT, Shahriar H, Islam T, Rahman M, Yunus M, Graziano J, Gamble MV, Ahsan H, Pierce BL (2020) Assessing the impact of arsenic metabolism efficiency on DNA methylation using Mendelian randomization. Environ Epidemiol 4(2):e083. https://doi.org/10.1097/EE9.0000000000000083
167. Marsit CJ, Eddy K, Kelsey KT (2006) MicroRNA responses to cellular stress. Cancer Res 66(22):10843–10848. https://doi.org/10.1158/0008-5472.CAN-06-1894
168. Tumolo MR, Panico A, De Donno A, Mincarone P, Leo CG, Guarino R, Bagordo F, Serio F, Idolo A, Grassi T, Sabina S (2020) The expression of microRNAs and exposure to environmental contaminants related to human health: a review. Int J Environ Health Res:1–23. https://doi.org/10.1080/09603123.2020.1757043
169. Ramirez T, Brocher J, Stopper H, Hock R (2008) Sodium arsenite modulates histone acetylation, histone deacetylase activity and HMGN protein dynamics in human cells. Chromosoma 117(2):147–157. https://doi.org/10.1007/s00412-007-0133-5
170. Howe CG, Gamble MV (2016) Influence of arsenic on global levels of histone posttranslational modifications: a review of the literature and challenges in the field. Curr Environ Health Rep 3(3):225–237. https://doi.org/10.1007/s40572-016-0104-1
171. (IARC) IAfRoC (1993) Summaries & Evaluations: cadmium and cadmium compounds. Summaries & Evaluations 58
172. Registry AfTSaD (2019) ATSDR's Substance Priority List. https://www.atsdr.cdc.gov/spl/index.html
173. Benbrahim-Tallaa L, Waterland RA, Dill AL, Webber MM, Waalkes MP (2007) Tumor suppressor gene inactivation during cadmium-induced malignant transformation of human prostate cells correlates with overexpression of de novo DNA methyltransferase. Environ Health Perspect 115(10):1454–1459. https://doi.org/10.1289/ehp.10207
174. Zhou ZH, Lei YX, Wang CX (2012) Analysis of aberrant methylation in DNA repair genes during malignant transformation of human bronchial epithelial cells induced by cadmium. Toxicol Sci 125(2):412–417. https://doi.org/10.1093/toxsci/kfr320
175. Yuan D, Ye S, Pan Y, Bao Y, Chen H, Shao C (2013) Long-term cadmium exposure leads to the enhancement of lymphocyte proliferation via down-regulating p16 by DNA hypermethylation. Mutat Res 757(2):125–131. https://doi.org/10.1016/j.mrgentox.2013.07.007
176. Xiao C, Liu Y, Xie C, Tu W, Xia Y, Costa M, Zhou X (2015) Cadmium induces histone H3 lysine methylation by inhibiting histone demethylase activity. Toxicol Sci 145(1):80–89. https://doi.org/10.1093/toxsci/kfv019
177. Liang ZZ, Zhu RM, Li YL, Jiang HM, Li RB, Tang LY, Wang Q, Ren ZF (2020) Differential epigenetic and transcriptional profile in MCF-7 breast cancer cells exposed to cadmium. Chemosphere 261:128148. https://doi.org/10.1016/j.chemosphere.2020.128148
178. Chandravanshi L, Shiv K, Kumar S (2021) Developmental toxicity of cadmium in infants and children: a review. Environ Anal Health Toxicol 36(1):e2021003–e2021000. https://doi.org/10.5620/eaht.2021003
179. Vilahur N, Vahter M, Broberg K (2015) The epigenetic effects of prenatal cadmium exposure. Curr Environ Health Rep 2(2):195–203. https://doi.org/10.1007/s40572-015-0049-9
180. Kippler M, Engstrom K, Mlakar SJ, Bottai M, Ahmed S, Hossain MB, Raqib R, Vahter M, Broberg K (2013) Sex-specific effects of early life cadmium exposure on DNA methylation and implications for birth weight. Epigenetics 8(5):494–503. https://doi.org/10.4161/epi.24401
181. Everson TM, Punshon T, Jackson BP, Hao K, Lambertini L, Chen J, Karagas MR, Marsit CJ (2018) Cadmium-associated differential methylation throughout the placental genome: epigenome-wide association study of two U.S. birth cohorts. Environ Health Perspect 126(1):017010. https://doi.org/10.1289/EHP2192
182. Cowley M, Skaar DA, Jima DD, Maguire RL, Hudson KM, Park SS, Sorrow P, Hoyo C (2018) Effects of cadmium exposure on DNA methylation at imprinting control regions and

genome-wide in mothers and Newborn children. Environ Health Perspect 126(3):037003. https://doi.org/10.1289/EHP2085
183. Millership SJ, Van de Pette M, Withers DJ (2019) Genomic imprinting and its effects on postnatal growth and adult metabolism. Cell Mol Life Sci 76(20):4009–4021. https://doi.org/10.1007/s00018-019-03197-z
184. Wilkinson LS, Davies W, Isles AR (2007) Genomic imprinting effects on brain development and function. Nat Rev Neurosci 8(11):832–843. https://doi.org/10.1038/nrn2235
185. Everson TM, Marable C, Deyssenroth MA, Punshon T, Jackson BP, Lambertini L, Karagas MR, Chen J, Marsit CJ (2019) Placental expression of imprinted genes, overall and in sex-specific patterns, associated with placental cadmium concentrations and birth size. Environ Health Perspect 127(5):57005. https://doi.org/10.1289/EHP4264
186. Vidal AC, Semenova V, Darrah T, Vengosh A, Huang Z, King K, Nye MD, Fry R, Skaar D, Maguire R, Murtha A, Schildkraut J, Murphy S, Hoyo C (2015) Maternal cadmium, iron and zinc levels, DNA methylation and birth weight. BMC Pharmacol Toxicol 16:20. https://doi.org/10.1186/s40360-015-0020-2
187. Xu P, Wu Z, Yang W, Wang L (2017) Dysregulation of DNA methylation and expression of imprinted genes in mouse placentas of fetal growth restriction induced by maternal cadmium exposure. Toxicology 390:109–116. https://doi.org/10.1016/j.tox.2017.08.003
188. Zhu J, Huang Z, Yang F, Zhu M, Cao J, Chen J, Lin Y, Guo S, Li J, Liu Z (2021) Cadmium disturbs epigenetic modification and induces DNA damage in mouse preimplantation embryos. Ecotoxicol Environ Saf 219:112306. https://doi.org/10.1016/j.ecoenv.2021.112306
189. Brooks SA, Fry RC (2017) Cadmium inhibits placental trophoblast cell migration via miRNA regulation of the transforming growth factor beta (TGF-beta) pathway. Food Chem Toxicol 109(Pt 1):721–726. https://doi.org/10.1016/j.fct.2017.07.059
190. Brooks SA, Martin E, Smeester L, Grace MR, Boggess K, Fry RC (2016) miRNAs as common regulators of the transforming growth factor (TGF)-beta pathway in the preeclamptic placenta and cadmium-treated trophoblasts: links between the environment, the epigenome and preeclampsia. Food Chem Toxicol 98(Pt A):50–57. https://doi.org/10.1016/j.fct.2016.06.023
191. Yang K, Julan L, Rubio F, Sharma A, Guan H (2006) Cadmium reduces 11 beta-hydroxysteroid dehydrogenase type 2 activity and expression in human placental trophoblast cells. Am J Physiol Endocrinol Metab 290(1):E135–E142. https://doi.org/10.1152/ajpendo.00356.2005
192. Bollati V, Marinelli B, Apostoli P, Bonzini M, Nordio F, Hoxha M, Pegoraro V, Motta V, Tarantini L, Cantone L, Schwartz J, Bertazzi PA, Baccarelli A (2010) Exposure to metal-rich particulate matter modifies the expression of candidate microRNAs in peripheral blood leukocytes. Environ Health Perspect 118(6):763–768. https://doi.org/10.1289/ehp.0901300
193. Wright RO, Schwartz J, Wright RJ, Bollati V, Tarantini L, Park SK, Hu H, Sparrow D, Vokonas P, Baccarelli A (2010) Biomarkers of lead exposure and DNA methylation within retrotransposons. Environ Health Perspect 118(6):790–795. https://doi.org/10.1289/ehp.0901429
194. Pilsner JR, Hu H, Ettinger A, Sanchez BN, Wright RO, Cantonwine D, Lazarus A, Lamadrid-Figueroa H, Mercado-Garcia A, Tellez-Rojo MM, Hernandez-Avila M (2009) Influence of prenatal lead exposure on genomic methylation of cord blood DNA. Environ Health Perspect 117(9):1466–1471. https://doi.org/10.1289/ehp.0800497
195. Wu S, Hivert MF, Cardenas A, Zhong J, Rifas-Shiman SL, Agha G, Colicino E, Just AC, Amarasiriwardena C, Lin X, Litonjua AA, DeMeo DL, Gillman MW, Wright RO, Oken E, Baccarelli AA (2017) Exposure to Low levels of Lead in utero and umbilical cord blood DNA methylation in project viva: an epigenome-wide association study. Environ Health Perspect 125(8):087019. https://doi.org/10.1289/EHP1246
196. Cardenas A, Rifas-Shiman SL, Agha G, Hivert MF, Litonjua AA, DeMeo DL, Lin X, Amarasiriwardena CJ, Oken E, Gillman MW, Baccarelli AA (2017) Persistent DNA methylation changes associated with prenatal mercury exposure and cognitive performance during childhood. Sci Rep 7(1):288. https://doi.org/10.1038/s41598-017-00384-5

197. Cardenas A, Koestler DC, Houseman EA, Jackson BP, Kile ML, Karagas MR, Marsit CJ (2015) Differential DNA methylation in umbilical cord blood of infants exposed to mercury and arsenic in utero. Epigenetics 10(6):508–515. https://doi.org/10.1080/15592294.2015.1046026
198. Cardenas A, Rifas-Shiman SL, Godderis L, Duca RC, Navas-Acien A, Litonjua AA, DeMeo DL, Brennan KJ, Amarasiriwardena CJ, Hivert MF, Gillman MW, Oken E, Baccarelli AA (2017) Prenatal exposure to mercury: associations with global DNA methylation and Hydroxymethylation in cord blood and in childhood. Environ Health Perspect 125(8):087022. https://doi.org/10.1289/EHP1467
199. Lozano M, Yousefi P, Broberg K, Soler-Blasco R, Miyashita C, Pesce G, Kim WJ, Rahman M, Bakulski KM, Haug LS, Ikeda-Araki A, Huel G, Park J, Relton C, Vrijheid M, Rifas-Shiman S, Oken E, Dou JF, Kishi R, Gutzkow KB, Annesi-Maesano I, Won S, Hivert MF, Fallin MD, Vafeiadi M, Ballester F, Bustamante M, Llop S (2021) DNA methylation changes associated with prenatal mercury exposure: a meta-analysis of prospective cohort studies from PACE consortium. Environ Res 204(Pt B):112093. https://doi.org/10.1016/j.envres.2021.112093
200. Gibb HJ, Lees PS, Pinsky PF, Rooney BC (2000) Lung cancer among workers in chromium chemical production. Am J Ind Med 38(2):115–126. https://doi.org/10.1002/1097-0274(200008)38:2<115::aid-ajim1>3.0.co;2-y
201. Zhitkovich A, Shrager S, Messer J (2000) Reductive metabolism of Cr(VI) by cysteine leads to the formation of binary and ternary Cr – DNA adducts in the absence of oxidative DNA damage. Chem Res Toxicol 13(11):1114–1124. https://doi.org/10.1021/tx0001169
202. Shi H, Hudson LG, Liu KJ (2004) Oxidative stress and apoptosis in metal ion-induced carcinogenesis. Free Radic Biol Med 37(5):582–593. https://doi.org/10.1016/j.freeradbiomed.2004.03.012
203. Kondo K, Takahashi Y, Hirose Y, Nagao T, Tsuyuguchi M, Hashimoto M, Ochiai A, Monden Y, Tangoku A (2006) The reduced expression and aberrant methylation of p16 (INK4a) in chromate workers with lung cancer. Lung Cancer 53(3):295–302. https://doi.org/10.1016/j.lungcan.2006.05.022
204. Schnekenburger M, Talaska G, Puga A (2007) Chromium cross-links histone deacetylase 1-DNA methyltransferase 1 complexes to chromatin, inhibiting histone-remodeling marks critical for transcriptional activation. Mol Cell Biol 27(20):7089–7101. https://doi.org/10.1128/MCB.00838-07
205. Vineis P, Husgafvel-Pursiainen K (2005) Air pollution and cancer: biomarker studies in human populations. Carcinogenesis 26(11):1846–1855. https://doi.org/10.1093/carcin/bgi216
206. Samet JM, Dominici F, Curriero FC, Coursac I, Zeger SL (2000) Fine particulate air pollution and mortality in 20 U.S. cities, 1987–1994. N Engl J Med 343(24):1742–1749. https://doi.org/10.1056/NEJM200012143432401
207. Tarantini L, Bonzini M, Apostoli P, Pegoraro V, Bollati V, Marinelli B, Cantone L, Rizzo G, Hou L, Schwartz J, Bertazzi PA, Baccarelli A (2009) Effects of particulate matter on genomic DNA methylation content and iNOS promoter methylation. Environ Health Perspect 117(2):217–222. https://doi.org/10.1289/ehp.11898
208. Baccarelli A, Wright RO, Bollati V, Tarantini L, Litonjua AA, Suh HH, Zanobetti A, Sparrow D, Vokonas PS, Schwartz J (2009) Rapid DNA methylation changes after exposure to traffic particles. Am J Respir Crit Care Med 179(7):572–578. https://doi.org/10.1164/rccm.200807-1097OC
209. Wang C, O'Brien KM, Xu Z, Sandler DP, Taylor JA, Weinberg CR (2020) Long-term ambient fine particulate matter and DNA methylation in inflammation pathways: results from the sister study. Epigenetics 15(5):524–535. https://doi.org/10.1080/15592294.2019.1699894
210. Chi GC, Liu Y, MacDonald JW, Barr RG, Donohue KM, Hensley MD, Hou L, McCall CE, Reynolds LM, Siscovick DS, Kaufman JD (2016) Long-term outdoor air pollution and DNA methylation in circulating monocytes: results from the multi-ethnic study of atherosclerosis (MESA). Environ Health 15(1):119. https://doi.org/10.1186/s12940-016-0202-4

211. Madrigano J, Baccarelli A, Mittleman MA, Wright RO, Sparrow D, Vokonas PS, Tarantini L, Schwartz J (2011) Prolonged exposure to particulate pollution, genes associated with glutathione pathways, and DNA methylation in a cohort of older men. Environ Health Perspect 119(7):977–982. https://doi.org/10.1289/ehp.1002773
212. Wu Y, Qie R, Cheng M, Zeng Y, Huang S, Guo C, Zhou Q, Li Q, Tian G, Han M, Zhang Y, Wu X, Li Y, Zhao Y, Yang X, Feng Y, Liu D, Qin P, Hu D, Hu F, Xu L, Zhang M (2021) Air pollution and DNA methylation in adults: a systematic review and meta-analysis of observational studies. Environ Pollut 284:117152. https://doi.org/10.1016/j.envpol.2021.117152
213. Liu X, Ye Y, Chen Y, Li X, Feng B, Cao G, Xiao J, Zeng W, Li X, Sun J, Ning D, Yang Y, Yao Z, Guo Y, Wang Q, Zhang Y, Ma W, Du Q, Zhang B, Liu T (2019) Effects of prenatal exposure to air particulate matter on the risk of preterm birth and roles of maternal and cord blood LINE-1 methylation: a birth cohort study in Guangzhou. China Environ Int 133 (Pt A):105177. https://doi.org/10.1016/j.envint.2019.105177
214. Yang P, Gong YJ, Cao WC, Wang RX, Wang YX, Liu C, Chen YJ, Huang LL, Ai SH, Lu WQ, Zeng Q (2018) Prenatal urinary polycyclic aromatic hydrocarbon metabolites, global DNA methylation in cord blood, and birth outcomes: a cohort study in China. Environ Pollut 234:396–405. https://doi.org/10.1016/j.envpol.2017.11.082
215. Breton CV, Yao J, Millstein J, Gao L, Siegmund KD, Mack W, Whitfield-Maxwell L, Lurmann F, Hodis H, Avol E, Gilliland FD (2016) Prenatal air pollution exposures, DNA methyl transferase genotypes, and associations with Newborn LINE1 and Alu methylation and childhood blood pressure and carotid intima-media thickness in the Children's health study. Environ Health Perspect 124(12):1905–1912. https://doi.org/10.1289/EHP181
216. Cai J, Zhao Y, Liu P, Xia B, Zhu Q, Wang X, Song Q, Kan H, Zhang Y (2017) Exposure to particulate air pollution during early pregnancy is associated with placental DNA methylation. Sci Total Environ 607–608:1103–1108. https://doi.org/10.1016/j.scitotenv.2017.07.029
217. Ren C, Park SK, Vokonas PS, Sparrow D, Wilker E, Baccarelli A, Suh HH, Tucker KL, Wright RO, Schwartz J (2010) Air pollution and homocysteine: more evidence that oxidative stress-related genes modify effects of particulate air pollution. Epidemiology 21(2):198–206. https://doi.org/10.1097/EDE.0b013e3181cc8bfc
218. Baccarelli A, Zanobetti A, Martinelli I, Grillo P, Hou L, Lanzani G, Mannucci PM, Bertazzi PA, Schwartz J (2007) Air pollution, smoking, and plasma homocysteine. Environ Health Perspect 115(2):176–181. https://doi.org/10.1289/ehp.9517
219. Eze IC, Jeong A, Schaffner E, Rezwan FI, Ghantous A, Foraster M, Vienneau D, Kronenberg F, Herceg Z, Vineis P, Brink M, Wunderli JM, Schindler C, Cajochen C, Roosli M, Holloway JW, Imboden M, Probst-Hensch N (2020) Genome-wide DNA methylation in peripheral blood and Long-term exposure to source-specific transportation noise and air pollution: the SAPALDIA study. Environ Health Perspect 128(6):67003. https://doi.org/10.1289/EHP6174
220. Gondalia R, Baldassari A, Holliday KM, Justice AE, Mendez-Giraldez R, Stewart JD, Liao D, Yanosky JD, Brennan KJM, Engel SM, Jordahl KM, Kennedy E, Ward-Caviness CK, Wolf K, Waldenberger M, Cyrys J, Peters A, Bhatti P, Horvath S, Assimes TL, Pankow JS, Demerath EW, Guan W, Fornage M, Bressler J, North KE, Conneely KN, Li Y, Hou L, Baccarelli AA, Whitsel EA (2019) Methylome-wide association study provides evidence of particulate matter air pollution-associated DNA methylation. Environ Int 132:104723. https://doi.org/10.1016/j.envint.2019.03.071
221. Panni T, Mehta AJ, Schwartz JD, Baccarelli AA, Just AC, Wolf K, Wahl S, Cyrys J, Kunze S, Strauch K, Waldenberger M, Peters A (2016) Genome-wide analysis of DNA methylation and fine particulate matter air pollution in three study populations: KORA F3, KORA F4, and the normative aging study. Environ Health Perspect 124(7):983–990. https://doi.org/10.1289/ehp.1509966
222. Sayols-Baixeras S, Fernandez-Sanles A, Prats-Uribe A, Subirana I, Plusquin M, Kunzli N, Marrugat J, Basagana X, Elosua R (2019) Association between long-term air pollution

exposure and DNA methylation: the REGICOR study. Environ Res 176:108550. https://doi.org/10.1016/j.envres.2019.108550
223. Li H, Chen R, Cai J, Cui X, Huang N, Kan H (2018) Short-term exposure to fine particulate air pollution and genome-wide DNA methylation: a randomized, double-blind, crossover trial. Environ Int 120:130–136. https://doi.org/10.1016/j.envint.2018.07.041
224. Gruzieva O, Xu CJ, Yousefi P, Relton C, Merid SK, Breton CV, Gao L, Volk HE, Feinberg JI, Ladd-Acosta C, Bakulski K, Auffray C, Lemonnier N, Plusquin M, Ghantous A, Herceg Z, Nawrot TS, Pizzi C, Richiardi L, Rusconi F, Vineis P, Kogevinas M, Felix JF, Duijts L, den Dekker HT, Jaddoe VWV, Ruiz JL, Bustamante M, Anto JM, Sunyer J, Vrijheid M, Gutzkow KB, Grazuleviciene R, Hernandez-Ferrer C, Annesi-Maesano I, Lepeule J, Bousquet J, Bergstrom A, Kull I, Soderhall C, Kere J, Gehring U, Brunekreef B, Just AC, Wright RJ, Peng C, Gold DR, Kloog I, DeMeo DL, Pershagen G, Koppelman GH, London SJ, Baccarelli AA, Melen E (2019) Prenatal particulate air pollution and DNA methylation in Newborns: an epigenome-wide meta-analysis. Environ Health Perspect 127(5):57012. https://doi.org/10.1289/EHP4522
225. Gruzieva O, Xu CJ, Breton CV, Annesi-Maesano I, Anto JM, Auffray C, Ballereau S, Bellander T, Bousquet J, Bustamante M, Charles MA, de Kluizenaar Y, den Dekker HT, Duijts L, Felix JF, Gehring U, Guxens M, Jaddoe VV, Jankipersadsing SA, Merid SK, Kere J, Kumar A, Lemonnier N, Lepeule J, Nystad W, Page CM, Panasevich S, Postma D, Slama R, Sunyer J, Soderhall C, Yao J, London SJ, Pershagen G, Koppelman GH, Melen E (2017) Epigenome-wide meta-analysis of methylation in children related to prenatal NO2 air pollution exposure. Environ Health Perspect 125(1):104–110. https://doi.org/10.1289/EHP36
226. Ferrari L, Carugno M, Bollati V (2019) Particulate matter exposure shapes DNA methylation through the lifespan. Clin Epigenetics 11(1):129. https://doi.org/10.1186/s13148-019-0726-x
227. Isaevska E, Moccia C, Asta F, Cibella F, Gagliardi L, Ronfani L, Rusconi F, Stazi MA, Richiardi L (2021) Exposure to ambient air pollution in the first 1000 days of life and alterations in the DNA methylome and telomere length in children: a systematic review. Environ Res 193:110504. https://doi.org/10.1016/j.envres.2020.110504
228. Rider CF, Carlsten C (2019) Air pollution and DNA methylation: effects of exposure in humans. Clin Epigenetics 11(1):131. https://doi.org/10.1186/s13148-019-0713-2
229. White AJ, Kresovich JK, Keller JP, Xu Z, Kaufman JD, Weinberg CR, Taylor JA, Sandler DP (2019) Air pollution, particulate matter composition and methylation-based biologic age. Environ Int 132:105071. https://doi.org/10.1016/j.envint.2019.105071
230. Robinson BW, Lake RA (2005) Advances in malignant mesothelioma. N Engl J Med 353(15): 1591–1603. https://doi.org/10.1056/NEJMra050152
231. Sugarbaker DJ, Richards WG, Gordon GJ, Dong L, De Rienzo A, Maulik G, Glickman JN, Chirieac LR, Hartman ML, Taillon BE, Du L, Bouffard P, Kingsmore SF, Miller NA, Farmer AD, Jensen RV, Gullans SR, Bueno R (2008) Transcriptome sequencing of malignant pleural mesothelioma tumors. Proc Natl Acad Sci U S A 105(9):3521–3526. https://doi.org/10.1073/pnas.0712399105
232. Christensen BC, Godleski JJ, Marsit CJ, Houseman EA, Lopez-Fagundo CY, Longacker JL, Bueno R, Sugarbaker DJ, Nelson HH, Kelsey KT (2008) Asbestos exposure predicts cell cycle control gene promoter methylation in pleural mesothelioma. Carcinogenesis 29(8): 1555–1559. https://doi.org/10.1093/carcin/bgn059
233. He B, Lee AY, Dadfarmay S, You L, Xu Z, Reguart N, Mazieres J, Mikami I, McCormick F, Jablons DM (2005) Secreted frizzled-related protein 4 is silenced by hypermethylation and induces apoptosis in beta-catenin-deficient human mesothelioma cells. Cancer Res 65(3): 743–748
234. Hirao T, Bueno R, Chen CJ, Gordon GJ, Heilig E, Kelsey KT (2002) Alterations of the p16 (INK4) locus in human malignant mesothelial tumors. Carcinogenesis 23(7):1127–1130. https://doi.org/10.1093/carcin/23.7.1127

235. Lee AY, He B, You L, Dadfarmay S, Xu Z, Mazieres J, Mikami I, McCormick F, Jablons DM (2004) Expression of the secreted frizzled-related protein gene family is downregulated in human mesothelioma. Oncogene 23(39):6672–6676. https://doi.org/10.1038/sj.onc.1207881
236. Ohta Y, Shridhar V, Kalemkerian GP, Bright RK, Watanabe Y, Pass HI (1999) Thrombospondin-1 expression and clinical implications in malignant pleural mesothelioma. Cancer 85(12):2570–2576
237. Tsou JA, Galler JS, Wali A, Ye W, Siegmund KD, Groshen S, Laird PW, Turla S, Koss MN, Pass HI, Laird-Offringa IA (2007) DNA methylation profile of 28 potential marker loci in malignant mesothelioma. Lung Cancer 58(2):220–230. https://doi.org/10.1016/j.lungcan.2007.06.015
238. Christensen BC, Houseman EA, Godleski JJ, Marsit CJ, Longacker JL, Roelofs CR, Karagas MR, Wrensch MR, Yeh RF, Nelson HH, Wiemels JL, Zheng S, Wiencke JK, Bueno R, Sugarbaker DJ, Kelsey KT (2009) Epigenetic profiles distinguish pleural mesothelioma from normal pleura and predict lung asbestos burden and clinical outcome. Cancer Res 69(1):227–234. https://doi.org/10.1158/0008-5472.CAN-08-2586
239. Kettunen E, Hernandez-Vargas H, Cros MP, Durand G, Le Calvez-Kelm F, Stuopelyte K, Jarmalaite S, Salmenkivi K, Anttila S, Wolff H, Herceg Z, Husgafvel-Pursiainen K (2017) Asbestos-associated genome-wide DNA methylation changes in lung cancer. Int J Cancer 141(10):2014–2029. https://doi.org/10.1002/ijc.30897
240. Guarrera S, Viberti C, Cugliari G, Allione A, Casalone E, Betti M, Ferrante D, Aspesi A, Casadio C, Grosso F, Libener R, Piccolini E, Mirabelli D, Dianzani I, Magnani C, Matullo G (2019) Peripheral blood DNA methylation as potential biomarker of malignant pleural mesothelioma in Asbestos-exposed subjects. J Thorac Oncol 14(3):527–539. https://doi.org/10.1016/j.jtho.2018.10.163
241. Cugliari G, Allione A, Russo A, Catalano C, Casalone E, Guarrera S, Grosso F, Ferrante D, Sculco M, La Vecchia M, Pirazzini C, Libener R, Mirabelli D, Magnani C, Dianzani I, Matullo G (2021) New DNA methylation signals for malignant pleural mesothelioma risk assessment. Cancers (Basel) 13(11). https://doi.org/10.3390/cancers13112636
242. Oner D, Ghosh M, Moisse M, Duca RC, Coorens R, Vanoirbeek JAJ, Lambrechts D, Godderis L, Hoet PHM (2018) Global and gene-specific DNA methylation effects of different asbestos fibres on human bronchial epithelial cells. Environ Int 115:301–311. https://doi.org/10.1016/j.envint.2018.03.031
243. Oner D, Ghosh M, Coorens R, Bove H, Moisse M, Lambrechts D, Ameloot M, Godderis L, Hoet PHM (2020) Induction and recovery of CpG site specific methylation changes in human bronchial cells after long-term exposure to carbon nanotubes and asbestos. Environ Int 137:105530. https://doi.org/10.1016/j.envint.2020.105530
244. Wang J, Tian X, Zhang J, Tan L, Ouyang N, Jia B, Chen C, Ge C, Li J (2021) Postchronic single-walled carbon nanotube exposure causes irreversible malignant transformation of human bronchial epithelial cells through DNA methylation changes. ACS Nano 15(4):7094–7104. https://doi.org/10.1021/acsnano.1c00239
245. Feinberg AP, Cui H, Ohlsson R (2002) DNA methylation and genomic imprinting: insights from cancer into epigenetic mechanisms. Semin Cancer Biol 12(5):389–398. https://doi.org/10.1016/s1044-579x(02)00059-7
246. Feinberg AP (2000) DNA methylation, genomic imprinting and cancer. Curr Top Microbiol Immunol 249:87–99. https://doi.org/10.1007/978-3-642-59696-4_6
247. Feinberg AP (2004) The epigenetics of cancer etiology. Semin Cancer Biol 14(6):427–432. https://doi.org/10.1016/j.semcancer.2004.06.005
248. Tilghman SM (1999) The sins of the fathers and mothers: genomic imprinting in mammalian development. Cell 96(2):185–193. https://doi.org/10.1016/s0092-8674(00)80559-0
249. Wolffe AP (2000) Transcriptional control: imprinting insulation. Curr Biol 10(12):R463–R465. https://doi.org/10.1016/s0960-9822(00)00534-0

250. Dean W, Bowden L, Aitchison A, Klose J, Moore T, Meneses JJ, Reik W, Feil R (1998) Altered imprinted gene methylation and expression in completely ES cell-derived mouse fetuses: association with aberrant phenotypes. Development 125(12):2273–2282
251. Young LE, Sinclair KD, Wilmut I (1998) Large offspring syndrome in cattle and sheep. Rev Reprod 3(3):155–163. https://doi.org/10.1530/ror.0.0030155
252. Cox GF, Burger J, Lip V, Mau UA, Sperling K, Wu BL, Horsthemke B (2002) Intracytoplasmic sperm injection may increase the risk of imprinting defects. Am J Hum Genet 71(1):162–164. https://doi.org/10.1086/341096
253. DeBaun MR, Niemitz EL, Feinberg AP (2003) Association of in vitro fertilization with Beckwith-Wiedemann syndrome and epigenetic alterations of LIT1 and H19. Am J Hum Genet 72(1):156–160. https://doi.org/10.1086/346031
254. Jones PA, Liang G (2009) Rethinking how DNA methylation patterns are maintained. Nat Rev Genet 10(11):805–811. https://doi.org/10.1038/nrg2651
255. Rakyan VK, Down TA, Maslau S, Andrew T, Yang TP, Beyan H, Whittaker P, McCann OT, Finer S, Valdes AM, Leslie RD, Deloukas P, Spector TD (2010) Human aging-associated DNA hypermethylation occurs preferentially at bivalent chromatin domains. Genome Res 20(4):434–439. https://doi.org/10.1101/gr.103101.109
256. Masuda T, Shinoara H, Kondo M (1975) Reactions of hydroxyl radicals with nucleic acid bases and the related compounds in gamma-irradiated aqueous solution. J Radiat Res 16(3): 153–161. https://doi.org/10.1269/jrr.16.153
257. Valinluck V, Tsai HH, Rogstad DK, Burdzy A, Bird A, Sowers LC (2004) Oxidative damage to methyl-CpG sequences inhibits the binding of the methyl-CpG binding domain (MBD) of methyl-CpG binding protein 2 (MeCP2). Nucleic Acids Res 32(14):4100–4108. https://doi.org/10.1093/nar/gkh739
258. Valinluck V, Sowers LC (2007) Endogenous cytosine damage products alter the site selectivity of human DNA maintenance methyltransferase DNMT1. Cancer Res 67(3):946–950. https://doi.org/10.1158/0008-5472.CAN-06-3123
259. Henderson JP, Byun J, Takeshita J, Heinecke JW (2003) Phagocytes produce 5-chlorouracil and 5-bromouracil, two mutagenic products of myeloperoxidase, in human inflammatory tissue. J Biol Chem 278(26):23522–23528. https://doi.org/10.1074/jbc.M303928200
260. Sun FL, Elgin SC (1999) Putting boundaries on silence. Cell 99(5):459–462. https://doi.org/10.1016/s0092-8674(00)81534-2
261. Turker MS (2002) Gene silencing in mammalian cells and the spread of DNA methylation. Oncogene 21(35):5388–5393. https://doi.org/10.1038/sj.onc.1205599
262. Turker MS, Bestor TH (1997) Formation of methylation patterns in the mammalian genome. Mutat Res 386(2):119–130. https://doi.org/10.1016/s1383-5742(96)00048-8
263. Muse ME, Titus AJ, Salas LA, Wilkins OM, Mullen C, Gregory KJ, Schneider SS, Crisi GM, Jawale RM, Otis CN, Christensen BC, Arcaro KF (2020) Enrichment of CpG island shore region hypermethylation in epigenetic breast field cancerization. Epigenetics 15(10): 1093–1106. https://doi.org/10.1080/15592294.2020.1747748
264. Schwartz DA, Freedman JH, Linney EA (2004) Environmental genomics: a key to understanding biology, pathophysiology and disease. Hum Mol Genet 13 Spec No 2:R217–224. https://doi.org/10.1093/hmg/ddh228

Part II

Epigenetic Epidemiology of Complex Diseases

Epigenetic Epidemiology of Cancer

13

Zdenko Herceg, Akram Ghantous, and Felicia Fei-Lei Chung

Contents

Competing financial interest declaration statement: The authors declare no competing financial interests. The authors further certify that their freedom to design, conduct, interpret, and publish research is not compromised by any controlling sponsor.

Disclaimer: Where authors are identified as personnel of the International Agency for Research on Cancer/World Health Organization, the authors alone are responsible for the views expressed in this article and they do not necessarily represent the decisions, policy, or views of the International Agency for Research on Cancer/World Health Organization.

Z. Herceg (✉) · A. Ghantous
International Agency for Research on Cancer (IARC), Lyon, France
e-mail: hercegz@iarc.fr

F. F.-L. Chung (✉)
Department of Medical Sciences, School of Medical and Life Sciences, Sunway University, Petaling Jaya, Malaysia
e-mail: feliciacfl@sunway.edu.my

K. B. Michels (ed.), *Epigenetic Epidemiology*,
https://doi.org/10.1007/978-3-030-94475-9_13

Abstract

The epigenome has been proposed as a biosensor of past or cumulative exposures and could also be a disease mediator. Human cancers exhibit a wide range of epigenetic alterations characterized by progressive acquisition during tumorigenesis and potential reversibility. Epigenetic changes may occur early in cancer development, supporting the notion that disrupted epigenetic mechanisms precede and promote malignant transformation. Recent exciting advances in epigenomics that allow the analysis of the epigenome with unprecedented resolution have galvanized investigations in epigenetic epidemiology of cancer. Epigenome states are regulated by three basic mechanisms: DNA methylation, posttranslational histone modifications, and non-coding RNAs (ncRNAs). DNA methylation is the best characterized epigenetic modification, and it is the most extensively studied in epigenetic epidemiology. Whereas it has long been established that DNA methylation (and other epigenetic) changes are ubiquitous in tumour tissue, many recent studies provided evidence that cancer risk- and exposure-associated epigenetic changes can be detected in non-malignant adjacent tissues or surrogate tissues (such as peripheral blood), providing attractive targets for discovering novel biomarkers of exposure and risk stratification. In this chapter, we review evidence from retrospective and prospective studies supporting the utility of epigenetic markers as predictors of predisposition to cancer and risk stratification. We also discuss changes in the "epigenetic clock" associated with cancer susceptibility as well as the potential of identifying epigenetic markers from negative surgical margins as predictors of cancer recurrence risk.

Abbreviations

450k	Illumina Infinium HumanMethylation450k BeadChip
850k	Infinium MethylationEPIC BeadChip
AHRR	Aryl-Hydrocarbon Receptor Repressor gene
BRCA1	BReast CAncer gene 1
CCGA	Circulating Cell-free Genome Atlas
cfDNA	cell-free DNA
CHARM	Comprehensive high-throughput arrays for relative methylation
CpG	Cytosine followed by a Guanine
DMR	Differential Methylation Region
dmrff	Method for identifying differentially methylated regions
EPIC	The European Prospective Investigation into Cancer and Nutrition
EWAS	Epigenome-Wide Association Studies
HNSCC	head and neck squamous cell carcinoma

LASSO	Least Absolute Shrinkage and Selection Operator
LINE-1	long interspersed nuclear elements
lncRNA	long non-coding RNA molecules
MCCS	Melbourne Collaborative Cohort Study
MeDIP-seq	methylated DNA immunoprecipitation sequencing
miRNAs	microRNAs or
mRNA	messenger RNA
ncRNAs	non-coding RNAs
DNMT	DNA methyltransferase
PLCO	Prostate, Lung, Colorectal, and Ovarian Cancer Screening Trial
PLS-DA	Partial Least Squares Discriminant Analysis
RCC	renal cell carcinoma
RLM	Robust Linear Regression
RNAi	RNA interference
RRBS	reduced representation bisulphite-sequencing
seqlm	method for identifying differentially methylated regions in high density methylation data
Ten-eleven translocation	TET
WGBS	whole genome bisulphite-sequencing

13.1 Introduction

Epigenetic alterations are central events in virtually all human cancers. It is now known that epigenetic alterations, as in the classic view of cancer evolution, occur progressively throughout tumorigenesis. Distinct to epigenetic alterations, however, is their potential reversibility. Furthermore, epigenome deregulation may occur early in cancer development, supporting the notion that disrupted epigenetic mechanisms precede and promote oncogenic transformation.

Epigenetic modifications are involved in the recruitment of various proteins to DNA, such as members of the transcription machinery, which affect gene expression and chromatin organization. These mechanisms are crucial for the establishment and maintenance of transcriptional programmes that in turn control processes of embryonic development, maintenance of stem state (pluripotency), cellular differentiation, and protection of the organism from exogenous genomes (e.g. viral DNA) [1–3]. Epigenome states are regulated by three basic mechanisms: DNA methylation, posttranslational histone modifications, and non-coding RNAs (ncRNAs).

13.1.1 DNA Methylation

In the eukaryotic genome, the best characterized epigenetic modification is DNA methylation at CpG dinucleotides, that is, covalent addition of a methyl group to a fifth carbon cytosine atom (C) that is linked through a phosphodiester bond (p) to a guanine (G). Genomic regions rich in these dinucleotides are known as CpG islands, which are typically found in gene promoter regions where they play a key role in controlling gene activity through DNA methylation. Changes in DNA methylation trigger a reorganization of chromatin structure consequently altering gene expression. DNA methylation was the first epigenetic mechanism recognized as a crucial cellular mechanism during embryogenesis as well as the development of many diseases, notably cancer [3].

Unscheduled promoter hypomethylation often results in increased gene activity while aberrant hypermethylation leads to decreased gene activity or complete silencing [4, 5]. Beside its role in gene promoter regulation, the maintenance of the methylation profile in repetitive sequences is important for genomic stability due to the suppression of transposon and retrotransposon element movement throughout the genome [6]. DNA methylation can be divided into three phases: establishment (*de novo* DNA methylation), maintenance of the methylation profile, and demethylation (removal of the methyl group) [6]. Enzymes whose role is to establish and maintain a methylation pattern within CpG dinucleotides are called DNA methyltransferases. In mammals, two de novo DNA methyltransferases, DNMT3A and DNMT3B, have been described, which are essential for establishing DNA methylation, particularly during early embryonic development, thereby, regulating tissue-specific gene transcription and cell differentiation [6]. During mitotic division, daughter cells inherit a methylation profile from the mother cell through DNMT1 activity, also called maintenance methyltransferase [4]. The methyl group on cytosine can be removed passively during successive rounds of replication in the absence of DNA methylation maintenance machinery or actively through the action of enzymes from the TET family (Ten-eleven translocation) [7]. Mutations in TET enzymes disrupt DNA methylation homeostasis resulting in global hypermethylation and susceptibility to diseases (most commonly malignant diseases of the haematopoietic system) [8].

13.1.2 Histone Modifications

Eukaryotic DNA is condensed and "packaged" into chromatin using histone and non-histone proteins. The basic building block of chromatin is called the nucleosome and consists of a 147-strand DNA strand wrapped around a histone octamer, a protein complex made up of an H3-H4 dimer surrounded by a tetramer composed of two H2A-H2B dimers [9]. Posttranslational modifications of the N- and C-terminal tails of histones involve their covalent addition of various chemical groups, such as acetylation, methylation, phosphorylation, ubiquitination, sumoylation, ADP-ribosylation, and many others [10]. The combination of different modifications on histones is called the "histone code" [11], and these modifications

dictate different cellular processes, including transcription, DNA replication, and DNA repair. An intimate interplay exists between DNA methylation and histone modifications during gene regulation [1].

13.1.3 Non-coding RNAs

Regulation of gene transcription *via* non-coding RNA (ncRNA) molecules is a relatively recently described epigenetic mechanism (RNAi). They are involved in the regulation of numerous cellular signalling pathways, including those related to tumour initiation and progression [12]. Some ncRNAs, the microRNAs or miRNAs, are small in size, encompassing 20 to 22 nucleotides which bind to complementary messenger RNA (mRNA) sequences leading to their degradation; hence, the translation of mRNA into a protein is inhibited [13]. There are also long non-coding RNA molecules (lncRNA) that participate in the regulation of larger regions of chromatin. These ncRNAs participate in chromatin remodelling. For example, in the embryonic development of female mammals, non-coding XIST RNA coats the future inactive X chromosome and triggers a cascade of events, which includes specific posttranslational histone modifications and DNA methylation, leading to the stable silencing of the entire chromosome [14]. Regulation of transcription by modifying ncRNA molecules may play a role in cancer initiation and progression, and represents an attractive target for the development of new therapies [15].

13.2 Epigenetic Markers in Cancer Epidemiology

Whereas a predominant proportion of studies on cancer epigenetics investigated epigenetic changes in tumour tissues (demonstrating that epigenetic changes are ubiquitous in cancers), it has been hypothesized that cancer risk- and exposure-associated epigenetic changes can be detected in non-malignant tissues (such as normal adjacent tissue) or surrogate tissues (such as peripheral blood), providing attractive targets for discovering novel biomarkers of exposure and risk stratification. This is supported by several lines of evidence, as discussed below.

Firstly, some aberrant epigenetic alterations might result from environmental exposures and stochastic events that may occur in stem/progenitor cells, particularly during embryonic development, hence, propagating over many cell generations or even an entire lifespan. These epigenetic alterations may generate constitutional "epimutations" throughout the body or create mosaic patterns, that is, epigenetic heterogeneity that is not simply determined by tissue-specific epigenetic patterns (Fig. 13.1) [16, 17].

Secondly, some gene families, such as metastable epialleles and imprinted genes, tend to exhibit similar methylation levels across different tissues of the body, as demonstrated in our work [18]; this is partly due to the fact that such genes escape the wave of demethylation that occurs just after fertilization and erases the methylation marks inherited from the gametes [19].This "resetting" of the epigenome paves the way to the subsequent establishment of the DNA methylome landscapes of the

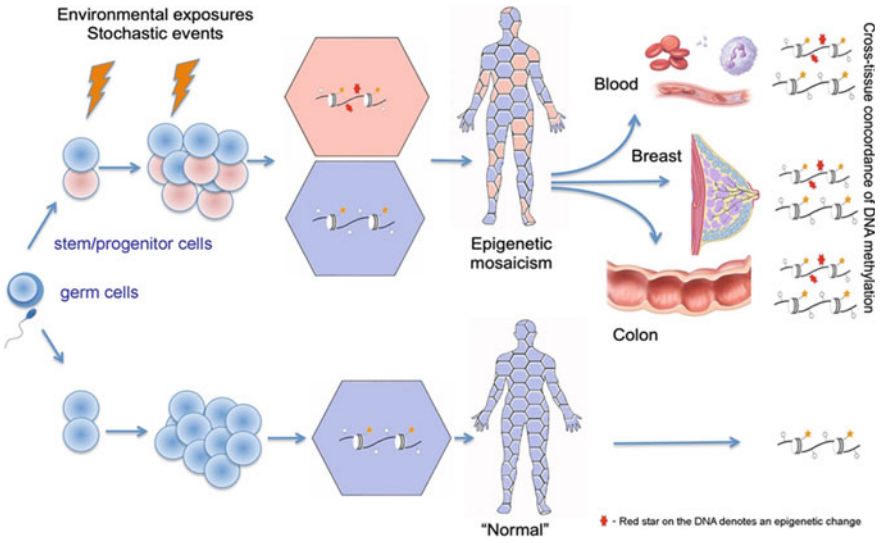

Fig. 13.1 Hypothetical model of epigenetic mosaicism as a mechanism of cancer causality and targets for biomarker discovery. (Adapted from Herceg et al. International Journal of Cancer 142(5):874–882, 2018; CC Attribution-NonCommercial 3.0 IGO)

various embryonic tissues. Interestingly, metastable epialleles and imprinted genes are known to function as key molecular sensors of exposures and to play central roles in carcinogenesis [18].

Thirdly, there is increasing evidence from recent studies showing that a gradual accumulation of changes in epigenetic patterns over time ("epigenetic drift"), through environmental hits and/or stochastic events, may exhibit a remarkable level of consistency across human tissue types and contribute to the risk of developing cancer [16] (and our unpublished data). One promising metric that captures, at least partly, epigenetic drift is the epigenetic clock, which is an algorithm based on sets of CpGs that are differentially methylated with ageing in multiple tissues and can be used to estimate a person's biological age (pan-tissue epigenetic clock) [20–23]. Subsequent successes in developing similar pan-tissue clocks for other species even hint at the universality of the epigenetic ageing mechanisms, which may be evolutionary conserved, particularly given their implication in developmental processes across several species [24]. The burden of many chronic diseases increases with chronological age, where changes at the tissue, cellular, and molecular levels (including epigenetic changes) accompany ageing in humans. Beyond its functioning as a clock, DNA methylation measures of ageing have also been used to track the accumulation of mutations, a key risk factor for cancers, as well as clinical outcomes, the rate of decline in physiological integrity, and health/life span [25]. For example, we have recently shown that epigenetic age acceleration is significantly associated with breast cancer susceptibility in postmenopausal women [26]. Collectively, measures of epigenetic drift represent a powerful toolkit for epidemiologic research

to understand the mechanisms underlying the effects of biological ageing and environmental and/or stochastic events on cancer development.

Epigenetic changes, with their great potential for modulation, have become a major focus in biomarker discovery. Recent studies on a variety of tumour samples, normal tissues, and surrogate tissues revealed that epigenetic changes may be risk factor-specific ("epigenetic signatures" of exposures), predictors (epigenetic biomarkers) of cancer risk and clinical outcome, and "mediators" between the exposure and the outcome [21, 27–30]. Furthermore, recent studies showed lifetime persistence or reversibility of specific epigenetic changes associated with environmental exposures (e.g. tobacco [27, 29]), highlighting the potential of epigenome patterns in surrogate tissues to serve as long-term or actionable biomarkers [16].

13.2.1 Evidence from Retrospective Studies

The interest in using epigenetic alterations as biomarkers for cancer or cancer predisposition has been explored in multiple contexts—using a variety of biospecimen types such as peripheral blood DNA, leukocyte DNA, or circulating free DNA, from patients with cancer or in pre-diagnostic samples. The field also spans leaps in analytical technology, including gene-specific, array-based, and sequencing-based technology. While these studies all share similar aims of identifying epigenetic marks of cancer or cancer predisposition, these variations in study design, biospecimen type, and methodological approach allow for different facets of this overarching question to be addressed.

Studies identifying epigenetic alterations in peripheral blood samples from symptomatic cancer patients provide compelling evidence that epigenomic biomarkers in the blood may aid the detection of solid cancers. Forming a large body of evidence, such studies have demonstrated comprehensively that epigenetic alterations in peripheral blood samples can be detected in multiple cancer types, including ovarian [31], bladder [32], head and neck [33], colorectal [34], and breast cancers [35].

While these studies often report epigenetic alterations in regulatory regions of known cancer-related genes, differential methylation is also commonly observed in genomic regions not associated with CpG islands, or in genomic regions not associated with known cancer-related genes. In a case–control study involving ovarian cancer patients, methylome profiling was conducted using DNA extracted from the blood cells of 113 individuals with ovarian cancer and 148 apparently healthy individuals [31]. This study led to the description of a DNA methylation signature that could predict for active ovarian cancer, and the identification of 2714 CpG sites that could constitute a list of cancer diagnostic CpGs [31]. Interestingly, the authors observed that the list was enriched with non-CpG island CpGs. Similarly, Langevin et al. [33] described a DNA methylation signature indicative of head and neck squamous cell carcinoma (HNSCC) based on methylome-wide profiles of peripheral blood samples from 92 patients with HNSCC and 92 cancer-free individuals. Amongst the six CpGs which allowed for discrimination between cases and controls, only one of the above was in a CpG island.

Notably, while findings from individual studies may report striking differences in DNA methylation patterns between cases and controls, these differences are not always replicable across independent studies. This could be attributed to differences in biospecimen types, sample processing methods, or analytical methods, but also highlights the need for a better understanding of the interindividual variability of DNA methylation. This heterogeneity could stem from heterogeneity in carcinogenic mechanisms, or heterogeneity across tumour types. For instance, Cappetta et al. [36] reported that none of the differentially methylated sites observed in their study, which was conducted with participants of European descent, coincided with CpGs detected in blood samples in similar studies in European populations. Not only does this underscore the need for population-specific studies and the detection of population-specific biomarkers, it also suggests that seemingly similar populations may display distinctly different epigenomes.

As such, a relatively low number of blood-based epigenetic biomarkers have emerged as being robust across populations to date. Arguably one of the most well-studied blood-based biomarker of breast cancer is that of the *BRCA1* promoter. Elevated levels of DNA methylation on the *BRCA1* promoter were observed more frequently in DNA extracted from peripheral blood samples of breast cancer patients compared to healthy controls, and were found to be associated with a significantly higher risk of breast cancer in a case–control study involving a total of 400 participants [37]. The authors observed methylation at the *BRCA1* promoter in 21.5% of the breast cancer patients, compared to 13.5% of the controls, and determined that women with promoter methylation of BRCA1 had a significantly higher risk (odds ratio = 1.73) of breast cancer compared to those not harbouring promoter methylation at the same sites [37]. Similar findings were reported in a separate population, where BRCA1 promoter methylation was examined in leukocyte DNA from 155 breast cancer patients and 143 cancer-free controls [38].

Interestingly, there was a significant association between a family history of breast cancer and the presence of *BRCA1* promoter methylation [38]. In addition, *BRCA1* promoter methylation may represent an epimutation which, like *BRCA1* mutations, may predispose to breast cancer. This is supported by the observation that cancer-free individuals who are carriers of a methylated *BRCA1* promoter in their blood display epigenetic, RNA, and plasma protein expression patterns similar to those of *BRCA1*-methylated breast cancer patients. This further suggests that promoter methylation of *BRCA1* may be informative as a blood biomarker for detecting predisposed individuals [38]. These findings complement those of an earlier case-only study, which reported substantial heterogeneity in BRCA1 methylation states amongst BRCA1-wild-type breast cancer patients [39]. The authors noted that tumours exhibiting more BRCA1 mutation-associated features (such as a strong family history, high mitotic index, and being hormone receptor negative) were more likely to show high levels of BRCA1 promoter methylation in their peripheral blood DNA and tumour DNA [39]. These findings highlight the importance of considering tumour heterogeneity within and between cancer types in the evaluation of epigenetic biomarkers.

Recent technological developments allowing for epigenome-wide analysis of cell-free DNA (cfDNA) have led to exciting breakthroughs in the field. A case–

control study investigating cell-free methylome profiles in pancreatic cancer patients demonstrated that there is a high degree of concordance between the epigenome of the cfDNA and that of the tumour tissue [40]. Machine learning classifiers developed using cfDNA were able to accurately differentiate acute myeloid leukaemia, lung and pancreatic cancer patients from all other cancer patients [40]. A similar study using samples from patients with intracranial tumours reported the development of a classifier that could reliably discriminate between patients with malignancy and healthy participants, and between five classes of intracranial tumours (IDH-mutant gliomas, IDH-wildtype gliomas, low-grade glial-neuronal tumours, meningiomas, and hemangiopericytomas) [41]. This technology has also been applied to the analysis of both plasma and urinary cfDNA from renal cell carcinoma (RCC) patients [42]. Machine learning classifiers developed using regions differentially methylated in the plasma and urinary cfDNA between cases and controls were able to discriminate between the two patient groups, underscoring the promise of cfDNA methylome-based biomarkers as screening tools for RCC detection [42].

Two recent multi-cancer studies on the cfDNA epigenomics have separately demonstrated the tissue-type specificity of cfDNA. The CancerLocator classifier was developed using plasma cfDNA methylome data from cancer-free individuals and large pools of publicly-available DNA methylomics data generated from solid tumour tissue [43]. The classifier was able to accurately classify normal, breast, colon, kidney, liver, and lung samples (with an error rate of 0.265) [43]. However, it should be noted that this study was conducted on a relatively small number of plasma samples, and was developed by comparing the methylome profiles of cfDNA from cancer-free individuals and tumour tissue DNA, as opposed to cfDNA from cancer patients. A large-scale study, the Circulating Cell-free Genome Atlas (CCGA) followed, involving 2482 cancer patients and 4207 cancer-free individuals, representing more than 50 cancer types across all stages [44]. The authors reported that the CCGA classifier performed at consistently high specificity (>99%) across cancer types and stages, though performance in terms of sensitivity was higher in late-stage compared to early-stage disease [44]. As the largest study on cell-free DNA methylomics to date, the findings of this study underscore the utility of methylation-based methods for population-level screening of cancer.

13.2.2 Evidence from Prospective Studies

The findings presented above offer promising evidence for the utility of epigenetic assays as tools for early detection or diagnostics. While it might be tempting to apply some of these markers as predictors of cancer predisposition, particularly as many of the discussed assays are minimally invasive, it should be noted that biomarkers developed and tested using specimens collected at diagnosis or from symptomatic individuals may exhibit reduced performance in a prospective setting.

For instance, one of the best-known blood-based epigenetic biomarkers is the *SEPT9* gene methylation assay. *SEPT9* assays have been tested in multiple retrospective case–control studies for the detection of colorectal cancer, where performance in terms of sensitivity has ranged between 71.1–95.6%, and 81.5% to 99% for

specificity [45]. However, the Prospective Evaluation of SEPTin 9 (PRESEPT) study, which was the first to evaluate the accuracy of the *SEPT9* gene methylation assay in a screening population, reported that while test specificity was 91.5%, standardized sensitivity was relatively low, at 48.2% [46]. The authors concluded that improvements to assay sensitivity would need to be made, particularly for advanced adenomas (11.2%), before the assay could be utilized as a test for population screening of colorectal cancer [46].

Despite the logistical challenges involved in conducting prospective cohort studies, epigenome-wide profiling has been applied to large prospective cohort studies in recognition of the value risk stratification markers could have if successful. As could be expected, however, most studies have reported that probe-wise epigenomic differences, if any, between nested cases and controls in prospectively collected specimens are relatively subtle.

Early studies utilized surrogate measurements of global methylation, such as LINE-1 DNA methylation levels. An early, large prospective study reported no differences in LINE-1 methylation between breast cancer cases and controls [47]. However, similar analyses in the Sister Study, which enrols women who had a biological sister that was diagnosed with breast cancer, indicated that LINE-1 methylation levels were associated with breast cancer risk [48].

Subsequent studies favoured the use of mean genome-wide methylation levels from array-based methylation assays. In a nested case–control study within the prospective Melbourne Collaborative Cohort Study (MCCS), Severi et al. [49] reported that cases exhibited lower levels of epigenome-wide methylation compared to controls. While the authors did not report the findings of probe-wise analyses, they noted that increased methylation within functional promoters was associated with an increased risk of breast cancer. Moreover, this study utilized multiple types of blood specimens, namely dried blood spots, buffy coats, or lymphocyte samples for DNA extraction, showing that the differences between cases and controls were similar for all specimen types.

Similarly, a study combining epigenome-wide data from three separate prospective nested case–control studies, one of which was the study MCCS reported by Severi et al. [49] above, demonstrated that epigenome-wide hypomethylation was associated with an increased risk of breast cancer [50]. On a probe-wise level, however, the authors reported that there was no overlap across the participating cohorts with regard to probe signatures associated with breast cancer risk.

These findings seemed to contrast those of an earlier report on the Sister Study, which identified 250 differentially methylated CpGs between cases and controls using the Illumina Infinium HumanMethylation27 BeadChip arrays [51]. A subsequent case-cohort study utilizing a newer iteration of the BeadChip arrays, the Infinium HumanMethylation450 BeadChips, identified 9601 CpG markers associated with invasive breast cancer, of which 2095 were replicated in the EPIC-Italy dataset mentioned above [52]. This study suggested that there are extensive DNA methylation differences between nested cases and controls, and that the DNA methylation profile in peripheral blood may be altered years before the tumour is clinically detected [52].

However, in a more recent meta-analysis combining nested case–control studies from four prospective cohorts, namely the Prostate, Lung, Colorectal, and Ovarian Cancer Screening Trial (PLCO), EPIC-International Agency for Research on Cancer (EPIC-IARC), and the aforementioned MCCS, EPIC-Italy studies, Bodelon et al. [53] reported no association between blood DNA methylation at individual CpG sites and breast cancer risk, with stratification by age, oestrogen receptor status, stage, or time since blood collection. The same meta-analysis found no evidence of an association between global methylation levels and breast cancer risk [53]. In gastric cancers, it was similarly reported that there were no significant associations between blood DNA methylation at individual CpG sites or regions with gastric cancer risk, based on the findings of a prospective case–control study nested within the MCCS [54]. Moreover, there was no evidence of association between global methylation levels and gastric cancer risk [54].

With the exception of reports from the Sister Study, current reports seem to indicate that DNA methylome differences between cases and controls in a prospective setting are subtle, if detected at all. As no detailed comparisons have been made between the individual prospective cohorts to date, we can only postulate that the cases identified in the Sister Study may harbour inherited or environmentally-induced epigenomic signatures unique to that study, as participants in the Sister Study were required to have had a biological sister with breast cancer. It is also worth noting that the prospective studies discussed above had utilized DNA isolated from buffy coat, dried blood spots, whole blood samples, or leukocyte samples, which may not directly reflect the molecular alterations occurring in the sites where cancers eventually develop. Instead, epigenomic changes in such samples could reflect changes in leukocyte cell composition, resulting from early cancer-related processes such as inflammation or immune response [55]. In contrast, cfDNA is generally considered to be a mixture of DNA shed as a result of cell death, including that occurring in cancer cells, and is known to be elevated in cancer patients [56–58]. Thus, cfDNA would be expected to better recapitulate the epigenome of the tissue from which it originated. It could also be possible that the above studies were limited by their use of microarray-based methods. Array-based methods, while informative and widely used, are relatively limited in coverage compared with sequencing-based profiling methods [59].

13.3 Identifying Epigenetic Markers from Negative Surgical Margins as Predictors of Cancer Recurrence Risk

The application of epigenetic marks as predictors of postoperative recurrence has emerged as a promising avenue for improving patient survival. While securing a negative surgical margin (one in which microscopically normal-appearing tissue is observed adjacent to the resect cancerous tissue) is an important goal in curative surgical treatment, a significant fraction of cancer patients with pathologically negative surgical margins suffer from postoperative local recurrence [60–63]. It is thus important that robust and reliable molecular markers capable of identifying individuals at high risk of postoperative recurrence are developed and used in

conjunction with existing clinical and pathological predictors to better allow for precise tailoring of treatment, preventive, and monitoring strategies for high-risk individuals. To date, the integration of molecular techniques with pathology practice has led to improved disease definitions and more accurate diagnostics [64]. Expanding this integration into the analysis of negative surgical margins may lead to the identification of molecular aberrations that are apparent across different pathological stages.

This is because the tissue surrounding a tumour may harbour molecular changes due to the presence of residual cancer cells resulting from the tumour budding, the phenomenon defined as cluster of cancer cells protruding into the tissues surrounding the invasive front. The presence and accumulation of these molecular changes in normal-appearing tissue, which is a part of the process known as the field cancerization, may be associated with the development of local recurrence or second primary tumours [65–67]. Both concepts imply an insufficient histological evaluation of normal-appearing adjacent tissues, advocating the need for additional methods with greater accuracy and sensitivity.

Early work on genetic alterations (mutation status) in histologically negative surgical margins showed that the probability of local recurrence of head and neck cancer was significantly correlated with positive molecular margins (the presence of mutation in TP53 gene) [68]. These findings were subsequently corroborated in multiple studies in head and neck cancer patients [69, 70]. However, a universal and robust predictive marker of cancer recurrence has yet to be reported.

Epigenetic alterations have also been detected in the tumour-adjacent normal tissues although it is unclear whether these changes develop prior to the onset of cancer. These changes may be detected in the negative margins that are undetectable on routine histopathological examination [67, 71]. Taken together, these findings support the notion that negative surgical margins might harbour specific epigenetic and other molecular aberrations that promote cancer development at a later stage [66, 72].

A number of studies focused on specific DNA methylation events with a targeted or genome-wide approach in several cancer types. Goldenberg et al. examined DNA methylation alterations of specific genes in the tumour and surgical margin samples of 13 HNSCC patients using a quantitative methylation-specific polymerase chain reaction (QMSP) assay and found that histologically negative margins exhibited tumour-specific methylation changes of the cyclin-dependent kinase inhibitor 2A (*p16*) and O-6-methylguanine-DNA methyltransferase (*MGMT*) genes [73]. Methylation status in other genes including E-cadherin type 1 (ECAD), death-associated protein kinase (DAPK) [74], Paired Box 5 (PAX5), Potassium Two Pore Domain Channel Subfamily K Member 12 (KCNK12), CD1d (CD1D) [75], and the deleted in colorectal cancer (DCC) gene [76] have also been associated with cancer recurrence. Major drawbacks of these studies are that they were based on retrospective patient cohorts, examining single genes or a small panel of genes. Moreover, in most cases the methylation status in matched tumour tissues was not examined.

To further refine the approach of molecular analysis of negative margins and avoid potential bias in sample selection, Hayashi et al. [67] used prospectively collected margin samples and matched tumour samples from a large series of

consecutive patients with head and neck cancer. By analysing methylation status of a selected set of genes in deep surgical margins, the authors reported that the methylation statuses of endothelin receptor type B (*EDNRB*) and homeobox protein Hox-A9 (*HOXA9*) in combination significantly predicted for locoregional recurrence [67].

In a more recent study, Sorroche et al. [77] conducted high-coverage methylome profiling (Infinium MethylationEPIC BeadChip) of negative surgical margins of 32 oral squamous cell carcinoma (OSCC) patients, half of whom suffered from local recurrence within five years after initial treatment. The authors identified 2512 recurrence-associated Differentially Methylated Positions (DMPs) and 392 Differentially Methylated Regions (DMRs) between recurrent and non-recurrent cases. Further analysis identified a set of 14-CpG markers capable of discriminating recurrent and non-recurrent cases with high specificity and sensitivity. A risk score was generated based on the 14-CpG marker panel, and was applied to identify individuals at significantly higher risk of recurrence-free survival. Interestingly, these findings were replicated on a larger series of tumour-adjacent normal samples from The Cancer Genome Atlas (TCGA) [77], reinforcing the notion that the residual DNA methylation changes in the negative surgical margins could serve as predictor markers for tumour recurrence.

Together, accumulating evidence supports the notion that DNA methylation changes are present in the histologically normal tissues adjacent to the primary tumours and that aberrant DNA methylation patterns may prove useful markers to differentiate individuals with a higher risk of developing recurrences and predicting prognosis. Further studies should be focused on validation and assessment of the utility of these marker candidates. Finally, promising marker candidates identified from postoperative negative margin tissues may be further tested on surrogate tissues (such as oral exfoliated cells) to validate its efficacy.

13.4 Conclusions

Aberrations to the epigenome, which may occur in response to environmental exposures or endogeneous cellular processes, have important implications for gene regulation and genomic stability. By virtue of their reversibility, epigenetic patterns are widely considered to be reflective of current cellular states. Simultaneously, it has been demonstrated that epigenetic marks formed in response to certain exposures, such as tobacco smoke, are stable and can persist across multiple cell divisions. As such, epigenetic events have been put forward as promising biomarkers of exposure and disease. The large number of high-throughput epigenomic data collections available, and the sheer diversity in analytical techniques and bioinformatics tools that have been developed are testaments to the promise and the complexity of the field of epigenetic epidemiology.

However, it is worth noting that there has been limited overlap between the biomarkers identified from different studies. A possible explanation for this is that there are substantial variations between studies with regard to sample type, storage conditions, and analytical approaches. As epigenomic signatures are sensitive to the cellular context, the diversity between studies could hinder subsequent efforts at

replication or validation. Robust validation might require standardization with regard to sample collection or preparation methods. Alternatively, cfDNA-based epigenetic marks may be pursued as they have already demonstrated utility as robust, minimally invasive diagnostic tools. However, the feasibility of these methods has yet to be tested extensively in a prospective setting.

It is also important to keep in mind that EWAS studies have predominantly focused on adult cancers, and a lot is yet to be done on childhood cancers, which develop much earlier in life and may have an epigenetic *in utero* origin [78].

Future studies applying new generation arrays or deep sequencing and powerful and analytical approaches in conjunction with large samples sizes from prospective cohorts practicing consensus sample collection and preparation techniques should solidify the finding that epigenetic signatures and acceleration of the "epigenetic clock" may predict cancer susceptibility, and set the stage for the utility of epigenomic signatures as biomarkers for risk stratification in cancer risk-prediction models [79].

Acknowledgment The work in the Epigenomics and Mechanisms Branch (EGM) at IARC is supported by grants from the Institut National du Cancer (INCa, France), the European Commission (EC) Seventh Framework Programme (FP7) Translational Cancer Research (TRANSCAN) Framework, the Foundation ARC pour la Recherche sur le Cancer (France), and Plan Cancer-Eva-Inserm research grant to Z.H. FC was supported by a Postdoctoral Fellowship from the International Agency for Research on Cancer, partially supported by the EC FP7 Marie Curie Actions—People—Co-funding of regional, national, and international programmes (COFUND).

References

1. Vaissiere T, Sawan C, Herceg Z (2008) Epigenetic interplay between histone modifications and DNA methylation in gene silencing. Mutat Res 659:40–48
2. Jones PA, Issa JP, Baylin S (2016) Targeting the cancer epigenome for therapy. Nat Rev Genet 17:630–641
3. Jones AP (2012) Functions of DNA methylation: islands, start sites, gene bodies and beyond. Nat Rev Genet 13:484–492
4. Smith ZD, Meissner A (2013) DNA methylation: roles in mammalian development. Nat Rev Genet 14:204–220
5. Jones PA (2012) Functions of DNA methylation: islands, start sites, gene bodies and beyond. Nat Rev Genet 13:484–492
6. Greenberg MVC, Bourc'his D (2019) The diverse roles of DNA methylation in mammalian development and disease. Nat Rev Mol Cell Biol 20:590–607
7. Pastor WA, Aravind L, Rao A (2013) TETonic shift: biological roles of TET proteins in DNA demethylation and transcription. Nat Rev Mol Cell Biol 14:341–356
8. Williams K, Christensen J, Helin K (2011) DNA methylation: TET proteins-guardians of CpG islands? EMBO Rep 13:28–35
9. Vanzan L, Sklias A, Herceg Z, Murr R (2017) Mechanisms of histone modifications. In: Tollefsbol T (ed) *Handbook of epigenetics: the new molecular and medical genetics*, 2nd edn. Academic Press (Elsevier), pp 25–46
10. Zhao Z, Shilatifard A (2019) Epigenetic modifications of histones in cancer. Genome Biol 20: 245
11. Sawan C, Herceg Z (2011) Histone modifications and cancer. Adv Genet 70:57–85

12. Yang X, Liu M, Li M, Zhang S, Hiju H, Sun J, Mao Z, Zheng M, Feng B (2020) Epigenetic modulations of noncoding RNA: a novel dimension of Cancer biology. Mol Cancer 19:64
13. Krutovskikh VA, Herceg Z (2010) Oncogenic microRNAs (OncomiRs) as a new class of cancer biomarkers. Bioessays 32:894–904
14. Holoch D, Moazed D (2015) RNA-mediated epigenetic regulation of gene expression. Nat Rev Genet 16:71–84
15. Rupaimoole R, Slack FJ (2017) MicroRNA therapeutics: towards a new era for the management of cancer and other diseases. Nat Rev Drug Discov 16:203–222
16. Herceg Z, Ghantous A, Wild CP, Sklias A, Casati L, Duthie SJ, Fry R, Issa JP, Kellermayer R, Koturbash I et al (2018) Roadmap for investigating epigenome deregulation and environmental origins of cancer. Int J Cancer 142:874–882
17. Hitchins MP (2015) Constitutional epimutation as a mechanism for cancer causality and heritability? Nat Rev Cancer 15:625–634
18. Van Baak TE, Coarfa C, Dugue PA, Fiorito G, Laritsky E, Baker MS, Kessler NJ, Dong J, Duryea JD, Silver MJ et al (2018) Epigenetic supersimilarity of monozygotic twin pairs. Genome Biol 19:2
19. Zeng Y, Chen T (2019) DNA methylation reprogramming during mammalian development. Genes (Basel) 10
20. Horvath S, Raj K (2018) DNA methylation-based biomarkers and the epigenetic clock theory of ageing. Nat Rev Genet 19:371–384
21. Horvath S (2013) DNA methylation age of human tissues and cell types. Genome Biol 14:R115
22. Zhang Q, Vallerga CL, Walker RM, Lin T, Henders AK, Montgomery GW, He J, Fan D, Fowdar J, Kennedy M et al (2019) Improved precision of epigenetic clock estimates across tissues and its implication for biological ageing. Genome Med 11:54
23. Hannum G, Guinney J, Zhao L, Zhang L, Hughes G, Sadda S, Klotzle B, Bibikova M, Fan JB, Gao Y et al (2013) Genome-wide methylation profiles reveal quantitative views of human aging rates. Mol Cell 49:359–367
24. MAMMALIAN METHYLATION CONSORTIUM, Lu AT, Fei Z, Haghani A, Robeck TR, Zoller JA, Li CZ, Zhang J, Ablaeva J, Adams DM et al (2021) Universal DNA methylation age across mammalian tissues. BioRxiv
25. Yu M, Hazelton WD, Luebeck GE, Grady WM (2020) Epigenetic aging: more than just a clock when it comes to cancer. Cancer Res 80:367–374
26. Ambatipudi S, Horvath S, Perrier F, Cuenin C, Hernandez-Vargas H, Calvez-Kelm FL, Durand G, Byrnes G, Ferrari P, Bouaoun L et al (2017) DNA methylome analysis identifies accelerated epigenetic aging associated with postmenopausal breast cancer susceptibility. Eur J Cancer 75:299–307
27. Ambatipudi S, Cuenin C, Hernandez-Vargas H, Ghantous A, Calvez-Kelm FL, Kaaks R, Barrdahl M, Boeing H, Aleksandrova K, Trichopoulou A et al (2016) Tobacco smoking-associated genome-wide DNA methylation changes in the EPIC study. Epigenomics 8:599–618
28. Fasanelli F, Baglietto L, Ponzi E, Guida F, Campanella G, Johansson M, Grankvist K, Johansson M, Assumma BM, Naccarati A et al (2015) Hypomethylation of smoking-related genes is associated with future lung cancer in four prospective cohorts. Nat Commun 6:10192
29. Joehanes R, Just AC, Marioni RE, Pilling LC, Reynolds LM, Mandaviya PR, Guan W, Xu T, Elks CE, Aslibekyan S et al (2016) Epigenetic signatures of cigarette smoking. Circ Cardiovasc Genet 9:436–447
30. Kresovich JK, Xu Z, O'Brien KM, Weinberg CR, Sandler DP, Taylor JA (2019) Methylation-based biological age and breast cancer risk. J Natl Cancer Inst 111:1051–1058
31. Teschendorff AE, Menon U, Gentry-Maharaj A, Ramus SJ, Gayther SA, Apostolidou S, Jones A, Lechner M, Beck S, Jacobs IJ, Widschwendter M (2009) An epigenetic signature in peripheral blood predicts active ovarian cancer. PLoS One 4:e8274
32. Marsit CJ, Koestler DC, Christensen BC, Karagas MR, Houseman EA, Kelsey KT (2011) DNA methylation array analysis identifies profiles of blood-derived DNA methylation associated with bladder cancer. J Clin Oncol 29:1133–1139

33. Langevin SM, Koestler DC, Christensen BC, Butler RA, Wiencke JK, Nelson HH, Houseman EA, Marsit CJ, Kelsey KT (2012) Peripheral blood DNA methylation profiles are indicative of head and neck squamous cell carcinoma: an epigenome-wide association study. Epigenetics 7: 291–299
34. Jung G, Hernandez-Illan E, Moreira L, Balaguer F, Goel A (2020) Epigenetics of colorectal cancer: biomarker and therapeutic potential. Nat Rev Gastroenterol Hepatol 17:111–130
35. Yang R, Stocker S, Schott S, Heil J, Marme F, Cuk K, Chen B, Golatta M, Zhou Y, Sutter C et al (2017) The association between breast cancer and S100P methylation in peripheral blood by multicenter case-control studies. Carcinogenesis 38:312–320
36. Cappetta M, Fernandez L, Brignoni L, Artagaveytia N, Bonilla C, Lopez M, Esteller M, Bertoni B, Berdasco M (2021) Discovery of novel DNA methylation biomarkers for non-invasive sporadic breast cancer detection in the Latino population. Mol Oncol 15:473–486
37. Iwamoto T, Yamamoto N, Taguchi T, Tamaki Y, Noguchi S (2011) BRCA1 promoter methylation in peripheral blood cells is associated with increased risk of breast cancer with BRCA1 promoter methylation. Breast Cancer Res Treat 129:69–77
38. Al-Moghrabi N, Nofel A, Al-Yousef N, Madkhali S, Bin Amer SM, Alaiya A, Shinwari Z, Al-Tweigeri T, Karakas B, Tulbah A, Aboussekhra A (2014) The molecular significance of methylated BRCA1 promoter in white blood cells of cancer-free females. BMC Cancer 14:830
39. Wong EM, Southey MC, Fox SB, Brown MA, Dowty JG, Jenkins MA, Giles GG, Hopper JL, Dobrovic A (2011) Constitutional methylation of the BRCA1 promoter is specifically associated with BRCA1 mutation-associated pathology in early-onset breast cancer. Cancer Prev Res (Phila) 4:23–33
40. Shen SY, Singhania R, Fehringer G, Chakravarthy A, Roehrl MHA, Chadwick D, Zuzarte PC, Borgida A, Wang TT, Li T et al (2018) Sensitive tumour detection and classification using plasma cell-free DNA methylomes. Nature 563:579–583
41. Nassiri F, Chakravarthy A, Feng S, Shen SY, Nejad R, Zuccato JA, Voisin MR, Patil V, Horbinski C, Aldape K et al (2020) Detection and discrimination of intracranial tumors using plasma cell-free DNA methylomes. Nat Med 26:1044–1047
42. Nuzzo PV, Berchuck JE, Korthauer K, Spisak S, Nassar AH, Abou Alaiwi S, Chakravarthy A, Shen SY, Bakouny Z, Boccardo F et al (2020) Detection of renal cell carcinoma using plasma and urine cell-free DNA methylomes. Nat Med 26:1041–1043
43. Kang S, Li Q, Chen Q, Zhou Y, Park S, Lee G, Grimes B, Krysan K, Yu M, Wang W et al (2017) CancerLocator: non-invasive cancer diagnosis and tissue-of-origin prediction using methylation profiles of cell-free DNA. Genome Biol 18:53
44. Liu MC, Oxnard GR, Klein EA, Swanton C, Seiden MV, Consortium C (2020) Sensitive and specific multi-cancer detection and localization using methylation signatures in cell-free DNA. Ann Oncol 31:745–759
45. Song L, Jia J, Peng X, Xiao W, Li Y (2017) The performance of the SEPT9 gene methylation assay and a comparison with other CRC screening tests: a meta-analysis. Sci Rep 7:3032
46. Church TR, Wandell M, Lofton-Day C, Mongin SJ, Burger M, Payne SR, Castanos-Velez E, Blumenstein BA, Rosch T, Osborn N et al (2014) Prospective evaluation of methylated SEPT9 in plasma for detection of asymptomatic colorectal cancer. Gut 63:317–325
47. Brennan K, Garcia-Closas M, Orr N, Fletcher O, Jones M, Ashworth A, Swerdlow A, Thorne H, Investigators KC, Riboli E et al (2012) Intragenic ATM methylation in peripheral blood DNA as a biomarker of breast cancer risk. Cancer Res 72:2304–2313
48. Deroo LA, Bolick SC, Xu Z, Umbach DM, Shore D, Weinberg CR, Sandler DP, Taylor JA (2014) Global DNA methylation and one-carbon metabolism gene polymorphisms and the risk of breast cancer in the Sister Study. Carcinogenesis 35:333–338
49. Severi G, Southey MC, English DR, Jung CH, Lonie A, McLean C, Tsimiklis H, Hopper JL, Giles GG, Baglietto L (2014) Epigenome-wide methylation in DNA from peripheral blood as a marker of risk for breast cancer. Breast Cancer Res Treat 148:665–673
50. van Veldhoven K, Polidoro S, Baglietto L, Severi G, Sacerdote C, Panico S, Mattiello A, Palli D, Masala G, Krogh V et al (2015) Epigenome-wide association study reveals decreased average methylation levels years before breast cancer diagnosis. Clin Epigenetics 7:67

51. Xu Z, Bolick SC, DeRoo LA, Weinberg CR, Sandler DP, Taylor JA (2013) Epigenome-wide association study of breast cancer using prospectively collected sister study samples. J Natl Cancer Inst 105(10):694–700
52. Xu Z, Sandler DP, Taylor JA (2020) Blood DNA methylation and breast cancer: a prospective case-cohort analysis in the Sister Study. J Natl Cancer Inst 112:87–94
53. Bodelon C, Ambatipudi S, Dugue PA, Johansson A, Sampson JN, Hicks B, Karlins E, Hutchinson A, Cuenin C, Chajes V et al (2019) Blood DNA methylation and breast cancer risk: a meta-analysis of four prospective cohort studies. Breast Cancer Res 21:62
54. Chamberlain JA, Dugue PA, Bassett JK, Milne RL, Joo JE, Wong EM, Brinkman MT, Stuart GW, Boussioutas A, Southey MC et al (2021) DNA methylation in peripheral blood and risk of gastric cancer: a prospective nested case-control study. Cancer Prev Res (Phila) 14:233–240
55. Houseman EA, Kelsey KT, Wiencke JK, Marsit CJ (2015) Cell-composition effects in the analysis of DNA methylation array data: a mathematical perspective. BMC Bioinform 16:95
56. Lin LH, Chang KW, Kao SY, Cheng HW, Liu CJ (2018) Increased plasma circulating cell-free DNA could be a potential marker for oral cancer. Int J Mol Sci 19
57. Valpione S, Gremel G, Mundra P, Middlehurst P, Galvani E, Girotti MR, Lee RJ, Garner G, Dhomen N, Lorigan PC, Marais R (2018) Plasma total cell-free DNA (cfDNA) is a surrogate biomarker for tumour burden and a prognostic biomarker for survival in metastatic melanoma patients. Eur J Cancer 88:1–9
58. Yan L, Chen Y, Zhou J, Zhao H, Zhang H, Wang G (2018) Diagnostic value of circulating cell-free DNA levels for hepatocellular carcinoma. Int J Infect Dis 67:92–97
59. Carmona JJ, Accomando WP Jr, Binder AM, Hutchinson JN, Pantano L, Izzi B, Just AC, Lin X, Schwartz J, Vokonas PS et al (2017) Empirical comparison of reduced representation bisulfite sequencing and Infinium BeadChip reproducibility and coverage of DNA methylation in humans. NPJ Genom Med 2:13
60. Sawabata N, Maeda H, Matsumura A, Ohta M, Okumura M (2012) Clinical implications of the margin cytology findings and margin/tumor size ratio in patients who underwent pulmonary excision for peripheral non-small cell lung cancer. Surg Today 42:238–244
61. Leemans CR, Tiwari R, Nauta JJ, van der Waal I, Snow GB (1994) Recurrence at the primary site in head and neck cancer and the significance of neck lymph node metastases as a prognostic factor. Cancer 73:187–190
62. Lee JY, Cho KB, Kim ES, Park KS, Lee YJ, Lee YS, Jang BK, Chung WJ, Hwang JS (2016) Risk factors for local recurrence after en bloc endoscopic submucosal dissection for early gastric cancer. World J Gastrointest Endosc 8:330–337
63. Hinerman RW, Mendenhall WM, Morris CG, Amdur RJ, Werning JW, Villaret DB (2004) Postoperative irradiation for squamous cell carcinoma of the oral cavity: 35-year experience. Head Neck 26:984–994
64. Romeo S, Dei Tos AP, Hogendoorn PCW (2012) The clinical impact of molecular techniques on diagnostic pathology of soft tissue and bone tumours. Diagn Histopathol 18:81–85
65. Caygill CP, Gatenby PA, Herceg Z, Lima SC, Pinto LF, Watson A, Wu MS (2014) Comparative genomic analysis of esophageal cancers. Ann N Y Acad Sci 1325:69–76
66. Slaughter DP, Southwick HW, Smejkal W (1953) Field cancerization in oral stratified squamous epithelium; clinical implications of multicentric origin. Cancer 6:963–968
67. Hayashi M, Wu G, Roh JL, Chang X, Li X, Ahn J, Goldsmith M, Khan Z, Bishop J, Zhang Z et al (2015) Correlation of gene methylation in surgical margin imprints with locoregional recurrence in head and neck squamous cell carcinoma. Cancer 121:1957–1965
68. Brennan JA, Mao L, Hruban RH, Boyle JO, Eby YJ, Koch WM, Goodman SN, Sidransky D (1995) Molecular assessment of histopathological staging in squamous-cell carcinoma of the head and neck. N Engl J Med 332:429–435
69. van Houten VM, Leemans CR, Kummer JA, Dijkstra J, Kuik DJ, van den Brekel MW, Snow GB, Brakenhoff RH (2004) Molecular diagnosis of surgical margins and local recurrence in head and neck cancer patients: a prospective study. Clin Cancer Res 10:3614–3620
70. Partridge M, Li SR, Pateromichelakis S, Francis R, Phillips E, Huang XH, Tesfa-Selase F, Langdon JD (2000) Detection of minimal residual cancer to investigate why oral tumors recur despite seemingly adequate treatment. Clin Cancer Res 6:2718–2725

71. Clark DJ, Mao L (2017) Understanding the surgical margin: a molecular assessment. Oral Maxillofac Surg Clin North Am 29:245–258
72. Braakhuis BJ, Bloemena E, Leemans CR, Brakenhoff RH (2010) Molecular analysis of surgical margins in head and neck cancer: more than a marginal issue. Oral Oncol 46:485–491
73. Goldenberg D, Harden S, Masayesva BG, Ha P, Benoit N, Westra WH, Koch WM, Sidransky D, Califano JA (2004) Intraoperative molecular margin analysis in head and neck cancer. Arch Otolaryngol Head Neck Surg 130:39–44
74. Supic G, Kozomara R, Jovic N, Zeljic K, Magic Z (2011) Prognostic significance of tumor-related genes hypermethylation detected in cancer-free surgical margins of oral squamous cell carcinomas. Oral Oncol 47:702–708
75. Sunagawa Y, Hayashi M, Yamada S, Tanabe H, Kurimoto K, Tanaka N, Sonohara F, Inokawa Y, Takami H, Kanda M et al (2021) Impact of molecular surgical margin analysis on the prediction of pancreatic cancer recurrences after pancreaticoduodenectomy. Clin Epigenetics 13:172
76. Tan H, Saulnier P, AupÈrin A, Lacroix L, Casiraghi O, Janot F, Fouret P, Temam S (2008) Quantitative methylation analyses of resection margins predict local recurrences and disease-specific deaths in patients with head and neck squamous cell carcinomas. Br J Cancer 99:357–363
77. Sorroche BP, Talukdar FR, Lima SCS, Melendez ME, de Carvalho AC, de Almeida GC, De Marchi P, Lopes M, Ribeiro Pinto LF, Carvalho AL et al (2021) DNA methylation markers from negative surgical margins can predict recurrence of oral squamous cell carcinoma. Cancers (Basel) 13
78. Ghantous A, Hernandez-Vargas H, Byrnes G, Dwyer T, Herceg Z (2015) Characterising the epigenome as a key component of the fetal exposome in evaluating in utero exposures and childhood cancer risk. Mutagenesis 30:733–742
79. Pashayan N, Antoniou AC, Ivanus U, Esserman LJ, Easton DF, French D, Sroczynski G, Hall P, Cuzick J, Evans DG et al (2020) Personalized early detection and prevention of breast cancer: ENVISION consensus statement. Nat Rev Clin Oncol

Epigenetic Epidemiology of Infectious Diseases

14

Toshikazu Ushijima, Yumi Furuichi, Hideyuki Takeshima, and Naoko Hattori

Contents

T. Ushijima (✉) · Y. Furuichi · H. Takeshima · N. Hattori
Division of Epigenomics, National Cancer Center Research Institute, Tokyo, Japan
e-mail: tushijim@ncc.go.jp

K. B. Michels (ed.), *Epigenetic Epidemiology*,
https://doi.org/10.1007/978-3-030-94475-9_14

Abstract

Many human cancers are induced by infectious agents, such as *Helicobacter pylori*, hepatitis viruses, Epstein-Barr virus (EBV), papilloma virus, liver flukes, and Schistosoma, and such cancers are often associated with extensive aberrant DNA methylation, including methylation of promoter CpG islands of tumor-suppressor genes and enhancers. The infectious agents cause aberrant DNA methylation via chronic inflammation and also by directly interacting with cellular proteins involved in DNA methylation regulation. Chronic inflammation induces repression of TETs, via NF-kB activation, and increased DNMT activity, via nitric oxide exposure, simultaneously, and these simultaneous changes lead to induction of extensive DNA methylation. Some target genes for methylation induction are dependent upon inducers and tissues, and can be used as a methylation fingerprint of exposure to an infectious agent. Infection-exposed tissues accumulate aberrant DNA methylation, and the degree of accumulation is often associated with cancer risk, forming an epigenetic field for cancerization. The usefulness of measurement of aberrant DNA methylation for cancer risk diagnosis has been demonstrated by a prospective clinical study.

