Luke B. Hesson · Antonia L. Pritchard
Editors

Clinical Epigenetics

Springer

Clinical Epigenetics

Luke B. Hesson • Antonia L. Pritchard
Editors

Clinical Epigenetics

Springer

Editors
Luke B. Hesson
Faculty of Medicine
UNSW
NSW, NSW, Australia

Antonia L. Pritchard
Oncogenomics
QIMR Berghofer Medical Research Institute
Brisbane, QLD, Australia

ISBN 978-981-13-8957-3 ISBN 978-981-13-8958-0 (eBook)
https://doi.org/10.1007/978-981-13-8958-0

This Springer imprint is published by the registered company Springer Nature Singapore Pte Ltd.
The registered company address is: 152 Beach Road, #21-01/04 Gateway East, Singapore 189721, Singapore

Preface

The purpose of this book *Clinical Epigenetics* is threefold: (1) to introduce epigenetics to a non-expert scientific audience, (2) to describe current and future clinical utility of epigenetics, and (3) to help professionals (scientists and clinicians) to understand, in a practical sense, how epigenetics is relevant in a clinical context. In the discipline of genetic pathology, genetic testing informs the clinical management of patients, including their diagnosis. Genetic testing is widely used in a variety of clinical contexts, whereas epigenetics is currently underappreciated and underutilised. This book emphasises the importance of epigenetics in health and disease with the aim of bringing it closer to being more regularly used in diagnostic laboratories.

Chapter 1 provides a historical perspective on the origins and development of the field of epigenetics and describes the pivotal research findings of early pioneers. The timeline of major discoveries will be discussed in the context of genetic knowledge at the time. Key terms will be introduced and defined in footnotes to allow uninterrupted flow of the narrative. These have been included to inform and guide the reader through an ever-evolving lexicon. As we will learn in more detail in Chaps. 1, 2 and 3, one of the most important epigenetic modifications is DNA methylation. This chemical modification to DNA affects the way it is packaged within the nucleus of the cell. Higher levels of packaging can make DNA inaccessible and therefore unreadable. Chapter 4 describes the composition of chromatin (the combination of DNA and protein within the nucleus of a cell) and details the various ways in which chromatin can be modified to regulate DNA function. The remainder of the book focuses on the clinical relevance of epigenetics. This will begin with the importance of non-coding RNA in disease (Chap. 5), the role of histone variants in cancer (Chap. 6), epigenetic changes in cancer development (Chap. 7), the role of specific regions of the genome known as *cis*-regulatory elements that regulate genes over long distances (Chap. 8), the importance of epigenetics in imprinting disorders (Chap. 9) and the clinical utility of measuring the methylation of circulating tumour DNA (Chap. 10). Finally, Chap. 11 provides a case study to illustrate the use of epigenetic testing in genetic pathology.

Compiling this book was an international effort with contributions from around the world including Australia, France, Hong Kong, the UK and the USA. This book would not have been possible without the expertise of these world-leading scientists

who have given their time to write chapters focused on this theme. In addition, as editors we are also indebted to experts who provided critique and invaluable discussion, in particular Mathew A. Sloane, Neil A. Youngson, Marina Berbic, Hayley Espanol and Dilshan Kalpage for providing feedback during the writing of this book.

Epigenetics: A Lay Description

Of the many analogies that have been proposed to introduce the topic of epigenetics, by far the clearest is proposed by Laura Bonetta. She states, "If our genome is referred to as the book of life, written in A's, T's, G's, and C's, then the epigenome represents the spaces and punctuation that format the text." This beautifully simple analogy can be taken further; epigenetics tells the cell how to interpret our DNA, where to start reading, where to stop, where each word begins and ends and even when to read so that the appropriate sentence is read at the appropriate time. Extending this further, each sentence represents a gene and the linear order of sentences combines to form a chapter (chromosome). The collection of chapters constitutes the book (genome). Typographical errors may be represented by misspelt words (sequence variants) or errors in punctuation (epigenetic differences). In either case, these may or may not change the meaning of a sentence. Punctuation can dramatically alter the way we read a sentence just as epigenetics can dramatically alter the way cells read our DNA. As an example, consider the following two sentences:

A woman without her man is nothing
A woman: without her, man is nothing!

The addition of punctuation not only alters the meaning; the two sentences offer totally opposing viewpoints. Likewise, there are examples throughout our genomes where the presence or absence of epigenetic marks produces opposing effects, having dramatic effects on the behaviour of cells.

So, how does the cell punctuate our genomes? This is achieved through chemical tags that decorate our chromosomes and serve as signposts to instruct the cell what to read, or more accurately, what parts of the genome are readable. In each type of cell in the body, these chemical tags archive specific sections of the genome making them unreadable. The complement of genes that are readable and unreadable or switched on and off is important for defining the function of that cell type. This is known as cell lineage, and once a cell is committed to a particular lineage, each daughter cell thereafter carries the memory of this identity. In other words, each time a liver cell (e.g. hepatocyte) divides, its daughter cells will only be capable of being hepatocytes, rather than a different type of liver cell (e.g. a Kupffer cell) or a different cell type from another part of the body. Epigenetics is therefore responsible for defining and maintaining cellular identity. It does this by ensuring that once sections of the genome are archived they remain that way, even after the cell divides (i.e. the

catalogue of genes switched off is heritable across cell division). When we consider how our cells use and read our DNA, the sequence of our genomes is therefore only part of the story. As eloquently summarised by Siddhartha Mukherjee in his book *The Gene*, "each cell can write a unique novel from the same basic script".

Epigenetics: A Scientific Description

The term epigenetics was first proposed by Conrad Waddington to refer to "*all those events which lead to the unfolding of the genetic program for development*". Epigenetics is derived from the term epigenesis, a much older term that was used to refer to the process by which the separate tissues of an organism develop gradually through a series of stages of differentiation. When using the term epigenetics, Waddington was referring to the interaction between genetic mutations and cell fate during development. David Nanney further developed the field by suggesting that the term should also invoke the concepts of cellular memory, persistent homeostasis and changes to cell fate that are not attributable to changes in DNA sequence. These definitions focused on cellular properties, but over time scientists increasingly sought to understand the mechanistic basis of how cell fate was determined. In the 1970s, two papers from Arthur Riggs, Robin Holliday and John Pugh began the transition of the use of the term epigenetics towards a molecular definition. The identification of DNA methylation as a major epigenetic mark, its heritability during cell division and its relationship with gene activity offered a mechanistic understanding of processes such as X-inactivation, cell fate and differentiation. Today the term is used in reference to Waddington's and Nanney's original definitions of cellular properties and in Riggs' and Holliday's definition of molecular or mechanistic processes. It is therefore important that the definition of epigenetics we are using in this book is clearly stated to avoid ambiguity. Throughout this book the term is used in reference to molecular or mechanistic processes and is defined as *heritable chemical or structural modifications to chromatin that alter the function of DNA*. This definition encompasses chemical modifications to DNA and histones, histone variants, nucleosome positioning and nuclear organisation, but omits factors not specifically associated with chromatin. This definition also distinguishes stable heritable modification to chromatin from transient changes that might be part of acute transcriptional regulation. In contrast to many chromatin modifications, DNA methylation in the context of CpG dinucleotides has a clear mechanism for heritability across cell division, and this has led some to argue that it should be considered the only truly epigenetic mark. However, the regulation of gene expression involves the coordination of multiple molecular changes to chromatin. These include the repositioning and occupancy of nucleosomes and the presence of specific histone modifications and histone variants. These different layers of regulation are tightly linked with DNA methylation. Moreover, when considering the factors that give rise to stable gene silencing, DNA methylation is in fact a relatively late molecular event that is thought to consolidate and lock down a state of transcriptional inactivity. Therefore, DNA methylation is just one step in a hierarchy of epigenetic events that

regulates gene activity. Finally, the definition we have given does not contain the prerequisite that DNA sequence is unaltered. This is simply because there are many well-characterised examples of epigenetic changes, many of which cause or predispose to disease, which have a genetic cue; in other words, the epigenetic alteration is instructed by a genetic alteration in either *cis* or *trans* (i.e. on the same or different chromosomes). Specific examples and the mechanisms that illustrate this are provided in the final chapter of this book.

Sydney, NSW — Luke B. Hesson
Brisbane, QLD — Antonia L. Pritchard

Contents

About the Editors

Luke B. Hesson is a Clinical Scientist who has studied the genetic and epigenetic causes of cancer for more than 18 years. He holds a doctorate in Clinical Genetics from the University of Birmingham, UK, and is a former Director of the Australian Society for Medical Research (2012–2015). In 2016, he was admitted as a Fellow of the Faculty of Science of the Royal College of Pathologists of Australasia. He has lectured at the University of New South Wales, Sydney, on genetics, epigenetics, cancer, pathology and personalised medicine. Luke is currently the Head of Cancer Diagnostics and Laboratory Director at the Garvan Institute of Medical Research DNA sequencing facility in Sydney, which provides clinically accredited genome sequencing.

Antonia L. Pritchard is currently a Senior Lecturer at the University of the Highlands and Islands, Scotland, UK, and a visiting scientist at the QIMR Berghofer Medical Research Institute, Brisbane, Australia. She received her PhD in Psychiatric Genetics from the University of Birmingham in 2005. Her research interests chiefly concern genetic influences on disease (including Alzheimer's disease, asthma and cancer) and on the human immune system. Dr Pritchard has published over 60 peer-reviewed articles on these topics and continues to work on the interaction of genetics, genomics and immunology in cancer, particularly with regard to melanoma. She is a member of the British Society of Immunology (BSI), American Association for Cancer Research (AACR), Association of Cancer Immunotherapy (CIMT) and the Society for Melanoma Research (SMR). Dr Pritchard's research group is part of the GenoMEL Consortium and the BAP1 Interest Group (BIG) Consortium, both of which are highly productive, collaborative and multidisciplinary organisations bringing together researchers from around the world.

Genetics and Epigenetics: A Historical Overview

1

Luke B. Hesson and Antonia L. Pritchard

Abstract

Epigenetics has a fascinating and convoluted history steeped within the fields of embryology and genetics. Here, we introduce genetics and epigenetics to the non-expert reader and give an account of the pivotal discoveries that have helped shape these fields. The significance of major discoveries will be explained and their impact on our understanding of epigenetic mechanisms in health and disease will be discussed.

Keywords

Genetics · Epigenetics · History · Timeline · DNA methylation · X-chromosome · Imprinting · Chromatin

1.1 The Early Origins of Genetics

Long before the term 'genetics' was conceived, the basis of inheritance had been extensively debated. It is difficult to choose a specific point in history where the true basis of inheritance began to take shape; however, much can be said of the contributions of the mathematician and scientific philosopher Pierre Louis de Maupertuis. In 1751, he hypothesised that both parents contributed equally to their

L. B. Hesson (✉)
Garvan Institute for Medical Research, Darlinghurst, NSW, Australia

Prince of Wales Clinical School, UNSW, Sydney, NSW, Australia
e-mail: l.hesson@unsw.edu.au

A. L. Pritchard (✉)
Genetics and Immunology Research Group, Centre for Health Science, University of the Highlands and Islands, Inverness, UK
e-mail: Antonia.Pritchard@uhi.ac.uk

L. B. Hesson, A. L. Pritchard (eds.), *Clinical Epigenetics*,
https://doi.org/10.1007/978-981-13-8958-0_1

offspring and proposed a particulate basis of heredity. The concept of particulate inheritance was opposed to the beliefs of other philosophers of the time who argued that the characteristics of each parent were blended together in the next generation. Instead, his theory of particulate inheritance proposed that it had a physical basis caused by discrete particles (now known as genes) that are not diluted or diminished in the next generation. Shortly after, in 1753, Maupertuis was the first to apply probability estimates to predict disease risk in his study of a family with polydactyly and is credited with striking insights into the possibility of natural selection (Stubbe 1972). In his book *Essai de Cosmologie* published in 1751 (de Maupertuis 1751), he argued that variation in animals and plants arose spontaneously but that only a small proportion of individuals showed fitness and survival.

Decades later in the early nineteenth century, Jean-Baptiste Lamarck proposed the theory of inheritance of acquired characteristics, in which he suggested that species gradually developed characteristics that suited the physical conditions of life. In his publications *Recherches sur l'organisation des corps vivants* (de Monet de Lamarck 1802), and later in *Philosophie Zoologique* (de Monet de Lamarck 1809), Lamarck argued that a key driving force of his theory of inheritance was the effects of use and disuse and that the memory of these acquired characteristics were passed to future generations and therefore perpetuated. He used the example of the blind mole rat to illustrate loss of function through disuse. Though this theory is now widely disregarded as a major contributor to inheritance, it was nevertheless one of the first attempts to provide a tangible theory for biological evolution. Lamarck also asserted in his book *Philosophie Zoologique* that species, including man, are descended from other species.

The theory of natural selection is of course attributed to Charles Darwin in his book *On the Origin of Species by Means of Natural Selection or the Preservation of Favoured Races in the Struggle for Life* (often referred to simply as *Origin of Species*), published in 1859 (Darwin 1859). In the introduction of *Origin of Species*, two sentences perfectly summarise his theory: 'As many more individuals of each species are born than can possibly survive; and as, consequently, there is a frequently recurring struggle for existence, it follows that any being, if it vary however slightly in any manner profitable to itself, under the complex and sometimes varying conditions of life, will have a better chance of surviving, and thus be naturally selected. From the strong principle of inheritance, any selected variety will tend to propagate its new and modified form' (Darwin 1859). *Origin of Species* challenged the popular belief at the time that species were immutable, static in nature and as God, the Creator, had designed. Darwin himself grappled with the implications of his theories of natural selection and evolution, as he had also believed that life was created in its present form. However, *Origin of Species* presented the objective evidence, collected by himself and others over many decades, that particular species had adapted to their environments over many generations.

By Darwin's own admission, his theory of evolution was imperfect; however, Darwin did not try to hide the faults in his theory and in fact discussed them extensively in his book. He clearly communicated that evolution and inheritance are inextricably linked but also acknowledged that his greatest problem was the absence of a mechanism by which traits were inherited from parent to offspring. This

came in Gregor Mendel's solution to the problem of inheritance in pea plants in 1865–1866 (Mendel 1866); however, throughout his lifetime, Darwin remained unaware of Gregor Mendel and his work. Darwin continued to publish updated editions of *Origin of Species*, with the sixth and final edition published in 1872, and published his work on human evolution in *The Descent of Man and Selection in Relation to Sex* in 1872 (Darwin 1872).

Gregor Mendel is now widely regarded as the founder of the modern science of genetics, with the rules he established on the basis of heredity now referred to as Mendelian inheritance. Initially overlooked, Mendel's work on the inheritance pattern of pea plant characteristics was rediscovered in 1900, which lead to the establishment of genetics as a scientific field of study. The term genetics was coined by William Bateson in 1905, with the word originating from the Greek γεννώ (gennó), which translates as 'to give birth'. He proposed the term should be used to describe the study of heredity, or how characteristics are transferred from parent to offspring.

1.2 Discovery of DNA

Deoxyribonucleic acid was first isolated from the nuclei of pus cells from surgical bandages in 1869 by the Swiss biologist Friedrich Meischer. Meischer was a mentee of the German scientist Felix Hoppe-Seyler, who is widely regarded as the founder of the disciplines of biochemistry and molecular biology. Meischer showed that the substance he had isolated was acidic, abundant in phosphorous and resistant to enzymes that degrade proteins. Given that he had isolated this substance from nuclei, he named it 'nuclein'. Meischer submitted his work describing nuclein to Hoppe-Seyler, who was also editor of the journal *Medizinisch-chemische Untersuchungen* (Medical and Chemical Analysis). The unusual properties of nuclein described by Meischer were so unlike anything Hoppe-Seyler had seen previously that he decided to delay publication for 2 years whilst he repeated Meischer's experiments with the help of two students Pal Plósz and Nikolai Lübavin. Once repeated, and confident of the accuracy of Meischer's results, he finally published the original paper in 1871, another from Meischer describing a similar substance isolated from egg yolk and work from his own lab showing the isolation of nuclein from yeast and other cells. Neither Meischer nor Hoppe-Seyler appreciated the importance of nuclein in heredity; however, they recognised that its abundance and unusual chemical properties suggested it was important to the biology of the cell nucleus.