Abbreviations

DNMT	DNA methyltransferase
DSS	Dextran sulfate sodium
EBV	Epstein-Barr virus
ESCC	Esophageal squamous cell carcinoma
H. pylori	*Helicobacter pylori*
H3Ac	Acetylation of histone H3
H3K27me3	Trimethylation of histone H3 lysine 27
H3K4me3	Trimethylation of histone H3 lysine 4
H3K9me3	Trimethylation of histone H3 lysine 9
HBV	Hepatitis B virus
HBx	The X protein of HBV
HCC	Hepatocellular carcinoma
HCV	Hepatitis C virus
HPV	Human papilloma virus
LMP1	Latent membrane protein 1
MeDIP	Methylated DNA immunoprecipitation
MSP	Methylation-specific PCR
NO	Nitric oxide
Pol II	RNA polymerase II
PRC	Polycomb repressive complex

14.1 Introduction

Infectious agents are some of the most well-known inducers of human cancers and aberrant DNA methylation. Infection with *Helicobacter pylori* (*H. pylori*), a bacterial strain causally involved in gastric carcinogenesis, is known to induce aberrant DNA methylation in gastric epithelial cells [1, 2]. Infection with hepatitis C and B viruses (HCV and HBV), both of which are involved in development of hepatocellular carcinomas, is associated with aberrant DNA methylation in cancer tissues and surrounding non-cancer tissues [3–5]. Infection with Epstein-Barr virus (EBV), associated with lymphomas, nasopharyngeal cancers, and gastric cancers, is also associated with frequent DNA methylation in tumor tissue [6–8]. Infection with papilloma virus, associated with cervical cancers and head and neck cancers [9, 10], is associated with aberrant DNA methylation in cancer tissues and surrounding non-cancer tissues [11–13].

The frequent presence of aberrant DNA methylation is accompanied by methylation of specific genomic regions or promoter CpG islands, producing a methylation signature or fingerprint [11, 14, 15]. This may provide an excellent tool for molecular epidemiology revealing past exposure to specific infectious agents, even when these agents are no longer present and serum antibody titer has declined. Also, accumulation levels of aberrant DNA methylation can be associated with cancer risk, reflecting the severity of past exposure to infection, host response to it, and build-up of tissue damage [15, 16].

In this chapter, we will introduce how we interpret DNA methylation data in cancer and non-cancer tissues, how aberrant DNA methylation and its fingerprint are induced by infectious agents in non-cancer tissues, and how we apply the methylation accumulation in non-cancer tissues to cancer risk estimation.

14.2 Interpretation of DNA Methylation Changes

Massive aberrant DNA methylation is present in cancer and non-cancer tissues exposed to infectious agents. However, such a picture has only recently been established. In addition, methylation patterns in monoclonal tissues (cancer tissues) polyclonal tissues (non-cancer tissues) need to be interpreted in a different manner.

14.2.1 Historical Aspects of the DNA Methylation Landscape in Cancers and in Non-cancer Tissues

In the 1990s, tumor-suppressor genes were preferentially analyzed for their aberrant DNA methylation. Resultantly, many investigators felt that methylation-silencing of a gene, due to methylation of a promoter CpG island, was a rare event, that such methylation is rarely present in non-cancer tissues, and that the growth advantage conferred by methylation-silencing makes a cell with it dominant in cancer tissue. In the 2000s, early-stage genome-wide analyses of differentially methylated genes were introduced. It was recognized that methylation of promoter CpG islands,

especially those of genes with no or little expression, was unexpectedly prevalent, and that we have to be cautious to interpret the significance of aberrant DNA methylation of promoter CpG islands [17]. In the 2010s, owing to the availability of robust genome-wide DNA methylation analysis techniques, it became evident that a large number of promoter CpG islands are aberrantly methylated, even in non-cancer tissues, such as gastric mucosa with *H. pylori* infection and liver tissues with hepatitis [4, 18].

14.2.2 The Meaning of DNA Methylation in Cancer and Non-cancer Tissues

Cancer is composed of cancer cells after monoclonal growth, although recent cancer genome sequencing has revealed extensive heterogeneity. In contrast, non-cancer tissues are composed of many clonal patches. A clonal patch in the intestine and stomach is composed of a gland (Fig. 14.1), and many other tissues, including the liver, esophagus, and skin [19–21], are known to consist of similar clonal patches.

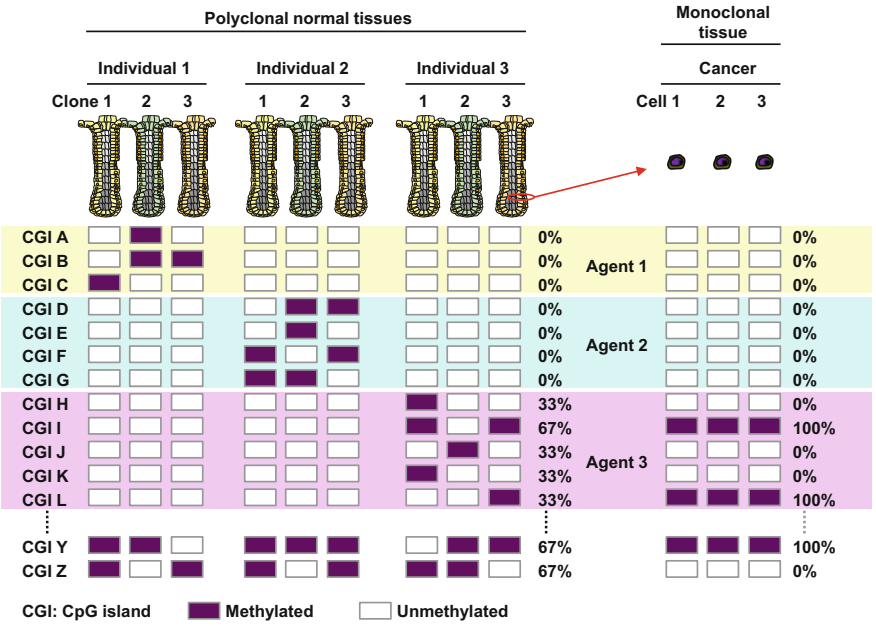

Fig. 14.1 Interpretation of methylation levels and patterns in polyclonal and monoclonal tissues. Many human tissues consist of clonal patches, as represented by clonal crypts (clones 1, 2, and 3), and cancer cells develop from one precursor cell. Suppose that specific methylation patterns of CpG islands (CGIs) are induced by different agents (agents 1, 2, and 3) in normal tissues of individuals. Polyclonal tissues (normal tissues) consistently reflect the patterns, but monoclonal tissues (cancer tissues) may or may not reflect the patterns. Methylation levels in polyclonal tissues reflect the fraction of clonal patches with methylation, and those in monoclonal tissues reflect the methylation status of the precursor cell

The monoclonal and polyclonal natures make important differences. All the cancer cells in a monoclonal tissue are theoretically derived from one precursor cell, and have a uniform methylation status for any promoter CpG islands. Therefore, the methylation level of a CpG island is expected to be 0, 50, or 100% when copy number abnormality is not present. In reality, this is often obscured by tumor heterogeneity and variable methylation statuses of contaminating normal and stromal cells. In contrast, a polyclonal tissue has many clones, each of which has its own methylation status of a CpG island, and the methylation level in such a tissue reflects the fraction of clones with methylation of the CpG island. Usually, a sample from a normal tissue contains a large number of clones (clonal patches), and its methylation level reflects the overall methylation level of the CpG island.

In addition to methylation levels, patterns of methylated CpG islands have different meanings in monoclonal and polyclonal tissues. A pattern in a monoclonal tissue reflects that of its precursor cell, and many CpG islands happen to be or not to be methylated in the cell. Supposing that specific CpG islands tend to be methylated, the precursor cell or resultant monoclonal cancer cells may or may not reflect the pattern. In our daily analysis of clinical samples, one monoclonal sample from an individual may or may not reflect the pattern. This makes it essential to analyze a group of samples with one etiology to assess its methylation pattern. In contrast, polyclonal tissues have many clones, all of which have been exposed to an agent. Supposing that specific CpG islands tend to be methylated by the agent, such methylation patterns are likely to be detected in one polyclonal sample from an individual.

14.3 Induction Mechanisms of Aberrant DNA Methylation

Infection is consistently associated with inflammation, and some inflammation has a mechanism to induce aberrant DNA methylation. In addition, some viruses have direct effects on cellular proteins involved in DNA methylation regulation.

14.3.1 Role of a Specific Type of Chronic Inflammation

The role of chronic inflammation in methylation induction was originally proposed based on the observation that aberrant methylation of specific genes was present in colonic mucosae of patients with ulcerative colitis [22, 23]. Direct evidence for the role of chronic inflammation in methylation induction was provided by an animal model of methylation induction by *H. pylori* infection. When inflammation by *H. pylori* infection was suppressed by treating Mongolian gerbils with an immunosuppressant, cyclosporin A, methylation induction was markedly suppressed without affecting colonization of *H. pylori* [24]. This showed that it is inflammation, not *H. pylori* itself, that is involved in methylation induction. Also in the liver, while infection of hepatocytes by HCV in vitro produced no DNA methylation changes, infection of humanized mice liver by HCV induced methylation [5]. This again showed that inflammation is important for methylation induction.

In addition, specific inflammation appears to have great power to induce aberrant DNA methylation. Regarding the stomach, Mongolian gerbils were treated for 20 weeks with multiple types of inflammation, namely inflammation induced by *H. pylori* infection, a high concentration of salt, or a high concentration of ethanol [25]. All the three treatments induced severe tissue damage and strong cell proliferation. Nevertheless, aberrant DNA methylation was induced only by *H. pylori* infection triggered inflammation. Also, in the liver, HCV has been consistently associated with stronger methylation induction [3, 5].

14.3.2 Mechanisms of Aberrant DNA Methylation Induction by Chronic Inflammation

In *H. pylori* infection triggered gerbil gastritis and dextran sulfate sodium (DSS)-induced mouse colitis, high expression of IL-1β, TNF-α, and NOS2 was consistently associated with DNA methylation induction. IL-1β/TNF-α is known to activate the NF-κB pathway, and the activation was shown to induce expression of multiple TET-targeting microRNAs, such as *miR-20a*, *miR-26b*, and *miR-29c* [26, 27]. Regarding NOS2, exposure to nitric oxide (NO) led to increased DNA methyltransferase (DNMT) activity [27, 28]. Importantly, either *TET* repression or increased DNMT activity-induced DNA methylation only very weakly, but simultaneous induction of both of them caused extensive aberrant DNA methylation (Fig. 14.2) [27]. Since simultaneous induction of IL-1β/TNF-α and production of nitric oxide are observed in many types of infection-induced chronic inflammation, this mechanism may have widespread involvement.

14.3.3 Direct Effect of Infectious Agents

In addition to the signals from chronic inflammation to epithelial cells, some infectious agents have a direct effect on cellular proteins involved in DNA methylation regulation [1, 29]. The X protein of HBV (HBx) upregulates DNMT1 and DNMT3A, and their increased activity led to hypermethylation of some tumor-suppressor genes, such as *CDKN2A* [30, 31]. In addition, increased production of reactive oxygen species in mitochondria by HBV is also reported to be involved in DNA methylation induction [32]. The HCV core protein upregulates DNMT1 and DNMT3B [33], and can recruit DNMT1 to a specific gene promoter, such as *SFRP1* [34]. Human papilloma virus 16 (HPV-16) E7 can increase DNMT1 activity, and can induce hypermethylation of tumor-suppressor genes [13, 35, 36]. Kaposi's sarcoma-associated herpes virus LANA protein interacts with DNMT3A and upregulates its activity [37].

The major EBV oncogene, latent membrane protein 1 (LMP1), upregulates protein expression and activity of DNMT1, DNMT3A, and DNMT3B in nasopharyngeal carcinoma cell lines and induces methylation of the tumor-suppressor genes *RARB* and *CDH1* in these cell lines [38, 39]. LMP2A, another EBV latent gene,

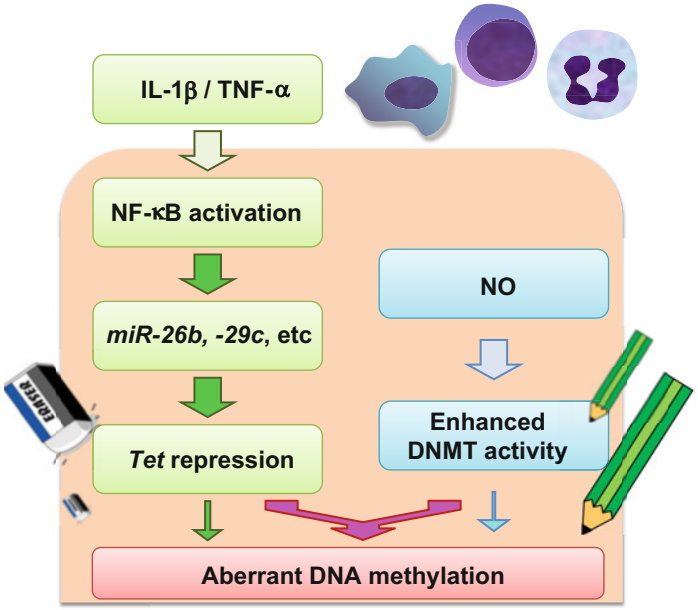

Fig. 14.2 Methylation induction mechanism by chronic inflammation. Tet-targeting microRNAs are induced by NF-κB activation, due to the IL-1β/TNF-α signalling, and Tet expression is repressed (eraser becomes small). Exposure to nitric oxide increased DNMT activity (writer becomes large). However, either of the changes induces only mild aberrant DNA methylation, but their combination potently induces aberrant DNA methylation of extensive genomic loci

upregulates DNMT1 and induces methylation of a tumor-suppressor gene, *PTEN*, in gastric cancer cell lines [40, 41]. LMP2A and seven human miRNAs, which were upregulated by EBV infection, can downregulate TET2 [24]. An extensive enhancer reprogramming by EBV was discovered recently [42].

14.4 Methylation Signature Produced by Specific Agents

It is now known that specific point mutations in specific sequence contexts are induced by specific agents, such as smoking, ultraviolet light, and aristolochic acid, and such specific profiles are known as mutation signatures [43]. Regarding DNA methylation, a specific agent induces methylation of specific group of genes, producing methylation signature or methylation fingerprints [15].

14.4.1 The Presence of Methylation Signature

Historically, the presence of target gene specificity in DNA methylation induction was initially indicated by the presence of methylation of specific genes in cancer

cells [44–46]. A pioneering study of 1184 CpG islands using restriction landmark genomic scanning, an early-stage genome-wide scanning technique for differences in DNA methylation, revealed that certain CpG islands were more frequently methylated in specific tumor types [44]. Analyses of promoter CpG islands of mostly tumor-suppressor genes suggested that some CpG islands were methylated at high incidences in specific tumor types [45]. Methylated DNA immunoprecipitation (MeDIP) microarray analysis of colon cancer tissue revealed that most methylated genes were located within defined genomic clusters [46]. Nevertheless, it remained unclear whether the methylation profiles were due to tissue types or inducers.

Afterward, a comparison of hepatocellular carcinomas (HCCs) associated with either HBV or HCV revealed that some genes are preferentially methylated in HCCs, depending on the specific hepatitis virus [3, 47], which was confirmed in a recent comprehensive genome-wide study [48]. Analysis of non-cancer liver tissues with HBV infection, HCV infection, alcohol, or non-alcoholic steatohepatitis clearly showed that each etiology produces a unique pattern of methylated CpG sites or CpG islands [4, 49]. A study using mice with humanized liver also showed that specific genes, in addition to common genes, are methylated by HCV and HBV [5].

Another example of methylation signature was reported for esophageal tissues of smokers and drinkers. Among the 13 promoter CpG islands methylated in squamous cell carcinomas (ESCCs), methylation levels of 5 genes (*HOXA9*, *MT1M*, *NEFH*, *RSPO4*, and *UCHL1*) in esophageal mucosae were significantly correlated with smoking duration [50]. Although smoking is not an infectious agent, this finding in non-cancerous esophageal mucosae supports the notion that a specific agent induces methylation of specific genes, leaving a methylation signature. Furthermore, oropharyngeal squamous cell carcinomas (OPSCCs) with HPV displayed specific methylation patterns different from OPSCCs without HPV [11].

14.4.2 Differences in Methylation Induction Between Aging and Chronic Inflammation

Aging is known to be an important inducer of aberrant DNA methylation, and inflammation is believed to accelerate age-related methylation [51]. Taking advantage of extensive induction of aberrant DNA methylation by *H. pylori* infection-triggered chronic inflammation, characteristics of CpG sites and promoter CpG islands methylated by aging and chronic inflammation were compared (Fig. 14.3) [18]. When gastric mucosa of non-infected old people was compared with that of non-infected young people (baseline), both hypermethylation and hypomethylation were observed. When gastric mucosa of infected young people was compared with the baseline, extensive hypermethylation and hypomethylation were observed. Comparison of age-related methylation and inflammation-induced methylation suggested that the age-related changes were accelerated by chronic inflammation. However, even when genomic regions not affected by aging were analyzed, extensive hypermethylation and hypomethylation were induced by chronic inflammation. This showed that chronic inflammation produces its signature by methylating unique

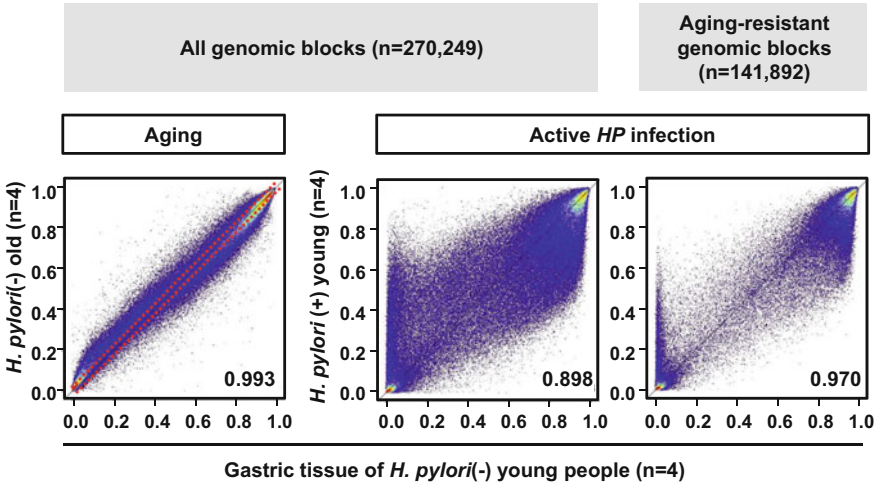

Fig. 14.3 The presence of specific targets of methylation induction by *H. pylori* infection differs from those by aging. 482,421 CpG sites were analyzed by Infinium microarray, and were assembled into 270,249 genomic blocks. To analyze age-related methylation (left panel), gastric tissues of elderly people without *H. pylori* infection were compared with those of young people without *H. pylori* infection (baseline). Slight hypermethylation and hypomethylation were observed. To analyze methylation induced by *H. pylori* infection (middle panel), gastric tissues of young people with *H. pylori* infection were compared with baseline. Extensive hypermethylation and hypomethylation were observed, indicating acceleration of age-related methylation by inflammation. However, when only genomic blocks unaffected by aging were analyzed (genomic blocks between the two red broken lines in the left panel; n = 141,892), still extensive hypermethylation and hypomethylation were observed, showing the presence of inflammation-specific methylation. Modified from Yamashita et al. [18] on an open access license (CC BY 4.0)

genomic regions. Such genomic regions were characterized by the lack of "seeds of methylation," and promoter CpG islands of highly expressed genes can be methylated by chronic inflammation.

14.4.3 Determinants for Target Genomic Regions for DNA Methylation Induction

Genomic regions and promoter CpG islands that undergo DNA methylation are instructed by preexisting epigenetic statuses, such as trimethylation of histone H3 lysine 27 (H3K27me3), gene expression level, and RNA polymerase II (Pol II), even if a gene has only low or no transcription (Fig. 14.4) [14].

The most well-known determinant is the presence of H3K27me3, which was proposed based upon analysis of representative genes [52–54] and confirmed by genome-wide analyses [55–57]. Mechanistically, H3K27me3 is recognized by a polycomb repressive complex (PRC) containing EZH2 [58–60], and EZH2 recruits DNA methyltransferases (DNMTs) [61, 62]. In contrast with the repressive

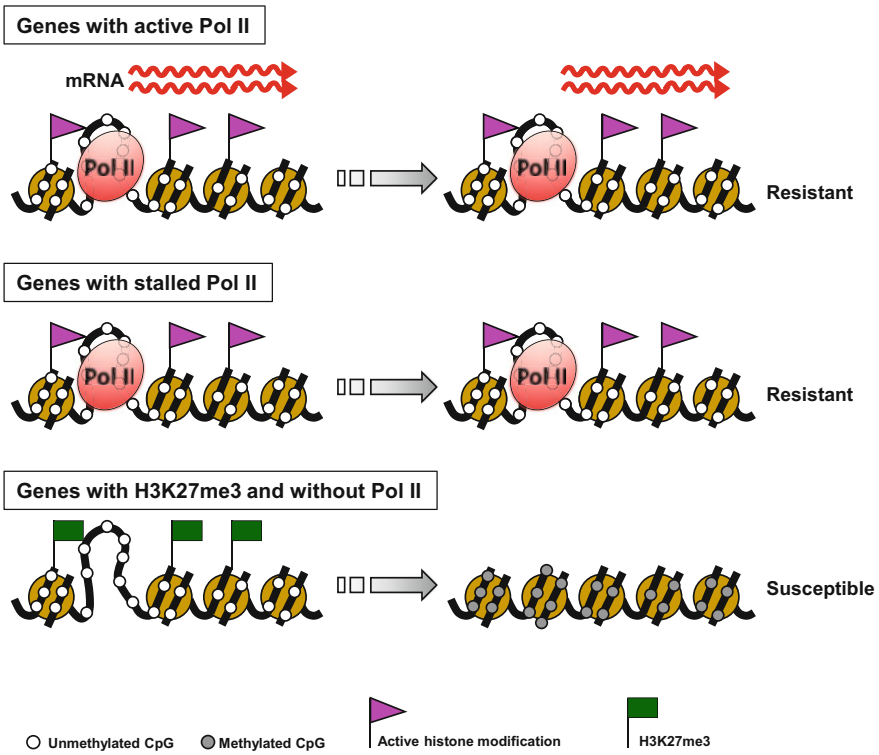

Fig. 14.4 Players involved in induction of DNA methylation in specific genes. Genes with active Pol II, 68% of which have active histone modifications (H3Ac), are resistant to DNA methylation induction. Genes with stalled Pol II, 19% of which also have active histone modifications (H3Ac), are also resistant. On the other hand, genes without Pol II, 90% of which are associated with H3K27me3 modification, are susceptible to methylation induction. Modified from Takeshima et al [14]

H3K27me3 modification, active histone modification, such as acetylation of histone H3 (H3Ac) and trimethylation of histone H3 lysine 4 (H3K4me3), are weakly associated with resistance to DNA methylation [57].

A low transcription level of a gene was proposed to be involved in methylation induction in the early 2000s [63–66]. Song et al. demonstrated that disruption of promoter activity (thus low transcription levels) of a transfected gene leads to aberrant DNA methylation of promoter CpG islands in a cancer cell line [63]. de Smet et al. demonstrated that a gene demethylated by a DNA demethylating agent, 5-aza-2′-deoxycytidine, tends to be re-methylated when it is not transcribed [64]. The majority of genes methylated in cancer tend to have low transcription levels in normal cells [46, 57, 65, 66]. Among genes methylated in non-cancerous tissues, genes susceptible to methylation induction had lower transcription levels than resistant genes [67].

Even among the genes with low transcription levels, some genes are still resistant to methylation induction. We have demonstrated that binding of Pol II to promoter CpG islands (stalled Pol II) is associated with resistance to methylation induction [57]. Multivariate analysis of transcription levels, H3K27me3, H3Ac, and Pol II binding suggested that Pol II binding had a stronger effect on DNA methylation resistance than active histone modifications. Taken together with the fact that transcribed genes are resistant, binding of Pol II, active or stalled, is associated with resistance to methylation induction during carcinogenesis.

In addition, the basal level of DNA methylation, or "seeds of methylation," was associated with a high chance of DNA methylation induction [63]. When a plasmid was introduced into a cell with preexisting sparse CpG methylation, it was readily DNA methylated. In contrast, when a plasmid was introduced without preexisting sparse CpG methylation, it was resistant to DNA methylation induction. Regarding genomic architecture, methylation-prone genes are located further apart from SINE and LINE retrotransposons compared with methylation-resistant genes [68].

14.4.4 Mechanism for Formation of a Methylation Signature of an Agent

As a mechanism of how a specific agent induces methylation of specific genes, the agent first induces changes in transcription, H3K27me3, and binding of Pol II in its target genes, and the changes then can lead to methylation induction of the genes that acquired a susceptible epigenetic status (Fig. 14.5). Since the first changes in epigenetic status are consistently induced depending upon the agent, DNA methylation is expected to be induced in specific genes, forming a methylation signature. This model was supported by time-course analysis of mouse colon tissues exposed to inflammation [69] and human gastric tissues with and without *H. pylori* infection [18]. Chemicals, such as cobalt compounds and cigarette smoke condensate, were also reported to induce alterations of H3K27me3 status [70, 71].

14.5 Application to Cancer Risk Estimation

A level of aberrant DNA methylation in a tissue shows overall epigenomic damage in the tissue [15]. Multiple cross-sectional studies showed that DNA methylation levels in normal tissues are correlated with cancer risk, and at least one prospective clinical study showed that DNA methylation levels in a normal tissue can predict future cancer risk.

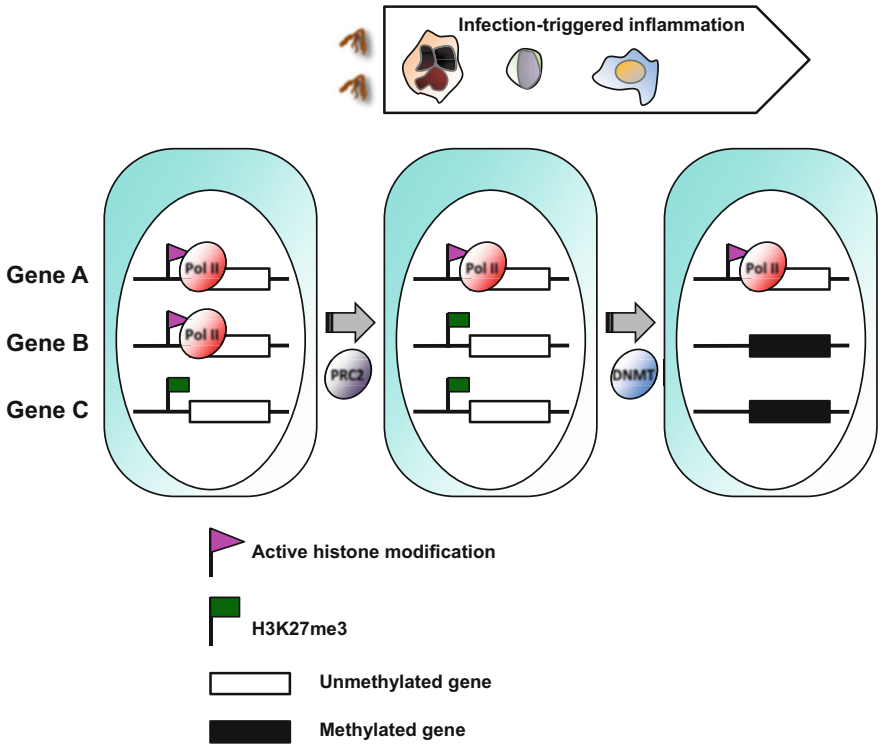

Fig. 14.5 A model for formation of a methylation signature by an infectious agent. An agent induces changes in epigenetic statuses, including H3K27me3 and binding of Pol II. A gene that acquires a susceptible epigenetic status by exposure to an agent, such as *H. pylori* infection, is expected to become methylated (Gene B). Since the changes in epigenetic statuses are consistently induced by an agent, specific genes are expected to be methylated by the agent. A gene with Pol II (Gene A) and one with H3K27me3 (Gene C), regardless of the exposure to an agent, are expected to be resistant and susceptible, respectively, to methylation induction

14.5.1 Association Between the Accumulation Level of Aberrant DNA Methylation and Cancer Risk

Association between the accumulation level of aberrant DNA methylation and cancer risk can be achieved by quantitative analysis of a large number of samples with low cancer risk and those with high cancer risk. Such association was observed for gastric tissues with *H. pylori* infection [72–76], liver tissues with HCV infection [77], and cervical tissues with HPV infection [78–80]. All these show that a predisposed tissue, namely an epigenetic field, is formed by infection (Fig. 14.6).

The accumulation level is influenced by the duration of exposure to the inducer [81] and host response to the inflammation. For example, a polymorphism in the promoter of IL-1β determines individuals who secrete more IL-1β in response to *H. pylori*-triggered inflammation, and thus have a high risk of gastric cancer

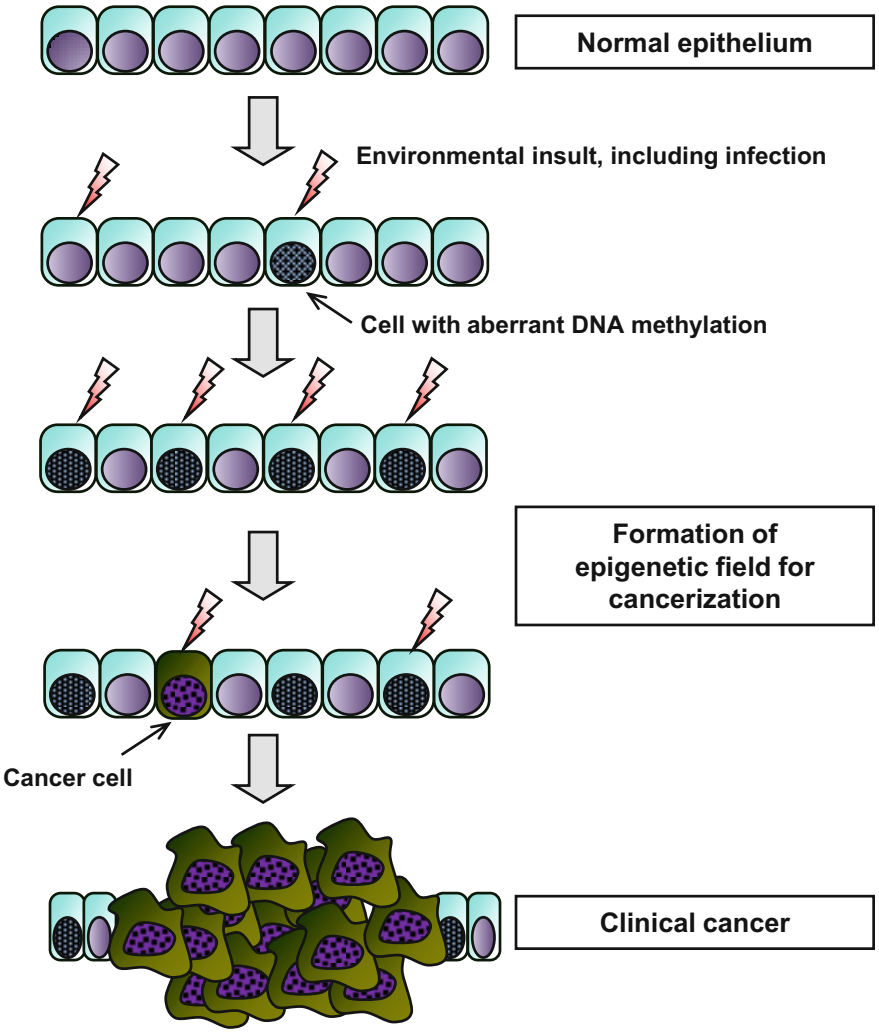

Fig. 14.6 Epigenetic field for cancerization. Exposure to infectious agents induces extensive aberrant DNA methylation in normal-appearing tissues. When infection continues, aberrant DNA methylation continues to be accumulated, and an epigenetic field for cancerization is produced. The DNA methylation accumulation is often associated with cancer risk

[82, 83]. The polymorphism was associated with DNA methylation levels in normal gastric tissue [84]. In contrast with gene expression levels, which fluctuate reflecting the degree of inflammation and host response, aberrant methylation levels simply increase because aberrant DNA methylation induced in stem cells will not be erased, and provides a reliable marker for tissue damage.

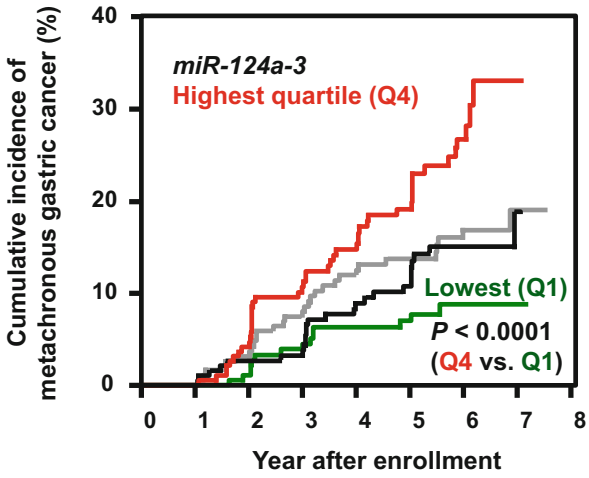

Fig. 14.7 The result of a prospective study to predict the risk of a second primary (metachronous) gastric cancer in gastric cancer patients cured by endoscopic treatment of their primary cancer. The patients with the highest methylation level had a threefold higher risk of developing a metachronous gastric cancer compared with those with the lowest methylation level. Modified from Maeda et al [86] on an open access license (CC BY-NC 4.0)

14.5.2 Validity Shown by a Prospective Clinical Study

To translate the finding of an epigenetic field, a prospective clinical study has been conducted in gastric cancer patients whose primary cancer was cured by endoscopic treatment [85, 86]. Using the methylation levels of three pre-selected genes at the time of enrollment after the endoscopic treatment, the risk of a second independent (metachronous) gastric cancer was successfully predicted (Fig. 14.7). The quartile with the highest methylation level had a threefold higher risk of developing a metachronous gastric cancer than the quartile with the lowest methylation levels after a median follow-up period of 5.46 years. This study showed that the accumulation level of aberrant DNA methylation can be used as a cancer risk marker.

14.6 Epigenetic Impacts on the Course of Infectious Disorders, Represented by COVID-19

Although cancer-causing infectious agents are mainly discussed in this chapter, some infectious agents can cause severe social and health damage, such as COVID-19 caused by SARS-COV-2. From this viewpoint, it is noteworthy that preexisting epigenetic alterations in human body can affect the course of infectious disorders. For example, expression of angiotensin converting enzyme-2 (ACE-2), a cellular receptor of SARS-COV-2, can be regulated by epigenetic alterations [87]. Genome-wide association analysis between the severity of COVID-19 and methylation levels of 850,000 CpG sites in peripheral lymphocytes revealed that DNA methylation of multiple genes involved in interferon response was associated with clinical severity [88, 89]. These findings illuminate another aspect of epigenetic epidemiology, which may be valid for other infectious agents severely involved in human health.

14.7 Epilogue

Aberrant DNA methylation is induced by various infectious agents, and inflammation is an important mechanism for the induction. Specific genes are methylated by specific inducers, forming methylation signatures. The potential application of methylation signatures in molecular epidemiology is to identify past exposure to infectious agents. Serum antibody titer against an infectious agent gradually decreases after elimination of the agent, but a methylation signature in a tissue will remain. Some efforts have been initiated to assess methylation signatures in DNA from peripheral blood leucocytes. If methylation signatures of various infectious agents can be assessed in blood DNA, they will be extremely useful to assess the involvement of an infectious agent in various human disorders. Further studies are needed.

Acknowledgment The original research described here was supported by AMED under Grant Number JP21ck0106552.

References

1. Hattori N, Ushijima T (2016) Epigenetic impact of infection on carcinogenesis: mechanisms and applications. Genome Med 8:10
2. Ushijima T, Hattori N (2012) Molecular pathways: involvement of *helicobacter pylori*-triggered inflammation in the formation of an epigenetic field defect, and its usefulness as cancer risk and exposure markers. Clin Cancer Res 18:923–929
3. Nishida N, Nagasaka T, Nishimura T et al (2008) Aberrant methylation of multiple tumor suppressor genes in aging liver, chronic hepatitis, and hepatocellular carcinoma. Hepatology 47: 908–918
4. Kuramoto J, Arai E, Tian Y et al (2017) Genome-wide DNA methylation analysis during non-alcoholic steatohepatitis-related multistage hepatocarcinogenesis: comparison with hepatitis virus-related carcinogenesis. Carcinogenesis 38:261–270
5. Okamoto Y, Shinjo K, Shimizu Y et al (2014) Hepatitis virus infection affects DNA methylation in mice with humanized livers. Gastroenterology 146:562–572
6. Kaneda A, Matsusaka K, Aburatani H, Fukayama M (2012) Epstein-Barr virus infection as an epigenetic driver of tumorigenesis. Cancer Res 72:3445–3450
7. Leong MML, Lung ML (2021) The impact of Epstein-Barr virus infection on epigenetic regulation of host cell gene expression in epithelial and lymphocytic malignancies. Front Oncol 11:629780
8. Yang J, Liu Z, Zeng B et al (2020) Epstein-Barr virus-associated gastric cancer: a distinct subtype. Cancer Lett 495:191–199
9. Mody MD, Rocco JW, Yom SS et al (2021) Head and neck cancer. Lancet
10. Xia C, Li S, Long T et al (2021) Current updates on cancer-causing types of human papillomaviruses (HPVs) in east, southeast, and South Asia. Cancers (Basel) 13:2691
11. Camuzi D, Buexm LA, Lourenco SQC et al (2021) HPV infection leaves a DNA methylation signature in oropharyngeal cancer affecting both coding genes and transposable elements. Cancers (Basel) 13:3621
12. Degli Esposti D, Sklias A, Lima SC et al (2017) Unique DNA methylation signature in HPV-positive head and neck squamous cell carcinomas. Genome Med 9:33
13. Nakagawa T, Kurokawa T, Mima M et al (2021) DNA methylation and HPV-associated head and neck cancer. Microorganisms 9:801

14. Takeshima H, Ushijima T (2010) Methylation destiny: Moira takes account of histones and RNA polymerase II. Epigenetics 5:89–95
15. Takeshima H, Ushijima T (2019) Accumulation of genetic and epigenetic alterations in normal cells and cancer risk. NPJ Precis Oncol 3:7
16. Ushijima T (2007) Epigenetic field for cancerization. J Biochem Mol Biol 40:142–150
17. Ushijima T (2005) Detection and interpretation of altered methylation patterns in cancer cells. Nat Rev Cancer 5:223–231
18. Yamashita S, Nanjo S, Rehnberg E et al (2019) Distinct DNA methylation targets by aging and chronic inflammation: a pilot study using gastric mucosa infected with helicobacter pylori. Clin Epigenetics 11:191
19. Adam AC, Faudou V, Paschen SA et al (2012) Hepatocarcinogenesis in non-cirrhotic liver is associated with a reduced number of clonal hepatocellular patches in non-tumorous liver parenchyma. J Pathol 228:333–340
20. Martincorena I, Fowler JC, Wabik A et al (2018) Somatic mutant clones colonize the human esophagus with age. Science 362:911–917
21. Martincorena I, Roshan A, Gerstung M et al (2015) Tumor evolution. High burden and pervasive positive selection of somatic mutations in normal human skin. Science 348:880–886
22. Hsieh CJ, Klump B, Holzmann K et al (1998) Hypermethylation of the p16INK4a promoter in colectomy specimens of patients with long-standing and extensive ulcerative colitis. Cancer Res 58:3942–3945
23. Issa JP, Ahuja N, Toyota M et al (2001) Accelerated age-related CpG island methylation in ulcerative colitis. Cancer Res 61:3573–3577
24. Niwa T, Tsukamoto T, Toyoda T et al (2010) Inflammatory processes triggered by *helicobacter pylori* infection cause aberrant DNA methylation in gastric epithelial cells. Cancer Res 70: 1430–1440
25. Hur K, Niwa T, Toyoda T et al (2011) Insufficient role of cell proliferation in aberrant DNA methylation induction and involvement of specific types of inflammation. Carcinogenesis 32: 35–41
26. Namba-Fukuyo H, Funata S, Matsusaka K et al (2016) TET2 functions as a resistance factor against DNA methylation acquisition during Epstein-Barr virus infection. Oncotarget 7:81512–81526
27. Takeshima H, Niwa T, Yamashita S et al (2020) TET repression and increased DNMT activity synergistically induce aberrant DNA methylation. J Clin Invest 130:5370–5379
28. Hmadcha A, Bedoya FJ, Sobrino F, Pintado E (1999) Methylation-dependent gene silencing induced by interleukin 1beta via nitric oxide production. J Exp Med 190:1595–1604
29. Pietropaolo V, Prezioso C, Moens U (2021) Role of virus-induced host cell epigenetic changes in cancer. Int J Mol Sci 22:8346
30. Jung JK, Arora P, Pagano JS, Jang KL (2007) Expression of DNA methyltransferase 1 is activated by hepatitis B virus X protein via a regulatory circuit involving the p16INK4a-cyclin D1-CDK 4/6-pRb-E2F1 pathway. Cancer Res 67:5771–5778
31. Park IY, Sohn BH, Yu E et al (2007) Aberrant epigenetic modifications in hepatocarcinogenesis induced by hepatitis B virus X protein. Gastroenterology 132:1476–1494
32. Yuan K, Lei Y, Chen HN et al (2016) HBV-induced ROS accumulation promotes hepatocarcinogenesis through snail-mediated epigenetic silencing of SOCS3. Cell Death Differ 23:616–627
33. Arora P, Kim EO, Jung JK, Jang KL (2008) Hepatitis C virus core protein downregulates E-cadherin expression via activation of DNA methyltransferase 1 and 3b. Cancer Lett 261:244–252
34. Quan H, Zhou F, Nie D et al (2014) Hepatitis C virus core protein epigenetically silences SFRP1 and enhances HCC aggressiveness by inducing epithelial-mesenchymal transition. Oncogene 33:2826–2835
35. Laurson J, Khan S, Chung R et al (2010) Epigenetic repression of E-cadherin by human papillomavirus 16 E7 protein. Carcinogenesis 31:918–926

36. Ekanayake Weeramange C, Tang KD, Vasani S et al (2020) DNA methylation changes in human papillomavirus-driven head and neck cancers. Cell 9:1359
37. Shamay M, Krithivas A, Zhang J, Hayward SD (2006) Recruitment of the de novo DNA methyltransferase Dnmt3a by Kaposi's sarcoma-associated herpesvirus LANA. Proc Natl Acad Sci U S A 103:14554–14559
38. Seo SY, Kim EO, Jang KL (2008) Epstein-Barr virus latent membrane protein 1 suppresses the growth-inhibitory effect of retinoic acid by inhibiting retinoic acid receptor-beta2 expression via DNA methylation. Cancer Lett 270:66–76
39. Tsai CN, Tsai CL, Tse KP et al (2002) The Epstein-Barr virus oncogene product, latent membrane protein 1, induces the downregulation of E-cadherin gene expression via activation of DNA methyltransferases. Proc Natl Acad Sci U S A 99:10084–10089
40. Chang MS, Uozaki H, Chong JM et al (2006) CpG island methylation status in gastric carcinoma with and without infection of Epstein-Barr virus. Clin Cancer Res 12:2995–3002
41. Hino R, Uozaki H, Murakami N et al (2009) Activation of DNA methyltransferase 1 by EBV latent membrane protein 2A leads to promoter hypermethylation of PTEN gene in gastric carcinoma. Cancer Res 69:2766–2774
42. Okabe A, Huang KK, Matsusaka K et al (2020) Cross-species chromatin interactions drive transcriptional rewiring in Epstein-Barr virus-positive gastric adenocarcinoma. Nat Genet 52: 919–930
43. Alexandrov LB, Kim J, Haradhvala NJ et al (2020) The repertoire of mutational signatures in human cancer. Nature 578:94–101
44. Costello JF, Fruhwald MC, Smiraglia DJ et al (2000) Aberrant CpG-island methylation has non-random and tumour-type-specific patterns. Nat Genet 24:132–138
45. Esteller M, Corn PG, Baylin SB, Herman JG (2001) A gene hypermethylation profile of human cancer. Cancer Res 61:3225–3229
46. Keshet I, Schlesinger Y, Farkash S et al (2006) Evidence for an instructive mechanism of de novo methylation in cancer cells. Nat Genet 38:149–153
47. Lambert MP, Paliwal A, Vaissiere T et al (2011) Aberrant DNA methylation distinguishes hepatocellular carcinoma associated with HBV and HCV infection and alcohol intake. J Hepatol 54:705–715
48. Sun S, Li Y, Han S et al (2019) A comprehensive genome-wide profiling comparison between HBV and HCV infected hepatocellular carcinoma. BMC Med Genet 12:147
49. Hlady RA, Tiedemann RL, Puszyk W et al (2014) Epigenetic signatures of alcohol abuse and hepatitis infection during human hepatocarcinogenesis. Oncotarget 5:9425–9443
50. Oka D, Yamashita S, Tomioka T et al (2009) The presence of aberrant DNA methylation in noncancerous esophageal mucosae in association with smoking history: a target for risk diagnosis and prevention of esophageal cancers. Cancer 115:3412–3426
51. Issa JP (2014) Aging and epigenetic drift: a vicious cycle. J Clin Invest 124:24–29
52. Ohm JE, McGarvey KM, Yu X et al (2007) A stem cell-like chromatin pattern may predispose tumor suppressor genes to DNA hypermethylation and heritable silencing. Nat Genet 39:237–242
53. Schlesinger Y, Straussman R, Keshet I et al (2007) Polycomb-mediated methylation on Lys27 of histone H3 pre-marks genes for de novo methylation in cancer. Nat Genet 39:232–236
54. Widschwendter M, Fiegl H, Egle D et al (2007) Epigenetic stem cell signature in cancer. Nat Genet 39:157–158
55. Gal-Yam EN, Egger G, Iniguez L et al (2008) Frequent switching of Polycomb repressive marks and DNA hypermethylation in the PC3 prostate cancer cell line. Proc Natl Acad Sci U S A 105:12979–12984
56. Hahn MA, Hahn T, Lee DH et al (2008) Methylation of polycomb target genes in intestinal cancer is mediated by inflammation. Cancer Res 68:10280–10289
57. Takeshima H, Yamashita S, Shimazu T et al (2009) The presence of RNA polymerase II, active or stalled, predicts epigenetic fate of promoter CpG islands. Genome Res 19:1974–1982

58. Fischle W, Wang Y, Jacobs SA et al (2003) Molecular basis for the discrimination of repressive methyl-lysine marks in histone H3 by Polycomb and HP1 chromodomains. Genes Dev 17: 1870–1881
59. Min J, Zhang Y, Xu RM (2003) Structural basis for specific binding of Polycomb chromodomain to histone H3 methylated at Lys 27. Genes Dev 17:1823–1828
60. Hansen KH, Bracken AP, Pasini D et al (2008) A model for transmission of the H3K27me3 epigenetic mark. Nat Cell Biol 10:1291–1300
61. Vire E, Brenner C, Deplus R et al (2006) The Polycomb group protein EZH2 directly controls DNA methylation. Nature 439:871–874
62. Mohammad HP, Cai Y, McGarvey KM et al (2009) Polycomb CBX7 promotes initiation of heritable repression of genes frequently silenced with cancer-specific DNA hypermethylation. Cancer Res 69:6322–6330
63. Song JZ, Stirzaker C, Harrison J et al (2002) Hypermethylation trigger of the glutathione-S-transferase gene (GSTP1) in prostate cancer cells. Oncogene 21:1048–1061
64. De Smet C, Loriot A, Boon T (2004) Promoter-dependent mechanism leading to selective hypomethylation within the 5′ region of gene MAGE-A1 in tumor cells. Mol Cell Biol 24: 4781–4790
65. Hagihara A, Miyamoto K, Furuta J et al (2004) Identification of 27 5' CpG islands aberrantly methylated and 13 genes silenced in human pancreatic cancers. Oncogene 23:8705–8710
66. Furuta J, Nobeyama Y, Umebayashi Y et al (2006) Silencing of peroxiredoxin 2 and aberrant methylation of 33 CpG islands in putative promoter regions in human malignant melanomas. Cancer Res 66:6080–6086
67. Nakajima T, Yamashita S, Maekita T et al (2009) The presence of a methylation fingerprint of helicobacter pylori infection in human gastric mucosae. Int J Cancer 124:905–910
68. Estecio MR, Gallegos J, Vallot C et al (2010) Genome architecture marked by retrotransposons modulates predisposition to DNA methylation in cancer. Genome Res 20:1369–1382
69. Takeshima H, Ikegami D, Wakabayashi M et al (2012) Induction of aberrant trimethylation of histone H3 lysine 27 by inflammation in mouse colonic epithelial cells. Carcinogenesis 33: 2384–2390
70. Li Q, Ke Q, Costa M (2009) Alterations of histone modifications by cobalt compounds. Carcinogenesis 30:1243–1251
71. Liu F, Killian JK, Yang M et al (2010) Epigenomic alterations and gene expression profiles in respiratory epithelia exposed to cigarette smoke condensate. Oncogene 29:3650–3664
72. Maekita T, Nakazawa K, Mihara M et al (2006) High levels of aberrant DNA methylation in *helicobacter pylori*-infected gastric mucosae and its possible association with gastric cancer risk. Clin Cancer Res 12:989–995
73. Nakajima T, Maekita T, Oda I et al (2006) Higher methylation levels in gastric mucosae significantly correlate with higher risk of gastric cancers. Cancer Epidemiol Biomark Prev 15: 2317–2321
74. Kim HJ, Kim N, Kim HW et al (2021) Promising aberrant DNA methylation marker to predict gastric cancer development in individuals with family history and long-term effects of H. pylori eradication on DNA methylation. Gastric Cancer 24:302–313
75. Shin CM, Kim N, Park JH et al (2012) Prediction of the risk for gastric cancer using candidate methylation markers in the non-neoplastic gastric mucosae. J Pathol 226:654–665
76. Suzuki R, Yamamoto E, Nojima M et al (2014) Aberrant methylation of microRNA-34b/c is a predictive marker of metachronous gastric cancer risk. J Gastroenterol 49:1135–1144
77. Wijetunga NA, Pascual M, Tozour J et al (2017) A pre-neoplastic epigenetic field defect in HCV-infected liver at transcription factor binding sites and polycomb targets. Oncogene 36: 2030–2044
78. Lorincz AT, Brentnall AR, Vasiljevic N et al (2013) HPV16 L1 and L2 DNA methylation predicts high-grade cervical intraepithelial neoplasia in women with mildly abnormal cervical cytology. Int J Cancer 133:637–644

79. Kremer WW, Steenbergen R, Heideman D et al (2021) The use of host cell DNA methylation analysis in the detection and management of women with advanced cervical intraepithelial neoplasia: a review. BJOG 128:504–514
80. Wijetunga NA, Ben-Dayan M, Tozour J et al (2016) A polycomb-mediated epigenetic field defect precedes invasive cervical carcinoma. Oncotarget 7:62133–62143
81. Takeshima H, Niwa T, Toyoda T et al (2017) Degree of methylation burden is determined by the exposure period to carcinogenic factors. Cancer Sci 108:316–321
82. El-Omar EM, Carrington M, Chow WH et al (2000) Interleukin-1 polymorphisms associated with increased risk of gastric cancer. Nature 404:398–402
83. Loh M, Koh KX, Yeo BH et al (2009) Meta-analysis of genetic polymorphisms and gastric cancer risk: variability in associations according to race. Eur J Cancer 45:2562–2568
84. Yoo EJ, Park SY, Cho NY et al (2010) Influence of IL1B polymorphism on CpG island hypermethylation in *helicobacter pylori*-infected gastric cancer. Virchows Arch 456:647–652
85. Asada K, Nakajima T, Shimazu T et al (2015) Demonstration of the usefulness of epigenetic cancer risk prediction by a multicentre prospective cohort study. Gut 64:388–396
86. Maeda M, Nakajima T, Oda I et al (2017) High impact of methylation accumulation on metachronous gastric cancer: 5-year follow-up of a multicentre prospective cohort study. Gut 66:1721–1723
87. Rath S, Perikala V, Jena AB, Dandapat J (2021) Factors regulating dynamics of angiotensin-converting enzyme-2 (ACE2), the gateway of SARS-CoV-2: epigenetic modifications and therapeutic interventions by epidrugs. Biomed Pharmacother 143:112095
88. Castro de Moura M, Davalos V, Planas-Serra L et al (2021) Epigenome-wide association study of COVID-19 severity with respiratory failure. EBioMedicine 66:103339
89. Balnis J, Madrid A, Hogan KJ et al (2021) Blood DNA methylation and COVID-19 outcomes. Clin Epigenetics 13:118

Epigenetic Epidemiology of Inflammation and Rheumatoid Arthritis

15

Caroline Ospelt and Steffen Gay

Contents

Abstract

The field of epigenetic research has evolved dramatically in recent years and has given valuable insight into regulation and dysregulation of gene expression in health and disease. The knowledge of the mechanisms controlling epigenetic changes increased, and in more and more diseases, a role for epigenetics could be found. Rheumatoid arthritis is a chronic inflammatory, autoimmune disease that leads to joint destruction but also affects other tissues and organs. Its etiology is not yet clarified, but a combination of the genetic background and environmental factors is proposed to trigger the onset of the disease. Epigenetic changes might be the link between non-genetic risk factors and development of symptoms. Changes in DNA methylation and histone modifications have been detected in patients with RA and have been found to promote inflammation and joint destruction. Future studies will have to show which epigenetic changes are causative factors and which are induced at a later stage by the chronic

C. Ospelt (✉) · S. Gay
Department of Rheumatology, Center of Experimental Rheumatology, University Hospital Zurich, University of Zurich, Zurich, Switzerland
e-mail: caroline.ospelt@usz.ch; steffen.gay@usz.ch

K. B. Michels (ed.), *Epigenetic Epidemiology*,
https://doi.org/10.1007/978-3-030-94475-9_15

inflammatory environment seen in the disease. This knowledge holds the potential for new preventive, predictive, and therapeutic opportunities in RA.

Abbreviations

ACPA	anti-citrullinated peptide autoantibodies
DNMT	DNA methyltransferase
EBV	Epstein Barr virus
H3	histone 3
H4	histone 4
HAT	histone acetylases
HDACs	histone deacetylases
HERV	human endogenous retrovirus
IGF	Insulin-like growth factor
IL	interleukin
MHC	major histocompatibility complex
PBMCs	peripheral blood mononuclear cells
RA	rheumatoid arthritis
SFN	Sulforaphane
SLE	systemic lupus erythematodes
TLRs	Toll-like receptors
Xi	inactivated X chromosome

15.1 Introduction

The classic definition of epigenetics describes epigenetic changes as mitotically heritable changes in gene function that, in contrast to mutations, do not alter the sequence of the DNA. However, the more epigenetics is studied in health and disease, the more facets and extensions have been added to this definition. Although an intriguing concept, the stable and mitotically heritable changes induced by the environment exclude a number of chromatin modifications and transcriptional regulators, which, while short lived, profoundly influence gene expression and might initiate changes that are passed on to the next generation of cells. In the current review, we follow the definition of Adrian Bird, describing epigenetic events as "the structural adaption of chromosomal regions so as to register, signal or perpetuate altered activity states" [1]. This definition includes more general mechanisms that regulate gene expression without claiming to have transgenerational effects. Furthermore, it emphasizes the responsive nature of the epigenome, which is of particular interest in the context of the contribution of epigenetics to the development of diseases. Epigenetics may be the missing link

between risk factors such as aging, environmental toxins or co-morbidities and the development of disease symptoms.

15.2 Epigenetics in Immune Function and Inflammation

The initiation of an immune response demands activation of regulatory processes to keep the frail balance between efficient defense and self-destruction. Epigenetic modifications have been found to play an important role in the transcriptional regulation of genes modulating and coordinating the immune response at various levels.

The innate immune system is activated via binding of pathogen-associated molecular patterns, so-called PAMPs by innate immune receptors. One of the best-studied innate immune receptors is Toll-like receptors (TLRs) which initiate a fast pro-inflammatory response against invading pathogens. It has long been known that repeated stimulation of TLR pathways, in particular of TLR4 pathways leads to the induction of tolerance against the inducing agent [2]. This mechanism controls the pro-inflammatory response and protects the organism from development of a septic shock. Foster et al. could show that there are two categories of genes that are induced by TLR4 stimulation [3]. The first set of genes is only induced at the first response, becomes then silenced and is not any more expressed after further TLR4 stimulations. The second set of genes is induced after every TLR4 stimulation. This set includes genes responsible for elimination of pathogens, expression of further immune receptors and activators of the adaptive immune response. After an initial TLR stimulation both sets of genes undergo specific epigenetic modifications such as histone acetylation and methylation, which shape innate immune responses, improve the elimination of the pathogen and prevent inflammatory tissue damage. This epigenetically imprinted memory to an unspecific stimulus is called trained immunity and was not only found in immune cells, but also in stromal cells, such as fibroblasts [4].

Also adaptive immunity gets modulated by epigenetic mechanisms. The specific cytokine pattern produced by Th1 and Th2 T cell subtypes could be shown to be regulated by histone acetylation and methylation. While in Th1 cells expression of interferon γ and interleukin (IL)-4 is repressed by histone methylation, Th2 cells have no histone methylation at this locus but instead show hyperacetylation of histones at the IL-4-IL-13 promoter site facilitating gene expression [5–8]. In addition, expression of IL-2, a key regulator of T cell differentiation and survival is regulated by histone deacetylases (HDACs) [9].

The expression of a variety of pro-inflammatory cytokines, e.g., IL-1, IL-8, and IL-12 has been found to be controlled by HDACs. Thus, it can be assumed that modulation of the histone code is an important general mechanism to regulate the inflammatory response [10–12]. Accordingly, disturbances in epigenetic control mechanisms might promote excessive, chronic, or self-directed immune responses resulting in chronic inflammatory diseases and autoimmunity.

15.3 Rheumatoid Arthritis

To scale the impact of epigenetic changes on disease development, studies of identical twins with discordant disease are most valuable. In celiac disease concordance in monozygotic twins is more than 80%, suggesting a strong genetic component in this disease [13, 14]. In rheumatoid arthritis (RA), the concordance rate between monozygotic twins is about 15% [15, 16]. However, concordance rates are in general lower in diseases with low prevalence. Heritability, which estimates the genetic contribution, is independent of disease prevalence and is about 60% for RA [17]. Nevertheless, these data suggest a substantial role of non-genetic factors in RA etiology. In the following, we will discuss current evidence on the contribution of non-genetic factors to the development of RA and the role of epigenetic mechanisms.

15.3.1 Epidemiology of RA

RA is a systemic autoimmune disease, which simultaneously affects multiple joints and ultimately leads to irreversible damage of articular structures. Destructive arthritis primarily affects small joints of the hands and feet, but larger joints are also affected during the course of the disease. In a murine disease model, circulating synovial fibroblasts in the blood were connected to this typical pattern of joint involvement [18]. The most common extra-articular manifestations of RA are subcutaneous nodules, serositis, and amyloidosis.

15.3.1.1 Geography

Worldwide, 0.5–1% of the population is affected by RA. Incidence rates vary between countries; RA was reported to newly affect 9/100,000 individuals per year in France and 45/100,000 individuals per year in the USA [19, 20]. The prevalence of RA is generally lower in less economically developed countries and regions [21, 22]. These suggest that genetic variance, environmental and socioeconomic factors affect disease prevalence, however, data from epidemiologic studies also may reflect underdiagnosis due to lack of access to healthcare. In addition, lower life expectancy may also contribute to the low prevalence of RA in developing countries. Furthermore, these differences may be, at least partly, due to methodological differences, making it difficult to compare data from different studies. Studies measuring disease rates within a country are particularly useful in separating the contribution of genetic and non-genetic factors. High prevalence rates in specific population subgroups point to genetic components, whereas higher rates of disease in specific geographical areas are indicators of environmental influences. Examples for the first scenario are studies indicating a very high RA prevalence in Native Americans [23]. A larger genetic contribution can be assumed in these population groups, although social habits and living conditions may also play a role. Since the incidence and prevalence of RA in Pima Indians have significantly decreased over a 25-year time period, an important environmental contribution cannot be excluded

[24]. A Finnish study suggested geographical variation in the incidence of RA in Finland with more than 2.5 fold higher incidence rate in north-eastern compared to north-western Finland [25]. These variances could not be explained by genetic differences or by differences in health care or diagnostics. In central Finland, RA incidence rates declined from 45/100′000 in 1980 to 27/100′000 in 1995. Worldwide the incidence of RA has declined in many countries, e.g., in the USA or Japan [20, 26]. Since the genetic background in a population hardly changes in such a short time period, these studies support a role of environmental factors affecting RA etiology.

15.3.1.2 Sex

Like most autoimmune diseases, RA is more common in women than in men, however, the ratio changes with age. While in women, the RA incidence rate peaks at around 55 years of age and then remains stable, it increases in men up to an age of about 85 years. Accordingly, the sex ratio is highest in the mid-forties with 4:1, declines in the mid-fifties to about 2:1 and levels in women and men over the age of 70 years [20, 27]. A possible explanation is the role of sex hormones on inflammation and immune response, but mechanisms are insufficiently understood. Studies explored the influence of reproductive factors on RA risk, but provided inconsistent results. Pregnancy is generally considered as ameliorating factor in RA disease activity, and some studies identified breastfeeding as a possible inducer of postpartum RA flares [28, 29]. However, others could not confirm these results [30]. Furthermore, a protective role of breastfeeding on the risk to develop RA was found [31]. Similarly, reports on the influence of age at menarche, use of oral contraceptives and exogenous hormones are inconclusive [31–33].

15.3.1.3 Lifestyle

Occupational risk to develop RA was mainly described in professions where workers were exposed to silica [34]. Also exposure to mineral oils and mineral dust seems to lead to an increased risk of RA [35].

It remains unresolved whether infectious agents may influence the development of RA. One of the most prominent candidates is the Epstein Barr virus (EBV) but studies analyzing the presence of EBV in RA patients were often limited by technical shortcomings and results were conflicting [36–38]. However, RA patients consistently have higher titers of EBV antibodies compared to healthy controls [39, 40]. These may result from a generally dysregulated immune response in RA patients, since a clear link between EBV infection and RA is missing [41]. A variety of other infectious agents, e.g., parvovirus B19, *Mycoplasma pneumoniae* or *Proteus mirabilis* have also been implicated in increased risk of RA, but none provide strong and consistent evidence for a causal role in RA [42].

Studies on the role of dietary factors in RA etiology are compromised by the complexity of diet, making it difficult to single out the possible contribution of individual foods or nutrients and by the presence of unidentified confounding factors that cannot be corrected for. Conflicting results do not permit conclusions on the role of consumption of fruits and vegetables, dairy products, coffee, or alcohol

[43, 44]. Evidence that is more consistent links antioxidants and RA. High dietary intake of antioxidants was found to decrease the risk for RA in the Iowa Women's Health Study cohort of older women [45], and several studies found lower serum levels of antioxidants in individuals who subsequently developed RA compared to healthy individuals [46–48]. However, no association was found between intake of antioxidants from foods and supplements and the risk of RA in the large prospective Nurses' Health Study [49].

The role of smoking in RA etiology is supported by several studies and an elevated risk to develop RA was described in past smokers until at least 10 years after cessation [50–53]. A gene-environment interaction was identified in smokers that carry risk alleles within the *HLA-DRB1* locus. The *HLA-DRB1* gene codes for the β-chain of the major histocompatibility complex (MHC) class II receptor on antigen-presenting cells and is highly allelic diverse. *HLA-DRB1* RA risk alleles have a common amino acid sequence, which is important in peptide binding and is known as "shared epitope." Depending on the type of risk allele, individuals carrying shared epitope alleles are 1.5 to 3 times as likely to develop RA compared to non-carriers. Sub-classification of RA patients into those with anti-citrullinated peptide autoantibodies (ACPA) and without revealed that shared epitope alleles mainly influence the development of ACPA positive RA [54]. With an odds ratio 5.27 compared to shared epitope-negative non-smokers, smoking is a strong risk factor for the development of ACPA positive RA in individuals with the shared epitope but not in individuals without any risk alleles [55, 56].

15.3.2 Epigenetics in RA

In many rheumatic diseases, and in particular in RA, epigenetic mechanisms are believed to contribute to the pathogenesis and to be connected to the above-mentioned risk factors for RA. Several studies could show epigenetic changes in cells and tissues of RA patients, however, the causality of these changes in disease development is difficult to assess. Nevertheless, changes in DNA methylation of PBMCs in RA might be used as biomarkers to predict the response to therapies [57].

15.3.2.1 DNA Methylation

Transposable Elements

Transposable elements are mobile DNA sequences that can move to different positions in the genome either by transcription/reverse transcription (retrotransposons) or by a cut and paste mechanism (DNA transposons). These remnants of ancient infections comprise around 45% of the human genome and have a sustained effect on human cell biology and evolution by influencing gene expression and genomic organization. In humans the expression of transposable elements is among others regulated by DNA methylation [58]. Demethylation of transposal elements leads to their transcription, which can increase expression of neighboring genes by providing promoter functions or can disturb cell physiology by translation of transposon gene products.

The product of the retrotransposon *HRES-1*, which belongs to the family of human endogenous retrovirus (HERV), was shown to suppress CD4 expression on human CD4+ T cells [59]. Most important for autoimmune diseases, transposon products, in particular HERV proteins can induce an antibody response and these antibodies can cross-react with other auto-antigens [60]. In patients with the autoimmune disease systemic lupus erythematodes (SLE), expression of the HERV-E clone 4-1 *gag* transcripts correlated with the presence of the anti-nuclear antibodies anti-U1 ribonucleoprotein and anti-Sm [61]. Furthermore, HERV-E clone 4-1 transcripts were only found to be expressed in peripheral blood mononuclear cells (PBMCs) from SLE patients but not from healthy controls and their expression could be induced in vitro by DNA demethylation with 5-azacytidine [62].

In RA, the retrotransposable element LINE-1 is expressed in synovial fibroblasts from RA patients but not from patients with osteoarthritis (OA) and its expression can be induced by in vitro DNA demethylation [63]. Expression of transposable elements by loss of DNA methylation might be linked to decreased expression of the enzyme that promotes methylation of DNA, DNA methyltransferase (DNMT)-1, in RA synovial fibroblasts [64]. Functional consequences of increased LINE-1 expression in RA are however not analyzed in detail.

Imprinted Genes

In addition to transposable elements, imprinted genes are particularly sensitive to changes in DNA methylation. A subset of RA synovial fibroblasts was found to have loss of imprinting at the Insulin-like growth factor (IGF-2) gene locus [65]. High expression of IGF-2 was further found to correlate with a low-inflammatory type of RA, suggesting that in a subgroup of RA patients, loss of imprinting of the IGF-2 gene in synovial fibroblasts leads to higher expression of IGF-2, promoting synovial fibroblast proliferation and an inflammation-independent type of synovial invasion. A general role of loss imprinting in RA pathogenesis is however disputed by studies that have failed to show any influence of genomic imprinting on susceptibility to RA [66, 67].

Genome-Wide Changes in DNA Methylation

Genome-wide changes in DNA methylation have been found in stromal as well as immune cell populations in RA patients [68–70]. Pathway analysis implicated that the genes that are affected by changes in DNA methylation in RA patients are not random, but are involved in relevant pathogenic pathways previously shown to be active in RA. Since biospecimen in these studies were obtained from patients with manifested RA, the temporal relation between changes in DNA methylation and RA remains unclear. Studies analyzing DNA methylation in synovial fibroblasts could, however, show that changes in DNA methylation occur early in disease [71], even before the clinical diagnosis [72] and that DNA methylation patterns change in the course of the disease.

Promoter-Specific Changes in DNA Methylation

In addition to genome-wide changes in DNA methylation, promoter-specific changes in DNA methylation were described in RA patients. Increased production of the chemokine CXCL12 by RA synovial fibroblasts could be shown to be caused by hypomethylation of the CXCL12 promoter in these cells [73] Takami et al. found higher methylation rates in the promoter of the death receptor DR3 in synovial fibroblasts of RA patients compared to OA patients. Accordingly protein levels of the DR3 receptor were found to be lower in RA synovial fibroblasts compared to OA synovial fibroblasts, which might be the explanation for the lack of apoptosis seen in RA synovial fibroblasts after application of the physiologic ligand of DR3 [74].

Analysis of the "senescent" subset of T cells, namely CD4+ CD28-, which are mainly found in elderly people and in people with chronic inflammatory diseases such as RA revealed that demethylation leads to the specific gene expression pattern seen in these cells. Due to repetitive stress, these T cells downregulate ERK and JNK signaling pathways which was connected to loss of methylation via downregulation of DNMT1 and DNMT3a [75]. In PBMCs, it could furthermore been shown that in RA patients, methylation is lost at a specific site in the IL-6 gene promoter. Lack of methylation at this site was connected to higher levels of IL-6 transcripts after stimulation [76]. Most interestingly, differential DNA methylation was found in the MHC region, which is intimately connected to the genetic risk to develop RA, in PBMCs of RA patients [77]. The variability of DNA methylation in this region might mediate the genetic risk of the MHC region to develop RA.

Some of the epigenetic changes found in RA might result from repetitive stress during chronic inflammation and may evolve during disease and perpetuate the inflammatory process. Since the enzymes transferring methylation marks after mitosis have a much higher error rate compared to DNA polymerases, differences in methylation patterns accumulate with every mitotic cycle; this could explain why the risk of developing a variety of diseases, including RA increases with age [78]. Some epigenetic changes may already be induced years before the onset of disease, since peri- and postnatal nutrition and environment affect the epigenome in animal models. Even though some epidemiological studies suggest a connection between incidence rates of RA and year of birth, the underlying events are not clear, and studies on the early life origins of RA patients are scarce and some results are conflicting [20, 79]. Associations between being breastfed and risk of RA could be shown in one study, but were not confirmed by others [80, 81]. Whereas no relation between birth weight and Rheumatoid Factor positivity could be found, there seems to be a higher risk of developing RA associated with high birth weight [80, 82, 83].

Another lifestyle factor that affects DNA methylation is smoking which was shown to cause changes in DNA methylation in PBMCs of healthy smokers [84]. Interestingly, maternal smoking increased the risk to develop RA [85]. In conjunction with the genetic background, smoking-induced changes in DNA methylation may lay the groundwork for the development of RA.

15.3.2.2 Histone Modifications

In addition to DNA methylation, epigenetic regulation occurs via histone modifications. These modifications are more dynamic and varied and include acetyl, methyl, phosphate, and ubiquitin residues, which can be placed at different sites on the histone tails. The various histone modifications are interconnected and are influenced by DNA methylation creating an intricate system. The combination of different modifications at specific positions at a certain stage of transcription tips the scale for transcriptional repression or activation [86]. Unfortunately, the complexity of this system often leads to studies focussing on specific modifications at specific sites of the genome. Such data can only provide a limited picture of the complex interactions in vivo. Most intensely studied are acetylation and methylation of histone 3 (H3) and histone 4 (H4). Histone acetylation is commonly associated with transcriptional activation, whereas deacetylation by histone deacetylases (HDACs) leads to transcriptional repression.

In CD4+ T cells of SLE, patients global H3/H4 acetylation was decreased, which suggests transcriptional repression. Accordingly, levels of histone acetylases (HAT) were significantly lower in these patients [87]. Nevertheless, also HDAC2 and HDAC7 levels were lower in SLE patients than in controls. Histones are not the only targets of HDACs for deacetylation. Therefore, the expression of single HDAC is not necessarily correlating with global histone acetylation levels. In synovial tissues of RA patients, the HDAC/HAT balance was found to be shifted towards higher levels of HAT with hyperacetylated histones promoting gene expression [88]. Other studies found increased activity and expression of HDAC1 in RA synovial tissues [89]. In peripheral blood cells of RA patients, total HDAC activity was higher than in healthy controls, but activity of HATs was similar between patient PBMCs and healthy controls, thus suggesting a shift of the HDAC/HAT balance towards increased HDAC activity [90, 91].

In synovial fibroblasts, changes in histone acetylation are suggested to have a major impact on the inflammatory response. Stimulation with the pro-inflammatory factor TNF as well as with the TLR4 ligand LPS led to changes in histone acetylation that were connected to prolonged expression of pathogenic factors by synovial fibroblasts, propagating chronic inflammation in RA synovium [92, 93]. Downregulation of HDAC5 and upregulation of HDAC3 were identified as important factors regulating the production of interferon response genes in synovial fibroblasts [94, 95].

Histone modifications can be regulated by the activation of the transcription factor NF-κB. NF-κB is a key transcription factor in inflammation and cell growth and its activation leads to expression of a variety of pro-inflammatory cytokines and chemokines [96]. Saccani et al. observed in murine macrophages that not all NF-κB target genes are accessible for NF-κB binding and that H3 phosphorylation/phosphoacetylation or H4 acetylation needs to be increased before NF-κB can bind to the promoter of IL-6. Their data also suggest that different stimuli induce histone modifications at different gene locations, thereby shaping the resulting inflammatory response [97, 98]. H3 methylation seems to have a crucial role in basal and post-induction repression of a subgroup of NF-κB inducible genes, tightly

regulating gene expression and leaving only a small time window for activation [99]. These examples of NF-κB regulation by histone modifications underline the importance of epigenetics in the regulation of inflammation and indicate that small but lasting changes in the epigenetic code may have a sustained effect on inflammatory and immune responses.

Another hypothesis linking histone modifications to the development of autoimmune diseases suggests that changes in the histone code lead to the appearance of neo-epitopes which after cell death might be recognized and induce an autoimmune response [100]. In SLE and in chronic inflammatory bowel disease autoantibodies against the mono-ubiquitinated H2A are regularly found and also autoantibodies against poly (ADP) ribose, another histone modification, are present in the serum of these patients [101–103]. In RA patients, autoantibodies to citrullinated histones are found and a pathogenic role for these autoantibodies in the development of RA was suggested [104]. However, a causal role in disease pathogenesis has not been convincingly shown yet, and the appearance of these autoantibodies may be the consequence and not the cause of an aberrantly activated and disturbed immune system.

Histone modifications are much more dynamic than DNA methylation and a variety of external stimuli likely have an impact on the histone code. Sulforaphane (SFN) is an organosulfur compound that is present at high levels in cruciferous vegetables such as cabbage, broccoli and horseradish and is a naturally occurring HDAC inhibitor. Study participants were asked to consume one cup of broccoli sprouts after 48 h of a cruciferous vegetable-free diet and HDAC activity and H3/H4 acetylation was measured in PBMCs [105]. Three hours after broccoli consumption, a strong downregulation of HDAC activity was detected and H3 and H4 were hyperacetylated accordingly. This effect was still detectable 48 h after consumption of the broccoli sprouts. SFN is not the only naturally occurring HDAC inhibitor and the authors speculate that long-term intake of such modulators of the histone code may prime cells for an appropriate response to exogenous insults [106]. A combination of different dietary modulators of chromatin remodeling may interact to induce constant subtle changes in the epigenetic state, and this dynamic state may reduce susceptibility to diseases with epigenetic components such as chronic inflammation, autoimmunity, or cancer.

15.3.2.3 Sex Difference

One of the main questions regarding the occurrence of autoimmunity is why women are generally more frequently affected than men. Among different explanations including hormone status in particular estrogen levels and differences in environmental exposures, hypotheses relating to the sex chromosomes have drawn increasing attention. During embryogenesis, one of the two X chromosomes in the cells of female mammals gets inactivated. In humans, the choice between paternal or maternal X chromosome is randomly, but once made, the choice is permanent for the cells. Silencing is achieved by different epigenetic mechanisms. The future inactivated X chromosome (Xi) expresses a large non-coding RNA from the *Xist* (X inactive specific transcript) gene, which binds to Xi and suppresses gene

expression. Inactivation of Xi is completed by histone ubiquitination, methylation and loss of acetylation and DNA methylation [107]. The combination of these epigenetic modifications makes Xi silencing irreversible, albeit not complete. Only 75% of the genes on the Xi are constantly silenced, 15% are constantly expressed and another 10% are differential expressed between individuals [108]. The fact that there are genes that are constantly expressed at a higher level in females than in males and that the expression of some of these genes also varies between females might be of importance in the development of diseases. Furthermore, Xi chromosome reactivation may play a role. Methylation of the promoter region of the *CD40LG* gene on the X chromosome was diminished in women with SLE and transcript levels of this B cell co-stimulatory molecule were higher in affected women compared to healthy women but interestingly also compared to affected men [109]. However, direct proof that silencing in vivo was lost at this specific site of Xi in female SLE patients is lacking. Finally, skewed X chromosome inactivation has been implicated in disease development. In this case, the inactivation of the X chromosomes is not random, but either the maternal or the paternal X chromosome is preferentially silenced. The cause of skewed inactivation is not clear yet. Higher frequency of skewed inactivation patterns was found in peripheral blood cells of patients with scleroderma and RA compared to healthy controls [110, 111].

15.4 Conclusions and Perspectives

In recent years, epigenetic research has drastically increased and provided interesting and exciting insights into gene regulation and transcriptional control mechanisms. However, we are only beginning to understand how cells react to an altered environment via the intricate network of epigenetic modifications. Depending on the cell type, every individual gene has a particular pattern of DNA methylation and histone modifications, which change with its state of activation and must be tightly regulated. Environmental factors such as nutrition, toxins, or infections that can interfere with the epigenome may induce epigenetic changes over time that persist even in the absence of the triggering factor and shape response mechanisms.

Some epigenetic changes that are seen in chronic inflammation and autoimmunity may be induced by consistently aberrant activation of certain signaling pathways and perpetuate disease by facilitating accessibility to these pathways like a well-trodden path. As illustrated in Fig. 15.1 epigenetic changes might also be induced before disease onset by seemingly disease-unrelated factors shifting cellular pathways towards a vulnerable state that together with a susceptible genetic background or additional exogenous factors triggers disease. One key question for future research will therefore be which epigenetic changes occur before disease onset and which occur because of disease. Answers to this question will not only offer novel insights in disease pathogenesis, but also provide new preventive, predictive, and therapeutic opportunities.

Future epidemiologic studies will benefit from additional epigenetic insights and start to integrate epigenetic data. The combination of epidemiologic, genetic,

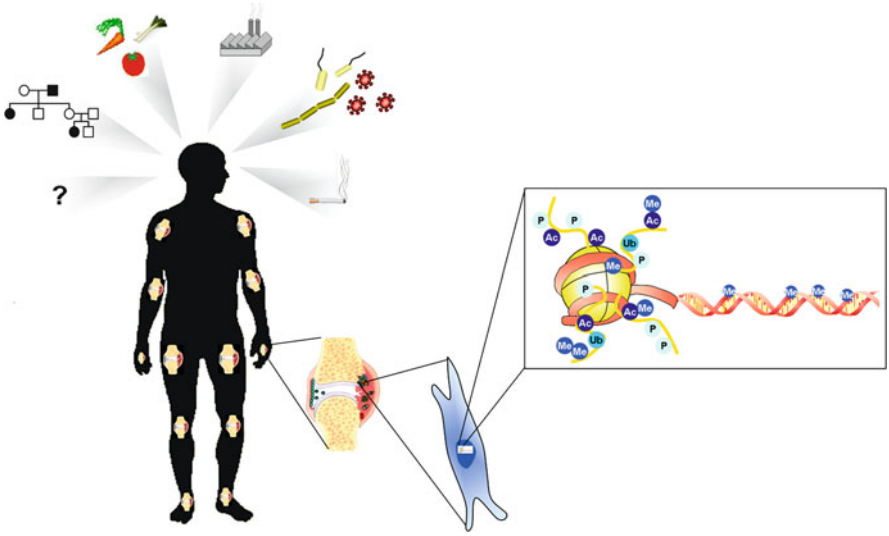

Fig. 15.1 The combination of a variety of different factors such as smoking, infections, pollution, nutrition, genetic predisposition and other factors may lead to changes in histone modifications and DNA methylation in local cells of the joints, thereby supporting the onset of chronic inflammation and autoimmunity. P = phosphorylation; Ac = acetylation; Me = methylation; Ub = ubiquitination

and epigenetic data can be a powerful tool to clarify susceptibility, outcome, and therapeutic response not only in RA but also in a variety of chronic inflammatory diseases.

References

1. Bird A (2007) Perceptions of epigenetics. Nature 447(7143):396–398
2. West MA, Heagy W (2002) Endotoxin tolerance: a review. Crit Care Med 30(1 Suppl):S64–S73
3. Foster SL, Hargreaves DC, Medzhitov R (2007) Gene-specific control of inflammation by TLR-induced chromatin modifications. Nature 447(7147):972–978
4. Dominguez-Andres J, Fanucchi S, Joosten LAB, Mhlanga MM, Netea MG (2020) Advances in understanding molecular regulation of innate immune memory. Curr Opin Cell Biol 63:68–75
5. Avni O, Lee D, Macian F, Szabo SJ, Glimcher LH, Rao A (2002) T(H) cell differentiation is accompanied by dynamic changes in histone acetylation of cytokine genes. Nat Immunol 3(7):643–651
6. Baguet A, Bix M (2004) Chromatin landscape dynamics of the Il4-Il13 locus during T helper 1 and 2 development. Proc Natl Acad Sci U S A 101(31):11410–11415
7. Koyanagi M, Baguet A, Martens J, Margueron R, Jenuwein T, Bix M (2005) EZH2 and histone 3 trimethyl lysine 27 associated with Il4 and Il13 gene silencing in Th1 cells. J Biol Chem 280(36):31470–31477

8. Chang S, Aune TM (2007) Dynamic changes in histone-methylation 'marks' across the locus encoding interferon-gamma during the differentiation of T helper type 2 cells. Nat Immunol 8(7):723–731
9. Wang J, Lee S, Teh CE, Bunting K, Ma L, Shannon MF (2009) The transcription repressor, ZEB1, cooperates with CtBP2 and HDAC1 to suppress IL-2 gene activation in T cells. Int Immunol 21(3):227–235
10. Enya K, Hayashi H, Takii T, Ohoka N, Kanata S, Okamoto T et al (2008) The interaction with Sp1 and reduction in the activity of histone deacetylase 1 are critical for the constitutive gene expression of IL-1 alpha in human melanoma cells. J Leukoc Biol 83(1):190–199
11. Lu J, Sun H, Wang X, Liu C, Xu X, Li F et al (2005) Interleukin-12 p40 promoter activity is regulated by the reversible acetylation mediated by HDAC1 and p300. Cytokine 31(1):46–51
12. Schmeck B, Lorenz J, N'Guessan PD, Opitz B, van Laak V, Zahlten J et al (2008) Histone acetylation and flagellin are essential for legionella pneumophila-induced cytokine expression. J Immunol 181(2):940–947
13. Greco L, Romino R, Coto I, Di Cosmo N, Percopo S, Maglio M et al (2002) The first large population based twin study of coeliac disease. Gut 50(5):624–628
14. Nistico L, Fagnani C, Coto I, Percopo S, Cotichini R, Limongelli MG et al (2006) Concordance, disease progression, and heritability of coeliac disease in Italian twins. Gut 55(6): 803–808
15. Aho K, Koskenvuo M, Tuominen J, Kaprio J (1986) Occurrence of rheumatoid arthritis in a nationwide series of twins. J Rheumatol 13(5):899–902
16. Silman AJ, MacGregor AJ, Thomson W, Holligan S, Carthy D, Farhan A et al (1993) Twin concordance rates for rheumatoid arthritis: results from a nationwide study. Br J Rheumatol 32(10):903–907
17. MacGregor AJ, Snieder H, Rigby AS, Koskenvuo M, Kaprio J, Aho K et al (2000) Characterizing the quantitative genetic contribution to rheumatoid arthritis using data from twins. Arthritis Rheum 43(1):30–37
18. Lefevre S, Knedla A, Tennie C, Kampmann A, Wunrau C, Dinser R et al (2009) Synovial fibroblasts spread rheumatoid arthritis to unaffected joints. Nat Med 15(12):1414–1420
19. Guillemin F, Briancon S, Klein JM, Sauleau E, Pourel J (1994) Low incidence of rheumatoid arthritis in France. Scand J Rheumatol 23(5):264–268
20. Doran MF, Pond GR, Crowson CS, O'Fallon WM, Gabriel SE (2002) Trends in incidence and mortality in rheumatoid arthritis in Rochester, Minnesota, over a forty-year period. Arthritis Rheum 46(3):625–631
21. Beighton P, Solomon L, Valkenburg HA (1975) Rheumatoid arthritis in a rural south African negro population. Ann Rheum Dis 34(2):136–141
22. Silman AJ, Ollier W, Holligan S, Birrell F, Adebajo A, Asuzu MC et al (1993) Absence of rheumatoid arthritis in a rural Nigerian population. J Rheumatol 20(4):618–622
23. Peschken CA, Esdaile JM (1999) Rheumatic diseases in North America's indigenous peoples. Semin Arthritis Rheum 28(6):368–391
24. Jacobsson LT, Hanson RL, Knowler WC, Pillemer S, Pettitt DJ, McCance DR et al (1994) Decreasing incidence and prevalence of rheumatoid arthritis in Pima Indians over a twenty-five-year period. Arthritis Rheum 37(8):1158–1165
25. Kaipiainen-Seppanen O, Aho K, Nikkarinen M (2001) Regional differences in the incidence of rheumatoid arthritis in Finland in 1995. Ann Rheum Dis 60(2):128–132
26. Shichikawa K, Inoue K, Hirota S, Maeda A, Ota H, Kimura M et al (1999) Changes in the incidence and prevalence of rheumatoid arthritis in Kamitonda, Wakayama, Japan, 1965–1996. Ann Rheum Dis 58(12):751–756
27. Wiles N, Symmons DP, Harrison B, Barrett E, Barrett JH, Scott DG et al (1999) Estimating the incidence of rheumatoid arthritis: trying to hit a moving target? Arthritis Rheum 42(7): 1339–1346
28. Barrett JH, Brennan P, Fiddler M, Silman A (2000) Breast-feeding and postpartum relapse in women with rheumatoid and inflammatory arthritis. Arthritis Rheum 43(5):1010–1015

29. Brennan P, Silman A (1994) Breast-feeding and the onset of rheumatoid arthritis. Arthritis Rheum 37(6):808–813
30. Jorgensen C, Picot MC, Bologna C, Sany J (1996) Oral contraception, parity, breast feeding, and severity of rheumatoid arthritis. Ann Rheum Dis 55(2):94–98
31. Karlson EW, Mandl LA, Hankinson SE, Grodstein F (2004) Do breast-feeding and other reproductive factors influence future risk of rheumatoid arthritis? Results from the nurses' health study. Arthritis Rheum 50(11):3458–3467
32. Linos A, Worthington JW, O'Fallon WM, Kurland LT (1983) Case-control study of rheumatoid arthritis and prior use of oral contraceptives. Lancet 1(8337):1299–1300
33. Vandenbroucke JP, Witteman JC, Valkenburg HA, Boersma JW, Cats A, Festen JJ et al (1986) Noncontraceptive hormones and rheumatoid arthritis in perimenopausal and postmenopausal women. JAMA 255(10):1299–1303
34. Turner S, Cherry N (2000) Rheumatoid arthritis in workers exposed to silica in the pottery industry. Occup Environ Med 57(7):443–447
35. Olsson AR, Skogh T, Axelson O, Wingren G (2004) Occupations and exposures in the work environment as determinants for rheumatoid arthritis. Occup Environ Med 61(3):233–238
36. Mousavi-Jazi M, Bostrom L, Lovmark C, Linde A, Brytting M, Sundqvist VA (1998) Infrequent detection of cytomegalovirus and Epstein-Barr virus DNA in synovial membrane of patients with rheumatoid arthritis. J Rheumatol 25(4):623–628
37. Niedobitek G, Lisner R, Swoboda B, Rooney N, Fassbender HG, Kirchner T et al (2000) Lack of evidence for an involvement of Epstein-Barr virus infection of synovial membranes in the pathogenesis of rheumatoid arthritis. Arthritis Rheum 43(1):151–154
38. Newkirk MM, Watanabe Duffy KN, Leclerc J, Lambert N, Shiroky JB (1994) Detection of cytomegalovirus, Epstein-Barr virus and herpes virus-6 in patients with rheumatoid arthritis with or without Sjogren's syndrome. Br J Rheumatol 33(4):317–322
39. Kouri T, Petersen J, Rhodes G, Aho K, Palosuo T, Heliovaara M et al (1990) Antibodies to synthetic peptides from Epstein-Barr nuclear antigen-1 in sera of patients with early rheumatoid arthritis and in preillness sera. J Rheumatol 17(11):1442–1449
40. Fox RI, Luppi M, Pisa P, Kang HI (1992) Potential role of Epstein-Barr virus in Sjogren's syndrome and rheumatoid arthritis. J Rheumatol Suppl 32:18–24
41. Toussirot E, Roudier J (2007) Pathophysiological links between rheumatoid arthritis and the Epstein-Barr virus: an update. Joint Bone Spine 74(5):418–426
42. Carty SM, Snowden N, Silman AJ (2004) Should infection still be considered as the most likely triggering factor for rheumatoid arthritis? Ann Rheum Dis 63(Suppl 2):ii46–ii49
43. Pattison DJ, Harrison RA, Symmons DP (2004) The role of diet in susceptibility to rheumatoid arthritis: a systematic review. J Rheumatol 31(7):1310–1319
44. Pattison DJ, Winyard PG (2008) Dietary antioxidants in inflammatory arthritis: do they have any role in etiology or therapy? Nat Clin Pract Rheumatol 4(11):590–596
45. Cerhan JR, Saag KG, Merlino LA, Mikuls TR, Criswell LA (2003) Antioxidant micronutrients and risk of rheumatoid arthritis in a cohort of older women. Am J Epidemiol 157(4):345–354
46. Heliovaara M, Knekt P, Aho K, Aaran RK, Alfthan G, Aromaa A (1994) Serum antioxidants and risk of rheumatoid arthritis. Ann Rheum Dis 53(1):51–53
47. Comstock GW, Burke AE, Hoffman SC, Helzlsouer KJ, Bendich A, Masi AT et al (1997) Serum concentrations of alpha tocopherol, beta carotene, and retinol preceding the diagnosis of rheumatoid arthritis and systemic lupus erythematosus. Ann Rheum Dis 56(5):323–325
48. Knekt P, Heliovaara M, Aho K, Alfthan G, Marniemi J, Aromaa A (2000) Serum selenium, serum alpha-tocopherol, and the risk of rheumatoid arthritis. Epidemiology 11(4):402–405
49. Costenbader KH, Kang JH, Karlson EW (2010) Antioxidant intake and risks of rheumatoid arthritis and systemic lupus erythematosus in women. Am J Epidemiol 172:205–216
50. Vessey MP, Villard-Mackintosh L, Yeates D (1987) Oral contraceptives, cigarette smoking and other factors in relation to arthritis. Contraception 35(5):457–464

51. Hernandez Avila M, Liang MH, Willett WC, Stampfer MJ, Colditz GA, Rosner B et al (1990) Reproductive factors, smoking, and the risk for rheumatoid arthritis. Epidemiology 1(4): 285–291
52. Criswell LA, Merlino LA, Cerhan JR, Mikuls TR, Mudano AS, Burma M et al (2002) Cigarette smoking and the risk of rheumatoid arthritis among postmenopausal women: results from the Iowa Women's health study. Am J Med 112(6):465–471
53. Stolt P, Bengtsson C, Nordmark B, Lindblad S, Lundberg I, Klareskog L et al (2003) Quantification of the influence of cigarette smoking on rheumatoid arthritis: results from a population based case-control study, using incident cases. Ann Rheum Dis 62(9):835–841
54. Huizinga TW, Amos CI, van der Helm-van Mil AH, Chen W, van Gaalen FA, Jawaheer D et al (2005) Refining the complex rheumatoid arthritis phenotype based on specificity of the HLA-DRB1 shared epitope for antibodies to citrullinated proteins. Arthritis Rheum 52(11): 3433–3438
55. Linn-Rasker SP, van der Helm-van Mil AH, van Gaalen FA, Kloppenburg M, de Vries RR, le Cessie S et al (2006) Smoking is a risk factor for anti-CCP antibodies only in rheumatoid arthritis patients who carry HLA-DRB1 shared epitope alleles. Ann Rheum Dis 65(3):366–371
56. Klareskog L, Stolt P, Lundberg K, Kallberg H, Bengtsson C, Grunewald J et al (2006) A new model for an etiology of rheumatoid arthritis: smoking may trigger HLA-DR (shared epitope)-restricted immune reactions to autoantigens modified by citrullination. Arthritis Rheum 54(1): 38–46
57. Liebold I, Grutzkau A, Gockeritz A, Gerl V, Lindquist R, Feist E et al (2021) Peripheral blood mononuclear cells are hypomethylated in active rheumatoid arthritis and methylation correlates with disease activity. Rheumatology (Oxford) 60(4):1984–1995
58. Schulz WA, Steinhoff C, Florl AR (2006) Methylation of endogenous human retroelements in health and disease. Curr Top Microbiol Immunol 310:211–250
59. Nagy G, Ward J, Mosser DD, Koncz A, Gergely P Jr, Stancato C et al (2006) Regulation of CD4 expression via recycling by HRES-1/RAB4 controls susceptibility to HIV infection. J Biol Chem 281(45):34574–34591
60. Hishikawa T, Ogasawara H, Kaneko H, Shirasawa T, Matsuura Y, Sekigawa I et al (1997) Detection of antibodies to a recombinant gag protein derived from human endogenous retrovirus clone 4-1 in autoimmune diseases. Viral Immunol 10(3):137–147
61. Piotrowski PC, Duriagin S, Jagodzinski PP (2005) Expression of human endogenous retrovirus clone 4-1 may correlate with blood plasma concentration of anti-U1 RNP and anti-Sm nuclear antibodies. Clin Rheumatol 24(6):620–624
62. Okada M, Ogasawara H, Kaneko H, Hishikawa T, Sekigawa I, Hashimoto H et al (2002) Role of DNA methylation in transcription of human endogenous retrovirus in the pathogenesis of systemic lupus erythematosus. J Rheumatol 29(8):1678–1682
63. Neidhart M, Rethage J, Kuchen S, Kunzler P, Crowl RM, Billingham ME et al (2000) Retrotransposable L1 elements expressed in rheumatoid arthritis synovial tissue: association with genomic DNA hypomethylation and influence on gene expression. Arthritis Rheum 43(12):2634–2647
64. Karouzakis E, Gay RE, Michel BA, Gay S, Neidhart M (2009) DNA hypomethylation in rheumatoid arthritis synovial fibroblasts. Arthritis Rheum 60(12):3613–3622
65. Martin-Trujillo A, van Rietschoten JG, Timmer TC, Rodriguez FM, Huizinga TW, Tak PP et al (2010) Loss of imprinting of IGF2 characterises high IGF2 mRNA-expressing type of fibroblast-like synoviocytes in rheumatoid arthritis. Ann Rheum Dis 69(6):1239–1242
66. Blunk I, Thomsen H, Reinsch N, Mayer M, Forsti A, Sundquist J et al (2020) Genomic imprinting analyses identify maternal effects as a cause of phenotypic variability in type 1 diabetes and rheumatoid arthritis. Sci Rep 10(1):11562
67. Zhou X, Chen W, Swartz MD, Lu Y, Yu R, Amos CI et al (2007) Joint linkage and imprinting analyses of GAW15 rheumatoid arthritis and gene expression data. BMC Proc 1(Suppl 1):S53
68. Nakano K, Whitaker JW, Boyle DL, Wang W, Firestein GS (2013) DNA methylome signature in rheumatoid arthritis. Ann Rheum Dis 72(1):110–117

69. Whitaker JW, Shoemaker R, Boyle DL, Hillman J, Anderson D, Wang W et al (2013) An imprinted rheumatoid arthritis methylome signature reflects pathogenic phenotype. Genome Med 5(4):40
70. Glossop JR, Emes RD, Nixon NB, Haworth KE, Packham JC, Dawes PT et al (2014) Genome-wide DNA methylation profiling in rheumatoid arthritis identifies disease-associated methylation changes that are distinct to individual T- and B-lymphocyte populations. Epigenetics 9(9): 1228–1237
71. Ai R, Whitaker JW, Boyle DL, Tak PP, Gerlag DM, Wang W et al (2015) DNA Methylome signature in Synoviocytes from patients with early rheumatoid arthritis compared to Synoviocytes from patients with longstanding rheumatoid arthritis. Arthritis Rheumatol 67(7):1978–1980
72. Karouzakis E, Raza K, Kolling C, Buckley CD, Gay S, Filer A et al (2018) Analysis of early changes in DNA methylation in synovial fibroblasts of RA patients before diagnosis. Sci Rep 8(1):7370
73. Karouzakis E, Rengel Y, Jungel A, Kolling C, Gay RE, Michel BA et al (2011) DNA methylation regulates the expression of CXCL12 in rheumatoid arthritis synovial fibroblasts. Genes Immun 12(8):643–652
74. Takami N, Osawa K, Miura Y, Komai K, Taniguchi M, Shiraishi M et al (2006) Hypermethylated promoter region of DR3, the death receptor 3 gene, in rheumatoid arthritis synovial cells. Arthritis Rheum 54(3):779–787
75. Chen Y, Gorelik GJ, Strickland FM, Richardson BC (2010) Decreased ERK and JNK signaling contribute to gene overexpression in "senescent" CD4+CD28- T cells through epigenetic mechanisms. J Leukoc Biol 87(1):137–145
76. Nile CJ, Read RC, Akil M, Duff GW, Wilson AG (2008) Methylation status of a single CpG site in the IL6 promoter is related to IL6 messenger RNA levels and rheumatoid arthritis. Arthritis Rheum 58(9):2686–2693
77. Liu Y, Aryee MJ, Padyukov L, Fallin MD, Hesselberg E, Runarsson A et al (2013) Epigenome-wide association data implicate DNA methylation as an intermediary of genetic risk in rheumatoid arthritis. Nat Biotechnol 31(2):142–147
78. Wong AH, Gottesman II, Petronis A (2005) Phenotypic differences in genetically identical organisms: the epigenetic perspective. Hum Mol Genet 14 Spec No 1:R11–8. https://doi.org/10.1093/hmg/ddi116
79. Enzer I, Dunn G, Jacobsson L, Bennett PH, Knowler WC, Silman A (2002) An epidemiologic study of trends in prevalence of rheumatoid factor seropositivity in Pima Indians: evidence of a decline due to both secular and birth-cohort influences. Arthritis Rheum 46(7):1729–1734
80. Jacobsson LT, Jacobsson ME, Askling J, Knowler WC (2003) Perinatal characteristics and risk of rheumatoid arthritis. BMJ 326(7398):1068–1069
81. Simard JF, Costenbader KH, Hernan MA, Liang MH, Mittleman MA, Karlson EW (2010) Early life factors and adult-onset rheumatoid arthritis. J Rheumatol 37(1):32–37
82. Edwards CJ, Goswami R, Goswami P, Syddall H, Dennison EM, Arden NK et al (2006) Growth and infectious exposure during infancy and the risk of rheumatoid factor in adult life. Ann Rheum Dis 65(3):401–404
83. Mandl LA, Costenbader KH, Simard JF, Karlson EW (2009) Is birthweight associated with risk of rheumatoid arthritis? Data from a large cohort study. Ann Rheum Dis 68(4):514–518
84. Li S, Wong EM, Bui M, Nguyen TL, Joo JE, Stone J et al (2018) Causal effect of smoking on DNA methylation in peripheral blood: a twin and family study. Clin Epigenetics 10:18
85. Jaakkola JJ, Gissler M (2005) Maternal smoking in pregnancy as a determinant of rheumatoid arthritis and other inflammatory polyarthropathies during the first 7 years of life. Int J Epidemiol 34(3):664–671
86. Strahl BD, Allis CD (2000) The language of covalent histone modifications. Nature 403(6765):41–45
87. Hu N, Qiu X, Luo Y, Yuan J, Li Y, Lei W et al (2008) Abnormal histone modification patterns in lupus CD4+ T cells. J Rheumatol 35(5):804–810

88. Huber LC, Brock M, Hemmatazad H, Giger OT, Moritz F, Trenkmann M et al (2007) Histone deacetylase/acetylase activity in total synovial tissue derived from rheumatoid arthritis and osteoarthritis patients. Arthritis Rheum 56(4):1087–1093
89. Kawabata T, Nishida K, Takasugi K, Ogawa H, Sada K, Kadota Y et al (2010) Increased activity and expression of histone deacetylase 1 in relation to tumor necrosis factor-alpha in synovial tissue of rheumatoid arthritis. Arthritis Res Ther 12(4):R133
90. Gillespie J, Savic S, Wong C, Hempshall A, Inman M, Emery P et al (2012) Histone deacetylases are dysregulated in rheumatoid arthritis and a novel histone deacetylase 3-selective inhibitor reduces interleukin-6 production by peripheral blood mononuclear cells from rheumatoid arthritis patients. Arthritis Rheum 64(2):418–422
91. Meng Q, Pan B, Sheng P (2021) Histone deacetylase 1 is increased in rheumatoid arthritis synovium and promotes synovial cell hyperplasia and synovial inflammation in the collagen-induced arthritis mouse model via the microRNA-124-dependent MARCKS-JAK/STAT axis. Clin Exp Rheumatol 39:970–981
92. Klein K, Frank-Bertoncelj M, Karouzakis E, Gay RE, Kolling C, Ciurea A et al (2017) The epigenetic architecture at gene promoters determines cell type-specific LPS tolerance. J Autoimmun 83:122–133
93. Sohn C, Lee A, Qiao Y, Loupasakis K, Ivashkiv LB, Kalliolias GD (2015) Prolonged tumor necrosis factor alpha primes fibroblast-like synoviocytes in a gene-specific manner by altering chromatin. Arthritis Rheumatol 67(1):86–95
94. Angiolilli C, Grabiec AM, Ferguson BS, Ospelt C, Malvar Fernandez B, van Es IE et al (2016) Inflammatory cytokines epigenetically regulate rheumatoid arthritis fibroblast-like synoviocyte activation by suppressing HDAC5 expression. Ann Rheum Dis 75(2):430–438
95. Angiolilli C, Kabala PA, Grabiec AM, Van Baarsen IM, Ferguson BS, Garcia S et al (2017) Histone deacetylase 3 regulates the inflammatory gene expression programme of rheumatoid arthritis fibroblast-like synoviocytes. Ann Rheum Dis 76(1):277–285
96. Muller-Ladner U, Gay RE, Gay S (2002) Role of nuclear factor kappaB in synovial inflammation. Curr Rheumatol Rep 4(3):201–207
97. Saccani S, Pantano S, Natoli G (2001) Two waves of nuclear factor kappaB recruitment to target promoters. J Exp Med 193(12):1351–1359
98. Saccani S, Pantano S, Natoli G (2002) p38-dependent marking of inflammatory genes for increased NF-kappa B recruitment. Nat Immunol 3(1):69–75
99. Saccani S, Natoli G (2002) Dynamic changes in histone H3 Lys 9 methylation occurring at tightly regulated inducible inflammatory genes. Genes Dev 16(17):2219–2224
100. Dieker J, Muller S (2010) Epigenetic histone code and autoimmunity. Clin Rev Allergy Immunol 39(1):78–84
101. Plaue S, Muller S, van Regenmortel MH (1989) A branched, synthetic octapeptide of ubiquitinated histone H2A as target of autoantibodies. J Exp Med 169(5):1607–1617
102. Reumaux D, Meziere C, Colombel JF, Duthilleul P, Mueller S (1995) Distinct production of autoantibodies to nuclear components in ulcerative colitis and in Crohn's disease. Clin Immunol Immunopathol 77(3):349–357
103. Kanai Y, Kawaminami Y, Miwa M, Matsushima T, Sugimura T (1977) Naturally-occurring antibodies to poly(ADP-ribose) in patients with systemic lupus erythematosus. Nature 265(5590):175–177
104. Sohn DH, Rhodes C, Onuma K, Zhao X, Sharpe O, Gazitt T et al (2015) Local joint inflammation and histone citrullination in a murine model of the transition from preclinical autoimmunity to inflammatory arthritis. Arthritis Rheumatol 67(11):2877–2887
105. Myzak MC, Tong P, Dashwood WM, Dashwood RH, Ho E (2007) Sulforaphane retards the growth of human PC-3 xenografts and inhibits HDAC activity in human subjects. Exp Biol Med (Maywood) 232(2):227–234
106. Dashwood RH, Ho E (2007) Dietary histone deacetylase inhibitors: from cells to mice to man. Semin Cancer Biol 17(5):363–369
107. Wutz A, Gribnau J (2007) X inactivation Xplained. Curr Opin Genet Dev 17(5):387–393

108. Carrel L, Willard HF (2005) X-inactivation profile reveals extensive variability in X-linked gene expression in females. Nature 434(7031):400–404
109. Lu Q, Wu A, Tesmer L, Ray D, Yousif N, Richardson B (2007) Demethylation of CD40LG on the inactive X in T cells from women with lupus. J Immunol 179(9):6352–6358
110. Ozbalkan Z, Bagislar S, Kiraz S, Akyerli CB, Ozer HT, Yavuz S et al (2005) Skewed X chromosome inactivation in blood cells of women with scleroderma. Arthritis Rheum 52(5): 1564–1570
111. Chabchoub G, Uz E, Maalej A, Mustafa CA, Rebai A, Mnif M et al (2009) Analysis of skewed X-chromosome inactivation in females with rheumatoid arthritis and autoimmune thyroid diseases. Arthritis Res Ther 11(4):R106

Asthma Epigenetics: Elucidating an Expanding Paradigm

16

Rachel L. Miller and James Chen

Contents

Abstract

This chapter will cover key concepts related to epigenetic regulation of asthma and the disease's environmental triggers, development, underlying immune pathways, and clinical course. Recent studies that include epigenetic mediation of the effects of exposures on risk factors on asthma will be stressed. Epigenetic regulation of asthma treatments and patient responses to interventions will be described. Among these topics, emphasis will be placed on the time windows of susceptibility. In general, this review suggests that despite continuing progress in this field, further rigorous studies of epigenetic mechanisms in asthma and its management are needed. Future studies may consider the temporality and

R. L. Miller (✉)
Division of Clinical Immunology, Department of Medicine, Icahn School of Medicine at Mount Sinai, New York, NY, USA
e-mail: rachel.miller2@mssm.edu

J. Chen
Harvard College, Cambridge, MA, USA
e-mail: jameschen@college.harvard.edu

K. B. Michels (ed.), *Epigenetic Epidemiology*,
https://doi.org/10.1007/978-3-030-94475-9_16

duration of epigenetic regulation to strengthen the evolving understandings of epigenetics in asthma.

Abbreviations

ACT	Asthma Control Test
AEC	Airway epithelial cells
ASMC	Airway smooth muscle cell
Aza	5-aza-2′-deoxycytidine
BBP	Butyl benzyl phthalate
BMI	Body mass index
CANDLE	Conditions Affecting Neurocognitive Development and Learning in Early Childhood Study
Der p	*Dermatophagoides pteronyssinus*
DMR	Differentially methylated region
EPIC	European Prospective Investigation into Cancer and Nutrition
eQTM	Expression Quantitative Trait Methylation
FEF	Forced expiratory flow
FeNO	Fractional exhaled nitric oxide
FEV_1	Forced expiratory volume in one second
FVC	Forced vital capacity
H3K27me3	H3 lysine 27 tri-methylation
IQR	Interquartile range
IT	Immunotherapy
LSC-R	Life Stressor Checklist-Revised survey
miRNA	MicroRNA
mMRC	Modified Medical Research Council
NEC	Nasal epithelial cells
OC	Organic carbon
PACE	Pregnancy and Childhood Epigenetics Consortium
PAH	Polycyclic aromatic hydrocarbon
PBMC	Peripheral blood mononuclear cell
PM_{10}	Coarse particulate matter
$PM_{2.5}$	Fine particulate matter
PSQ	Perceived stress questionnaire
PUFA	Polyunsaturated fatty acids
ROC	Receiver operating characteristic
scaRNA	Small Cajal body RNA
snoRNA	Small nucleolar RNA
SNP	Single-nucleotide polymorphism
Th	T-helper
WGCNA	Weighted gene correlation network analysis
α-SMA	α-smooth muscle actin

16.1 Introduction

Asthma is a complex disorder whose mechanistic underpinnings continue to be elucidated. Estimates indicate that non-genetic factors, such as environmental exposures, may contribute to approximately 60% of these mechanisms [1]. The immunological pathways that are induced and associated with the development of asthma and features of its clinical course (i.e., asthma exacerbations) are varied and include the stimulation of T helper (Th) 2 proallergic pathways, altered T regulation, and airway smooth muscle functional impairments. They all have been shown to be regulated by multiple environmental exposures, such as cigarette smoke, fungi, house dust mites, and bisphenol A [2–7], suggesting that the impact of the environment on asthma immunopathogenesis is substantial.

Epigenetic regulation, briefly defined as involving mechanisms that are potentially heritable modifications that may influence gene expression without alterations in the DNA sequence, provides a fundamental framework for considering asthma immunopathogenesis for many reasons. Both epigenetic regulation and the asthma phenotype change over time and with aging [8]. Others have shown that changing levels of environmental exposures were associated with persistent vs. improved asthma symptomatology. These are best represented in randomized trials of environmental remediation on asthma-related outcomes [9]. Hence, environmental epigenetic regulation may explain the temporal relation between the changing environment exposures and changing asthma symptoms [10].

This chapter will focus on delineating the complex relation between environmental exposures and the impact of epigenetic regulation on asthma. This will include a review of the role for epigenetic regulation in environmental inducers of asthma, as well as in asthma immune pathways and clinical features of asthma. Evidence for epigenetic regulation of asthma treatments also will be reviewed. We will conclude with itemizing challenges and opportunities for future investigations to fill in the many research gaps in this expanding paradigm.

16.2 Asthma Environmental Triggers Appear Epigenetically Regulated

Traffic-related air pollution describes a wide array of emissions, including gases, coarse and fine particulate matter (PM_{10}, $PM_{2.5}$), and diesel, all of which have been implicated in asthma. While some studies have attributed epigenetic effects of measures of traffic as a surrogate for exposure to air pollution, both prenatally [11] and during childhood [12], others have determined associations with specific pollutants. These include several studies that relied on maternal residential addresses to estimate prenatal levels. For example, Abraham and coworkers examined prenatal NO_2 exposure using atmospheric dispersion modeling. DNA methylation in placental samples collected at delivery showed that higher estimated exposure to NO_2 through pregnancy, and particularly during trimesters 1 and 2, was associated with lower methylation at two CpG sites in the developmental gene Adenosine A2b

Receptor (*ADORA2B*) [13]. Further, Ladd-Acosta and coworkers used air quality monitoring data and spatial modeling to estimate NO_2 and ozone exposures. Higher levels of prenatal NO_2 or O_3 exposure were associated with alterations in nine differentially methylated regions (DMRs), including in genes related to vaccination response and substrate metabolism in cord blood samples, as well as mitochondrial function and gastrointestinal inflammation in placental samples [14]. Combined, these studies argue for the importance of the prenatal time window of exposure on NO_2 and ozone-induced epigenetic regulation of genes involved in multiple metabolic, developmental, and inflammatory processes in cord blood and the placenta.

Nonetheless, additional works point to the importance of exposures to NO_2 and ozone during adulthood. These include epidemiological studies in humans, using estimates of long-term exposure to NO_2 from land-use regression models, that found evidence of global demethylation on somatic chromosomes and several functional regions such as CpG shores and shelves in Italian and Dutch adults participating in the European Prospective Investigation into Cancer and Nutrition (EPIC) cohort [15]; many were associated with altered immune regulatory gene transcription. Bind and coworkers exposed healthy adult study participants to either clean air or ozone for two two-hour exposures and observed differences in bronchial epithelial cell DNA methylation of 19 CpG sites in the Phospholipid Scramblase 1 (*PLSCR1*), Hydroxycarboxylic Acid Receptor 1 (*HCAR1*), and Long Intergenic Non-Protein Coding RNA 336 (*LINC00336*) genes by experimental arm [16], indicating that short-term exposures to ozone also can induce epigenetic effects.

Particulate matter has emerged as another pollutant that appears to induce epigenetic alterations, including during the prenatal window. This has been found for global demethylation and mitochondrial DNA methylation measured in placental samples [17, 18]. In a meta-analysis of a subset of participants in the Pregnancy and Childhood Epigenetics consortium (PACE), prenatal exposure to $PM_{2.5}$ and PM_{10} was estimated using dispersion modeling at the maternal addresses and was associated with differential methylation at 14 and 6 CpG sites in cord blood samples, respectively. Four of the CpGs associated with prenatal PM_{10} exposure, including in pulmonary-related genes, remained associated in peripheral blood samples measured in a separate cohort of older children [19], suggesting that specific epigenetic differences related to prenatal PM exposures may persist later into childhood. Associations between $PM_{2.5}$ exposures during adulthood and epigenetic alterations similarly suggest that periods beyond pregnancy may be susceptible to epigenetic regulation. These include reports from adult men enrolled in the US Department of Veterans Affairs Normative Aging Study. Long-term residential $PM_{2.5}$ exposures over one year, estimated through modeling based on satellite-derived aerosol measurements and land-use regression variables, were associated with 20 DNA-methylation-age-related CpGs of 20 different genes measured in blood cells [20]. Studies of more acute exposures, including those measured from monitoring sites over 2, 7, and 28 days in one study [21], and from personal backpacks over one day [22], and 14-day averages of PM_{10} collected from an air quality monitoring station [23], reported associations with altered levels of DNA methylation in multiple CpG sites and DMRs and altered histone modifications in adult

peripheral blood. While these studies suggest that epigenetic regulation can occur over the short term, they do not investigate whether these alterations are durable or clinically impactful.

Beyond traffic-related air pollution, exposure to pollution from secondhand cigarette smoke also has shown evidence of epigenetic regulation. Christensen and coworkers first exposed pregnant mice, daily throughout gestation, to either filtered air or tobacco smoke. Then, they experimentally induced dust mite allergic inflammation in a subset of both exposure groups. This process yielded four exposure categories: controls, single exposure to either tobacco smoke or house dust mite (tobacco smoke-only and house-dust-mite-only, respectively), or combined-exposure to both (combined-exposure). Tobacco smoke-only mice showed lower methylation of *IL-13* as compared to controls. Combined-exposure mice had lower methylation in *IL-4* and *IL-13*, as well as higher *IFN-γ* methylation, in comparisons with both tobacco smoke-only and house-dust-mite-only mice [24], suggesting incrementally further epigenetic regulation following two experimental exposures. In a subset of cohorts from the PACE consortium, prenatal exposure to cigarette smoke throughout pregnancy was measured via questionnaires. Through meta-analysis, prenatal smoke exposure was associated with 568 CpGs in cord blood samples that remained significant after Bonferroni multiple test correction [25]. This study highlighted associations between prenatal tobacco smoke and epigenetic differences, including many in developmental pathways, further emphasizing the importance of the prenatal window of exposure. In addition, the top 1511 prenatal smoking-related CpGs identified by Joubert and coworkers, including those with some nominal associations, were further examined by Rauschert and coworkers. They measured methylation at those prenatally identified CpG sites in three cohorts of one-time adolescent and adult blood collections and then combined that methylation data with self-reported smoking collected from the participants' mothers during pregnancy. They assessed, trained, and tested previously established machine learning algorithms for their accuracy in predicting past prenatal smoke exposure based on methylation in the adolescent and adult blood samples; their final, best-performing score utilized 204 CpGs to make this prediction [26], suggesting that a subset of prenatal smoking-related epigenetic marks identified by Joubert and colleagues may generalize to other cohorts of older individuals.

In addition to physical exposures to pollutants, exposures to the asthma trigger of stress also have shown epigenetic associations in epidemiological studies. Brunst and coworkers assessed maternal lifetime exposure to stress and potentially traumatic events using the Life Stressor Checklist-Revised (LSC-R) survey. A cumulative score of all included stressful events, weighted by self-reported negative impact of each event, was calculated and associated with placental methylation of 12 CpG sites, including sites involved in lysine degradation and fetal development [27]. These findings suggest that exposures that may occur even prior to the prenatal window of susceptibility window may affect methylation, although this study design did not isolate that time period.

As environmental exposures do not occur in isolation, emerging studies have sought to test the epigenetic impacts of combined or sequential environmental

exposures. In one example, Clifford and coworkers examined sequential exposures to diesel exhaust, a major component of PM emission mixtures, and allergens. They performed a crossover experiment with 2 hours of randomly assigned filtered air or diesel exhaust exposure. Subsequently, the adult participants were administered an allergen challenge in a lower lobe bronchial segment, paired with a saline control that was instilled in a contralateral bronchial segment. After a four-week washout period, the process was repeated with reversed exposures and reversed sides. Bronchial epithelial samples were collected 48 hours after each allergen challenge for measures of DNA methylation. They found that seven CpGs changed methylation levels following diesel exposure alone and co-exposure to diesel and allergen after the initial exposures. Allergen challenge four weeks after diesel exposure changed methylation in 75 CpG sites, and diesel challenge four weeks after allergen exposure changed methylation in 548 CpG sites [28]. These results suggest that sequential and differing toxicants may have synergistic epigenetic impacts, with levels that vary depending on the timing and order of the sequence.

16.3 Asthma and its Underlying Immune Pathways Appear Epigenetically Regulated

Asthma is a heterogeneous disorder characterized by different underlying immunological pathways, some of which are still being defined. Accordingly, epigenetic research has focused both on agnostic approaches using high output methylation arrays and more directed investigation of epigenetic regulation of known asthma pathways. Both strategies have yielded substantial and complementary insights into asthma mechanisms. While epigenetic studies in blood yielded some early evidence of epigenetic regulation in multiple genes and pathways related to T cell function, NK cell function, and eosinophil activation [29–31], more recent studies suggest that epigenetic modifications in other tissues or cells closer to the airways are very relevant, and may be more relevant, to asthma and its mechanistic underpinnings, such as sputum, nasal epithelial, and airway epithelial cells, as described below.

Groth and coworkers found physician-diagnosed asthma-associated DMRs in immune-related, small nucleolar RNA (snoRNA), and small Cajal body RNA (scaRNA) genes in sputum in their agnostic approach. They further utilized computer algorithms (i.e., deconvolution processing) for statistically inferring cell-type-specific methylation profiles given the mixed cellular profile of sputum. The more robust findings were specific among macrophages, as opposed to neutrophils or eosinophils, where DMRs in asthma samples, as compared to non-asthmatic samples, were located within the *IL23A* and *CCL24* genes; these two loci relate to macrophage polarization and eosinophil stimulation, respectively [32].

Nasal epithelial cells have been studied increasingly due to their relative noninvasive collection techniques and even closer proximity to the lower airway. These include epigenome-wide studies in children that appear to have implicated pathways or genes epigenetically regulated in asthma. In several, differential methylation of genes and pathways, including those involved in eosinophilic activity, allergic

responses, and oxidative stress, were cross-sectionally associated with asthma diagnosis and medication use [33–35]. Further, Forno and coworkers compared pediatric seroatopic vs. non-seroatopic physician-diagnosed asthmatics, reporting 8644 CpGs that were cross-sectionally differentially methylated [36]. These included genes important to epithelial barrier function, gap junction signaling, and type-1 and type-2 helper T cell (Th1/Th2) immune responses. Agnostic approaches also have examined epigenetic influences on gene transcription, using expression quantitative trait methylation (eQTM) analysis to probe more than 8.5 million pairs of methylation loci and gene expression of nearby genes. In nasal cells from pediatric physician-diagnosed atopic asthmatics, Kim and coworkers found 16,867 eQTM pairs, based on 9103 CpGs associated with the expression of 3512 genes, that were differentially identified in atopic asthma patients as compared with non-atopic controls [37]. While a relation of methylation regulating gene expression is well-established, this methodology that measures sites that demonstrate both altered methylation and gene expression pinpoints potentially influential loci with more specificity than other studies that rely on one molecular platform alone.

Airway epithelial cells (AECs) have been assessed in a limited number of studies and yielded new insights into their epigenetic regulation in asthma [38]. In one example, Nicodemus-Johnson used an array-based approach to assess genome-wide AEC methylation in adult asthma, defined based on doctor's diagnosis and current medication use. They found 40,892 CpG sites that were cross-sectionally differentially methylated in asthmatics as compared to non-asthmatics. Further, by clustering CpG sites with highly correlated methylation levels using weighted gene correlation network analysis (WGCNA), they reported that the differentially methylated loci were enriched in pathways related to airway remodeling, leukocyte attraction, and response to nitric oxide [39].

Studies that have targeted loci in established asthma pathways include those centered on the regulation of allergic sensitization and allergic inflammation. These too have occurred across tissue types. For example, Barton and colleagues targeted analyses of the proallergic Th2 transcription factor GATA Binding Protein 3 (*GATA3*). Higher cord blood methylation at birth in two *GATA3* CpGs were associated prospectively with reduced risk of physician-diagnosed asthma at ages 3 and 6 years [40]. Further studies have utilized buccal cells, easily and noninvasively accessible and in proximity to the airways, to examine cytokines that regulate IgE synthesis, such as IL-4. Jung and coworkers observed buccal *IL-4* promoter hypomethylation 5 days following higher levels of short-term black carbon exposures, measured using a backpack-based device over two 24-hour periods [41]. These studies collectively indicate that key regulatory components of the Th2 proallergic pathway may be epigenetically influenced, albeit in various tissue types with different apparent time courses of epigenetic regulation. Nicodemus-Johnson and coworkers also studied AECs to elucidate the epigenetic impact of allergic immune responses induced by the Th2 cytokine: IL-13. First, they treated cells ex vivo with either 10 ng/mL IL-13 or vehicle (controls) for 24 hours. Using an array-based approach, they reported 6522 CpG sites that were differentially methylated with the IL-13-exposed cells, as compared to the controls. They

subsequently compared these *ex vivo* loci to *in vivo* methylation in AEC samples collected from patients during bronchoscopy; 2020 of the 6522 IL-13-responsive CpGs from their cell culture model also were differentially methylated in samples from asthmatics vs. non-asthmatics [42]. The authors interpreted these results to indicate the biological relevance of their cell culture model of allergic immune responses in demonstrating and discovering methylation patterns that contribute to asthma pathogenesis.

The T regulatory arm of adaptive immunity, often signaled by altered activity of the transcription factor Forkhead Box P3 (*FOXP3*), is another asthma pathway that appears epigenetically regulated. For example, experiments in mice performed by Kyburz and coworkers compared the epigenetic regulation of protection from perinatal and postnatal *Helicobacter pylori* exposure following house dust mite-induced allergic airway inflammation. Maternal *H. pylori* exposure, orally administered twice-weekly through pregnancy, was associated with lower methylation of the Treg-specific demethylated region in the *FOXP3* locus in the offspring [43], demonstrating *H. pylori*-induced epigenetic alterations in the regulatory T pathway. The Infant Immune Study cohort, which includes children of asthmatic mothers, identified 589 cord blood DMRs that distinguished children who did and did not develop asthma later in childhood. In particular, *SMAD3* promoter hypermethylation, paired with IL-1β production (both measured in cord blood), was associated prospectively with childhood risk of asthma by age 9. Both *SMAD3* hypermethylation and IL-1β production have been implicated in destabilization of T regulation and greater Th17 differentiation [44]. As such, these results affirm the prenatal window of epigenetic susceptibility on asthma risk as it applies to T regulatory pathways.

Finally, pathways related to airway smooth muscle function may be epigenetically regulated in asthma. Yu and coworkers developed a house dust mite-induced mouse model of allergic airway inflammation with airway remodeling and described lower H3 lysine 27 tri-methylation (H3K27me3) levels in lung tissues from sensitized mice vs. controls. They further examined the effects of GSK-J4, a selective inhibitor of H3K27me3, and found that sensitized mice that were pre-treated with GSK-J4 showed lower levels of collagen deposition in the lungs and lower α-smooth muscle actin (α-SMA) levels in the peribronchial area, as compared to sensitized mice without the pre-treatment [45]. As collagen deposition and higher α-SMA expression are both features of asthmatic airway remodeling, these results elucidate some epigenetic underpinnings of airway smooth muscle phenotypes in asthma. Perry and coworkers reported 12,383 CpGs in adult airway smooth muscle cells (ASMCs) that were differentially methylated among non-severe asthmatics, severe asthmatics, and healthy controls. These sites have established functions in cell proliferation, apoptosis, and ASMC contraction [46], further linking epigenetic regulation presumably to altered airway smooth function. MicroRNAs (miRNAs), including let-7a and miRNA-744, appear cross-sectionally differentially expressed in airway smooth muscle in diagnosed asthmatic ASMCs, as compared to non-asthmatic samples [47]. Finally, one small study of severe asthmatic vs. normal control samples extended this differential expression of miRNA-744 to bronchial

epithelial cells. Overexpression of miRNA-744 in asthmatics reduced bronchial epithelial cell proliferation, apparently through phosphorylation of *SMAD3* and modulation of *TGF-β1* [48], suggesting that the airway smooth muscle cells may not be the only airway cells susceptible to epigenetic regulation.

16.4 Clinical Features of Asthma Appear Epigenetically Regulated

Epigenetic regulation may underlie diverse clinical presentations of asthma. For example, among pediatric sibling pairs discordant for asthma symptoms, determined based on a questionnaire and using an array-based approach, six CpG sites in NECs were associated cross-sectionally with current asthma status [49], including loci related to the gene regulation-involved *TET1* gene, the lung function-related *LAMA5* gene, and the inflammation-associated *NLRP3* gene. All three have been implicated previously in asthma development through Th2 proallergic or airway remodeling pathways. Other studies suggest that asthma exacerbations may be epigenetically regulated. In one example, Wardzyńska and coworkers studied blood serum from adults who were admitted for an asthma exacerbation. They longitudinally compared expression levels of seven selected miRNAs at admission with measurements in samples collected at a subsequent follow-up visit, when patients largely showed improvements in clinical symptoms evaluated by the Asthma Control Test (ACT) and modified Medical Research Council (mMRC) score. They reported higher expression of miRNA-126a, miRNA-16, and miRNA-21 during the initial exacerbation visit, as compared to expression levels measured at the follow-up visit [50]. These findings include some miRNAs that have been associated previously with asthma and suggest the reversal of a potentially adverse epigenetic pattern in patients' recoveries following asthma exacerbation.

Further studies have examined the epigenetic underpinning of impaired lung function and airway inflammation. The Wardzyńska study assessed two airway parameters at both patient visit timepoints: forced expiratory volume in one second (FEV_1) and fractional exhaled nitric oxide (FeNO), the latter as a biomarker for airway inflammation. They observed lower expression of miRNA-21 in subjects concurrently with lower FEV_1 at the initial exacerbation visit. Additionally, they cross-sectionally found higher miRNA-126a expression measured at follow-up among patients whose FeNO levels improved between the two visits [50]. In another pediatric study, meta-analyzing data across five cohorts, den Dekker and coworkers found 22, 15, and 22 cord blood DMRs that were associated prospectively and inversely with indicators of obstructive lung function measures among children ages 7–13: FEV_1, FEV_1/forced vital capacity (FVC), and forced expiratory flow at 75% volume ($FEF_{75\%}$), respectively [51]. Epigenetic mechanisms of trajectories of lung function development from pre-adolescence to adulthood also have been studied, such as by Sunny and coworkers. They measured peripheral blood methylation at age 10 and assessed trajectories of FVC, FEV_1, and FEV_1/FVC ratio at ages 10, 18, and 26 that were classified as "high" and "low;" They reported 96 CpGs that were

associated prospectively with lung function trajectories [52]. In another study, Wu and coworkers examined methylation of the Integrin β4 (*ITGB4*) gene that has been implicated previously in asthma pathogenesis in response to oxidative stress or stimulation by inflammatory mediators [53]. Among adult subjects, they identified two peripheral blood CpG methylation levels in *ITGB4* that were cross-sectionally and negatively correlated with lung function parameters, including FEV_1, FEV_1/FVC, and $FEF_{25\%}$. Additionally, based on analysis of the area under the curve for receiver operating characteristic (ROC) curves, they reported that methylation levels at these two loci showed high positive predictive value for identifying patients with physician-diagnosed current asthma [54], although its utility as a screening tool for asthma needs to be tested more rigorously in future studies.

16.5 Assessing Environmental Exposures and Asthma via Epigenetic Regulation

Emerging research has focused more on the pathway from environmental exposure to asthma-related outcomes via epigenetic regulation (Fig. 16.1). These include studies that have addressed multiple components of this paradigm in the same cohorts. For example, Prunicki and coworkers examined exposure to $PM_{2.5}$, NO_2, and CO in children and epigenetic associations, the former measured using air quality monitoring stations. Physician-diagnosed asthma was associated cross-sectionally with higher peripheral blood methylation in the *FOXP3* promoter, as

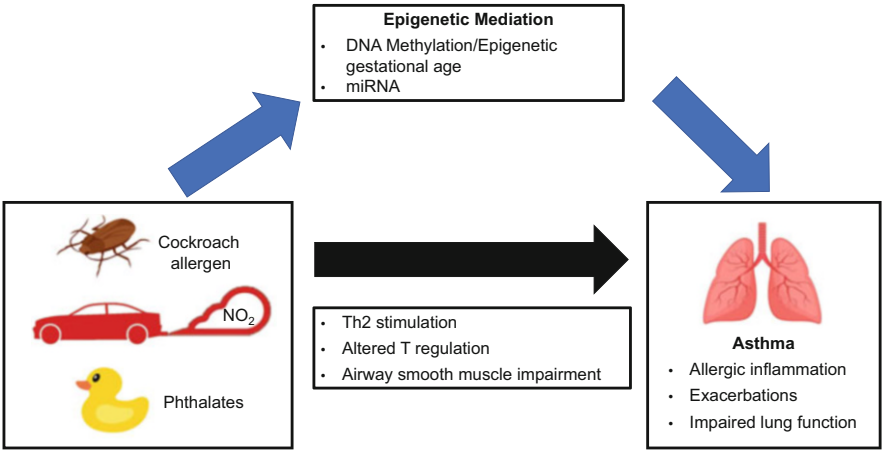

Fig. 16.1 Emerging studies suggest environmental exposures induce asthma-related outcomes via epigenetic regulation. Many epidemiological and experimental studies to date have demonstrated that environmental toxicants induce asthma or exacerbate asthma symptoms through pathways such as proallergic Th2 stimulation, altered T regulation, and airway smooth muscle impairment. New studies suggest that may occur by epigenetic regulation, including altered DNA methylation and induction of miRNA

compared to the absence of asthma. In the same study cohort, 90-day ambient exposures to $PM_{2.5}$, NO_2, and CO were associated positively with *FOXP3* promoter methylation [55]. In a similar approach, Somineni and coworkers examined epigenetic regulation by the modulator enzyme Ten-Eleven Translocation 1 (*TET1*), which helps regulate allergic airway inflammation through the interferon signaling and aryl hydrocarbon receptor pathways [56]. They cross-sectionally found lower methylation of cg23602092 in the *TET1* promoter in NECs from children with reported physician-diagnosed asthma. Meanwhile, higher current exposure to elemental carbon (a proxy for traffic-related diesel particles that was estimated using land-use regression models) was associated with higher NEC methylation of that CpG site in the non-asthmatics, but not the asthmatics [57], paradoxically uncovering opposite directionality of the methylation related to asthma and to exposure. In adults, further studies have examined associations of epigenetic markers with airway inflammation. Zhang and coworkers measured buccal methylation in the arginase-nitric oxide synthase pathway from four measures across 4½ months among nonsmoking healthy college students in response to acute measures of 24-hour $PM_{2.5}$ constituents (measured at an air quality monitoring station). Methylation levels also were compared with FeNO. Multiple $PM_{2.5}$ constituents (i.e., organic carbon (OC), EC, K+, Si, K, Fe, Zn, Ba, Cr, Se, and Pb) were associated with higher levels of FeNO measured one day later. Concurrently, in the single-constituent model, OC, EC, NO3−, Mg2+, K, Fe, Mn, As, Cr, and Se were associated with lower *NOS2A* methylation. Also, concurrently, an interquartile range (IQR) increase in OC, K+, Si, K, Fe, Zn, Mn, Ba, Cu, As, Cr, Se, and Pb was associated with increases in *ARG2* methylation [58]. Combined, these results show associations of acute measures of $PM_{2.5}$ individual constituents with both buccal methylation in the arginase-nitric oxide synthase pathway and FeNO, although they too did not analyze whether these methylation levels mediated the effects of exposure on airway inflammation.

In addition to air pollutants, exposure to stress has been examined jointly in relation to both asthma and epigenetic changes. Trump and coworkers reported that higher maternal stress during pregnancy, determined using a perceived stress questionnaire (PSQ), was associated prospectively with increased risk of persistent childhood wheeze. Additionally, by comparing mother-child pairs with low and high maternal stress levels, they reported 2306 stress-dependent cord blood DMRs in newborns and 2495 venous blood DMRs in the mothers collected at the 36th week of gestation, with limited overlap between the two sets. Further, in children who subsequently and persistently wheezed, they reported stress-dependent differential blood mRNA expression at age four years in downstream targets of several DMRs identified in newborns. These occurred predominantly among lung maturation-related calcium—and Wnt-signaling pathways [59], suggesting a mechanistic epigenetic link between higher prenatal stress childhood wheeze.

This paradigm has been tested following combined or sequential exposures. In children ages 9–14 years, our group assessed physical activity across six days with an accelerometer and personal exposures to black carbon through two 24-hour periods with a vest-based personal air sampling device. Higher levels of activity

were associated concurrently with lower *FOXP3* promoter methylation in the subgroup of children exposed to high levels of black carbon. *FOXP3* promoter methylation also was associated inversely with obstructive lung function indicators reduced, such as FEV_1/FVC and $FEF_{25–75\%}$ [60]. These findings indicate that physical activity and black carbon exposure may have additive effects on methylation, with corresponding effects on lung function. In another study, Rider and coworkers sequentially exposed adults to diesel exhaust and allergens. Their methodology was identical to those utilized by Clifford and coworkers, described in an earlier section [28]. Allergen exposure alone modulated the expression of five miRNAs, including miRNA-183-5p that targets Forkhead Box O1 (*FOXO1*), an oxidative stress-related transcription factor. The authors, however, did not observe effects from the sequential diesel and allergen exposures [61].

Emerging studies have relied on experimental approaches or analyses for statistical mediation to help determine more directly whether environmentally induced asthma outcomes occur via epigenetic regulation. In one study, Jahreis and coworkers focused on exposures to phthalates, a group of chemicals found in a wide range of consumer products, including butyl benzyl phthalate (BBP). In an ovalbumin-induced mouse model of allergic airway inflammation, they administered BBP to the mothers through their drinking water. Prenatal BBP exposure was associated with a higher number of eosinophils measured in the bronchoalveolar lavage and greater lung resistance, and splenic T cell genome-wide DNA hypermethylation. To demonstrate that epigenetic regulation was the mechanism underlying the phthalate exposure effects on allergic airway inflammation, DNA methyltransferase inhibitor 5-aza-2′-deoxycytidine (Aza) was administered for two weeks to a cohort of one-week-old offspring. Although methylation was not remeasured following administration of Aza, the Aza-treated mice showed lower BBP-induced bronchoalveolar lavage eosinophils and attenuated lung resistance that was attributed to BBP-induced DNA methylation among the untreated mice [62]. In another example, Qiu and coworkers developed a cockroach allergen-induced asthma model by sensitizing mice for four days by intratracheal inhalation of cockroach extract followed by airway challenge to cockroach extract days later. Compared to wild-type sensitized mice, sensitized miRNA-155-knockout mice showed lower lung tissue expression levels of Cyclo-Oxygenase Isoenzyme 2 (*COX-2*), as well as lower eosinophil, IL-4, and IL-13 levels in bronchoalveolar fluid. Further transfection of miRNA-155-knockout mice with an adeno-associated virus-carrying miRNA-155 reversed those phenotypical changes, as compared to sensitized, knockout mice that received a mock adeno-associated virus [63]. Together, these suggest that miRNA-155 regulates cockroach allergen-induced Th2-associated airway inflammation, possibly by modifying the expression of the *COX-2* gene.

In one epidemiological study, Sbihi and coworkers observed that greater differences between children's chronological age and a calculated epigenetic gestational age, determined using cord blood DNA methylation, correlated with higher risk of allergic sensitization in young children following prenatal exposure to NO_2. The latter exposure was measured by land-use regression models at the maternal

residences. Subsequently, the study utilized counterfactual framework-based mediation analysis to dissect the total effect of the NO_2 exposure on allergic sensitization into direct and indirect effects. Through this approach, 31% of the association between prenatal NO_2 exposure and risk of allergic sensitization was mediated by differences in epigenetic gestational age [64]. Wang and coworkers examined phthalate exposure, approximated using levels of the phthalate metabolite mono (2-ethyl-5-hydroxyhexyl)phthalate (5OH-MEHP) in urine, among children. From a targeted assessment of 21 candidate genes in whole blood, *TNF-α* methylation was associated inversely with 5OH-MEHP levels. Further, both higher 5OH-MEHP concentrations and lower *TNF-α* methylation were associated cross-sectionally with physician-diagnosed asthma at age 3 years. Subsequently, they utilized a similar analysis strategy as Sbihi and coworkers and calculated that 20% of the phthalate exposure effect on asthma was mediated by *TNF- α* methylation [65]. In the adult LifeLines cohort study, ambient NO_2 exposures estimated using land-use regression models were associated with methylation in seven CpG sites across seven genes. In addition, NO_2 exposures were associated cross-sectionally with lower FVC and higher FEV_1/FVC ratio. In mediation analyses, two of the seven most significantly differentially methylated CpGs following NO_2 exposure mediated 27 to 31% of the association between exposure and FVC and FEV_1/FVC [66], suggesting that a small number of methylation sites may modify the relation between NO_2 exposure and lung function.

Epigenome-wide analyses found effect modification of the relation between body mass index (BMI), a measure of obesity important to the pathogenesis of some asthma phenotypes, and physician-diagnosed adult-onset asthma by DNA methylation of genes in glucocorticoid signaling pathways and multiple signal transduction pathways [67]. However, mediation of the relation BMI and lung function was not observed when analyzed by epigenetic Mendelian randomization. Some relation of BMI on peripheral blood DNA methylation was observed at select CpG sites, but a casual effect of DNA methylation at such sites on lung function parameters was not identified across four population-based cohorts [68]. Hence, the number of studies demonstrating epigenetic mediation or effect modification of exposures or asthma risk factors and asthma-related outcomes remains limited and a substantial research gap.

16.6 Asthma Treatments and Responses Appear Epigenetically Regulated

So far, a few established and supplemental treatments for asthma have shown indications of epigenetic regulation. In a small study, Wang and coworkers compared methylation patterns from allergic asthmatic children following immunotherapy (IT) against the dust mite *Dermatophagoides pteronyssinus* (*Der p*) with dust mite allergic children who were not treated with IT and nonallergic controls. They identified 108 peripheral blood DMRs that were differentially methylated between the allergic asthmatics according to IT treatment, including in genes related to

extracellular matrix remodeling, oxidative stress pathways, and airway inflammation [69]. While small and possibly experiencing bias related to indication for prescribing IT, the study does hint that IT may have potential to alter epigenetic regulation in allergic asthmatics. Other studies have queried whether asthma treatments used as supplements, including vitamins and polyunsaturated fatty acids (PUFAs), may induce epigenetic alterations. For example, Fussbroich and coworkers assessed this in a mouse model of dust mite-induced allergic inflammation. 62 miRNAs exhibited differential expression levels in sensitized mice as compared to non-sensitized control mice. Long chain PUFA supplementation for 24 days (although not compared with vehicle-only negative control) diminished the expression of eight of those dysregulated miRNAs to levels observed among the non-sensitized mice. Six of those miRNAs were functionally implicated in airway remodeling-related pathways (TGF-β signaling, extracellular matrix-receptor interaction) as well as fatty acid synthesis [70]. In one randomized, double-blinded, placebo-controlled intervention study, Shorey-Kendrick and coworkers established three groups of pregnant women at mid-gestation: a non-smoker control group, a group of mothers who were unwilling/unable to stop smoking during the pregnancy and were randomized to receive a daily 500-mg vitamin C oral supplement until delivery, and a group of smoking mothers who were randomized to receive a placebo. They measured methylation in the placenta and cord blood samples collected at delivery, as well as buccal cells collected from the children between ages 3 and 6 years. Overall, from 200,000 CpGs, the study identified 1408 with at least 10% methylation difference in the children based on whether the mother smoked during pregnancy. With prenatal vitamin C supplementation, 69% of those identified CpG sites reverted to levels 50% closer to those measured among children from the non-smoker controls [71]. As such, these investigators presented one of the first lines of evidence that a highly accessible early intervention can correct epigenetic consequences of an environmental toxicant associated with asthma. In another, Zhong and coworkers assigned each adult on three different days, separated by four weeks, to a two-hour filtered air exposure and then two two-hour $PM_{2.5}$ exposure. In the four weeks between the first two exposures, the participants received a daily oral placebo, and between the latter two exposures, the participants received a daily oral 53.5-mg vitamin B supplement containing folic acid, vitamin B6 and vitamin B12. Peripheral T-helper cells showed evidence of differential DNA methylation following the first $PM_{2.5}$ vs. filtered air exposure, although they did not enumerate the total number of differentially methylated loci due to the limited statistical power of their study. Nonetheless, at the top ten exposure-related differentially methylated loci, vitamin B supplements attenuated the $PM_{2.5}$ effect by 28–76% [72]. Although their study was performed in a small number of healthy adults, the results suggest another low-risk and low-cost supplement may epigenetically modify the effects of an asthma-related environmental exposure.

Other studies focused on epigenetic regulation of clinical responses to asthma treatments. Xiao and coworkers, for example, assessed children who presented to the emergency room with asthma exacerbations and recorded gene expression prior to and 18–24 hours following treatment with corticosteroids. Eight genes appeared to

be differentially expressed according to treatment response. Additional experiments focused on *Vanin-1* gene expression; this alteration was accompanied by differences in methylation at one of the 5 CpG sites tested that varied according to corticosteroid treatment response [73]. Subsequent studies by Zhang and coworkers utilized an array-based approach to examine genome-wide DNA methylation at the same two timepoints among 20 children. The more robust findings were measured at the second timepoint, following a process that removed differentially methylated CpGs where single nucleotide polymorphisms may be present in the array probes. Then they performed hierarchical clustering and assessment of the biological function of the genes near the CpG sites to yield 32 CpG sites that separated the two responder groups. Two sites (cg17187762 and cg00802903) reached genome-wide statistical significance, suggestive of a small epigenetic signal in association with improvement following corticosteroid treatment [74]. In another study, Wang and coworkers reported, among children with mild-to-moderate persistent asthma, that 545 whole blood CpG sites were cross-sectionally associated with improved FEV_1 eight weeks after initiation of inhaled corticosteroid treatment [75].