1.3 Early Characterisation of DNA

From 1878, the biochemist Albrecht Kossel, also a mentee of Hoppe-Seyler, began efforts to identify and characterise the chemical composition of the nucleus in greater detail. Kossel's first major discovery in 1884 was the isolation of a type of protein (which he named histone) from the red blood corpuscles of birds that purified with

nucleic acid (Kossel 1884). Following Meischer's discovery of nuclein, many considered it chemically non-distinct from protein or closely related. This was due to methodological limitations at the time that prevented the isolation of pure protein-free nuclein. In the 1880s and 1890s, methods developed by Kossel and others (including Richard Altmann) allowed the isolation of protein-free nuclein and definition of its chemical composition. In 1889, following the discovery[1] that protein-free nuclein was acidic, Altmann proposed the term nucleic acid (Altmann 1889). Efforts then turned to defining the building blocks of nucleic acid by characterising the products produced by breaking the chemical down into its constituent parts in a process called hydrolysis. Between 1885 and 1901, Kossel's laboratory identified the nucleobases present in nucleic acid, beginning with adenine (A) in 1885 and later thymine (T), cytosine (C), guanine (G) and uracil (U).[2] These nucleobases, known as pyrimidines (cytosine, thymine and uracil) and purines (adenine and guanine), combine with phosphate and sugar in the form of ribose or deoxyribose. By the beginning of the twentieth century, Kossel had therefore helped define the basic building blocks of nucleic acids, eventually leading to the terms deoxyribonucleic acid (DNA) and ribonucleic acid (RNA).

1.4 Discovering That Genes Are Made of DNA

Based on its chemical properties, Meischer initially considered that DNA functioned to store phosphorous within the cell. Although its abundance in sperm lead Meischer and others to speculate a role in heredity, this was considered improbable due to its limited chemical composition (Friedrich 1874). In this section, we will describe the scientific discoveries that lead to the realisation that genes are made of DNA.

An appropriate starting point is the work of embryologist Thomas Morgan and his students Alfred Sturtevant, Calvin Bridges and Hermann Muller whose work in the early years of the twentieth century first demonstrated that inherited characteristics within the model organism *Drosophila melanogaster* were determined by physical units carried within chromosomes. Morgan's team were the first to describe the role of chromosomes in sex-linked inheritance, to create the first genetic linkage map, to fully describe crossing over (now known as recombination) and to describe chromosomal abnormalities including non-disjunction, duplications and translocations (Morgan 2018). These were all monumental milestones that firmly asserted the importance of chromosomes in inheritance and that the linear order of genes on a chromosome could be defined.

[1] Interestingly, the separation of nucleic acid from protein was so difficult at the time that when Richard Altmann achieved this, and described the acidic derivative (nucleic acid), he believed he was describing a novel subcomponent of nuclein. Friedrich Meischer had in fact defined nuclein as acidic in his original paper.

[2] Uracil is usually found only in RNA.

Throughout the first half of the twentieth century, the term gene was used without an understanding of its chemical basis. In fact, most biologists thought that proteins, which can be composed of up to 20 different amino acids, were much more likely to carry hereditary information (McCarty 2003). Part of the reason for this was the tetranucleotide hypothesis proposed by the Russian biochemist Phoebus Levene in ~1910, which suggested that DNA was composed of equal amounts of guanine, cytosine, thymine and adenine in a repeating ring structure (Levene and Mandel 1908). This simple repeating configuration of DNA was to remain the accepted model of DNA structure until the 1940s and for many years contributed to the perception that DNA was far too simple in structure to be responsible for the molecular basis of inheritance. However, in the 1940s, it was discovered that the proportions of nucleobases in DNA can be different across species (Chargaff et al. 1949). This challenged the tetranucleotide hypothesis and began resurgence in research to define the molecular structure of DNA, which we shall return to later in this chapter.

In 1928, the British bacteriologist Frederick Griffith described the first widely accepted example of bacterial transformation[3] (Avery 1941). Griffith was a meticulous scientist who painstakingly characterised the different types of bacteria isolated from patients with pneumonia (the leading cause of death at the time). What particularly intrigued Griffith was that some strains of these bacteria could switch from virulent to non-virulent forms and vice versa. In crucial work, often referred to as Griffith's Experiment, he injected mice with heat-inactivated preparations of the virulent strain and showed they did not develop pneumonia; however, when he injected both non-virulent and heat-inactivated virulent pneumococci, the mice developed pneumonia. This experiment suggested that a *transforming factor* persisted after heat-inactivation of virulent bacteria and that this transmitted virulence to the non-virulent strain (Griffith 1928). Though Griffith's Experiment did not identify DNA as the transforming factor, his work showed that the characteristics of bacteria, including virulence, could be transformed and inspired others to identify the transforming factor, including Oswald Avery, Colin MacLeod and Maclyn McCarty at the Rockefeller Institute for Medical Research in New York City. The Avery-MacLeod-McCarty experiment isolated or enzymatically destroyed the different components of bacteria to identify the chemical that retained this transforming power. They first killed bacteria with heat, isolated the saline soluble components that contained protein, polysaccharides and DNA, then purified DNA using alcohol precipitation. To test whether DNA was the transforming factor, they treated extracts with enzymes that selectively destroyed protein or DNA. The transforming power of the extract was lost only after DNA was destroyed, thereby providing the definitive proof[4] that the function of DNA was to carry genetic information (Avery et al. 1944).

[3]The term *transformation* describes the uptake of genetic material by a cell from its surroundings. This can be measured by a change in the form and function of the cell, such as the transformation of a non-virulent strain of bacteria into a virulent strain.

[4]Though others quickly confirmed the findings of the Avery-MacLeod-McCarty experiment the conclusion that DNA was the genetic material was still met with resistance from those that believed genes were made of protein. Further work in 1948 by Rollin Hotchkiss, a mentee of Avery, would

1.5 The Birth and Evolution of Epigenetics

In the late 1930s and early 1940s, the term epigenetics was coined by the British embryologist Conrad Waddington (Waddington 1939, 1940, 2012) to describe 'the branch of biology that studies the causal interactions between genes and their products which bring the phenotype into being' (Dupont et al. 2009; Waddington 1942). Waddington used the term broadly to describe the interaction between genetic mutations and cell differentiation (epigenesis); he was interested in how embryonic development unfolded at the molecular level. The term was further defined by David Nanney in the 1950s to include the concept of persistent homeostasis in the absence of genetic influence (Nanney 1958). This represented the first recognition of the importance of cellular memory in maintaining cell lineage and tissue type. To underscore the significance of epigenetics in developmental processes, Nanney highlighted some important cellular phenomenon that helped guide the use of the term (Nanney 1958), specifically:

1. Cells with the same genetic material may manifest different phenotypes.
2. Cellular properties are determined by the activity of an integrated set of genes.
3. Specific patterns of gene activity can be induced.
4. Epigenetic systems show stability (i.e. are heritable).
5. Epigenetic control systems are localised in the nucleus of the cell.[5]

These formed the basic tenets of epigenetics; however, the precise mechanisms that underpinned these cellular properties were not identified until decades later.

In 1975, two papers from Robin Holliday, John Pugh and Arthur Riggs, inextricably linked epigenetics with DNA methylation. Holliday, Pugh and Riggs proposed that DNA methylation was an epigenetic modification that regulated X-inactivation and gene expression (Holliday and Pugh 1975; Riggs 1975). For the first time, scientists could link cellular and phenotypic properties with differences at the molecular level. Consequently, the use of the term began to change from a description of cellular properties to one of molecular properties, beginning with DNA methylation and expanding to include all chromatin and DNA modifications that alter DNA function. The precise definition of the term epigenetics and its use (and misuse) have been extensively debated over many years (examples include Deans and Maggert 2015; Greally 2018; Haig 2004; Ptashne 2007). Throughout this book, the term epigenetics is used to refer to *heritable chemical or structural modifications to chromatin that alter the function of DNA* (see the Preface of this book for a more detailed description of the definition of epigenetics).

demonstrate that transformation was not the result of contaminating protein and that DNA was indeed responsible. Work by Alfred Hershey and Martha Chase in 1952 would provide the final confirmation by showing that it is the DNA from bacteriophage that enters the host bacterium.

[5]These cellular phenomena are not taken verbatim from Nanney's original 1958 paper and have been adapted for clarity.

1.6 The Double Helix Structure of DNA

In 1951, Erwin Chargaff and others built upon Levene's tetranucleotide hypothesis by showing that the relative amounts of A, T, C and G can differ between species and that, despite this diversity, the amount of adenine always equalled the amount of thymine and the amount of cytosine always equalled the amount of guanine (Chargaff 1951; Chargaff et al. 1949). Chargaff's findings were a pivotal piece of information that complemented crystallography studies by Rosalind Franklin, Maurice Wilkins and others, and collectively, these findings enabled the chemical structure of DNA to be deduced in 1953 (Watson and Crick 1953b; Wilkins et al. 1953). Their model described the double helix structure of DNA containing two antiparallel strands in which adenine on one strand pairs with thymine on the other and cytosine pairs with guanine. This pairing occurs through hydrogen bonds that can be broken, thereby allowing the two molecules to unwind and separate. This structure reconciled Franklin's crystallography observations of a symmetrical molecule and Chargaff's observations that the amount of guanine is equal to cytosine and the amount of adenine is equal to thymine. The double helix structure is hugely significant in genetics—but why? Firstly, it finally dispelled any doubts that genes were made of DNA by casting aside the argument that DNA was too simple in structure to encode something as complex as inheritance. The helical structure of DNA is remarkably elegant and has an immense capacity for complexity. As the length of a DNA molecule increases, the amount of information it can contain increases exponentially. At any given position on one strand of a DNA molecule, there are four possible letters, A, T, C or G. A two-base-pair-long piece of DNA has 16 possible combinations (AA, TT, CC, GG, AT, AC, AG, TA, TC, TG, CA, CT, CG, GA, GT, GC or 4×4 combinations). A three-base-pair-long piece of DNA has 64 possible combinations ($4 \times 4 \times 4$), four base pairs gives 256 combination, five base pairs has 1024 combinations and so on. At 150 base pairs long, a piece of DNA has more possible combinations than there are atoms in the observable universe (which is currently estimated to be between 10^{78} and 10^{82}). The true content of information is doubled when one considers that there is also a second strand within the double helix structure that contains a different sequence of complementary letters. This new model therefore made it abundantly clear that DNA had the requisite complexity to encode the instructions for life. Secondly, the double helix structure provided the answer to the molecular basis of inheritance by explaining how a DNA molecule can be copied for transfer between generations. As Watson and Crick explained in one of their 1953 papers, the DNA molecule 'is, in effect a pair of templates, each of which is complementary to the other. We imagine that prior to duplication the hydrogen bonds are broken down and the two chains unwind and separate. Each chain then acts as a template for the formation onto itself of a new companion chain so that eventually we shall have two pairs of chains, where we only had one before [...]. Moreover the sequence of pairs of bases will have been duplicated exactly' (Watson and Crick 1953a). The chemical structure of DNA is described in greater detail in Chap. 2.

1.7 The Discovery of DNA Methylation

One might anticipate that the discovery of DNA methylation occurred after the chemical structure of DNA was deduced; however, the existence of 5-methylcytosine was recognised long before.

Between 1898 and 1910, collaboration of the US chemists Henry Wheeler and Treat Johnson at Yale University helped to characterise the chemical natures of the nucleobases.[6] In 1904, Wheeler and Johnson hypothesised the existence of 5-methylcytosine and synthesised this artificially to characterise its chemical properties (Wheeler and Johnson 1904). Several years later in 1925, Johnson and Robert Coghill confirmed the presence of 5-methylcytosine as a natural constituent of DNA from the Mycobacterium tuberculosis (Johnson and Coghill 1925). However, the existence of 5-methylcytosine in mammalian cells was a serendipitous discovery of the American biochemist Rollin Hotchkiss. In one of his papers, Hotchkiss described a small amount of an additional base that separated from cytosine. This 'minor constituent' of calf thymus DNA was distinct from cytosine yet shared similar chemical characteristics suggesting a modified form, which he referred to as 'epicytosine' (Hotchkiss 1948). He correctly deduced this base was not uracil and that it was pre-existing in the nucleic acid before extraction from cells and therefore present naturally. He also noted that the properties of this additional base could be distinguished from cytosine in a similar way that 5-methyluracil (otherwise known as thymine) could be distinguished from uracil. Though he referred to 5-methylcytosine and the earlier work of Johnson and Coghill, he was cautious in his conclusions noting that, 'More than this cannot be said until further study of epicytosine has been made' (Hotchkiss 1948). This discovery did not precipitate an immediate impact; however, studies over the next two decades would find 5-methylcytosine within DNA from all vertebrate and plant species (Hall 1971). In mammalian DNA, approximately 2–7% of cytosine is converted to 5-methylcytosine (Vanyushin et al. 1970). This ubiquity suggested an important function and several hypotheses were proposed, including the regulation of protein-DNA interactions, gene regulation and differentiation, protection from eukaryotic restriction enzymes, the regulation of DNA replication, chromosome folding, packing and sorting, and recombination. However, its functional significance would remain unclear until the 1980s.

[6] A total of 48 research papers from this collaboration were published as a book titled *Papers on Pyrimidines* in 1910 and can be accessed freely at The Internet Archive at https://archive.org/details/papersonpyrimidi00wheerich.

1.8 The X-Chromosome and Its Unique Place in Genetics and Epigenetics

The story of the X-chromosome is unique in human genetics and, as we will see, has a particular place in the history of epigenetics. Its unique properties have enabled several conceptual leaps by providing the chromosomal basis of sex determination, X-linked inheritance and X-inactivation, as well as precedents for dosage compensation and the role of DNA methylation in gene activity.

The X-chromosome was first identified in 1891 by the German cytologist Hermann Henking. Henking studied the firebug *Pyrrhocoris apterus* and was intrigued by a heavily stained chromatin body at the periphery of the nucleus that seemed to distinguish cells from males *versus* females. Females contained 24 chromosomes that arranged into 12 pairs, whereas males had 11 pairs and one solitary chromosome. In one of the figures of his papers, he labelled this body of chromatin with an *X* and referred to it as the X-element, primarily because he was unsure if it was a chromosome.[7] Henking observed that sperm cells from *Pyrrhocoris* were of two kinds: those with the X-element and those without, in approximately equal numbers. This observable difference between the chromatin of sperm cells and the growing evidence at the time that it was the chromatin of cells that contained the hereditary information, led Clarence McClung to propose a chromosomal basis of sex determination in 1902 (McClung 1902). McClung's theory was significant because it provided early support for the chromosomal theory of inheritance, which was proposed the same year by Walter Sutton and Theodor Boveri. This introduced the concept that the physical characteristics of chromosomes (e.g. the number or type of chromosome) can drastically alter the physical characteristics of an organism, including something as fundamental as sex. Though McClung had identified the specific chromosome that appeared to be associated with sex determination he had mistakenly proposed that it was the number of chromosomes that determines sex. However, this was soon to be corrected in 1905 by the American cytologist Nettie Stevens, whose work helped solve the basis of what would become known as the XY sex determination system (Brush 1978). As a former student of Thomas Morgan, she had studied the fruit fly *Drosophila* several years before Morgan adopted this species as a model organism. In her studies of more than 50 species of beetles and flies, she investigated the chromosomal basis of sex determination. In the mealworm beetle *Tenebrio molitor*, she showed that males produced two kinds of sperm that contained either a small chromosome or a large chromosome, and that offspring that inherit the small chromosome were invariably male whereas those that inherited the large chromosome were invariably female. From this, she deduced that the chromosomal basis of sex depended on the presence

[7] Work by Clarence McClung at the University of Kansas in 1902 showed the X-element was indeed a chromosome (though McClung referred to it as the accessory chromosome). Edmund Wilson and Nettie Stevens are credited with designating the names *X-chromosome* and *Y-chromosome* following the discovery of the XY sex determination system.

or absence of the smaller chromosome. Though this mechanism of sex determination would prove relevant for many other species, including humans, she found that this was not true for all species. Around the same time, and independently, Edmund Wilson also confirmed that sex determination was much more complex and varied, depending on which species of insect was studied. Further contributions by both Stevens and Wilson showed that in the context of XY sex determination, females inherit two copies of the larger chromosome (and were referred to as XX), whilst males have one small and one large chromosome (and were referred to as XY). Therefore, it is the presence or absence of the Y-chromosome, rather than the number of chromosomes, that determines sex. As we now know, a consequence of this is 'sex-related' inheritance, which describes the inheritance of specific traits preferentially in one sex. In 1910, Morgan published his observations of the inheritance patterns of several traits in *Drosophila*, including the inheritance of white eyes instead of the normal red eyes. He first noted that white eyes occurred exclusively in males, but on further crosses, white-eyed females could be observed. From this he implied a physical relationship between the X-chromosome and the white-eyed trait. Morgan's paper was therefore the first to propose X-linked inheritance, and in doing so, the field of genetics made an important conceptual leap by showing that genes were physical entities that reside on chromosomes. At this point, the concept of the gene ceased to be a theoretical term with no physical evidence and the modern theory of the gene was born. This modern theory posited that genes were located on chromosomes, that they could be studied and experimentally manipulated, and that the linear order of genes on any given chromosome could be mapped, which Morgan and his team (in particular Alfred Sturtevant) did to great effect in the years that followed. Finally, in 1911, Wilson published a famous review in which he described the XY sex determination system and predicted its consequences for human X-linked inheritance in the context of haemophilia and colour blindness (Kingsland 2007; Wilson 1911). Prior to this, the patterns of inheritance of these disorders and their predominance in males had puzzled geneticists.