16.7 Research Gaps in Asthma Epigenetics

While notable progress has been made in elucidating epigenetic underpinnings of asthma, its environmental triggers, its clinical course, and its treatment, many gaps in knowledge remain. One is the temporality of epigenetic regulation in asthma, particularly across the life course. One of the first and few studies to address this, by Michel and coworkers, collected both cord blood samples at birth and whole blood samples at age 4.5 years. Measuring methylation in 23 regions of ten candidate genes associated with asthma, Th2 and T regulatory pathways, they reported changes in methylation between the two sampling timepoints in 15 of those regions. Those differences were clustered particularly in the asthma-associated *ORMDL* family of genes and the Th2-related *RAD50*, *IL-13*, and *IL-4* genes [76]. Few other studies have reported such repeat measures. Another under-investigated area is the durability of epigenetic regulation. Some intriguing studies have demonstrated this by examining epigenetic measures across generations, mostly in rodent models. Gregory and coworkers exposed pregnant female mice at gestational days 14–15 to a single intranasal instillation of diesel exhaust particles. This one-time prenatal diesel exhaust exposure was associated with differential dendritic cell methylation at 14,480 CpGs in the F_1, 9413 CpGs in F_2, and 6239 CpGs in F_3 generation; of these CpGs, 402 were altered in all three generations. In addition, F_2 and F_3 offspring from the diesel exhaust-exposed lineage showed higher levels of eosinophils in bronchoalveolar fluid, compared to unexposed controls, consistent with an airway allergic inflammation that persisted across generations. This was prevented by intraperitoneal treatment of a subset of F_1 females with a DNA methyltransferase inhibitor 3 days prior to mating, implicating broad DNA methylation as the mechanism that underlies the persistent allergic phenotype [77]. Still, many questions remain regarding the different potential roles on temporality and duration of

epigenetic regulation by paternal vs. maternal prenatal and even pre-conception exposures.

While AECs and other cells derived from the lungs are presumed to inform most accurately on epigenetic regulation in asthma, few studies have directly compared epigenetic measures across cell types. This gap was tackled in part by Brugha and coworkers who collected peripheral blood mononuclear cells (PBMCs), buccal, nasal, and airway epithelial samples from children ages five to 13, half of whom had physician-diagnosed asthma. Using hierarchical clustering of the array-based methylation measurements across each of the different samples, they reported among both asthmatics and healthy controls that methylation profiles in NECs were most similar to those of AECs. The profiles in buccal cells were moderately similar, and those in blood were the least similar to those measured in AECs [78]. Combined, these results highlight the importance of sampling as close to the respiratory tract as possible to assess airway regulation. In another, Lin and coworkers obtained AEC measures from adults, and NEC and PBMC from children, and relied on cluster analyses to compare which classified better with physician-diagnosed asthma. These showed that both AEC and NEC data, compared to PBMC data, better-classified patients with asthma and with a similar degree of accuracy, highlighting the potential of the more accessible tissues from the nasal epithelium [79].

Epigenetic regulation likely diverges across ancestries; few studies have considered this sufficiently. Moen and coworkers used an array-based approach to examine methylation in lymphoblastoid cell lines derived from individuals of European or African descent, and they found that 36,597 CpGs (13% of the loci assessed) showed differential methylation between the two populations [80]. Mozhui and coworkers similarly reported ancestry-related methylation differences in a subset of African American and European American neonates from the Conditions Affecting Neurocognitive Development and Learning in Early Childhood (CANDLE) Study cohort. 3802 CpGs (>18% of loci assessed) showed differential methylation between the two ancestry groups [81]. Barfield and coworkers sought to address these epigenetic divergences through computational approaches. First, from African American and Caucasian adults, they found 12,827 whole blood CpG sites that were nominally differentially methylated based on race. They adjusted for confounding in the methylation-association tests due to population stratification by utilizing principal component analysis to identify covariates that considered genome-wide SNP data and information on the proximity of SNP variants to the measured methylation probes. Their adjustment methods led to virtually total elimination of the differential methylation previously observed [82], suggesting that analysis of ancestry may impact measures of epigenetic regulation.

Finally, as well-delineated by Breton and colleagues, most studies of epigenetic regulation in environmental exposures and disease have demonstrated small effect sizes thus far, potentially challenging the functional importance of measured epigenetic differences [83]. As reported in the aforementioned Barfield study, some of the differences have been reported as nominal. These studies highlight the vastness of epigenetic regulation and remind us that differences in the epigenome, transcriptome, and proteome, as well as other metabolic processes, likely interact

to exert effects on asthma. Few studies to date have the scope to consider these interactions.

In summary, the future of research into asthma epigenetics has the potential to elucidate novel mechanisms underlying the disease, including those induced by environmental exposures, and may eventually more accurately pinpoint diagnosis and treatment of this heterogeneous condition. Further refinement and progression of current methods—including those that consider cell specificity, changes over time in exposures and their effects, and molecular interactions—can strengthen the clinical relevance and robustness of these future studies.

References

1. Rappaport SM (2016) Genetic factors are not the major causes of chronic diseases. PLoS One 11(4):e0154387. https://doi.org/10.1371/journal.pone.0154387
2. Aravamudan B, Thompson M, Sieck GC, Vassallo R, Pabelick CM, Prakash YS (2017) Functional effects of cigarette smoke-induced changes in airway smooth muscle mitochondrial morphology. J Cell Physiol 232(5):1053–1068. https://doi.org/10.1002/jcp.25508
3. Cayrol C, Duval A, Schmitt P, Roga S, Camus M, Stella A, Burlet-Schiltz O, Gonzalez-de-Peredo A, Girard JP (2018) Environmental allergens induce allergic inflammation through proteolytic maturation of IL-33. Nat Immunol 19(4):375–385. https://doi.org/10.1038/s41590-018-0067-5
4. Cho YJ, Park SB, Park JW, Oh SR, Han M (2018) Bisphenol A modulates inflammation and proliferation pathway in human endometrial stromal cells by inducing oxidative stress. Reprod Toxicol 81:41–49. https://doi.org/10.1016/j.reprotox.2018.06.016
5. Hrusch CL, Stein MM, Gozdz J, Holbreich M, von Mutius E, Vercelli D, Ober C, Sperling AI (2019) T-cell phenotypes are associated with serum IgE levels in Amish and Hutterite children. J Allergy Clin Immunol 144(5):1391–1401. https://doi.org/10.1016/j.jaci.2019.07.034
6. Munakata S, Ishimori K, Kitamura N, Ishikawa S, Takanami Y, Ito S (2018) Oxidative stress responses in human bronchial epithelial cells exposed to cigarette smoke and vapor from tobacco- and nicotine-containing products. Regul Toxicol Pharmacol 99:122–128. https://doi.org/10.1016/j.yrtph.2018.09.009
7. Szoka P, Lachowicz J, Cwiklińska M, Lukaszewicz A, Rybak A, Baranowska U, Holownia A (2019) Cigarette smoke-induced oxidative stress and autophagy in human alveolar epithelial cell line (A549 cells). Adv Exp Med Biol 1176:63–69. https://doi.org/10.1007/5584_2019_373
8. Schoos A-MM, Jelding-Dannemand E, Stokholm J, Bønnelykke K, Bisgaard H, Chawes BL (2019) Single and multiple time-point allergic sensitization during childhood and risk of asthma by age 13. Pediatr Allergy Immunol 30(7):716–723. https://doi.org/10.1111/pai.13109
9. Matsui EC, Perzanowski M, Peng RD, Wise RA, Balcer-Whaley S, Newman M, Cunningham A, Divjan A, Bollinger ME, Zhai S, Chew G, Miller RL, Phipatanakul W (2017) Effect of an integrated pest management intervention on asthma symptoms among mouse-sensitized children and adolescents with asthma: a randomized clinical trial. JAMA 317(10):1027–1036. https://doi.org/10.1001/jama.2016.21048
10. Wu X, Huang Q, Javed R, Zhong J, Gao H, Liang H (2019) Effect of tobacco smoking on the epigenetic age of human respiratory organs. Clin Epigenetics 11(1):183. https://doi.org/10.1186/s13148-019-0777-z
11. Kingsley SL, Eliot MN, Whitsel EA, Huang YT, Kelsey KT, Marsit CJ, Wellenius GA (2016) Maternal residential proximity to major roadways, birth weight, and placental DNA methylation. Environ Int 92–93:43–49. https://doi.org/10.1016/j.envint.2016.03.020
12. Commodore A, Mukherjee N, Chung D, Svendsen E, Vena J, Pearce J, Roberts J, Arshad SH, Karmaus W (2018) Frequency of heavy vehicle traffic and association with DNA methylation at

age 18 years in a subset of the Isle of Wight birth cohort. Environ Epigenet 4(4):dvy028. https://doi.org/10.1093/eep/dvy028
13. Abraham E, Rousseaux S, Agier L, Giorgis-Allemand L, Tost J, Galineau J, Hulin A, Siroux V, Vaiman D, Charles MA, Heude B, Forhan A, Schwartz J, Chuffart F, Bourova-Flin E, Khochbin S, Slama R, Lepeule J (2018) Pregnancy exposure to atmospheric pollution and meteorological conditions and placental DNA methylation. Environ Int 118:334–347. https://doi.org/10.1016/j.envint.2018.05.007
14. Ladd-Acosta C, Feinberg JI, Brown SC, Lurmann FW, Croen LA, Hertz-Picciotto I, Newschaffer CJ, Feinberg AP, Fallin MD, Volk HE (2019) Epigenetic marks of prenatal air pollution exposure found in multiple tissues relevant for child health. Environ Int 126:363–376. https://doi.org/10.1016/j.envint.2019.02.028
15. Plusquin M, Guida F, Polidoro S, Vermeulen R, Raaschou-Nielsen O, Campanella G, Hoek G, Kyrtopoulos SA, Georgiadis P, Naccarati A, Sacerdote C, Krogh V, Bas Bueno-de-Mesquita H, Monique Verschuren WM, Sayols-Baixeras S, Panni T, Peters A, Hebels D, Kleinjans J, Vineis P, Chadeau-Hyam M (2017) DNA methylation and exposure to ambient air pollution in two prospective cohorts. Environ Int 108:127–136. https://doi.org/10.1016/j.envint.2017.08.006
16. Bind MC, Rubin DB, Cardenas A, Dhingra R, Ward-Caviness C, Liu Z, Mirowsky J, Schwartz JD, Diaz-Sanchez D, Devlin RB (2020) Heterogeneous ozone effects on the DNA methylome of bronchial cells observed in a crossover study. Sci Rep 10(1):15739. https://doi.org/10.1038/s41598-020-72068-6
17. Maghbooli Z, Hossein-Nezhad A, Adabi E, Asadollah-Pour E, Sadeghi M, Mohammad-Nabi S, Zakeri Rad L, Malek Hosseini AA, Radmehr M, Faghihi F, Aghaei A, Omidifar A, Aghababei Y, Behzadi H (2018) Air pollution during pregnancy and placental adaptation in the levels of global DNA methylation. PLoS One 13(7):e0199772. https://doi.org/10.1371/journal.pone.0199772
18. Vos S, Nawrot TS, Martens DS, Byun HM, Janssen BG (2020) Mitochondrial DNA methylation in placental tissue: a proof of concept study by means of prenatal environmental stressors. Epigenetics:1–11. https://doi.org/10.1080/15592294.2020.1790923
19. Gruzieva O, Xu CJ, Yousefi P, Relton C, Merid SK, Breton CV, Gao L, Volk HE, Feinberg JI, Ladd-Acosta C, Bakulski K, Auffray C, Lemonnier N, Plusquin M, Ghantous A, Herceg Z, Nawrot TS, Pizzi C, Richiardi L, Rusconi F, Vineis P, Kogevinas M, Felix JF, Duijts L, den Dekker HT, Jaddoe VWV, Ruiz JL, Bustamante M, Antó JM, Sunyer J, Vrijheid M, Gutzkow KB, Grazuleviciene R, Hernandez-Ferrer C, Annesi-Maesano I, Lepeule J, Bousquet J, Bergström A, Kull I, Söderhäll C, Kere J, Gehring U, Brunekreef B, Just AC, Wright RJ, Peng C, Gold DR, Kloog I, DeMeo DL, Pershagen G, Koppelman GH, London SJ, Baccarelli AA, Melén E (2019) Prenatal particulate air pollution and DNA methylation in newborns: an epigenome-wide meta-analysis. Environ Health Perspect 127(5):57012. https://doi.org/10.1289/ehp4522
20. Nwanaji-Enwerem JC, Colicino E, Trevisi L, Kloog I, Just AC, Shen J, Brennan K, Dereix A, Hou L, Vokonas P, Schwartz J, Baccarelli AA (2016) Long-term ambient particle exposures and blood DNA methylation age: findings from the VA normative aging study. Environ Epigenet 2(2). https://doi.org/10.1093/eep/dvw006
21. Panni T, Mehta AJ, Schwartz JD, Baccarelli AA, Just AC, Wolf K, Wahl S, Cyrys J, Kunze S, Strauch K, Waldenberger M, Peters A (2016) Genome-wide analysis of DNA methylation and fine particulate matter air pollution in three study populations: KORA F3, KORA F4, and the Normative Aging Study. Environ Health Perspect 124(7):983–990. https://doi.org/10.1289/ehp.1509966
22. Mostafavi N, Vermeulen R, Ghantous A, Hoek G, Probst-Hensch N, Herceg Z, Tarallo S, Naccarati A, Kleinjans JCS, Imboden M, Jeong A, Morley D, Amaral AFS, van Nunen E, Gulliver J, Chadeau-Hyam M, Vineis P, Vlaanderen J (2018) Acute changes in DNA methylation in relation to 24 h personal air pollution exposure measurements: a panel study in four European countries. Environ Int 120:11–21. https://doi.org/10.1016/j.envint.2018.07.026

23. Zheng Y, Sanchez-Guerra M, Zhang Z, Joyce BT, Zhong J, Kresovich JK, Liu L, Zhang W, Gao T, Chang D, Osorio-Yanez C, Carmona JJ, Wang S, McCracken JP, Zhang X, Chervona Y, Díaz A, Bertazzi PA, Koutrakis P, Kang CM, Schwartz J, Baccarelli AA, Hou L (2017) Traffic-derived particulate matter exposure and histone H3 modification: a repeated measures study. Environ Res 153:112–119. https://doi.org/10.1016/j.envres.2016.11.015
24. Christensen S, Jaffar Z, Cole E, Porter V, Ferrini M, Postma B, Pinkerton KE, Yang M, Kim YJ, Montrose L, Roberts K, Holian A, Cho YH (2017) Prenatal environmental tobacco smoke exposure increases allergic asthma risk with methylation changes in mice. Environ Mol Mutagen 58(6):423–433. https://doi.org/10.1002/em.22097
25. Joubert BR, Felix JF, Yousefi P, Bakulski KM, Just AC, Breton C, Reese SE, Markunas CA, Richmond RC, Xu CJ, Küpers LK, Oh SS, Hoyo C, Gruzieva O, Söderhäll C, Salas LA, Baïz N, Zhang H, Lepeule J, Ruiz C, Ligthart S, Wang T, Taylor JA, Duijts L, Sharp GC, Jankipersadsing SA, Nilsen RM, Vaez A, Fallin MD, Hu D, Litonjua AA, Fuemmeler BF, Huen K, Kere J, Kull I, Munthe-Kaas MC, Gehring U, Bustamante M, Saurel-Coubizolles MJ, Quraishi BM, Ren J, Tost J, Gonzalez JR, Peters MJ, Håberg SE, Xu Z, van Meurs JB, Gaunt TR, Kerkhof M, Corpeleijn E, Feinberg AP, Eng C, Baccarelli AA, Benjamin Neelon SE, Bradman A, Merid SK, Bergström A, Herceg Z, Hernandez-Vargas H, Brunekreef B, Pinart M, Heude B, Ewart S, Yao J, Lemonnier N, Franco OH, Wu MC, Hofman A, McArdle W, Van der Vlies P, Falahi F, Gillman MW, Barcellos LF, Kumar A, Wickman M, Guerra S, Charles MA, Holloway J, Auffray C, Tiemeier HW, Smith GD, Postma D, Hivert MF, Eskenazi B, Vrijheid M, Arshad H, Antó JM, Dehghan A, Karmaus W, Annesi-Maesano I, Sunyer J, Ghantous A, Pershagen G, Holland N, Murphy SK, DeMeo DL, Burchard EG, Ladd-Acosta C, Snieder H, Nystad W, Koppelman GH, Relton CL, Jaddoe VW, Wilcox A, Melén E, London SJ (2016) DNA methylation in newborns and maternal smoking in pregnancy: genome-wide consortium meta-analysis. Am J Hum Genet 98(4):680–696. https://doi.org/10.1016/j.ajhg.2016.02.019
26. Rauschert S, Melton PE, Heiskala A, Karhunen V, Burdge G, Craig JM, Godfrey KM, Lillycrop K, Mori TA, Beilin LJ, Oddy WH, Pennell C, Järvelin MR, Sebert S, Huang RC (2020) Machine learning-based DNA methylation score for fetal exposure to maternal smoking: development and validation in samples collected from adolescents and adults. Environ Health Perspect 128(9):97003. https://doi.org/10.1289/ehp6076
27. Brunst KJ, Tignor N, Just A, Liu Z, Lin X, Hacker MR, Bosquet Enlow M, Wright RO, Wang P, Baccarelli AA, Wright RJ (2018) Cumulative lifetime maternal stress and epigenome-wide placental DNA methylation in the PRISM cohort. Epigenetics 13(6):665–681. https://doi.org/10.1080/15592294.2018.1497387
28. Clifford RL, Jones MJ, MacIsaac JL, McEwen LM, Goodman SJ, Mostafavi S, Kobor MS, Carlsten C (2017) Inhalation of diesel exhaust and allergen alters human bronchial epithelium DNA methylation. J Allergy Clin Immunol 139(1):112–121. https://doi.org/10.1016/j.jaci.2016.03.046
29. Reese SE, Xu CJ, den Dekker HT, Lee MK, Sikdar S, Ruiz-Arenas C, Merid SK, Rezwan FI, Page CM, Ullemar V, Melton PE, Oh SS, Yang IV, Burrows K, Söderhäll C, Jima DD, Gao L, Arathimos R, Küpers LK, Wielscher M, Rzehak P, Lahti J, Laprise C, Madore AM, Ward J, Bennett BD, Wang T, Bell DA, Vonk JM, Håberg SE, Zhao S, Karlsson R, Hollams E, Hu D, Richards AJ, Bergström A, Sharp GC, Felix JF, Bustamante M, Gruzieva O, Maguire RL, Gilliland F, Baïz N, Nohr EA, Corpeleijn E, Sebert S, Karmaus W, Grote V, Kajantie E, Magnus MC, Örtqvist AK, Eng C, Liu AH, Kull I, Jaddoe VWV, Sunyer J, Kere J, Hoyo C, Annesi-Maesano I, Arshad SH, Koletzko B, Brunekreef B, Binder EB, Räikkönen K, Reischl E, Holloway JW, Jarvelin MR, Snieder H, Kazmi N, Breton CV, Murphy SK, Pershagen G, Anto JM, Relton CL, Schwartz DA, Burchard EG, Huang RC, Nystad W, Almqvist C, Henderson AJ, Melén E, Duijts L, Koppelman GH, London SJ (2019) Epigenome-wide meta-analysis of DNA methylation and childhood asthma. J Allergy Clin Immunol 143(6):2062–2074. https://doi.org/10.1016/j.jaci.2018.11.043

30. Xu CJ, Söderhäll C, Bustamante M, Baïz N, Gruzieva O, Gehring U, Mason D, Chatzi L, Basterrechea M, Llop S, Torrent M, Forastiere F, Fantini MP, Carlsen KCL, Haahtela T, Morin A, Kerkhof M, Merid SK, van Rijkom B, Jankipersadsing SA, Bonder MJ, Ballereau S, Vermeulen CJ, Aguirre-Gamboa R, de Jongste JC, Smit HA, Kumar A, Pershagen G, Guerra S, Garcia-Aymerich J, Greco D, Reinius L, McEachan RRC, Azad R, Hovland V, Mowinckel P, Alenius H, Fyhrquist N, Lemonnier N, Pellet J, Auffray C, van der Vlies P, van Diemen CC, Li Y, Wijmenga C, Netea MG, Moffatt MF, Cookson W, Anto JM, Bousquet J, Laatikainen T, Laprise C, Carlsen KH, Gori D, Porta D, Iñiguez C, Bilbao JR, Kogevinas M, Wright J, Brunekreef B, Kere J, Nawijn MC, Annesi-Maesano I, Sunyer J, Melén E, Koppelman GH (2018) DNA methylation in childhood asthma: an epigenome-wide meta-analysis. Lancet Respir Med 6(5):379–388. https://doi.org/10.1016/s2213-2600(18)30052-3
31. Hoang TT, Sikdar S, Xu CJ, Lee MK, Cardwell J, Forno E, Imboden M, Jeong A, Madore AM, Qi C, Wang T, Bennett BD, Ward JM, Parks CG, Beane-Freeman LE, King D, Motsinger-Reif A, Umbach DM, Wyss AB, Schwartz DA, Celedón JC, Laprise C, Ober C, Probst-Hensch N, Yang IV, Koppelman GH, London SJ (2020) Epigenome-wide association study of DNA methylation and adult asthma in the Agricultural Lung Health Study. Eur Respir J 56(3). https://doi.org/10.1183/13993003.00217-2020
32. Groth EE, Weber M, Bahmer T, Pedersen F, Kirsten A, Börnigen D, Rabe KF, Watz H, Ammerpohl O, Goldmann T (2020) Exploration of the sputum methylome and omics deconvolution by quadratic programming in molecular profiling of asthma and COPD: the road to sputum omics 2.0. Respir Res 21(1):274. https://doi.org/10.1186/s12931-020-01544-4
33. Cardenas A, Sordillo JE, Rifas-Shiman SL, Chung W, Liang L, Coull BA, Hivert MF, Lai PS, Forno E, Celedón JC, Litonjua AA, Brennan KJ, DeMeo DL, Baccarelli AA, Oken E, Gold DR (2019) The nasal methylome as a biomarker of asthma and airway inflammation in children. Nat Commun 10(1):3095. https://doi.org/10.1038/s41467-019-11058-3
34. Qi C, Jiang Y, Yang IV, Forno E, Wang T, Vonk JM, Gehring U, Smit HA, Milanzi EB, Carpaij OA, Berg M, Hesse L, Brouwer S, Cardwell J, Vermeulen CJ, Acosta-Pérez E, Canino G, Boutaoui N, van den Berge M, Teichmann SA, Nawijn MC, Chen W, Celedón JC, Xu CJ, Koppelman GH (2020) Nasal DNA methylation profiling of asthma and rhinitis. J Allergy Clin Immunol 145(6):1655–1663. https://doi.org/10.1016/j.jaci.2019.12.911
35. Yang IV, Pedersen BS, Liu AH, O'Connor GT, Pillai D, Kattan M, Misiak RT, Gruchalla R, Szefler SJ, Khurana Hershey GK, Kercsmar C, Richards A, Stevens AD, Kolakowski CA, Makhija M, Sorkness CA, Krouse RZ, Visness C, Davidson EJ, Hennessy CE, Martin RJ, Togias A, Busse WW, Schwartz DA (2017) The nasal methylome and childhood atopic asthma. J Allergy Clin Immunol 139(5):1478–1488. https://doi.org/10.1016/j.jaci.2016.07.036
36. Forno E, Wang T, Qi C, Yan Q, Xu CJ, Boutaoui N, Han YY, Weeks DE, Jiang Y, Rosser F, Vonk JM, Brouwer S, Acosta-Perez E, Colón-Semidey A, Alvarez M, Canino G, Koppelman GH, Chen W, Celedón JC (2019) DNA methylation in nasal epithelium, atopy, and atopic asthma in children: a genome-wide study. Lancet Respir Med 7(4):336–346. https://doi.org/10.1016/s2213-2600(18)30466-1
37. Kim S, Forno E, Zhang R, Park HJ, Xu Z, Yan Q, Boutaoui N, Acosta-Pérez E, Canino G, Chen W, Celedón JC (2020) Expression quantitative trait methylation analysis reveals methylomic associations with gene expression in childhood asthma. Chest 158(5):1841–1856. https://doi.org/10.1016/j.chest.2020.05.601
38. Stefanowicz D, Ullah J, Lee K, Shaheen F, Olumese E, Fishbane N, Koo HK, Hallstrand TS, Knight DA, Hackett TL (2017) Epigenetic modifying enzyme expression in asthmatic airway epithelial cells and fibroblasts. BMC Pulm Med 17(1):24. https://doi.org/10.1186/s12890-017-0371-0
39. Nicodemus-Johnson J, Myers RA, Sakabe NJ, Sobreira DR, Hogarth DK, Naureckas ET, Sperling AI, Solway J, White SR, Nobrega MA, Nicolae DL, Gilad Y, Ober C (2016) DNA methylation in lung cells is associated with asthma endotypes and genetic risk. JCI Insight 1(20):e90151. https://doi.org/10.1172/jci.insight.90151

40. Barton SJ, Ngo S, Costello P, Garratt E, El-Heis S, Antoun E, Clarke-Harris R, Murray R, Bhatt T, Burdge G, Cooper C, Inskip H, van der Beek EM, Sheppard A, Godfrey KM, Lillycrop KA (2017) DNA methylation of Th2 lineage determination genes at birth is associated with allergic outcomes in childhood. Clin Exp Allergy 47(12):1599–1608. https://doi.org/10.1111/cea.12988
41. Jung KH, Lovinsky-Desir S, Yan B, Torrone D, Lawrence J, Jezioro JR, Perzanowski M, Perera FP, Chillrud SN, Miller RL (2017) Effect of personal exposure to black carbon on changes in allergic asthma gene methylation measured 5 days later in urban children: importance of allergic sensitization. Clin Epigenetics 9:61. https://doi.org/10.1186/s13148-017-0361-3
42. Nicodemus-Johnson J, Naughton KA, Sudi J, Hogarth K, Naurekas ET, Nicolae DL, Sperling AI, Solway J, White SR, Ober C (2016) Genome-wide methylation study identifies an IL-13-induced epigenetic signature in asthmatic airways. Am J Respir Crit Care Med 193(4):376–385. https://doi.org/10.1164/rccm.201506-1243OC
43. Kyburz A, Fallegger A, Zhang X, Altobelli A, Artola-Boran M, Borbet T, Urban S, Paul P, Münz C, Floess S, Huehn J, Cover TL, Blaser MJ, Taube C, Müller A (2019) Transmaternal Helicobacter pylori exposure reduces allergic airway inflammation in offspring through regulatory T cells. J Allergy Clin Immunol 143(4):1496–1512. https://doi.org/10.1016/j.jaci.2018.07.046
44. DeVries A, Wlasiuk G, Miller SJ, Bosco A, Stern DA, Lohman IC, Rothers J, Jones AC, Nicodemus-Johnson J, Vasquez MM, Curtin JA, Simpson A, Custovic A, Jackson DJ, Gern JE, Lemanske RF Jr, Guerra S, Wright AL, Ober C, Halonen M, Vercelli D (2017) Epigenome-wide analysis links SMAD3 methylation at birth to asthma in children of asthmatic mothers. J Allergy Clin Immunol 140(2):534–542. https://doi.org/10.1016/j.jaci.2016.10.041
45. Yu Q, Yu X, Zhao W, Zhu M, Wang Z, Zhang J, Huang M, Zeng X (2018) Inhibition of H3K27me3 demethylases attenuates asthma by reversing the shift in airway smooth muscle phenotype. Clin Exp Allergy 48(11):1439–1452. https://doi.org/10.1111/cea.13244
46. Perry MM, Lavender P, Kuo CS, Galea F, Michaeloudes C, Flanagan JM, Fan Chung K, Adcock IM (2018) DNA methylation modules in airway smooth muscle are associated with asthma severity. Eur Respir J 51(4). https://doi.org/10.1183/13993003.01068-2017
47. Chen Y, Qiao L, Zhang Z, Hu G, Zhang J, Li H (2019) Let-7a inhibits proliferation and promotes apoptosis of human asthmatic airway smooth muscle cells. Exp Ther Med 17(5):3327–3334. https://doi.org/10.3892/etm.2019.7363
48. Huang H, Lu H, Liang L, Zhi Y, Huo B, Wu L, Xu L, Shen Z (2019) MicroRNA-744 inhibits proliferation of bronchial epithelial cells by regulating SMAD3 pathway via targeting Transforming Growth Factor-β1 (TGF-β1) in severe asthma. Med Sci Monit 25:2159–2168. https://doi.org/10.12659/msm.912412
49. Zhang X, Biagini Myers JM, Burleson JD, Ulm A, Bryan KS, Chen X, Weirauch MT, Baker TA, Butsch Kovacic MS, Ji H (2018) Nasal DNA methylation is associated with childhood asthma. Epigenomics 10(5):629–641. https://doi.org/10.2217/epi-2017-0127
50. Wardzyńska A, Pawełczyk M, Rywaniak J, Kurowski M, Makowska JS, Kowalski ML (2020) Circulating microRNAs and T-cell cytokine expression are associated with the characteristics of asthma exacerbation. Allergy, Asthma Immunol Res 12(1):125–136. https://doi.org/10.4168/aair.2020.12.1.125
51. den Dekker HT, Burrows K, Felix JF, Salas LA, Nedeljkovic I, Yao J, Rifas-Shiman SL, Ruiz-Arenas C, Amin N, Bustamante M, DeMeo DL, Henderson AJ, Howe CG, Hivert MF, Ikram MA, de Jongste JC, Lahousse L, Mandaviya PR, van Meurs JB, Pinart M, Sharp GC, Stolk L, Uitterlinden AG, Anto JM, Litonjua AA, Breton CV, Brusselle GG, Sunyer J, Smith GD, Relton CL, Jaddoe VWV, Duijts L (2019) Newborn DNA-methylation, childhood lung function, and the risks of asthma and COPD across the life course. Eur Respir J 53(4). https://doi.org/10.1183/13993003.01795-2018
52. Sunny SK, Zhang H, Mzayek F, Relton CL, Ring S, Henderson AJ, Ewart S, Holloway JW, Arshad SH (2021) Pre-adolescence DNA methylation is associated with lung function

trajectories from pre-adolescence to adulthood. Clin Epigenetics 13(1):5. https://doi.org/10.1186/s13148-020-00992-5
53. Yuan L, Du X, Tang S, Wu S, Wang L, Xiang Y, Qu X, Liu H, Qin X, Liu C (2019) ITGB4 deficiency induces senescence of airway epithelial cells through p53 activation. FEBS J 286(6):1191–1203. https://doi.org/10.1111/febs.14749
54. Wu M, Yang Y, Yuan L, Yang M, Wang L, Du X, Qin L, Wu S, Xiang Y, Qu X, Liu H, Qin X, Liu C (2020) DNA methylation down-regulates integrin β4 expression in asthmatic airway epithelial cells. Clin Exp Allergy 50(10):1127–1139. https://doi.org/10.1111/cea.13697
55. Prunicki M, Stell L, Dinakarpandian D, de Planell-Saguer M, Lucas RW, Hammond SK, Balmes JR, Zhou X, Paglino T, Sabatti C, Miller RL, Nadeau KC (2018) Exposure to NO(2), CO, and PM(2.5) is linked to regional DNA methylation differences in asthma. Clin. Epigenetics 10(2). https://doi.org/10.1186/s13148-017-0433-4
56. Burleson JD, Siniard D, Yadagiri VK, Chen X, Weirauch MT, Ruff BP, Brandt EB, Hershey GKK, Ji H (2019) TET1 contributes to allergic airway inflammation and regulates interferon and aryl hydrocarbon receptor signaling pathways in bronchial epithelial cells. Sci Rep 9(1):7361. https://doi.org/10.1038/s41598-019-43767-6
57. Somineni HK, Zhang X, Biagini Myers JM, Kovacic MB, Ulm A, Jurcak N, Ryan PH, Khurana Hershey GK, Ji H (2016) Ten-eleven translocation 1 (TET1) methylation is associated with childhood asthma and traffic-related air pollution. J Allergy Clin Immunol 137(3):797–805. https://doi.org/10.1016/j.jaci.2015.10.021
58. Zhang Q, Wang W, Niu Y, Xia Y, Lei X, Huo J, Zhao Q, Zhang Y, Duan Y, Cai J, Ying Z, Li W, Chen R, Fu Q, Kan H (2019) The effects of fine particulate matter constituents on exhaled nitric oxide and DNA methylation in the arginase-nitric oxide synthase pathway. Environ Int 131:105019. https://doi.org/10.1016/j.envint.2019.105019
59. Trump S, Bieg M, Gu Z, Thürmann L, Bauer T, Bauer M, Ishaque N, Röder S, Gu L, Herberth G, Lawerenz C, Borte M, Schlesner M, Plass C, Diessl N, Eszlinger M, Mücke O, Elvers HD, Wissenbach DK, von Bergen M, Herrmann C, Weichenhan D, Wright RJ, Lehmann I, Eils R (2016) Prenatal maternal stress and wheeze in children: novel insights into epigenetic regulation. Sci Rep 6:28616. https://doi.org/10.1038/srep28616
60. Lovinsky-Desir S, Jung KH, Jezioro JR, Torrone DZ, de Planell-Saguer M, Yan B, Perera FP, Rundle AG, Perzanowski MS, Chillrud SN, Miller RL (2017) Physical activity, black carbon exposure, and DNA methylation in the FOXP3 promoter. Clin Epigenetics 9:65. https://doi.org/10.1186/s13148-017-0364-0
61. Rider CF, Yamamoto M, Günther OP, Hirota JA, Singh A, Tebbutt SJ, Carlsten C (2016) Controlled diesel exhaust and allergen coexposure modulates microRNA and gene expression in humans: effects on inflammatory lung markers. J Allergy Clin Immunol 138(6):1690–1700. https://doi.org/10.1016/j.jaci.2016.02.038
62. Jahreis S, Trump S, Bauer M, Bauer T, Thürmann L, Feltens R, Wang Q, Gu L, Grützmann K, Röder S, Averbeck M, Weichenhan D, Plass C, Sack U, Borte M, Dubourg V, Schüürmann G, Simon JC, von Bergen M, Hackermüller J, Eils R, Lehmann I, Polte T (2018) Maternal phthalate exposure promotes allergic airway inflammation over 2 generations through epigenetic modifications. J Allergy Clin Immunol 141(2):741–753. https://doi.org/10.1016/j.jaci.2017.03.017
63. Qiu L, Zhang Y, Do DC, Ke X, Zhang S, Lambert K, Kumar S, Hu C, Zhou Y, Ishmael FT, Gao P (2018) miR-155 modulates cockroach allergen- and oxidative stress-induced Cyclooxygenase-2 in asthma. J Immunol 201(3):916–929. https://doi.org/10.4049/jimmunol.1701167
64. Sbihi H, Jones MJ, MacIsaac JL, Brauer M, Allen RW, Sears MR, Subbarao P, Mandhane PJ, Moraes TJ, Azad MB, Becker AB, Brook JR, Kobor MS, Turvey SE (2019) Prenatal exposure to traffic-related air pollution, the gestational epigenetic clock, and risk of early-life allergic sensitization. J Allergy Clin Immunol 144(6):1729–1731. https://doi.org/10.1016/j.jaci.2019.07.047

65. Wang IJ, Karmaus WJ, Chen SL, Holloway JW, Ewart S (2015) Effects of phthalate exposure on asthma may be mediated through alterations in DNA methylation. Clin Epigenetics 7(1):27. https://doi.org/10.1186/s13148-015-0060-x
66. Lichtenfels AJ d FC, van der Plaat DA, de Jong K, van Diemen CC, Postma DS, Nedeljkovic I, van Duijn CM, Amin N, la Bastide-van Gemert S, de Vries M, Ward-Caviness CK, Wolf K, Waldenberger M, Peters A, Stolk RP, Brunekreef B, Boezen HM, Vonk JM (2018) Long-term air pollution exposure, genome-wide DNA methylation and lung function in the LifeLines Cohort study. Environ Health Perspect 126(2):027004. https://doi.org/10.1289/ehp2045
67. Jeong A, Imboden M, Ghantous A, Novoloaca A, Carsin AE, Kogevinas M, Schindler C, Lovison G, Herceg Z, Cuenin C, Vermeulen R, Jarvis D, Amaral AFS, Kronenberg F, Vineis P, Probst-Hensch N (2019) DNA methylation in inflammatory pathways modifies the association between BMI and adult-onset non-atopic asthma. Int J Environ Res Public Health 16(4). https://doi.org/10.3390/ijerph16040600
68. Amaral AFS, Imboden M, Wielscher M, Rezwan FI, Minelli C, Garcia-Aymerich J, Peralta GP, Auvinen J, Jeong A, Schaffner E, Beckmeyer-Borowko A, Holloway JW, Jarvelin MR, Probst-Hensch NM, Jarvis DL (2020) Role of DNA methylation in the association of lung function with body mass index: a two-step epigenetic Mendelian randomisation study. BMC Pulm Med 20(1):171. https://doi.org/10.1186/s12890-020-01212-9
69. Wang CM, Chang CB, Lee SP, Chan MWY, Wu SF (2020) Differential DNA methylation profiles of peripheral blood mononuclear cells in allergic asthmatic children following dust mite immunotherapy. J Microbiol Immunol Infect 53(6):986–995. https://doi.org/10.1016/j.jmii.2020.06.004
70. Fussbroich D, Kohnle C, Schwenger T, Driessler C, Dücker RP, Eickmeier O, Gottwald G, Jerkic SP, Zielen S, Kreyenberg H, Beermann C, Chiocchetti AG, Schubert R (2020) A combination of LCPUFAs regulates the expression of miRNA-146a-5p in a murine asthma model and human alveolar cells. Prostaglandins Other Lipid Mediat 147:106378. https://doi.org/10.1016/j.prostaglandins.2019.106378
71. Shorey-Kendrick LE, McEvoy CT, Ferguson B, Burchard J, Park BS, Gao L, Vuylsteke BH, Milner KF, Morris CD, Spindel ER (2017) Vitamin C prevents offspring DNA methylation changes associated with maternal smoking in pregnancy. Am J Respir Crit Care Med 196(6):745–755. https://doi.org/10.1164/rccm.201610-2141OC
72. Zhong J, Karlsson O, Wang G, Li J, Guo Y, Lin X, Zemplenyi M, Sanchez-Guerra M, Trevisi L, Urch B, Speck M, Liang L, Coull BA, Koutrakis P, Silverman F, Gold DR, Wu T, Baccarelli AA (2017) B vitamins attenuate the epigenetic effects of ambient fine particles in a pilot human intervention trial. Proc Natl Acad Sci U S A 114(13):3503–3508. https://doi.org/10.1073/pnas.1618545114
73. Xiao C, Biagini Myers JM, Ji H, Metz K, Martin LJ, Lindsey M, He H, Powers R, Ulm A, Ruff B, Ericksen MB, Somineni HK, Simmons J, Strait RT, Kercsmar CM, Khurana Hershey GK (2015) Vanin-1 expression and methylation discriminate pediatric asthma corticosteroid treatment response. J Allergy Clin Immunol 136(4):923–931. https://doi.org/10.1016/j.jaci.2015.01.045
74. Zhang X, Biagini Myers JM, Yadagiri VK, Ulm A, Chen X, Weirauch MT, Khurana Hershey GK, Ji H (2017) Nasal DNA methylation differentiates corticosteroid treatment response in pediatric asthma: a pilot study. PLoS One 12(10):e0186150. https://doi.org/10.1371/journal.pone.0186150
75. Wang AL, Qiu W, DeMeo DL, Raby BA, Weiss ST, Tantisira KG (2019) DNA methylation is associated with improvement in lung function on inhaled corticosteroids in pediatric asthmatics. Pharmacogenet Genomics 29(3):65–68. https://doi.org/10.1097/fpc.0000000000000366
76. Michel S, Busato F, Genuneit J, Pekkanen J, Dalphin JC, Riedler J, Mazaleyrat N, Weber J, Karvonen AM, Hirvonen MR, Braun-Fahrländer C, Lauener R, von Mutius E, Kabesch M, Tost J (2013) Farm exposure and time trends in early childhood may influence DNA methylation in genes related to asthma and allergy. Allergy 68(3):355–364. https://doi.org/10.1111/all.12097

77. Gregory DJ, Kobzik L, Yang Z, McGuire CC, Fedulov AV (2017) Transgenerational transmission of asthma risk after exposure to environmental particles during pregnancy. Am J Phys Lung Cell Mol Phys 313(2):L395–l405. https://doi.org/10.1152/ajplung.00035.2017
78. Brugha R, Lowe R, Henderson AJ, Holloway JW, Rakyan V, Wozniak E, Mahmud N, Seymour K, Grigg J, Shaheen SO (2017) DNA methylation profiles between airway epithelium and proxy tissues in children. Acta Paediatr 106(12):2011–2016. https://doi.org/10.1111/apa.14027
79. Lin PI, Shu H, Mersha TB (2020) Comparing DNA methylation profiles across different tissues associated with the diagnosis of pediatric asthma. Sci Rep 10(1):151. https://doi.org/10.1038/s41598-019-56310-4
80. Moen EL, Zhang X, Mu W, Delaney SM, Wing C, McQuade J, Myers J, Godley LA, Dolan ME, Zhang W (2013) Genome-wide variation of cytosine modifications between European and African populations and the implications for complex traits. Genetics 194(4):987–996. https://doi.org/10.1534/genetics.113.151381
81. Mozhui K, Smith AK, Tylavsky FA (2015) Ancestry dependent DNA methylation and influence of maternal nutrition. PLoS One 10(3):e0118466. https://doi.org/10.1371/journal.pone.0118466
82. Barfield RT, Almli LM, Kilaru V, Smith AK, Mercer KB, Duncan R, Klengel T, Mehta D, Binder EB, Epstein MP, Ressler KJ, Conneely KN (2014) Accounting for population stratification in DNA methylation studies. Genet Epidemiol 38(3):231–241. https://doi.org/10.1002/gepi.21789
83. Breton CV, Marsit CJ, Faustman E, Nadeau K, Goodrich JM, Dolinoy DC, Herbstman J, Holland N, LaSalle JM, Schmidt R, Yousefi P, Perera F, Joubert BR, Wiemels J, Taylor M, Yang IV, Chen R, Hew KM, Freeland DM, Miller R, Murphy SK (2017) Small-magnitude effect sizes in epigenetic end points are important in children's environmental health studies: the Children's Environmental Health and Disease Prevention Research Center's Epigenetics Working Group. Environ Health Perspect 125(4):511–526. https://doi.org/10.1289/ehp595

Epigenetic Epidemiology of Autism and Other Neurodevelopmental Disorders

17

Sung Eun Wang and Yong-Hui Jiang

Contents

Abstract

The prevalence of neurodevelopmental disorders (NDD) has been rising gradually over the last two decades. These developmental disabilities are caused by genetic and environmental factors, reciprocally. Recent extensive epigenetic epidemiological studies suggest that epigenetic dysregulations (Epimutation) such as abnormal DNA methylation may contribute to the etiology of NDD

S. E. Wang
Department of Genetics, Yale University School of Medicine, New Haven, CT, USA
e-mail: sungeun.wang@yale.edu

Y.-H. Jiang (✉)
Department of Genetics, Yale University School of Medicine, New Haven, CT, USA

Department of Neuroscience and Pediatrics, Yale University School of Medicine, New Haven, CT, USA
e-mail: yong-hui.jiang@yale.edu

K. B. Michels (ed.), *Epigenetic Epidemiology*,
https://doi.org/10.1007/978-3-030-94475-9_17

including autism spectrum disorder (ASD), attention-deficit/hyperactivity disorder (ADHD), and intellectual disability (ID). Epimutations of DNA methylation and histone modification have been found in genetic loci under epigenetic regulation but also allele- or tissue-specific patterns in individuals with NDD. Epigenetic reprograming and remodeling occur from embryonic development to throughout life. Recent technical advances in epigenome-wide association studies (EGWAS) and genome-wide differentially methylated regions (DMRs) analyses have established correlations between abnormal DNA methylation and histone modification and neurodevelopmental dysfunctions. However, it remains a challenge to establish a concrete causative evidence that is implicated in the pathogenesis of NDD.

Abbreviations

5caC	5-carboxylcytosine
5fC	5-formylcytosine
5hmC	5-hydroxymethylcytosine
5mC	5-methylcytosine
ADHD	Attention-deficit/hyperactivity disorder
ASD	Autism spectrum disorder
CDC	Centers for Disease Control and Prevention
CHD	Chromodomain helicase DNA binding protein
ChIP-seq	Chromatin immunoprecipitation sequencing
CNVs	Copy number variants
CpG	CG dinucleotides
DA	Differential acetylation
DhMRs	Differentially hydroxymethylated regions
DMRs	Differentially methylated regions
DNMT	DNA methyltransferase
EWAS	Epigenome-wide association studies
GWAS	Genome-wide associated study
HAWAS	Histone acetylome-wide association study
ID	Intellectual disabilities
IQ	Intelligence quotient
KAT	lysine acetyltransferase
KDM	lysine demethylase
KMT	lysine methyltransferase
NDD	Neurodevelopmental disorders
NGS	Next generation sequencing
PFC	Prefrontal cortex
PTVs	Protein-truncating variants
SAM	S-adenylsylmethionine
SCZ	Schizophrenia
SNP	Single nucleotide polymorphism

SNVs	Single nucleotide variants
TC	Temporal cortex
TET	Ten-eleven translocation
TSS	transcription start sites

17.1 Introduction

Neurodevelopmental disorders (NDD) are characterized by difficulties in social communication/interaction, motor skills, learning, memory, or other neurological functions [1, 2]. The phenotypes of NDD evolve over the developmental stage. Both genetic and non-genetic factors are known to implicate the etiology of NDD. It has been recognized that epigenetic modification has a critical function in brain development [3–5]. Widespread DNA methylation and histone modifications reconfigure from fetal brain to young adult, simultaneously with synaptogenesis [4, 6, 7]. Fine regulation of epigenetic processes is essential for normal brain development during prenatal and postnatal time periods [5, 7–9].

Epigenetic mutation/Epimutation is caused by epigenetic dysregulation of DNA, histone, and chromatin remodeling [10]. DNA methylation is considered the primary target because of its molecular nature [2]. Genome-wide disruption of DNA methylation is reported closely associated with NDD accompanying neuropsychiatric disorders [11–13]. Genetic mutations of basic enzymes and transcription factors related to chromatin remodeling and DNA methylation are implicated in epigenetic dysregulation leading to a list of human genetic diseases including NDD [2]. Thus, therapeutic target for NDD faces the enzymes targeting histone or DNA, although the underlying mechanism is not still clear.

Among NDD, autism spectrum disorder (ASD) has shown alarming increase over the last two decades. In 2020, the CDC reported that 1 in 54 children in the USA is diagnosed with ASD, according to 2016 data (32214087). ASD is more than 4 times more common among boys than among girls [14]. ASD prevalence is reported in all race/ethnic and socioeconomic groups [15]. Although most ASD is not inherited, the high concordance rate among identical twins of ASD has been well documented. However, the exact genetic mechanism remains unknown in the majority of cases [16]. Genetic and genomic studies over the two decades have supported a hypothesis that genetic defects play a significant etiological risk in ASD [17, 18]. Approximately, ~20% of ASD cases can be attributed to genetic abnormalities, mostly rare and de novo mutations definitively [19, 20].

Large-scale exome or whole genome sequencing study of >11,000 patients have implicated recurrent copy number variants of gain or loss (CNVs) or protein-truncating variants (PTVs) in ~100 genes in ASD [17–27]. The majority of ASD risk genes are classified into two distinct functional categories: (1) genes encoding proteins directly involved in the development and synaptic function and (2) genes encoding proteins or enzymes associated with epigenetic modulation and chromatin remodeling [20, 28, 29]. The discoveries of synapse affecting genes in ASD patients

are expected as a class of NDD. But, a significant number of ASD risk genes highly enriched for chromatin modifications and DNA methylation are unexpected. This observation suggests that epigenetic dysregulations resulted from nongenetic factors such as epigenetics and environment interaction could also contribute to brain development and pathophysiology of NDD [2, 28, 30].

Attention-deficit/hyperactivity disorder (ADHD) is one of the most common childhood NDDs. Clinical features of ADHD are impulsivity, excessive activity, and attention problems persisting into adulthood [31]. Strong genetic factors are implicated in etiology of ADHD because of higher heritability but the exact genetic basis remains elusive. Like other NDDs, gene and postnatal environment interplay are considered an important mechanism for disease susceptibility [32]. Additionally, environmental factors such as prenatal maternal smoking or lead exposure are also implicated in ADHD [33–35]. Extensive studies have suggested an association between altered DNA methylation of individual genes or loci and ADHD, implicating that changes of DNA methylation pattern in specific loci can be a novel ADHD marker [31, 35].

Intellectual disability (ID) is a prevalent global NDD affecting 1–3% of the world's population. ID is defined by intelligence quotient (IQ) below 70; limitations both in intellectual functioning and in adaptive behavior, including social and practical skills [36]. ID represents a large and heterogeneous group of disorders, sharing a common etiology with other NDD disorders such as ASD and epilepsy. Both genetics and acquired etiological factors are implicated in ID traits, in the same line with other NDD. Most prominent causes of ID are genetic variations such as aneuploidies, copy number variants (CNVs), single nucleotide variants (SNVs), and or dynamic mutation in specific genes. Dynamic mutation of CGG triplet repeat expansion of *FMR1* in fragile X syndrome and de novo mutations of *MECP2* in Rett syndrome is the most common single gene cause in male and female with ID [37, 38]. In fact, epigenetic mechanism is implicated in both *FMR1* and *MECP2* causing ID. The DNA methylation is implicated in the full mutation of CGG triplet of *FMR1* gene in fragile X syndrome. MECP2 gene encodes a protein that is a well-characterized epigenetic modifier [39]. Among the ID related genes >400, approximately ~10 ID genes encode proteins directly regulating chromatin modification and epigenetic writers [40, 41]. Growing evidence indicates the important role of environmental factors such as hazardous chemical exposures, infections during pregnancy, malnutrition, childhood diseases, severe head injury, and infection of central nervous system [42]. Here, we describe the current knowledge of how epigenetic dysregulation, especially DNA modification, contributes to ASD, ADHD, and ID.

17.2 Epigenetic Machinery of DNA Modification

Among the epimutations, chemical modification of DNA has an important role in brain development [2]. As an epigenetic mark, the cytosine modification 5-methylcytosine (5mC) has been well studied in mammalian genome [43]. 5mC occurs in symmetrical CG dinucleotides (CpG) at 5mCpG sites, and more frequently at a genomic region enriched with CpGs referred to as a CpG island, containing

minimum 50% CG content, more than 200 base pairs in length, and often closed to regulatory elements, such as promoters in the genome [44]. The epigenetic "writers" modifying 5' position of cytosine ring include six DNA methyltransferases (DNMT1, DNMT2, DNMT3A, DNMT3B, DNMT3C, and DNMT3L). These DNMTs use S-adenylsylmethionine (SAM) as a methyl donor from the one-carbon metabolism pathway to form 5mC [45]. Extensive studies have suggested DNMT1, DNMT3A, and DNMT3B have a primary role in establishing DNA methylation. DNMT1 maintains DNA methylation by remethylating hemi-methylated DNA during genome replication [46]. Both DNMT3A and DNMT3B are essential for de novo methylation covalently linking a methyl group to the 5′ position of the cytosine ring to create 5mC.

The enzymatic "erasers" targeting 5mC are the ten-eleven translocation (TET) family of methyl cytosine dioxygenases [43]. TET1, TET2, and TET3 oxidize 5mC to 5-hydroxymethylcytosine (5hmC) leading to active DNA demethylation pathway (Fig. 17.1). TET family can further oxidize 5hmC to create 5-formylcytosine (5fC) and 5-carboxylcytosine (5caC). These oxidized forms of 5mC can be excised and

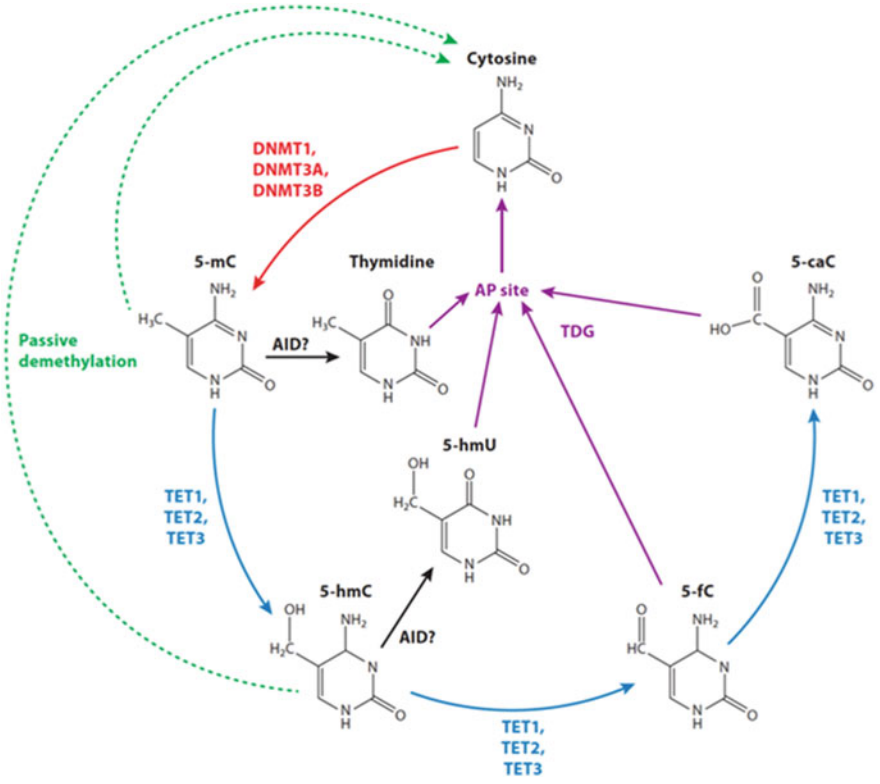

Fig. 17.1 DNA methylation pathway (Note. This image from Annual Review of Medicine "DNA Methylation and Susceptibility to Autism Spectrum Disorder," by Tremblay and Jiang, Copyright 2019 by Annual Reviews. Used under the permission by copyright clearance center)

repaired at any point through the base excision repair pathway toward an unmodified cytosine [43, 47]. This active DNA demethylation, as a primary form occurs in postmitotic cells including mature neurons in the brain [48]. While, passive demethylation loss of 5mC by reduced activity of DNMTs during replication process for mitosis [49] (Fig. 17.1).

The 5mC in CpG islands is widely considered a repressive marker of gene transcription [46]. Besides this canonical function of DNA methylation, it has also important role in fine tuning transcriptional activation/silencing and alternative splicing. Methylated DNA as a "reader" recruits DNA binding proteins with other epigenetic enzymes [10, 50]. Depending on the binding protein, DNA–protein complex can act as a repressor or an enhancer for gene transcription by altering the binding affinity of polymerases to the DNA. Many human disease studies show negative correlation between DNA methylation and RNA polymerase 2 density. Increased DNA methylation is negatively correlated with active histone marks such as H3K4me2/3, H3K9ac, and H3K14ac [51].

As an oxidative derivative of 5mC, 5hmC has a function of regulating gene transcription but its function is elusive [4]. 5hmC peaks are correlated with higher gene expression levels in gene bodies of ESC, but with lower expression levels in NPC. Invariably, 5hmC distribution is widely different in ESCs and NPCs [52]. Besides, technical advances have made it possible to do 5 fc/5caC-specific sequencing but their specific functions related to transcription for individual genes have not been fully elucidated [53]. Other oxidative derivatives of 5mC, such as N (4)-methylcytosine and N(6)-methyladenine, also have not been studied extensively in mammalian systems [46, 54].

17.3 Dynamic Reprogramming of DNA Methylation During Brain Development

Active DNA demethylation is an important epigenetic process to change the cellular distribution of 5mC, 5hmC, and the enzymes related to their formation, especially during early embryonic development [55–58] (Fig. 17.2). Genome-wide reconfigurations of DNA methylation occur at two stages, one immediately after fertilization, with the exception of parental origin-specific imprinting region [59, 60]. Another occurs during early primordial germ cells for gametogenesis [59]. This dynamic epigenomic process of global DNA demethylation and the cell type- or gene-specific reconfiguration of methyl marks during early embryonic development may be highly susceptible to errors in epigenetic modification for brain development. Impaired DNA methylation in animal models deficient *Dnmt1* or *Tet3* has shown that maintaining methylation pattern is critical for cell survival and differentiation [61, 62].

Dynamic regulation of DNA methylation continues even postnatal development and aging. The peak of synaptogenesis in humans (the first five years, varying with brain region) and mice (2–4 weeks) is coincidence with increased 5mC or 5 mH (H = A, C, or T) [4]. In mice, enrichment of 5hmC is shown in actively transcribed

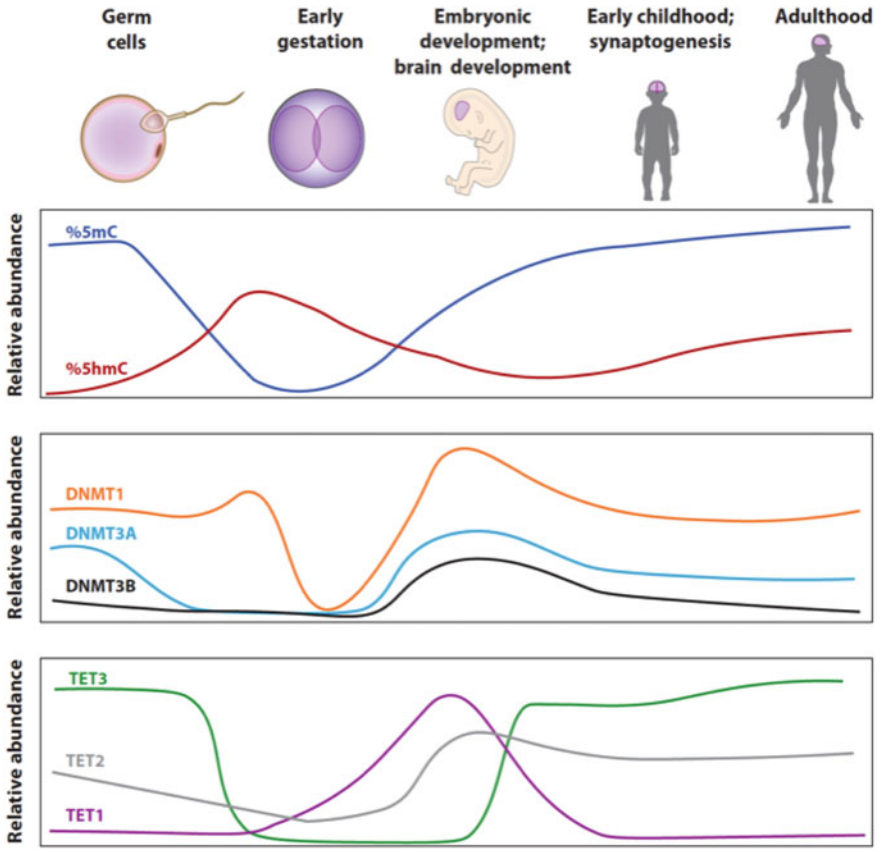

Fig. 17.2 DNA methylation dynamics during brain development. (Note. this image from Annual Review of Medicine "DNA Methylation and Susceptibility to Autism Spectrum Disorder", by Tremblay and Jiang, Copyright 2019 by Annual Reviews. Used under the permission by copyright clearance center)

gene bodies. Dynamic changes in 5mC and 5hmC have been suggested as a crucial factor for synaptogenesis and early brain development.

Epimutations can occur at multiple developmental stages, in contrast to a genetic mutation. Epimutation hot spots (yellow stars in Fig. 17.3) include pronuclear reprogramming following fertilization, primordial germ cell reprogramming and brain development in the embryo, early childhood and post-gestational synaptogenesis, and adult maintenance of synapse function and plasticity. Transgenerational inheritance of epimutations from parental germ cells has also been proposed (dashed lines) [2] (Fig. 17.3). Genome-wide reconfigurations of DNA methylation at embryonic and postnatal developmental stages concrete the possibility that these points in development may be hot spots for the interface of gene and environmental interactions for any disease mutation mechanism.

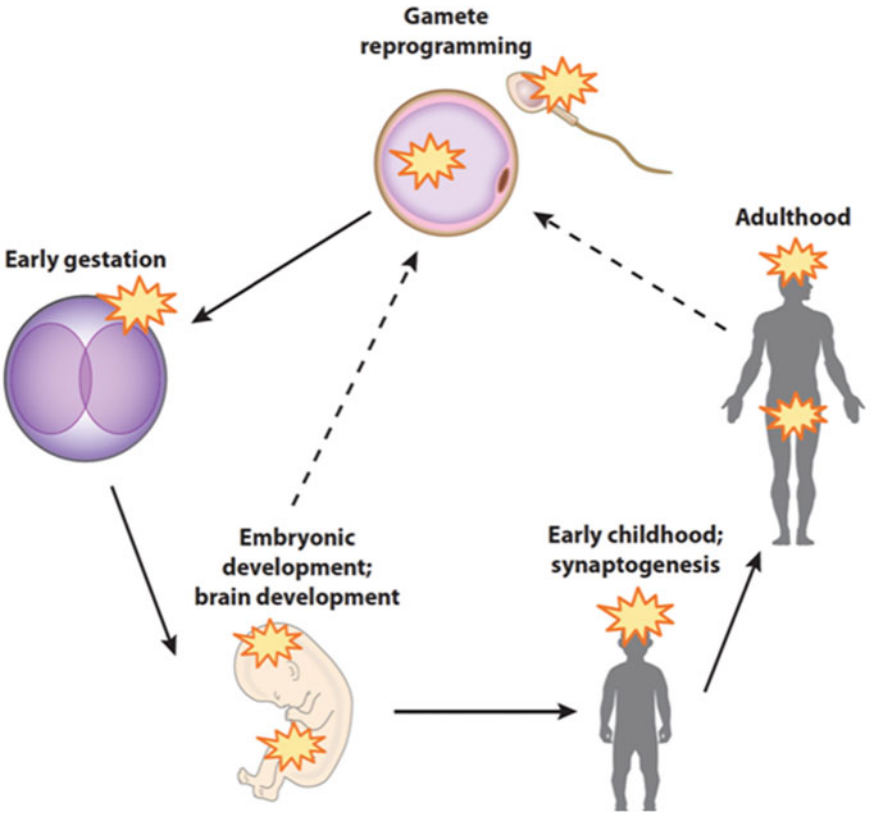

Fig. 17.3 Proposed epimutation hot spot model. (Note. this image from Annual Review of Medicine "DNA Methylation and Susceptibility to Autism Spectrum Disorder," by Tremblay and Jiang, Copyright 2019 by Annual Reviews. Used under the permission by copyright clearance center)

17.4 Evidence of DNA Methylation Dysregulation in ASD

17.4.1 Genetic Alteration in DNA Methylation Machinery Correlated with ASD

Cumulative large-scale next generation sequencing (NGS) studies of patients have supported epigenetic's implications in ASD. Those studies have shown genetic mutations associated with dysregulation of DNA methylation in ASD, existing on multiple levels: genes encoding epigenetic machinery, abnormal methylation in specific loci, and genome-wide correlations of hyper- and hypomethylation. Whole exome and genome sequencing in ASD patients revealed de novo mutations in DNA methylation machinery involving writers, readers, and erasers (*DNMT3A, TET2, MECP2, MBD5*) (Table 17.1). The similar findings also occur to the genes encoding

Table 17.1 Genetic mutations in genes encoding the protein of DNA methylation machinery in patients with NDD

Role	Gene	Description	SFARI score*	Comorbidities	References
Writer	*DNMT3A*	De novo methyltransferase	3	Tatton-Brown-Rahman syndrome, acute myeloid leukemia	[19, 20, 26]
Eraser	*TET3*	DNA methylcytosine dioxygenase	3	Neurodevelopmental disorder	[19, 63]
Reader	*MECP2*	Methylation-dependent transcriptional repressor	2	Rett syndrome, X-linked intellectual disability, encephalopathy	[64]
Reader	*MBD5*	Methyl-CpG-binding domain family member, interacts with the polycomb repressive complex	3	Intellectual disability, microcephaly, intellectual disabilities, severe speech impairment, and seizure	[65, 66]

the histone modifications and chromatin remodeling such as lysine acetyltransferases (KATs), lysine methyltransferases (KMTs), lysine demethylases (KDMs), and chromodomain Helicase DNA binding proteins (CHDs) (Table 17.2). These genetic mutations typically result in deficiency of epigenetic modifying enzymes. These genetic findings allow significant molecular evidence to support epigenetic dysfunction and the epigenetic's implication in ASD [2].

17.4.2 Changes in DNA Methylation of ASD Risk Genes

Multiple studies have reported aberrant methylation patterns in ASD patients, mainly hypermethylation of promoter region including CpG island. Hypermethylation of MECP2 and UBE3A promoter was found with loss-of-function mutations in postmortem ASD patients' brains, at initial studies [87, 88]. Following those discoveries, many other genes such as *OXTR, SNRPN, MAGEL2, FMR1, RELN*, and *GAD1* have been found as targeted genes of hypermethylated transcription start sites (TSS), with reduced gene expression in ASD patients brains compared to normal brain [89–93]. In comparison, hypomethylation pattern was shown in several genes including *RORA, ERMN, USP24, METTL21C, PDE10A, STX16*, and *DBT* in peripheral blood DNA of ASD patients [94]. In the same line with hypomethylation, abundant 5hmC in promoters has been shown a positive correlation with increased transcription of genes encoding Engrailed-2, *GAD1*, and *MECP2* [64, 95]. These genetic discoveries support the epigenetic dysregulation in ASD, but the causality and origin of these changes remain to be elucidated.

Table 17.2 Genetic mutations in genes encoded in proteins of chromatin machinery in ASD patients

Role	Gene	Description	SFARI score[a]	Comorbidities	Reference
Writer	*ASH1L*	H3K36 methyltransferase, associates with HOX genes	1	ID/ASD	[19, 20, 67–71]
Writer	*EHMT1*	H3K9 methyltransferase, E2F6 transcriptional repression complex component	3	Kleefstra syndrome	[72–76]
Writer	*KAT2B*	Histone acetyltransferase for lysine, associates with p300/CBP	2		[19, 67, 77]
Writer	*KAT6A*	H3K9 histone acetyltransferase	3	ID	[19, 67]
Writer/ reader	*KMT2A*	H3K4 methyltransferase, binds to unmethylated promoter CpGs to maintain unmethylated state	1	MLL, Wiedemann-Steiner syndrome	[19, 20, 67, 69, 78]
Writer/ reader	*KMT2C*	H3K4 methyltransferase, contains a DNA binding domain	2	Kleefstra syndrome/ ASD/ID	[19, 20, 25, 67–69, 79, 80]
Writer	*KMT5B*	H4K20 methyltransferase	1	ID	[20, 26, 27, 67, 69, 70]
Eraser	*KDM5B (JARID1B)*	H3K4 demethylase	2	ID	[19, 20, 69, 79, 81]
Eraser	*KDM5C (JARID1C)*	H3K4 demethylase	3	X linked ID	[72, 82, 83]
Chromatin	*ASXL3*	Polycomb protein	1	Bainbridge-Ropers Syndrome/ ID/ASD	[20, 67, 69, 84]
Chromatin	*ARID1B*	Chromatin remodeling, SWI/SNF complexes	1	Coffin-Siris syndrome/ ASD/ID	[25, 67–70, 78, 79, 85, 86]

(continued)

Table 17.2 (continued)

Role	Gene	Description	SFARI score[a]	Comorbidities	Reference
Chromatin	*CHD2*, *CHD8*	Chromatin remodeling	1		[19, 20, 25, 67–70, 78, 80, 81, 84, 86]

Note. This table is modified from the Annual Review of Medicine "DNA Methylation and Susceptibility to Autism Spectrum Disorder," by Tremblay and Jiang, Copyright 2019 by Annual Reviews. Used under the permission by copyright clearance center

[a]SFARI score based on clinical data: 1–2, strong evidence; 3–4, suggestive/minimal evidence

However, DNA methylation is not a one-size-fits-all rule, acting as an on-off switch at promoters. Abundant DNA methylation in gene bodies has an important function of alternative splicing to generate isoform. Our recent studies have shown that alternative TSS can be regulated by hypermethylation in specific intragenic CpG islands for isoform-specific expression of *SHANK3* in ASD patients' brains [96]. This finding supports an important role of DNA methylation regulating alternative promoters for gene products in the brain [97, 98]. Besides, DNA methylation in gene bodies or intragenic regions can slow down the elongation rate of RNA polymerase II by recruiting methylation-dependent splicing factors [99]. Some studies have suggested dysregulation of splicing pre-mRNAs in ASD patients' brains [100, 101]. Integrative analysis between isoform-specific data and DNA methylome should be established in the brain, because alternative splicing is present in 95% of neuronal genes and a crucial mechanism to contribute to the protein diversity encoded by neuronal genes [102, 103].

17.4.3 Differentially Methylated Regions of DNA Methylome in ASD

Genome-wide DNA methylation profiling has emerged as a popular approach to identify convergent molecular candidates. Many studies reported 5mC levels in blood samples of ASD patients, due to scarcity of ASD postmortem brain tissues [89, 104–106]. One group has shown that differentially methylated regions (DMRs) in sperm DNA are associated with early signatures of ASD risk in one cohort [107]. In addition, DNA methylome analysis has been attempted in placental tissue at birth [108].

Current technical advances bring the extensive DNA methylome dataset to interpret wide DMRs in >200 postmortem ASD brain tissues [104, 109, 110]. A systematic view of DNA methylation in prefrontal cortex, temporal cortex, and cerebellum from 43 ASD patients has shown that a number of co-methylated modules are enriched for genes involved in immune system, synaptic signaling, and neuronal function [109, 110]. Analyzing single nucleotide polymorphism (SNP)

has also reported genetic variation associated with the changes of DNA methylation in the cases analyzed [104].

Profiling genome-wide distribution of 5hmC has identified differentially hydroxmethylated regions (DhMRs) in postmortem ASD cerebellums [111, 112]. DhMRs between ASD patients and normal controls were shown only in a young group (age $\leq$ 18), suggesting that genome-wide disruption of 5hmC may be specific to early developmental stage. Additionally, these DhMRs associated genes are involved in psychiatric disease and NDD such as ASD, ID, and SCZ [111, 112]. Recent studies have suggested that altered intragenic DMRs/DhMRs in postmortem ASD brains associate with aberrant precursor messenger RNA splicing [113]. However, the concrete evidence for the function of DhMRs in noncoding region remains unknown.

17.4.4 Impact of Environmental Contributor on ASD Pathogenesis

Epigenome is known to be modifiable by environmental factors. Many studies have supported the hypothesis using environmental chemicals in rodents [30]. Maternal diet involving methyl-donor precursors affects increasing DNA methylation on specific loci such as agouti, changing the coat color in offspring [114]. This finding provides direct evidence that environmental factors exposure to in utero can contribute to DNA methylation of offspring. Regarding maternal care of pups such as licking and grooming, glucocorticoid receptor gene promoter was shown in hypomethylation status. These 5mC patterns can be reversed by alteration of care and persist into adult age [115]. The environmental chemicals such as polychlorinated biphenyls 95 and bisphenolin have been found in postmortem ASD brain tissues, placenta, cord blood, suggesting a positive correlation with abnormal DNA methylation in ASD risk genes loci. Furthermore, maternal exposures to air pollution, obesity, asthma, stress, alcohol, and tobacco in utero have been shown to impact DNA methylome of offspring [28, 30, 116–120]. It is poorly understood whether there is a critical window during development and hot spots in genome that are more susceptible to environmental insults and resulting in the accumulations of epimutations.

17.5 Abnormal DNA Methylation Associated with ADHD

ADHD has been considered as a childhood disease but is also recognized now that 20% of cases with ADHD persist into teens and adults [121]. Cumulative data have suggested that the prevalence of ADHD is greater in males than females, globally [121, 122]. Gender ratios of ADHD prevalence in childhood are extremely varied by countries ranging from 1:3 to 1:16 in females to males [122]. More than 10 studies of worldwide meta-regression analysis in adults ($\geq$19 years) found that the prevalence rate has shown a similar tendency in males to females (1.6:1) [121]. Besides, the symptoms of ADHD are different between individuals with ages [123]. A worldwide

meta-analysis of around 100 studies in children, adolescents, and adults has shown that the hyperactive-impulsive type was the most common symptom in preschool children. Otherwise, inattentive type was the most predominant common symptom in adults and kids over 6 years old [121].

Likewise other NDD, ADHD is affected by interaction of gene and environmental burdens including maternal stress, lead, nicotine, alcohol, polychlorinated biphenyls, poor maternal diet, and drug exposure such as paracetamol [34, 35, 124, 125]. GWAS and Epigenome-wide association studies (EWAS) have identified that DNA methylation is associated with ADHD symptoms. Initial methylome studies found the altered DNA methylation in *VIPR2* gene from salivary DNA in children with ADHD [126]. Accumulating large-scale methylome datasets in ADHD patients has suggested the correlation between DNA methylation from peripheral blood and postmortem brain samples with ADHD [33, 127]. Recent studies provided specific candidate genes loci associated with DMRs. Prenatal smoking exposure to fetus impact altered DNA methylation of *TARBP1* and *GFI1* gene promoters, linking between environment and ADHD [128, 129]. Prenatal unhealthy diet also influences hypermethylation of *IGF2* promoter in offspring, increasing risk of ADHD [130]. Aberrant DNA methylation of *DRD4* and *KLDR1* genes loci contributed to persistent ADHD status [128]. Hypermethylated *SLC6A4* promoter was positively correlated with certain features of ADHD, such as behavioral disinhibition [131].

17.6 Emerging Role of Chromatin Structural Protein and Modifiers in Neurodevelopmental Disorders

As genetic heterogenecity of ID, many studies have found high comorbidity between ID and other NDD including ASD, ADHD, and SCZ [36]. About 40% of ID cases have a comorbidity of ASD and in the opposite ID is associated with 30% of children with ASD [36, 132]. The symptomatic overlap between ID and ASD/other NDD is reflected in shared genetic susceptibility among them. A significant fraction of known ID causing genes have been implicated in ASD, ADHD, and other neuropsychiatric disorders [36]. Not surprisingly, mutations in genes encoding synaptic scaffold proteins or neuronal channels/receptors in the synapse are identified in ID with variable NDD. Growing NGS data with advent analyses indicate a certain causative variant in $>17\%$ of ~100 patients with ID and ASD, simultaneously. Interestingly, 16 genes shared in ID and ASD are chromatin remodeling factors [133].

DNA methylome analysis in blood samples from ID patients found a mutation of histone modifying enzymes such as SETD1B, SETD5, and EHMT1 including SET domain, which is necessary for histone methylation. SETD1B and EHMT1 have a specific target of H3K4 and H3K9, respectively [133, 134]. Besides, loss of function in SETD1B brings DMRs in gene bodies associated with regulatory clusters in syndromic ID patients [134]. Emerging evidence indicates that disrupted H3K36me3, acting as a docking signal for other chromatin remodeling enzymes, causes ID common symptoms in patients with haploinsufficient SETD5 or loss of

function in SETD2 [135]. Regarding DNA demethylation, TET3 deficiency in 11 patients shows common phenotypes of ID and developmental delay with disruption of DNA demethylation [63].

In addition to discovering aberrant DNA methylation from EWAS, histone acetylome-wide association study (HAWAS) provided evidence of aberrant histone modifications in postmortem ASD brains [136]. Histone acetylomes analysis was performed by H3K27ac chromatin immunoprecipitation sequencing (ChIP-seq) using prefrontal cortex, temporal cortex, and cerebellum tissues from ASD and controls. H3K27ac is useful to highlight active enhancers and promoters, in the same line with gene expression and transcription factor binding. Functional enrichment analysis showed a quite similar pattern of differential acetylation peak in prefrontal cortex and temporal cortex. Upregulated differential acetylation peaks were significantly enriched in synaptic transmission and cation transport activity such as *SLC30A5* zinc transporter gene. While, downregulated differential acetylation peaks showed enrichment in immune-related genes or -signaling pathway. Interestingly, histone deacetylase genes including *HDAC2* and *HDAC4* had downregulated differential acetylation peaks in prefrontal cortex and temporal cortex. Furthermore, ASD-related increased differential acetylation peaks were different depending on developmental stages, not reduced differential acetylation peaks. Despite the causes for ASD are heterogeneous, a large cohort of ASD analysis shows a convergence on shared downstream epimutations.

17.7 Conclusion

The role of epigenetics has been increasingly recognized in the field of epidemiology because epigenetic modification serves as a molecular interface to measure the impact of gene and environment interaction. The studies of epigenetic epidemiology of NDDs have drawn much attention over the last decade. While substantial progress has been made to understand the genetic contributions to the NDD, the cause for a significant fraction or the majority of NDD cases remains elusive. Emerging evidence have supported the epigenetic dysregulation during early development could play a significant etiological role for NDD. However, in contrast to the studies of genetic contributions, challenges to understand epigenetic dysregulation in NDD remains substantial. Technically, because the epimutations are frequently tissue- and cell-type specific, the availability of targeted disease tissues such as brains become critically important. Unfortunately, the high-quality postmortem brain tissues are very limited. Conceptually, it has been difficulty to establish a causal relationship between observed epigenetic dysregulations and diseases implicated because the functional consequences of associated epimutations are hard to determine in humans. New techniques that are capable of profiling epigenome and transcriptome at single cell level simultaneously may overcome some of these challenges in future studies. A study of a mechanistic framework how epimutations at single, multiple loci, as well genome-wide levels contribute to the disease pathogenesis remains to be explored.