In humans, and other species where sex is determined by the presence or absence of a Y-chromosome,[8] females contain two X-chromosomes, whereas males contain only one. Therefore, female cells contain one extra copy of all X-linked genes that could, theoretically, lead to a profound imbalance in the expression of those genes between males and females. To overcome this, all somatic cells in female eutherian mammals[9] adopt a mechanism of *dosage compensation*, where one of the X-chromosomes is inactivated. In 1948, the Canadian researcher and physician Murray Barr and his student Ewart Bertram reported that cells from male and female cats could be distinguished by simply staining chromatin and viewing using a compound microscope. They discovered that somatic, non-dividing cells (specifically nerve cells) of male and female cats could be distinguished by the presence or absence of densely staining chromatin at the periphery of the nucleolus (Barr and

[8]In some species, it is the number of X chromosomes that determines sex.

[9]Eutherians include all placental mammals.

Bertram 1949). This densely staining chromatin was, at the time, referred to as 'sex chromatin', and later as the 'Barr body'. Barr and Bertram demonstrated that this was present in all female somatic cells, much like the structure described by Henking in insects over 50 years earlier. In their paper, they postulated that the Barr body was most likely comprised of the X-chromosomes and that in female cells, the presence of two X-chromosomes explained its visibility. For the next decade, the theory persisted that the Barr body represented the tight pairing of the two X-chromosomes; however, its precise nature remained a topic of intense investigation. The next major breakthrough was offered by Susumu Ohno and Theodore Hauschka in 1960 whose investigation of the Barr body in cells from female mice suggested it was comprised of a single X-chromosome (Ohno and Hauschka 1960).[10] This raised an interesting question: was it the paternal X or the maternal X? A theory favoured by Ohno and Hauschka was that with each cell division this 'alternate[d] between the two X's ... regardless of their parental derivation' (Ohno and Hauschka 1960). However, this was soon superseded by a radical theory that was to have major implications for the field of epigenetics.

Mary Lyon was a visionary British geneticist whose work had a profound impact in clinical genetics. In 1961, soon after Ohno and Hauschka had proposed the theory that the Barr body was comprised of a single X-chromosome, Lyon proposed that this X-chromosome 'can be either paternal or maternal in origin, in different cells of the same animal', and that 'it is genetically inactivated' (Lyon 1961). Lyon's hypothesis was derived from her years of experience in mouse genetics and mouse cytology and was based on two key pieces of genetic evidence:

1. That mice with only one X-chromosome and no Y-chromosome (known as XO) are normal, fertile females (Welshons and Russell 1959)[11] and show no evidence of a Barr body, which shows that only one active X-chromosome is necessary for normal development of the female mouse.
2. Specific X-linked traits in female animals, including the patchy appearance of different coat colours in female cats, suggested different X-chromosomes were active in different cells of the same animal—also known as *genetic mosaicism*.[12]

Lyon's theory was significant and far reaching in its implications. It proposed the concept of X-inactivation, a form of dosage compensation that renders one of the two X-chromosomes in XX females genetically inert. By indicating that the inactive X could be either paternal or maternal in origin, and that this could differ in different cells of the same animal (selected at random), Lyon's theory also further explained

[10]In 1991, Barbara Migeon and colleagues used fluorescence in situ hybridisation to show that the Barr body is comprised of a single X-chromosome folded in two and attached at both ends to the periphery of the nucleus (Walker et al. 1991).

[11]Though XO female mice are phenotypically normal and fertile, in humans, individuals with only one X-chromosome are infertile.

[12]The term mosaicism is used in genetics to describe genetic differences between cells of the same organism.

the unusual characteristics of X-linked traits and diseases. These included colour blindness and haemophilia, traits that are rarely seen in females due to the presence of two X-chromosomes and the random inactivation of one, which prevents their penetrance.[13] Importantly, X-inactivation is a clear example of gene activity that is heritable through cell division because once one of the X-chromosomes is inactivated, this 'choice' is then fixed in all daughter cells thereafter. As an example, in cats, coat colour is determined by a gene on the X-chromosome and the pigment cells in each patch are descended from one cell with inactivation of a specific X-chromosome. Lyon was exacting in her hypothesis, going so far as stating, 'Patches of intermediate color would arise by cell mingling in development, and the shape of the patches would depend on cell movement during growth' and that this would 'vary from one individual to another by chance' (Lyon 1962). Interestingly, Lyon also showed that when genes from other chromosomes are abnormally translocated to the X-chromosome they too can become inactivated. This showed that the mechanism of X-inactivation was chromosome-wide and occurred independently of DNA sequence. At the time, it was unclear at what stage of embryonic development X-inactivation occurred, but it was known to be an early event that was already established by the late blastocyst stage in humans (Austin and Amoroso 1957; Park 1957). In honour of Lyon's contribution, X-inactivation was for many years referred to as *Lyonisation*, though this term is no longer used.[14]

The concept of X-inactivation was initially met with some resistance, with sceptics of the theory arguing that if surplus X-chromosomes are completely inert due to inactivation then why would sex chromosome aneuploidy[15] result in any clinical symptoms? These include females with a single X, known as Turner syndrome (referred to by cytogeneticists as 45,XO), who show developmental delay, infertility and short stature with extra folds of skin on the neck and low set ears. Males with XXY, known as Klinefelter syndrome (47,XXY) are characterised by low muscle tone, underdeveloped testicles, infertility and the development of breast tissue (gynaecomastia). There are various other described combinations (47, XXX, 48,XXXX, etc.), with each exhibiting variable clinical symptoms. In 45,XO cells, no Barr body is observed, whereas in cells containing 47,XXY a single Barr body is observed, and in cells with more than two X-chromosomes, multiple Barr bodies are found. During the 1960s, various explanations were proposed including incomplete inactivation and the effects of abnormal dosage in early development prior to inactivation (Lyon 1963). However, it was soon realised that specific genes on the X-chromosome that are involved in development escape inactivation, thereby

[13]The term penetrance is used in genetics to describe the proportion of individuals carrying a particular variant (or allele) of a gene (the genotype) that also express an associated trait (the phenotype).

[14]In 2017, The Royal Society paid tribute to Mary Lyon with a series of review articles in the journal *Philosophical Transactions of the Royal Society B* (http://rstb.royalsocietypublishing.org/content/372/1733) and audio recordings (https://royalsociety.org/science-events-and-lectures/2016/10/x-chromosome-inactivation/), for readers interested in further details regarding X-inactivation.

[15]The term aneuploidy refers to an abnormal number of chromosomes.

resulting in an imbalance of some genes in individuals with sex chromosome aneuploidy (Burch and Burwell 1963).[16]

X-inactivation posed a molecular puzzle; how are most of the thousands of genes located on only one of the X-chromosomes coordinately inactivated? Two landmark papers published independently by Arthur Riggs (1975), Robin Holliday and John Pugh (1975) would provide the stimulus for several conceptual advances toward answering this question.

Riggs was a researcher in the department of Susumu Ohno, one of the discoverers of X-inactivation, at the City of Hope Medical Centre, Los Angeles and had a keen interest in understanding gene regulation in *E.coli*. During a short visit to Herbert Boyer's laboratory at the University of California, San Francisco, Riggs learned about a group of proteins known as restriction enzymes. One enzyme in particular, named EcoK, was a methyltransferase in *E.coli* and had a preference for hemimethylated DNA. Riggs realised a similar enzyme in mammalian cells could provide a mechanism of cellular heredity by maintaining patterns of methylation following DNA replication. In his paper, he proposed that DNA methylation was important for the X-inactivation process and that there existed a hitherto unrecognised information-coding system based on methylation patterns. His theory referred specifically to the X-inactivation process but clearly had broader implications for gene regulation. Interestingly, Riggs' paper was promptly rejected by the first journal and Ohno had to convince him to persist by submitting his theory for publication elsewhere (Riggs 1988). Independently, Robin Holliday and his student John Pugh were based at the National Institute for Medical Research in London. Similar to Riggs, Pugh and Holliday had concluded that methylation of DNA could be a mechanism for gene regulation, but also that gene expression patterns could be stably inherited during cell division. Moreover, Holliday and Pugh clearly conveyed the concept that methylation might also switch genes on and off during development. At the time, it was well appreciated that 5-methylcytosine was abundant in higher organisms (Doskocil and Sorm 1962; Wyatt 1951); however, its functional effects, if any, were unclear. An additional concept discussed in the paper by Holliday and Pugh was that DNA mutations could be caused by the deamination of 5-methylcytosine to generate thymine, which had earlier been proposed by Eduardo Scarano (1971). We now know that the methylation of cytosine is an important cause of transition to thymine and a major mutation mechanism in cancer.

How the cell carries out X-inactivation is an extremely complex process that is still under investigation, despite decades of research. The precise mechanisms of X-inactivation likely differ between species (Migeon 2017). The critical role of DNA methylation was confirmed by experiments showing that inhibition of

[16]In humans, approximately 15% of X-linked genes escape X-inactivation, whereas in mice it is only 3% (Berletch et al. 2011), which explains why XO mice remain fertile. Interestingly, species comparison of genes on the X-chromosome suggests genes that escape X-inactivation once originated from autosomal chromosomes (chromosomes other than the sex chromosomes X or Y), which are not subject to dosage compensation (Ballabio and Willard 1992).

methylation with the drug 5-azacytidine resulted in the reactivation of genes on the inactive X-chromosome (Mohandas et al. 1981). 1991 saw a major breakthrough with several publications in the journal *Nature* from laboratories in the UK and the USA (Borsani et al. 1991; Brockdorff et al. 1991; Brown et al. 1991a, b). These showed that X-inactivation was dependent on a region on the X-chromosome, designated the X-inactivation centre (XIC), from which the inactivation signal is initiated and spreads throughout the rest of the chromosome. DNA elements within the XIC regulate the different aspects of X-inactivation, namely the counting of X-chromosomes, choosing one for silencing, initiating the silencing and finally maintenance of the inactive state (Lu et al. 2017). Within the XIC, the gene *XIST* is expressed specifically from the inactive X-chromosome. The RNA produced from this gene does not encode protein and is known as a long non-coding RNA (lncRNA). Instead, the *XIST* lncRNA coats the X-chromosome and, through interactions with a range of different proteins, results in exclusion of the transcriptional machinery, chromatin modifications and tethering of the future inactive X to the inner nuclear membrane where most of the genes on the chromosome are silenced by DNA methylation.[17] The importance of the *XIST* transcript is clear from observations that mutations within the promoter of the gene that cause changes to *XIST* expression are associated with skewed X-inactivation (Plenge et al. 1997) whereby one of the X-chromosomes is preferentially inactivated.

Up to 200 different proteins interact with *XIST* through repetitive sequences in the lncRNA (Chu et al. 2015; Lu et al. 2017; Minajigi et al. 2015). One important interacting protein is known as LBR (laminin B receptor), which serves as a bridge to tether the *XIST*-coated X-chromosome to the inner nuclear membrane. Other interacting proteins include various histone modifying enzymes, which prompt the condensation of the X-chromosome and extinguish gene expression. Recent research suggests the existence of additional genes on some autosomes[18] that regulate X-inactivation. Evidence for this theory includes the fact that in diploid[19] cells with three X-chromosomes (47,XXX), all but one is subject to X-inactivation, whereas triploid cells (69,XXX and 69,XXY) contain two active X's (Migeon et al. 1979, 2008; Weaver and Gartler 1975). Maintenance of more than one active X-chromosome suggests that ploidy (the number of sets of autosomes) is important as a counting mechanism. This has led some to suggest that in human cells there exists a dosage-sensitive *XIST* repressor, encoded by an autosomal gene (Migeon et al. 2008). In this respect, one or more of several genes on chromosome 19 may be important (Migeon 2017).

[17]Methylation of the inactive X chromosome is dependent on the gene *SMCHD1* (Blewitt et al. 2008), which is an important regulator of DNA methylation genome-wide.

[18]The term autosome refers to a chromosome other than one of the sex chromosomes. In humans, there are 22 pairs of autosomes and two sex chromosomes.

[19]The term diploid is used to define a cell or nucleus containing two complete sets of chromosomes, one from each parent. In a triploid cell or nucleus, three complete sets of chromosomes are present.

1.9 Heritability of DNA Methylation

DNA methylation helps to determine cellular identity by controlling the complement of genes that are expressed in a cell. As such, the heritability of DNA methylation across cell division has major biological implications for ensuring maintenance of this identity and gene expression patterns in daughter cells. How the levels and precise location of DNA methylation are maintained remained unclear until relatively recently; however, a famous experiment from 1958 would later provide a crucial piece of this puzzle.

Following the discovery of the structure of DNA, there were three hypotheses for the mechanism by which replication might occur. The *conservative* hypothesis proposed that an entirely new DNA molecule was formed using the original as a template. The *dispersive* hypothesis, proposed by Max Delbrück, was a complicated combination of DNA unwinding, double-stranded DNA breaks, copying and the end-to-end ligation to form two copies comprised of both newly synthesised and original template DNA. Finally, in the *semi-conservative* hypothesis, proposed by Watson and Crick, the two strands of a DNA molecule separate and act as the template for the synthesis of complementary strands. This generates two DNA molecules each consisting of one original template strand and one newly synthesised strand. The Meselson-Stahl experiment, which illuminated which of these hypotheses was correct, has been called 'the most beautiful experiment in biology' (Judson 1979). Meselson and Stahl cultured *E. coli* for several generations in the presence of the nitrogen isotope ^{15}N ('heavy' nitrogen), which contains an additional neutron when compared with the naturally abundant ^{14}N. During cell division, this isotope became incorporated into newly synthesised DNA, which would later allow it to be distinguished by virtue of molecular weight using a method known as density centrifugation. *E.coli* cultured in the presence of ^{14}N contained 'light' DNA, whereas in the presence of ^{15}N the DNA became 'heavy'. They then transferred the *E.coli* back to ^{14}N and observed that after one cell division the DNA showed an intermediate weight. Though this result excluded conservative replication, which would have given equal amounts of heavy and light DNA, it was consistent with both the semi-conservative and dispersive hypotheses. However, after two rounds of cell division equal amounts of intermediate and light DNA were observed, which is consistent only with semi-conservative DNA replication; dispersive replication would have resulted in DNA with a weight between intermediate and light, since the heavier ^{15}N DNA would have been further diluted by ^{14}N DNA and would be evenly distributed throughout all DNA molecules. Meselson and Stahl had managed to distil the complex process of DNA replication into a simple readout (the presence or absence of DNA at a specific weight) with predictable results.

Many years later in the 1990s and 2000s, the model of semi-conservative replication was to help answer the question of how DNA methylation was maintained during cell division. However, another crucial aspect of DNA methylation was yet to be discovered, specifically, how DNA methylation is established; in other words, how does cytosine become 5-methylcytosine? In 1959, the American biochemist Arthur Kornberg, who in the same year won the Nobel Prize for helping to decipher the mechanisms of DNA and RNA synthesis, suggested the possibility of an enzymatic mechanism for the methylation of DNA as a pre-formed polymer (Kornberg et al. 1959). In this model, DNA is firstly

synthesised then an enzyme methylates specific sites along the DNA polymer. In 1963, this model proved accurate when the first descriptions of the enzymatic methylation of DNA in bacteria emerged (Gold et al. 1963) and by 1965 it was realised that bacteria contained at least two enzymes: those that methylate adenine to generate 6-methyladenine and those that methylate cytosine to generate 5-methylcytosine (Fujimoto et al. 1965). These enzymes became known as DNA methyltransferases, and though much of the early research of these enzymes focused on those found in bacterial species, it was anticipated that similar enzymes would exist in human cells.

In the 1980s, Timothy Bestor and colleagues at the Massachusetts Institute of Technology (MIT), Cambridge, USA, identified, cloned and characterised the first mammalian DNA methyltransferase from mouse cells, now known as DNMT1 (DNA methyltransferase 1), and showed it had a marked preference for DNA when only one strand is methylated (Bestor et al. 1988). Referred to as hemimethylated DNA, this describes DNA molecules that contain methylated cytosine on only one DNA strand (hence 'hemi', or 'half', methylated). The 1980s, 1990s and 2000s saw huge strides in our understanding of the cellular machinery that regulates DNA methylation and the identification of a family of DNA and RNA methyltransferases (described further in Chaps. 2 and 3).

We now know that during the cell cycle, DNMT1 expression dramatically increases and becomes localised to sites of DNA replication where it binds to newly synthesised DNA molecules (Leonhardt et al. 1992; Szyf et al. 1985). Due to the semi-conservative nature of DNA replication, each newly formed DNA molecule contains one original DNA strand that is modified by a specific pattern of DNA methylation, whereas the newly synthesised strand is unmethylated. In 2004, Albert Jeltsch and colleagues confirmed Bestor's earlier findings that DNMT1 preferentially methylates cytosine at hemimethylated targets sites (Hermann et al. 2004). DNMT1 uses the original methylated DNA strand as a template to replicate the pattern of methylation on the newly synthesised DNA strand. If no methylation is present on the original strand then DNMT1 will not methylated the newly synthesised strand. The essential role of DNMT1 in maintaining DNA methylation has led to its description as the maintenance methyltransferase.

1.10 Genomic Imprinting

Each somatic cell in our body contains two sets of autosomes, one set that is maternally inherited and one set that is paternally inherited, and therefore each parent contributes a copy of every gene.[20] For some time, it was therefore assumed that the genetic

[20]In the human genome, duplicated or deleted regions known as copy number variants can result in changes to the number of copies of specific genes. This may be responsible for genotypic and phenotypic variation between individuals. An example is the gene *CYP2D6*, which is responsible for the metabolism of a range of different drugs. The presence of more than two copies of this gene can cause individuals to metabolise some drugs more rapidly than individuals with two copies, which has implications for the dose of drug they should receive.

contribution of each parent was equal. However, we now know some genes show parental-specific expression and, as a result, the contribution of each parent is different. Genomic imprinting is an epigenetic mechanism that restricts the expression of a gene to one of the two parental chromosomes (Barlow and Bartolomei 2014). Mechanistically, genomic imprinting is a complex phenomenon and, as with many pioneering discoveries, its existence was met with a large degree of scepticism for many years.

During the 1920s and 1930s, the laboratory of Charles Metz at the Carnegie Institution of Washington in Baltimore published several studies describing the embryonic development of Sciara, a type of *Diptera* (two-winged fly) that was amenable to genetic and cytological study at the time. As an embryologist, Metz was particularly interested in the unusual behaviour of chromosomes during development and spermatogenesis in this species, and the role this plays in sex determination (Metz 1938). During spermatogenesis all paternal chromosomes are selectively eliminated, but two identical copies of the maternal X-chromosome are retained. Therefore, the sperm contributes two X-chromosomes to the zygote whereas the oocyte contributes one X-chromosome and a complete set of autosomes. During development of the embryo, one or both of the X-chromosomes derived from the sperm are selectively eliminated so either one remains (XO) to derive a male or two remain (XX) to derive a female (Crouse 1960). This selective elimination of paternal chromosomes is dependent on a remarkable feature; the ability of the cell to somehow distinguish the parental origin of each chromosome. Helen Crouse, a student of Metz, realised that the mechanism by which the cell achieves this must be erasable, so that chromosomes that are recognised as paternal in origin in one generation may be recognised as maternal in the next, and vice versa. This unorthodox behaviour of chromosomes, driven by their parental origin, seemed to violate apparently well-established principles of heredity and revealed a new mechanism of inheritance. It was Crouse who first used the term 'imprint' to describe this unusual behaviour of chromosomes, commenting: '. . .the dramatic chromosome unorthodoxies in Sciara are clearly unrelated to the genic make-up of the chromosomes: a chromosome which passes through the male germ line acquires an "imprint" which will result in behavior exactly opposite to the "imprint" conferred on the same chromosome by the female germ line. In other words, the "imprint" a chromosome bears is unrelated to the genic constitution of the chromosome and is determined only by the sex of the germ line through which the chromosome has been inherited' (Crouse 1960). Though the exact nature of the 'imprint' remained mysterious, Crouse's own research had suggested that a specific segment of the X-chromosome, which she termed the 'controlling element', might be important in this process (Crouse 1960). This early work in Sciara is significant because it foreshadowed the discovery of similar mechanisms of inheritance in other species, including humans.

Chromosomal imprinting in mammals was first recognised in 1971 when it was shown that the paternal X-chromosome was inactivated in all cells of female marsupials and in extra-embryonic tissues[21] during mouse development (Cooper et al. 1971). Further work during the 1970s described specific traits in mice when

[21]Extra-embryonic tissues are those that contribute to the placenta.

they were engineered to contain two maternal or paternal copies of specific chromosomes (Barlow and Bartolomei 2014). These experiments suggested that the expression of only the paternal or the maternal copy of specific genes was important for normal mouse development (Searle and Beechey 1978).

A major breakthrough came in the 1980s when two papers were published showing that mouse zygotes[22] generated from entirely paternal or maternal DNA were not viable (McGrath and Solter 1984; Surani and Barton 1983). These studies confirmed that both parental genomes are required for normal mouse development. The experiments involved taking fertilised zygotes containing maternal and paternal pronuclei[23] and using nuclear transfer techniques to engineer zygotes containing either two maternal pronuclei, two paternal pronuclei or one of each. Embryos derived from zygotes with both a maternal and paternal pronuclei survived, as expected. However, embryos derived from only maternal pronuclei were defective in extraembryonic tissues, whereas embryos derived from only paternal pronuclei were defective in embryonic tissue. This suggested that the development of extra-embryonic tissues requires genes specifically expressed in the paternal genome, whereas development of the embryo requires genes specifically expressed in the maternal genome (Barton et al. 1984). The definitive proof of the existence of genomic imprinting in mammals came in 1991 when several papers published in the journals *Cell* and *Nature* identified three genes (*Igf2, Igf2r* and *H19*) that were expressed specifically from either the maternal or paternal chromosomes in mice (Barlow et al. 1991; Bartolomei et al. 1991; DeChiara et al. 1991; Ferguson-Smith et al. 1991).

1.11 Why Do Genes Become Imprinted?

Since the initial discovery of genomic imprinting in mammals, more than 100 imprinted genes have been identified. Most of these genes have a known role in embryonic and neonatal development and regulate the growth of embryonic or extra-embryonic tissues. Inquiries in different species show that imprinting occurs in placental mammals and marsupials, but not in egg-laying mammals. A distinguishing feature between these species is dependence on maternal nutrition during gestation. In placental mammals, a developing embryo can access maternal resources, whereas embryos developing within an egg cannot. Interestingly, some egg-laying animals, for example, some lizards, can undergo asexual reproduction whereby the embryo develops from a single oocyte following duplication of the maternal genome. This suggests that in egg-laying animals there is no prerequisite

[22]A zygote is a diploid cell that results from the fusion of two haploid gametes (i.e. a fertilised oocyte).

[23]The term *pronucleus* describes the structure within the oocyte or sperm that contains the maternal or paternal chromosomes, respectively. Around 12 h after fertilisation of the oocyte by a sperm the maternal and paternal pronuclei fuse to form the nucleus.

for both parental genomes, and therefore no imprinting, again indicating that the dependence of mammals on maternal resources during gestation may be key to understanding why imprinting exists. In 1991, Tom Moore and David Haig proposed the 'parental conflict' theory to explain the emergence of imprinting during evolution (Moore and Haig 1991). In this theory, Moore and Haig suggested: 'imprinting has evolved in mammals because of the conflicting interests of maternal and paternal genes in relation to the transfer of nutrients from the mother to her offspring' (Moore and Haig 1991). This theory posits that paternal imprints activate or repress the expression of genes to promote the uptake of nutrients from the mother, whereas maternal imprints limit this uptake. Another theory, named the 'trophoblast defense' theory, was proposed in 1994 by Susannah Varmuza and Mellissa Mann. This theory suggests that imprinting arose to protect females from spontaneous pregnancy and malignant trophoblast disease (Varmuza and Mann 1994). However, neither theory fully explains the imprinting of all genes, since some are specifically involved in neuronal development during the neonatal period.[24]

1.12 How Do Genes Become Imprinted?

How is a cell able to distinguish the maternal and paternal copies of the same gene and what is the precise nature of the 'imprint'? During gamete formation, the maternal and paternal chromosomes are separate, thereby providing a window of opportunity to establish parental imprints before fertilisation. Once fertilisation has occurred, these imprints allow the cell to distinguish the maternal and paternal copies of the same chromosome. Therefore, it is thought that the maternally and paternally derived genomes are already differently marked (imprinted) before the two genomes combine to form the zygote. After the discovery of imprinted genes in mammals, scientists began to focus on the mechanisms responsible. It was concluded that it must involve epigenetic modification of DNA because imprinting had been observed in inbred strains of mice wherein the two parental chromosomes contained identical DNA sequences. Crouse's description of the switching of parental identity of the same chromosome in successive generations added weight to this hypothesis (see above). As a molecular imprint, DNA methylation fulfils several necessary criteria. Firstly, de novo methyltransferases present in the sperm or oocyte can methylate specific regions[25] thereby providing a means to establish an imprint. Secondly, the maintenance methyltransferase can maintain an imprint following cell division and throughout development. Finally, DNA can also be demethylated (described further in Chap. 3), thereby allowing the imprint to be erased in the germline and reset in the

[24]It is possible that neuronal changes may affect behaviour and competition for maternal resources during the neonatal period, including the regulation of appetite for maternal milk, which would support the parental conflict model.

[25]It is currently unclear how a specific region is selected for methylation.

next generation. To qualify as an imprint, the epigenetic mark must be present in only one of the gametes, persist after fertilisation and be maintained throughout development. Scientists therefore suspected the mechanism was *cis*-acting,[26] thereby allowing targeting of only one chromosome. The role of DNA methylation in X-inactivation and the availability of methods to measure its presence or absence provided a clear and testable candidate.

DNA methylation, it turned out, was a key factor governing genomic imprinting. A crucial insight was the realisation that imprinted genes are often found in clusters. In fact, around 80% of imprinted genes are clustered together within imprinted domains (Reik and Walter 2001), which provides strong evidence that they are coordinately regulated. Within imprinted gene clusters that have been well characterised, a common theme is methylation of at least one, and in some instances two, regions specifically on one of the parental chromosomes within the gametes, which regulates the imprinted expression of the entire cluster. These *differentially methylated regions* are maintained after fertilisation and in all cells of the developing embryo. The necessity of these differentially methylated regions is clear from experiments that show imprinting is lost following their deletion. For the majority of clusters, it is the maternal chromosome that contains the methylation imprint. Another common theme is the presence of at least one lncRNA within the imprinted cluster, which is also expressed specifically from one of the parental chromosomes (lncRNAs have been observed in all but one of the well-characterised clusters). Regions that are essential for establishing imprinting within an imprinted gene cluster are known as *imprint control elements*. Though DNA methylation of imprint control elements plays a critical role in imprinting, the precise mechanisms at play differ depending on the cluster. In several instances, the imprint control element overlaps with the promoter of the lncRNA within the cluster, and methylation silences its expression. Importantly, the expression of the lncRNA within an imprinted domain is crucial to establishing imprinted gene expression patterns. The specific mechanisms involved in imprinting at several important regions and the relevance of disorders of imprinting in human disease are described in Chap. 9.

1.13 Histones, Nucleosomes and Chromatin Structure

Despite knowledge from the 1800s that chromatin was composed of histones and DNA, the fine structure and three-dimensional arrangement of chromatin within the cell nucleus remained an enigma until the 1970s. As late as 1972, the prevailing hypothesis was of a superhelical structure in which DNA was coated with a layer of histones (Olins and Olins 2003). However, the 1960s and 70s saw a period of discovery that revolutionised our understanding of chromatin structure, and ultimately how this regulated DNA function. These discoveries included the

[26]The term *cis* in genetics refers to two genetic features on the same chromosome. In this context, the element controlling imprinting and the imprinted gene must be on the same chromosome.

demonstration that chromosomes are uninemic, i.e. that they represent a single DNA molecule running from end to end (Gall 1963); the fractionation, purification and characterisation of the different histones (Johns 1969); the recognition that modification of histones (acetylation and methylation) may regulate gene expression (Allfrey et al. 1964); improvements to methods for preparing and visualising chromatin using the transmission electron microscope (Zubay and Doty 1959); and the demonstration that approximately 50% of the DNA in chromatin was accessible to enzymes that degrade DNA and therefore not covered with proteins (Clark and Felsenfeld 1971; Itzhaki 1971; Mirsky 1971). However, Ada and Donald Olins, and Christopher Woodcock, who in 1973 independently visualised DNA and spherical histone particles in a repeating unit like beads on a string (Fig. 1.1a–c), made by far the biggest conceptual leap in this productive period (Olins and Olins 1973, 1974; Woodcock 1973; Woodcock et al. 1976). A year later Roger Kornberg and Jean Thomas described the chromatin subunit model (R. D. Kornberg 1974; Kornberg and Thomas 1974), in which DNA wraps around an octamer of histone proteins containing one histone $(H3–H4)_2$ tetramer and two histone H2A–H2B dimers. This model was supported by independent evidence showing interactions between histones (D'Anna and Isenberg 1974; Roark et al. 1974) and by biochemical data showing that chromatin structure is a repeating unit (Oudet et al. 1975). In 1975, this repeating unit of chromatin was named the nucleosome (Oudet et al. 1975).

The nucleosome is the basic functional unit of chromatin and its discovery revolutionised the perception of how DNA function is regulated. As described by Ada and Donald Olins: 'Higher-order packaging of chromosomal DNA and DNA-based processes, such as transcription, replication and repair, were now all viewed through a different lens. DNA was no longer seen as being coated by histones (superhelical models), but conceived as being coiled on the outside of a globular histone core, which is accessible to the binding of other nuclear proteins. The nucleosome became the 'quantum' of chromatin structure, the fundamental unit for the modulation of chromatin function' (Olins and Olins 2003).

Over the next two decades the crystal structure of the nucleosome core particle would be solved with ever increasing resolution. In 1997, stunning high-resolution images of the nucleosome revealed the orientation of histones at the core and the histone amino-terminal tails that can be chemically modified (Fig. 1.1d, Luger et al. 1997). These protein tails emanate from the nucleosome core particle where they are accessible to a range of enzymes capable of adding or removing a wide range of post-translational modifications, including acetylation and methylation (Fig. 1.1e). The next challenges included defining the histone code and its relationship with gene expression, nucleosome occupancy, nucleosome positioning and DNA methylation (Chaps. 4 and 5 describe these aspects of epigenetics in greater detail) and deciphering the three-dimensional organisation of chromatin in the nucleus and gene regulation through long-range chromatin interactions.

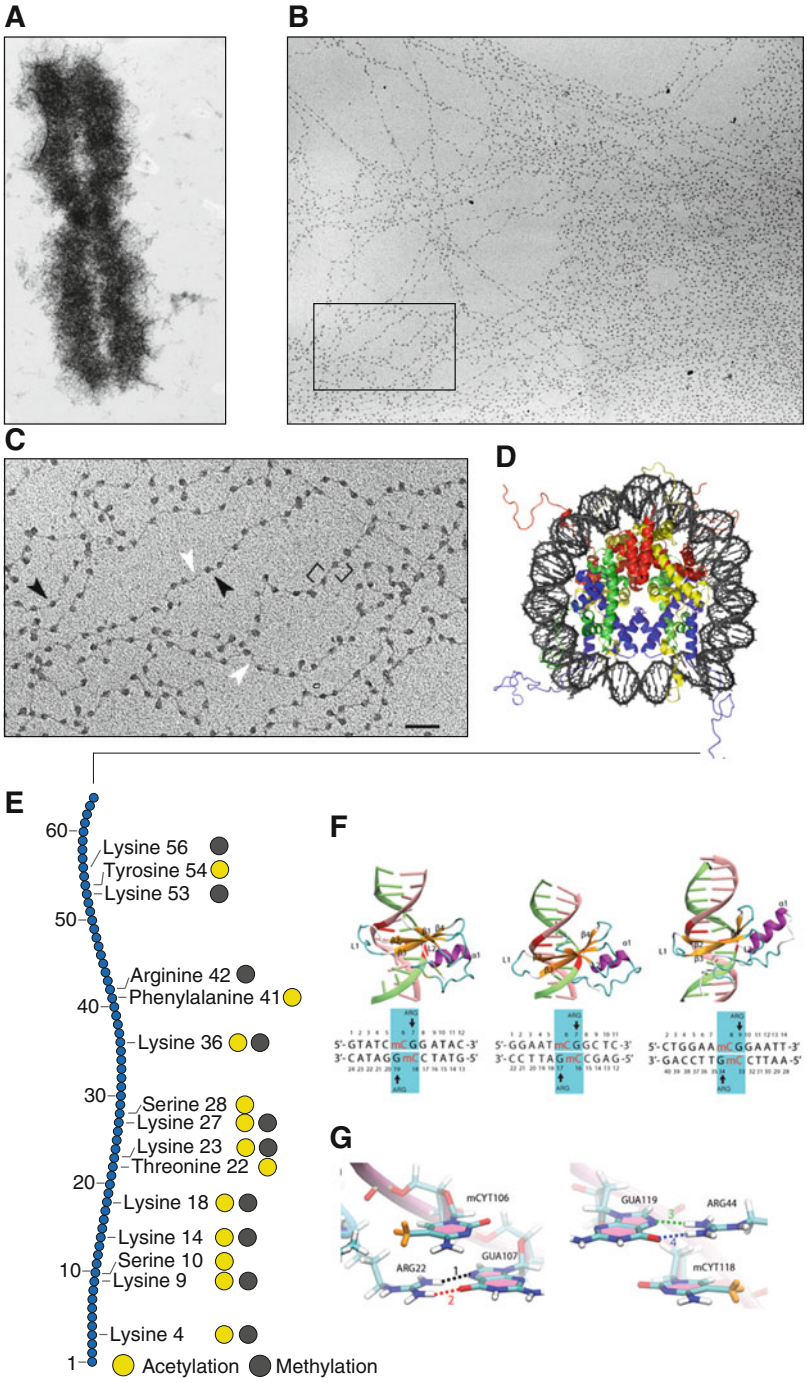

Fig. 1.1 Chromatin structure. (**a**) An isolated metaphase chromatid pair from a mouse cell. (**b**) Transmission electron microscopy image of chromatin prepared from chicken erythrocytes. (**c**) A closer look at chromatin from the box in panel b. Black arrow heads indicate nucleosome core particles. White arrow heads indicate linker DNA between nucleosomes. Black brackets indicate nucleosomes and linker

1.14 Cancer Epigenetics

Two non-exclusive hypotheses have been proposed for the origin of cancer: (1) alterations to stem cells[27] that result in loss of controlled proliferation and (2) the dedifferentiation of mature cells with specialised functions to stem-like cells that retain the ability to proliferate. In either case, cells must undergo genetic and epigenetic changes that enable these altered behaviours. Cancer can therefore be considered as a disease of altered differentiation.