References

1. America's Children and the Environment 2013. Third Edition ed. Washington, D.C.: U.-S. Environmental Protection Agency.
2. Tremblay MW, Jiang YH (2019) DNA methylation and susceptibility to autism Spectrum disorder. Annu Rev Med 70:151–166. https://doi.org/10.1146/annurev-med-120417-091431
3. de la Torre-Ubieta L, Stein JL, Won H, Opland CK, Liang D, Lu D et al (2018) The dynamic landscape of open chromatin during human cortical neurogenesis. Cell 172(1–2):289–304 e18. https://doi.org/10.1016/j.cell.2017.12.014
4. Lister R, Mukamel EA, Nery JR, Urich M, Puddifoot CA, Johnson ND et al (2013) Global epigenomic reconfiguration during mammalian brain development. Science 341(6146): 1237905. https://doi.org/10.1126/science.1237905
5. Bale TL (2015) Epigenetic and transgenerational reprogramming of brain development. Nat Rev Neurosci 16(6):332–344. https://doi.org/10.1038/nrn3818
6. Lamonica JM, Zhou Z (2019) Disentangling chromatin architecture to gain insights into the etiology of brain disorders. Curr Opin Genet Dev 55:76–81. https://doi.org/10.1016/j.gde.2019.06.009
7. Jeong H, Mendizabal I, Berto S, Chatterjee P, Layman T, Usui N et al (2021) Evolution of DNA methylation in the human brain. Nat Commun 12(1):2021. https://doi.org/10.1038/s41467-021-21917-7
8. Bludau A, Royer M, Meister G, Neumann ID, Menon R (2019) Epigenetic regulation of the social brain. Trends Neurosci 42(7):471–484. https://doi.org/10.1016/j.tins.2019.04.001
9. Jakovcevski M, Akbarian S (2012) Epigenetic mechanisms in neurological disease. Nat Med 18(8):1194–1204. https://doi.org/10.1038/nm.2828
10. Allis CD, Jenuwein T (2016) The molecular hallmarks of epigenetic control. Nat Rev Genet 17(8):487–500. https://doi.org/10.1038/nrg.2016.59
11. Madrid A, Papale LA, Alisch RS (2016) New hope: the emerging role of 5-hydroxymethylcytosine in mental health and disease. Epigenomics 8(7):981–991. https://doi.org/10.2217/epi-2016-0020
12. Kuehner JN, Bruggeman EC, Wen Z, Yao B (2019) Epigenetic regulations in neuropsychiatric disorders. Front Genet 10:268. https://doi.org/10.3389/fgene.2019.00268
13. Moore LD, Le T, Fan G (2013) DNA methylation and its basic function. Neuropsychopharmacology 38(1):23–38. https://doi.org/10.1038/npp.2012.112
14. Maenner MJ, Shaw KA, Baio J, Washington A, Patrick M et al (eds) (2020) Prevalence of autism Spectrum disorder among children aged 8 years – autism and developmental disabilities monitoring network, 11 sites, United States, 2016. MMWR Surveill Summ 69(4):1–12. https://doi.org/10.15585/mmwr.ss6904a1
15. Baio J, Wiggins L, Christensen DL, Maenner MJ, Daniels J, Warren Z et al (2018) Prevalence of autism Spectrum disorder among children aged 8 years – autism and developmental disabilities monitoring network, 11 sites, United States, 2014. MMWR Surveill Summ 67(6):1–23. https://doi.org/10.15585/mmwr.ss6706a1
16. Folstein SE, Rosen-Sheidley B (2001) Genetics of autism: complex aetiology for a heterogeneous disorder. Nat Rev Genet 2(12):943–955. https://doi.org/10.1038/35103559
17. Jiang YH, Wang Y, Xiu X, Choy KW, Pursley AN, Cheung SW (2014) Genetic diagnosis of autism spectrum disorders: the opportunity and challenge in the genomics era. Crit Rev Clin Lab Sci 51(5):249–262. https://doi.org/10.3109/10408363.2014.910747
18. Willsey AJ, State MW (2015) Autism spectrum disorders: from genes to neurobiology. Curr Opin Neurobiol 30:92–99. https://doi.org/10.1016/j.conb.2014.10.015
19. Iossifov I, O'Roak BJ, Sanders SJ, Ronemus M, Krumm N, Levy D et al (2014) The contribution of de novo coding mutations to autism spectrum disorder. Nature 515(7526): 216–221. https://doi.org/10.1038/nature13908

20. De Rubeis S, He X, Goldberg AP, Poultney CS, Samocha K, Cicek AE et al (2014) Synaptic, transcriptional and chromatin genes disrupted in autism. Nature 515(7526):209–215. https://doi.org/10.1038/nature13772
21. Satterstrom FK, Kosmicki JA, Wang J, Breen MS, De Rubeis S, An JY et al (2020) Large-scale exome sequencing study implicates both developmental and functional changes in the neurobiology of autism. Cell 180(3):568–84 e23. https://doi.org/10.1016/j.cell.2019.12.036
22. Buxbaum JD, Daly MJ, Devlin B, Lehner T, Roeder K, State MW et al (2012) The autism sequencing consortium: large-scale, high-throughput sequencing in autism spectrum disorders. Neuron 76(6):1052–1056. https://doi.org/10.1016/j.neuron.2012.12.008
23. Neale BM, Kou Y, Liu L, Ma'ayan A, Samocha KE, Sabo A et al (2012) Patterns and rates of exonic de novo mutations in autism spectrum disorders. Nature 485(7397):242–245. https://doi.org/10.1038/nature11011
24. O'Roak BJ, Deriziotis P, Lee C, Vives L, Schwartz JJ, Girirajan S et al (2011) Exome sequencing in sporadic autism spectrum disorders identifies severe de novo mutations. Nat Genet 43(6):585–589. https://doi.org/10.1038/ng.835
25. O'Roak BJ, Vives L, Girirajan S, Karakoc E, Krumm N, Coe BP et al (2012) Sporadic autism exomes reveal a highly interconnected protein network of de novo mutations. Nature 485(7397):246–250. https://doi.org/10.1038/nature10989
26. Sanders SJ, Murtha MT, Gupta AR, Murdoch JD, Raubeson MJ, Willsey AJ et al (2012) De novo mutations revealed by whole-exome sequencing are strongly associated with autism. Nature 485(7397):237–241. https://doi.org/10.1038/nature10945
27. Iossifov I, Ronemus M, Levy D, Wang Z, Hakker I, Rosenbaum J et al (2012) De novo gene disruptions in children on the autistic spectrum. Neuron 74(2):285–299. https://doi.org/10.1016/j.neuron.2012.04.009
28. Tran NQV, Miyake K (2017) Neurodevelopmental disorders and environmental toxicants: epigenetics as an underlying mechanism. Int J Genomics 2017:7526592. https://doi.org/10.1155/2017/7526592
29. Duffney LJ, Valdez P, Tremblay MW, Cao X, Montgomery S, McConkie-Rosell A et al (2018) Epigenetics and autism spectrum disorder: a report of an autism case with mutation in H1 linker histone HIST1H1E and literature review. Am J Med Genet B Neuropsychiatr Genet 177(4):426–433. https://doi.org/10.1002/ajmg.b.32631
30. Keil KP, Lein PJ (2016) DNA methylation: a mechanism linking environmental chemical exposures to risk of autism spectrum disorders? Environ Epigenet 2(1). https://doi.org/10.1093/eep/dvv012
31. Goodman SJ, Burton CL, Butcher DT, Siu MT, Lemire M, Chater-Diehl E et al (2020) Obsessive-compulsive disorder and attention-deficit/hyperactivity disorder: distinct associations with DNA methylation and genetic variation. J Neurodev Disord 12(1):23. https://doi.org/10.1186/s11689-020-09324-3
32. Thapar A, Stergiakouli E (2008) An overview on the genetics of ADHD. Xin Li Xue Bao 40(10):1088–1098. https://doi.org/10.3724/SP.J.1041.2008.01088
33. Neumann A, Walton E, Alemany S, Cecil C, Gonzalez JR, Jima DD et al (2020) Association between DNA methylation and ADHD symptoms from birth to school age: a prospective meta-analysis. Transl Psychiatry 10(1):398. https://doi.org/10.1038/s41398-020-01058-z
34. Marceau K, Cinnamon Bidwell L, Karoly HC, Evans AS, Todorov AA, Palmer RH et al (2018) Within-family effects of smoking during pregnancy on ADHD: the importance of phenotype. J Abnorm Child Psychol 46(4):685–699. https://doi.org/10.1007/s10802-017-0320-7
35. Daneshparvar M, Mostafavi SA, Zare Jeddi M, Yunesian M, Mesdaghinia A, Mahvi AH et al (2016) The role of Lead exposure on attention-deficit/ hyperactivity disorder in children: a systematic review. Iran J Psychiatry 11(1):1–14
36. van Bokhoven H (2011) Genetic and epigenetic networks in intellectual disabilities. Annu Rev Genet 45:81–104. https://doi.org/10.1146/annurev-genet-110410-132512

37. Jin P, Warren ST (2000) Understanding the molecular basis of fragile X syndrome. Hum Mol Genet 9(6):901–908. https://doi.org/10.1093/hmg/9.6.901
38. Sajan SA, Jhangiani SN, Muzny DM, Gibbs RA, Lupski JR, Glaze DG et al (2017) Enrichment of mutations in chromatin regulators in people with Rett syndrome lacking mutations in MECP2. Genet Med 19(1):13–19. https://doi.org/10.1038/gim.2016.42
39. Brasa S, Mueller A, Jacquemont S, Hahne F, Rozenberg I, Peters T et al (2016) Reciprocal changes in DNA methylation and hydroxymethylation and a broad repressive epigenetic switch characterize FMR1 transcriptional silencing in fragile X syndrome. Clin Epigenetics 8:15. https://doi.org/10.1186/s13148-016-0181-x
40. van Bokhoven H, Kramer JM (2010) Disruption of the epigenetic code: an emerging mechanism in mental retardation. Neurobiol Dis 39(1):3–12. https://doi.org/10.1016/j.nbd.2010.03.010
41. Inlow JK, Restifo LL (2004) Molecular and comparative genetics of mental retardation. Genetics 166(2):835–881. https://doi.org/10.1534/genetics.166.2.835
42. Ilyas M, Mir A, Efthymiou S, Houlden H (2020) The genetics of intellectual disability: advancing technology and gene editing. F1000Res 9. https://doi.org/10.12688/f1000research.16315.1
43. Wu X, Zhang Y (2017) TET-mediated active DNA demethylation: mechanism, function and beyond. Nat Rev Genet 18(9):517–534. https://doi.org/10.1038/nrg.2017.33
44. Bird A (2002) DNA methylation patterns and epigenetic memory. Genes Dev 16(1):6–21. https://doi.org/10.1101/gad.947102
45. Mentch SJ, Locasale JW (2016) One-carbon metabolism and epigenetics: understanding the specificity. Ann N Y Acad Sci 1363:91–98. https://doi.org/10.1111/nyas.12956
46. Lyko F (2018) The DNA methyltransferase family: a versatile toolkit for epigenetic regulation. Nat Rev Genet 19(2):81–92. https://doi.org/10.1038/nrg.2017.80
47. Cortellino S, Xu J, Sannai M, Moore R, Caretti E, Cigliano A et al (2011) Thymine DNA glycosylase is essential for active DNA demethylation by linked deamination-base excision repair. Cell 146(1):67–79. https://doi.org/10.1016/j.cell.2011.06.020
48. Wu SC, Zhang Y (2010) Active DNA demethylation: many roads lead to Rome. Nat Rev Mol Cell Biol 11(9):607–620. https://doi.org/10.1038/nrm2950
49. Chen T, Li E (2004) Structure and function of eukaryotic DNA methyltransferases. Curr Top Dev Biol 60:55–89. https://doi.org/10.1016/S0070-2153(04)60003-2
50. Yin Y, Morgunova E, Jolma A, Kaasinen E, Sahu B, Khund-Sayeed S et al (2017) Impact of cytosine methylation on DNA binding specificities of human transcription factors. Science 356(6337). https://doi.org/10.1126/science.aaj2239
51. Lorincz MC, Dickerson DR, Schmitt M, Groudine M (2004) Intragenic DNA methylation alters chromatin structure and elongation efficiency in mammalian cells. Nat Struct Mol Biol 11(11):1068–1075. https://doi.org/10.1038/nsmb840
52. Shi DQ, Ali I, Tang J, Yang WC (2017) New insights into 5hmC DNA modification: generation. Distribution and Function Front Genet 8:100. https://doi.org/10.3389/fgene.2017.00100
53. Liu Y, Hu Z, Cheng J, Siejka-Zielinska P, Chen J, Inoue M et al (2021) Subtraction-free and bisulfite-free specific sequencing of 5-methylcytosine and its oxidized derivatives at base resolution. Nat Commun 12(1):618. https://doi.org/10.1038/s41467-021-20920-2
54. Luo GZ, Blanco MA, Greer EL, He C, Shi Y (2015) DNA N(6)-methyladenine: a new epigenetic mark in eukaryotes? Nat Rev Mol Cell Biol 16(12):705–710. https://doi.org/10.1038/nrm4076
55. Cao Z, Zhou N, Zhang Y, Zhang Y, Wu R, Li Y et al (2014) Dynamic reprogramming of 5-hydroxymethylcytosine during early porcine embryogenesis. Theriogenology 81(3):496–508. https://doi.org/10.1016/j.theriogenology.2013.10.025
56. Feng J, Zhou Y, Campbell SL, Le T, Li E, Sweatt JD et al (2010) Dnmt1 and Dnmt3a maintain DNA methylation and regulate synaptic function in adult forebrain neurons. Nat Neurosci 13(4):423–430. https://doi.org/10.1038/nn.2514

57. Wossidlo M, Nakamura T, Lepikhov K, Marques CJ, Zakhartchenko V, Boiani M et al (2011) 5-Hydroxymethylcytosine in the mammalian zygote is linked with epigenetic reprogramming. Nat Commun 2:241. https://doi.org/10.1038/ncomms1240
58. Szwagierczak A, Bultmann S, Schmidt CS, Spada F, Leonhardt H (2010) Sensitive enzymatic quantification of 5-hydroxymethylcytosine in genomic DNA. Nucleic Acids Res 38(19):e181. https://doi.org/10.1093/nar/gkq684
59. Morgan HD, Santos F, Green K, Dean W, Reik W (2005) Epigenetic reprogramming in mammals. Hum Mol Genet 14(1):R47–R58. https://doi.org/10.1093/hmg/ddi114
60. Reik W, Walter J (2001) Genomic imprinting: parental influence on the genome. Nat Rev Genet 2(1):21–32. https://doi.org/10.1038/35047554
61. Gu TP, Guo F, Yang H, Wu HP, Xu GF, Liu W et al (2011) The role of Tet3 DNA dioxygenase in epigenetic reprogramming by oocytes. Nature 477(7366):606–610. https://doi.org/10.1038/nature10443
62. von Meyenn F, Iurlaro M, Habibi E, Liu NQ, Salehzadeh-Yazdi A, Santos F et al (2016) Impairment of DNA methylation maintenance is the Main cause of global demethylation in naive embryonic stem cells. Mol Cell 62(6):848–861. https://doi.org/10.1016/j.molcel.2016.04.025
63. Beck DB, Petracovici A, He C, Moore HW, Louie RJ, Ansar M et al (2020) Delineation of a human Mendelian disorder of the DNA demethylation machinery: TET3 deficiency. Am J Hum Genet 106(2):234–245. https://doi.org/10.1016/j.ajhg.2019.12.007
64. Zhubi A, Chen Y, Dong E, Cook EH, Guidotti A, Grayson DR (2014) Increased binding of MeCP2 to the GAD1 and RELN promoters may be mediated by an enrichment of 5-hmC in autism spectrum disorder (ASD) cerebellum. Transl Psychiatry 4:e349. https://doi.org/10.1038/tp.2013.123
65. Cukier HN, Rabionet R, Konidari I, Rayner-Evans MY, Baltos ML, Wright HH et al (2010) Novel variants identified in methyl-CpG-binding domain genes in autistic individuals. Neurogenetics 11(3):291–303. https://doi.org/10.1007/s10048-009-0228-7
66. Talkowski ME, Mullegama SV, Rosenfeld JA, van Bon BW, Shen Y, Repnikova EA et al (2011) Assessment of 2q23.1 microdeletion syndrome implicates MBD5 as a single causal locus of intellectual disability, epilepsy, and autism spectrum disorder. Am J Hum Genet 89(4):551–563. https://doi.org/10.1016/j.ajhg.2011.09.011
67. Iossifov I, Levy D, Allen J, Ye K, Ronemus M, Lee YH et al (2015) Low load for disruptive mutations in autism genes and their biased transmission. Proc Natl Acad Sci U S A 112(41):E5600–E5607. https://doi.org/10.1073/pnas.1516376112
68. Wang T, Guo H, Xiong B, Stessman HA, Wu H, Coe BP et al (2016) De novo genic mutations among a Chinese autism spectrum disorder cohort. Nat Commun 7:13316. https://doi.org/10.1038/ncomms13316
69. RK CY, Merico D, Bookman M, JLH, Thiruvahindrapuram B, Patel RV, et al (2017) Whole genome sequencing resource identifies 18 new candidate genes for autism spectrum disorder. Nat Neurosci 20(4):602–611. https://doi.org/10.1038/nn.4524
70. Stessman HA, Xiong B, Coe BP, Wang T, Hoekzema K, Fenckova M et al (2017) Targeted sequencing identifies 91 neurodevelopmental-disorder risk genes with autism and developmental-disability biases. Nat Genet 49(4):515–526. https://doi.org/10.1038/ng.3792
71. Willsey AJ, Sanders SJ, Li M, Dong S, Tebbenkamp AT, Muhle RA et al (2013) Coexpression networks implicate human midfetal deep cortical projection neurons in the pathogenesis of autism. Cell 155(5):997–1007. https://doi.org/10.1016/j.cell.2013.10.020
72. Alvarez-Mora MI, Calvo Escalona R, Puig Navarro O, Madrigal I, Quintela I, Amigo J et al (2016) Comprehensive molecular testing in patients with high functioning autism spectrum disorder. Mutat Res 784–785:46–52. https://doi.org/10.1016/j.mrfmmm.2015.12.006
73. Talkowski ME, Rosenfeld JA, Blumenthal I, Pillalamarri V, Chiang C, Heilbut A et al (2012) Sequencing chromosomal abnormalities reveals neurodevelopmental loci that confer risk across diagnostic boundaries. Cell 149(3):525–537. https://doi.org/10.1016/j.cell.2012.03.028

74. Kleefstra T, Brunner HG, Amiel J, Oudakker AR, Nillesen WM, Magee A et al (2006) Loss-of-function mutations in euchromatin histone methyl transferase 1 (EHMT1) cause the 9q34 subtelomeric deletion syndrome. Am J Hum Genet 79(2):370–377. https://doi.org/10.1086/505693
75. Jiang YH, Yuen RK, Jin X, Wang M, Chen N, Wu X et al (2013) Detection of clinically relevant genetic variants in autism spectrum disorder by whole-genome sequencing. Am J Hum Genet 93(2):249–263. https://doi.org/10.1016/j.ajhg.2013.06.012
76. Balan S, Iwayama Y, Maekawa M, Toyota T, Ohnishi T, Toyoshima M et al (2014) Exon resequencing of H3K9 methyltransferase complex genes, EHMT1, EHTM2 and WIZ, in Japanese autism subjects. Mol Autism 5(1):49. https://doi.org/10.1186/2040-2392-5-49
77. Pinto D, Delaby E, Merico D, Barbosa M, Merikangas A, Klei L et al (2014) Convergence of genes and cellular pathways dysregulated in autism spectrum disorders. Am J Hum Genet 94(5):677–694. https://doi.org/10.1016/j.ajhg.2014.03.018
78. (2015) Deciphering developmental disorders S. large-scale discovery of novel genetic causes of developmental disorders. Nature 519(7542):223–228. https://doi.org/10.1038/nature14135
79. Krumm N, Turner TN, Baker C, Vives L, Mohajeri K, Witherspoon K et al (2015) Excess of rare, inherited truncating mutations in autism. Nat Genet 47(6):582–588. https://doi.org/10.1038/ng.3303
80. Lim ET, Uddin M, De Rubeis S, Chan Y, Kamumbu AS, Zhang X et al (2017) Rates, distribution and implications of postzygotic mosaic mutations in autism spectrum disorder. Nat Neurosci 20(9):1217–1224. https://doi.org/10.1038/nn.4598
81. Chen R, Davis LK, Guter S, Wei Q, Jacob S, Potter MH et al (2017) Leveraging blood serotonin as an endophenotype to identify de novo and rare variants involved in autism. Mol Autism 8:14. https://doi.org/10.1186/s13229-017-0130-3
82. Toma C, Hervas A, Torrico B, Balmana N, Salgado M, Maristany M et al (2013) Analysis of two language-related genes in autism: a case-control association study of FOXP2 and CNTNAP2. Psychiatr Genet 23(2):82–85. https://doi.org/10.1097/YPG.0b013e32835d6fc6
83. Adegbola A, Gao H, Sommer S, Browning M (2008) A novel mutation in JARID1C/SMCX in a patient with autism spectrum disorder (ASD). Am J Med Genet A 146A(4):505–511. https://doi.org/10.1002/ajmg.a.32142
84. Li J, Wang L, Guo H, Shi L, Zhang K, Tang M et al (2017) Targeted sequencing and functional analysis reveal brain-size-related genes and their networks in autism spectrum disorders. Mol Psychiatry 22(9):1282–1290. https://doi.org/10.1038/mp.2017.140
85. Doan RN, Bae BI, Cubelos B, Chang C, Hossain AA, Al-Saad S et al (2016) Mutations in human accelerated regions disrupt cognition and social behavior. Cell 167(2):341–54 e12. https://doi.org/10.1016/j.cell.2016.08.071
86. D'Gama AM, Pochareddy S, Li M, Jamuar SS, Reiff RE, Lam AN et al (2015) Targeted DNA sequencing from autism Spectrum disorder brains implicates multiple genetic mechanisms. Neuron 88(5):910–917. https://doi.org/10.1016/j.neuron.2015.11.009
87. Jiang YH, Sahoo T, Michaelis RC, Bercovich D, Bressler J, Kashork CD et al (2004) A mixed epigenetic/genetic model for oligogenic inheritance of autism with a limited role for UBE3A. Am J Med Genet A 131(1):1–10. https://doi.org/10.1002/ajmg.a.30297
88. Nagarajan RP, Patzel KA, Martin M, Yasui DH, Swanberg SE, Hertz-Picciotto I et al (2008) MECP2 promoter methylation and X chromosome inactivation in autism. Autism Res 1(3):169–178. https://doi.org/10.1002/aur.24
89. Wong CC, Meaburn EL, Ronald A, Price TS, Jeffries AR, Schalkwyk LC et al (2014) Methylomic analysis of monozygotic twins discordant for autism spectrum disorder and related behavioural traits. Mol Psychiatry 19(4):495–503. https://doi.org/10.1038/mp.2013.41
90. Gregory SG, Connelly JJ, Towers AJ, Johnson J, Biscocho D, Markunas CA et al (2009) Genomic and epigenetic evidence for oxytocin receptor deficiency in autism. BMC Med 7:62. https://doi.org/10.1186/1741-7015-7-62

91. Labouesse MA, Dong E, Grayson DR, Guidotti A, Meyer U (2015) Maternal immune activation induces GAD1 and GAD2 promoter remodeling in the offspring prefrontal cortex. Epigenetics 10(12):1143–1155. https://doi.org/10.1080/15592294.2015.1114202
92. Hogart A, Leung KN, Wang NJ, Wu DJ, Driscoll J, Vallero RO et al (2009) Chromosome 15q11-13 duplication syndrome brain reveals epigenetic alterations in gene expression not predicted from copy number. J Med Genet 46(2):86–93. https://doi.org/10.1136/jmg.2008.061580
93. Lintas C, Sacco R, Persico AM (2016) Differential methylation at the RELN gene promoter in temporal cortex from autistic and typically developing post-puberal subjects. J Neurodev Disord 8:18. https://doi.org/10.1186/s11689-016-9151-z
94. Homs A, Codina-Sola M, Rodriguez-Santiago B, Villanueva CM, Monk D, Cusco I et al (2016) Genetic and epigenetic methylation defects and implication of the ERMN gene in autism spectrum disorders. Transl Psychiatry 6(7):e855. https://doi.org/10.1038/tp.2016.120
95. James SJ, Shpyleva S, Melnyk S, Pavliv O, Pogribny IP (2014) Elevated 5-hydroxymethylcytosine in the Engrailed-2 (EN-2) promoter is associated with increased gene expression and decreased MeCP2 binding in autism cerebellum. Transl Psychiatry 4:e460. https://doi.org/10.1038/tp.2014.87
96. Zhu L, Wang X, Li XL, Towers A, Cao X, Wang P et al (2014) Epigenetic dysregulation of SHANK3 in brain tissues from individuals with autism spectrum disorders. Hum Mol Genet 23(6):1563–1578. https://doi.org/10.1093/hmg/ddt547
97. Maunakea AK, Chepelev I, Cui K, Zhao K (2013) Intragenic DNA methylation modulates alternative splicing by recruiting MeCP2 to promote exon recognition. Cell Res 23(11):1256–1269. https://doi.org/10.1038/cr.2013.110
98. Neri F, Rapelli S, Krepelova A, Incarnato D, Parlato C, Basile G et al (2017) Intragenic DNA methylation prevents spurious transcription initiation. Nature 543(7643):72–77. https://doi.org/10.1038/nature21373
99. Lev Maor G, Yearim A, Ast G (2015) The alternative role of DNA methylation in splicing regulation. Trends Genet 31(5):274–280. https://doi.org/10.1016/j.tig.2015.03.002
100. Quesnel-Vallieres M, Dargaei Z, Irimia M, Gonatopoulos-Pournatzis T, Ip JY, Wu M et al (2016) Misregulation of an activity-dependent splicing network as a common mechanism underlying autism Spectrum disorders. Mol Cell 64(6):1023–1034. https://doi.org/10.1016/j.molcel.2016.11.033
101. Irimia M, Weatheritt RJ, Ellis JD, Parikshak NN, Gonatopoulos-Pournatzis T, Babor M et al (2014) A highly conserved program of neuronal microexons is misregulated in autistic brains. Cell 159(7):1511–1523. https://doi.org/10.1016/j.cell.2014.11.035
102. Li Q, Lee JA, Black DL (2007) Neuronal regulation of alternative pre-mRNA splicing. Nat Rev Neurosci 8(11):819–831. https://doi.org/10.1038/nrn2237
103. Vuong CK, Black DL, Zheng S (2016) The neurogenetics of alternative splicing. Nat Rev Neurosci 17(5):265–281. https://doi.org/10.1038/nrn.2016.27
104. Andrews SV, Ellis SE, Bakulski KM, Sheppard B, Croen LA, Hertz-Picciotto I et al (2017) Cross-tissue integration of genetic and epigenetic data offers insight into autism spectrum disorder. Nat Commun 8(1):1011. https://doi.org/10.1038/s41467-017-00868-y
105. Tsang SY, Ahmad T, Mat FW, Zhao C, Xiao S, Xia K et al (2016) Variation of global DNA methylation levels with age and in autistic children. Hum Genomics 10(1):31. https://doi.org/10.1186/s40246-016-0086-y
106. Liang S, Li Z, Wang Y, Li X, Yang X, Zhan X et al (2019) Genome-wide DNA methylation analysis reveals epigenetic pattern of SH2B1 in Chinese monozygotic twins discordant for autism Spectrum disorder. Front Neurosci 13:712. https://doi.org/10.3389/fnins.2019.00712
107. Feinberg JI, Bakulski KM, Jaffe AE, Tryggvadottir R, Brown SC, Goldman LR et al (2015) Paternal sperm DNA methylation associated with early signs of autism risk in an autism-enriched cohort. Int J Epidemiol 44(4):1199–1210. https://doi.org/10.1093/ije/dyv028

108. Schroeder DI, Schmidt RJ, Crary-Dooley FK, Walker CK, Ozonoff S, Tancredi DJ et al (2016) Placental methylome analysis from a prospective autism study. Mol Autism 7:51. https://doi.org/10.1186/s13229-016-0114-8
109. Wong CCY, Smith RG, Hannon E, Ramaswami G, Parikshak NN, Assary E et al (2019) Genome-wide DNA methylation profiling identifies convergent molecular signatures associated with idiopathic and syndromic autism in post-mortem human brain tissue. Hum Mol Genet 28(13):2201–2211. https://doi.org/10.1093/hmg/ddz052
110. Nardone S, Sams DS, Zito A, Reuveni E, Elliott E (2017) Dysregulation of cortical neuron DNA methylation profile in autism Spectrum disorder. Cereb Cortex 27(12):5739–5754. https://doi.org/10.1093/cercor/bhx250
111. Cheng Y, Li Z, Manupipatpong S, Lin L, Li X, Xu T et al (2018) 5-Hydroxymethylcytosine alterations in the human postmortem brains of autism spectrum disorder. Hum Mol Genet 27(17):2955–2964. https://doi.org/10.1093/hmg/ddy193
112. Wang T, Pan Q, Lin L, Szulwach KE, Song CX, He C et al (2012) Genome-wide DNA hydroxymethylation changes are associated with neurodevelopmental genes in the developing human cerebellum. Hum Mol Genet 21(26):5500–5510. https://doi.org/10.1093/hmg/dds394
113. Corley MJ, Vargas-Maya N, Pang APS, Lum-Jones A, Li D, Khadka V et al (2019) Epigenetic delay in the neurodevelopmental trajectory of DNA methylation states in autism Spectrum disorders. Front Genet 10:907. https://doi.org/10.3389/fgene.2019.00907
114. Waterland RA, Jirtle RL (2003) Transposable elements: targets for early nutritional effects on epigenetic gene regulation. Mol Cell Biol 23(15):5293–5300. https://doi.org/10.1128/MCB.23.15.5293-5300.2003
115. Weaver IC, Cervoni N, Champagne FA, D'Alessio AC, Sharma S, Seckl JR et al (2004) Epigenetic programming by maternal behavior. Nat Neurosci 7(8):847–854. https://doi.org/10.1038/nn1276
116. Dolinoy DC, Huang D, Jirtle RL (2007) Maternal nutrient supplementation counteracts bisphenol A-induced DNA hypomethylation in early development. Proc Natl Acad Sci U S A 104(32):13056–13061. https://doi.org/10.1073/pnas.0703739104
117. Ruiz-Hernandez A, Kuo CC, Rentero-Garrido P, Tang WY, Redon J, Ordovas JM et al (2015) Environmental chemicals and DNA methylation in adults: a systematic review of the epidemiologic evidence. Clin Epigenetics 7:55. https://doi.org/10.1186/s13148-015-0055-7
118. Sharp GC, Salas LA, Monnereau C, Allard C, Yousefi P, Everson TM et al (2017) Maternal BMI at the start of pregnancy and offspring epigenome-wide DNA methylation: findings from the pregnancy and childhood epigenetics (PACE) consortium. Hum Mol Genet 26(20):4067–4085. https://doi.org/10.1093/hmg/ddx290
119. Maccani JZ, Maccani MA (2015) Altered placental DNA methylation patterns associated with maternal smoking: current perspectives. Adv Genomics Genet 2015(5):205–214. https://doi.org/10.2147/AGG.S61518
120. McGowan PO, Sasaki A, D'Alessio AC, Dymov S, Labonte B, Szyf M et al (2009) Epigenetic regulation of the glucocorticoid receptor in human brain associates with childhood abuse. Nat Neurosci 12(3):342–348. https://doi.org/10.1038/nn.2270
121. Willcutt EG (2012) The prevalence of DSM-IV attention-deficit/hyperactivity disorder: a meta-analytic review. Neurotherapeutics 9(3):490–499. https://doi.org/10.1007/s13311-012-0135-8
122. Novik TS, Hervas A, Ralston SJ, Dalsgaard S, Rodrigues Pereira R, Lorenzo MJ et al (2006) Influence of gender on attention-deficit/hyperactivity disorder in Europe – ADORE. Eur Child Adolesc Psychiatry 15(Suppl 1):I15–I24. https://doi.org/10.1007/s00787-006-1003-z
123. Caci H, Asherson P, Donfrancesco R, Faraone SV, Hervas A, Fitzgerald M et al (2015) Daily life impairments associated with childhood/adolescent attention-deficit/hyperactivity disorder as recalled by adults: results from the European lifetime impairment survey. CNS Spectr 20(2):112–121. https://doi.org/10.1017/S1092852914000078

124. Gervin K, Nordeng H, Ystrom E, Reichborn-Kjennerud T, Lyle R (2017) Long-term prenatal exposure to paracetamol is associated with DNA methylation differences in children diagnosed with ADHD. Clin Epigenetics 9:77. https://doi.org/10.1186/s13148-017-0376-9
125. Mill J, Petronis A (2008) Pre- and peri-natal environmental risks for attention-deficit hyperactivity disorder (ADHD): the potential role of epigenetic processes in mediating susceptibility. J Child Psychol Psychiatry 49(10):1020–1030. https://doi.org/10.1111/j.1469-7610.2008.01909.x
126. Wilmot B, Fry R, Smeester L, Musser ED, Mill J, Nigg JT (2016) Methylomic analysis of salivary DNA in childhood ADHD identifies altered DNA methylation in VIPR2. J Child Psychol Psychiatry 57(2):152–160. https://doi.org/10.1111/jcpp.12457
127. Meijer M, Klein M, Hannon E, van der Meer D, Hartman C, Oosterlaan J et al (2020) Genome-wide DNA methylation patterns in persistent attention-deficit/hyperactivity disorder and in association with impulsive and callous traits. Front Genet 11:16. https://doi.org/10.3389/fgene.2020.00016
128. Weiss AL, Meijer M, Budeus B, Pauper M, Hakobjan M, Groothuismink J et al (2021) DNA methylation associated with persistent ADHD suggests TARBP1 as novel candidate. Neuropharmacology 184:108370. https://doi.org/10.1016/j.neuropharm.2020.108370
129. Miyake K, Miyashita C, Ikeda-Araki A, Miura R, Itoh S, Yamazaki K et al (2021) DNA methylation of GFI1 as a mediator of the association between prenatal smoking exposure and ADHD symptoms at 6 years: the Hokkaido study on environment and Children's health. Clin Epigenetics 13(1):74. https://doi.org/10.1186/s13148-021-01063-z
130. Rijlaarsdam J, Cecil CA, Walton E, Mesirow MS, Relton CL, Gaunt TR et al (2017) Prenatal unhealthy diet, insulin-like growth factor 2 gene (IGF2) methylation, and attention deficit hyperactivity disorder symptoms in youth with early-onset conduct problems. J Child Psychol Psychiatry 58(1):19–27. https://doi.org/10.1111/jcpp.12589
131. Park S, Lee JM, Kim JW, Cho DY, Yun HJ, Han DH et al (2015) Associations between serotonin transporter gene (SLC6A4) methylation and clinical characteristics and cortical thickness in children with ADHD. Psychol Med 45(14):3009–3017. https://doi.org/10.1017/S003329171500094X
132. Thurm A, Farmer C, Salzman E, Lord C, Bishop S (2019) State of the field: differentiating intellectual disability from autism Spectrum disorder. Front Psychiatry 10:526. https://doi.org/10.3389/fpsyt.2019.00526
133. Aspromonte MC, Bellini M, Gasparini A, Carraro M, Bettella E, Polli R et al (2019) Characterization of intellectual disability and autism comorbidity through gene panel sequencing. Hum Mutat 40(9):1346–1363. https://doi.org/10.1002/humu.23822
134. Krzyzewska IM, Maas SM, Henneman P, Lip KVD, Venema A, Baranano K et al (2019) A genome-wide DNA methylation signature for SETD1B-related syndrome. Clin Epigenetics 11(1):156. https://doi.org/10.1186/s13148-019-0749-3
135. Zaghi M, Broccoli V, Sessa A (2019) H3K36 methylation in neural development and associated diseases. Front Genet 10:1291. https://doi.org/10.3389/fgene.2019.01291
136. Sun W, Poschmann J, Cruz-Herrera Del Rosario R, Parikshak NN, Hajan HS, Kumar V et al (2016) Histone Acetylome-wide association study of autism Spectrum disorder. Cell 167(5):1385–97 e11. https://doi.org/10.1016/j.cell.2016.10.031

Epigenome-Wide Association Studies in Psychiatry: Achievements and Problems

18

Katherine Bassil, Nadya Ali, Ehsan Pishva, and Daniel L. A. van den Hove

Contents

Katherine Bassil and Nadya Ali have been contributed equally to this work.

K. Bassil · N. Ali · E. Pishva
Department of Psychiatry and Neuropsychology, School for Mental Health and Neuroscience (MHeNs), Maastricht University, Maastricht, The Netherlands
e-mail: k.bassil@maastrichtuniversity.nl; na.ali@student.maastrichtuniversity.nl; e.pishva@maastrichtuniversity.nl

D. L. A. van den Hove (✉)
Department of Psychiatry and Neuropsychology, School for Mental Health and Neuroscience (MHeNs), Maastricht University, Maastricht, The Netherlands

Laboratory of Translational Neuroscience, Department of Psychiatry, Psychosomatics and Psychotherapy, University of Würzburg, Würzburg, Germany
e-mail: d.vandenhove@maastrichtuniversity.nl

K. B. Michels (ed.), *Epigenetic Epidemiology*,
https://doi.org/10.1007/978-3-030-94475-9_18

Abstract

Genetic variation can only explain a small portion of risk to psychiatric disorders, including major depressive disorder, schizophrenia, and post-traumatic stress disorder. Epidemiological studies are increasingly showing a link between environmental factors and the development of various psychiatric disorders, mainly mediated by underlying epigenetic mechanisms. DNA methylation is one of the most studied epigenetic mechanisms in psychiatric disorders. Epigenome-wide association studies (EWAS) typically used to study changes in DNA methylation still face methodological challenges and limitations at both the fundamental, technical, and data analysis levels. In this chapter, we offer a brief overview of some EWAS studies in different psychiatric disorders and discuss the current challenges, pitfalls, and future considerations for this field.

Abbreviations

5hmC	5-hydroxymethylcytosine
5mC	5-methylcytosine
BP	Bipolar disorder
CpG	Cytosine-phosphate-guanine
CRISPR-dCas9	Clustered regularly interspaced short palindromic repeat-deficient Cas9
DLPFC	Dorsolateral prefrontal cortex
DMP	Differentially methylated position
DMR	Differentially methylated region
DNMT	DNA methyltransferase
EWAS	Epigenome-wide association studies
GABBR1	GABA-B receptor subunit 1 gene
GWAS	Genome-wide association studies
HAT	Histone acetyltransferase
HDAC	Histone deacetylase
MBD-Seq	Methyl-binding domain sequencing
MDD	Major Depressive Disorder
MeDIP-Seq	Methylated DNA immunoprecipitation sequencing
miRNA	microRNA
MOBP	Myelin-associated oligodendrocyte basic protein
mQTL	Methylation quantitative trait loci
MZ	Monozygotic
NGS	Next-generation sequencing
PMPS	Poly-methylomic profile scores
PTSD	Post-traumatic stress disorder
RRBS	Reduced representation bisulfite sequencing

RT-qPCR	Reverse transcription quantitative polymerase chain reaction
SCZ	Schizophrenia
TET	Ten-eleven translocation
WGBS	Whole-genome bisulphite sequencing

18.1 Introduction

Over the past two decades, a growing number of genome-wide association studies (GWAS) has shed light onto the etiology and progression of a range of psychiatric disorders. However, the estimates on the contribution of genetic variation obtained from these studies only explain a portion of the heritability (ranging from 40% to 60%) observed in psychiatric disorders such as major depressive disorder (MDD), anxiety disorders, and schizophrenia (SCZ) [1]. Whereas the symptomatology of these types of disorders is too heterogeneous to be linked to one gene as an underlying cause [2], a large number of epidemiological studies has revealed the importance of environmental factors, both pre- and postnatally, on the development of mental disorders [3, 4].

Epigenetic mechanisms are key players in mediating the complex interplay between genetic and environmental factors and, as such, might explain, at least in part, the missing heritability is psychiatric genetics. Epigenetic processes mediate dynamic alterations in the expression of genes without affecting the DNA sequence itself and are the result of developmental, environmental, or stochastic influences [2]. Epigenetic research offers a promising avenue in filling the knowledge gap in understanding the complex etiology and symptomatology of psychiatric disorders and could guide future research into the development of advanced treatment options.

This chapter provides a short overview of the recent literature on the involvement of epigenetic mechanisms in the development and course of psychiatric disorders, with a focus on the challenges associated with this line of research and perspectives for future research in this respect. The first section provides a brief summary of the most commonly studied epigenetic mechanisms, their relevance for brain development, and potential pitfalls associated with psychiatric epigenetics research. Subsequently, this chapter summarizes findings from recent epigenetic epidemiological research on psychiatric disorders, considering the potential challenges of this line of research and the extent to which these challenges have been addressed in existing studies. Finally, this chapter discusses future perspectives of psychiatric epigenetics research.

18.1.1 Epigenetic Mechanisms in Psychiatry

Epigenetic modifications consist of dynamic processes that can regulate gene expression, but without involving changes in the underlying DNA sequence [5]. They are crucial both for the development of the brain, as well as for experience-driven transcriptional changes [6, 7]. Such modifications affect every level of transcription, starting with regulating access of, e.g., the transcriptional machinery to the DNA by means of histone modifications or DNA methylation changes, as well as post-transcriptional modifications, mediated by, e.g., microRNAs (miRNAs), thereby affecting subsequent translation. Often, epigenetic regulation involves stable changes in the chromatin structure, which is comprised of DNA wrapped around a histone octamer. The nucleosome consists of two copies of the core histones H2A, H2B, H3, and H4 [5]. Chromatin structure is modified via remodeling enzymes such as histone acetyltransferases (HATs) and histone deacetylases (HDACs), causing either an open, i.e., active or closed, i.e., inactive chromatin state, respectively. HATs open up the chromatin by adding acetyl groups to lysine residues of the histone tail, which ultimately allows the transcriptional machinery access to the associated DNA sequence. HDACs, on the other hand, promote transcriptional repression by removing those acetyl groups [5]. In addition to acetylation, many other types of post-translational modifications have been found at histone residues, such as phosphorylation, sumoylation, methylation, and ubiquitination. The focus of this chapter will be on DNA methylation, mainly due to the fact that it is the most and best-studied epigenetic mechanism in relation to psychiatric disorders to date [8].

DNA methylation of the fifth position of cytosine is one of the best understood epigenetic modifications and targets the DNA bases directly [5, 9]. DNA is methylated via DNA methyltransferases (DNMTs), typically occurring at cytosine-phosphate-guanine (CpG) sites, resulting in the formation of 5-methylcytosine (5mC). DNA methylation is the most extensively studied mechanism of epigenetic regulation and has been implicated in the regulation of gene transcription, maintenance of genomic imprinting, X chromosome inactivation, chromatin structure, and the silencing of transposable elements [10]. Recent findings have shown the importance of another, closely related, type of DNA modification of the fifth position of cytosine, i.e., 5-hydroxymethylcytosine (5hmC). It occurs as a result of the oxidation of 5mC, catalyzed by the ten-eleven translocation (TET) family of enzymes. While 5hmC can be stable on its own, it can also contribute to the process of DNA demethylation, with two mechanisms that can convert 5hmC back into unmodified cytosine [11]. While 5hmC is detected in all tissue types, it is most abundant in the brain, suggesting its important role in brain function [10, 11]. Although an increasing number of studies suggest a role of 5hmC in the brain distinct from that of 5mC, traditionally used DNA methylation detection methods that make use of sodium bisulfite treatment cannot distinguish between 5mC and 5hmC. Whereas the great majority of those papers report the results as if reflecting true DNA methylation, i.e., 5mC, in reality, these data reflect the combined levels of 5mC and 5hmC.

In addition to histone modifications and DNA methylation, noncoding RNAs, specifically miRNAs, have been shown to exert an important role in epigenetic regulation by modifying protein levels [12]. MiRNAs are 18–25 nucleotide-long noncoding RNAs that regulate gene expression at the post-transcriptional stage, where they bind to the untranslated regions of mRNA molecules in order to suppress protein translation or to stimulate the breakdown of the associated mRNA [12]. Not only do miRNAs target key enzymes such as DNMTs, HDACs, and histone methyltransferases, but the expression of miRNAs is also regulated via epigenetic mechanisms, such as DNA methylation, RNA modifications and histone modification, creating a miRNA-epigenetic feedback loop [12, 13]. A single miRNA can regulate hundreds of mRNAs, while a single mRNA can also be targeted by numerous miRNAs, making the interpretation of findings from research in this area particularly challenging. Similarly to the area of DNA methylation, there are methodological challenges associated with the comparability of miRNA studies, such as the lack of consensus on the target tissue (such as peripheral blood, cerebrospinal fluid, or brain tissue), as well as the numerous techniques used for miRNA detection, such as quantitative reverse transcription polymerase chain reaction (RT-qPCR), microarray, Northern Blotting, and next-generation sequencing [14].

18.1.2 Methodological Considerations in Epigenome-Wide Studies

While the field of psychiatry epigenetics offers valuable insight into the underlying mechanisms of disorders extending beyond the genome, there are important challenges in epigenetics research that should be considered when conducting and interpreting epigenome-wide association studies (EWAS) to better understand how these changes impact key genes underlying psychiatric disorders [5, 15]. Such potential pitfalls can be found in every step of planning and conducting an EWAS study, starting with our basic understanding of the epigenome and study design, through specific methodological considerations regarding the laboratory analysis of samples and subsequent data pre-processing and statistical analyses. Understanding the limitations of epigenome-wide research is, therefore, crucial for both the careful interpretation of existing findings, as well as for the improvement of future studies in the field (Fig. 18.1).

18.1.2.1 Fundamental Understanding of the Epigenome and Study Design

Current understanding of the epigenome is still limited, which comes along with important considerations in terms of designing an EWAS and interpreting associated findings.

For example, mechanisms involving histone modifications are typically neglected likely due to challenges in either detecting them or in terms of interpreting the functional implications of the findings [16]. Accordingly, most of the current epigenome-wide research has focused on DNA methylation for both biological

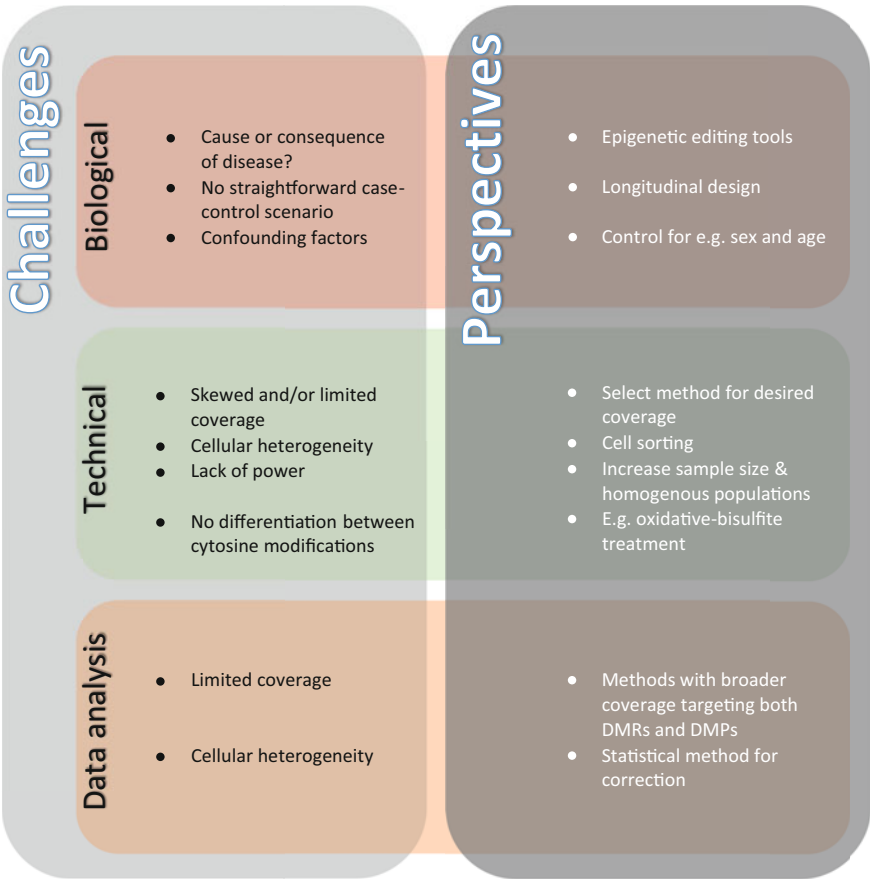

Fig. 18.1 Challenges of epigenetic-wide association studies (EWAS) in psychiatric disorders. Several challenges at the biological, technical, and data analysis levels aimed at unraveling epigenetic mechanisms, in particular DNA methylation in psychiatric disorders, remain. There are several avenues worth exploring in overcoming these challenges in order to allow optimal interpretation of the epigenetic data in psychiatry. See text for more details

and practical reasons. While most challenges in terms of detecting DNA methylation have been overcome by now (see below), the functional relevance of differential methylation is still often unclear though. As such, differential methylation could indicate increased or reduced transcription depending on the location of the methylated position, such as the promoter region, which is commonly investigated. As in certain cases, differential methylation at intergenic CpG shores and intra-genic CpG islands may be even more relevant for phenotypic variation when compared to methylation changes at CpG islands located in promoter regions, selecting the appropriate method to detect DNA methylation, e.g., through arrays or by means of sequencing, is crucial in view of the desired coverage and genomic location.

Furthermore, as the epigenome is not dynamic, a prospective study design often offers the most robust findings, allowing for the interrogation of epigenetic changes, e.g., in response to environmental variation, over time [15]. The timing of sample collection is crucial in this respect, as epigenetic modifications can be transient in certain cases. In the case of research into the underlying effects of treatments, as well as exposure to certain environments, such as military deployment [17], longitudinal studies can offer valuable insight into the epigenetic modifications in human populations. Analysis of disease-discordant monozygotic (MZ) twins offers an excellent tool in the study of epigenetic modifications associated with the disease phenotype, while its combination with a longitudinal design can provide valuable insight into the environmental variation impacting upon the etiology of the disease [15]. Moreover, in view of the important role of the early environment in programming adult mental health and disease, applying a longitudinal design including the assessment of developmental epigenetic changes, could be of great added value [15]. A major challenge to this type of design, however, lies in the lack of detailed phenotyping, a thorough collection of detailed information regarding exposure to environmental variation and repetitive follow-up and sampling throughout life [15]. Of note, generally, DNA methylation changes originating during early development may have higher inter-tissue concordance than those induced later in life [16].

In addition, researchers must be wary that the disease process itself as well as its (e.g., pharmacological) treatment can also cause epigenetic changes and causal inference into the etiology of the disease, and is therefore problematic when based solely on findings from human populations [15]. For this purpose, additional in vitro or in vivo animal studies are often needed [16].

18.1.2.2 Technological Challenges

A major challenge of epigenetic research in humans lies in the access to the tissue of interest, i.e., the brain, which is typically not available. Unlike genetic research where the sample tissue is not as relevant, in epigenetics, differences in, e.g., DNA methylation can be confounded by the sample's source and its cellular composition, especially in the case of whole blood samples [16]. While cells share the same DNA sequence, it is the epigenome that differentiates cell types and, therefore, cellular composition of the tissue of interest. This may in fact even differ between samples of the same type of tissue, which is of great importance in epigenetic analysis. Hence, taking into account potential differences in cell-type composition, either a priori by using cell sorting or post hoc using bioinformatics tools, is a must. Moreover, using peripheral tissue as a replacement of the brain can be problematic, especially that research shows that association between blood or saliva and the brain is quiet limited [18], and may respond differently to environmental challenges, such as exposure to stress [19]. One possible solution to optimize the translational ability of EWAS lies in the use of DNA methylation as a biomarker for mental illnesses [20].

The use of different techniques to detect DNA methylation raises an additional problem in comparability across studies [16]. Due to the demand for large sample sizes, researchers have to weigh the benefits of coverage versus precision. While

there are various methods available to capture methylated DNA, the 'golden standard' in current epigenetic research is the use of bisulfite treatment, followed by either an array- or next-generation sequencing-based detection [16, 21, 22]. So far, based on the delicate balance between precision and coverage, most labs involved in EWAS have made use of commercially available Illumina 450k and EPIC (850k) Methylation Beadchip microarrays. The 450k array, which is not available anymore, assessed DNA methylation at over 480 000 CG dinucleotides, mainly covering promoter regions [22], and providing an affordable option for large-scale epidemiological findings. Its successor, the Infinium MethylationEPIC array, targets over 850 000 CpGs. While being of clear added value, next to the limited coverage, these platforms lack the possibility to thoroughly assess methylation at non-CpG sites [20]. Recently, more frequently, next-generation sequencing (NGS) is being used, which allows for flexibility in this respect, which evidently is more costly and time consuming. Whole-genome bisulfite sequencing (WGBS) offers almost full coverage [16, 21, 22]. Reduced representation bisulfite sequencing (RRBS), while granting single-base resolution as in WGBS, has lower coverage, including the great majority of promoters and CpG islands, implying a lower number of reads necessary to yield accurate sequencing, in addition to lower costs and processing time compared with WGBS. Other approaches, such as methylated DNA immunoprecipitation sequencing (MeDIP-Seq) or methyl-binding domain sequencing (MBD-Seq), make use of means to capture methylated pieces of DNA prior to sequencing. It is worth noting that while there is high correlation between different platforms, which is related to the fact that most of the genome is either simply methylated or unmethylated, there may be substantial differences within loci that display intermediate methylation levels [16].

As indicated above, accumulating evidence suggests an important role for other cytosine modifications such as 5hmC in the human brain. However, the great majority of EWAS published to date made use of sodium bisulfite-treated DNA, which does not allow to discriminate between 5mC and 5hmC, which may lead to incorrect interpretation of associated findings. Making use of oxidative-bisulfite DNA treatment (parallel to classical bisulfite treatment) now allows for simultaneously assessing DNA methylation and hydroxymethylation [23].

18.1.2.3 Data Analysis

There are various data analysis strategies available for analysing and interpreting epigenetic data and, similarly to the methodological problems raised earlier, to date, no clear consensus on the most optimal approach exists [15]. Most of the current studies in the literature have investigated individual CpGs to identify differentially methylated positions (DMPs), as that is also the method most suitable for the Illumina array-based platforms, owing to their modest coverage. Another approach is to explore differentially methylated regions (DMRs), which is based on the assumption that several differentially methylated cytosines in close proximity with each other are likely to affect chromatin formation and thus the transcription of the accompanying gene(s) [15]. Methods with greater coverage may be more suitable for the investigation of DMRs, as there is not clear consensus yet whether the

number of CpGs covered by the Illumina arrays is sufficient for this type of analysis. The advantage of DMR analysis is the reduced risk of false-positives due to artifacts affecting single CpGs, as well as higher power because of the reduced multiple testing error [15]. Additionally, DMRs may be easier to be interpreted biologically. It should be noted though, that differential methylation at single CpG methylation sites can also have important functional consequences. In addition to these two commonly used methods, Mendelian randomization can be useful for causal inference and is being applied more regularly nowadays. As such, a methylation quantitative trait loci (mQTL) analysis aims at identifying genetic variants that affect DNA methylation patterns [24], knowledge of which is of extreme importance, as a certain degree of epigenetic variation may be caused by genetic variation. Another critical point to consider is accounting for cellular heterogeneity, especially in blood, which is crucial in EWAS. For instance, one study showed that cellular composition accounts for the majority of the detected variability in DNA methylation. It appears that DNA methylation profiles were both cell type- and age-dependent. This can be corrected for using a statistical approach as described in Ref. [25].

18.2 Epigenetic Profiles in Psychiatric Disorders

While the following sections include a discussion on numerous EWAS performed to date, it should be noted that this does not concern a systematic review of the available literature in this respect, as its aim is to highlight strengths and limitations of certain approaches in this respect.

18.2.1 Major Depressive Disorder

MDD is one of the most heterogeneous and prevalent psychiatric disorders and findings from EWAS studies suggest a wide variety of changes in epigenetic regulation which may underlie MDD [26]. A recent longitudinal study utilizing the 450k Illumina platform and examining tissue from the dorsolateral prefrontal cortex (DLPFC) of 608 participants showed epigenome-wide significant association between DNA methylation changes in certain genes and late-life MDD [27]. The most significant association found in this study was observed for the *YOD1/PFKB2* locus. *YOD1* has been found to be involved in the maintenance of the correct conformation of proteins, more specifically related to inflammatory responses. Although this study made use of a tissue-specific analysis and a longitudinal design, it is worth noting that the researchers employed tissue bulk analysis, which was later corrected for cell-type composition [27]. This notion is relevant, as cell-type composition can be different between, e.g., patients and controls in EWAS studies, e.g., due to cell loss, when assessing brain tissue, or inflammation, when assessing blood. This challenges the correct interpretation of results, as cell-type-specific modifications in one cell-type could, for example, be masked by changes in another.

In another study, on a cohort of 724 MZ Danish twins, making use of whole blood samples and the Illumina 450k platform, Starnawska and colleagues showed associations between DNA methylation and MDD for the gene encoding neuropsin, which is involved in synaptogenesis and has previously been implicated in schizophrenia and bipolar disorder (BP) [28]. The authors also found differential methylation for the *DAZAP2* gene, which is known for inducing stress granule formation. Although the study utilized peripheral blood samples, it has the strong advantage of a discordant MZ design, as well as a relatively large sample size [28]. Additionally, the researchers investigated depression severity amongst the general population, as opposed to clinical cases. An additional strength of the study is that blood cell composition proportions were estimated using flow cytometry for a part of the individuals ($n = 471$). In another study using a relatively large sample size ($N = 844$), followed by a replication study ($N = 1339$), the relationship between umbilical cord DNA methylation and maternal depression throughout pregnancy was investigated, using the Illumina 450k platform [29]. Results from the first cohort showed a relationship between maternal depression at any point during pregnancy and 7 DMRs, which was, however, not replicated in the second cohort. The DMRs were located within genes related to brain development and the formation of the nervous system, as well as the *LYNX1* gene, which was previously shown to be hypermethylated in the hippocampus of patients with MDD [29].

18.2.2 Suicide

Suicide is the fourth leading cause of death amongst 15–29 year old, with nearly one million deaths per year globally [30]. The heterogeneity of disorders accompanying suicide completion poses additional challenges in this line of research, where researchers investigating the epigenetic changes associated with suicide may either focus on a specific disorder, exclude individuals with a particular underlying disorder, or in some cases compare disorders. Additionally, depending on whether suicide attempts or suicide completion is being investigated, the sample methods differ, with tissue-specific samples being the typical choice in the latter case. It should be noted that while a history of suicide attempts is the strongest predictor of suicide completion, some studies focus on comparing individuals with a history of suicidal ideation versus healthy controls, whereas other studies examine post-mortem tissue in suicide completers. As there may be differences in epigenetic modifications between the two groups, findings should be interpreted with caution.

Due to high suicide attempt rates amongst individuals with BP, it is the most studied comorbid disorder in individuals with a history of suicide attempts [31]. In one such study, the authors investigated the prefrontal cortex of 23 individuals with BP who died of suicide, 27 who died of other causes, and 31 non-psychiatric controls [31]. Overall, BP subjects had more hypomethylated DMRs compared to controls, pooled for cause of death, whereas within the BP group, individuals who died of suicide showed increased methylation than those who did not. Of specific interest, *ARHGEF38* was hypomethylated in BP associated with suicide, which, while

relatively unknown, is believed to be involved in the GTPase cycle [31]. It should be noted that those findings were primarily driven by males with BP who died of suicide, which suggests that sex differences may be a factor that should be included in future studies. This is especially relevant since there are known sex differences in the rates and method of suicide, with males being more likely to die of suicide and opting for violent methods, whereas females tend to gravitate towards non-violent tools [32, 33]. Additionally, aggression and impulsivity are known risk factors for suicide [34], which may also be relevant as there have been sex differences demonstrated in aggressive behavior, impulsivity, and violent acts [35]. The difference in suicidal ideation and suicidal completion, therefore, with the former being more common in women, whereas the latter is more common in men, may suggest distinct molecular changes associated with each. Work on suicide is further complicated by the heterogeneity of disorders accompanying suicidal ideation, such as MDD, BP, SCZ, anxiety disorders, substance abuse, and personality disorders [31, 36, 37]. Finally, the authors showed differences in methylation in axonal guidance signaling pathways, cardiac beta-adrenergic signaling, and opioid signaling [31].

Another study by Jokinen and colleagues [37] investigated DNA methylation in whole blood of individuals with a history of suicide attempt. The researchers used both the Illumina 450k and the EPIC 850k platforms to investigate DMPs associated with suicide attempts. They excluded factors such as schizophrenia spectrum disorders, intravenous drug abuse, dementia, and mental retardation ($N = 88$ for cohort 1; $N = 129$ in the second cohort and $N = 93$ for the third cohort). Their findings suggest that individuals with a higher risk of suicide attempt had reduced levels of methylation in the promoter region of the *CRH* gene, which is crucial in HPA-axis regulation [37]. A strength of this study is that the initial findings were also replicated in the following cohort of adolescents. Yet again, most of the participants in the study were previously treated with antidepressants which also influences DNA methylation profiles [38], and there was no clear distinction in diagnosis in view of the underlying psychiatric illness(es).

In a comparison of depressed individuals who committed suicide versus non-psychiatric sudden-death controls ($N = 75$ in total), using brain tissue from Brodmann areas 11 and 25, Murphy and colleagues [39] showed methylation changes in several genes. The researchers observed hypomethylation in both cortical regions in MDD suicide cases for the *PSORS1C3* gene, which is thought to be involved in the regulation of nearby immune system-related genes. In addition, two other DMRs related to antigen processing (*TAPBP*) and mitochondrial ATP synthase function (*ATP5G2*) were identified. It is worth to note that a limitation of this kind of design is the lack of medication data, which can be especially crucial when comparing groups with an underlying psychiatric illness with controls, as some of the methylation changes that are identified may be related to treatment or even result from the disease progress itself, as opposed to be causally associated with suicidal ideation specifically. A design including a group of individuals with MDD who died of other causes could have tackled those limitations. In addition to individuals with MDD and BP, individuals with SCZ are at increased risk for committing suicide.

Bani-Fatemi and colleagues conducted two studies in peripheral blood investigating suicide attempts in individuals with SCZ, in the first of which they did not find changes in methylation between the group with a history of suicide attempts and those without [40]. Findings from the second study did show hypermethylation in both *SLC20A1*, a sodium-dependent phosphate transporter, and *SMPD2*, which encodes for the enzyme sphingomyelin phosphodiesterase 2, in those individuals displaying suicidal ideation. In both these studies, no records on medication were available, and a relatively small sample size was used ($N = 123$ and $N = 107$, respectively).

In addition, in another study investigating suicidal behavior in individuals with BP, the authors included not only a typical DMP and DMR analysis, but also correlated DNA methylation age with suicidal behavior [41]. In this study, individuals with a history of suicidal behavior displayed hypomethylation in *MPP4*, which is known to regulate the activity of membrane calcium ATPases, and *TBC1D16*, which represents a known activator of Rab4a involved in cell growth and survival, as well as hypermethylation in *NUP133*, which is known to be involved in spindle assembly and nuclear mRNA export [41]. Furthermore, age-related signatures of DNA methylation showed a weaker correlation in DNA methylation age and chronological age when compared to controls, where the DNA of individuals with a history of suicidal behavior was hypothesized to exhibit accelerated (epigenetic) aging. This study was the first to investigate the relationship between suicidal ideation and DNA methylation age, although it is worth to note that the authors did not specifically control for factors that have previously been associated with changes in DNA methylation age [42].

18.2.3 Schizophrenia

SCZ is a neurodevelopmental disorder characterized by psychosis and altered cognitive function, which was one of the first targets in the investigation of both genetic risk factors as well as epigenetic changes associated with disease risk and progression [43]. Initial twin and family studies have shown the heritability of the disease, although it does not fully explain its etiology. Together with MDD, it is also one of the disorders for which there are studies with relatively large sample sizes. In one such project, Hannon and colleagues [44] investigated methylation changes in three cohorts, the first of which was a discovery cohort ($N = 675$), followed by a replication one ($N = 847$), and an MZ pair cohort ($N = 96$ pairs). Findings show the top-ranked group of pathways associated with SCZ was related to immune functioning, with the second-ranked group involved in neuronal proliferation and brain development [45]. The large sample size, multiple phases and the inclusion of an MZ cohort make this project an excellent example of a study that aimed to address the methodological challenges of this line of research.

A more recent study by Wakeys and colleagues [46] examined 171 individuals, of which 57 diagnosed with SCZ, 59 with BP, and 55 healthy controls. In addition to the DMP analyses, the researchers computed poly-methylomic profile scores

(PMPS), based on EWAS data and clinical status. All five PMPS computed were associated with SCZ, where individuals with SCZ showed heightened PMPS relative to controls. While similarly to the Hannon study [45], the researchers used peripheral blood for the methylation analyses and the sample size for this cohort was not very large, a strength of this study was the inclusion of polygenic risk score analysis, as well as the comparison between individuals with SCZ not only to controls, but also to BP patients.

A big challenge in the identification of epigenetic changes in SCZ is defining whether DNA methylation changes are a cause or a consequence of the disease [47]. That could also be due to the fact that antipsychotics also carry an effect on DNA methylation profiles [48, 49], and hence any observed EWAS association in SCZ could also be due to the effects of antipsychotics. That being said, future EWAS studies must work to include both medicated and drug-naïve patients to detect true disease phenotypes.

18.2.4 Post-Traumatic Stress Disorder

Post-traumatic stress disorder (PTSD) is a debilitating psychiatric disorder that typically develops as a result of direct or indirect exposure to a traumatic event [17]. While most individuals are resilient to the long-term effects of trauma exposure, some develop PTSD with no effective pharmacotherapy being developed to date. Epigenetic findings from this area can, therefore, not only shed light on risk and resilience factors, but also on the underlying mechanisms of the disorder directing further research into advanced treatment options. Due to the nature of the disorder and its cause, researchers often select military cohorts as their sample, often comprised exclusively of male individuals. The advantage of such a design is the possibility to analyze the methylome prior to, as well as in the short and long run after exposure to trauma, as well as the possibility to identify potential predictors of resilience. Additionally, the type of stressor that the individuals are exposed to is typically similar of nature, often related to combat. Evidently, there are also disadvantages related to the use of military samples, such as the higher number of male individuals, the potential underreporting of symptoms, as well as the relatively homogenous type of trauma. Samples drawn from the general population could shed more light on epigenetic mechanisms underlying PTSD following various types of traumas, as well as over different periods of time. Thus, while military cohorts may be more convenient from a methodological perspective, a well-devised longitudinal study of a community sample could also provide valuable insight into epigenetic modifications associated with the disorder.

In a large study including three military cohorts, Snijders and colleagues [17] investigated the peripheral blood of soldiers exposed to trauma, comparing those who did develop PTSD symptoms with resilient individuals and those not exposed to trauma. Participants with PTSD symptoms showed an association between DNA methylation profiles and post-deployment symptoms [17]. The top replicating DMP

represented an intergenic site near a gene (*SPRY4*) that codes for a member of the Sprouty proteins involved in the inhibition of tyrosine kinase signaling. Additionally, the strongest association for the DMR analysis was in a region related to immune functioning [17]. It should be noted that the individuals in this study were males only, with those who developed PTSD having been exposed to more traumatic events. Another study that focused on adolescents ($N = 39$), the majority of which female (84.6%), showed two DMPs associated with PTSD symptom severity, one of which was in the *MAML* gene, which has been implicated in neuronal plasticity [50]. In addition, a DMR in the *MOBP* (myelin-associated oligodendrocyte basic protein) gene was identified, which has previously been associated with PTSD in the literature.

While the previous two studies both investigated peripheral blood, Logue and colleagues [51] used samples from two blood-based cohorts and a brain bank cohort ($N = 513$; replication cohort $N = 1253$; and brain tissue (prefrontal cortex) cohort $N = 72$). Analysing DMPs and DMRs, the authors found a negative correlation between the effect size estimations across the blood-based cohort and the brain bank samples, suggesting that effects investigated using peripheral blood should be interpreted with caution. This difference was mainly driven by one site in *OR2AG1*, which interestingly displayed the largest effect size in both blood and brain, but in the opposite direction. *OR2AG1* is a gene coding for an olfactory receptor, whose activation carries consequences for serotonin release [52]. Additionally, in the brain bank cohort, a probe in the *CHST11* gene, which was also among the top 10 results in the discovery cohort, was associated with PTSD. This gene is involved in, e.g., neuronal plasticity, fear learning, and neuroinflammation. Finally, an epigenome-wide association in the *G0/G1 Switch 2 (GOS2)* gene was identified, where higher methylation was associated with a PTSD diagnosis. Previous findings have shown cortisol to suppress G0S2 [53] and its expression has been linked to PTSD in the literature [54, 55]. It should be noted that while this study employs a much larger sample size and has the strength of multi-tissue type comparison, the researchers used different platforms for the discovery and replication cohort, namely Illumina 450k and EPIC 850k arrays, respectively.

In another recent study, Katrinli and colleagues [56] investigated methylation changes associated with PTSD in two large cohorts ($N = 554$ and 780), including correlating the levels of DNA methylation and gene expression. The authors found two DMRs associated with PTSD, for one of which (*HLA-DPB1*) the direction of the effect was in the opposite direction for the two cohorts, with this gene having been previously associated with PTSD in the literature [17] and is thought to be related to immune system dysregulation [57, 58]. The function of the other gene (*SPATC1L*) is thought to be related to the protection of cells from cell death induced by DNA-damaging alkylating agents [59], which are by-products of normal cellular function, as well as oxidative stress and chronic inflammation [60]. Similar to the previously discussed multi-cohort project, however, this study also used different platforms for the two cohorts investigated.

18.2.5 Anxiety Disorders

Due to the heterogeneity of anxiety disorders, the scope of studies that investigate epigenetic mechanisms associated with anxiety is relatively wide. Research in this area, therefore, varies both in terms of the type of anxiety investigated, with some studies including individuals with panic attacks, as well as in view of the stage of life at which they are being studied, where for example, some studies focus on transgenerational effects of maternal anxiety on the offspring.

Using a population-based ($N = 1522$) and a clinical cohort ($N = 300$), Emeny and colleagues [61] identified an association between severe anxiety and increased methylation at a CpG site located in the promoter of *ASB1*, a gene associated with ubiquitin degradation pathways, as well as scaffolding, neurogenesis, and neuroprotection [62]. While the population-based sample as well as the large sample size for both cohorts is a particular strength of this study, it should be noted that the authors used two different questionnaires to assess anxiety and there was no clear distinction between panic disorder and anxiety symptoms across the cohorts. In addition to studies that investigate epigenetic changes associated with anxiety in adults, some focus on the effects of maternal anxiety on DNA methylation in newborns. In one such study, Vangeel and colleagues [63] evaluated changes in DNA methylation derived from the umbilical cord blood of newborns whose mothers exhibited more symptoms of anxiety and correlated that with the cortisol awakening response measured at 2, 4, and 12 months of age of the infant. Of the 10 DMRs, the top DMR was found in the GABA-B receptor subunit 1 gene (*GABBR1*) and associated with prenatal anxiety especially in male newborns. *GABBR1* methylation profile was also associated with newborn cortisol levels at 4 months. These findings highlight the role of *GABBR1* in prenatal anxiety, particularly in influencing neuronal plasticity in the face of environmental stressors [64]. Another important observation points towards taking into account gender effects when analyzing differential methylation profiles [65].

18.3 Concluding Remarks

Epigenome-wide approaches have garnered much attention in the recent years, with the primary focus being on DNA methylation. While there is increasing interest in the role of epigenetic dysregulation in mental health and disease, results from EWAS studies should be interpreted with caution (Fig. 18.1). Most importantly, for various reasons, several studies still make use of low sample sizes. In addition, studies often make use of bulk brain tissue, which does not always allow correcting for cellular heterogeneity, an issue that can be overcome by, for example, cell sorting. Moreover, particularly when assessing DNA derived from brain tissue, specific approaches that are able to discriminate between 5mC and 5hmC should be employed. Future EWAS studies should also include sex as a determining factor owing to known sex differences in epigenetic mechanisms. Another consideration would be to include homogenous populations of a certain disorder, or addressing specific (e.g.,

behavioral) phenotypes (e.g., panic attacks) so as not to risk missing potential associations. In the near future, the field could benefit from validating candidate signatures, by for instance employing (e.g., CRISPR-dCas9-based) epigenetic editing tools, to determine whether signatures identified in EWAS in fact represent a causal association. Moreover, multi-omics approaches centered around an integrative analysis of various layers including genetic, epigenetic, proteomic, and metabolomic data show great promise in the development of novel treatment strategies and diagnostic and prognostic biomarkers for psychiatric disorders.