By the early 1980s, reports that DNA methylation played a role in differentiation were beginning to emerge (Ehrlich et al. 1982; Jones and Taylor 1980; Mandel and Chambon 1979; McGhee and Ginder 1979; Razin and Riggs 1980; Shen and Maniatis 1980; Taylor and Jones 1979; van der Ploeg and Flavell 1980). These studies showed that methylation patterns were tissue-specific, heritable following cell division, and that inhibition of the enzymes that methylate DNA (DNA methyltransferases) disrupts methylation patterns and can alter the differentiation of cells. Given that cancer can be considered a disease of altered differentiation it seemed a natural progression to question whether DNA methylation was altered between normal and cancer tissues. In 1983, two papers from Andrew Feinberg, Bert Vogelstein and the laboratory of Melanie Ehrlich showed that tumours had lower levels of DNA methylation than matched normal tissue (Feinberg and Vogelstein 1983; Gama-Sosa et al. 1983). This hypomethylation was universal across tumour types and was evident in both benign and malignant tumours, suggesting it was an early event in the development of tumours (Feinberg et al. 1988; Goelz et al. 1985). It was also clear by this point that there was a relationship between methylation levels and gene activity; in normal tissues, active (expressed) genes were always less methylated (Razin and Riggs 1980). Accordingly, reports began to emerge of genes showing hypomethylation and overexpression (relative to normal tissue) or hypermethylation and transcriptional silencing (Feinberg and Tycko 2004). In

Fig. 1.1 (continued) DNA. (**d**) The crystal structure of the nucleosome core particle consisting of H2A (yellow), H2B (red), H3 (blue) and H4 (green) core histones, and DNA. (**e**) Schematic of the N-terminal tails of Histone H3 showing the amino acid positions that can be methylated or acetylated. (**f**) Nuclear Magnetic Resonance (NMR) images of the binding of methyl-CpG binding domain (MBD) proteins bound to methylated DNA. Left, MBD1; centre, MBD2; right, MeCP2. (**g**) Interaction of MBD1 and methylated DNA at the atomic level. Left, hydrogen bonding between the H- and N-atoms (black line) and the H- and O-atoms (red line) in arginine 22 (ARG22) and guanine at position 107 (GUA107). Right, hydrogen bonding between the H- and N-atoms in the ARG44 and GUA119 (green line), H- and O-atoms in the ARG44-GUA119 pair (blue line). Images from Cell Image Library (http://www.cellimagelibrary.org) using the following accession numbers: Chromosome (panel A), CIL:40682, chromatin (panels B and C), CIL:709 (provided by Christopher Woodcock). Image of nucleosome core particle in panel D taken from https://en.wikipedia.org/wiki/Nucleosome. NMR images in panels F and G taken from Zou et al. (2012)

[27]Stem cells have the ability to perpetually self-renew and can differentiate into specific cell types.

1989, the first example of epigenetic inactivation[28] of a known tumour suppressor was described for the retinoblastoma (*RB1*) gene (Greger et al. 1989; Sakai et al. 1991). This led to widespread acceptance that epigenetic alterations can contribute to the development of cancer and the subsequent description of epigenetic activation or inactivation of many cancer-related genes (Baylin and Jones 2016; see Chap. 7 for a detailed review of the role DNA methylation changes in cancer).

1.15 A Molecular Definition of the Term 'Gene'

As discussed above, the term 'gene' was conceived long before its molecular attributes were known. The term itself has evolved over many years and has been reviewed previously (Portin and Wilkins 2017). In this section, we will briefly describe this evolution and offer a contemporary molecular definition.

When proposed by Wilhelm Johannsen in 1909, the word gene was an abstract term used to refer to a 'unit of inheritance'. Thomas Morgan and his team confirmed that genes were physical entities in the early twentieth century by demonstrating that they reside on chromosomes in a linear order. However, the term was still somewhat abstract in that it had now become a dimensionless point on a chromosome (Portin and Wilkins 2017). In 1941, George Beadle and Edward Tatum showed that each gene directs the synthesis of a protein, also known as the one gene, one enzyme hypothesis (Beadle and Tatum 1941). Confirmation by Avery, MacLeod and McCarty that DNA was the hereditary molecule in 1944 and the discovery of the structure of DNA in 1953 gave the gene its chemical identity. In the 1950s and 60s, the central dogma of molecular biology was formulated primarily by Francis Crick to describe the unidirectional transfer of information from DNA to messenger RNA (mRNA) and protein during processes known as transcription (DNA to mRNA) and translation (mRNA to protein, (F. Crick 1970; F. H. Crick 1958)). Inherent to this model is the genetic code, whereby triplets (codons) of nucleobases composed of A, T, C and G encode information in the form of amino acids or stop codons. During the 1960s, Robert Holley, Har Gobind Khorana and Marshall Nirenberg cracked the genetic code by deciphering which codons encoded which amino acids, for which they received the Nobel Prize in 1968. By this time it was thought that a gene was a contiguous segment of DNA that encodes a protein.

In the late 1970s, it was realised that genes can contain many exons (coding DNA) interrupted by introns [non-coding DNA, Fig. 1.2 (Berget and Sharp 1977; Berk and Sharp 1977; Chow et al. 1977)]. Before the discovery of introns, it was thought that mRNA molecules were faithful copies of the DNA sequence from the genome; however, on average, around 90% of a gene is comprised of intronic sequence which is spliced out to generate mature mRNA. This discovery changed the way scientists thought about the architecture of the human genome and revealed

[28]In the context of cancer, epigenetic inactivation refers to the hypermethylation of the promoter of a gene and its transcriptional silencing in tumour tissue relative to matched normal tissue.

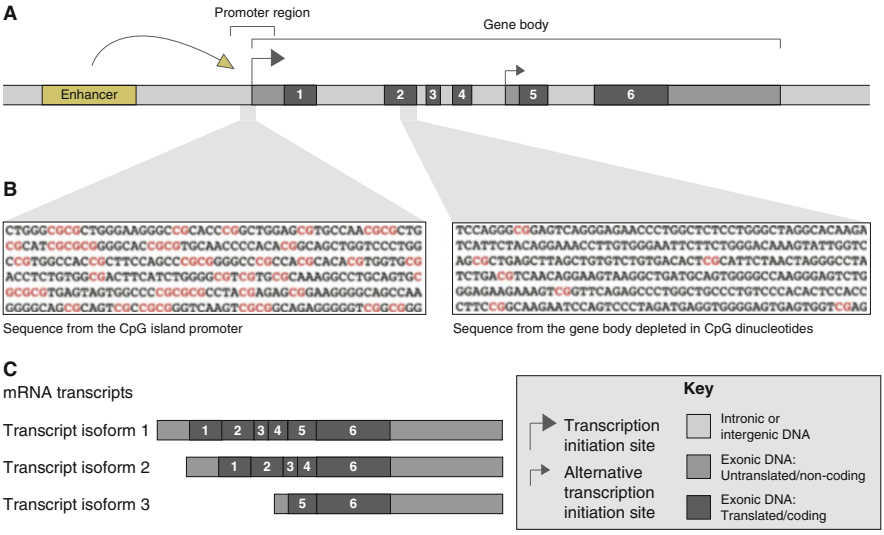

Fig. 1.2 Molecular definition of the gene. (**a**) A gene can be defined as a DNA sequence (whose component segments do not need to be physically contiguous) that includes the regulatory sequences that can control expression and that produces one or more sequence-related RNAs/proteins. Depicted is an enhancer upstream of a multi-exon gene with two transcription initiation sites and a promoter overlapping a CpG island. Genes can reside on either the positive or the negative DNA strand. *Cis*-regulatory elements (CREs) such as enhancers and silencers usually operate irrespective of the orientation or strand relative to their target genes. One gene can have multiple CREs and the distance between them is highly variable. The positioning of a CRE may be upstream, downstream or within the gene body. A gene may have multiple promoters and transcription initiation sites. (**b**) CpG islands are stretches of DNA that are enriched in CpG dinucleotides (see section titled 'CpG islands'). In approximately 72% of human genes the promoter region has a high CpG content. (**c**) A gene can give rise to multiple different mRNA transcript isoforms though differential promoter usage and alternative splicing

that genes can have multiple isoforms due to alternative splicing[29] and the use of different transcription initiation sites. In humans, around 95% of genes with more than one exon are alternatively spliced (Pan et al. 2008). Alternative splicing can vary across tissue types, developmental stages or disease states and represents a mechanism of increasing the functionality encoded within a single gene.

In addition, there are other non-coding DNA sequences that are important in gene function and should be included in the definition of a gene. These are known as *cis*-regulatory elements and include gene promoters, enhancers and silencers. Gene promoters are found near the transcription initiation site of a gene and contain specific DNA sequences that recruit proteins (transcription factors, transcriptional activators or repressors) that are important in the initiation of gene transcription.

[29]Alternative mRNA splicing describes the joining together of different combinations of exons. This can vary the sequence of an mRNA, and thus protein. The discovery of alternative splicing earned Richard Roberts and Phil Sharp the Nobel Prize in 1993.

Enhancers and silencers are often many kilobases away from a target gene but can regulate their activity by physical interaction with the promoter via by chromatin looping, which bring two distant loci on the same chromosome into physical proximity. Identifying the enhancer for a specific gene can be extremely difficult because their positioning relative to each other can be highly variable. Also, one enhancer may interact with many genes, one gene may interact with multiple enhancers and these interactions may be tissue- or stimulus-specific (Fig. 1.3). The first enhancer was discovered at the immunoglobulin heavy chain locus (Banerji et al. 1983; Gillies et al. 1983; Mercola et al. 1983); however, it is now estimated that there are hundreds of thousands of enhancers scattered throughout the human genome (Pennacchio et al. 2013). A large part of the prolific discovery of these regulatory regions has been due to the development of methods for mapping three-dimensional chromatin structure, massively parallel sequencing technologies and international initiatives such as the Encyclopedia of DNA elements (ENCODE) and the International Human Epigenome Consortium (IHEC). These are described in greater detail later in this chapter.

To conclude, many modern definitions of the term gene have been proposed (Burian 2004; Griffiths and Stotz 2006; Keller and Harel 2007; Moss 2003; Pesole 2008; Portin and Wilkins 2017; Scherrer and Jost 2007; Stadler et al. 2009). Here, we use the term to refer to *a DNA sequence (whose component segments do not need to be physically contiguous) that includes the regulatory sequences that can control expression of the gene and that produces one or more sequence-related RNAs/proteins*. However, biology and genetics are seldom simple and some exceptions that challenge this general definition of a gene include RNA editing,[30] gene sharing,[31] gene fusions events[32] and pseudogenes[33] (Portin and Wilkins 2017).

1.16 CpG Islands

Levels of DNA methylation vary widely across the human genome, which is divided into heavily methylated and non-methylated domains. This is because the primary target of DNA methylation in human cells is cytosine that precedes guanine, also known as the CpG dinucleotide, which is not evenly distributed across the genome. The term CpG is used to distinguish the single-stranded linear sequence (where the 'p' represents the phosphate backbone of DNA) from the complementary base pairing of C and G on opposite strands. Approximately 70–80% of CpG cytosines are methylated in mammalian DNA (Jabbari and Bernardi 2004). Throughout

[30]Post-transcriptional alterations to RNA sequence.

[31]Gene sharing describes rare circumstances whereby the same protein sequence derived from the same gene can assume different conformations and functions in different cellular contexts.

[32]The physiological transcription of genes in tandem to produce hybrid mRNA molecules.

[33]A pseudogene is a non-functioning copy of another gene which may under some circumstances produce mRNA.

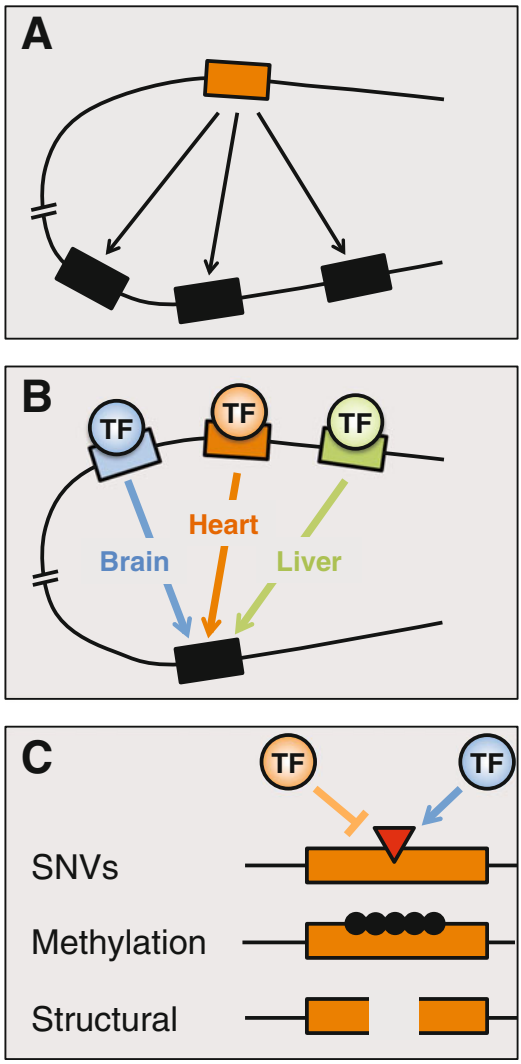

Fig. 1.3 *Cis*-regulatory elements. *Cis*-regulatory elements (CREs) include enhancers and silencers that can be located many kilobases from a target gene. Black boxes represent genes. Coloured boxes represent CREs. TF, transcription factor. (**a**) One CRE may target interact with multiple genes. (**b**) Multiple CREs may interact with one gene and this may depend on the tissue type or developmental stage. This may occur due to the expression of a specific transcription factor in a particular tissue type or stage of development. (**c**) The regulation of CREs may be altered by DNA sequence variants (red triangle), DNA methylation (black circles) or structural variants including copy number alterations. In either case, these can modify the binding of transcription factors to the CRE

mammalian genomes, CpG dinucleotides are under-represented and cluster within regions known as CpG islands. These islands are defined as stretches of DNA that are at least 200 bp in length and contain a GC percentage greater than 50% and an observed-to-expected CpG ratio greater than 60% (Gardiner-Garden and Frommer 1987). CpG islands are often found at the start of a gene, but many exist within repetitive DNA sequences spread throughout the genome, including more than one million copies of a repetitive sequence known as the Alu element (Szmulewicz et al. 1998). Importantly, CpG islands frequently overlap gene promoters (Fig. 1.2b). In the human genome, 72% of gene promoters contain a high CpG content (Saxonov et al. 2006). It has been known for some time that CpG islands overlapping gene

promoters are protected from methylation (Bird et al. 1985). This ensures these promoters retain an open chromatin structure and that the underlying DNA sequence is accessible to the transcriptional machinery. However, hypermethylation of a CpG island promoter leads to transcriptional silencing of the linked gene.

1.17 How Does DNA Methylation Cause Transcriptional Silencing?

Although the functional effect of DNA methylation is context dependent, the hypermethylation of a CpG island promoter is usually associated with transcriptional silencing of the gene (Jones 2012). However, there are multiple interdependent layers within chromatin that regulate DNA function and gene expression, including nucleosome occupancy and positioning, post-translational histone modifications, and histone variants. Stable transcriptional repression of a gene involves the remodeling of chromatin structure, which renders the underlying promoter DNA sequence inaccessible to the transcriptional machinery (Ng and Bird 1999). In 1997, key experiments showed that DNA methylation directed a time-dependent repression of transcription (Kass et al. 1997). This was demonstrated by showing that naked methylated DNA displays equivalent expression than non-methylated DNA; however, as chromatin is assembled, the methylated template becomes transcriptionally silent and DNA becomes inaccessible. This indicated that transcriptional silencing by methylation involves a hierarchy of epigenetic events. A year later it was shown that methylated CpG dinucleotides serve as docking sites for the recruitment of a range of proteins containing methyl-CpG binding domains including MBD1 and MeCP2 (Fig. 1.1f and g, Chandler et al. 1999; Nan et al. 1998). MeCP2 interacts with a histone deacetylase complex that catalyses the removal of acetyl groups from histones thereby restoring a positive charge to lysine residues and increasing the affinity between histones and negatively charged DNA (Nan et al. 1998). These changes are accompanied by increased nucleosome occupancy and the addition and removal of a range of other histone modifications and histone variants. The chromatin structure of silent and active gene promoters are described in more detail in Chap. 4.

1.18 Epigenomics

Human genetics and epigenetics has been revolutionised by recent technological advances and by international efforts to democratise data. An example of this is the International Human Epigenome Consortium (IHEC)[34]—an international effort to produce reference maps of at least 1000 human epigenomes from different cellular states, including cells from different tissues and diseases. Contributions are from

[34]Further information about IHEC can be accessed at http://ihec-epigenomes.org/.

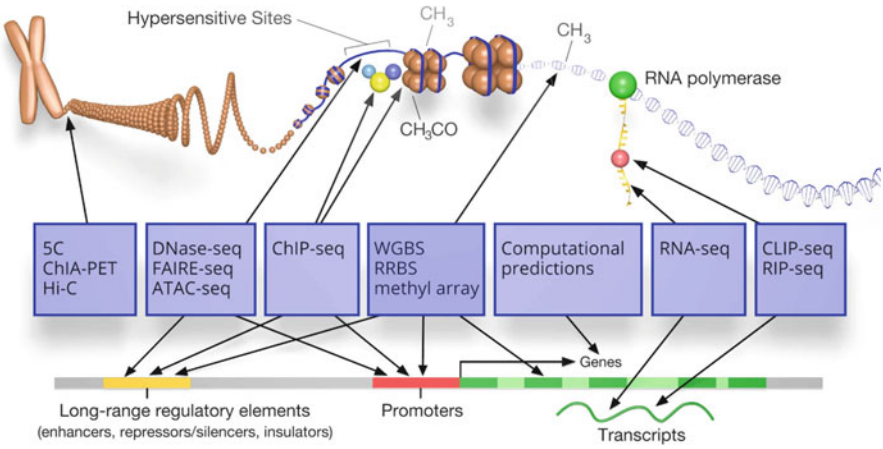

Fig. 1.4 Data and technologies used by the Encyclopedia of DNA Elements (ENCODE) Consortium to discover and annotate functional DNA elements. Taken from the ENCODE website December 2018 (https://www.encodeproject.org/)

leading scientists from the European Union, the USA, Canada, Australia, Japan, South Korea, Hong Kong and Singapore and encompasses the Encyclopedia of DNA elements (ENCODE)[35] project led by the National Institute of Health in the USA. This includes mapping DNA binding proteins (including transcription factors, histone modifications and histone variants), gene transcription, DNA accessibility, RNA binding proteins, DNA methylation, replication timing, three-dimensional chromatin structure and RNA structure. The profiling of genetic and epigenetic characteristics across tissues on this huge scale is possible only with the coordination of global expertise and with the use of massively parallel sequencing and microarray technologies. The types of chromatin modifications and the technologies used to investigate them are summarised in Fig. 1.4.

The value of these reference maps and the amount of information they contain is immense; they allow us understand the epigenetic marks that characterise healthy and disease states. They are a reference source that allows scientists to understand how the different layers of epigenetic information enable different interpretations of the same genome. This also allows identification of epigenetic differences that characterise healthy and diseased states. An example of how this information can be mined to understand epigenetic regulation in a specific region of the human genome is shown in Fig. 1.5. In this example, DNA methylation, histone modifications, gene expression, DNA accessibility, CTCF binding and DNA sequence conservation across a cluster of genes and several nearby enhancers is interpreted to convey some of the principles of epigenetic regulation in the human haemoglobin gene locus.

[35]Further information about ENCODE can be accessed at https://www.encodeproject.org/.

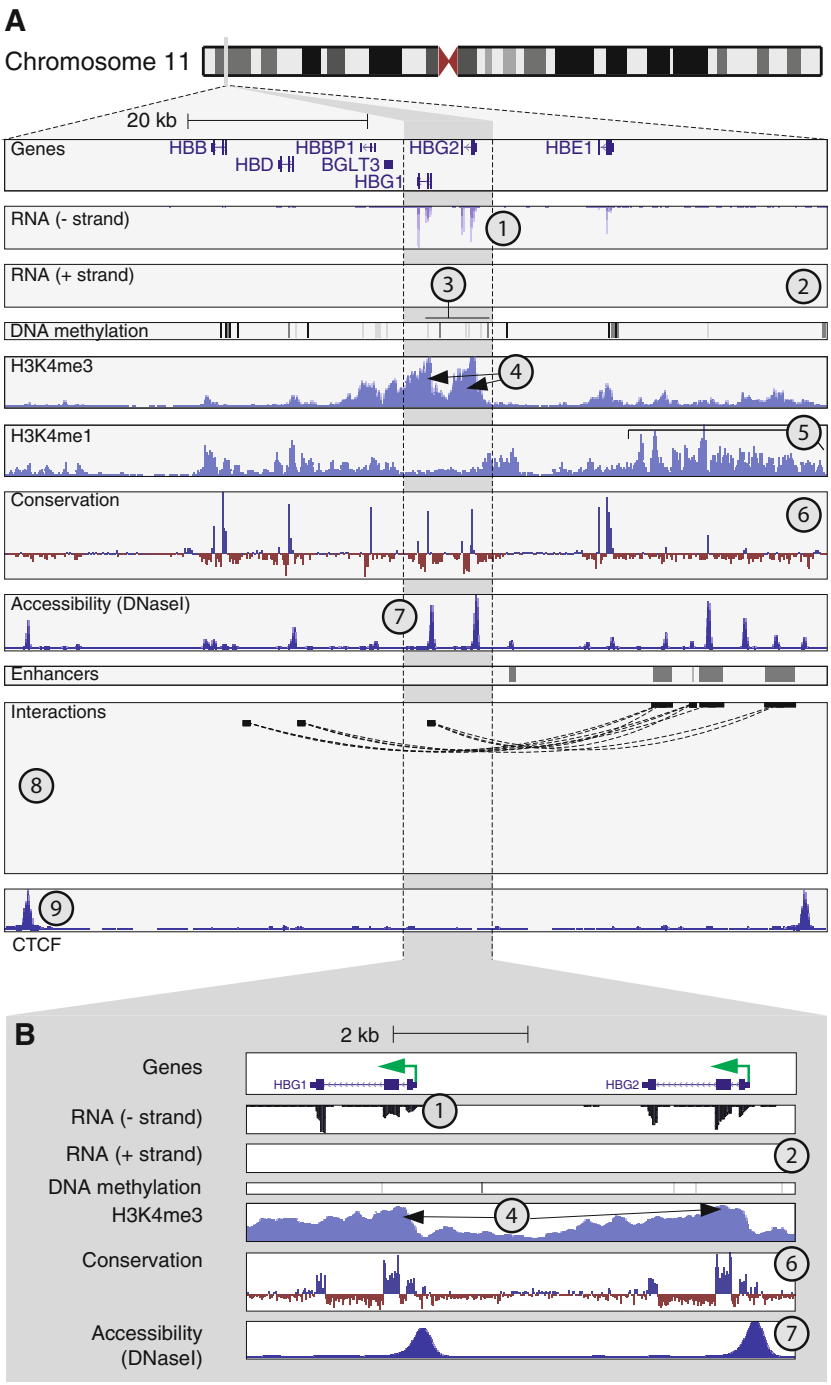

Fig. 1.5 Visualising epigenetic characteristics: the human haemoglobin locus. Data shown are from the myeloid cell line K562 taken from ENCODE and visualised using the UCSC Human Genome Browser (https://genome.ucsc.edu). (**a**) The haemoglobin locus at human chr11:5,223,622-5,315,172 (GRCh37/hg19; February 2009 freeze). Shown are the locations of

1.19 Key Milestones in Genetics and Epigenetics

This introductory chapter has provided an overview of the history of genetic and epigenetic discoveries that have brought us to modern technologies. Below is a timeline of some of the key milestones in genetics and epigenetics with important terms underlined and milestones specific to epigenetics shown in *italicised* text. Due to the pace of discovery the timeline ends with the launch of The International Human Epigenome Consortium in 2010. An abridged version focused on epigenetics discoveries is shown in Fig. 1.6.

Fig. 1.5 (continued) several haemoglobin genes (*HBB*, *HBD*, *HBG1*, *HBG2* and *HBE1*), a pseudogene of *HBB1* known as *HBBP1* and a long non-coding RNA known as *BGLT3*. (**b**) A closer view of the *HBG1* and *HBG2* genes in the region chr11:5,267,292-5,277,485. Green arrows indicate the direction and gene transcription and the transcription start sites. (1) RNA data showing expression of three genes *HBG1*, *HBG2* and *HBE1* from the negative DNA strand. (2) RNA data from the positive DNA strand showing no detectable gene expression, as expected due to the fact that all genes in this region reside on the negative strand. (3) DNA methylation at specific sites across the region determined using Illumina Infinium Human Methylation 450 Bead Array technology. Black bars = methylated sites, dark grey bars = partially methylated sites and light grey bars = unmethylated sites. (4) Levels of trimethylation of lysine 4 on histone H3 (H3K4me3). The abundance of H3K4me3 immediately downstream of the transcription start site (as indicated by black arrows) is typical in actively transcribed genes. (5) Levels of monomethylation of lysine 4 on histone H3 (H3K4me1). H3K4me1 is abundant in regulatory regions known as enhancers and less abundant in actively transcribed genes. Note the abundance of H3K4me1 in the region indicated by black lines. This region is a known regulatory region containing several enhancer sites (as indicated by ENCODE GeneHancer data). (6) DNA sequence conservation across vertebrate species. Peaks above the line indicate sites showing evolutionary conservation of DNA sequence, peaks below the line indicate sites where DNA sequence is not conserved. DNA sequence conservation is greatest within the coding regions of genes (exons, see panel b), but conservation in non-coding DNA may indicate important gene regulatory regions such as enhancers. Note that the exons (coding DNA) of all genes across the region are conserved, as well as some sites within enhancer regions. (7) Sensitivity of DNA to cleavage by the enzyme DNaseI, which is evidence of DNA accessibility and an open chromatin structure. DNaseI accessibility can identify important regulatory regions and is often high at the transcription start sites of highly expression genes and at enhancers. Note that the highest levels of DNA accessibility are the transcription start sites of the two genes that are highly expressed (*HGB1* and *HGB2*, see panel B), but also at several conserved sites within enhancer regions. (8) Sites of physical interaction between regions on the same chromosome (as indicated by ENCODE GeneHancer data). Note that several regions of high DNA accessibility physical interact. This includes the interaction of enhancers with several genes across the region including the highly expressed *HGB1* gene. (9) The entire haemoglobin locus is bookended by two sites enriched for CTCF binding. CTCF is a protein that plays a key role in gene regulation by binding specific sites in the genome and forming boundaries that demarcate chromosome domains. This partitioning allows the independent regulation of different domains within a chromosome

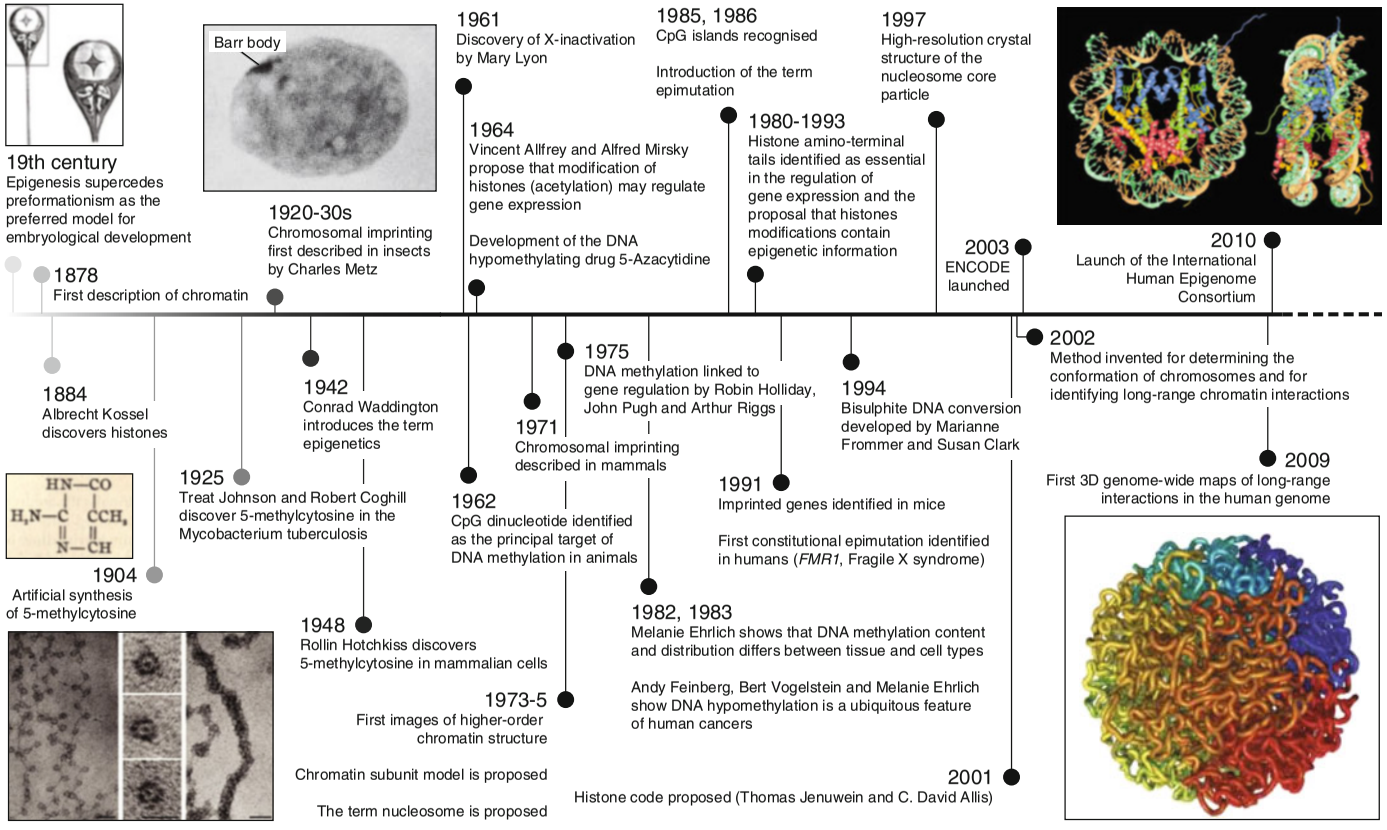

Fig. 1.6 Timeline of epigenetic discoveries. Image sources: Homunculus, (Hartsoeker 1694); 5-methylcytosine, (Wheeler and Johnson 1904); Barr body, (http://glencoe.mcgraw-hill.com/sites/dl/free/0078664276/281029/ccq_ch13_q3.gif), Electron micrographs of chromatin, (Olins and Olins 2003); Crystal structure of the nucleosome core particle to 2.8 Angstrom resolution (Luger et al. 1997); Fractal globule of the three-dimensional structure of chromatin in the nucleus (Lieberman-Aiden et al. 2009)

1.20 Key Discoveries

1751	Pierre Louis de Maupertuis hypothesises equal contribution of both parents to their offspring and a particulate basis of heredity
1800s	*Advances in microscopy results in epigenesis superseding preformationism as the preferred model of embryological development*
1809	Jean-Baptiste Lamarck proposes the model of inheritance of acquired characteristics
1814	Physician and surgeon Joseph Adams classifies hereditary disorders as hereditary and congenital and distinguishes between predisposition and disposition
1859	Charles Darwin Publishes *Origin of Species*
1866	Gregor Mendel describes patterns of particulate inheritance in pea plants and introduces the terms dominant and recessive
1866	Ernst Heinrich Haeckel proposes that the nucleus of a cell transmits its hereditary information
1869	Charles Darwin publishes Variation in Animals and Plants
1871	Friedrich Miescher isolates nuclein (DNA)
1871	Charles Darwin publishes Descent of Man
1873	First descriptions of cell division by Anton Schneider
1875	Francis Galton uses twins to study characteristics
1875	Oscar Hertwig recognises that fertilisation represents the union of the nuclei contributed by male and female germ cells
1876	Francis Galton offers a statistical approach to heredity
1878	*Walther Flemming discovers a substance he calls chromatin and identifies it as a constituent of chromosomes*
1882	Walther Flemming introduces the term mitosis
1882	Eduard Strasburger introduces the terms cytoplasm and nucleoplasm
1883	Edouard van Beneden recognises that the sperm and egg contain fewer chromosomes and that chromosome number is combined after fertilisation
1883	August Weismann makes the distinction between somatic cells and germ cells and proposes that only germ cells carry information that can be transmitted to offspring
1884	Eduard Strasburger introduces the terms prophase, metaphase and anaphase to describe stages of cell division
1884	*Albrecht Kossel discovers histones and protamines*
1885	Hans Driesch clones the first animal (sea urchin) using a process known as embryo splitting
1887	August Weismann deduces the existence of a reduction division (now known as meiosis) in all sexual organisms, which is observed during germ cell maturation by Edouard van Beneden later in the same year