References

1. Thibaut F (2019) Epigenetics: the missing link between genes and psychiatric disorders? Dialogues Clin Neurosci 21(4):337–338
2. Kular L, Kular S (2018) Epigenetics applied to psychiatry: clinical opportunities and future challenges. Psychiatry Clin Neurosci 72(4):195–211
3. Susser E et al (2006) Psychiatric epidemiology: searching for the causes of mental disorders. Oxford University Press
4. Cooper B (2001) Nature, nurture and mental disorder: old concepts in the new millennium. Br J Psychiatry 178(S40):s91–s101
5. Mahgoub M, Monteggia LM (2013) Epigenetics and psychiatry. Neurotherapeutics 10(4): 734–741
6. Kravitz SN, Gregg C (2019) New subtypes of allele-specific epigenetic effects: implications for brain development, function and disease. Curr Opin Neurobiol 59:69–78
7. Nagy C, Vaillancourt K, Turecki G (2018) A role for activity-dependent epigenetics in the development and treatment of major depressive disorder. Genes Brain Behav 17(3):e12446
8. Starnawska A, Demontis D (2021) Role of DNA methylation in mediating genetic risk of psychiatric disorders. Front Psychiatry 12:596821
9. Smith ZD, Meissner A (2013) DNA methylation: roles in mammalian development. Nat Rev Genet 14(3):204–220
10. Wen L, Tang F (2014) Genomic distribution and possible functions of DNA hydroxymethylation in the brain. Genomics 104(5):341–346
11. Rustad SR, Papale LA, Alisch RS (2019) DNA methylation and hydroxymethylation and behavior. Curr Top Behav Neurosci 42:51–82
12. Yao Q, Chen Y, Zhou X (2019) The roles of microRNAs in epigenetic regulation. Curr Opin Chem Biol 51:11–17
13. Van den Hove DL et al (2014) Epigenetically regulated microRNAs in Alzheimer's disease. Neurobiol Aging 35(4):731–745
14. Hunt EA et al (2015) MicroRNA detection: current technology and research strategies. Annu Rev Anal Chem 8(1):217–237
15. Mill J, Heijmans BT (2013) From promises to practical strategies in epigenetic epidemiology. Nat Rev Genet 14(8):585–594
16. Heijmans BT, Mill J (2012) Commentary: the seven plagues of epigenetic epidemiology. Int J Epidemiol 41(1):74–78
17. Snijders C et al (2020) Longitudinal epigenome-wide association studies of three male military cohorts reveal multiple CpG sites associated with post-traumatic stress disorder. Clin Epigenetics 12(1):11
18. Lin D et al (2018) Characterization of cross-tissue genetic-epigenetic effects and their patterns in schizophrenia. Genome Med 10(1):1–12
19. Jones MJ, Moore SR, Kobor MS (2018) Principles and challenges of applying epigenetic epidemiology to psychology. Annu Rev Psychol 69:459–485

20. Barker ED, Roberts S, Walton E (2019) Hidden hypotheses in 'hypothesis-free' genome-wide epigenetic associations. Curr Opin Psychol 27:13–17
21. Khodadadi E et al (2021) Current Advances in DNA methylation analysis methods. Biomed Res Int 2021:8827516
22. Kurdyukov S, Bullock M (2016) DNA methylation analysis: choosing the right method. Biology 5(1):3
23. Zeng T et al (2015) Detection of 5-methylcytosine and 5-hydroxymethylcytosine in DNA via host–guest interactions inside α-hemolysin nanopores. Chem Sci 6(10):5628–5634
24. Gao X et al (2017) The impact of methylation quantitative trait loci (mQTLs) on active smoking-related DNA methylation changes. Clin Epigenetics 9(1):87
25. Jaffe AE, Irizarry RA (2014) Accounting for cellular heterogeneity is critical in epigenome-wide association studies. Genome Biol 15(2):1–9
26. Lin E, Tsai S-J (2019) Epigenetics and depression: an update. Psychiatry Investig 16(9):654
27. Hüls A et al (2020) Association between DNA methylation levels in brain tissue and late-life depression in community-based participants. Transl Psychiatry 10(1):262
28. Starnawska A et al (2019) Epigenome-wide association study of depression symptomatology in elderly monozygotic twins. Transl Psychiatry 9(1):214
29. Viuff AC et al (2018) Maternal depression during pregnancy and cord blood DNA methylation: findings from the Avon longitudinal study of parents and children. Transl Psychiatry 8(1):244
30. Organization, W.H (2014) Preventing suicide: a global imperative. World Health Organization
31. Gaine ME et al (2019) Differentially methylated regions in bipolar disorder and suicide. Am J Med Genet B Neuropsychiatr Genet 180(7):496–507
32. Fox KR et al (2018) Examining the role of sex in self-injurious thoughts and behaviors. Clin Psychol Rev 66:3–11
33. Skogman K, Alsén M, Öjehagen A (2004) Sex differences in risk factors for suicide after attempted suicide. Soc Psychiatry Psychiatr Epidemiol 39(2):113–120
34. Coryell W et al (2018) Aggression, impulsivity and inflammatory markers as risk factors for suicidal behavior. J Psychiatr Res 106:38–42
35. Cross CP, Copping LT, Campbell A (2011) Sex differences in impulsivity: a meta-analysis. Psychol Bull 137(1):97
36. Hayashi N et al (2012) Post-hospitalization course and predictive signs of suicidal behavior of suicidal patients admitted to a psychiatric hospital: a 2-year prospective follow-up study. BMC Psychiatry 12(1):186
37. Jokinen J et al (2018) Epigenetic changes in the CRH gene are related to severity of suicide attempt and a general psychiatric risk score in adolescents. EBioMedicine 27:123–133
38. Li M et al (2019) What do DNA methylation studies tell us about depression? A systematic review. Transl Psychiatry 9(1):1–14
39. Murphy T et al (2017) Methylomic profiling of cortex samples from completed suicide cases implicates a role for PSORS1C3 in major depression and suicide. Transl Psychiatry 7(1):e989–e989
40. Bani-Fatemi A et al (2018) Epigenome-wide association study of suicide attempt in schizophrenia. J Psychiatr Res 104:192–197
41. Jeremian R et al (2017) Investigation of correlations between DNA methylation, suicidal behavior and aging. Bipolar Disord 19(1):32–40
42. Dhingra R et al (2018) DNA methylation age—Environmental influences, health impacts, and its role in environmental epidemiology. Curr Environ Health Rep 5(3):317–327
43. Viana J et al (2017) Schizophrenia-associated methylomic variation: molecular signatures of disease and polygenic risk burden across multiple brain regions. Hum Mol Genet 26(1): 210–225
44. Hannon E, Dempster EL, Mansell G, Burrage J, Bass N, Bohlken MM, Corvin A, Curtis CJ, Dempster D, Di Forti M, Dinan TG, Donohoe G, Gaughran F, Gill M, Gillespie A, Gunasinghe C, Hulshoff HE, Hultman CM, Johansson V, Kahn RS, Kaprio J, Kenis G, Kowalec K, MacCabe J, McDonald C, McQuillin A, Morris DW, Murphy KC, Mustard CJ, Nenadic I,

O'Donovan MC, Quattrone D, Richards AL, Rutten BP, St Clair D, Therman S, Toulopoulou T, Van Os J, Waddington JL, Wellcome Trust Case Control Consortium (WTCCC); CRESTAR consortium, Sullivan P, Vassos E, Breen G, Collier DA, Murray RM, Schalkwyk LS, Mill J (2021) DNA methylation meta-analysis reveals cellular alterations in psychosis and markers of treatment-resistant schizophrenia. Elife Feb 26;10:e58430. https://doi.org/10.7554/eLife.58430. PMCID: PMC8009672

45. Hannon E et al (2016) An integrated genetic-epigenetic analysis of schizophrenia: evidence for co-localization of genetic associations and differential DNA methylation. Genome Biol 17(1):176
46. Watkeys OJ et al (2020) Derivation of poly-methylomic profile scores for schizophrenia. Prog Neuro-Psychopharmacol Biol Psychiatry 101:109925
47. Cariaga-Martinez A, Saiz-Ruiz J, Alelú-Paz R (2016) From linkage studies to epigenetics: what we know and what we need to know in the neurobiology of schizophrenia. Front Neurosci 10:202
48. Ovenden ES et al (2018) DNA methylation and antipsychotic treatment mechanisms in schizophrenia: progress and future directions. Prog Neuro-Psychopharmacol Biol Psychiatry 81:38–49
49. Smigielski L et al (2020) Epigenetic mechanisms in schizophrenia and other psychotic disorders: a systematic review of empirical human findings. Mol Psychiatry 25(8):1718–1748
50. Sheerin CM et al (2021) Epigenome-wide study of posttraumatic stress disorder symptom severity in a treatment-seeking adolescent sample. J Trauma Stress 34:607–615
51. Logue MW et al (2020) An epigenome-wide association study of posttraumatic stress disorder in US veterans implicates several new DNA methylation loci. Clin Epigenetics 12(1):1–14
52. Maßberg D, Hatt H (2018) Human olfactory receptors: novel cellular functions outside of the nose. Physiol Rev 98:1739–1763
53. Stimson RH et al (2017) Acute physiological effects of glucocorticoids on fuel metabolism in humans are permissive but not direct. Diabetes Obes Metab 19(6):883–891
54. Daskalakis NP et al (2014) Expression profiling associates blood and brain glucocorticoid receptor signaling with trauma-related individual differences in both sexes. Proc Natl Acad Sci 111(37):13529–13534
55. Bam M et al (2016) Dysregulated immune system networks in war veterans with PTSD is an outcome of altered miRNA expression and DNA methylation. Sci Rep 6(1):1–13
56. Katrinli S et al (2021) PTSD is associated with increased DNA methylation across regions of HLA-DPB1 and SPATC1L. Brain Behav Immun 91:429–436
57. Michopoulos V et al (2017) Inflammation in fear-and anxiety-based disorders: PTSD, GAD, and beyond. Neuropsychopharmacology 42(1):254–270
58. Eraly SA et al (2014) Assessment of plasma C-reactive protein as a biomarker of posttraumatic stress disorder risk. JAMA Psychiatry 71(4):423–431
59. Fry RC et al (2008) Genomic predictors of interindividual differences in response to DNA damaging agents. Genes Dev 22(19):2621–2626
60. La D, Upton P, Swenberg J (2010) Carcinogenic alkylating agents. Elsevier
61. Emeny RT et al (2018) Anxiety associated increased CpG methylation in the promoter of Asb1: a translational approach evidenced by epidemiological and clinical studies and a murine model. Neuropsychopharmacology 43(2):342–353
62. Leussis MP, Madison JM, Petryshen TL (2012) Ankyrin 3: genetic association with bipolar disorder and relevance to disease pathophysiology. Biol Mood Anxiety Disord 2(1):1–13
63. Vangeel EB et al (2017) Newborn genome-wide DNA methylation in association with pregnancy anxiety reveals a potential role for GABBR1. Clin Epigenetics 9(1):1–12
64. Bains JS, Cusulin JIW, Inoue W (2015) Stress-related synaptic plasticity in the hypothalamus. Nat Rev Neurosci 16(7):377–388
65. Bale TL (2011) Sex differences in prenatal epigenetic programing of stress pathways. Stress 14(4):348–356

Epigenetic Epidemiology of Obesity and Type 2 Diabetes

19

Charlotte Ling, Sabrina Ruhrmann, Johanna Säll, Karl Bacos, and Tina Rönn

Contents

Abstract

Type 2 diabetes (T2D) and obesity are multifactorial and polygenic metabolic diseases. Combinations of genetic and non-genetic risk factors such as risk SNPs, age, unhealthy diets, and physical inactivity increase the risk for these diseases. Emerging data also support a key role for epigenetic mechanisms in the pathogenesis of T2D and obesity. In this chapter, we summarize current knowledge of

C. Ling (✉) · S. Ruhrmann · J. Säll · K. Bacos · T. Rönn
Epigenetics and Diabetes Unit, Department of Clinical Sciences, Lund University Diabetes Centre, Lund University, Scania University Hospital, Malmö, Sweden
e-mail: charlotte.ling@med.lu.se; sabrina.ruhrmann@med.lu.se; johanna.sall_sernevi@med.lu.se; karl.bacos@med.lu.se; tina.ronn@med.lu.se

K. B. Michels (ed.), *Epigenetic Epidemiology*,
https://doi.org/10.1007/978-3-030-94475-9_19

epigenetic alterations found in individuals with T2D and obesity. We present studies performed in blood, as well as human tissues important for metabolism, i.e., adipose tissue, skeletal muscle, liver, and pancreatic islets. These studies have found differential DNA methylation associated with both T2D and obesity. Although some studies exist, there is still limited information regarding histone modifications in human tissues linked to metabolic diseases. We finally explore how epigenetic mechanisms may be targeted by epigenetic editing and inhibitors of epigenetic enzymes for future therapies and precision medicine in T2D and obesity.

Abbreviations

BMI	Body mass index
BWS	Beckwith–Wiedemann syndrome
Cas	CRISPR associated system
ChIP	Chromatin immunoprecipitation
CRISPR	Clustered regularly interspaced palindromic repeats
DMR	Differentially methylated region
DNMT	DNA methyltransferase
EWAS	Epigenome-wide association study
GWAS	Genome-wide association study
HDAC	Histone deacetylase
ICR	Imprinting control region
mQTL	Methylation quantitative trait locus
PBL	Peripheral blood lymphocyte
PBMC	Peripheral blood mononuclear cells
SAM	S-adenosylmethionine
SAT	Subcutaneous adipose tissue
sgRNA	Single guide RNA
T2D	Type 2 diabetes
TALES	Transcription activator-like effectors
TET	Ten eleven translocation
VAT	Visceral adipose tissue
WB	Western blot
WGBS	Whole-genome bisulfite sequencing
ZF	Zinc finger

19.1 Introduction

This chapter will discuss recent advances in epigenomics of human obesity and type 2 diabetes (T2D), common metabolic disorders with a continuously increasing prevalence worldwide. Metabolic disorders develop when the metabolism of

nutrients is inappropriate to the body's needs, with insufficient or excess intake of essential substances. The underlying cause may be dysfunctions in organs involved in metabolism, like the pancreas, skeletal muscle, adipose tissue, or liver, triggered by, e.g., inherited genetic defects, consuming too much of certain nutrients, or a mixture of these factors. According to the World Health Organization (WHO), more than 1.9 billion adults, or 39% of the world's population aged 18 years and over, were overweight (BMI $\geq$25) in 2016. Additionally, the International Diabetes Federation (IDF) predicts 578 million people to be affected by T2D by 2030 [1]. This development is likely the result of increased availability and intake of food with high fat and sugar content, in combination with the more sedentary lifestyle that comes with urbanization and changed modes of work and transportation.

Although both genetic and non-genetic factors, *e.g.* diet and physical inactivity, are known to increase the risk for obesity and T2D, these explain only a modest part of disease prevalence [2, 3]. Epigenetic changes are likely to mediate part of the missing heritability for metabolic disorders and may occur in response to the fast changes of the environment. The Dutch Hunger Winter is one example of how an altered environment, in this case prenatal famine, leads to epigenetic changes that are visible decades later [4]. Also, exercise seems to affect DNA methylation genome-wide [5, 6]. Not only environmental exposures need to be considered when analyzing and interpreting epigenetic changes, but also variables such as age and sex, which are known to impact the epigenome [7, 8].

Finally, as the epigenome is tissue specific, there is also a need to consider what tissue to study when connecting a metabolic phenotype to epigenetic disturbances, or vice versa. For obesity and T2D, most relevant are the hormone-secreting cells of the pancreas controlling blood glucose levels, the liver with a central role in all metabolic processes, adipose tissue, and skeletal muscle. However, in some cases, blood may mirror the epigenetic changes observed in metabolically active tissues [8, 9], which can then be used as a biomarker with the advantage of being easy to analyze in the clinic. Additionally, the life span of different cell types may also affect the dynamics of the epigenetic pattern.

The research field of epigenetics of obesity and T2D is fast emerging, and the accumulation of data from larger cohorts and adequate tissues will move the field from detecting genome-wide associations to providing evidence for causality.

19.2 DNA Methylation in Human Blood and Association with Type 2 Diabetes and Obesity

Blood is one of the easiest human tissues to use for analyses due to availability, commonly drawn at clinical visits and available in various biobanks. However, as mentioned above, the epigenetic state of blood cells is not as straightforward to connect to a metabolic phenotype as that of other, more metabolically active tissues. Still, metabolic diseases such as obesity are often associated with low-grade inflammation and increased white blood cell (WBC) count [10]. Also, DNA methylation

markers in the blood may respond to processes or disturbances throughout the different tissues of the human body.

The absolute changes in DNA methylation are often modest, hence large sample sizes as well as replication or collaborative efforts are needed to establish a true association with a disease. Chambers et al. [11] performed a nested case–control study based on DNA methylation measured using the Illumina 450k array [12] in DNA from blood drawn at baseline, and analyzed the risk of developing T2D. The study included both Indian Asians ($n = 1074$ with incident T2D and 1590 controls) and Europeans ($n = 377$ with incident T2D and 746 controls). Five methylation markers, annotated to *ABCG1*, *PHOSPHO1*, *SOCS3*, *SREBF1,* and *TXNIP*, were found to consistently associate with T2D development, both individually (relative T2D risk between top and bottom quartiles: 1.77–2.14), but also when combining the results to a methylation risk score with a relative risk for future T2D (top vs bottom quartile) of 3.51 in Indian Asians and 2.49 in Europeans (Table 19.1). The higher methylation risk score in Indian Asians may suggest an explanation for the higher T2D incidence in this population and provide an opportunity for risk stratification and disease prevention [11]. Dayeh et al. also found that DNA methylation in blood at *ABCG1* was associated with an increased risk, while DNA methylation at *PHOSPHO1* was associated with a decreased risk of future T2D [13]. A systematic review for differential DNA methylation associated with T2D or related traits was performed by Walaszczyk et al., followed by replication of 100 top hits in blood from 100 T2D and 100 control individuals [14]. In their case-control cohort, significance was obtained for five of those markers (in *ABCG1*, *LOXL2*, *TXNIP*, *SLC1A5*, and *SREBF1*), and hence these were suggested as disease biomarkers. The approach of combining data from different cohorts and studies increase generalizability, which is important for use in a random population. Still, the absolute changes in those markers range from 1.8 to 3.6% [14], and more markers need to be combined before it can be brought into the clinic as a useful tool. Another drawback of this study is that the replicated T2D markers are also associated with age, and as the control group is significantly younger than the T2D cases, it is hard to determine if the methylation differences are driven by the disease or rather by increasing age. In a different study, Cardona et al. identified 15 novel sites where DNA methylation was associated with incident T2D, and they confirmed the three sites annotated to *TXNIP*, *ABCG1*, and *SREBF1* [15].

A review from 2018 summarized the role of DNA methylation in blood and T2D, based on 37 studies. Both candidate gene, global and genome-wide approaches were considered, supporting evidence for reproducible associations between T2D and DNA methylation of *TCF7L2*, *KCNQ1*, *ABCG1*, *TXNIP*, *PHOSPHO1*, *SREBF1*, *SLC30A8*, and *FTO* [16]. Again, confounding factors are highlighted, e.g., sex, tissue heterogeneity, age, and ethnicity, which may all influence methylation levels by themselves and may mask the results of many studies. More recently, a meta-analysis of DNA methylation in peripheral blood from four European cohorts was performed, including methylation data from the Illumina 450k array and phenotypic traits in a total of 3428 individuals, of which 10% were diagnosed with T2D [17]. This study replicated previous associations, *e.g.* with *TXNIP*, *ABCG1*, and

Table 19.1 DNA methylation in blood and association with T2D and obesity

Study	Cohort (*n*)	Readout
Chambers et al. [11]	Prospective study: Indian Asians (1074 incident T2D; 1590 controls) Europeans (377 incident T2D; 746 controls).	*T2D* prediction: *ABCG1*, *PHOSPHO1*, *SOCS3*, *SREBF1* and *TXNIP*
Dayeh et al. [13]	Prospective study (129 incident T2D; 129 controls)	Association with T2D: *ABCG1* and *PHOSPHO1*
Walaszczyk et al. [14]	Review of EWAS publications followed by replication in 100 T2D and 100 controls	Association with T2D: *ABCG1*, *LOXL2*, *SLC1A5, SREBF1*, and *TXNIP*
Cardona et al. [15]	Main (563 incident T2D; 701 controls) Replication (1074 incident T2D; 1590 controls and 403 prevalent T2D; 2204 controls)	Association with T2D: *ABCG1*, *CPT1A*, *SREBF1*, and *TXNIP*
Willmer et al. [16]	Review of 37 studies	Association with T2D: *ABCG1*, *FTO*, *KCNQ1*, *PHOSPHO1*, *SLC30A8*, *SREBF1*, *TCF7L2*, and *TXNIP*
Juvinao-Quintero et al. [17]	Meta-analysis of four cohorts (3428 individuals; 10% diagnosed with T2D)	Association with T2D: *ABCG1*, *CPT1A*, *HDAC4*, *MIR23A*, *SYNM*, and *TXNIP*
Mendelson et al. [18]	Main (3743 individuals) Replication (4055 individuals)	Association with BMI: 83 CpG sites, of which 19 associated with gene expression (*ABCG1*, *CACNA2D3*, *CPT1A*, *DHCR24*, *LGALS3BP*, *NOD2*, *PHGDH*, *SARS*, *SLC1A*, and *SREBF1*). Two CpGs with causal effect on BMI
Wahl et al. [19]	Main (5387 individuals of European and Indian Asian ancestry) Replication (4874 European and Indian Asian men and women)	Association with BMI: 187 CpG sites, 62 of them also associated with T2D
Campanella et al. [20]	Population-based European cohorts (1941)	Association with adiposity: 40 CpG sites. CpG in *ABCG1* associated with four adiposity phenotypes, lipid levels and transcriptional activity

CPT1A, but also identified novel DNA methylation markers associated with T2D. These include a CpG site in *HDAC4*, and sites nearby *SYNM* and *MIR23A*. Together, these six markers captured 11% of the variation in T2D. Additional analyses showed that many of those top six sites were also associated with glucose tolerance, age, sex, and white-blood-cell type [17].

Considering obesity, several studies with thousands of samples run on the Illumina 450k array have presented robust associations between DNA methylation in blood and BMI, and also connected these alterations to gene expression and obesity-related traits [18–20]. Mendelson et al. presented 83 CpG sites associated with BMI after analyzing DNA methylation from whole blood in 3743 individuals

followed by replication in three cohorts with a total of 4055 individuals [18]. These CpG sites were further connected to gene expression, with significant associations for 19 CpG-expression pairs, representing 10 unique genes (*ABCG1*, *CPT1A*, *SREBF1*, *LGALS3BP*, *DHCR24*, *PHGDH*, *SARS*, *NOD2*, *CACNA2D3*, and *SLC1A*). Mendelian randomization analysis was used to determine casual support; 16 CpGs showed evidence for differential methylation secondary to BMI, whereas two CpGs, including the one in *SREBF1*, showed causal support for effect on BMI [18]. Another study based on genome-wide DNA methylation analysis in blood from 10,261 samples found 187 CpG sites significantly associated with BMI [19]. Also in this study there was more support for altered methylation secondary to obesity than the opposite, but still, a few CpG sites seem to influence obesity. Among the 187 epigenetic markers associated with obesity, 62 were also associated with incident T2D [19]. In an attempt to determine if DNA methylation may be an intermediate marker between obesity and obesity-related diseases, Campanella et al. connected DNA methylation in blood from 1941 individuals with four different adiposity measurements, and then tested if the significant marks were associated with some cancers and myocardial infarction [20]. First, 40 CpG sites were found associated with one or more adiposity measures (BMI, waist circumference, waist-hip, and waist-height ratio). One site in *ABCG1* is associated with all four phenotypes, and also with lipid levels and transcriptional activity. Next, among the 40 markers associated with adiposity, two were also associated with colorectal cancer and one with myocardial infarction [20].

19.3 DNA Methylation in Human Tissues of Importance for Metabolism, and Association with Type 2 Diabetes

A rapidly increasing number of studies show that epigenetic changes contribute to T2D pathophysiology by altering cell function in pancreatic islets, liver, skeletal muscle, and adipose tissue. These studies were initially based on candidate gene approaches, but as methods for genome-wide analysis of DNA methylation have been developed this has shifted towards epigenome-wide association studies (EWAS).

19.3.1 Pancreatic Islets

Mitochondrial metabolism is key for proper insulin secretion from the pancreatic β-cells, and dysfunction in this organelle therefore contributes to T2D [21]. Peroxisome proliferator-activated receptor gamma coactivator-1 alpha (PGC-1α), encoded by *PPARGC1A*, regulates the expression of genes involved in energy metabolism and has been implicated in T2D [22]. An early study on human islet epigenetics showed that this gene is not only vital for proper insulin secretion but also that promoter methylation at *PPARGC1A* was greatly increased in pancreatic islets in T2D [23]. Similar DNA methylation findings were later done in *INS*, the gene for

insulin [24], and *PDX1* [25] encoding a key transcription factor for β-cell development and function [26]. The expression of these three genes is greatly reduced in human islets in T2D, and altered promoter methylation has been experimentally shown to affect transcription of *INS* and *PDX1* [25, 27]. This indicates that the differential methylation patterns identified in T2D contribute to impaired β-cell function and therefore T2D pathophysiology. The genome-wide DNA methylation analyses in pancreatic islets in T2D have been based on Illumina methylation arrays or whole-genome bisulfite sequencing (WGBS). The earliest such study was published in 2012 and was based on the first generation of arrays, interrogating ~27,000 methylation sites. In this study, the authors found changes that are associated with β-cell survival and function [28]. A later study used the second generation of arrays, which allows the analysis of more than 450,000 methylation sites. This study also included a larger sample size, thus increasing the power greatly. More than 1500 CpG sites were found to be differentially methylated in islets from donors with T2D compared to non-diabetic controls [29]. More than 100 differentially methylated genes also exhibited altered expression in T2D, and it was experimentally shown that altered methylation can underlie the expression changes. Further in vitro follow-up showed that altering the expression of these genes in a similar fashion as in T2D, perturbed insulin and glucagon secretion from pancreatic β- and α-cells, respectively. A later study utilizing WGBS increased the coverage even more (~23 million CpG sites), albeit with lower power due to the high costs associated with sequencing at the depth needed for analysis of DNA methylation. Instead of identifying differential methylation at individual sites, this study focused on finding differentially methylated regions (DMRs), thus increasing the biological relevance [30]. DMRs were defined as stretches of DNA containing three or more consecutive differentially methylated CpG sites with an average absolute methylation difference of at least 5% in human islets from individuals with T2D vs controls. Almost 26,000 such DMRs were identified. Top hits were annotated to *PDX1*, *ARX*, a gene involved in α-cell development [31], and *TFAM*, a mediator of PDX1 function in β-cells [32]. In fact, there were seven DMRs annotated to *PDX1* and, on average, these were 1075 base pairs long, contained 46 CpG sites, and exhibited a mean difference in methylation of 52%, strongly supporting the notion that epigenetic dysregulation of *PDX1* is important in T2D. Additionally, the significant DMRs overlap several SNPs associated with T2D and were also enriched for binding sites for islet-specific transcription factors, further supporting an important role for altered DNA methylation in islet cell function in T2D. Finally, a few genes exhibiting both altered methylation and expression (*NR4A3*, *PARK2*, *PID1*, and *SOCS2*) were experimentally shown to regulate glucose-stimulated insulin secretion [30].

There are strong interactions between our genetic and epigenetic architecture, also in pancreatic islets. Importantly, approximately 25% of all SNPs introduce or remove a CpG site, hence directly affecting the possibility for DNA methylation to occur in the human genome. A methylation quantitative trait locus (mQTL) analysis showed that among 70,000 SNP-CpG pairs (i.e., SNPs associated with methylation at CpG site(s)), some were located in T2D candidate genes identified by genome-

wide association studies (GWAS), such as *ADCY5*, *KCNJ11*, *INS*, and *PDX1* [33]. A causal inference test showed that DNA methylation in some instances mediates the effect of the SNP on gene expression and insulin secretion. Interestingly, another study showed that close to 50% of the then known T2D-associated SNPs introduce or remove a CpG site and thereby strongly affect methylation and gene expression [34].

19.3.2 Liver

The liver is responsible for taking up glucose when levels are high and store it as glycogen, and releasing glucose from the breakdown of glycogen, or gluconeogenesis, to replenish low glucose levels in the circulation when needed. The decreased insulin and increased glucagon levels seen in T2D create an imbalance with enhanced glucose output from the liver as a result. Epigenetic modifications may contribute to this imbalance. Array-based EWAS have identified significant epigenetic changes in liver biopsies from individuals with T2D compared to non-diabetic controls [35–38]. Interestingly, the vast majority of CpG sites showed lower methylation in T2D, potentially due to lower levels of the methyl donor folate in T2D [38]. These studies identified altered methylation and expression of genes involved in, e.g., glucose and lipid metabolism, and cell cycle regulation. More specifically, the findings included T2D candidate genes as well as the genes for protein kinase Cε (PRKCE) [37] and platelet-derived growth factor α (PDGFA) [35] which were both hypomethylated and overexpressed in T2D. These proteins have been found to be involved in the development of insulin resistance in the liver [35, 39], supporting an important role for these epigenetic changes in T2D.

19.3.3 Skeletal Muscle

The skeletal muscle tissue is where most of the insulin-stimulated glucose uptake occurs, and well-functioning skeletal muscle cells are therefore important for glucose homeostasis. Studies on muscle biopsies as well as cell cultures established from human muscle have identified important roles of epigenetics in regulating this tissue in T2D. Barrès et al. used methylated DNA immunoprecipitation (MeDIP) to investigate promoter methylation and found that the promoter for *PPARGC1A* was hypermethylated in vastus lateralis from individuals with T2D [40]. Methylation of the promoter also correlated negatively with expression of the gene and number of mitochondria in the biopsies. Furthermore, exposing human myotubes to inflammatory molecules or fatty acids, i.e., a diabetogenic environment, resulted in increased methylation of the *PPARGC1A* promoter. These epigenetic changes were mediated by the DNA methyltransferase 3B (DNMT3B). Similar epigenetic findings in *PPARGC1A* were observed in monozygotic twins discordant for T2D [41]. Another study found altered methylation and expression of genes encoding mitochondrial enzymes in muscle biopsies from individuals with T2D. These included the gene for

pyruvate dehydrogenase kinase 4 (*PDK4*), which the authors also found hypomethylated in T2D. Knockout of *PDK4* results in improved glucose homeostasis in mice given a high-fat diet [42]. Interestingly, expression of both *PPARGC1A* and *PDK4* is upregulated in muscle by exercise [43, 44] and this might be due to epigenetic changes [45, 46]. In donors with T2D, it was further found that individuals who responded to exercise with a large epigenetic change at the *PPARGC1A* locus also showed a significant reduction in intramyocellular lipids [45]. A study investigating muscle biopsies from healthy individuals with and without a family history of T2D found differential methylation and expression of genes involved in, e.g., insulin signaling [5]. Similar epigenetic differences were evident in monozygotic twins discordant for T2D, further increasing the relevance of the findings to T2D. A more recent investigation found methylation and expression differences in muscle stem cells from normoglycemic controls compared with donors with T2D, and abnormal regulation of methylation and expression during muscle cell differentiation in T2D [47]. Close to 600 genes were differentially expressed in T2D myoblasts, including genes with known roles in muscle differentiation (e.g., *FBN2*, *TEAD4*, and *STAT3*) as well as previously unrecognized regulators of human myogenesis, e.g., *VPS39*. DNA methylation of 331 CpG sites correlated with expression of these differentially expressed genes, including *FBN2* and *VPS39*. Additionally, *VPS39* was among the genes that were differentially expressed between the two groups after differentiation to myotubes. Functional follow-up showed that mimicking the lower expression of *VPS39* seen in T2D via siRNA-mediated knockdown in human myoblasts resulted in a severely impaired formation of myotubes due to impaired autophagy and epigenetic mechanisms. Heterozygous knockout in mice also impaired glucose uptake in skeletal muscle, strongly implicating a role for *VPS39* deficiency in T2D.

19.3.4 Adipose Tissue

The adipose tissue is our main storage facility for energy. It takes up nutrients from the blood postprandially and stores the energy as lipids. When nutrient levels in the blood drop, and the levels of insulin and glucagon change, the adipose tissue will instead release nutrients for direct energy production in our cells or for glucose production in the liver.

In T2D there are DNA methylation changes in adipose tissue, as indicated by several studies. In 2012, Ribel-Madsen et al. used arrays interrogating ~27,000 methylation sites and bisulfite sequencing in a small cohort of monozygotic twins discordant for T2D to show that there were larger methylation differences in repetitive regions (e.g., LINE1) than in gene promoters. However, when they looked at specific sites they found significantly higher methylation of, e.g., the *HNF4A* promoter [41]. Mutation of *HNF4A* leads to maturity onset diabetes of the young (MODY) 1 due to β-cell defects [48]. Nilsson et al. [49] analyzed two T2D case-control cohorts and found differential methylation at almost 16,000 CpG sites, including in several T2D candidate genes, *e.g. IRS1, KCNQ1, TCF7L2*, and

PPARG. Methylation of a subset of the significant CpG sites is also associated with BMI and fasting glucose, two risk factors for T2D. Importantly, it was also shown that matching the donors for age and sex was necessary to detect significant differences that stood for correction for multiple testing. Similarly, another study investigated both visceral (VAT) and subcutaneous (SAT) adipose tissue and identified methylation changes in T2D [36]. Again, this included sites in candidate genes for T2D and obesity, e.g., *IRS1*, *FOXA2,* and *KCNQ1*. A study investigating a smaller cohort, and thereby with weaker power, found significant differences on 24 CpG sites including in *HOOK2* [50]. A more detailed methylation analysis of this gene showed further methylation differences between women and men with T2D. No functional data was presented however and whether *HOOK2*, encoding a dynein adaptor [51], plays a role in adipocyte function in T2D remains unknown.

Similarly to the findings in muscle stem cells described above, there seem to be DNA methylation differences in adipocyte precursors in T2D. This was shown in a study by Andersen et al., where they analyzed preadipocytes from VAT of lean individuals and obese individuals with and without T2D [52]. Adipocyte differentiation, as determined by expression of the differentiation marker Fatty acid-binding protein 4 (*FABP4*), was impaired in obese individuals with T2D. This was accompanied by both expression and DNA methylation changes. Genes downregulated only in obese individuals with T2D were enriched for metabolic pathways as well as the insulin and PPAR signaling pathways. While only a cell cycle pathway was enriched in the methylation data, the authors found epigenetic differences associated with adipocyte differentiation. The authors constructed an adipogenic score based on the expression of adipocyte-specific genes. Almost 4000 DMRs were associated with the adipogenic score, and when comparing lean normoglycemic and obese with T2D they found many more DMRs associating with the adipogenic score ($n = 305$) than when comparing normoglycemic lean and obese individuals ($n = 100$). Similarly, analysis of adipose tissue biopsies from first-degree relatives of individuals with T2D identified DNA methylation differences [53]. The most affected gene was *PTPRD*, a T2D candidate gene [54], with six hypomethylated DMRs. The first-degree relatives also exhibited higher *PTPRD* expression in adipose tissue compared to the control group, and methylation at the top DMR was shown to regulate transcription. *PTPRD* expression also correlated with adipocyte size. *In vitro* analysis further showed that overexpression of *Ptprd* in 3T3-L1 pre-adipocytes led to impaired adipogenesis.

19.4 DNA Methylation in Human Tissues and Association with Obesity

Although multiple, well-powered studies have presented robust associations between DNA methylation in blood and BMI, less is known about the relation between obesity and the epigenome of metabolically active organs, i.e., adipose tissue, skeletal muscle, pancreas, and the liver. In 2015, Rönn et al. published an EWAS based on DNA methylation in SAT and association with BMI [8]. As the

body composition differs between men and women, and the epigenome displays sex-specific differences [7], male ($n = 96$) and female ($n = 94$) cohorts were analyzed separately. The analysis based on the Illumina 450k array showed that methylation of 33,058 individual CpG sites in the male and 39,533 sites in the female cohort was associated with BMI. In the male cohort, BMI-associated methylation was further linked to differential expression of 2825 genes. Additionally, among the ~5000 CpG sites associated with BMI in both cohorts, a large proportion have also been associated with T2D, which may indicate that differential DNA methylation due to BMI predispose to T2D [8]. Another study of adipose tissue DNA methylation compared monozygotic twins discordant for BMI. Even if the methylome was more similar within twin pairs than between unrelated individuals, the authors found 22 CpG sites differentially methylated between co-twins with high and low BMI [55]. The genes discovered point to a downregulation of lipogenesis and adipogenesis, together with upregulation of inflammation and extracellular matrix remodeling in adipose tissue from the twins with a high BMI. Also, SAT DNA methylation of some CpG sites correlated with markers of unhealthy obesity, i.e., intra-abdominal and liver fat, insulin resistance, dyslipidemia, and low-grade inflammation.

In 2014, Horvath et al. presented evidence for epigenetic biomarkers of aging and hypothesized that obesity may increase biological age in a tissue-specific manner [56]. Indeed, they found a strong correlation between BMI and epigenetic age in the liver, which may partly explain why obese people have an increased risk of age-related diseases. The association between BMI and epigenetic age could not be confirmed in SAT, skeletal muscle or blood, while all tissues displayed a significant correlation between chronological age and the age determined by DNA methylation markers. To follow up on the acceleration of epigenetic age induced by obesity, Toro-Martin et al. used a reduced version of Horvath's epigenetic clock in VAT. Also in this metabolically active tissue, they found a correlation between BMI and epigenetic age. They also confirmed previous findings of DNA methylation age and association with BMI in the liver, but not in blood [57].

Dick et al. reported DNA methylation sites in *HIF3A* to be associated with BMI [58]. These sites were first identified in blood, but also replicated in adipose tissue. Furthermore, in adipose tissue the methylation level correlated with *HIF3A* gene expression. DNA methylation and gene expression of *HIF3A* was further investigated in several tissues by Main et al. [59], confirming the association between promoter methylation of *HIF3A* and BMI. Their results indicate a tissue-specific role for *HIF3A*, where SAT is the more important with low expression negatively affecting whole-body insulin sensitivity. In another candidate gene approach, DNA methylation levels of the adipokines *LEP* and *ADIPOQ* were investigated and related to, e.g., BMI [60]. These genes are involved in energy balance and the development of obesity and related complications. In SAT, *ADIPOQ* DNA methylation was associated with BMI, whereas an association between *LEP* DNA methylation and BMI was only found in blood. Moreover, there was an association between DNA methylation of both genes in SAT as well as VAT and LDL cholesterol [60].

Hall et al. investigated DNA methylation of the *GLP1R* promoter in human pancreatic islets [61]. The glucagon-like peptide-1 receptor (GLP1R) is needed for proper insulin secretion and its expression is downregulated in islets from T2D patients. In the human islets, one CpG site of the *GLP1R* promoter was positively associated with BMI, while negatively associated with *GLP1R* expression. If obesity is a cause of altered DNA methylation, which in turn affects gene expression and insulin secretion, this again provides a mechanism of how obesity increases risk of T2D. In another study by Hall et al., human pancreatic islets were exposed to palmitate *in vitro* to mimic the environment in obese individuals [62]. On a global level, there was a small significant increase in DNA methylation after palmitate treatment, but due to the limited number of human islets, changes on individual CpG sites only reached nominal significance. Anyhow, combining DNA methylation and mRNA expression data from islets exposed to palmitate revealed several T2D candidate genes, supporting how an obesogenic environment contributes to the development of T2D. In a larger set of human islets, it was also shown that BMI affected gene expression of several genes that were also altered by the palmitate treatment [62].

Furthermore, Davegårdh et al. found abnormal epigenetic changes during the myogenesis of human myoblasts from obese individuals compared with controls. Almost four times as many changes in DNA methylation took place when myoblasts from obese differentiated into myofibers compared to myoblasts from lean individuals, supporting an epigenetic memory from the obese *in vivo* situation [63]. Still, the impact of obesity on genome-wide DNA methylation in human pancreatic islets as well as skeletal muscle remains to be further investigated.

19.5 Chromatin Structure and Histone Modifications in Humans, and Association with Type 2 Diabetes

DNA is wound around histone proteins in order to facilitate its compaction to chromatin. Post-translational modifications of histone tails (e.g., methylation and acetylation) enable a complex regulation of chromatin structure and thereby gene activity. The identification of histone mark positioning over the genome and the co-localization of multiple, different histone marks are used to define and predict distinct chromatin states.

Studies on histone modifications associated with T2D in humans are limited compared to corresponding data on DNA methylation. Direct comparisons of histone marks in individuals with T2D vs non-diabetic individuals have only been performed in blood cells and using biased approaches (ChIP-PCR) for a single histone mark and one or two gene(s) of interest (summarized in Table 19.2). These studies demonstrate that the diabetes-associated inflammation in part can be attributed to altered histone modifications and consequently altered expression of inflammatory genes. Miao et al. [64] compared the levels of histone 3 acetylation at lysines 9 and 14 (H3K9ac and H3K14ac) at the *TNF-α* and *COX-2* promoters in blood monocytes from four individuals with diabetes (of which two with T2D) vs

Table 19.2 Studies of histone modifications in individuals with type 2 diabetes

Study	*n* (diabetes/ control)	Cell type	Readout (vs control)
Miao et al. [64]	4 (of which 2 T2D)/2	Monocyte	↑ H3K9ac levels at *TNF-α* and *COX-2* promoters (ChIP-PCR)
Miao et al. [65]	6 (also 7 T1D)/6	Monocyte	↑ H3K9me2 levels at *IL-1A* promoter and *PTEN* coding region (ChIP-PCR)
Hou et al. [66]	12/12	PBMC	↑ Histone H3 acetylation at *TNF-α* and *COX-2* promoters (ChIP-PCR)
Paneni et al. [67]	38/12	Monocyte	↑ H3K4me1 levels at *NFκB p65* promoter (ChIP-PCR)

two controls. They observed increased H3K9ac at the promoters of these inflammatory genes in individuals with diabetes. Similarly, Hou et al. [66] demonstrated an increased overall H3 acetylation at the *TNF-α* and *COX-2* promoters, and subsequent increased mRNA expression of these genes, in peripheral blood mononuclear cells (PBMC) from 12 individuals with T2D vs 12 controls. In another study, Miao et al. [65] compared the levels of the repressive mark H3K9me2 at the *IL-1A* promoter and *PTEN* coding region in blood monocytes from 6 individuals with T2D and 6 controls. They observed increased H3K9me2 levels at these regions in individuals with T2D vs controls. In this study, they also profiled histone H3 dimethylation at lysines 4 and 9 (H3K4me2 and H3K9me2) using DNA microarrays (ChIP-on-chip), and corresponding mRNA expression, in high glucose-treated THP-1 monocytes to mimic the diabetic environment. Paneni et al. [67] demonstrated that upregulation of the methyltransferase Set7 caused the enrichment of H3K4me1 at the *NFκB p65* promoter and subsequently increased *NFκB p65* mRNA expression in blood monocytes from 38 individuals with T2D vs 12 controls.

Global analyses of histone modifications in target tissues for T2D, e.g., pancreatic islets, have only been performed in individuals without diabetes. Thus, comprehensive genome-wide analyses of histone modifications in diabetes-relevant tissues from individuals with T2D are missing. To bridge this gap, efforts have been made to instead integrate data on chromatin structure and histone marks in non-diabetic individuals, to construct maps of chromatin states (e.g., active, repressed, bivalent), regulatory elements (e.g., enhancers, TSSs) and open chromatin. Using these approaches, insights on the role of histone modifications in T2D are gained by overlaying these maps with known T2D-associated loci [68–73], also at a single-cell level [74, 75]. Together, these studies have demonstrated that, in pancreatic islets, T2D-associated GWAS SNPs to a large extent are localized to distal regulatory elements (enhancers) characterized by enrichment of H3K4me1 in open chromatin (identified by FAIRE-seq, DNase-seq, and ATAC-seq). Furthermore, the generation of islet chromatin 3D structure enables the identification of candidate target genes to T2D-associated variants located in distal enhancers [76, 77]. Also, ATAC-seq has been used in islets from donors with T2D and non-diabetic controls to identify alterations in open chromatin regions (OCR) associated with diabetes [78]. These OCRs have further been linked to numerous histone modifications associated with open and closed chromatin, as well as enhancer regions.

19.6 Histone Modifications in Humans and Association with Obesity

Similarly to T2D, comprehensive studies on the role of histone modifications for human obesity are lacking and this is an area that needs further investigation. Existing studies have been limited with regards to the methodologies used (indirect or biased) and small sample sizes that could confound the results (summarized in Table 19.3). Jufvas et al. [79] used immunoblotting to compare the overall levels of H3K4me2, H3K4me3, and H3K9me2 in mature adipocytes isolated from 19 overweight individuals, 10 overweight individuals with T2D, and 14 control individuals. They observed lower H3K4me2 levels in the overweight non-diabetic individuals, and higher H3K4me3 levels in the overweight individuals with T2D, compared to control individuals, while no differences were observed for the heterochromatin mark H3K9me3 [79]. This suggests that being overweight or having T2D is associated with dynamic regulation of H3 lysine 4 methylation status. However, the underlying mechanisms for this regulation as well as the downstream effects on gene expression and regulatory networks were not explored. Castellano-Castillo et al. [80] compared the levels of H3K4me3 at the promoters of several genes involved in adipogenesis, lipid metabolism and inflammation (*LEP*, *LPL*, *SREBF2*, *SCD1*, *PPARG*, *IL6*, *TNF*, *E2F1*) in VAT from 10 individuals with morbid obesity, 9 morbidly obese individuals with prediabetes and 10 control individuals. They found enrichment of H3K4me3 at the promoters of *LPL*, *SREBF2*, *SCD1*, *PPARG*, *IL6*, and *E2F1* in morbidly obese individuals with prediabetes compared to control individuals, and there was a positive correlation between H3K4me3 levels at these promoters and BMI, HOMA-IR and glucose. Moreover, H3K4me3 enrichment at the *E2F1* promoter was also associated with increased *E2F1* mRNA expression [80].

Väremo et al. [82] performed gene-set analysis of genes that were differentially expressed in muscle cells isolated from individuals with either obesity or T2D, obese individuals with T2D, and control individuals ($n = 6$ per group). Using histone modification gene-sets obtained from the Epigenomics Roadmap and ENCODE projects, they identified one histone mark (H3K27me3) that was enriched in all groups compared to control for five muscle-specific pathways, and associated with

Table 19.3 Studies of histone modifications in individuals with obesity

Study	*n* (obesity/control)	Cell type/ tissue	Readout (vs control)
Jufvas et al. [79]	19 overweight (also 10 overweight with T2D)/ 14	Primary adipocytes	↓ H3K4me2 in overweight, ↑ H3K4me3 in T2D (WB)
Castellano-Castillo et al. [80]	10 morbid obesity (also 9 morbid obesity with prediabetes)/10	Visceral adipose tissue	↑ H3K4me3 levels at *LPL*, *SREBF2*, *SCD1*, *PPARG*, *IL6*, and *E2F1* promoters (ChIP-PCR)
Puig et al. [81]	6/6	Skeletal muscle	→ H3K27ac (WB)

downregulation or unspecific regulation of contributing genes. This histone mark likely influences the transcriptional signatures in these cells [82]. Puig et al. used immunoblotting to compare the overall levels of H3K27 acetylation (H3K27ac) in skeletal muscle biopsies from individuals with obesity and control individuals (n = 6 per group), and observed no differences between the groups [81]. Williams et al. [83] performed genome-wide profiling of H3K4me1 and H3K27ac in human muscle cells exposed to TNF-α or palmitate to mimic the metabolic stress associated with increasing adiposity. Using this approach, they were able to construct maps of enhancer elements and observed that acetylation of H3K27, marking active enhancers, was dynamically regulated in response to the treatments. By combining data on enhancer activity with enhancer-promoter interactions and gene expression in the same cells, and overlapping with GWAS SNPs associated with T2D, insulin resistance, BMI or waist-hip-ratio (WHR), they identified 13 putative target genes associated with these phenotypes. Based on eQTL analysis of human skeletal muscle biopsies they also demonstrated that 12 of these genes (associated with BMI and/or WHR) displayed differential expression based on SNP genotype (BMI: *GAB2*, *BTBD1*, *PABPC4*, *MACF1*; WHR: *LAMB1*, *TCEA3*, *FILIP1L*, *NRP1*, *ZHX3*, *TBX15*, *TNFAIP8*; BMI+WHR: *EIF6*) [83].

Obesity is tightly linked to diet, and several factors derived from nutrient catabolism act as substrates for histone-modifying enzymes, e.g., S-adenosylmethionine (SAM) for methylation and acetyl-CoA for acetylation. Studies in humans have shown that plasma levels of SAM are associated with adiposity [84], and serum levels of SAM increased in response to overfeeding [85]. McDonnell et al. demonstrated that acetyl-CoA derived from cellular fatty acid oxidation directly promotes histone acetylation [86]. In line, Malmgren et al. showed that palmitate exposure increased the activity of histone acetyl transferase (HAT), increased H3K9ac at candidate genes and reduced glucose-stimulated insulin secretion in β-cells [87]. Diet and nutrient availability likely influence histone methylation and acetylation patterns in metabolic diseases, but this remains to be further elucidated.

19.7 Epigenetic Editing and Impact on Metabolic Disease

Existing evidence suggests a key role of epigenetics in the pathogenesis of metabolic disease, as summarized above. However, up to today, there is limited proof of whether disease-associated epigenetic modifications cause the disease. Some studies have used causal mediation analysis to show that, e.g., DNA methylation may cause metabolic diseases and/or directly affect metabolic traits [15, 33, 88, 89]. However, these are mathematical models and not experimental proof. In order to identify causal epigenetic changes in metabolic disease, different tools for epigenetic editing have been developed. These technical platforms utilize DNA binding molecules and epigenetic modifiers to recognize the target sequence and actively re-write the epigenetic signature, thus allowing for the investigation of a specific epigenetic modification at a specific gene locus [90].

The epigenetic editing tools currently in use are (i) zinc finger (ZF) proteins, (ii) transcription activator-like effectors (TALES), and (iii) the clustered regularly interspaced palindromic repeats (CRISPR) system. Both ZF proteins and TALES consist of several modules of amino acids that recognize three base pairs and one single base pair, respectively, of the target DNA sequence [91, 92]. Both ZF proteins and TALES are expensive and labor intensive to produce. In addition, many off-target effects are observed for the ZF proteins. The CRISPR-CRISPR associated protein system (Cas) originates from the bacterial defense system. Here, the CRISPR-Cas system detects foreign DNA, which is then cleaved by the nuclease activity of Cas9. Cas9 is guided to the specific sequence by single-guide RNAs (sgRNAs). For the purpose of epigenetic editing, an enzymatically deactivated Cas9 (dCas9) is used in order to not change the actual DNA sequence [93]. The design and redesign of the CRISPR-dCas9 system is relatively simple and not very cost-intensive in comparison to the other two platforms. Furthermore, off-target effects can be predicted. Epigenetic modifiers like DNA methyltransferases (DNMTs), Ten Eleven Translocation (TET), or histone modification enzymes are then bound either directly to the DNA binding platform (ZF proteins and TALES) or to the deactivated Cas9 protein, which is recruited to the target region by the sgRNA (CRISPR-dCas9 system). When delivered into the target cells, the DNA binding platform finds its DNA sequence and the epigenetic modifier exerts its repressing or activating function by *e.g.* adding methylation to cytosines (DNMT3a) or promoting DNA demethylation (TET1), respectively. In addition to epigenetic modifiers that affect gene expression by modulating DNA methylation, a histone acetyltransferase (p300) for instance can also be fused to the DNA binding platform and mediate gene activation [94, 95].

Existing studies, trying to identify causal epigenetic changes in metabolic disease, utilized the TALES [96, 97] and CRISPR-dCas9 [98] platforms in order to investigate the epigenetic landscape of their genetic region of interest. Kameswaran et al. investigated the *DLK1-MEG3* locus, which is misregulated in islets from donors with T2D [96]. They describe hypermethylation of the *MEG3* promoter in T2D islets when compared to control islets. Hypermethylation of the *MEG3* promoter also correlated with decreased expression of *MEG3* and its associated miRNAs. To investigate if increased methylation at this region is responsible for decreased expression of *Meg3*, Kameswaran et al. used TALES to target DNMTs (DNMT3a and 3L) to the Meg3-DMR sequence, located 360 bp upstream of the *Meg3* transcription start site, in mouse ßTC6 β-cells. Here, they observed that hypermethylation of the *Meg3*-DMR sequence using epigenetic editing led to a repressed expression of *Meg3*. Furthermore, they found a 20% increase in methylation at the targeted *Meg3*-DMR in TALE-DNMT expressing mouse ßTC6 β-cells, showing that DNA methylation is controlling *Meg3* expression. All in all, Kameswaran et al. were able to identify potential causal epigenetic changes at the Dlk1-Meg3 locus using targeted epigenetic modifiers in mouse ßTC6 β-cells.

The same group also studied the imprinted *CDKN1C* gene in human islets [97]. CDKN1C (p57) inhibits the cell cycle and is thereby a negative regulator of cell proliferation. Here, they focused on the methylation status at the imprinting

control region 2 (ICR2), which regulates the expression of *CDKN1C* and is hypomethylated on both alleles in Beckwith-Wiedemann-syndrome (BWS). These patients display decreased p57 protein levels, a deactivation of the *CDKN1C* gene, and exhibit massive β-cell proliferation, a desirable outcome in diabetes research in order to generate an alternative strategy to the use of β-cell replacement therapy. To investigate if epigenetic editing can be used to induce β-cell proliferation, a TALE-TET1 fusion protein that targets the ICR2 region was designed. Targeted demethylation of the ICR2 region in human islets led to decreased *CDKN1C* transcript levels and p57 protein levels, both resulting in increased β-cell replication. With this approach, Ou et al. demonstrated that epigenetic editing cannot only be applied to identify causal epigenetic changes, but also to mimic the molecular alterations of an imprinting disorder in order to induce β-cell proliferation [97].

Another group, Liu et al., utilized the CRISPR-dCas9 epigenetic editing platform to investigate MyoD and its role in muscle development in mouse fibroblasts [98]. In their study, they fused TET1 to dCas9, which with the help of sgRNAs is targeted to one (DMR5) of the six DMRs that have been described for the MyoD gene, in order to induce demethylation. They reported that targeting TET1 to the DMR5 region upstream of the MyoD gene resulted in a substantial reduction of methylation in the targeted region and moderate induction of MyoD expression. Within this study, Liu et al. demonstrated that binding an epigenetic modifier like TET1 to dCas9 allows for the targeted erasure of DNA methylation.

The described studies demonstrate how epigenetic editing can currently be applied in the field of metabolic disease (Fig. 19.1). They show that with the help of epigenetic editing we will be able to identify causal epigenetic changes or to take

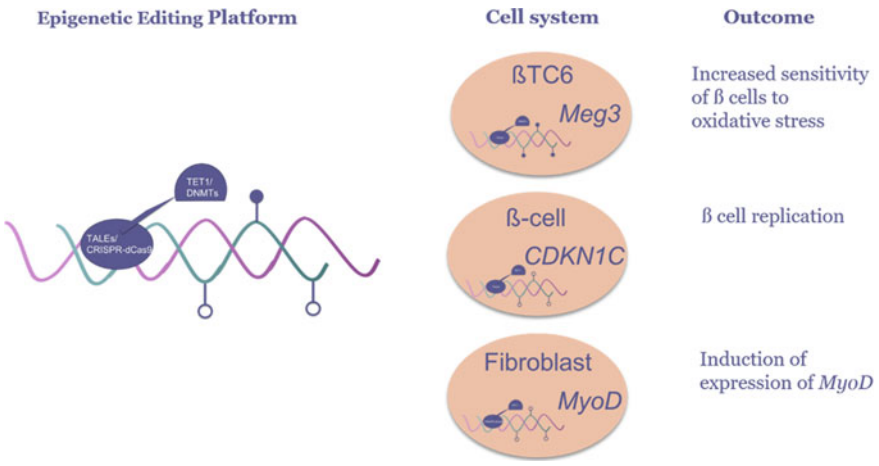

Fig. 19.1 Schematic overview of conducted studies by Kameswaran et al. [96], Ou et al. [97], and Liu et al. [98] in the field of epigenetic editing in metabolic disease. From left to right the figure shows the different epigenetic editing platforms used in the different studies, the cell system they were delivered to and the outcome of the studies. Filled and unfilled circles display methylated and unmethylated CpG sites, respectively

advantage of imprinting disorders to develop new treatment strategies. Furthermore, one of the hallmarks of epigenetics is reversibility. Utilizing epigenetic editing, this reversibility can be exerted in a targeted manner and unfavorable epigenetic states in metabolic disease can be corrected. Being able to correct those unfavorable epigenetic changes will offer patients with metabolic diseases a chance for a more personalized treatment approach and lead us to less treatment side effects in the future.

19.8 Epigenetic Inhibitors and Impact on Metabolic Disease

Previous studies have identified epigenetic modifications in individuals with metabolic disease, *e.g.* T2D or obesity [99]. These results support alterations in the activity of epigenetic enzymes. Indeed, some studies have shown changes in either expression, protein levels, or activity of enzymes known to alter the epigenome in tissues from individuals with metabolic diseases [47, 100, 101]. It is hence possible that inhibitors of epigenetic enzymes could affect metabolic phenotypes and potentially be used for therapy. There are numerous inhibitors of histone deacetylases (HDAC) and methyltransferases as well as of DNMTs [102]. Some of these inhibitors are commercially available and some are approved by the FDA or in clinical trials for cancer therapy [102]. Recent data also support that TSA, MC1568 and GSK-J4, inhibitors of histone-modifying enzymes, can improve β-cell function or insulin secretion in T2D islets [101, 103–105]. For example, the group of Mandrup-Poulsen has shown that HDAC inhibition prevents cytokine-induced toxicity in pancreatic β-cells, that oral administration of the HDAC inhibitor ITF2357 to streptozotocin (STZ) treated mice normalize STZ-induced hyperglycemia, and that selective inhibition of HDAC3 is a potential therapeutic path forward to protect β-cells from inflammatory cytokines and nutritional overload in diabetes [106–110]. Moreover, our group has shown that *HDAC7* is overexpressed in pancreatic islets from patients with T2D compared with normoglycemic controls. In addition, overexpression of *HDAC7* in clonal β-cells resulted in reduced glucose-stimulated insulin secretion, while simultaneous exposure to TSA or MC1568 restored the impaired insulin secretion in *HDAC7*-overexpressing cells [101]. We also exposed human pancreatic islets from donors with T2D to MC1568, which resulted in increased glucose-stimulated insulin secretion [103].

Some studies have also investigated epigenetic enzymes in target tissues for insulin *e.g.* muscle cells. For instance, we demonstrated abnormal levels of DNA methyltransferases, DNMT3A and 3B, during myogenesis of myoblasts from patients with T2D compared with non-diabetic individuals [47]. Also, obesity affected the expression of DNMTs during myogenesis [100]. However, whether inhibitors of DNMTs improve myogenesis in these patients remains to be tested. On the other hand, HDAC3 inhibition protected C2C12 myotubes and skeletal muscle against lipotoxicity, and the authors suggested that such inhibitors may be effective for the treatment of obesity and insulin resistance [111]. Moreover, adipose tissue phenotypes are affected by HDAC inhibitors, *e.g.* specific epigenome modifiers

affect white adipocyte differentiation, and inhibition of class I HDACs from the very first stage of differentiation stimulates the differentiation process and imprints cells toward a highly oxidative phenotype [112]. Together these data support that inhibitors of epigenetic enzymes affect tissues involved in both insulin secretion and sensitivity and that they may be specifically developed for future diabetes therapies.

19.9 Conclusion

Although more epigenetic studies are needed, in particular investigating histone modifications and individual cell types, it is well accepted that epigenetic alterations have a key role in the pathogenesis of T2D and obesity. Case-control and prospective human studies have identified numerous epigenetic differences that seem to affect both insulin sensitivity and secretion. The majority of these studies have focused on DNA methylation, an epigenetic mark that is easy to study using array-based approaches. Hence, future studies should focus on the large number of histone modifications that most likely play an important role in metabolic disease. Also, the use of epigenetic editing and inhibitors of epigenetic enzymes may make it possible to target epigenetic signatures and enzymes for future therapies. Nevertheless, a better understanding of the specific epigenetic mechanisms that cause T2D and obesity is needed before such therapies may be developed.

References

1. Saeedi P, Petersohn I, Salpea P, Malanda B, Karuranga S, Unwin N, Colagiuri S, Guariguata L, Motala AA, Ogurtsova K, Shaw JE, Bright D, Williams R, Committee IDFDA (2019) Global and regional diabetes prevalence estimates for 2019 and projections for 2030 and 2045: results from the International Diabetes Federation Diabetes Atlas, 9 (th) edition. Diabetes Res Clin Pract 157:107843. https://doi.org/10.1016/j.diabres.2019.107843
2. Locke AE, Kahali B, Berndt SI, Justice AE, Pers TH, Day FR, Powell C, Vedantam S, Buchkovich ML, Yang J, Croteau-Chonka DC, Esko T, Fall T, Ferreira T, Gustafsson S, Kutalik Z, Luan J, Magi R, Randall JC, Winkler TW, Wood AR, Workalemahu T, Faul JD, Smith JA, Zhao JH, Zhao W, Chen J, Fehrmann R, Hedman AK, Karjalainen J, Schmidt EM, Absher D, Amin N, Anderson D, Beekman M, Bolton JL, Bragg-Gresham JL, Buyske S, Demirkan A, Deng G, Ehret GB, Feenstra B, Feitosa MF, Fischer K, Goel A, Gong J, Jackson AU, Kanoni S, Kleber ME, Kristiansson K, Lim U, Lotay V, Mangino M, Leach IM, Medina-Gomez C, Medland SE, Nalls MA, Palmer CD, Pasko D, Pechlivanis S, Peters MJ, Prokopenko I, Shungin D, Stancakova A, Strawbridge RJ, Sung YJ, Tanaka T, Teumer A, Trompet S, van der Laan SW, van Setten J, Van Vliet-Ostaptchouk JV, Wang Z, Yengo L, Zhang W, Isaacs A, Albrecht E, Arnlov J, Arscott GM, Attwood AP, Bandinelli S, Barrett A, Bas IN, Bellis C, Bennett AJ, Berne C, Blagieva R, Bluher M, Bohringer S, Bonnycastle LL, Bottcher Y, Boyd HA, Bruinenberg M, Caspersen IH, Chen YI, Clarke R, Daw EW, de Craen AJM, Delgado G, Dimitriou M, Doney ASF, Eklund N, Estrada K, Eury E, Folkersen L, Fraser RM, Garcia ME, Geller F, Giedraitis V, Gigante B, Go AS, Golay A, Goodall AH, Gordon SD, Gorski M, Grabe HJ, Grallert H, Grammer TB, Grassler J, Gronberg H, Groves CJ, Gusto G,

Haessler J, Hall P, Haller T, Hallmans G, Hartman CA, Hassinen M, Hayward C, Heard-Costa NL, Helmer Q, Hengstenberg C, Holmen O, Hottenga JJ, James AL, Jeff JM, Johansson A, Jolley J, Juliusdottir T, Kinnunen L, Koenig W, Koskenvuo M, Kratzer W, Laitinen J, Lamina C, Leander K, Lee NR, Lichtner P, Lind L, Lindstrom J, Lo KS, Lobbens S, Lorbeer R, Lu Y, Mach F, Magnusson PKE, Mahajan A, McArdle WL, McLachlan S, Menni C, Merger S, Mihailov E, Milani L, Moayyeri A, Monda KL, Morken MA, Mulas A, Muller G, Muller-Nurasyid M, Musk AW, Nagaraja R, Nothen MM, Nolte IM, Pilz S, Rayner NW, Renstrom F, Rettig R, Ried JS, Ripke S, Robertson NR, Rose LM, Sanna S, Scharnagl H, Scholtens S, Schumacher FR, Scott WR, Seufferlein T, Shi J, Smith AV, Smolonska J, Stanton AV, Steinthorsdottir V, Stirrups K, Stringham HM, Sundstrom J, Swertz MA, Swift AJ, Syvanen AC, Tan ST, Tayo BO, Thorand B, Thorleifsson G, Tyrer JP, Uh HW, Vandenput L, Verhulst FC, Vermeulen SH, Verweij N, Vonk JM, Waite LL, Warren HR, Waterworth D, Weedon MN, Wilkens LR, Willenborg C, Wilsgaard T, Wojczynski MK, Wong A, Wright AF, Zhang Q, LifeLines Cohort S, Brennan EP, Choi M, Dastani Z, Drong AW, Eriksson P, Franco-Cereceda A, Gadin JR, Gharavi AG, Goddard ME, Handsaker RE, Huang J, Karpe F, Kathiresan S, Keildson S, Kiryluk K, Kubo M, Lee JY, Liang L, Lifton RP, Ma B, McCarroll SA, McKnight AJ, Min JL, Moffatt MF, Montgomery GW, Murabito JM, Nicholson G, Nyholt DR, Okada Y, Perry JRB, Dorajoo R, Reinmaa E, Salem RM, Sandholm N, Scott RA, Stolk L, Takahashi A, Tanaka T, van't Hooft FM, Vinkhuyzen AAE, Westra HJ, Zheng W, Zondervan KT, Consortium AD, Group A-BW, Consortium CAD, Consortium CK, Glgc, Icbp, Investigators M, Mu TC, Consortium MI, Consortium P, ReproGen C, Consortium G, International Endogene C, Heath AC, Arveiler D, Bakker SJL, Beilby J, Bergman RN, Blangero J, Bovet P, Campbell H, Caulfield MJ, Cesana G, Chakravarti A, Chasman DI, Chines PS, Collins FS, Crawford DC, Cupples LA, Cusi D, Danesh J, de Faire U, den Ruijter HM, Dominiczak AF, Erbel R, Erdmann J, Eriksson JG, Farrall M, Felix SB, Ferrannini E, Ferrieres J, Ford I, Forouhi NG, Forrester T, Franco OH, Gansevoort RT, Gejman PV, Gieger C, Gottesman O, Gudnason V, Gyllensten U, Hall AS, Harris TB, Hattersley AT, Hicks AA, Hindorff LA, Hingorani AD, Hofman A, Homuth G, Hovingh GK, Humphries SE, Hunt SC, Hypponen E, Illig T, Jacobs KB, Jarvelin MR, Jockel KH, Johansen B, Jousilahti P, Jukema JW, Jula AM, Kaprio J, Kastelein JJP, Keinanen-Kiukaanniemi SM, Kiemeney LA, Knekt P, Kooner JS, Kooperberg C, Kovacs P, Kraja AT, Kumari M, Kuusisto J, Lakka TA, Langenberg C, Marchand LL, Lehtimaki T, Lyssenko V, Mannisto S, Marette A, Matise TC, McKenzie CA, McKnight B, Moll FL, Morris AD, Morris AP, Murray JC, Nelis M, Ohlsson C, Oldehinkel AJ, Ong KK, Madden PAF, Pasterkamp G, Peden JF, Peters A, Postma DS, Pramstaller PP, Price JF, Qi L, Raitakari OT, Rankinen T, Rao DC, Rice TK, Ridker PM, Rioux JD, Ritchie MD, Rudan I, Salomaa V, Samani NJ, Saramies J, Sarzynski MA, Schunkert H, Schwarz PEH, Sever P, Shuldiner AR, Sinisalo J, Stolk RP, Strauch K, Tonjes A, Tregouet DA, Tremblay A, Tremoli E, Virtamo J, Vohl MC, Volker U, Waeber G, Willemsen G, Witteman JC, Zillikens MC, Adair LS, Amouyel P, Asselbergs FW, Assimes TL, Bochud M, Boehm BO, Boerwinkle E, Bornstein SR, Bottinger EP, Bouchard C, Cauchi S, Chambers JC, Chanock SJ, Cooper RS, de Bakker PIW, Dedoussis G, Ferrucci L, Franks PW, Froguel P, Groop LC, Haiman CA, Hamsten A, Hui J, Hunter DJ, Hveem K, Kaplan RC, Kivimaki M, Kuh D, Laakso M, Liu Y, Martin NG, Marz W, Melbye M, Metspalu A, Moebus S, Munroe PB, Njolstad I, Oostra BA, Palmer CNA, Pedersen NL, Perola M, Perusse L, Peters U, Power C, Quertermous T, Rauramaa R, Rivadeneira F, Saaristo TE, Saleheen D, Sattar N, Schadt EE, Schlessinger D, Slagboom PE, Snieder H, Spector TD, Thorsteinsdottir U, Stumvoll M, Tuomilehto J, Uitterlinden AG, Uusitupa M, van der Harst P, Walker M, Wallaschofski H, Wareham NJ, Watkins H, Weir DR, Wichmann HE, Wilson JF, Zanen P, Borecki IB, Deloukas P, Fox CS, Heid IM, O'Connell JR, Strachan DP, Stefansson K, van Duijn CM, Abecasis GR, Franke L, Frayling TM, McCarthy MI, Visscher PM, Scherag A, Willer CJ, Boehnke M, Mohlke KL, Lindgren CM, Beckmann JS, Barroso I, North KE, Ingelsson E, Hirschhorn JN, Loos RJF, Speliotes EK

(2015) Genetic studies of body mass index yield new insights for obesity biology. Nature 518(7538):197–206. https://doi.org/10.1038/nature14177

3. Mahajan A, Taliun D, Thurner M, Robertson NR, Torres JM, Rayner NW, Payne AJ, Steinthorsdottir V, Scott RA, Grarup N, Cook JP, Schmidt EM, Wuttke M, Sarnowski C, Magi R, Nano J, Gieger C, Trompet S, Lecoeur C, Preuss MH, Prins BP, Guo X, Bielak LF, Below JE, Bowden DW, Chambers JC, Kim YJ, Ng MCY, Petty LE, Sim X, Zhang W, Bennett AJ, Bork-Jensen J, Brummett CM, Canouil M, Ec Kardt KU, Fischer K, Kardia SLR, Kronenberg F, Lall K, Liu CT, Locke AE, Luan J, Ntalla I, Nylander V, Schonherr S, Schurmann C, Yengo L, Bottinger EP, Brandslund I, Christensen C, Dedoussis G, Florez JC, Ford I, Franco OH, Frayling TM, Giedraitis V, Hackinger S, Hattersley AT, Herder C, Ikram MA, Ingelsson M, Jorgensen ME, Jorgensen T, Kriebel J, Kuusisto J, Ligthart S, Lindgren CM, Linneberg A, Lyssenko V, Mamakou V, Meitinger T, Mohlke KL, Morris AD, Nadkarni G, Pankow JS, Peters A, Sattar N, Stancakova A, Strauch K, Taylor KD, Thorand B, Thorleifsson G, Thorsteinsdottir U, Tuomilehto J, Witte DR, Dupuis J, Peyser PA, Zeggini E, Loos RJF, Froguel P, Ingelsson E, Lind L, Groop L, Laakso M, Collins FS, Jukema JW, Palmer CNA, Grallert H, Metspalu A, Dehghan A, Kottgen A, Abecasis GR, Meigs JB, Rotter JI, Marchini J, Pedersen O, Hansen T, Langenberg C, Wareham NJ, Stefansson K, Gloyn AL, Morris AP, Boehnke M, McCarthy MI (2018) Fine-mapping type 2 diabetes loci to single-variant resolution using high-density imputation and islet-specific epigenome maps. Nat Genet 50(11):1505–1513. https://doi.org/10.1038/s41588-018-0241-6
4. Heijmans BT, Tobi EW, Stein AD, Putter H, Blauw GJ, Susser ES, Slagboom PE, Lumey LH (2008) Persistent epigenetic differences associated with prenatal exposure to famine in humans. Proc Natl Acad Sci U S A 105(44):17046–17049. https://doi.org/10.1073/pnas.0806560105
5. Nitert MD, Dayeh T, Volkov P, Elgzyri T, Hall E, Nilsson E, Yang BT, Lang S, Parikh H, Wessman Y, Weishaupt H, Attema J, Abels M, Wierup N, Almgren P, Jansson PA, Ronn T, Hansson O, Eriksson KF, Groop L, Ling C (2012) Impact of an exercise intervention on DNA methylation in skeletal muscle from first-degree relatives of patients with type 2 diabetes. Diabetes 61(12):3322–3332. https://doi.org/10.2337/db11-1653
6. Ronn T, Volkov P, Davegardh C, Dayeh T, Hall E, Olsson AH, Nilsson E, Tornberg A, Dekker Nitert M, Eriksson KF, Jones HA, Groop L, Ling C (2013) A six months exercise intervention influences the genome-wide DNA methylation pattern in human adipose tissue. PLoS Genet 9(6):e1003572. https://doi.org/10.1371/journal.pgen.1003572
7. Hall E, Volkov P, Dayeh T, Esguerra JL, Salo S, Eliasson L, Ronn T, Bacos K, Ling C (2014) Sex differences in the genome-wide DNA methylation pattern and impact on gene expression, microRNA levels and insulin secretion in human pancreatic islets. Genome Biol 15(12):522. https://doi.org/10.1186/s13059-014-0522-z
8. Ronn T, Volkov P, Gillberg L, Kokosar M, Perfilyev A, Jacobsen AL, Jorgensen SW, Brons C, Jansson PA, Eriksson KF, Pedersen O, Hansen T, Groop L, Stener-Victorin E, Vaag A, Nilsson E, Ling C (2015) Impact of age, BMI and HbA1c levels on the genome-wide DNA methylation and mRNA expression patterns in human adipose tissue and identification of epigenetic biomarkers in blood. Hum Mol Genet 24(13):3792–3813. https://doi.org/10.1093/hmg/ddv124
9. Bacos K, Gillberg L, Volkov P, Olsson AH, Hansen T, Pedersen O, Gjesing AP, Eiberg H, Tuomi T, Almgren P, Groop L, Eliasson L, Vaag A, Dayeh T, Ling C (2016) Blood-based biomarkers of age-associated epigenetic changes in human islets associate with insulin secretion and diabetes. Nat Commun 7:11089. https://doi.org/10.1038/ncomms11089
10. Farhangi MA, Keshavarz SA, Eshraghian M, Ostadrahimi A, Saboor-Yaraghi AA (2013) White blood cell count in women: relation to inflammatory biomarkers, haematological profiles, visceral adiposity, and other cardiovascular risk factors. J Health Popul Nutr 31(1):58–64. https://doi.org/10.3329/jhpn.v31i1.14749
11. Chambers JC, Loh M, Lehne B, Drong A, Kriebel J, Motta V, Wahl S, Elliott HR, Rota F, Scott WR, Zhang W, Tan ST, Campanella G, Chadeau-Hyam M, Yengo L, Richmond RC,

Adamowicz-Brice M, Afzal U, Bozaoglu K, Mok ZY, Ng HK, Pattou F, Prokisch H, Rozario MA, Tarantini L, Abbott J, Ala-Korpela M, Albetti B, Ammerpohl O, Bertazzi PA, Blancher C, Caiazzo R, Danesh J, Gaunt TR, de Lusignan S, Gieger C, Illig T, Jha S, Jones S, Jowett J, Kangas AJ, Kasturiratne A, Kato N, Kotea N, Kowlessur S, Pitkaniemi J, Punjabi P, Saleheen D, Schafmayer C, Soininen P, Tai ES, Thorand B, Tuomilehto J, Wickremasinghe AR, Kyrtopoulos SA, Aitman TJ, Herder C, Hampe J, Cauchi S, Relton CL, Froguel P, Soong R, Vineis P, Jarvelin MR, Scott J, Grallert H, Bollati V, Elliott P, McCarthy MI, Kooner JS (2015) Epigenome-wide association of DNA methylation markers in peripheral blood from Indian Asians and Europeans with incident type 2 diabetes: a nested case-control study. Lancet Diabetes Endocrinol 3(7):526–534. https://doi.org/10.1016/S2213-8587(15)00127-8

12. Sandoval J, Heyn H, Moran S, Serra-Musach J, Pujana MA, Bibikova M, Esteller M (2011) Validation of a DNA methylation microarray for 450,000 CpG sites in the human genome. Epigenetics 6(6):692–702. https://doi.org/10.4161/epi.6.6.16196
13. Dayeh T, Tuomi T, Almgren P, Perfilyev A, Jansson PA, de Mello VD, Pihlajamaki J, Vaag A, Groop L, Nilsson E, Ling C (2016) DNA methylation of loci within ABCG1 and PHOSPHO1 in blood DNA is associated with future type 2 diabetes risk. Epigenetics 11(7):482–488. https://doi.org/10.1080/15592294.2016.1178418
14. Walaszczyk E, Luijten M, Spijkerman AMW, Bonder MJ, Lutgers HL, Snieder H, Wolffenbuttel BHR, van Vliet-Ostaptchouk JV (2018) DNA methylation markers associated with type 2 diabetes, fasting glucose and HbA1c levels: a systematic review and replication in a case-control sample of the Lifelines study. Diabetologia 61(2):354–368. https://doi.org/10.1007/s00125-017-4497-7
15. Cardona A, Day FR, Perry JRB, Loh M, Chu AY, Lehne B, Paul DS, Lotta LA, Stewart ID, Kerrison ND, Scott RA, Khaw KT, Forouhi NG, Langenberg C, Liu C, Mendelson MM, Levy D, Beck S, Leslie RD, Dupuis J, Meigs JB, Kooner JS, Pihlajamaki J, Vaag A, Perfilyev A, Ling C, Hivert MF, Chambers JC, Wareham NJ, Ong KK (2019) Epigenome-wide association study of incident type 2 diabetes in a british population: EPIC-Norfolk study. Diabetes 68(12):2315–2326. https://doi.org/10.2337/db18-0290
16. Willmer T, Johnson R, Louw J, Pheiffer C (2018) Blood-based DNA methylation biomarkers for type 2 diabetes: potential for clinical applications. Front Endocrinol (Lausanne) 9:744. https://doi.org/10.3389/fendo.2018.00744
17. Juvinao-Quintero DL, Marioni RE, Ochoa-Rosales C, Russ TC, Deary IJ, van Meurs JBJ, Voortman T, Hivert MF, Sharp GC, Relton CL, Elliott HR (2021) DNA methylation of blood cells is associated with prevalent type 2 diabetes in a meta-analysis of four European cohorts. Clin Epigenetics 13(1):40. https://doi.org/10.1186/s13148-021-01027-3
18. Mendelson MM, Marioni RE, Joehanes R, Liu C, Hedman AK, Aslibekyan S, Demerath EW, Guan W, Zhi D, Yao C, Huan T, Willinger C, Chen B, Courchesne P, Multhaup M, Irvin MR, Cohain A, Schadt EE, Grove ML, Bressler J, North K, Sundstrom J, Gustafsson S, Shah S, McRae AF, Harris SE, Gibson J, Redmond P, Corley J, Murphy L, Starr JM, Kleinbrink E, Lipovich L, Visscher PM, Wray NR, Krauss RM, Fallin D, Feinberg A, Absher DM, Fornage M, Pankow JS, Lind L, Fox C, Ingelsson E, Arnett DK, Boerwinkle E, Liang L, Levy D, Deary IJ (2017) Association of body mass index with DNA methylation and gene expression in blood cells and relations to cardiometabolic disease: a mendelian randomization approach. PLoS Med 14(1):e1002215. https://doi.org/10.1371/journal.pmed.1002215
19. Wahl S, Drong A, Lehne B, Loh M, Scott WR, Kunze S, Tsai PC, Ried JS, Zhang W, Yang Y, Tan S, Fiorito G, Franke L, Guarrera S, Kasela S, Kriebel J, Richmond RC, Adamo M, Afzal U, Ala-Korpela M, Albetti B, Ammerpohl O, Apperley JF, Beekman M, Bertazzi PA, Black SL, Blancher C, Bonder MJ, Brosch M, Carstensen-Kirberg M, de Craen AJ, de Lusignan S, Dehghan A, Elkalaawy M, Fischer K, Franco OH, Gaunt TR, Hampe J, Hashemi M, Isaacs A, Jenkinson A, Jha S, Kato N, Krogh V, Laffan M, Meisinger C, Meitinger T, Mok ZY, Motta V, Ng HK, Nikolakopoulou Z, Nteliopoulos G, Panico S, Pervjakova N, Prokisch H, Rathmann W, Roden M, Rota F, Rozario MA, Sandling JK,