1884–1888	Oscar Hertwig, Eduard Strasburger, Albrecht von Kölliker and August Weismann show the basis of inheritance is contained within the cell nucleus
1888	Heinrich Gottfried von Waldeyer-Hartz introduces the term chromosome
1888	Theodor Boveri first studies chromosomes and suggests they are involved in heredity
1880–1890	Walther Flemming, Eduard Strasburger, Edouard van Beneden and others fully describe cell division, including the equal separation of chromosomes to daughter cells
1890	Theodor Boveri and Jean-Louis-Léon Guignard recognise that there is equal contribution of paternal and maternal chromosomes at fertilisation
1891	Hermann Henking identifies the 'X' body in a proportion of germ cells in insects
1892	August Weismann publishes Das Keimplasma (The Germ Plasm), which provided a framework to study development, evolution and heredity
1894	William Bateson publishes his book Materials for the Study of Variation, which illustrates the significance of discontinuous characteristics (those that do not 'blend') for the understanding of heredity
1899	Clarence McClung finds the 'X body' in locusts and identifies it as a chromosome
1899	First International Congress of Genetics
1899	Richard Altmann renames nuclein as nucleic acid
1900	Hugo de Vries, Erich von Tschermak and Carl Correns rediscover Gregor Mendel's work
1885–1901	Albrecht Kossel isolates and names the constituents of the non-protein component of nucleic acids as adenine, cytosine, guanine, thymine and uracil
1901	T. H. Montgomery recognises the pairing of maternal and paternal chromosomes during meiosis
1902	First description of the theory of chromosomes independently by Theodor Boveri and Walter Sutton
1902	William Bateson, Edith Saunders, William Castle and William Farabee first describe Mendelian inheritance in human disease (alkaptonuria, albinism and brachydactyly)
1902	Clarence McClung proposes that particular chromosomes determine sex
1904	William Bateson and Reginald Punnett first describe gene linkage
1904	Hugo De Vries coins the term mutation in his book Species and Varieties: Their Origin by Mutation
1904	*Henry Wheeler and Treat Johnson artificially synthesise 5-methylcytosine*

1905	William Bateson proposes the term genetics to define the study of heredity
1905	Nettie Stevens first identifies the chromosomal basis for sex determination in flies and beetles, later referred to as the XY sex-determination system
1908	Godfrey Hardy and Wilhelm Weinberg propose the Hardy-Weinberg equilibrium to explain the mathematical relationship of genotype frequencies
1908	Archibald Garrod publishes Inborn Errors of Metabolism and proposes they are determined by genetics
1909	Willhelm Johannsen defines the terms gene, genotype and phenotype
1910	Thomas Morgan describes sex-limited inheritance in drosophila using the white-eyed mutant
1910	Albrecht Kossel is awarded the Nobel Prize in Physiology or Medicine for his research in cell biology, the chemical composition of the cell nucleus, and the isolation and description of nucleic acids
1910	Phoebus Levene introduces the tetranucleotide hypothesis for the structure of DNA
1910–1911	Thomas Morgan shows that chromosomes carry genes and describes the crossing over (recombination) theory of chromosomes
1911	Edmund Wilson uses knowledge of the XY sex-determination system to predict consequences for X-linked inheritance for haemophilia and colour blindness
1913	Alfred Sturtevant creates the first genetic linkage map
1914	Calvin Bridges describes non-disjunction of sex chromosomes, thereby proving the chromosome theory of heredity
1915	Morgan, Sturtevant, Muller and Bridges publish The Mechanism of Mendelian Heredity
1915	Frederick Twort discovers the first bacteriophage
1917	Independently of Twort, Felix Hubert D'Herelle discovers another virus capable of infecting and destroying bacteria and coins the term bacteriophage
1919	Calvin Bridges discovers duplications within a chromosome in drosophila
1923	Calvin Bridges discovers chromosome translocations in drosophila
1925	*Treat Johnson and Robert Coghill discover 5-methylcytosine in the Mycobacterium tuberculosis*
1926	Alfred Sturtevant discovers chromosome inversions in drosophila
1927	Hermann Muller mutates genes using X-rays
1928	Frederick Griffith's bacterial transformation experiments provides initial evidence that DNA contains hereditary information
1920s–1930s	*Chromosomal imprinting described in insects by Charles Metz*
1930	First description of position effect variegation
1941	George Beadle and Edward Tatum show that genes direct synthesis of proteins (one gene, one enzyme hypothesis)

1942	*Conrad Waddington introduces the term epigenetics*
1944	Demonstration that protein-free DNA carries genetic information (Oswald Avery, Colin MacLeod and Maclyn McCarty)
1946	Hermann Muller receives Nobel Prize in Medicine for his work in radiation genetics
1948	*Rollin Hotchkiss discovers 5-methylcytosine in mammalian cells*
1948	Hermann Muller introduces the term dosage compensation
1949	James Neel shows sickle cell anaemia is inherited in a Mendelian autosomal recessive manner
1950	Erwin Chargaff shows the proportion of nucleobases differs in different species and that the amount of adenine equals the amount of thymine and the amount of cytosine equals the amount of guanine (A = T and G = C)
1937–1951	X-ray diffraction of DNA reveals a regular repeating periodic structure (William Astbury, Maurice Wilkins, Rosalind Franklin and others)
1952	Fred Sanger determines the sequence of amino acids in the protein insulin
1952	Alfred Hershey and Martha Chase show that it is the DNA from bacteriophage that enters the host bacterium, thereby dispelling any remaining doubt that DNA contains the heredity information
1953	The double helix structure of DNA is elucidated by Francis Crick and James Watson, using X-ray diffraction data from Rosalind Franklin
1955	Definition of the human karyotype by Joe Hin Tjio
1956	Arthur Kornberg crystallises DNA polymerase, the enzyme required for synthesising DNA
1957	Francis Crick proposes the central dogma of molecular biology
1958	Semi-conservative mechanism of DNA replication elucidated by Matthew Meselson and Franklin Stahl
1958	George Beadle and Edward Tatum receive the Nobel Prize in Medicine 'for their discovery that genes act by regulating definite chemical events' with the other half to Joshua Lederberg 'for his discoveries concerning genetic recombination and the organization of the genetic material of bacteria'
1958	Fred Sanger receives Nobel Prize for Chemistry for his work on the structure of proteins, especially that of insulin
1959	Arthur Pardee, Francois Jacob and Jacques Monod publish their study of the lactose operon in *Escherichia coli*
1959	Trisomy 21 identified as the cause of Down syndrome by Jerome Lejeune, Martha Gautier and Raymond Turpin
1959	The Nobel Prize in Physiology or Medicine is awarded jointly to Severo Ochoa and Arthur Kornberg 'for their discovery of the mechanisms in the biological synthesis of ribonucleic acid and deoxyribonucleic acid'

1960	*Helen Crouse introduces the term imprint to describe the marking of chromosomes by parent-of-origin*
1960s	*Fractionation of histones by E. W. Johns and others*
1961	Messenger RNA (mRNA) identified as the intermediate that carries information from DNA in the nucleus to the cytoplasm where protein is made (Sydney Brenner, Francois Jacob and Matthew Meselson)
1961	Triplets of DNA bases proposed to code for one of the 20 amino acids by Sydney Brenner and Francis Crick
1961	*Discovery of X-inactivation by Mary Lyon*
1962	First patient diagnosed with a mitochondrial disease
1962	The Nobel Prize in Physiology or Medicine is awarded jointly to Francis Crick, James Watson and Maurice Wilkins 'for their discoveries concerning the molecular structure of nucleic acids and its significance for information transfer in living material'
1963	Mitochondrial DNA identified
1964	*Vincent Allfrey, Robert Faulkner and Alfred Mirsky propose that modification of histones (acetylation and methylation) can regulate gene expression*
1964	*Development of the DNA hypomethylating drug 5-azacytidine*
1964	Robin Holliday describes concept of gene conversion and the Holliday junction
1965	The Nobel Prize in Physiology or Medicine is awarded jointly to François Jacob, André Lwoff and Jacques Monod 'for their discoveries concerning genetic control of enzyme and virus synthesis'
1961–1966	Marshall Nirenberg cracks the genetic code
1967	Mary Weiss and Howard Green map human genes using somatic cell hybridisation
1968	First therapeutic abortion performed entirely on the basis of a genetic test (Down syndrome)
1968	The Nobel Prize in Physiology or Medicine is awarded jointly to Robert Holley, Har Gobind Khorana and Marshall Nirenberg 'for their interpretation of the genetic code and its function in protein synthesis'
1969	The Nobel Prize in Physiology or Medicine is awarded jointly to Max Delbrück, Alfred Hershey and Salvador Luria 'for their discoveries concerning the replication mechanism and the genetic structure of viruses'
1970	Ron Laskey and John Gurdon show that a somatic cell nucleus contains all the necessary information to direct embryogenesis when introduced into an enucleated egg. This represented definitive proof that cell differentiation and embryonic development was not driven by the loss of genetic material
1970	Howard Temin and David Baltimore discovery reverse transcriptase, an enzyme that makes DNA from RNA

1971	Alfred Knudson's proposes his two-hit hypothesis of cancer
1971	*Chromosomal imprinting described in mammals (X-chromosome in the kangaroo)*
1971	*Eduardo Scarano proposes that 5-methylcytosine can deamination to generate thymine*
1972	First recombinant DNA molecules
1973	*First images of higher-order chromatin structure by Ada and Donald Olins and Christopher Woodcock*
1974	*The chromatin subunit model is proposed by Ada and Donald Olins, Roger Kornberg and Jean Thomas*
1975	*The term nucleosome is proposed by Oudet, Gross-Bellard and Chambon*
1975	*Robin Holliday, John Pugh and Arthur Riggs propose that DNA methylation controls gene expression, is heritable following cell division and is an epigenetic mechanism that explains X-chromosome inactivation*
1977	First DNA sequencing methods developed by Walter Gilbert, Alan Maxam and Fred Sanger
1977	The enterobacteriaphage phiX174 becomes the first genome sequenced (Fred Sanger)
1977	Discovery of introns and the concept of splicing
1978	*Methylation sensitive restriction endonucleases first used to detect DNA methylation*
1978	David Botstein uses restriction enzymes to map human genes and identify genetic differences between individuals
1980	Concept of positional cloning of genes proposed by Ron Davis and David Botstein
1980	The Nobel Prize in Chemistry is divided, one half awarded to Paul Berg 'for his fundamental studies of the biochemistry of nucleic acids, with particular regard to recombinant-DNA' and the other half jointly to Walter Gilbert and Frederick Sanger 'for their contributions concerning the determination of base sequences in nucleic acids'
1981	Sequencing of the mitochondrial genome
1982	*Melanie Ehrlich shows that DNA methylation content and distribution differs amongst different tissue and cell types*
1983	Invention of PCR by Kary Mullis
1983	*Discovery of the first enhancer within the immunoglobulin heavy chain locus*
1983	*Andy Feinberg, Bert Vogelstein and Melanie Ehrlich demonstrate that DNA hypomethylation is a ubiquitous feature of human cancers*
1984	*Solution of the crystal structure of the nucleosome core particle to 7 Angstrom resolution*
1984	Alec Jeffreys develops DNA fingerprinting
1984	*Demonstration that both maternal and paternal genomes are essential for normal mammalian embryonic development*

1985	*CpG islands first recognised*
1985	*Discovery of an imprinted region*
1986	*Penny Jeggo and Robin Holliday introduce the term epimutation*
1986	First gene identified by positional cloning (chronic granulomatous disease)
1986	Development of the first automated DNA sequencer using fluorescent dyes (Leroy Hood)
1986	*First descriptions of epigenetic inactivation of a tumour suppressor gene (RB1 in retinoblastoma)*
1986–1990	Human genome project launched
1980–1991	*Experiments showing that histone amino-terminal tails are essential in the regulation of gene expression*
1987	*Robin Holliday proposes the inheritance of epigenetic defects*
1990	BRCA1 gene identified and linked to familial breast and ovarian cancer (Mary-Claire King and Mark Skolnick)
1991	*Abnormal hypermethylation of FMR1 recognised in Fragile X syndrome, representing the first description of a constitutional epimutation in humans*
1991	*Imprinted genes first identified in mice by Denise Barlow, Anne Ferguson-Smith and others*
1991	*Parental conflict theory proposed to explain the evolution of genomic imprinting*
1992	*Mechanistic basis of the heritability of DNA methylation shown when it's discovered that Dnmt1 is targeted to sites of DNA replication*
1993	Positional cloning of the gene mutated in Huntington's disease by Nancy Wexler and James Gusella
1993	*Bryan Turner proposes that post-translational modifications of the amino-terminal tails of histones encodes epigenetic information*
1993	The Nobel Prize in Chemistry is awarded 'for contributions to the developments of methods within DNA-based chemistry' jointly with one half to Kary Mullis 'for his invention of the polymerase chain reaction (PCR) method' and with one half to Michael Smith 'for his fundamental contributions to the establishment of oligonucleotide-based, site-directed mutagenesis and its development for protein studies'
1994	*Bisulphite DNA conversion method developed by Marianne Frommer and Susan Clark*
1994	Généthon publishes a map of the human genome based on microsatellite markers
1995	Development of DNA microarrays
1995	The genome of Haemophilus influenzae is sequenced (J. Craig Venter)
1995	Map of the human genome based on Sequence Tagged Sites (STSs)
1995	The Nobel Prize in Physiology or Medicine is awarded jointly to Edward Lewis, Christiane Nüsslein-Volhard and Eric Wieschaus 'for their discoveries concerning the genetic control of early embryonic development'

1996	Sequencing of the genome of Saccharomyces cerevisiae (baker's yeast)
1996	Dolly the sheep is cloned by Keith Campbell and Ian Wilmut
1997	*The crystal structure of the nucleosome core particle to 2.8 Angstrom resolution shows the histone amino terminal tails modified by post-translational modifications*
1997	*The relationship between DNA methylation, chromatin structure and gene silencing becomes clearer when it is shown that methylated DNA only becomes transcriptionally inactive after being packaged into chromatin*
1998	*Recognition that transposable and viral elements within the mammalian genome are hypermethylated*
1998	*Discovery that DNA methylation increases with age*
1999	*Recognition that DNA methyltransferase preferentially recognises hemimethylated DNA*
2000	*CTCF identified as a key mediator of imprinting at the H19/Igf2 locus*
2001	First draft of the human genome published
2001	*The histone code is proposed by Thomas Jenuwein and C. David Allis*
2002	*Development of method for capturing chromosome conformation and identifying long-range chromatin interactions*
2003	*The National Human Genome Research Institute launches ENCODE (Encyclopedia of DNA Elements) to identify all functional elements in the human genome*
2004	*First description of a constitutional epimutation predisposing to cancer by Robyn Ward and colleagues (MLH1 in Lynch syndrome)*
2009	*First description of the three-dimensional architecture of a human genome*
2010	*The International Human Epigenome Consortium is founded to lead efforts in understanding the human epigenome and to generate reference databases of healthy and disease-related cell types*

References

Allfrey VG et al (1964) Acetylation and methylation of histones and their possible role in the regulation of RNA synthesis. Proc Natl Acad Sci U S A 51:786–794

Altmann R (1889) Ueber Nucleinsäuren. Arch f Anatomie u Physiol:524–536

Austin CR, Amoroso EC (1957) Sex chromatin in early cat embryos. Exp Cell Res 13(2):419–421

Avery O (1941) Obituary: Frederick Griffith. Lancet:588–589

Avery OT et al (1944) Studies on the chemical nature of the substance inducing transformation of pneumococcal types: induction of transformation by a desoxyribonucleic acid fraction isolated from Pneumococcus type Iii. J Exp Med 79(2):137–158