Schafmayer C, Schramm K, Siebert R, Slagboom PE, Soininen P, Stolk L, Strauch K, Tai ES, Tarantini L, Thorand B, Tigchelaar EF, Tumino R, Uitterlinden AG, van Duijn C, van Meurs JB, Vineis P, Wickremasinghe AR, Wijmenga C, Yang TP, Yuan W, Zhernakova A, Batterham RL, Smith GD, Deloukas P, Heijmans BT, Herder C, Hofman A, Lindgren CM, Milani L, van der Harst P, Peters A, Illig T, Relton CL, Waldenberger M, Jarvelin MR, Bollati V, Soong R, Spector TD, Scott J, McCarthy MI, Elliott P, Bell JT, Matullo G, Gieger C, Kooner JS, Grallert H, Chambers JC (2017) Epigenome-wide association study of body mass index, and the adverse outcomes of adiposity. Nature 541(7635):81–86. https://doi.org/10.1038/nature20784

20. Campanella G, Gunter MJ, Polidoro S, Krogh V, Palli D, Panico S, Sacerdote C, Tumino R, Fiorito G, Guarrera S, Iacoviello L, Bergdahl IA, Melin B, Lenner P, de Kok T, Georgiadis P, Kleinjans JCS, Kyrtopoulos SA, Bueno-de-Mesquita HB, Lillycrop KA, May AM, Onland-Moret NC, Murray R, Riboli E, Verschuren M, Lund E, Mode N, Sandanger TM, Fiano V, Trevisan M, Matullo G, Froguel P, Elliott P, Vineis P, Chadeau-Hyam M (2018) Epigenome-wide association study of adiposity and future risk of obesity-related diseases. Int J Obes 42(12):2022–2035. https://doi.org/10.1038/s41366-018-0064-7
21. Fex M, Nicholas LM, Vishnu N, Medina A, Sharoyko VV, Nicholls DG, Spegel P, Mulder H (2018) The pathogenetic role of beta-cell mitochondria in type 2 diabetes. J Endocrinol 236(3): R145–R159. https://doi.org/10.1530/JOE-17-0367
22. Soyal S, Krempler F, Oberkofler H, Patsch W (2006) PGC-1alpha: a potent transcriptional cofactor involved in the pathogenesis of type 2 diabetes. Diabetologia 49(7):1477–1488. https://doi.org/10.1007/s00125-006-0268-6
23. Ling C, Del Guerra S, Lupi R, Ronn T, Granhall C, Luthman H, Masiello P, Marchetti P, Groop L, Del Prato S (2008) Epigenetic regulation of PPARGC1A in human type 2 diabetic islets and effect on insulin secretion. Diabetologia 51(4):615–622. https://doi.org/10.1007/s00125-007-0916-5
24. Yang BT, Dayeh TA, Kirkpatrick CL, Taneera J, Kumar R, Groop L, Wollheim CB, Nitert MD, Ling C (2011) Insulin promoter DNA methylation correlates negatively with insulin gene expression and positively with HbA(1c) levels in human pancreatic islets. Diabetologia 54(2): 360–367. https://doi.org/10.1007/s00125-010-1967-6
25. Yang BT, Dayeh TA, Volkov PA, Kirkpatrick CL, Malmgren S, Jing X, Renstrom E, Wollheim CB, Nitert MD, Ling C (2012) Increased DNA methylation and decreased expression of PDX-1 in pancreatic islets from patients with type 2 diabetes. Mol Endocrinol 26(7): 1203–1212. https://doi.org/10.1210/me.2012-1004
26. Fujimoto K, Polonsky KS (2009) Pdx1 and other factors that regulate pancreatic beta-cell survival. Diabetes Obes Metab 11(Suppl 4):30–37. https://doi.org/10.1111/j.1463-1326.2009.01121.x
27. Kuroda A, Rauch TA, Todorov I, Ku HT, Al-Abdullah IH, Kandeel F, Mullen Y, Pfeifer GP, Ferreri K (2009) Insulin gene expression is regulated by DNA methylation. PLoS One 4(9): e6953. https://doi.org/10.1371/journal.pone.0006953
28. Volkmar M, Dedeurwaerder S, Cunha DA, Ndlovu MN, Defrance M, Deplus R, Calonne E, Volkmar U, Igoillo-Esteve M, Naamane N, Del Guerra S, Masini M, Bugliani M, Marchetti P, Cnop M, Eizirik DL, Fuks F (2012) DNA methylation profiling identifies epigenetic dysregulation in pancreatic islets from type 2 diabetic patients. EMBO J 31(6):1405–1426. https://doi.org/10.1038/emboj.2011.503
29. Dayeh T, Volkov P, Salo S, Hall E, Nilsson E, Olsson AH, Kirkpatrick CL, Wollheim CB, Eliasson L, Ronn T, Bacos K, Ling C (2014) Genome-wide DNA methylation analysis of human pancreatic islets from type 2 diabetic and non-diabetic donors identifies candidate genes that influence insulin secretion. PLoS Genet 10(3):e1004160. https://doi.org/10.1371/journal.pgen.1004160
30. Volkov P, Bacos K, Ofori JK, Esguerra JL, Eliasson L, Ronn T, Ling C (2017) Whole-genome bisulfite sequencing of human pancreatic islets reveals novel differentially methylated regions

in type 2 diabetes pathogenesis. Diabetes 66(4):1074–1085. https://doi.org/10.2337/db16-0996
31. Gage BK, Asadi A, Baker RK, Webber TD, Wang R, Itoh M, Hayashi M, Miyata R, Akashi T, Kieffer TJ (2015) The role of ARX in human pancreatic endocrine specification. PLoS One 10(12):e0144100. https://doi.org/10.1371/journal.pone.0144100
32. Gauthier BR, Wiederkehr A, Baquie M, Dai C, Powers AC, Kerr-Conte J, Pattou F, MacDonald RJ, Ferrer J, Wollheim CB (2009) PDX1 deficiency causes mitochondrial dysfunction and defective insulin secretion through TFAM suppression. Cell Metab 10(2):110–118. https://doi.org/10.1016/j.cmet.2009.07.002
33. Olsson AH, Volkov P, Bacos K, Dayeh T, Hall E, Nilsson EA, Ladenvall C, Ronn T, Ling C (2014) Genome-wide associations between genetic and epigenetic variation influence mRNA expression and insulin secretion in human pancreatic islets. PLoS Genet 10(11):e1004735. https://doi.org/10.1371/journal.pgen.1004735
34. Dayeh TA, Olsson AH, Volkov P, Almgren P, Ronn T, Ling C (2013) Identification of CpG-SNPs associated with type 2 diabetes and differential DNA methylation in human pancreatic islets. Diabetologia 56(5):1036–1046. https://doi.org/10.1007/s00125-012-2815-7
35. Abderrahmani A, Yengo L, Caiazzo R, Canouil M, Cauchi S, Raverdy V, Plaisance V, Pawlowski V, Lobbens S, Maillet J, Rolland L, Boutry R, Queniat G, Kwapich M, Tenenbaum M, Bricambert J, Saussenthaler S, Anthony E, Jha P, Derop J, Sand O, Rabearivelo I, Leloire A, Pigeyre M, Daujat-Chavanieu M, Gerbal-Chaloin S, Dayeh T, Lassailly G, Mathurin P, Staels B, Auwerx J, Schurmann A, Postic C, Schafmayer C, Hampe J, Bonnefond A, Pattou F, Froguel P (2018) Increased hepatic PDGF-AA signaling mediates liver insulin resistance in obesity-associated type 2 diabetes. Diabetes 67(7):1310–1321. https://doi.org/10.2337/db17-1539
36. Barajas-Olmos F, Centeno-Cruz F, Zerrweck C, Imaz-Rosshandler I, Martinez-Hernandez A, Cordova EJ, Rangel-Escareno C, Galvez F, Castillo A, Maydon H, Campos F, Maldonado-Pintado DG, Orozco L (2018) Altered DNA methylation in liver and adipose tissues derived from individuals with obesity and type 2 diabetes. BMC Med Genet 19(1):28. https://doi.org/10.1186/s12881-018-0542-8
37. Kirchner H, Sinha I, Gao H, Ruby MA, Schonke M, Lindvall JM, Barres R, Krook A, Naslund E, Dahlman-Wright K, Zierath JR (2016) Altered DNA methylation of glycolytic and lipogenic genes in liver from obese and type 2 diabetic patients. Mol Metab 5(3):171–183. https://doi.org/10.1016/j.molmet.2015.12.004
38. Nilsson E, Matte A, Perfilyev A, de Mello VD, Kakela P, Pihlajamaki J, Ling C (2015) Epigenetic alterations in human liver from subjects with type 2 diabetes in parallel with reduced folate levels. J Clin Endocrinol Metab 100(11):E1491–E1501. https://doi.org/10.1210/jc.2015-3204
39. Samuel VT, Liu ZX, Qu X, Elder BD, Bilz S, Befroy D, Romanelli AJ, Shulman GI (2004) Mechanism of hepatic insulin resistance in non-alcoholic fatty liver disease. J Biol Chem 279(31):32345–32353. https://doi.org/10.1074/jbc.M313478200
40. Barres R, Osler ME, Yan J, Rune A, Fritz T, Caidahl K, Krook A, Zierath JR (2009) Non-CpG methylation of the PGC-1alpha promoter through DNMT3B controls mitochondrial density. Cell Metab 10(3):189–198. https://doi.org/10.1016/j.cmet.2009.07.011
41. Ribel-Madsen R, Fraga MF, Jacobsen S, Bork-Jensen J, Lara E, Calvanese V, Fernandez AF, Friedrichsen M, Vind BF, Hojlund K, Beck-Nielsen H, Esteller M, Vaag A, Poulsen P (2012) Genome-wide analysis of DNA methylation differences in muscle and fat from monozygotic twins discordant for type 2 diabetes. PLoS One 7(12):e51302. https://doi.org/10.1371/journal.pone.0051302
42. Jeoung NH, Harris RA (2008) Pyruvate dehydrogenase kinase-4 deficiency lowers blood glucose and improves glucose tolerance in diet-induced obese mice. Am J Physiol Endocrinol Metab 295(1):E46–E54. https://doi.org/10.1152/ajpendo.00536.2007

43. Lira VA, Benton CR, Yan Z, Bonen A (2010) PGC-1alpha regulation by exercise training and its influences on muscle function and insulin sensitivity. Am J Physiol Endocrinol Metab 299(2):E145–E161. https://doi.org/10.1152/ajpendo.00755.2009
44. Pilegaard H, Neufer PD (2004) Transcriptional regulation of pyruvate dehydrogenase kinase 4 in skeletal muscle during and after exercise. Proc Nutr Soc 63(2):221–226. https://doi.org/10.1079/pns2004345
45. Bajpeyi S, Covington JD, Taylor EM, Stewart LK, Galgani JE, Henagan TM (2017) Skeletal muscle PGC1alpha-1 nucleosome position and -260 nt DNA methylation determine exercise response and prevent ectopic lipid accumulation in men. Endocrinology 158(7):2190–2199. https://doi.org/10.1210/en.2017-00051
46. Barres R, Yan J, Egan B, Treebak JT, Rasmussen M, Fritz T, Caidahl K, Krook A, O'Gorman DJ, Zierath JR (2012) Acute exercise remodels promoter methylation in human skeletal muscle. Cell Metab 15(3):405–411. https://doi.org/10.1016/j.cmet.2012.01.001
47. Davegardh C, Sall J, Benrick A, Broholm C, Volkov P, Perfilyev A, Henriksen TI, Wu Y, Hjort L, Brons C, Hansson O, Pedersen M, Wurthner JU, Pfeffer K, Nilsson E, Vaag A, Stener-Victorin E, Pircs K, Scheele C, Ling C (2021) VPS39-deficiency observed in type 2 diabetes impairs muscle stem cell differentiation via altered autophagy and epigenetics. Nat Commun 12(1):2431. https://doi.org/10.1038/s41467-021-22068-5
48. Nkonge KM, Nkonge DK, Nkonge TN (2020) The epidemiology, molecular pathogenesis, diagnosis, and treatment of maturity-onset diabetes of the young (MODY). Clin Diabetes Endocrinol 6(1):20. https://doi.org/10.1186/s40842-020-00112-5
49. Nilsson E, Jansson PA, Perfilyev A, Volkov P, Pedersen M, Svensson MK, Poulsen P, Ribel-Madsen R, Pedersen NL, Almgren P, Fadista J, Ronn T, Klarlund Pedersen B, Scheele C, Vaag A, Ling C (2014) Altered DNA methylation and differential expression of genes influencing metabolism and inflammation in adipose tissue from subjects with type 2 diabetes. Diabetes 63(9):2962–2976. https://doi.org/10.2337/db13-1459
50. Rodriguez-Rodero S, Menendez-Torre E, Fernandez-Bayon G, Morales-Sanchez P, Sanz L, Turienzo E, Gonzalez JJ, Martinez-Faedo C, Suarez-Gutierrez L, Ares J, Diaz-Naya L, Martin-Nieto A, Fernandez-Morera JL, Fraga MF, Delgado-Alvarez E (2017) Altered intragenic DNA methylation of HOOK2 gene in adipose tissue from individuals with obesity and type 2 diabetes. PLoS One 12(12):e0189153. https://doi.org/10.1371/journal.pone.0189153
51. Dwivedi D, Kumari A, Rathi S, Mylavarapu SVS, Sharma M (2019) The dynein adaptor Hook2 plays essential roles in mitotic progression and cytokinesis. J Cell Biol 218(3):871–894. https://doi.org/10.1083/jcb.201804183
52. Andersen E, Ingerslev LR, Fabre O, Donkin I, Altintas A, Versteyhe S, Bisgaard T, Kristiansen VB, Simar D, Barres R (2019) Preadipocytes from obese humans with type 2 diabetes are epigenetically reprogrammed at genes controlling adipose tissue function. Int J Obes 43(2):306–318. https://doi.org/10.1038/s41366-018-0031-3
53. Parrillo L, Spinelli R, Longo M, Desiderio A, Mirra P, Nigro C, Fiory F, Hedjazifar S, Mutarelli M, Carissimo A, Formisano P, Miele C, Smith U, Raciti GA, Beguinot F (2020) Altered PTPRD DNA methylation associates with restricted adipogenesis in healthy first-degree relatives of type 2 diabetes subjects. Epigenomics 12(10):873–888. https://doi.org/10.2217/epi-2019-0267
54. Tsai FJ, Yang CF, Chen CC, Chuang LM, Lu CH, Chang CT, Wang TY, Chen RH, Shiu CF, Liu YM, Chang CC, Chen P, Chen CH, Fann CS, Chen YT, Wu JY (2010) A genome-wide association study identifies susceptibility variants for type 2 diabetes in Han Chinese. PLoS Genet 6(2):e1000847. https://doi.org/10.1371/journal.pgen.1000847
55. Pietilainen KH, Ismail K, Jarvinen E, Heinonen S, Tummers M, Bollepalli S, Lyle R, Muniandy M, Moilanen E, Hakkarainen A, Lundbom J, Lundbom N, Rissanen A, Kaprio J, Ollikainen M (2016) DNA methylation and gene expression patterns in adipose tissue differ significantly within young adult monozygotic BMI-discordant twin pairs. Int J Obes 40(4):654–661. https://doi.org/10.1038/ijo.2015.221

56. Horvath S, Erhart W, Brosch M, Ammerpohl O, von Schonfels W, Ahrens M, Heits N, Bell JT, Tsai PC, Spector TD, Deloukas P, Siebert R, Sipos B, Becker T, Rocken C, Schafmayer C, Hampe J (2014) Obesity accelerates epigenetic aging of human liver. Proc Natl Acad Sci U S A 111(43):15538–15543. https://doi.org/10.1073/pnas.1412759111
57. de Toro-Martin J, Guenard F, Tchernof A, Hould FS, Lebel S, Julien F, Marceau S, Vohl MC (2019) Body mass index is associated with epigenetic age acceleration in the visceral adipose tissue of subjects with severe obesity. Clin Epigenetics 11(1):172. https://doi.org/10.1186/s13148-019-0754-6
58. Dick KJ, Nelson CP, Tsaprouni L, Sandling JK, Aissi D, Wahl S, Meduri E, Morange PE, Gagnon F, Grallert H, Waldenberger M, Peters A, Erdmann J, Hengstenberg C, Cambien F, Goodall AH, Ouwehand WH, Schunkert H, Thompson JR, Spector TD, Gieger C, Tregouet DA, Deloukas P, Samani NJ (2014) DNA methylation and body-mass index: a genome-wide analysis. Lancet 383(9933):1990–1998. https://doi.org/10.1016/S0140-6736(13)62674-4
59. Main AM, Gillberg L, Jacobsen AL, Nilsson E, Gjesing AP, Hansen T, Pedersen O, Ribel-Madsen R, Vaag A (2016) DNA methylation and gene expression of HIF3A: cross-tissue validation and associations with BMI and insulin resistance. Clin Epigenetics 8:89. https://doi.org/10.1186/s13148-016-0258-6
60. Houde AA, Legare C, Biron S, Lescelleur O, Biertho L, Marceau S, Tchernof A, Vohl MC, Hivert MF, Bouchard L (2015) Leptin and adiponectin DNA methylation levels in adipose tissues and blood cells are associated with BMI, waist girth and LDL-cholesterol levels in severely obese men and women. BMC Med Genet 16:29. https://doi.org/10.1186/s12881-015-0174-1
61. Hall E, Dayeh T, Kirkpatrick CL, Wollheim CB, Dekker Nitert M, Ling C (2013) DNA methylation of the glucagon-like peptide 1 receptor (GLP1R) in human pancreatic islets. BMC Med Genet 14:76. https://doi.org/10.1186/1471-2350-14-76
62. Hall E, Volkov P, Dayeh T, Bacos K, Ronn T, Nitert MD, Ling C (2014) Effects of palmitate on genome-wide mRNA expression and DNA methylation patterns in human pancreatic islets. BMC Med 12:103. https://doi.org/10.1186/1741-7015-12-103
63. Davegardh C, Broholm C, Perfilyev A, Henriksen T, Garcia-Calzon S, Peijs L, Hansen NS, Volkov P, Kjobsted R, Wojtaszewski JF, Pedersen M, Pedersen BK, Ballak DB, Dinarello CA, Heinhuis B, Joosten LA, Nilsson E, Vaag A, Scheele C, Ling C (2017) Abnormal epigenetic changes during differentiation of human skeletal muscle stem cells from obese subjects. BMC Med 15(1):39. https://doi.org/10.1186/s12916-017-0792-x
64. Miao F, Gonzalo IG, Lanting L, Natarajan R (2004) In vivo chromatin remodeling events leading to inflammatory gene transcription under diabetic conditions. J Biol Chem 279(17):18091–18097. https://doi.org/10.1074/jbc.M311786200
65. Miao F, Wu X, Zhang L, Yuan YC, Riggs AD, Natarajan R (2007) Genome-wide analysis of histone lysine methylation variations caused by diabetic conditions in human monocytes. J Biol Chem 282(18):13854–13863. https://doi.org/10.1074/jbc.M609446200
66. Hou C, Zhao M, Li X, Li YJ, Lin Y, Lu QJ, Zhou ZG (2011) Histone H3 acetylation of tumor necrosis factor-alpha and cyclooxygenase-2 in patients with type 2 diabetes. Zhonghua Yi Xue Za Zhi 91(26):1805–1808
67. Paneni F, Costantino S, Battista R, Castello L, Capretti G, Chiandotto S, Scavone G, Villano A, Pitocco D, Lanza G, Volpe M, Luscher TF, Cosentino F (2015) Adverse epigenetic signatures by histone methyltransferase Set7 contribute to vascular dysfunction in patients with type 2 diabetes mellitus. Circ Cardiovasc Genet 8(1):150–158. https://doi.org/10.1161/CIRCGENETICS.114.000671
68. Bhandare R, Schug J, Le Lay J, Fox A, Smirnova O, Liu C, Naji A, Kaestner KH (2010) Genome-wide analysis of histone modifications in human pancreatic islets. Genome Res 20(4):428–433. https://doi.org/10.1101/gr.102038.109
69. Gaulton KJ, Nammo T, Pasquali L, Simon JM, Giresi PG, Fogarty MP, Panhuis TM, Mieczkowski P, Secchi A, Bosco D, Berney T, Montanya E, Mohlke KL, Lieb JD, Ferrer J

(2010) A map of open chromatin in human pancreatic islets. Nat Genet 42(3):255–259. https://doi.org/10.1038/ng.530
70. Parker SC, Stitzel ML, Taylor DL, Orozco JM, Erdos MR, Akiyama JA, van Bueren KL, Chines PS, Narisu N, Program NCS, Black BL, Visel A, Pennacchio LA, Collins FS, National Institutes of Health Intramural Sequencing Center Comparative Sequencing Program A, Authors NCSP (2013) Chromatin stretch enhancer states drive cell-specific gene regulation and harbor human disease risk variants. Proc Natl Acad Sci U S A 110(44):17921–17926. https://doi.org/10.1073/pnas.1317023110
71. Pasquali L, Gaulton KJ, Rodriguez-Segui SA, Mularoni L, Miguel-Escalada I, Akerman I, Tena JJ, Moran I, Gomez-Marin C, van de Bunt M, Ponsa-Cobas J, Castro N, Nammo T, Cebola I, Garcia-Hurtado J, Maestro MA, Pattou F, Piemonti L, Berney T, Gloyn AL, Ravassard P, Skarmeta JLG, Muller F, McCarthy MI, Ferrer J (2014) Pancreatic islet enhancer clusters enriched in type 2 diabetes risk-associated variants. Nat Genet 46(2):136–143. https://doi.org/10.1038/ng.2870
72. Stitzel ML, Sethupathy P, Pearson DS, Chines PS, Song L, Erdos MR, Welch R, Parker SC, Boyle AP, Scott LJ, Program NCS, Margulies EH, Boehnke M, Furey TS, Crawford GE, Collins FS (2010) Global epigenomic analysis of primary human pancreatic islets provides insights into type 2 diabetes susceptibility loci. Cell Metab 12(5):443–455. https://doi.org/10.1016/j.cmet.2010.09.012
73. Thurner M, van de Bunt M, Torres JM, Mahajan A, Nylander V, Bennett AJ, Gaulton KJ, Barrett A, Burrows C, Bell CG, Lowe R, Beck S, Rakyan VK, Gloyn AL, McCarthy MI (2018) Integration of human pancreatic islet genomic data refines regulatory mechanisms at type 2 diabetes susceptibility loci. elife 7:e31977. https://doi.org/10.7554/eLife.31977
74. Chiou J, Zeng C, Cheng Z, Han JY, Schlichting M, Miller M, Mendez R, Huang S, Wang J, Sui Y, Deogaygay A, Okino ML, Qiu Y, Sun Y, Kudtarkar P, Fang R, Preissl S, Sander M, Gorkin DU, Gaulton KJ (2021) Single-cell chromatin accessibility identifies pancreatic islet cell type- and state-specific regulatory programs of diabetes risk. Nat Genet 53(4):455–466. https://doi.org/10.1038/s41588-021-00823-0
75. Rai V, Quang DX, Erdos MR, Cusanovich DA, Daza RM, Narisu N, Zou LS, Didion JP, Guan Y, Shendure J, Parker SCJ, Collins FS (2020) Single-cell ATAC-Seq in human pancreatic islets and deep learning upscaling of rare cells reveals cell-specific type 2 diabetes regulatory signatures. Mol Metab 32:109–121. https://doi.org/10.1016/j.molmet.2019.12.006
76. Greenwald WW, Chiou J, Yan J, Qiu Y, Dai N, Wang A, Nariai N, Aylward A, Han JY, Kadakia N, Regue L, Okino ML, Drees F, Kramer D, Vinckier N, Minichiello L, Gorkin D, Avruch J, Frazer KA, Sander M, Ren B, Gaulton KJ (2019) Pancreatic islet chromatin accessibility and conformation reveals distal enhancer networks of type 2 diabetes risk. Nat Commun 10(1):2078. https://doi.org/10.1038/s41467-019-09975-4
77. Miguel-Escalada I, Bonas-Guarch S, Cebola I, Ponsa-Cobas J, Mendieta-Esteban J, Atla G, Javierre BM, Rolando DMY, Farabella I, Morgan CC, Garcia-Hurtado J, Beucher A, Moran I, Pasquali L, Ramos-Rodriguez M, Appel EVR, Linneberg A, Gjesing AP, Witte DR, Pedersen O, Grarup N, Ravassard P, Torrents D, Mercader JM, Piemonti L, Berney T, de Koning EJP, Kerr-Conte J, Pattou F, Fedko IO, Groop L, Prokopenko I, Hansen T, Marti-Renom MA, Fraser P, Ferrer J (2019) Human pancreatic islet three-dimensional chromatin architecture provides insights into the genetics of type 2 diabetes. Nat Genet 51(7):1137–1148. https://doi.org/10.1038/s41588-019-0457-0
78. Bysani M, Agren R, Davegardh C, Volkov P, Ronn T, Unneberg P, Bacos K, Ling C (2019) ATAC-seq reveals alterations in open chromatin in pancreatic islets from subjects with type 2 diabetes. Sci Rep 9(1):7785. https://doi.org/10.1038/s41598-019-44076-8
79. Jufvas A, Sjodin S, Lundqvist K, Amin R, Vener AV, Stralfors P (2013) Global differences in specific histone H3 methylation are associated with overweight and type 2 diabetes. Clin Epigenetics 5(1):15. https://doi.org/10.1186/1868-7083-5-15
80. Castellano-Castillo D, Denechaud PD, Fajas L, Moreno-Indias I, Oliva-Olivera W, Tinahones F, Queipo-Ortuno MI, Cardona F (2019) Human adipose tissue H3K4me3 histone

mark in adipogenic, lipid metabolism and inflammatory genes is positively associated with BMI and HOMA-IR. PLoS One 14(4):e0215083. https://doi.org/10.1371/journal.pone.0215083
81. Puig LS, Altıntaş A, Casaní-Galdón S, Gabriel BM, Barrès R, Conesa A, Chibalin AV, Näslund E, Krook A, Pillon NJ, Zierath JR (2021) Circadian transcriptomic and epigenomic remodeling in response to lipid overload and human obesity. bioRxiv:2021.02.23.432336. https://doi.org/10.1101/2021.02.23.432336
82. Varemo L, Henriksen TI, Scheele C, Broholm C, Pedersen M, Uhlen M, Pedersen BK, Nielsen J (2017) Type 2 diabetes and obesity induce similar transcriptional reprogramming in human myocytes. Genome Med 9(1):47. https://doi.org/10.1186/s13073-017-0432-2
83. Williams K, Ingerslev LR, Bork-Jensen J, Wohlwend M, Hansen AN, Small L, Ribel-Madsen-R, Astrup A, Pedersen O, Auwerx J, Workman CT, Grarup N, Hansen T, Barres R (2020) Skeletal muscle enhancer interactions identify genes controlling whole-body metabolism. Nat Commun 11(1):2695. https://doi.org/10.1038/s41467-020-16537-6
84. Elshorbagy AK, Nijpels G, Valdivia-Garcia M, Stehouwer CD, Ocke M, Refsum H, Dekker JM (2013) S-adenosylmethionine is associated with fat mass and truncal adiposity in older adults. J Nutr 143(12):1982–1988. https://doi.org/10.3945/jn.113.179192
85. Elshorbagy AK, Jerneren F, Samocha-Bonet D, Refsum H, Heilbronn LK (2016) Serum S-adenosylmethionine, but not methionine, increases in response to overfeeding in humans. Nutr Diabetes 6:e192. https://doi.org/10.1038/nutd.2015.44
86. McDonnell E, Crown SB, Fox DB, Kitir B, Ilkayeva OR, Olsen CA, Grimsrud PA, Hirschey MD (2016) Lipids reprogram metabolism to become a major carbon source for histone acetylation. Cell Rep 17(6):1463–1472. https://doi.org/10.1016/j.celrep.2016.10.012
87. Malmgren S, Spegel P, Danielsson AP, Nagorny CL, Andersson LE, Nitert MD, Ridderstrale M, Mulder H, Ling C (2013) Coordinate changes in histone modifications, mRNA levels, and metabolite profiles in clonal INS-1 832/13 beta-cells accompany functional adaptations to lipotoxicity. J Biol Chem 288(17):11973–11987. https://doi.org/10.1074/jbc.M112.422527
88. Jonsson J, Renault KM, Garcia-Calzon S, Perfilyev A, Estampador AC, Norgaard K, Lind MV, Vaag A, Hjort L, Michaelsen KF, Carlsen EM, Franks PW, Ling C (2021) Lifestyle intervention in pregnant women with obesity impacts cord blood DNA methylation, which associates with body composition in the offspring. Diabetes 70(4):854–866. https://doi.org/10.2337/db20-0487
89. Volkov P, Olsson AH, Gillberg L, Jorgensen SW, Brons C, Eriksson KF, Groop L, Jansson PA, Nilsson E, Ronn T, Vaag A, Ling C (2016) A genome-wide mQTL analysis in human adipose tissue identifies genetic variants associated with DNA methylation, gene expression and metabolic traits. PLoS One 11(6):e0157776. https://doi.org/10.1371/journal.pone.0157776
90. Thakore PI, Black JB, Hilton IB, Gersbach CA (2016) Editing the epigenome: technologies for programmable transcription and epigenetic modulation. Nat Methods 13(2):127–137. https://doi.org/10.1038/nmeth.3733
91. Choo Y, Klug A (1994) Selection of DNA binding sites for zinc fingers using rationally randomized DNA reveals coded interactions. Proc Natl Acad Sci U S A 91(23):11168–11172. https://doi.org/10.1073/pnas.91.23.11168
92. Zhang M, Wang F, Li S, Wang Y, Bai Y, Xu X (2014) TALE: a tale of genome editing. Prog Biophys Mol Biol 114(1):25–32. https://doi.org/10.1016/j.pbiomolbio.2013.11.006
93. Sander JD, Joung JK (2014) CRISPR-Cas systems for editing, regulating and targeting genomes. Nat Biotechnol 32(4):347–355. https://doi.org/10.1038/nbt.2842
94. Cano-Rodriguez D, Rots MG (2016) Epigenetic editing: on the verge of reprogramming gene expression at will. Curr Genet Med Rep 4(4):170–179. https://doi.org/10.1007/s40142-016-0104-3

95. Stolzenburg S, Goubert D, Rots MG (2016) Rewriting DNA methylation signatures at will: the curable genome within reach? Adv Exp Med Biol 945:475–490. https://doi.org/10.1007/978-3-319-43624-1_17
96. Kameswaran V, Golson ML, Ramos-Rodriguez M, Ou K, Wang YJ, Zhang J, Pasquali L, Kaestner KH (2018) The dysregulation of the DLK1-MEG3 locus in islets from patients with type 2 diabetes is mimicked by targeted epimutation of its promoter with TALE-DNMT constructs. Diabetes 67(9):1807–1815. https://doi.org/10.2337/db17-0682
97. Ou K, Yu M, Moss NG, Wang YJ, Wang AW, Nguyen SC, Jiang C, Feleke E, Kameswaran V, Joyce EF, Naji A, Glaser B, Avrahami D, Kaestner KH (2019) Targeted demethylation at the CDKN1C/p57 locus induces human beta cell replication. J Clin Invest 129(1):209–214. https://doi.org/10.1172/JCI99170
98. Liu XS, Wu H, Ji X, Stelzer Y, Wu X, Czauderna S, Shu J, Dadon D, Young RA, Jaenisch R (2016) Editing DNA methylation in the mammalian genome. Cell 167(1):233–247. https://doi.org/10.1016/j.cell.2016.08.056
99. Ling C, Ronn T (2019) Epigenetics in human obesity and type 2 diabetes. Cell Metab 29(5):1028–1044. https://doi.org/10.1016/j.cmet.2019.03.009
100. Davegårdh C, Broholm C, Perfilyev A, Henriksen T, García-Calzón S, Peijs L, Hansen NS, Volkov P, Kjøbsted R, Wojtaszewski JFP, Pedersen M, Pedersen BK, Ballak DB, Dinarello CA, Heinhuis B, Joosten LAB, Nilsson E, Vaag A, Scheele C, Ling C (2017) Abnormal epigenetic changes during differentiation of human skeletal muscle stem cells from obese subjects. BMC Med 15(1):39–39. https://doi.org/10.1186/s12916-017-0792-x
101. Daneshpajooh M, Bacos K, Bysani M, Bagge A, Ottosson Laakso E, Vikman P, Eliasson L, Mulder H, Ling C (2017) HDAC7 is overexpressed in human diabetic islets and impairs insulin secretion in rat islets and clonal beta cells. Diabetologia 60(1):116–125. https://doi.org/10.1007/s00125-016-4113-2
102. Cheng Y, He C, Wang M, Ma X, Mo F, Yang S, Han J, Wei X (2019) Targeting epigenetic regulators for cancer therapy: mechanisms and advances in clinical trials. Signal Transduction Targeted Ther 4(1):62. https://doi.org/10.1038/s41392-019-0095-0
103. Daneshpajooh M, Eliasson L, Bacos K, Ling C (2018) MC1568 improves insulin secretion in islets from type 2 diabetes patients and rescues beta-cell dysfunction caused by Hdac7 upregulation. Acta Diabetol 55(12):1231–1235. https://doi.org/10.1007/s00592-018-1201-4
104. Backe MB, Andersson JL, Bacos K, Christensen DP, Hansen JB, Dorosz JJ, Gajhede M, Dahlby T, Bysani M, Kristensen LH, Ling C, Olsen L, Mandrup-Poulsen T (2018) Lysine demethylase inhibition protects pancreatic beta cells from apoptosis and improves beta-cell function. Mol Cell Endocrinol 460:47–56. https://doi.org/10.1016/j.mce.2017.07.001
105. Backe MB, Jin C, Andreone L, Sankar A, Agger K, Helin K, Madsen AN, Poulsen SS, Bysani M, Bacos K, Ling C, Perone MJ, Holst B, Mandrup-Poulsen T (2019) The lysine demethylase KDM5B regulates islet function and glucose homeostasis. J Diabetes Res 2019:5451038. https://doi.org/10.1155/2019/5451038
106. Larsen L, Tonnesen M, Ronn SG, Storling J, Jorgensen S, Mascagni P, Dinarello CA, Billestrup N, Mandrup-Poulsen T (2007) Inhibition of histone deacetylases prevents cytokine-induced toxicity in beta cells. Diabetologia 50(4):779–789. https://doi.org/10.1007/s00125-006-0562-3
107. Lewis EC, Blaabjerg L, Storling J, Ronn SG, Mascagni P, Dinarello CA, Mandrup-Poulsen T (2011) The oral histone deacetylase inhibitor ITF2357 reduces cytokines and protects islet beta cells in vivo and in vitro. Mol Med 17(5-6):369–377. https://doi.org/10.2119/molmed.2010.00152
108. Christensen DP, Dahllof M, Lundh M, Rasmussen DN, Nielsen MD, Billestrup N, Grunnet LG, Mandrup-Poulsen T (2011) Histone deacetylase (HDAC) inhibition as a novel treatment for diabetes mellitus. Mol Med 17(5–6):378–390. https://doi.org/10.2119/molmed.2011.00021

109. Lundh M, Galbo T, Poulsen SS, Mandrup-Poulsen T (2015) Histone deacetylase 3 inhibition improves glycaemia and insulin secretion in obese diabetic rats. Diabetes Obes Metab 17(7): 703–707. https://doi.org/10.1111/dom.12470
110. Wagner FF, Lundh M, Kaya T, McCarren P, Zhang YL, Chattopadhyay S, Gale JP, Galbo T, Fisher SL, Meier BC, Vetere A, Richardson S, Morgan NG, Christensen DP, Gilbert TJ, Hooker JM, Leroy M, Walpita D, Mandrup-Poulsen T, Wagner BK, Holson EB (2016) An isochemogenic set of inhibitors to define the therapeutic potential of histone deacetylases in beta-cell protection. ACS Chem Biol 11(2):363–374. https://doi.org/10.1021/acschembio.5b00640
111. Lee SJ, Choi SE, Lee HB, Song MW, Kim YH, Jeong JY, Kang Y, Kim HJ, Kim TH, Jeon JY, Lee KW (2020) A class I histone deacetylase inhibitor attenuates insulin resistance and inflammation in palmitate-treated C2C12 myotubes and muscle of HF/HFr diet mice. Front Pharmacol 11:601448. https://doi.org/10.3389/fphar.2020.601448
112. Ferrari A, Longo R, Peri C, Coppi L, Caruso D, Mai A, Mitro N, De Fabiani E, Crestani M (2020) Inhibition of class I HDACs imprints adipogenesis toward oxidative and brown-like phenotype. Biochim Biophys Acta Mol Cell Biol Lipids 1865(4):158594. https://doi.org/10.1016/j.bbalip.2019.158594

Epigenetics and Vascular Disease

20

Ignacio Fernando Hall, Montserrat Climent, Floriana Maria Farina, and Leonardo Elia

Contents

Ignacio Fernando Hall, Montserrat Climent, Floriana Maria Farina have been equally contribute to this work.

I. F. Hall · M. Climent · F. M. Farina
Humanitas Clinical and Research Center—IRCCS, Rozzano, MI, Italy

L. Elia (✉)
Humanitas Clinical and Research Center—IRCCS, Rozzano, MI, Italy

Department of Molecular and Translational Medicine, University of Brescia, Brescia, Italy
e-mail: leonardo.elia@unibs.it

K. B. Michels (ed.), *Epigenetic Epidemiology*,
https://doi.org/10.1007/978-3-030-94475-9_20

Abstract

The mechanisms responsible for vascular disease development have been investigated for many decades, but we are far from a complete identification of all involved molecular processes. This still remains a major unmet need and despite significant improvements in diagnosis, prevention, and early intervention, cardiovascular pathologies are still the leading cause of death and disability worldwide. Epigenetics guides gene expression through the regulation of transcription independently of the genetic code. Those regulatory mechanisms are essential to numerous processes, such as cell growth, development, and differentiation, and they might depend on environmental adaptation, aging, and disease states. The current knowledge on the epigenetic mechanisms regulating vascular physiopathology has uncovered new potential targets for intervention. Herein, we provide an overview of the epigenetic landscape and its role in vascular diseases, highlighting the impact of DNA methylation and histone modification as well as non-coding RNA mechanisms.

Abbreviation

5mC	5-methyl-Cytosine
5-aza	5-azacytidine
3'UTR	3′-Untranslated region
circRNA	Circular RNA
CVD	Cardiovascular disease
DNMT	DNA methyltransferase
EC	Endothelial cell
EndMT	Endothelial-to-mesenchymal transition
HDL	High-density lipoprotein
LDL	Low-density lipoprotein
lncRNA	Long ncRNA
miRNA, miR	microRNA
MRE	miRNA Recognition Element
nt	Nucleotides
ORF	Open reading frame
PAH	Pulmonary artery hypertension
pre-miRNA	Precursor miRNA
pri-miRNA	Primary miRNA
PVOD	Venous-occlusive disease
SNP	Single nucleotide polymorphism
TSS	Transcription start site
VSMC	Vascular smooth muscle cell

20.1 Introduction: The Constant Burden of Cardiovascular Disease

All vasculature-associated diseases, including atherosclerosis and pulmonary hypertension, belong to the vast category of cardiovascular diseases (CVDs). CVDs are the principal cause of death and disability worldwide [1], and are indeed responsible for 31% of global deaths (https://www.who.int/en/news-room/fact-sheets/detail/cardiovascular-diseases-(cvds)). Over the last decade, progresses in diagnosis, prevention, and treatment have considerably improved overall survival. However, the number of people diagnosed with CVDs remains very constant worldwide [2]. This fact already triggering some innovative ideas many years ago, suggesting that not all cases of CVD morbidity could be explained by common risk factors, such as blood lipids, hypertension, and diabetes [3]. Nonetheless, the probability of developing vasculature-associated diseases might depend, besides the environmental risk factors cited above, also on genetic variations, which should not to be confused with genetic mutations, such as those observed in pathologies like familial hypercholesterolemia [4]. Among those variations are the single nucleotide polymorphisms (SNPs), whose association with vascular pathologies is a complex challenge that must be integrated with specific sets of risk factors. However, all this information is not sufficient to explain the great penetrance of vasculature-associated diseases in the global population; additional elements need to be included in the global picture. The first theorization of such innovative events was proposed in the middle of the twentieth century by the British developmental biologist Conrad Waddington. Indeed, with no idea of the mechanisms, he coined a new term: epigenetic landscape [5]. Waddington proposed the existence of biological phenomena on top of genetic variability, as essential modulators of cell fate determination [6].

In eukaryotic organisms, organ development is tightly regulated by coordinated steps of gene expression activation and repression that follow precise time and space events in cells sharing the same DNA sequence [7]. This phenomenon modulates the extreme level of folding of chromatin needed to fit chromosomes into the nucleus and also the changes necessary so that genes can be either accessible to regulator elements and be transcribed (areas called euchromatin) or tightly packed and inactive (heterochromatin) [8]. Transformation of euchromatin to heterochromatin and vice versa contributes to gene regulation and defines today's meaning of the term epigenetics, which is more complex compared to Waddington's definition: epigenetics now refers to those heritable changes that, rather than depending on changes of the DNA sequence, influence how chromatin structure affects gene expression. Those are principally based on chemical modification of DNA—in particular, on cytosine—and histone proteins; more recently, processes regulated by non-coding RNAs have also been added to those classified as epigenetic mechanisms [9, 10].

Although the understanding of the epigenetic mechanisms associated with vasculature-associated diseases is still in its infancy, it is advancing very fast due to great improvements in DNA sequencing capacity. Thus, today we foresee that the knowledge generated by studying the role of epigenetics in vasculature-associated

diseases may be soon translated into new therapeutic approaches [11, 12]. The concept of epigenetics is quite intriguing for the field of human epidemiology. As a plethora of studies has assessed the central role of epigenetic modifications in human disease etiology, it is becoming evident how epigenetics might link environmental factors and lifestyle to pathology onset and progression [13]. In light of the crucial role of epigenetics, the study of its epidemiology permits a better identification of the interindividual variables impacting differential gene expression and disease susceptibility, augmenting the armamentarium of physicians and researchers working in several scientific fields, including the cardiovascular one [14].

In this chapter, we give an overview of the role of epigenetics as the main non-genetic component of vascular disease risk, focusing on the chemical alteration of DNA and histones and the activity of non-coding RNAs in the development of these pathologies.

20.1.1 Etiology and Pathobiology of Vascular Diseases

Vasculature-associated diseases comprise a wide variety of conditions affecting primarily blood vessels, including atherosclerosis, restenosis, aneurysm, and different hypertensive syndromes, such as pulmonary artery hypertension (PAH), among others [15].

While the specific pathophysiologic processes underpinning injury development might differ from one pathology to another, they still share common genetic and non-genetic molecular determinants [16]. Disease initiation and progression are defined by alterations in the transcriptional program of the cells residing within the vessel wall—namely endothelial cells (ECs), vascular smooth muscle cells (VSMCs), and fibroblasts—and the impaired communication with inflammatory cells recruited to the site of injury. Indeed, sustained exposure to stressors fosters vascular remodeling through the establishment of disease-prone cellular phenotypes, by disrupting intra- and inter-cellular responses [11].

EC de-regulation and activation leads to endothelial dysfunction and deterioration [17]. Similarly, VSMCs undergo profound phenotypic changes, transitioning towards a synthetic, non-contractile state while gaining migratory properties along with excessive extracellular matrix (ECM) deposition and alteration of the apoptotic program [18]. Furthermore, immune cell (mainly monocytes/macrophages) differentiation, activation, and polarization within vascular lesions nurture disease progression [19].

Of note, several lines of research have elucidated that the impact of VSMC plasticity on vascular disease might vary in a context-dependent fashion. As such, VSMC proliferation and migration can be harmful or beneficial, whereas apoptosis, senescence, and switching to a more macrophage-like phenotype can promote inflammation and disease progression [20].

To further complicate matters, these pathologies can, in turn, lead to secondary injures, such as heart failure, myocardial infarction, and ventricular hypertrophy.

Additionally, their progression might be hastened by complications such as diabetes and metabolic disorders [21].

Here, we will discuss the role of epigenetics in two specific pathologies: atherosclerosis and PAH.

20.1.2 Atherosclerosis

Atherosclerosis is the main clinical manifestation of CVDs and underlies the majority of cardiovascular complications, remaining the leading cause of morbidity and mortality worldwide [22, 23]. It is a chronic inflammatory disease that preferentially develops in specific points of vessels, known as predisposition sites, such as the branching points of large arterial trees and the inner curvature of the aortic arch [24].

Atherogenesis onset is characterized by the infiltration in the arterial wall of low-density lipoproteins (LDLs), which are then retained in the endothelium and oxidized. These oxidized LDLs (oxLDLs) trigger endothelial dysfunction at atheroprone sites, impairing laminar flow, promoting inflammatory cell infiltration and adhesion, and rearranging VSMC phenotype, eventually leading to arterial wall thickening, narrowing of the vascular lumen, accumulation of lipids, and the formation of plaques [23, 25]. Lesion progression can trigger plaque rupture or erosion, causing ischemic events, such as ischemic stroke and myocardial infarction (MI). Indeed, as the atheromatous plaque evolves, it is stabilized by a fibrous cap, but at later stages becomes thinner and susceptible to rupture, resulting in the aforementioned acute events [11, 15].

Additionally, pathologies such as coronary artery disease (CAD) and restenosis display similar or related features, being therefore classified as atherosclerosis-related conditions. More in detail, CAD is a localized form of atherosclerosis affecting coronary vessels. The main treatment option is surgical angioplasty associated with drug-eluting stent placement [26]. However, the efficacy of this solution can be temporary, favoring the formation of in-stent obstructions, called stenosis/restenosis, in the first 6 months post-surgery, as reported in 25–50% of cases [27]. At the cellular level, restenotic obstructions mainly rely on EC activation and the re-activation/de-differentiation of VSMCs that migrate to the site of injury, causing neointimal hyperplasia [28].

20.1.3 Pulmonary Artery Hypertension

Hypertension is a disorder that depends on complex genetic and environmental factors. PAH is characterized by progressive pulmonary vascular remodeling that gradually leads to narrowing of the vessel, increased pulmonary artery resistance, and imbalance between vasoconstriction and vasodilatation, events that eventually result in right ventricular maladaptive hypertrophy, heart failure, and premature death [29–31].

PAH can be classified into different subcategories, including idiopathic PAH, heritable PAH, and PAH associated with other diseases [32]. However, as mentioned above, all types of PAH share common aspects, including PAEC proliferation; PASMC proliferation, migration, and contraction; inflammation; and fibroblast proliferation, activation, and migration [32].

During PAH pathogenesis, vascular remodeling is guided by several molecular processes, with dysregulation evident across the different layers of the vessel wall; such mechanisms include, but are not limited to, a hypoxic microenvironment, growth factor changes, signaling pathway modulation, metabolic imbalance, and epigenetic modifications [33]. At the cellular level, it is characterized by alterations of PAEC biology, leading to a dysfunctional endothelium and endothelial-to-mesenchymal transition (EndMT). Further alterations pertain to apoptosis resistance, phenotype conversion, proliferation, and migration of PASMCs, alongside fibroblast accumulation [34–36].

Of note, PASMCs and PAECs are the principal cell types contributing to the pathogenesis of PAH, and their dysregulation is significantly involved in the vascular remodeling occurring during systemic hypertension and PAH, correlating with the detection of specific sub-cellular populations [15].

20.2 DNA Methylation and Vascular Diseases

Through the chemical modification of nuclear DNA and its interactors, epigenetics translates the influence of the environment into gene expression regulation. Appropriate packaging of genomic DNA within the nucleus is of pivotal importance for proper gene transcription. DNA is tightly organized in wrapped structures, known as chromatin, containing an equal mass of proteins (histones) [37]. DNA rolls up with histone octamers, consisting of two H2A-H2B dimers surrounding H3-H4 dimers, to form nucleosomes [38]. The solid interactions among nuclear DNA, histones, and linker RNA lead to the formation of a nucleoprotein complex: this chromatin can be defined as euchromatin or heterochromatin, depending on its level of compaction [39].

DNA methylation is the best-characterized chemical modification. It is a ubiquitous epigenetic mechanism that occurs when a methyl group derived from S-adenosyl-methionine is bound to position 5 of the cytosine ring, forming 5-methyl-cytosine (5mC) [40]. This heritable modification can occur at cytosine-guanine dinucleotide sites, called CpG islands, and it is commonly associated with repression and gene silencing [41], especially if it is happening at transcription start sites (TSSs), by preventing transcription factors (TFs) from approaching DNA [42]. This modification is commonly catalyzed by the activity of a family of enzymes named DNA methyltransferases (DNMTs) [43]. This family comprises three active members, identified in mammalian cells as DNMT1, DNMT3A, and DNMT3B [43, 44]. Among them, DNMT1 is broadly expressed in mammalian cells and has a role in the maintenance of the mitotic inheritance of methylated DNA [45]. The Ubiquitin-like with PHD and Ring Finger Domains 1 (UHRF1) has a crucial role in

this process, favoring DNMT1 in the recognition of hemimethylated DNA [46]. On the other hand, DNMT3A and DNMT3B are the main factors responsible for de novo methylation [47], which commonly occurs in early embryos [48] and at different stages of development [49].

Conversely, the removal of 5mC, also known as DNA demethylation, leads to DNA transcription and can be achieved through passive or active mechanisms. As easily perceivable, passive demethylation occurs due to a lack of activity of DNMT1 during cell division [50]. On the other hand, active demethylation occurs through direct enzymatic removal of the methyl group from 5mC. First described in 2009, it was reported that ten-eleven translocation (TET) methylcytosine dioxygenases family plays a crucial role in DNA demethylation via the enzymatic oxidation of 5mC to generate 5-hydroxymethylcytosine (5hmC) in a Fe(II)- and α-ketoglutarate (α-KG)-dependent manner [51]. 5hmC is a crucial intermediate that is gradually replaced by unmethylated cytosines through several mechanisms. In particular, the oxidation reaction can proceed to 5-formylcytosine (5fC) and 5-carboxylcytosine (5caC) [52, 53]. Furthermore, the pathway is continued by thymine-DNA glycosylase (TDG), which recognizes and excises 5caC (and also 5fC) from DNA. The latter reaction depends on base excision repair (BER), engaged for completing the DNA demethylation cycle [54] (Fig. 20.1). Although 5hmC has broadly been characterized as an intermediate state of the demethylation process, it emerged to represent a crucial epigenetic mark; increasing data are suggesting that 5hmC may represent a stable epigenetic mark [55]. 5hmC distribution was reported to have a pivotal role in the acquisition of cellular imprinting during cellular differentiation, but of outmost importance, it is associated with the recruitment of TFs, and therefore is entitled to actively regulate gene transcription processes [56, 57].

Besides its involvement in physiological processes, the regulation of DNA methylation is one of the crucial epigenetic mechanisms in the development and progression of cancer, where its patterns are globally disrupted [58]. In recent decades, many other pathologies have broadly been related to DNA methylation impairment; among them, cardiovascular pathologies are prominent.

20.2.1 Atherosclerosis and DNA Methylation

Due to their strong dependence on environmental stimuli, epigenetic modifications have been progressively associated with atherosclerosis development. Thus, they are of pivotal interest for therapeutic and biomarker outcomes.

Decades of studies described that DNA hypomethylation seems to be preponderated during atherosclerosis progression [59, 60]. Nonetheless, focal DNA hypermethylation is now considered of crucial importance in atherosclerosis development [61, 62]. In particular, several reports describe how the regulation of DNA methylation can broadly affect atherosclerosis development and progression by directly targeting all the layers within the vessel, affecting ECs, immune cells, and VSMCs.

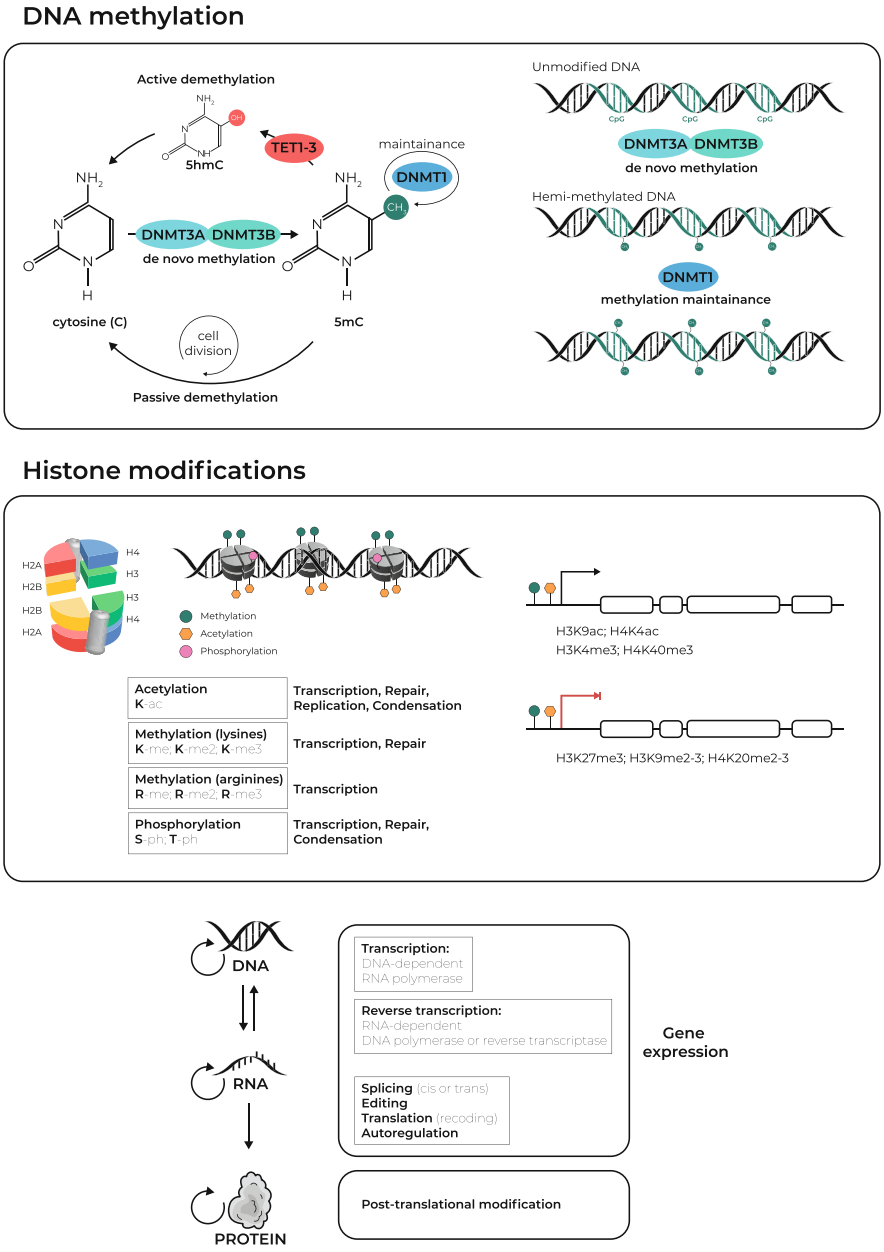

Fig. 20.1 Illustration of DNA methylation and histone modification effects on gene expression. Top: Mechanisms of DNA methylation and demethylation. During differentiation processes, DNA undergoes different chemical modifications. De novo methylation is mediated primarily by DNMT3A and DNMT3B. Upon DNA replication, newly synthesized strands lack methylation marks, but DNMT1 rapidly restores the correct methylation on a newly synthesized DNA. Adult patterns of methylation are erased by epigenetic mechanism involving TET enzymes. Bottom: The nucleosome is composed of double-stranded DNA wrapped around a core of histone proteins. Modifications such as acetylation and methylation of lysine residues on core histones are mutually exclusive. Acetylation on lysines is always associated with increased gene expression.

EC damage can be considered the first pathological trigger for atherosclerosis development, where oxidative stress, lipid deposition, and transcription of inflammatory mediators play a crucial role [63]. In the context of oxidative stress, the aberrant transcription of Src homology 2 domain-containing protein ($p66^{Shc}$) contributes to mitochondrial dysfunction, increased apoptosis, and endothelial functional alteration [64]. This pathological effect was prevented by the restoration of DNMT3B activity that resulted in $p66^{Shc}$ repressed transcription [65].

Treatment of ECs in vitro with oxLDL was reported to directly induce DNMT1-mediated methylation of Kruppel-like factor 2 (KLF2), abrogating endothelium-dependent vascular homeostasis, and so inducing a pro-atherogenic phenotype [66]. In addition, it induced increased methylation of the promoter regions of Src homology 2-containing protein tyrosine phosphatase 1 (SHP-1) [67]. Atherogenic stimuli are reported to induce a strong increase in expression of antioxidative enzymes (SOD2, catalase, and GPx) in infiltrating immune cells and migrating VSMCs, via altering 5mC status [68, 69].

Nonetheless, VSMCs were described to display a pro-atherogenic phenotype after SOD2 inhibition mediated by induction of DNA methylation; this effect was completely prevented with 5-azacytidine (5-aza), a very well-known DNMT1 inhibitor [70]. In line with this finding, UHRF1 was found to regulate the pro-atherogenic phenotype of VSMCs, directly cooperating with DNMT1, as a downstream target of platelet growth factor-BB (PDGF-BB) treatment [71].

20.2.2 Pulmonary Artery Hypertension and DNA Methylation

Epigenetics is involved in the development of PAH [72], so it may be a potential therapeutic target to ameliorate the clinical outcome of this severe disease [73]. Global alteration of DNA methylation status is reported to be clearly linked with PAH development and progression. In particular, the promoter region of the superoxide dismutase (SOD)2 gene was found to be hypermethylated through the activity of DNMT1 and DNMT3B in PASMCs isolated from spontaneously developing PAH (fawn-hooded) rats [74] and plexiform regions of PAH patients [75], overall contributing to the activation of hypoxia-inducible factor (HIF)1-α, and thus inducing a pro-proliferative and apoptosis-resistant state. This pathological phenotype can be abolished with the administration of 5-aza, which blunted DNA methylation on the SOD2 gene at different loci [75].

Furthermore, in PAECs, strong hypermethylation was found at the promoter of several genes and microRNAs involved in lipid metabolism, including ABCA1, ABCB4, ADIPOQ, miR-26A, and BCL2L11. ABCA1 was found to have reduced

Fig. 20.1 (continued) Lysine methylation is associated with gene activation when found at H3K4, and H4K40, but is associated with gene silencing at H3K27, H3K9, and H4K20

expression at the mRNA and protein levels, leading to novel therapeutic possibilities [76].

Global DNA methylation level reduction was also observed in fetal lambs exposed to long-term high-altitude hypoxia: this was associated with loss of the cyclin dependent kinase inhibitor (CDKN1A, p21), which caused aberrant PASMC proliferation in these fetuses, leading to PAH in the newborn [77].

Moreover, the presence of differential epigenetic marks was demonstrated between pulmonary venous-occlusive disease (PVOD) patients and PAH patients. This was due to hypermethylation of the granulysin (GNLY) gene, a cytosolic antimicrobial peptide found in the gDNA of explanted lungs and peripheral blood mononuclear cells [78]. This finding is noteworthy due to the lack of knowledge on this pathology, allowing early diagnosis and rapid discrimination of the two groups of patients.

20.3 The Histone Code and Vascular Diseases

As already described in previous paragraphs, genetic information is finely regulated by environmental stimuli and the accessibility of gene promoters to TFs.

Histones are the key component of chromatin and are rearranged in a spiral secondary structure known as a solenoid [79]. There are four types of histone protein: H2A, H2B, H3, and H4. They share a common structure, composed of a globular domain surrounded by flexible, protruding domains (also known as amino (N)- and carbon(C)-terminal tails). An octamer of dimers of each type of histone forms the main essence of chromatin: the nucleosome [37]. Within the three-dimensional structure of the solenoid, additional H1 histone proteins cooperate in the maintenance of chromosomal stability [80] As first discovered by the epigenetic pioneer Emil Heitz [81], there are two different chromatin states characterized by different accessibility capacities: euchromatin is related to active transcription because of its less compact and more accessible structure; on the other hand, heterochromatin is a tightly compact structure where transcription is prevented.

The protrusion of the N-terminal tails of all histones and of the C-terminal tails of H2A histones permits reversible covalent modifications, known as post-translational modifications (PTM), to control DNA accessibility and to regulate gene transcription [82]. Histone PTMs (hPTMs) include a wide variety of chemical modifications and constitute the so-called “histone code” [83] (Fig. 20.1). More in depth, lysine (K) was the first discovered to be acetylated or methylated [84], but further acylation, biotinylation, crotonylation, formylation, malonylation, sumoylation, and ubiquitination were discovered to collectively regulate conversions on this amino acid [85]. Arginine (R) can be either deiminated (converted to citrulline) or (mono-, di-, and tri-) methylated [86]. Serine (S), threonine (T), and tyrosine (Y) can be phosphorylated [87] and S and T can also be glycosylated [88]. In this complex scenario, hPTMs are the product of the dynamic activity of particular enzymes that deposit chemical modifications (writers), decipher the mark (readers), or proficiently remove the labels (erasers).

Mass spectrometry (MS) was the first approach able to decipher how epigenetic modifications could affect chromatin accessibility and, thus, regulate gene transcription. Taking advantage of the vastly technical benefits of high-throughput technologies, a more precise overview of histone marks has been possible in recent decades. Here, we will discuss the three main hPMTs: acetylation, methylation, and phosphorylation.

Histone Acetylation

Acetylation occurs predominantly on K residues of H3 and H4 histones. Histones exhibit a positive ionic charge that strongly interacts with negative acetyl groups, leading to a weak interaction with DNA, and enhancing chromatin accessibility. Acetylated histones are favorable binding sites for bromodomain-containing proteins (BRDPs), which orchestrate assembly of the transcriptional machinery [89]. Acetylation is predominantly enriched in peculiar regions inside the promoter and at the 5′ side of the coding sequences. Of note, H3 acetylation on K9 and K27 (H3K9ac and H3K27ac) is generally associated with enhancers and promoters of active genes.

Histone deacetylation is mediated by histone deacetylases (HDACs), generating a more compact structure of the nucleosome and, thus, leading to transcription impairment [90]. This mark can be reverted through the action of histone acetyl transferases (HATs; CBP/p300, MYST, and GNAT), which induce chromatin relaxation [90]. Of note, unbalance in the activity of these two families of "switchers" leads to aberrant acetylation patterns and the development of CVDs [91]. Indeed, abnormal histone acetylation is implicated in many diseases, including hypertension [92], PAH [93] and ventricular hypertrophy [94]. Therefore, approaches aimed at manipulating the acetylation/deacetylation balance, preventing an aberrant PTM state, might be an innovative therapeutic strategy for these pathologies. Overall, HDAC inhibitors have demonstrated to be protective agents in the development of several CVDs [95].

Histone Methylation

In contrast to acetylation, which is a clear mark of transcriptional activation, methylation of K and R residues can have different impacts on gene transcription. The most extensively studied histone lysine methylation sites are H3K4, H3K9, H3K27, H3K36, H3K79, and H4K20, although many other methylated lysine residues are present on H1, H2A, H2B, and at other H3 and H4 loci [96].

Histone methylation is mediated by lysine- and arginine-methyltransferases (KMTs, and PRMTs, respectively), taking advantage of S-adenosyl-L-methionine (SAM) as the methyl-donor. These enzymes can add up to three methyl groups, resulting in mono-, di-, or tri-methylation of lysines, and mono- or di-methylation of arginines. Generally, monomethylated lysine improves transcription, whereas the addition of further methyl groups can have a different effect depending on the residue (e.g., H3K4me2–3, H3K36me3, and H3K79me2–3). In addition, high plasticity in the recruitment of reader enzymes induces either gene transcription or

repression based on the methylation level at the residue (e.g., H3K4me2 and H3K4me3) [97].

Histone demethylases (HDMs) are the principal effectors of methyl group removal from arginine and lysine residues [98]. This process takes place through the activity of FAD-dependent amine oxidases [99] or FeII and α-KG-dependent dioxygenases [100].

Histone Phosphorylation

Despite being the first modification to be characterized, histone phosphorylation is somewhat less well understood than acetylation and methylation. Like with acetylation, phosphorylation produces a negative charge, reducing the positive charge carried by histones and leading to increased accessibility of chromatin. Histone phosphorylation is transient and dynamically coordinated by kinases and is mainly associated with gene activation mediated by specific TFs [101]. Histone phosphorylation has broadly been related to apoptosis, DNA repair, chromatin rearrangements during cellular division, and gene transcription [102]. Of note, one of the most studied phosphorylation sites is H3S10, which once phosphorylated by Aurora B, recruits enzymes that block H3K9 methylation; HATs then acetylate the lysine residue, further increasing transcriptional activity, in a mechanism known as phosphor-acetylation [103].

20.3.1 Atherosclerosis and Histone Modifications

Epigenetic de-regulation is a clear and well-known hallmark of atherosclerosis development. Global reduction in H3K27 tri-methylation was observed at a late stage in atherosclerotic plaques. Nevertheless, this was not related to alterations in global levels of the corresponding histone methyltransferase EZH2, the catalytic subunit of the polycomb repressive complex 2 (PRC2). Similarly, neither alterations in BMI1, a PRC1 complex component, which binds to H3K27me3, nor the expression of the histone demethylase JMJD3, which removes the methyl marks on H3K27, were reported [104]. In addition, H3K27me3 reduction was also reported to affect NF-κB activation in human coronary artery ECs, resulting in an increase in inflammatory status due to the aberrant transcription of adhesion molecules and cytokines [105].

G9a, another well-known epigenetic enzyme, is able to operate H3K9 mono- and di-methylation. In hyperhomocysteinemic ApoE$^{-/-}$ mice treated with methionine, G9a was reduced together with H3K9me2, resulting in the promotion of apoptosis in macrophages and increased plaque stability [106].

In VSMCs and macrophages of human carotid arteries, histone hyperacetylation—with a consequent reduction in methylation levels at H3K9 and H3K27—was reported to be a clear mark of atherosclerotic severity [107]. Furthermore, increased HAT-mediated histone acetylation was a potential regulator of VSMC differentiation due to the recruitment of serum response factor (SRF) and its cofactor, myocardin [108], that further mediate the activation of CArG box

elements of VSMC-specific marker genes [109]. Moreover, HDAC3 deficiency in myeloid cells was reported to improve plaque stability by increasing collagen deposition, and has been proposed as a macrophage polarization inducer, favoring the alternatively activated macrophage phenotype. These cells secrete TGFβ1, which can be considered a trigger promoting collagen deposition by VSMCs [110].

HDAC inhibitors have been proposed as intriguing suppressors of atherosclerosis development due to transcriptional reduction of a subset of inflammatory genes (e.g., TLR-induced IL-12p40 secretion in dendritic cells). They have been also described to induce other pro-atherogenic genes (e.g., COX-2 expression), suggesting that further efforts are needed to better describe their role in atherosclerosis [111].

Nonetheless, epigenetic mechanisms dynamically arrange gene expression, with the mutual relation between hPTMs and DNA modifications being of outmost importance. In particular, synergistic effects have been reported between DNMT3A/3B and H3K9 or H3K36 methylases (SETDB1, SUV39H1/2, EHMT1/2, and SETD2); in contrast, DNA methylation is completely abrogated by H3K4 mono-, di-, and tri-methylation [112].

In addition, H3K27 methylation (operated by EZH2) stimulated foam cell formation in $ApoE^{-/-}$ mice, improving atherosclerosis development. In particular, EZH2 enrolls DNMT1 and methyl CpG-binding protein-2 (MeCP2). The DNMT1-MeCP2 complex then promotes the methylation of the ATP-binding cassette transporter A1 (ABCA1) promoter, whose inhibition improves atherosclerosis [113].

20.3.2 Pulmonary Artery Hypertension and Histone Modifications

Over the past decades, increasing evidence has suggested that hPTMs, and especially aberrant histone acetylation, influences PAH development. Of note, epigenetics-based therapies directly targeting histone acetylation could be innovative approaches for the treatment of PAH.

Firstly, EZH2, which operates H3K27-specific methylation, was found upregulated in human PASMCs, regulating proliferation, migration, and anti-apoptotic processes [114].

Transcriptional regulator myocardin-related transcription factor A (MRTF-A) interacts with NF-κB, halting H3K4 methyltransferase on cell adhesion molecule (CAM) promoters in response to hypoxia. This complex then promotes ICAM-1, VCAM-1, and E-selectin transcription, improving leukocyte migration to the pulmonary vascular wall [115].

Moreover, in the monocrotaline (MCT)-induced PAH rat model, the histone methyltransferase nuclear receptor binding SET domain 2 (NSD2) was recently found to be involved in the development of the disease. In particular, NSD2 knockdown reduced H3K36me2, with an effect on pulmonary arterial remodeling through autophagy inhibition, pulmonary artery pressure normalization, and right ventricular hypertrophy reversion [116].

Histone acetylation also plays a critical role in PAH. Idiopathic PAH patients and chronically hypoxic rats had upregulated HDAC1 and HDAC5 in several tissues,

including heart, lung, and pulmonary arteries [117]. Increased histone acetylation is further induced by hypoxia and modulates hypoxia-inducible factor-1α (HIF-1α) binding levels in the endothelin-1 (ET-1) gene core promoter region in spontaneously hypertensive rats (SHR) [118]. In line with this, HDAC inhibitors were reported to attenuate hypertensive responses in SHRs [119]. Furthermore, valproic acid (VPA) prevented hypertension development via reduction of mineralocorticoid receptor (MR) transcription by increasing its acetylation [120]. In addition, trichostatin A (TSA), another HDAC inhibitor, effectively reduced blood pressure and vascular inflammation in SHRs [121].

An increased accumulation of HDAC4 and HDAC5 in the PAEC nuclei of PAH patients was reported to impair myocyte enhancer factor 2 (MEF2) activity, inducing connexin 37 (Cx37), connexin 40 (Cx40), KLF2, and KLF4, and resulting in an improvement in cell proliferation [122]. Moreover, PASMCs had increased levels of HDAC3 and HDAC6, regulating SOD3 expression and inhibiting BAX-induced cell death programs, and thus enhancing cell proliferation [123]. Furthermore, a global increase in acetylation was observed in Angiotensin (AII)-treated VSMCs. AII is a strong vasoconstrictor directly responsible for triggering hypertension. AII-treated VSMCs had elevated HDAC5 phosphorylation in a time- and dose-dependent manner. This led to HDAC5 export out of the nucleus, causing a reduction in MEF2 transcription [124].

20.4 ncRNAs and Vascular Diseases

Once acknowledged that only approximately 2% of the mammalian genome possesses a coding potential, the past decades have seen increased interest in non-coding transcripts. Initially considered to be artifacts, these molecules are now recognized as active players in the gene regulation machinery [125]. With the advent of new-generation sequencing, the catalog of non-coding RNAs (ncRNAs) and the understanding of their functions in cardiovascular health and disease has grown exponentially [126–128].

ncRNAs can be classified into two major groups based on their size: small (<200 nucleotides) and long (>200 nucleotides) ncRNAs [129, 130]. Small ncRNAs comprise regulatory elements, such as microRNAs (miRNAs, miRs) and Piwi RNAs (piRNAs), and the more stably expressed housekeeping ncRNAs, such as small nucleolar RNAs (snoRNAs) and transfer RNAs (tRNAs). The classification of long ncRNAs (lncRNAs) is more complex and can be based on genomic localization, specific function, and mechanism of biogenesis. Of note, a particular class of lncRNA defined as circular RNAs (circRNAs) has gained increasing interest in experimental biology because of its peculiar features [131–134]. Alternatively, ncRNAs can be classified on the basis of their role into infrastructural (i.e., small nuclear and nucleolar RNAs, ribosomal RNAs) and regulatory RNAs (i.e., miRNAs, lncRNAs, piRNAs, and small interfering RNAs) [135]. Notably, it has been reported that regulatory ncRNAs may actively participate in gene regulation by modulating

chromatin structure, adding another layer of complexity to the epigenetic landscape of vessel biology [136, 137].

Increasing evidence has demonstrated the importance of ncRNAs during the development of different pathologies, including those of the vascular system [15]. While the current knowledge on miRNAs is extensive, further investigation is required to elucidate the contribution of lncRNAs to vascular pathophysiology [125, 138].

miRNAs are small ncRNAs with a size of ~20 nucleotides (nt) that regulate gene expression at the post-transcriptional level. miRNAs play an essential role in controlling several pathways and biological processes, such as development, differentiation, apoptosis, and survival [139]. Because they have critical roles in the regulation of gene expression during normal physiology, being involved in many different biological processes, their malfunction has been directly associated with many diseases, including those of the vascular system [140, 141].

In the nucleus, miRNAs are first transcribed into a typical hairpin primary miRNA (pri-miRNA) structure by RNA polymerase enzymes, namely Pol II. Then, the pri-miRNA is specifically recognized and cleaved by the microprocessor complex DGCR8/Drosha, producing precursor miRNAs (pre-miRNAs). Pre-miRNAs are exported to the cytoplasm by the Exportin 5 (XPO5)/RanGTP complex, where they are further processed by Dicer, which removes the terminal loop, resulting in an intermediate miRNA duplex. Interestingly, the directionality of the strands determines the name of the mature miRNA form: the 5p strand arises from the 5′ end of the pre-miRNA hairpin, while the 3p originates from the 3′ end. This duplex is further processed by the Argonaute (AGO) protein family. Although both strands can be loaded onto AGO proteins, generally, the strand with lower 5′ stability or 5′ uracil is preferentially utilized and is considered the guide strand. The non-loaded strand, called the passenger strand, is unwound from the guide strand and eventually degraded. The guide strand is loaded onto the RNA inducing silencing complex (RISC), finally enabling mRNA target recognition. miRNA and RISC promote the recruitment of poly(A)-binding proteins, which shorten the poly (A) tail of mRNAs [141] (Fig. 20.2). Once the miRNA is loaded onto the RISC, miRNA:mRNA interaction can occur. The miRNA binds its mRNA target in a sequence-specific manner, suppressing gene expression by either blocking the translational mechanism or activating nucleolytic mRNA degradation mainly through targeting the 3′-untranslated region (3'UTR) of the target gene [129]. The specific sequence in the 3'UTR mRNA target recognized by the miRNA is known as the miRNA Recognition Element (MRE), which includes a region of ~8 nt at the 5′ end called the seed sequence [129, 142, 143] (Fig. 20.2).

LncRNAs represent a large and heterogeneous class of ncRNAs, comprising a wide catalog of transcripts that differ in their biogenesis, genomic origin, and function [126, 144]. Most lncRNAs are synthesized via spliceosome-mediated processes on Pol II and Pol III, and present 5′-caps as well as 3′ polyadenylated (polyA) tails [145, 146]. LncRNAs are usually subject to splicing events, being presumably transcribed and processed similarly to mRNAs [145], from which they differ mainly by the absence of translational coding potential. Mature transcripts are

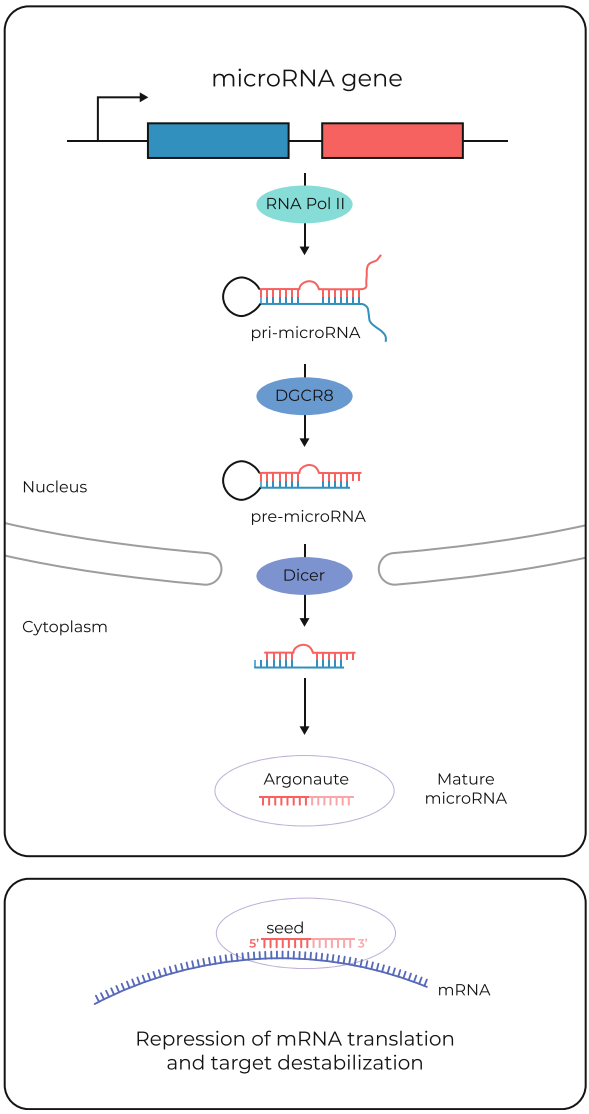

Fig. 20.2 MicroRNA biogenesis. A microRNA is transcribed as pri-microRNA, which matured first in a pre-microRNA and finally in the mature microRNA. The latest binds the 3'UTR of the target gene inhibiting gene expression

low-copy number molecules with specific sub-cellular localization [131, 147]. Of note, this is merely a general definition, as several exceptions have been reported so far. For instance, lncRNAs of intergenic origin differ from mRNA-like lncRNAs due to the lack of a polyA tail. Additionally, a few lncRNAs have been shown to contain functional cryptic open reading frames (ORFs) that may give rise to small peptides [148].

Generally, lncRNA sequences are poorly conserved among species; however, this might be compensated by the maintenance of structural properties warranting

conserved functionality [149]. Additionally, their expression is closely modulated at developmental stages, displaying tissue and cell specificity as well, this being suggestive of their involvement in gene regulation [130, 150].

Based on their direction with respect to the coding gene, lncRNAs transcripts are categorized as sense, antisense, exonic, and intronic. Other well-known forms are promoter bidirectional non-coding lncRNA, enhancer lncRNAs (eRNAs), intergenic lncRNA (lincRNAs), and circular RNAs (circRNAs) [151, 152] (Fig. 20.3). Moreover, these transcripts can act via *cis*- or *trans*-regulatory circuits [22, 153]. Herein, we present a partial description of their biological functions, which might be only the tip of the iceberg of the roles of lncRNAs in biological systems. LncRNAs can function as: (1) signals, responding to specific stimuli in the intra- and/or extracellular milieu; (2) decoys, sequestering biological active molecules and altering chromatin accessibility; (3) scaffolds, participating in specific functional multi-riboprotein complexes; (4) guides, orchestrating chromatin-modifying complexes; (5) mediators, contributing to 3D nuclear organization; and (6) molecular sponges, sequestering active miRNA molecules to inhibit and fine-tune their function [131, 144] (Fig. 20.3).

This class of ncRNA is, therefore, emerging as a fundamental player in the regulation of gene expression at the chromatin, DNA, transcriptional, and post-transcriptional levels in health and disease [15, 154]. As lncRNAs impact incidence and outcome of human diseases, a better understanding of their putative roles in vascular pathobiology merits further investigation [125, 155]. Indeed, key findings in the field warrant the potential use of these molecules in a clinical setting, as further described in this chapter.

Finally, among lncRNA transcripts, it is worth mentioning circRNAs in further detail. Like the vast majority of RNAs, circRNAs are generally transcribed by Pol II as a pre-messenger RNA (pre-mRNA) [156]. However, unlike linear transcripts, circRNAs undergo an alternative splicing event, namely back-splicing, promoting the formation of RNA loops. This tail-to-head folding exploits canonical splicing sites and proceeds from an upstream 5′ splice site (acceptor) to a downstream 3′ one (donor) [133, 134]. Consequently, covalently closed RNAs, displaying neither 5–3′ polarity nor a polyA tail, are generated. Given their structure, circRNAs are more stable than linear RNAs and less sensitive to RNAse R exoribonuclease degradation [157]. Finally, RNA loops can be secreted in exosome vesicles in response to specific stimuli and, subsequently, detectable in body fluids, such as serum and saliva [22]. Hence, circRNAs are promising candidates for human disease prognosis, prevention, and treatment.

In conclusion, compelling evidence has shown that ncRNAs such as miRNAs, lncRNAs, and circRNAs could be exploited for the development of new RNA-based therapeutics. Indeed, despite the limitations of such approaches, they represent a valuable tool for the generation of novel strategies in vascular disease.

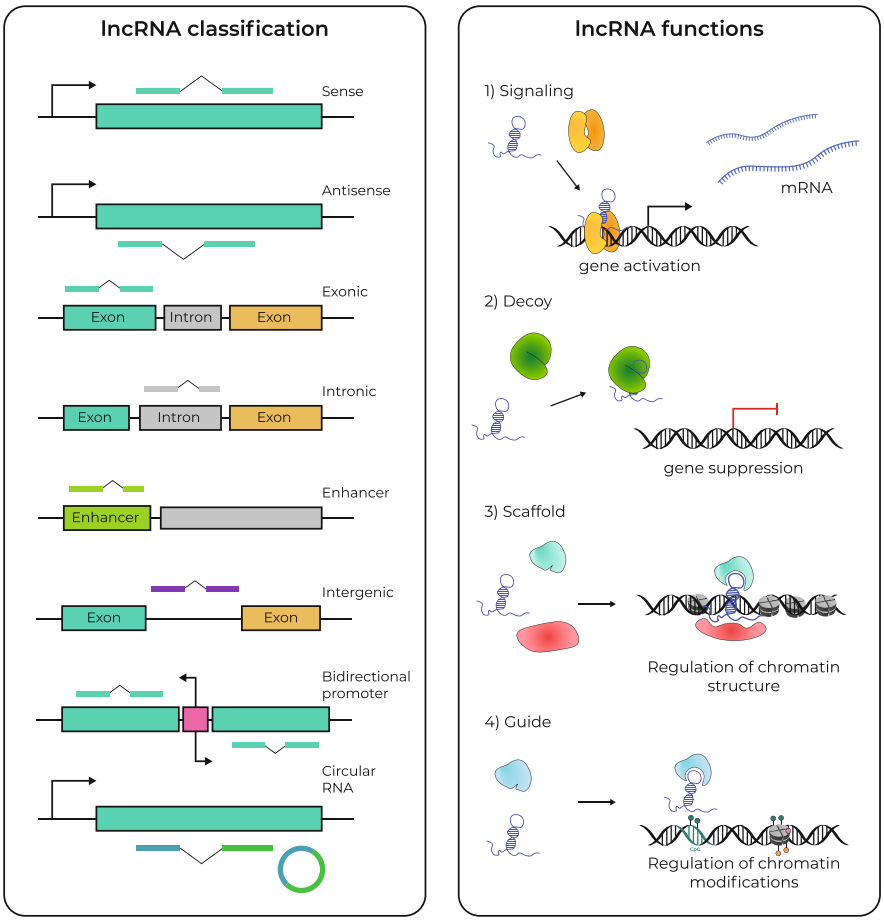

Fig. 20.3 Schematic classification of the different types and functions of lncRNAs. Left: LncRNAs are classified according to their genomic position: Exon are represented in green and orange, enhancer in light green, promoter in pink, while intron in gray. Arrows indicate transcription starting sites. Right: LncRNAs modulate gene expression by different mechanisms: in the nucleus they might guide TFs to promoter region (1) or sequester TFs inhibiting their function (2); They can also modulate chromatin structure by acting as scaffold (3) or guide (4)

20.4.1 Atherosclerosis and miRNAs

Atherosclerosis is an inflammatory disease in which hypercholesterolemia plays a central role in pathologically activating vascular cells and the immune system [158, 159]. These processes might be further accelerated by conditions such as diabetes mellitus and hypertension [160]. Carotid and coronary artery disease are two major atherosclerotic conditions and are the primary cause of stroke and heart attack, respectively. miRNAs are involved in every stage of the biological processes responsible for atherosclerosis development. Indeed, numerous miRNAs (miR-21-5p, −34a-5p, −146a-5p, −146b-5p, −210-3p) have been found to be upregulated in plaques from aorta, carotid, and femoral artery [161].

A detailed analysis of the single vessel layers shows how miRNAs are specific for the endothelium, infiltrating immune cells, and VSMCs. For instance, several miRNAs have been associated with dysfunction of the endothelium in vascular diseases. miRNAs involved in endothelium senescence include miR-34a, −217, and -146a, whereas others like miR-200c, −126, −10a, and -181b are modulated by oxidative stress and pro-inflammatory factors [160, 162]. Adhesion molecules are dysregulated by the misexpression of miRNAs such as miR-126, −31, and −17–3p [162]. Interestingly, the endothelium has been also found to secrete microparticles containing miRNAs, such as miR-19b, which affects other cellular components during plaque formation [163].

During atherosclerosis development, oxLDLs are loaded into monocytes/macrophages and VSMCs within the intimal layer, leading to the formation of foam cells and activated VSMCs. The role of macrophages during cellular adhesion, lipid uptake, and inflammatory responses has an important impact on plaque development. Several miRNAs, such as miR-99b, −152, −125a-5p, and −155, have been observed to be modulated in monocytes/macrophages within atherosclerotic plaques, influencing important steps, such as the accumulation of foam cells [160, 162]. Also, miR-10a plays a role in the pro-inflammatory contribution to the malfunctional endothelial phenotype during atherosclerosis [164]. In addition, miR-21 has been found to be increased in monocytes, driving atherosclerosis progression [165]. As for ECs, also macrophages are able to transfer miRNAs (i.e., miR-146a, −128) through extracellular vesicles [165]. However, macrophages might also secret miRNAs in conjunction with high-density lipoprotein (HDL); indeed, in human plasma, miR-210 has been found dysregulated in atherosclerosis patients [165].

Atherosclerotic lesions are also characterized by the presence of migratory VSMCs in the intimal layer. Some examples of miRNAs with an important impact on VSMC phenotypic switching during disease development are miR-143/145, −22, −21, −221/222, and − 128 [159, 160, 165]. Furthermore, during the final phase of atherosclerosis, involving plaque destabilization and rupture, those structures are characterized by a rich content of inflammatory macrophages and VSMCs. The levels of miR-322, −100, −127, −133a/b, and −145 are, for instance, significantly higher in patients with destabilized plaques [160, 165]. Moreover, VSMC differentiation—promoted by the progression of fatty streaks into a fibrous

cap—is regulated by miRNAs, such as miR-26a, which targets transforming growth factor-β (TGF β), a critical factor for VSMC differentiation and apoptosis [162]. Other short ncRNAs, such as miR-29b, are involved in the regulation of VSMC migration, by targeting metalloproteases (MMPs), and contractility, by targeting critical genes, such as KLF4 in the case of miR-145 or −128 [162, 166].

20.4.2 Atherosclerosis and lncRNAs

In contrast to miRNAs, lncRNA investigation has not delivered a wealth of information on their role in atherosclerosis. Nonetheless, based on a handful of key studies, it is understood that lncRNAs can alter the onset and progression of atherosclerosis by affecting a wide array of processes governing the transcriptional program of all three principal cell components of atherosclerotic plaques [28, 150].

A paradigmatic example is the antisense ncRNA known as ANRIL, found in the Inhibitors of Cyclin Dependent Kinase 4 (INK4) locus [167]. ANRIL functions as a guide lncRNA binding specific subunits of the Polycomb repression complexes 1 and 2 (PRC1 and 2) and mediates interaction with target promoters [167]. This fosters H3K27 tri-methylation and transcriptional repression of the INK4 locus in *cis*. Interestingly, ANRIL exerts its atheroprotective function also in *trans* [167, 168]. Additionally, it is expressed and modulated by all the three principal cell components, so is directly associated with the severity of atherogenesis and lesion progression [169, 170].

Different lncRNAs (i.e., PUNISHER, LEENE, sONE) have been found to participate in EC differentiation, homeostasis, and function [171–173]. Additionally, several others have been identified in atherogenesis, for instance, MALAT1, MEG3, LINC00323–003, and MIR503, as well as the circRNAs cZNF292, cAFF1, cDENND4C, and cTHSD1, which are differentially expressed upon hypoxic conditioning, a stimulus known to evoke EC dysfunction [22, 174, 175]. Nonetheless, only a few have been extensively characterized so far. Human metastasis-associated lung adenocarcinoma transcript 1 (MALAT1) was one of the first lncRNAs studied in human disease. It is highly expressed in vascular endothelium and is de-regulated in several CVDs upon distress signals. Indeed, its down-regulation produces a shift from a proliferative to a migratory phenotype, causing aberrant vessel growth in vivo [176]. Coherently, profound cell cycle dysregulation was revealed by gene expression profiling [176]. Nonetheless, recent literature suggests that its role on atherosclerositic deterioration is due to increased accumulation of hematopoietic cells [176, 177].

SENCR is a vascular cell-specific lncRNA expressed in human ECs and VSMCs and plays a role in EC differentiation and homeostasis [178, 179]. Interestingly, Baker's laboratory uncovered that SENCR may be involved also in endothelium dysfunction. Indeed, its expression is significantly reduced in ECs isolated from patients with premature CAD compared to control subjects, suggesting a central role in atherosclerosis [180]. In addition, SENCR protects against vascular damage by

participating in adherent junction maintenance, membrane integrity, and permeability of ECs [179].

Aberrant phenotypic transition and apoptosis impairment of VSMCs are central events in atherogenesis initiation and progression. Along with miRNAs, lncRNAs, and circRNAs participate in the fine-tuning of VSMC biology. This can be achieved either independently or through epigenetic networks. For instance, circLRP6/miR-145 axis modulation is essential for VSMC homeostasis and may play a role in atherogenesis as well [181].

Lnc-Ang362 is an AII-induced lncRNA that enhances aberrant cell growth in VSMCs. It is co-transcribed with miR-221 and -222, well known for their detrimental effect on vessel biology [182]. In light of this, the effects of lnc-Ang362 on proliferation may likely be due to its role as the host transcript of the 2 miRNAs. Noteworthy, as these molecules are modulated in the same manner, further studies might cement lnc-Ang362 as a therapeutic target in atherosclerosis [183]. Similarly, the lncRNA SMILR concurs in atherogenesis by promoting aberrant cell proliferation via modulation of CENPF mRNA. While its overexpression results in neointimal hyperplasia, its disruption ex vivo promoted adverse vascular remodeling [184]. The above-mentioned lncRNA SENCR was first described in human coronary artery VSMCs. In this context, SENCR contributes to maintain cell differentiation by acting on cell contraction-related genes such as myocardin (Myocd), while reducing the pro-migratory gene signature (Midkine/Pleiotrophin). Therefore, it can counteract the pathological migration of VSMCs to the neointima during atherosclerotic lesion formation [178].

Inflammation and innate immune responses are also crucial events in atherosclerosis and act mainly through the modulation of monocyte and macrophage function. Myocardial infarction associated transcript (MIAT) is upregulated in vulnerable atherosclerotic plaques and fosters persistent, aberrant immune responses. Of note, MIAT ablation alleviates injury progression by acting on the miR-149-5p/CD47 axis and promoting macrophage clearance capacity (efferocytosis), thus reducing the inflammatory burden [185]. Concordantly, the monocyte- and macrophage-enriched lncRNA PELATON is upregulated in vulnerable plaques and is localized at inflammation sites. Its knockdown ameliorated atherosclerosis progression by promoting macrophage phagocytosis by acting in *trans*, and unlike MIAT, PELATON did not affect efferocytosis. Further studies on its coding potential may provide deeper insight into its role in macrophages [186].

RAPIA is a lncRNA that is highly expressed in advanced stages of atherosclerosis; its silencing protects against the worsening of vascular conditions by reducing proliferation and triggering the apoptotic program in macrophages via the miR-183-5p/ITGB1 pathway [187].

Finally, the already mentioned lncRNA MALAT1 is a modulator of the innate immunity response and is atheroprotective. Further studies are needed to better define its mechanism of action; nonetheless, it seems to maintain vascular functionality via a network of ncRNA-mediated processes involving mascRNA and NEAT1 [188].

20.4.3 Pulmonary Artery Hypertension and miRNAs

Different studies have indicated the involvement of miRNAs during PAH development. Normalization of some of these ncRNAs has been reported to blunt experimental pulmonary hypertension.

Considering lung as a bulk tissue, screening analyses were performed to evaluate miRNA modulation in PAH patients. A first study, for instance, measured 337 miRNAs in PAH and control lungs, identifying six statistically significantly upregulated miRNAs (miR-450a, −145, −302b, −27b, −367, and − 138), while only one, miR-204, was found to be down-regulated in pathological specimens [189]. The authors then linked miR-204 to a fundamental pathway able to sustain PASMC proliferation and resistance to apoptosis, which involves the signal transducer and activator of transcription 3 (STAT3), the protein tyrosine phosphatase non-receptor type 11 (SHP2) expression, the proto-oncogene tyrosine-protein kinase (Src), and nuclear factor of activated T cells (NFAT) [189].

Among miRNAs playing a role in PASMCs, we also have the miR-17–92 cluster. It contains six miRNAs (miR-17, miR-18a, miR-19a, miR-19b, miR-20a, and miR-92a) transcribed as one common pri-miRNA [190]. miR-17–92 expression was found reduced in PASMCs from patients with PAH [191], and was associated with decreased levels of SMC differentiation proteins, such as α-smooth muscle actin (α-SMA), transgelin (SM22α), and calponin, indicating a direct correlation between an SMC differentiation phenotype and miR-17–92 expression [191].

Another polycistronic miRNA cluster involved in PAH development is miR-143/145, a highly specific ncRNA expressed by SMCs of different origins [192]. Its expression is increased in human PAH as well as in the hypoxia-induced mouse model of PAH [193]. miR-143/145 cluster expression in PAH relies on down-regulation of bone morphogenetic protein receptor type II (BMPR2). Indeed, transforming growth factor-β (TGF-β) and bone morphogenic protein 4 (BMP4) activate Myocd expression and nuclear translocation of Myocd-related transcription factors (MRTFs), respectively, resulting in increased expression of miR-143 or miR-145, with concomitant repression of KLF4 expression, leading to the activation of a contractile gene program in SMCs [194]. Finally, negative manipulation of miR-145, but not miR-143, inhibited hypoxia-induced PAH in mice [193].

The miR-17-92 cluster has been shown also to control endothelium dysfunction in PAH. The development of familial PAH might depend on mutations of BMPR2 [195]. The link between this gene and the pathology relies on miR-17–5p and miR-20a, two members of the miR-17–92 cluster that directly target BMPR2 in PAECs. Interleukin 6 (IL-6) upregulates miR-17/92 expression through increased transcriptional activity of STAT3 [196]. Thus, inhibition of BMPR2 upregulates miR-17–92 expression, which then induces PAEC proliferation and reduces their apoptotic rate, resulting in the development of PAH.

Very recently, a study from Zhang and colleagues demonstrated that miR-483 is down-regulated in the serum of idiopathic PAH patients, with a direct inverse correlation with the severity of the pathology. They also demonstrated that miR-483 directly regulates several genes, including TGF-β, TGF-β receptor

2 (TGFBR2), β-catenin, connective tissue growth factor (CTGF), interleukin-1β (IL-1β), and endothelin-1 (ET-1), all of which are involved in PAH pathogenesis. Finally, an in vivo approach demonstrated that EC-specific miR-483 overexpression in a model of PAH was able to blunt the typical clinical outcomes, such as increased pulmonary vascular pressure and right ventricular hypertrophy [197].

20.4.4 Pulmonary Artery Hypertension and lncRNAs

The modulation of lncRNAs evoked by stimuli such as hypoxia and oxLDL is involved in de-regulation of PAECs, causing endothelium dysfunction and EndMT. Some examples of modulated ncRNAs are GATA6-AS, H19, MIR210HG, MEG9, MALAT1, and MIR22HG [198, 199], although their contribution to pulmonary hypertrophy is still largely unknown.

Additionally, PASMCs contribute to PAH progression via aberrant proliferative and migration and impairment of apoptosis. These phenotypic changes are sustained by a variety of cytokines and growth factors (i.e., PDGF-BB, NOTCH, Ang II, IL-1, FGF-2, and IGF-1) and lead to aberrant vascular remodeling [33].

The hypoxia-induced lncRNA of the Hoxa cluster antisense RNA 3 (lncRNA Hoxaas3) is upregulated in pulmonary hypertension. It promotes pathology progression by boosting PASMC over-proliferation and accelerating the cell cycle; hence, it may be a suitable therapeutic target to weaken damage progression [200].

The lncRNA regulated by PDGF-BB and TGF-β, namely LnRPT, has been identified in rat PASMCs and described as a putative therapeutic tool. Mechanistically, LnRPT controls PASMC proliferation by inhibiting the Notch pathway. Coherently, this regulatory capacity is reduced upon PDGF-BB-mediated LnRPT down-regulation. Additional studies may provide further insight into the use of LnRPT in PAH treatment [201].

In contrast, the lncRNA H19 is upregulated upon PDGF-BB stimulation in a rodent model of hypoxic-induced damage. It participates in PAH progression by promoting cell proliferation and inflammation through the regulation of the miR-let7b/AT_1R axis [202].

The lincRNA Cox2 plays an important role in the modulation of innate immunity by activating or repressing specific responsive genes in mouse macrophages, thereby regulating inflammatory processes [203]. Interestingly, it is upregulated in the peripheral blood of PAH patients and is also modulated in PASMCs upon exposure to hypoxia. Its detrimental effect on PAH development is considered to function via miR-let7A/STAT3 axis regulation [204].

20.5 Conclusion

The body of evidence discussed in this chapter strongly indicates that chromatin modifications are directly involved in the pathogenesis of CVDs. This is particularly important since, over the past years, the incidence of CVDs has increased greatly.

Fortunately, the research focusing on CVDs has grown in step, with many studies aimed at understanding the role of epigenetic processes in CVD development. It is, then, easy to understand that the definition of personalized epigenetic patterns can help to diagnose the causes of such pathologies and, thus, to identify tailor-made therapeutic approaches. Furthermore, recent technologic advances both in terms of data generation and analysis have allowed the scientific community to create detailed epigenetic maps of CVDs, providing clinical practice with new tools to link environmental effects with traditional risk factors: this includes also the possibility to eventually predict individual response to drug treatments.

Fully deciphering the epigenetic networks of the vasculature is, however, still far from completion. Thorough comprehension of the links between epigenetic mechanisms and specific tissue transcriptional programs is needed. This is particularly important for the development of therapies aimed at normalizing altered programs in specific cell types of the vessels. Indeed, the final aim would be to target epigenetic mechanisms in specific tissue compartments to avoid undesirable effects provoked by the modulation of gene expression elsewhere. Therefore, further studies are critically needed in order to personalize these new types of therapies. Because they are able to target all the types of epigenetic mechanisms discussed above, current RNA-based therapeutics might be able to fulfill this unmet need: these include silencing RNAs (siRNAs), antisense oligonucleotides (ASOs), lnRNAs, miRNAs, RNA aptamers, single guide RNAs (sgRNAs) for CRISPR/Cas9 systems, etc. These methodologies can target any gene, including those codifying proteins with no enzymatic activity; this is in contrast with classical drugs, which are not able to do so. This sheds a very positive light on the possibility of being able to tackle the majority of diseases of the cardiovascular system in the near future.

Acknowledgment This work was supported by grants from the Horizon 2020 Research and Innovation Program under Grant Agreement (No. 828984), the Italian Ministry of Health (No. GR-2016-02364133), and Italian Ministry of Research (No. 2017HTKLRF) to L.E.

References

1. Murray CJ, Lopez AD (1997) Mortality by cause for eight regions of the world: global burden of disease study. Lancet 349(9061):1269–1276. https://doi.org/10.1016/S0140-6736(96)07493-4
2. Nabel EG, Braunwald E (2012) A tale of coronary artery disease and myocardial infarction. N Engl J Med 366(1):54–63. https://doi.org/10.1056/NEJMra1112570
3. Rose G (1964) Familial patterns in ischaemic heart disease. Br J Prev Soc Med 18:75–80. https://doi.org/10.1136/jech.18.2.75
4. Sharifi M, Futema M, Nair D, Humphries SE (2017) Genetic architecture of familial hypercholesterolaemia. Curr Cardiol Rep 19(5):44. https://doi.org/10.1007/s11886-017-0848-8
5. Waddington CH (2012) The epigenotype. 1942. Int J Epidemiol 41(1):10–13. https://doi.org/10.1093/ije/dyr184
6. H. WC (1956) The genetic assimilation of the bithorax phenotype. Evolution:1–13

7. Cantone I, Fisher AG (2013) Epigenetic programming and reprogramming during development. Nat Struct Mol Biol 20(3):282–289. https://doi.org/10.1038/nsmb.2489
8. Lanctôt C, Cheutin T, Cremer M, Cavalli G, Cremer T (2007) Dynamic genome architecture in the nuclear space: regulation of gene expression in three dimensions. Nat Rev Genet 8(2): 104–115. https://doi.org/10.1038/nrg2041
9. Egger G, Liang G, Aparicio A, Jones PA (2004) Epigenetics in human disease and prospects for epigenetic therapy. Nature 429(6990):457–463. https://doi.org/10.1038/nature02625
10. Quintavalle M, Condorelli G, Elia L (2011) Arterial remodeling and atherosclerosis: miRNAs involvement. Vasc Pharmacol 55(4):106–110. https://doi.org/10.1016/j.vph.2011.08.216S1537-1891(11)00318-1
11. van der Harst P, de Windt LJ, Chambers JC (2017) Translational perspective on epigenetics in cardiovascular disease. J Am Coll Cardiol 70(5):590–606. https://doi.org/10.1016/j.jacc.2017.05.067
12. Greco CM, Condorelli G (2015) Epigenetic modifications and noncoding RNAs in cardiac hypertrophy and failure. Nat Rev Cardiol 12(8):488–497. https://doi.org/10.1038/nrcardio.2015.71
13. Cavalli G, Heard E (2019) Advances in epigenetics link genetics to the environment and disease. Nature 571(7766):489–499. https://doi.org/10.1038/s41586-019-1411-0
14. Suades R, Cosentino F (2019) The environment, epigenetic landscape and cardiovascular risk. Cardiovasc Res 115(13):e147–ee50. https://doi.org/10.1093/cvr/cvz150
15. Stratton MS, Farina FM, Elia L (2019) Epigenetics and vascular diseases. J Mol Cell Cardiol 133:148–163. https://doi.org/10.1016/j.yjmcc.2019.06.010
16. Zarzour A, Kim HW, Weintraub NL (2019) Epigenetic regulation of vascular diseases. Arterioscler Thromb Vasc Biol 39(6):984–990. https://doi.org/10.1161/ATVBAHA.119.312193
17. Versari D, Daghini E, Virdis A, Ghiadoni L, Taddei S (2009) Endothelial dysfunction as a target for prevention of cardiovascular disease. Diabetes Care 32(Suppl 2):S314–S321. https://doi.org/10.2337/dc09-S330
18. Owens GK, Kumar MS, Wamhoff BR (2004) Molecular regulation of vascular smooth muscle cell differentiation in development and disease. Physiol Rev 84(3):767–801. https://doi.org/10.1152/physrev.00041.2003
19. Khoury MK, Yang H, Liu B (2020) Macrophage biology in cardiovascular diseases. Arterioscler Thromb Vasc Biol:ATVBAHA120313584. https://doi.org/10.1161/ATVBAHA.120.313584
20. Basatemur GL, Jørgensen HF, Clarke MCH, Bennett MR, Mallat Z (2019) Vascular smooth muscle cells in atherosclerosis. Nat Rev Cardiol 16(12):727–744. https://doi.org/10.1038/s41569-019-0227-9
21. Wilson PW, D'Agostino RB, Parise H, Sullivan L, Meigs JB (2005) Metabolic syndrome as a precursor of cardiovascular disease and type 2 diabetes mellitus. Circulation 112(20): 3066–3072. https://doi.org/10.1161/CIRCULATIONAHA.105.539528
22. Fasolo F, Di Gregoli K, Maegdefessel L, Johnson JL (2019) Non-coding RNAs in cardiovascular cell biology and atherosclerosis. Cardiovasc Res 115(12):1732–1756. https://doi.org/10.1093/cvr/cvz203
23. Matouk CC, Marsden PA (2008) Epigenetic regulation of vascular endothelial gene expression. Circ Res 102(8):873–887. https://doi.org/10.1161/CIRCRESAHA.107.171025
24. Chiu JJ, Chien S (2011) Effects of disturbed flow on vascular endothelium: pathophysiological basis and clinical perspectives. Physiol Rev 91(1):327–387. https://doi.org/10.1152/physrev.00047.2009
25. Libby P, Ridker PM, Maseri A (2002) Inflammation and atherosclerosis. Circulation 105(9): 1135–1143. https://doi.org/10.1161/hc0902.104353
26. Cassar A, Holmes DR, Rihal CS, Gersh BJ (2009) Chronic coronary artery disease: diagnosis and management. Mayo Clin Proc 84(12):1130–1146. https://doi.org/10.4065/mcp.2009.0391

27. Byrne RA, Joner M, Kastrati A (2015) Stent thrombosis and restenosis: what have we learned and where are we going? The Andreas Grüntzig lecture ESC 2014. Eur Heart J 36(47): 3320–3331. https://doi.org/10.1093/eurheartj/ehv511
28. Pierce JB, Feinberg MW (2020) Long noncoding RNAs in atherosclerosis and vascular injury: pathobiology, biomarkers, and targets for therapy. Arterioscler Thromb Vasc Biol 40(9): 2002–2017. https://doi.org/10.1161/ATVBAHA.120.314222
29. Galiè N, Humbert M, Vachiery JL, Gibbs S, Lang I, Torbicki A et al (2016) 2015 ESC/ERS guidelines for the diagnosis and treatment of pulmonary hypertension: the joint task force for the diagnosis and treatment of pulmonary hypertension of the European Society of Cardiology (ESC) and the European Respiratory Society (ERS): endorsed by: Association for European Paediatric and Congenital Cardiology (AEPC), International Society for Heart and Lung Transplantation (ISHLT). Eur Heart J 37(1):67–119. https://doi.org/10.1093/eurheartj/ehv317
30. Ryan JJ, Archer SL (2014) The right ventricle in pulmonary arterial hypertension: disorders of metabolism, angiogenesis and adrenergic signaling in right ventricular failure. Circ Res 115(1):176–188. https://doi.org/10.1161/CIRCRESAHA.113.301129
31. Lau EMT, Giannoulatou E, Celermajer DS, Humbert M (2017) Epidemiology and treatment of pulmonary arterial hypertension. Nat Rev Cardiol 14(10):603–614. https://doi.org/10.1038/nrcardio.2017.84
32. Simonneau G, Gatzoulis MA, Adatia I, Celermajer D, Denton C, Ghofrani A et al (2013) Updated clinical classification of pulmonary hypertension. J Am Coll Cardiol 62(25 Suppl): D34–D41. https://doi.org/10.1016/j.jacc.2013.10.029
33. Zahid KR, Raza U, Chen J, Raj UJ, Gou D (2020) Pathobiology of pulmonary artery hypertension: role of long non-coding RNAs. Cardiovasc Res 116(12):1937–1947. https://doi.org/10.1093/cvr/cvaa050
34. Dejana E, Hirschi KK, Simons M (2017) The molecular basis of endothelial cell plasticity. Nat Commun 8:14361. https://doi.org/10.1038/ncomms14361
35. Stenmark KR, Frid MG, Graham BB, Tuder RM (2018) Dynamic and diverse changes in the functional properties of vascular smooth muscle cells in pulmonary hypertension. Cardiovasc Res 114(4):551–564. https://doi.org/10.1093/cvr/cvy004
36. Bisserier M, Janostiak R, Lezoualc'h F, Hadri L (2020) Targeting epigenetic mechanisms as an emerging therapeutic strategy in pulmonary hypertension disease. Vasc Biol 2(1):R17–R34. https://doi.org/10.1530/vb-19-0030
37. Kornberg RD (1974) Chromatin structure: a repeating unit of histones and DNA. Science 184(4139):868–871. https://doi.org/10.1126/science.184.4139.868
38. Luger K, Mäder AW, Richmond RK, Sargent DF, Richmond TJ (1997) Crystal structure of the nucleosome core particle at 2.8 a resolution. Nature 389(6648):251–260. https://doi.org/10.1038/38444
39. Allis CD, Jenuwein T (2016) The molecular hallmarks of epigenetic control. Nat Rev Genet 17(8):487–500. https://doi.org/10.1038/nrg.2016.59
40. Bird A, Taggart M, Frommer M, Miller OJ, Macleod D (1985) A fraction of the mouse genome that is derived from islands of nonmethylated. CpG-rich DNA Cell 40(1):91–99. https://doi.org/10.1016/0092-8674(85)90312-5
41. Antequera F, Boyes J, Bird A (1990) High levels of de novo methylation and altered chromatin structure at CpG islands in cell lines. Cell 62(3):503–514. https://doi.org/10.1016/0092-8674(90)90015-7
42. Watt F, Molloy PL (1988) Cytosine methylation prevents binding to DNA of a HeLa cell transcription factor required for optimal expression of the adenovirus major late promoter. Genes Dev 2(9):1136–1143. https://doi.org/10.1101/gad.2.9.1136
43. Bestor TH (2000) The DNA methyltransferases of mammals. Hum Mol Genet 9(16): 2395–2402. https://doi.org/10.1093/hmg/9.16.2395
44. Yen RW, Vertino PM, Nelkin BD, Yu JJ, el-Deiry W, Cumaraswamy A et al (1992) Isolation and characterization of the cDNA encoding human DNA methyltransferase. Nucleic Acids Res 20(9):2287–2291. https://doi.org/10.1093/nar/20.9.2287

45. Hermann A, Goyal R, Jeltsch A (2004) The Dnmt1 DNA-(cytosine-C5)-methyltransferase methylates DNA processively with high preference for hemimethylated target sites. J Biol Chem 279(46):48350–48359. https://doi.org/10.1074/jbc.M403427200
46. Bostick M, Kim JK, Estève PO, Clark A, Pradhan S, Jacobsen SE (2007) UHRF1 plays a role in maintaining DNA methylation in mammalian cells. Science 317(5845):1760–1764. https://doi.org/10.1126/science.1147939
47. Okano M, Bell DW, Haber DA, Li E (1999) DNA methyltransferases Dnmt3a and Dnmt3b are essential for de novo methylation and mammalian development. Cell 99(3):247–257. https://doi.org/10.1016/s0092-8674(00)81656-6
48. Richard Albert J, Au Yeung WK, Toriyama K, Kobayashi H, Hirasawa R, Brind'Amour J et al (2020) Maternal DNMT3A-dependent de novo methylation of the paternal genome inhibits gene expression in the early embryo. Nat Commun 11(1):5417. https://doi.org/10.1038/s41467-020-19279-7
49. Yagi M, Kabata M, Tanaka A, Ukai T, Ohta S, Nakabayashi K et al (2020) Identification of distinct loci for de novo DNA methylation by DNMT3A and DNMT3B during mammalian development. Nat Commun 11(1):3199. https://doi.org/10.1038/s41467-020-16989-w
50. Bhutani N, Burns DM, Blau HM (2011) DNA demethylation dynamics. Cell 146(6):866–872. https://doi.org/10.1016/j.cell.2011.08.042
51. Tahiliani M, Koh KP, Shen Y, Pastor WA, Bandukwala H, Brudno Y et al (2009) Conversion of 5-methylcytosine to 5-hydroxymethylcytosine in mammalian DNA by MLL partner TET1. Science 324(5929):930–935. https://doi.org/10.1126/science.1170116
52. He YF, Li BZ, Li Z, Liu P, Wang Y, Tang Q et al (2011) Tet-mediated formation of 5-carboxylcytosine and its excision by TDG in mammalian DNA. Science 333(6047):1303–1307. https://doi.org/10.1126/science.1210944
53. Ito S, Shen L, Dai Q, Wu SC, Collins LB, Swenberg JA et al (2011) Tet proteins can convert 5-methylcytosine to 5-formylcytosine and 5-carboxylcytosine. Science 333(6047):1300–1303. https://doi.org/10.1126/science.1210597
54. Maiti A, Drohat AC (2011) Thymine DNA glycosylase can rapidly excise 5-formylcytosine and 5-carboxylcytosine: potential implications for active demethylation of CpG sites. J Biol Chem 286(41):35334–35338. https://doi.org/10.1074/jbc.C111.284620
55. Hahn MA, Szabó PE, Pfeifer GP (2014) 5-Hydroxymethylcytosine: a stable or transient DNA modification? Genomics 104(5):314–323. https://doi.org/10.1016/j.ygeno.2014.08.015
56. Hon GC, Song CX, Du T, Jin F, Selvaraj S, Lee AY et al (2014) 5mC oxidation by Tet2 modulates enhancer activity and timing of transcriptome reprogramming during differentiation. Mol Cell 56(2):286–297. https://doi.org/10.1016/j.molcel.2014.08.026
57. Pastor WA, Pape UJ, Huang Y, Henderson HR, Lister R, Ko M et al (2011) Genome-wide mapping of 5-hydroxymethylcytosine in embryonic stem cells. Nature 473(7347):394–397. https://doi.org/10.1038/nature10102
58. Jair KW, Bachman KE, Suzuki H, Ting AH, Rhee I, Yen RW et al (2006) De novo CpG island methylation in human cancer cells. Cancer Res 66(2):682–692. https://doi.org/10.1158/0008-5472.CAN-05-1980
59. Greißel A, Culmes M, Napieralski R, Wagner E, Gebhard H, Schmitt M et al (2015) Alternation of histone and DNA methylation in human atherosclerotic carotid plaques. Thromb Haemost 114(2):390–402. https://doi.org/10.1160/TH14-10-0852
60. Aavik E, Lumivuori H, Leppänen O, Wirth T, Häkkinen SK, Bräsen JH et al (2015) Global DNA methylation analysis of human atherosclerotic plaques reveals extensive genomic hypomethylation and reactivation at imprinted locus 14q32 involving induction of a miRNA cluster. Eur Heart J 36(16):993–1000. https://doi.org/10.1093/eurheartj/ehu437
61. Valencia-Morales MP, Zaina S, Heyn H, Carmona FJ, Varol N, Sayols S et al (2015) The DNA methylation drift of the atherosclerotic aorta increases with lesion progression. BMC Med Genet 8:7. https://doi.org/10.1186/s12920-015-0085-1

62. Zaina S, Heyn H, Carmona FJ, Varol N, Sayols S, Condom E et al (2014) DNA methylation map of human atherosclerosis. Circ Cardiovasc Genet 7(5):692–700. https://doi.org/10.1161/CIRCGENETICS.113.000441
63. Singh U, Jialal I (2006) Oxidative stress and atherosclerosis. Pathophysiology 13(3):129–142. https://doi.org/10.1016/j.pathophys.2006.05.002
64. Kim YR, Kim CS, Naqvi A, Kumar A, Kumar S, Hoffman TA et al (2012) Epigenetic upregulation of p66shc mediates low-density lipoprotein cholesterol-induced endothelial cell dysfunction. Am J Physiol Heart Circ Physiol 303(2):H189–H196. https://doi.org/10.1152/ajpheart.01218.2011
65. Costantino S, Paneni F, Mitchell K, Mohammed SA, Hussain S, Gkolfos C et al (2018) Hyperglycaemia-induced epigenetic changes drive persistent cardiac dysfunction via the adaptor p66. Int J Cardiol 268:179–186. https://doi.org/10.1016/j.ijcard.2018.04.082
66. Kumar A, Kumar S, Vikram A, Hoffman TA, Naqvi A, Lewarchik CM et al (2013) Histone and DNA methylation-mediated epigenetic downregulation of endothelial Kruppel-like factor 2 by low-density lipoprotein cholesterol. Arterioscler Thromb Vasc Biol 33(8):1936–1942. https://doi.org/10.1161/ATVBAHA.113.301765
67. Qi W, Li Q, Liew CW, Rask-Madsen C, Lockhart SM, Rasmussen LM et al (2017) SHP-1 activation inhibits vascular smooth muscle cell proliferation and intimal hyperplasia in a rodent model of insulin resistance and diabetes. Diabetologia 60(3):585–596. https://doi.org/10.1007/s00125-016-4159-1
68. Kobayashi S, Inoue N, Azumi H, Seno T, Hirata K, Kawashima S et al (2002) Expressional changes of the vascular antioxidant system in atherosclerotic coronary arteries. J Atheroscler Thromb 9(4):184–190. https://doi.org/10.5551/jat.9.184
69. Weitzman SA, Turk PW, Milkowski DH, Kozlowski K (1994) Free radical adducts induce alterations in DNA cytosine methylation. Proc Natl Acad Sci U S A 91(4):1261–1264. https://doi.org/10.1073/pnas.91.4.1261
70. Baccarelli A, Bollati V (2009) Epigenetics and environmental chemicals. Curr Opin Pediatr 21(2):243–251. https://doi.org/10.1097/mop.0b013e32832925cc
71. Elia L, Kunderfranco P, Carullo P, Vacchiano M, Farina FM, Hall IF et al (2018) UHRF1 epigenetically orchestrates smooth muscle cell plasticity in arterial disease. J Clin Invest 128(6):2473–2486. https://doi.org/10.1172/JCI96121
72. Thenappan T, Ormiston ML, Ryan JJ, Archer SL (2018) Pulmonary arterial hypertension: pathogenesis and clinical management. BMJ 360:j5492. https://doi.org/10.1136/bmj.j5492
73. Chelladurai P, Seeger W, Pullamsetti SS (2016) Epigenetic mechanisms in pulmonary arterial hypertension: the need for global perspectives. Eur Respir Rev 25(140):135–140. https://doi.org/10.1183/16000617.0036-2016
74. Bonnet S, Michelakis ED, Porter CJ, Andrade-Navarro MA, Thébaud B, Haromy A et al (2006) An abnormal mitochondrial-hypoxia inducible factor-1alpha-Kv channel pathway disrupts oxygen sensing and triggers pulmonary arterial hypertension in fawn hooded rats: similarities to human pulmonary arterial hypertension. Circulation 113(22):2630–2641. https://doi.org/10.1161/CIRCULATIONAHA.105.609008
75. Archer SL, Marsboom G, Kim GH, Zhang HJ, Toth PT, Svensson EC et al (2010) Epigenetic attenuation of mitochondrial superoxide dismutase 2 in pulmonary arterial hypertension: a basis for excessive cell proliferation and a new therapeutic target. Circulation 121(24):2661–2671. https://doi.org/10.1161/CIRCULATIONAHA.109.916098
76. Hautefort A, Chesné J, Preussner J, Pullamsetti SS, Tost J, Looso M et al (2017) Pulmonary endothelial cell DNA methylation signature in pulmonary arterial hypertension. Oncotarget 8(32):52995–53016. https://doi.org/10.18632/oncotarget.18031
77. Yang Q, Lu Z, Ramchandran R, Longo LD, Raj JU (2012) Pulmonary artery smooth muscle cell proliferation and migration in fetal lambs acclimatized to high-altitude long-term hypoxia: role of histone acetylation. Am J Physiol Lung Cell Mol Physiol 303(11):L1001–L1010. https://doi.org/10.1152/ajplung.00092.2012

78. Perros F, Cohen-Kaminsky S, Gambaryan N, Girerd B, Raymond N, Klingelschmitt I et al (2013) Cytotoxic cells and granulysin in pulmonary arterial hypertension and pulmonary veno-occlusive disease. Am J Respir Crit Care Med 187(2):189–196. https://doi.org/10.1164/rccm.201208-1364OC
79. Finch JT, Klug A (1976) Solenoidal model for superstructure in chromatin. Proc Natl Acad Sci U S A 73(6):1897–1901. https://doi.org/10.1073/pnas.73.6.1897
80. Hergeth SP, Schneider R (2015) The H1 linker histones: multifunctional proteins beyond the nucleosomal core particle. EMBO Rep 16(11):1439–1453. https://doi.org/10.15252/embr.201540749
81. Passarge E (1979) Emil Heitz and the concept of heterochromatin: longitudinal chromosome differentiation was recognized fifty years ago. Am J Hum Genet 31(2):106–115
82. Kouzarides T (2007) Chromatin modifications and their function. Cell 128(4):693–705. https://doi.org/10.1016/j.cell.2007.02.005
83. Strahl BD, Allis CD (2000) The language of covalent histone modifications. Nature 403(6765):41–45. https://doi.org/10.1038/47412
84. Martin C, Zhang Y (2005) The diverse functions of histone lysine methylation. Nat Rev Mol Cell Biol 6(11):838–849. https://doi.org/10.1038/nrm1761
85. Tan M, Luo H, Lee S, Jin F, Yang JS, Montellier E et al (2011) Identification of 67 histone marks and histone lysine crotonylation as a new type of histone modification. Cell 146(6):1016–1028. https://doi.org/10.1016/j.cell.2011.08.008
86. Cuthbert GL, Daujat S, Snowden AW, Erdjument-Bromage H, Hagiwara T, Yamada M et al (2004) Histone deimination antagonizes arginine methylation. Cell 118(5):545–553. https://doi.org/10.1016/j.cell.2004.08.020
87. Eckhart W, Hutchinson MA, Hunter T (1979) An activity phosphorylating tyrosine in polyoma T antigen immunoprecipitates. Cell 18(4):925–933. https://doi.org/10.1016/0092-8674(79)90205-8
88. Sakabe K, Wang Z, Hart GW (2010) Beta-N-acetylglucosamine (O-GlcNAc) is part of the histone code. Proc Natl Acad Sci U S A 107(46):19915–19920. https://doi.org/10.1073/pnas.1009023107
89. Candau R, Zhou JX, Allis CD, Berger SL (1997) Histone acetyltransferase activity and interaction with ADA2 are critical for GCN5 function in vivo. EMBO J 16(3):555–565. https://doi.org/10.1093/emboj/16.3.555
90. Brownell JE, Zhou J, Ranalli T, Kobayashi R, Edmondson DG, Roth SY et al (1996) Tetrahymena histone acetyltransferase a: a homolog to yeast Gcn5p linking histone acetylation to gene activation. Cell 84(6):843–851. https://doi.org/10.1016/s0092-8674(00)81063-6
91. Rosa-Garrido M, Chapski DJ, Vondriska TM (2018) Epigenomes in cardiovascular disease. Circ Res 122(11):1586–1607. https://doi.org/10.1161/CIRCRESAHA.118.311597
92. Mu S, Shimosawa T, Ogura S, Wang H, Uetake Y, Kawakami-Mori F et al (2011) Epigenetic modulation of the renal β-adrenergic-WNK4 pathway in salt-sensitive hypertension. Nat Med 17(5):573–580. https://doi.org/10.1038/nm.2337
93. Mumby S, Gambaryan N, Meng C, Perros F, Humbert M, Wort SJ et al (2017) Bromodomain and extra-terminal protein mimic JQ1 decreases inflammation in human vascular endothelial cells: implications for pulmonary arterial hypertension. Respirology 22(1):157–164. https://doi.org/10.1111/resp.12872
94. Cho YK, Eom GH, Kee HJ, Kim HS, Choi WY, Nam KI et al (2010) Sodium valproate, a histone deacetylase inhibitor, but not captopril, prevents right ventricular hypertrophy in rats. Circ J 74(4):760–770. https://doi.org/10.1253/circj.cj-09-0580
95. Yoon S, Eom GH (2016) HDAC and HDAC inhibitor: from cancer to cardiovascular diseases. Chonnam Med J 52(1):1–11. https://doi.org/10.4068/cmj.2016.52.1.1
96. Elia L, Condorelli G (2019) The involvement of epigenetics in vascular disease development. Int J Biochem Cell Biol 107:27–31. https://doi.org/10.1016/j.biocel.2018.12.005

97. Shi X, Hong T, Walter KL, Ewalt M, Michishita E, Hung T et al (2006) ING2 PHD domain links histone H3 lysine 4 methylation to active gene repression. Nature 442(7098):96–99. https://doi.org/10.1038/nature04835
98. Costantino S, Paneni F, Cosentino F (2015) Targeting chromatin remodeling to prevent cardiovascular disease in diabetes. Curr Pharm Biotechnol 16(6):531–543. https://doi.org/10.2174/1389201016061504071136 44
99. Shi Y, Lan F, Matson C, Mulligan P, Whetstine JR, Cole PA et al (2004) Histone demethylation mediated by the nuclear amine oxidase homolog LSD1. Cell 119(7):941–953. https://doi.org/10.1016/j.cell.2004.12.012
100. Tsukada Y, Fang J, Erdjument-Bromage H, Warren ME, Borchers CH, Tempst P et al (2006) Histone demethylation by a family of JmjC domain-containing proteins. Nature 439(7078):811–816. https://doi.org/10.1038/nature04433
101. Jenuwein T, Allis CD (2001) Translating the histone code. Science 293(5532):1074–1080. https://doi.org/10.1126/science.1063127
102. Rossetto D, Avvakumov N, Côté J (2012) Histone phosphorylation: a chromatin modification involved in diverse nuclear events. Epigenetics 7(10):1098–1108. https://doi.org/10.4161/epi.21975
103. Sawicka A, Seiser C (2014) Sensing core histone phosphorylation—a matter of perfect timing. Biochim Biophys Acta 1839(8):711–718. https://doi.org/10.1016/j.bbagrm.2014.04.013
104. Wierda RJ, Rietveld IM, van Eggermond MC, Belien JA, van Zwet EW, Lindeman JH et al (2015) Global histone H3 lysine 27 triple methylation levels are reduced in vessels with advanced atherosclerotic plaques. Life Sci 129:3–9. https://doi.org/10.1016/j.lfs.2014.10.010
105. Barroso M, Kao D, Blom HJ, Tavares de Almeida I, Castro R, Loscalzo J et al (2016) S-adenosylhomocysteine induces inflammation through NFkB: a possible role for EZH2 in endothelial cell activation. Biochim Biophys Acta 1862(1):82–92. https://doi.org/10.1016/j.bbadis.2015.10.019
106. Cong G, Yan R, Huang H, Wang K, Yan N, Jin P et al (2017) Involvement of histone methylation in macrophage apoptosis and unstable plaque formation in methionine-induced hyperhomocysteinemic ApoE. Life Sci 173:135–144. https://doi.org/10.1016/j.lfs.2017.02.003
107. Greißel A, Culmes M, Burgkart R, Zimmermann A, Eckstein HH, Zernecke A et al (2016) Histone acetylation and methylation significantly change with severity of atherosclerosis in human carotid plaques. Cardiovasc Pathol 25(2):79–86. https://doi.org/10.1016/j.carpath.2015.11.001
108. Manabe I, Owens GK (2001) Recruitment of serum response factor and hyperacetylation of histones at smooth muscle-specific regulatory regions during differentiation of a novel P19-derived in vitro smooth muscle differentiation system. Circ Res 88(11):1127–1134. https://doi.org/10.1161/hh1101.091339
109. McDonald OG, Wamhoff BR, Hoofnagle MH, Owens GK (2006) Control of SRF binding to CArG box chromatin regulates smooth muscle gene expression in vivo. J Clin Invest 116(1):36–48. https://doi.org/10.1172/JCI26505
110. Hoeksema MA, Gijbels MJ, Van den Bossche J, van der Velden S, Sijm A, Neele AE et al (2014) Targeting macrophage histone deacetylase 3 stabilizes atherosclerotic lesions. EMBO Mol Med 6(9):1124–1132. https://doi.org/10.15252/emmm.201404170
111. Bode KA, Schroder K, Hume DA, Ravasi T, Heeg K, Sweet MJ et al (2007) Histone deacetylase inhibitors decrease toll-like receptor-mediated activation of proinflammatory gene expression by impairing transcription factor recruitment. Immunology 122(4):596–606. https://doi.org/10.1111/j.1365-2567.2007.02678.x
112. Ooi SK, Qiu C, Bernstein E, Li K, Jia D, Yang Z et al (2007) DNMT3L connects unmethylated lysine 4 of histone H3 to de novo methylation of DNA. Nature 448(7154):714–717. https://doi.org/10.1038/nature05987
113. Lv YC, Tang YY, Zhang P, Wan W, Yao F, He PP et al (2016) Histone methyltransferase enhancer of Zeste homolog 2-mediated ABCA1 promoter DNA methylation contributes to the

progression of atherosclerosis. PLoS One 11(6):e0157265. https://doi.org/10.1371/journal.pone.0157265
114. Wang Y, Huang XX, Leng D, Li JF, Liang Y, Jiang T (2021) Effect of EZH2 on pulmonary artery smooth muscle cell migration in pulmonary hypertension. Mol Med Rep 23(2):1. https://doi.org/10.3892/mmr.2020.11768
115. Chen D, Yang Y, Cheng X, Fang F, Xu G, Yuan Z et al (2015) Megakaryocytic leukemia 1 directs a histone H3 lysine 4 methyltransferase complex to regulate hypoxic pulmonary hypertension. Hypertension 65(4):821–833. https://doi.org/10.1161/HYPERTENSIONAHA.114.04585
116. Zhou XL, Liu ZB, Zhu RR, Huang H, Xu QR, Xu H et al (2019) NSD2 silencing alleviates pulmonary arterial hypertension by inhibiting trehalose metabolism and autophagy. Clin Sci (Lond) 133(9):1085–1096. https://doi.org/10.1042/CS20190142
117. Chelladurai P, Dabral S, Basineni SR, Chen CN, Schmoranzer M, Bender N et al (2020) Isoform-specific characterization of class I histone deacetylases and their therapeutic modulation in pulmonary hypertension. Sci Rep 10(1):12864. https://doi.org/10.1038/s41598-020-69737-x
118. Xu XF, Lv Y, Gu WZ, Tang LL, Wei JK, Zhang LY et al (2013) Epigenetics of hypoxic pulmonary arterial hypertension following intrauterine growth retardation rat: epigenetics in PAH following IUGR. Respir Res 14:20. https://doi.org/10.1186/1465-9921-14-20
119. Cardinale JP, Sriramula S, Pariaut R, Guggilam A, Mariappan N, Elks CM et al (2010) HDAC inhibition attenuates inflammatory, hypertrophic, and hypertensive responses in spontaneously hypertensive rats. Hypertension 56(3):437–444. https://doi.org/10.1161/HYPERTENSIONAHA.110.154567
120. Lee HA, Lee DY, Cho HM, Kim SY, Iwasaki Y, Kim IK (2013) Histone deacetylase inhibition attenuates transcriptional activity of mineralocorticoid receptor through its acetylation and prevents development of hypertension. Circ Res 112(7):1004–1012. https://doi.org/10.1161/CIRCRESAHA.113.301071
121. Usui T, Okada M, Mizuno W, Oda M, Ide N, Morita T et al (2012) HDAC4 mediates development of hypertension via vascular inflammation in spontaneous hypertensive rats. Am J Physiol Heart Circ Physiol 302(9):H1894–H1904. https://doi.org/10.1152/ajpheart.01039.2011
122. Kim J, Hwangbo C, Hu X, Kang Y, Papangeli I, Mehrotra D et al (2015) Restoration of impaired endothelial myocyte enhancer factor 2 function rescues pulmonary arterial hypertension. Circulation 131(2):190–199. https://doi.org/10.1161/CIRCULATIONAHA.114.013339
123. Nozik-Grayck E, Woods C, Stearman RS, Venkataraman S, Ferguson BS, Swain K et al (2016) Histone deacetylation contributes to low extracellular superoxide dismutase expression in human idiopathic pulmonary arterial hypertension. Am J Physiol Lung Cell Mol Physiol 311(1):L124–L134. https://doi.org/10.1152/ajplung.00263.2015
124. Xu X, Ha CH, Wong C, Wang W, Hausser A, Pfizenmaier K et al (2007) Angiotensin II stimulates protein kinase D-dependent histone deacetylase 5 phosphorylation and nuclear export leading to vascular smooth muscle cell hypertrophy. Arterioscler Thromb Vasc Biol 27(11):2355–2362. https://doi.org/10.1161/ATVBAHA.107.151704
125. Fiedler J, Baker AH, Dimmeler S, Heymans S, Mayr M, Thum T (2018) Non-coding RNAs in vascular disease—from basic science to clinical applications: scientific update from the working Group of Myocardial Function of the European Society of Cardiology. Cardiovasc Res 114(10):1281–1286. https://doi.org/10.1093/cvr/cvy121
126. Wapinski O, Chang HY (2011) Long noncoding RNAs and human disease. Trends Cell Biol 21(6):354–361. https://doi.org/10.1016/j.tcb.2011.04.001
127. Poller W, Dimmeler S, Heymans S, Zeller T, Haas J, Karakas M et al (2018) Non-coding RNAs in cardiovascular diseases: diagnostic and therapeutic perspectives. Eur Heart J 39(29):2704–2716. https://doi.org/10.1093/eurheartj/ehx165

128. Pagiatakis C, Hall IF, Condorelli G (2020) Long non-coding RNA H19: a new avenue for RNA therapeutics in cardiac hypertrophy? Eur Heart J 41(36):3475–3476. https://doi.org/10.1093/eurheartj/ehaa663
129. Bartel DP (2009) MicroRNAs: target recognition and regulatory functions. Cell 136(2): 215–233. https://doi.org/10.1016/j.cell.2009.01.002
130. Mercer TR, Dinger ME, Mattick JS (2009) Long non-coding RNAs: insights into functions. Nat Rev Genet 10(3):155–159. https://doi.org/10.1038/nrg2521
131. Quinn JJ, Chang HY (2016) Unique features of long non-coding RNA biogenesis and function. Nat Rev Genet 17(1):47–62. https://doi.org/10.1038/nrg.2015.10
132. Jeck WR, Sorrentino JA, Wang K, Slevin MK, Burd CE, Liu J et al (2013) Circular RNAs are abundant, conserved, and associated with ALU repeats. RNA 19(2):141–157. https://doi.org/10.1261/rna.035667.112
133. Memczak S, Jens M, Elefsinioti A, Torti F, Krueger J, Rybak A et al (2013) Circular RNAs are a large class of animal RNAs with regulatory potency. Nature 495(7441):333–338. https://doi.org/10.1038/nature11928
134. Barrett SP, Salzman J (2016) Circular RNAs: analysis, expression and potential functions. Development 143(11):1838–1847. https://doi.org/10.1242/dev.128074
135. Elia L, Condorelli G (2015) RNA (epi)genetics in cardiovascular diseases. J Mol Cell Cardiol 89(Pt A):11–16. https://doi.org/10.1016/j.yjmcc.2015.07.012
136. Beerman I, Rossi DJ (2015) Epigenetic control of stem cell potential during homeostasis, aging, and disease. Cell Stem Cell 16(6):613–625. https://doi.org/10.1016/j.stem.2015.05.009
137. Elia L, Quintavalle M (2017) Epigenetics and vascular diseases: influence of non-coding RNAs and their clinical implications. Front Cardiovasc Med 4:26. https://doi.org/10.3389/fcvm.2017.00026
138. Uchida S, Dimmeler S (2015) Long noncoding RNAs in cardiovascular diseases. Circ Res 116(4):737–750. https://doi.org/10.1161/CIRCRESAHA.116.302521
139. Bartel DP (2004) MicroRNAs: genomics, biogenesis, mechanism, and function. Cell 116(2): 281–297
140. Lopez-Pedrera C, Barbarroja N, Patiño-Trives AM, Luque-Tévar M, Torres-Granados C, Aguirre-Zamorano MA et al (2020) Role of microRNAs in the development of cardiovascular disease in systemic autoimmune disorders. Int J Mol Sci 21(6). https://doi.org/10.3390/ijms21062012
141. O'Brien J, Hayder H, Zayed Y, Peng C (2018) Overview of MicroRNA biogenesis, mechanisms of actions, and circulation. Front Endocrinol (Lausanne) 9:402. https://doi.org/10.3389/fendo.2018.00402
142. Ghini F, Rubolino C, Climent M, Simeone I, Marzi MJ, Nicassio F (2018) Endogenous transcripts control miRNA levels and activity in mammalian cells by target-directed miRNA degradation. Nat Commun 9(1):3119. https://doi.org/10.1038/s41467-018-05182-9
143. Shukla GC, Singh J, Barik S (2011) MicroRNAs: processing, maturation, target recognition and regulatory functions. Mol Cell Pharmacol 3(3):83–92
144. Statello L, Guo CJ, Chen LL, Huarte M (2021) Gene regulation by long non-coding RNAs and its biological functions. Nat Rev Mol Cell Biol 22(2):96–118. https://doi.org/10.1038/s41580-020-00315-9
145. Erdmann VA, Szymanski M, Hochberg A, de Groot N, Barciszewski J (1999) Collection of mRNA-like non-coding RNAs. Nucleic Acids Res 27(1):192–195. https://doi.org/10.1093/nar/27.1.192
146. Kapranov P, Cheng J, Dike S, Nix DA, Duttagupta R, Willingham AT et al (2007) RNA maps reveal new RNA classes and a possible function for pervasive transcription. Science 316(5830):1484–1488. https://doi.org/10.1126/science.1138341
147. Salamon I, Saccani Jotti G, Condorelli G (2018) The long noncoding RNA landscape in cardiovascular disease: a brief update. Curr Opin Cardiol 33(3):282–289. https://doi.org/10.1097/HCO.0000000000000507

148. Yuan Y, Xu L, Geng Z, Liu J, Zhang L, Wu Y et al (2021) The role of non-coding RNA network in atherosclerosis. Life Sci 265:118756. https://doi.org/10.1016/j.lfs.2020.118756
149. Pang KC, Frith MC, Mattick JS (2006) Rapid evolution of noncoding RNAs: lack of conservation does not mean lack of function. Trends Genet 22(1):1–5. https://doi.org/10.1016/j.tig.2005.10.003
150. Leeper NJ, Maegdefessel L (2018) Non-coding RNAs: key regulators of smooth muscle cell fate in vascular disease. Cardiovasc Res 114(4):611–621. https://doi.org/10.1093/cvr/cvx249
151. Cabili MN, Trapnell C, Goff L, Koziol M, Tazon-Vega B, Regev A et al (2011) Integrative annotation of human large intergenic noncoding RNAs reveals global properties and specific subclasses. Genes Dev 25(18):1915–1927. https://doi.org/10.1101/gad.17446611
152. Boon RA, Jaé N, Holdt L, Dimmeler S (2016) Long noncoding RNAs: from clinical genetics to therapeutic targets? J Am Coll Cardiol 67(10):1214–1226. https://doi.org/10.1016/j.jacc.2015.12.051
153. Guttman M, Rinn JL (2012) Modular regulatory principles of large non-coding RNAs. Nature 482(7385):339–346. https://doi.org/10.1038/nature10887
154. Pagiatakis C, Musolino E, Gornati R, Bernardini G, Papait R (2019) Epigenetics of aging and disease: a brief overview. Aging Clin Exp Res. https://doi.org/10.1007/s40520-019-01430-0
155. Soler-Botija C, Gálvez-Montón C, Bayés-Genís A (2019) Epigenetic biomarkers in cardiovascular diseases. Front Genet 10:950. https://doi.org/10.3389/fgene.2019.00950
156. Li X, Yang L, Chen LL (2018) The biogenesis, functions, and challenges of circular RNAs. Mol Cell 71(3):428–442. https://doi.org/10.1016/j.molcel.2018.06.034
157. Lim TB, Lavenniah A, Foo RS (2020) Circles in the heart and cardiovascular system. Cardiovasc Res 116(2):269–278. https://doi.org/10.1093/cvr/cvz227
158. Aranda JF, Madrigal-Matute J, Rotllan N, Fernández-Hernando C (2013) MicroRNA modulation of lipid metabolism and oxidative stress in cardiometabolic diseases. Free Radic Biol Med 64:31–39. https://doi.org/10.1016/j.freeradbiomed.2013.07.014
159. Romaine SP, Tomaszewski M, Condorelli G, Samani NJ (2015) MicroRNAs in cardiovascular disease: an introduction for clinicians. Heart 101(12):921–928. https://doi.org/10.1136/heartjnl-2013-305402
160. Nishiguchi T, Imanishi T, Akasaka T (2015) MicroRNAs and cardiovascular diseases. Biomed Res Int 2015:682857. https://doi.org/10.1155/2015/682857
161. Collura S, Morsiani C, Vacirca A, Fronterrè S, Ciavarella C, Vasuri F et al (2020) The carotid plaque as paradigmatic case of site-specific acceleration of aging process: the microRNAs and the inflammaging contribution. Ageing Res Rev 61:101090. https://doi.org/10.1016/j.arr.2020.101090
162. Madrigal-Matute J, Rotllan N, Aranda JF, Fernández-Hernando C (2013) MicroRNAs and atherosclerosis. Curr Atheroscler Rep 15(5):322. https://doi.org/10.1007/s11883-013-0322-z
163. Li C, Li S, Zhang F, Wu M, Liang H, Song J et al (2018) Endothelial microparticles-mediated transfer of microRNA-19b promotes atherosclerosis via activating perivascular adipose tissue inflammation in apoE. Biochem Biophys Res Commun 495(2):1922–1929. https://doi.org/10.1016/j.bbrc.2017.11.195
164. Jamaluddin MS, Weakley SM, Zhang L, Kougias P, Lin PH, Yao Q et al (2011) miRNAs: roles and clinical applications in vascular disease. Expert Rev Mol Diagn 11(1):79–89. https://doi.org/10.1586/erm.10.103
165. Lu Y, Thavarajah T, Gu W, Cai J, Xu Q (2018) Impact of miRNA in atherosclerosis. Arterioscler Thromb Vasc Biol 38(9):e159–ee70. https://doi.org/10.1161/ATVBAHA.118.310227
166. Farina FM, Hall IF, Serio S, Zani S, Climent M, Salvarani N et al (2020) miR-128-3p is a novel regulator of vascular smooth muscle cell phenotypic switch and vascular diseases. Circ Res 126(12):e120–ee35. https://doi.org/10.1161/CIRCRESAHA.120.316489
167. Holdt LM, Hoffmann S, Sass K, Langenberger D, Scholz M, Krohn K et al (2013) Alu elements in ANRIL non-coding RNA at chromosome 9p21 modulate atherogenic cell

functions through trans-regulation of gene networks. PLoS Genet 9(7):e1003588. https://doi.org/10.1371/journal.pgen.1003588
168. Kotake Y, Nakagawa T, Kitagawa K, Suzuki S, Liu N, Kitagawa M et al (2011) Long non-coding RNA ANRIL is required for the PRC2 recruitment to and silencing of p15 (INK4B) tumor suppressor gene. Oncogene 30(16):1956–1962. https://doi.org/10.1038/onc.2010.568
169. Holdt LM, Beutner F, Scholz M, Gielen S, Gäbel G, Bergert H et al (2010) ANRIL expression is associated with atherosclerosis risk at chromosome 9p21. Arterioscler Thromb Vasc Biol 30(3):620–627. https://doi.org/10.1161/ATVBAHA.109.196832
170. Congrains A, Kamide K, Oguro R, Yasuda O, Miyata K, Yamamoto E et al (2012) Genetic variants at the 9p21 locus contribute to atherosclerosis through modulation of ANRIL and CDKN2A/B. Atherosclerosis 220(2):449–455. https://doi.org/10.1016/j.atherosclerosis.2011.11.017
171. Kurian L, Aguirre A, Sancho-Martinez I, Benner C, Hishida T, Nguyen TB et al (2015) Identification of novel long noncoding RNAs underlying vertebrate cardiovascular development. Circulation 131(14):1278–1290. https://doi.org/10.1161/CIRCULATIONAHA.114.013303
172. Miao Y, Ajami NE, Huang TS, Lin FM, Lou CH, Wang YT et al (2018) Enhancer-associated long non-coding RNA LEENE regulates endothelial nitric oxide synthase and endothelial function. Nat Commun 9(1):292. https://doi.org/10.1038/s41467-017-02113-y
173. Robb GB, Carson AR, Tai SC, Fish JE, Singh S, Yamada T et al (2004) Post-transcriptional regulation of endothelial nitric-oxide synthase by an overlapping antisense mRNA transcript. J Biol Chem 279(36):37982–37996. https://doi.org/10.1074/jbc.M400271200
174. Fiedler J, Breckwoldt K, Remmele CW, Hartmann D, Dittrich M, Pfanne A et al (2015) Development of Long noncoding RNA-based strategies to modulate tissue vascularization. J Am Coll Cardiol 66(18):2005–2015. https://doi.org/10.1016/j.jacc.2015.07.081
175. Boeckel JN, Jaé N, Heumüller AW, Chen W, Boon RA, Stellos K et al (2015) Identification and characterization of hypoxia-regulated endothelial circular RNA. Circ Res 117(10):884–890. https://doi.org/10.1161/CIRCRESAHA.115.306319
176. Michalik KM, You X, Manavski Y, Doddaballapur A, Zörnig M, Braun T et al (2014) Long noncoding RNA MALAT1 regulates endothelial cell function and vessel growth. Circ Res 114(9):1389–1397. https://doi.org/10.1161/CIRCRESAHA.114.303265
177. Cremer S, Michalik KM, Fischer A, Pfisterer L, Jaé N, Winter C et al (2019) Hematopoietic deficiency of the Long noncoding RNA MALAT1 promotes atherosclerosis and plaque inflammation. Circulation 139(10):1320–1334. https://doi.org/10.1161/CIRCULATIONAHA.117.029015
178. Bell RD, Long X, Lin M, Bergmann JH, Nanda V, Cowan SL et al (2014) Identification and initial functional characterization of a human vascular cell-enriched long noncoding RNA. Arterioscler Thromb Vasc Biol 34(6):1249–1259. https://doi.org/10.1161/ATVBAHA.114.303240
179. Lyu Q, Xu S, Lyu Y, Choi M, Christie CK, Slivano OJ et al (2019) Stabilizes vascular endothelial cell adherens junctions through interaction with CKAP4. Proc Natl Acad Sci U S A 116(2):546–555. https://doi.org/10.1073/pnas.1810729116
180. Boulberdaa M, Scott E, Ballantyne M, Garcia R, Descamps B, Angelini GD et al (2016) A role for the Long noncoding RNA SENCR in commitment and function of endothelial cells. Mol Ther 24(5):978–990. https://doi.org/10.1038/mt.2016.41
181. Hall IF, Climent M, Quintavalle M, Farina FM, Schorn T, Zani S et al (2019) Circ_Lrp6, a circular RNA enriched in vascular smooth muscle cells, acts as a sponge regulating miRNA-145 function. Circ Res 124(4):498–510. https://doi.org/10.1161/CIRCRESAHA.118.314240
182. Liu X, Cheng Y, Yang J, Xu L, Zhang C (2012) Cell-specific effects of miR-221/222 in vessels: molecular mechanism and therapeutic application. J Mol Cell Cardiol 52(1):245–255. https://doi.org/10.1016/j.yjmcc.2011.11.008
183. Leung A, Trac C, Jin W, Lanting L, Akbany A, Sætrom P et al (2013) Novel long noncoding RNAs are regulated by angiotensin II in vascular smooth muscle cells. Circ Res 113(3):266–278. https://doi.org/10.1161/CIRCRESAHA.112.300849

184. Mahmoud AD, Ballantyne MD, Miscianinov V, Pinel K, Hung J, Scanlon JP et al (2019) The human-specific and smooth muscle cell-enriched LncRNA SMILR promotes proliferation by regulating mitotic CENPF mRNA and drives cell-cycle progression which can be targeted to limit vascular Remodeling. Circ Res 125(5):535–551. https://doi.org/10.1161/CIRCRESAHA.119.314876
185. Ye ZM, Yang S, Xia YP, Hu RT, Chen S, Li BW et al (2019) LncRNA MIAT sponges miR-149-5p to inhibit efferocytosis in advanced atherosclerosis through CD47 upregulation. Cell Death Dis 10(2):138. https://doi.org/10.1038/s41419-019-1409-4
186. Hung J, Scanlon JP, Mahmoud AD, Rodor J, Ballantyne M, Fontaine MAC et al (2020) Novel plaque enriched Long noncoding RNA in atherosclerotic macrophage regulation (PELATON). Arterioscler Thromb Vasc Biol 40(3):697–713. https://doi.org/10.1161/ATVBAHA.119.313430
187. Sun C, Fu Y, Gu X, Xi X, Peng X, Wang C et al (2020) Macrophage-enriched lncRNA RAPIA: a novel therapeutic target for atherosclerosis. Arterioscler Thromb Vasc Biol 40(6):1464–1478. https://doi.org/10.1161/ATVBAHA.119.313749
188. Gast M, Rauch BH, Nakagawa S, Haghikia A, Jasina A, Haas J et al (2019) Immune system-mediated atherosclerosis caused by deficiency of long non-coding RNA MALAT1 in ApoE−/−mice. Cardiovasc Res 115(2):302–314. https://doi.org/10.1093/cvr/cvy202
189. Courboulin A, Paulin R, Giguère NJ, Saksouk N, Perreault T, Meloche J et al (2011) Role for miR-204 in human pulmonary arterial hypertension. J Exp Med 208(3):535–548. https://doi.org/10.1084/jem.20101812
190. Tanzer A, Stadler PF (2004) Molecular evolution of a microRNA cluster. J Mol Biol 339(2):327–335. https://doi.org/10.1016/j.jmb.2004.03.065
191. Chen T, Zhou G, Zhou Q, Tang H, Ibe JC, Cheng H et al (2015) Loss of microRNA-17∼92 in smooth muscle cells attenuates experimental pulmonary hypertension via induction of PDZ and LIM domain 5. Am J Respir Crit Care Med 191(6):678–692. https://doi.org/10.1164/rccm.201405-0941OC
192. Elia L, Quintavalle M, Zhang J, Contu R, Cossu L, Latronico MV et al (2009) The knockout of miR-143 and -145 alters smooth muscle cell maintenance and vascular homeostasis in mice: correlates with human disease. Cell Death Differ 16(12):1590–1598. https://doi.org/10.1038/cdd.2009.153
193. Caruso P, Dempsie Y, Stevens HC, McDonald RA, Long L, Lu R et al (2012) A role for miR-145 in pulmonary arterial hypertension: evidence from mouse models and patient samples. Circ Res 111(3):290–300. https://doi.org/10.1161/CIRCRESAHA.112.267591
194. Davis-Dusenbery BN, Chan MC, Reno KE, Weisman AS, Layne MD, Lagna G et al (2011) Down-regulation of Kruppel-like factor-4 (KLF4) by microRNA-143/145 is critical for modulation of vascular smooth muscle cell phenotype by transforming growth factor-beta and bone morphogenetic protein 4. J Biol Chem 286(32):28097–28110. https://doi.org/10.1074/jbc.M111.236950
195. Li W, Dunmore BJ, Morrell NW (2010) Bone morphogenetic protein type II receptor mutations causing protein misfolding in heritable pulmonary arterial hypertension. Proc Am Thorac Soc 7(6):395–398. https://doi.org/10.1513/pats.201002-024AW
196. Brock M, Trenkmann M, Gay RE, Michel BA, Gay S, Fischler M et al (2009) Interleukin-6 modulates the expression of the bone morphogenic protein receptor type II through a novel STAT3-microRNA cluster 17/92 pathway. Circ Res 104(10):1184–1191. https://doi.org/10.1161/CIRCRESAHA.109.197491
197. Zhang J, He Y, Yan X, Chen S, He M, Lei Y et al (2020) MicroRNA-483 amelioration of experimental pulmonary hypertension. EMBO Mol Med 12(5):e11303. https://doi.org/10.15252/emmm.201911303
198. Neumann P, Jaé N, Knau A, Glaser SF, Fouani Y, Rossbach O et al (2018) The lncRNA GATA6-AS epigenetically regulates endothelial gene expression via interaction with LOXL2. Nat Commun 9(1):237. https://doi.org/10.1038/s41467-017-02431-1
199. Voellenkle C, Garcia-Manteiga JM, Pedrotti S, Perfetti A, De Toma I, Da Silva D et al (2016) Implication of Long noncoding RNAs in the endothelial cell response to hypoxia revealed by RNA-sequencing. Sci Rep 6:24141. https://doi.org/10.1038/srep24141

200. Zhang H, Liu Y, Yan L, Wang S, Zhang M, Ma C et al (2019) Long noncoding RNA Hoxaas3 contributes to hypoxia-induced pulmonary artery smooth muscle cell proliferation. Cardiovasc Res 115(3):647–657. https://doi.org/10.1093/cvr/cvy250
201. Chen J, Guo J, Cui X, Dai Y, Tang Z, Qu J et al (2018) The Long noncoding RNA LnRPT is regulated by PDGF-BB and modulates the proliferation of pulmonary artery smooth muscle cells. Am J Respir Cell Mol Biol 58(2):181–193. https://doi.org/10.1165/rcmb.2017-0111OC
202. Su H, Xu X, Yan C, Shi Y, Hu Y, Dong L et al (2018) LncRNA H19 promotes the proliferation of pulmonary artery smooth muscle cells through AT. Respir Res 19(1):254. https://doi.org/10.1186/s12931-018-0956-z
203. Carpenter S, Aiello D, Atianand MK, Ricci EP, Gandhi P, Hall LL et al (2013) A long noncoding RNA mediates both activation and repression of immune response genes. Science 341(6147):789–792. https://doi.org/10.1126/science.1240925
204. Cheng G, He L, Zhang Y (2020) LincRNA-Cox2 promotes pulmonary arterial hypertension by regulating the let-7a-mediated STAT3 signaling pathway. Mol Cell Biochem 475(1–2): 239–247. https://doi.org/10.1007/s11010-020-03877-6