Ballabio A, Willard HF (1992) Mammalian X-chromosome inactivation and the XIST gene. Curr Opin Genet Dev 2(3):439–447
Banerji J et al (1983) A lymphocyte-specific cellular enhancer is located downstream of the joining region in immunoglobulin heavy chain genes. Cell 33(3):729–740
Barlow DP, Bartolomei MS (2014) Genomic imprinting in mammals. Cold Spring Harb Perspect Biol 6(2):a018382
Barlow DP et al (1991) The mouse insulin-like growth factor type-2 receptor is imprinted and closely linked to the Tme locus. Nature 349(6304):84–87
Barr ML, Bertram EG (1949) A morphological distinction between neurones of the male and female, and the behaviour of the nucleolar satellite during accelerated nucleoprotein synthesis. Nature 163(4148):676
Bartolomei MS et al (1991) Parental imprinting of the mouse H19 gene. Nature 351(6322):153–155
Barton SC et al (1984) Role of paternal and maternal genomes in mouse development. Nature 311 (5984):374–376
Baylin SB, Jones PA (2016) Epigenetic determinants of cancer. Cold Spring Harb Perspect Biol 8 (9):a019505
Beadle GW, Tatum EL (1941) Genetic control of biochemical reactions in neurospora. Proc Natl Acad Sci U S A 27(11):499–506
Berget SM, Sharp PA (1977) A spliced sequence at the 5'-terminus of adenovirus late mRNA. Brookhaven Symp Biol (29):332–344
Berk AJ, Sharp PA (1977) Sizing and mapping of early adenovirus mRNAs by gel electrophoresis of S1 endonuclease-digested hybrids. Cell 12(3):721–732
Berletch JB et al (2011) Genes that escape from X inactivation. Hum Genet 130(2):237–245
Bestor T et al (1988) Cloning and sequencing of a cDNA encoding DNA methyltransferase of mouse cells. The carboxyl-terminal domain of the mammalian enzymes is related to bacterial restriction methyltransferases. J Mol Biol 203(4):971–983
Bird A et al (1985) A fraction of the mouse genome that is derived from islands of nonmethylated, CpG-rich DNA. Cell 40(1):91–99
Blewitt ME et al (2008) SmcHD1, containing a structural-maintenance-of-chromosomes hinge domain, has a critical role in X inactivation. Nat Genet 40(5):663–669
Borsani G et al (1991) Characterization of a murine gene expressed from the inactive X chromosome. Nature 351(6324):325–329
Brockdorff N et al (1991) Conservation of position and exclusive expression of mouse Xist from the inactive X chromosome. Nature 351(6324):329–331
Brown CJ et al (1991a) A gene from the region of the human X inactivation centre is expressed exclusively from the inactive X chromosome. Nature 349(6304):38–44
Brown CJ et al (1991b) Localization of the X inactivation centre on the human X chromosome in Xq13. Nature 349(6304):82–84
Brush SG (1978) Nettie M. Stevens and the discovery of sex determination by chromosomes. Isis 69(247):163–172
Burch PR, Burwell RG (1963) Lyonisation of the X chromosome. Lancet 2(7319):1229
Burian RM (2004) Molecular epigenesis, molecular pleiotropy, and molecular gene definitions. Hist Philos Life Sci 26(1):59–80
Chandler SP et al (1999) The methyl-CpG binding transcriptional repressor MeCP2 stably associates with nucleosomal DNA. Biochemistry 38(22):7008–7018
Chargaff E (1951) Structure and function of nucleic acids as cell constituents. Fed Proc 10 (3):654–659
Chargaff E et al (1949) The composition of the desoxypentose nucleic acids of thymus and spleen. J Biol Chem 177(1):405–416
Chow LT et al (1977) A map of cytoplasmic RNA transcripts from lytic adenovirus type 2, determined by electron microscopy of RNA: DNA hybrids. Cell 11(4):819–836
Chu C et al (2015) Systematic discovery of Xist RNA binding proteins. Cell 161(2):404–416
Clark RJ, Felsenfeld G (1971) Structure of chromatin. Nat New Biol 229(4):101–106

Cooper DW et al (1971) Phosphoglycerate kinase polymorphism in kangaroos provides further evidence for paternal X inactivation. Nat New Biol 230(13):155–157
Crick FH (1958) On protein synthesis. Symp Soc Exp Biol 12:138–163
Crick F (1970) Central dogma of molecular biology. Nature 227(5258):561–563
Crouse HV (1960) The controlling element in sex chromosome behavior in Sciara. Genetics 45 (10):1429–1443
D'Anna JA Jr, Isenberg I (1974) A histone cross-complexing pattern. Biochemistry 13 (24):4992–4997
Darwin C (1859) On the origin of species by means of natural selection: or the preservation of favoured races in the struggle for life. D. Appleton, New York
Darwin C (1872) The descent of man, and selection in relation to sex. D. Appleton, New York
de Maupertuis PLM (1751) Essai de Cosmologie. Sorli
de Monet de Lamarck JB (1802) Recherches sur l'organisation des corps vivants et particulièrement sur son origine, sur la cause de son développement et des progrès de sa composition ... Précédé du discours d'ouverture du cours de zoologie donné dans le Muséum d'histoire naturelle, l'an X de la République. Par J.-B. Lamarck: L'auteur
de Monet de Lamarck JBPA (1809) Philosophie zoologique: ou Exposition des considérations relative à l'histoire naturelle des animaux. Dentu, Paris
Deans C, Maggert KA (2015) What do you mean, "epigenetic". Genetics 199(4):887–896
DeChiara TM et al (1991) Parental imprinting of the mouse insulin-like growth factor II gene. Cell 64(4):849–859
Doskocil J, Sorm F (1962) Distribution of 5-methylcytosine in pyrimidine sequences of deoxyribonucleic acids. Biochim Biophys Acta 55:953–959
Dupont C et al (2009) Epigenetics: definition, mechanisms and clinical perspective. Semin Reprod Med 27(5):351–357
Ehrlich M et al (1982) Amount and distribution of 5-methylcytosine in human DNA from different types of tissues of cells. Nucleic Acids Res 10(8):2709–2721
Feinberg AP, Tycko B (2004) The history of cancer epigenetics. Nat Rev Cancer 4(2):143–153
Feinberg AP, Vogelstein B (1983) Hypomethylation distinguishes genes of some human cancers from their normal counterparts. Nature 301(5895):89–92
Feinberg AP et al (1988) Reduced genomic 5-methylcytosine content in human colonic neoplasia. Cancer Res 48(5):1159–1161
Ferguson-Smith AC et al (1991) Embryological and molecular investigations of parental imprinting on mouse chromosome 7. Nature 351(6328):667–670
Friedrich M (1874) Die Spermatozoen einiger Wirbeltiere. Ein Beitrag zur Histochemie. Verhandlungen der naturforschenden Gesellschaft in Basel VI, pp 138–208
Fujimoto D et al (1965) On the nature of the deoxyribonucleic acid methylases. Biological evidence for the multiple nature of the enzymes. Biochemistry 4(12):2849–2855
Gall JG (1963) Kinetics of deoxyribonuclease action on chromosomes. Nature 198:36–38
Gama-Sosa MA et al (1983) The 5-methylcytosine content of DNA from human tumors. Nucleic Acids Res 11(19):6883–6894
Gardiner-Garden M, Frommer M (1987) CpG islands in vertebrate genomes. J Mol Biol 196 (2):261–282
Gillies SD et al (1983) A tissue-specific transcription enhancer element is located in the major intron of a rearranged immunoglobulin heavy chain gene. Cell 33(3):717–728
Goelz SE et al (1985) Hypomethylation of DNA from benign and malignant human colon neoplasms. Science 228(4696):187–190
Gold M et al (1963) The enzymatic methylation of RNA and DNA, Ii. On the species specificity of the methylation enzymes. Proc Natl Acad Sci U S A 50(1):164–169
Greally JM (2018) A user's guide to the ambiguous word 'epigenetics'. Nat Rev Mol Cell Biol 19 (4):207–208
Greger V et al (1989) Epigenetic changes may contribute to the formation and spontaneous regression of retinoblastoma. Hum Genet 83(2):155–158

Griffith F (1928) The significance of pneumococcal types. J Hyg (Lond) 27(2):113–159
Griffiths PE, Stotz K (2006) Genes in the postgenomic era. Theor Med Bioeth 27(6):499–521
Haig D (2004) The (dual) origin of epigenetics. Cold Spring Harb Symp Quant Biol 69:67–70
Hall RH (1971) The modified nucleosides in nucleic acids. Columbia University Press, New York
Hartsoeker N (1694) Essay de dioptrique. Kessinger, Whitefish
Hermann A et al (2004) The Dnmt1 DNA-(cytosine-C5)-methyltransferase methylates DNA processively with high preference for hemimethylated target sites. J Biol Chem 279 (46):48350–48359
Holliday R, Pugh JE (1975) DNA modification mechanisms and gene activity during development. Science 187(4173):226–232
Hotchkiss RD (1948) The quantitative separation of purines, pyrimidines, and nucleosides by paper chromatography. J Biol Chem 175(1):315–332
Itzhaki RF (1971) Studies on the accessibility of deoxyribonucleic acid in deoxyribonucleoprotein to cationic molecules. Biochem J 122(4):583–592
Jabbari K, Bernardi G (2004) Cytosine methylation and CpG, TpG (CpA) and TpA frequencies. Gene 333:143–149
Johns EW (1969) The histones, their interactions with DNA and some aspects of gene control. Churchill, London
Johnson TB, Coghill RD (1925) Researches on pyrimidines. C111. The discovery of 5-methyl-cytosine in tuberculinic acid, the nucleic acid of the tubercle bacillus. J Am Chem Soc 47 (11):2838–2844
Jones PA (2012) Functions of DNA methylation: islands, start sites, gene bodies and beyond. Nat Rev Genet 13(7):484–492. https://doi.org/10.1038/nrg3230
Jones PA, Taylor SM (1980) Cellular differentiation, cytidine analogs and DNA methylation. Cell 20(1):85–93
Judson HF (1979) The eighth day of creation: makers of the revolution in biology, n. edn. Cold Spring Harbor Laboratory Press, New York
Kass SU et al (1997) DNA methylation directs a time-dependent repression of transcription initiation. Curr Biol 7(3):157–165
Keller EF, Harel D (2007) Beyond the gene. PLoS One 2(11):e1231
Kingsland SE (2007) Maintaining continuity through a scientific revolution: a rereading of E. B. Wilson and T. H. Morgan on sex determination and Mendelism. Isis 98(3):468–488
Kornberg RD (1974) Chromatin structure: a repeating unit of histones and DNA. Science 184 (4139):868–871
Kornberg RD, Thomas JO (1974) Chromatin structure; oligomers of the histones. Science 184 (4139):865–868
Kornberg A et al (1959) Enzymatic synthesis of deoxyribonucleic acid. Influence of bacteriophage T2 on the synthetic pathway in host cells. Proc Natl Acad Sci U S A 45(6):772–785
Kossel A (1884) Ueber einen peptonartigen Bestandtheil des Zellkerns. Zschr Physiol Chem 8:511
Leonhardt H et al (1992) A targeting sequence directs DNA methyltransferase to sites of DNA replication in mammalian nuclei. Cell 71(5):865–873
Levene PA, Mandel J-A (1908) Über die darstellung und analyse einiger nucleinsäuren. XIII. Über ein verfahren zur gewinnung der purinbasen. Biochem Z 10:215–220
Lieberman-Aiden E et al (2009) Comprehensive mapping of long-range interactions reveals folding principles of the human genome. Science 326(5950):289–293
Lu Z et al (2017) Mechanistic insights in X-chromosome inactivation. Philos Trans R Soc Lond Ser B Biol Sci 372(1733):20160356
Luger K et al (1997) Crystal structure of the nucleosome core particle at 2.8 A resolution. Nature 389(6648):251–260
Lyon MF (1961) Gene action in the X-chromosome of the mouse (Mus musculus L.). Nature 190:372–373
Lyon MF (1962) Sex chromatin and gene action in the mammalian X-chromosome. Am J Hum Genet 14:135–148

Lyon MF (1963) Lyonisation of the X Chromosome. Lancet 2(7317):1120–1121
Mandel JL, Chambon P (1979) DNA methylation: organ specific variations in the methylation pattern within and around ovalbumin and other chicken genes. Nucleic Acids Res 7 (8):2081–2103
McCarty M (2003) Discovering genes are made of DNA. Nature 421(6921):406
McClung CE (1902) The accessory chromosome—Sex determinant? Biol Bull 3
McGhee JD, Ginder GD (1979) Specific DNA methylation sites in the vicinity of the chicken beta-globin genes. Nature 280(5721):419–420
McGrath J, Solter D (1984) Completion of mouse embryogenesis requires both the maternal and paternal genomes. Cell 37(1):179–183
Mendel G (1866) Versuche über Plflanzenhybriden. Verhandlungen des naturforschenden Vereines in Brünn, Bd. IV für das Jahr. Abhandlungen, 3–47
Mercola M et al (1983) Transcriptional enhancer elements in the mouse immunoglobulin heavy chain locus. Science 221(4611):663–665
Metz CW (1938) Chromosome behavior, inheritance and sex determination in Sciara. Am Nat 72 (743):485–520
Migeon BR (2017) Choosing the active X: the human version of X inactivation. Trends Genet 33 (12):899–909
Migeon BR et al (1979) Stability of the "two active X" phenotype in triploid somatic cells. Cell 18 (3):637–641
Migeon BR et al (2008) X inactivation in triploidy and trisomy: the search for autosomal transfactors that choose the active X. Eur J Hum Genet 16(2):153–162
Minajigi A et al (2015) Chromosomes. A comprehensive Xist interactome reveals cohesin repulsion and an RNA-directed chromosome conformation. Science 349(6245):aab2276
Mirsky AE (1971) The structure of chromatin. Proc Natl Acad Sci U S A 68(12):2945–2948
Mohandas T et al (1981) Reactivation of an inactive human X chromosome: evidence for X inactivation by DNA methylation. Science 211(4480):393–396
Moore T, Haig D (1991) Genomic imprinting in mammalian development: a parental tug-of-war. Trends Genet 7(2):45–49
Morgan TH (2018) The theory of the gene, 2nd edn. Creative Media Partners, LLC, Sacramento, CA
Moss L (2003) What genes can't do. MIT Press, Cambridge, MA
Nan X et al (1998) Transcriptional repression by the methyl-CpG-binding protein MeCP2 involves a histone deacetylase complex. Nature 393(6683):386–389
Nanney DL (1958) Epigenetic control systems. Proc Natl Acad Sci U S A 44(7):712–717
Ng HH, Bird A (1999) DNA methylation and chromatin modification. Curr Opin Genet Dev 9 (2):158–163
Ohno S, Hauschka TS (1960) Allocycly of the X-chromosome in tumors and normal tissues. Cancer Res 20:541–545
Olins DE, Olins AL (1973) Spheroid chromatin units (v bodies). J Cell Biol 59:A252
Olins DE, Olins AL (1974) Spheroid chromatin units (v bodies). Science 183(4122):330–332
Olins DE, Olins AL (2003) Chromatin history: our view from the bridge. Nat Rev Mol Cell Biol 4 (10):809–814
Oudet P et al (1975) Electron microscopic and biochemical evidence that chromatin structure is a repeating unit. Cell 4(4):281–300
Pan Q et al (2008) Deep surveying of alternative splicing complexity in the human transcriptome by high-throughput sequencing. Nat Genet 40(12):1413–1415
Park WW (1957) The occurrence of sex chromatin in early human and macaque embryos. J Anat 91 (3):369–373
Pennacchio LA et al (2013) Enhancers: five essential questions. Nat Rev Genet 14(4):288–295
Pesole G (2008) What is a gene? An updated operational definition. Gene 417(1–2):1–4
Plenge RM et al (1997) A promoter mutation in the XIST gene in two unrelated families with skewed X-chromosome inactivation. Nat Genet 17(3):353–356

Portin P, Wilkins A (2017) The evolving definition of the term "Gene". Genetics 205(4):1353–1364
Ptashne M (2007) On the use of the word 'epigenetic'. Curr Biol 17(7):R233–R236
Razin A, Riggs AD (1980) DNA methylation and gene function. Science 210(4470):604–610
Reik W, Walter J (2001) Genomic imprinting: parental influence on the genome. Nat Rev Genet 2 (1):21–32
Riggs AD (1975) X inactivation, differentiation, and DNA methylation. Cytogenet Cell Genet 14 (1):9–25
Riggs AD (1988) X-inactivation, differentiation, and DNA methylation. Curr Content 31(1):15
Roark DE et al (1974) A two-subunit histone complex from calf thymus. Biochem Biophys Res Commun 59(2):542–547
Sakai T et al (1991) Allele-specific hypermethylation of the retinoblastoma tumor-suppressor gene. Am J Hum Genet 48(5):880–888
Saxonov S et al (2006) A genome-wide analysis of CpG dinucleotides in the human genome distinguishes two distinct classes of promoters. Proc Natl Acad Sci U S A 103(5):1412–1417
Scarano E (1971) The control of gene function in cell differentiation and in embryogenesis. Adv Cytopharmacol 1:13–24
Scherrer K, Jost J (2007) Gene and genon concept: coding versus regulation. A conceptual and information-theoretic analysis of genetic storage and expression in the light of modern molecular biology. Theory Biosci 126(2–3):65–113
Searle AG, Beechey CV (1978) Complementation studies with mouse translocations. Cytogenet Cell Genet 20(1–6):282–303
Shen CK, Maniatis T (1980) Tissue-specific DNA methylation in a cluster of rabbit beta-like globin genes. Proc Natl Acad Sci U S A 77(11):6634–6638
Stadler PF et al (2009) Defining genes: a computational framework. Theory Biosci 128(3):165–170
Stubbe H (1972) History of genetics: from prehistoric times to the rediscovery of Mendel's laws. MIT Press, Cambridge, MA
Surani MA, Barton SC (1983) Development of gynogenetic eggs in the mouse: implications for parthenogenetic embryos. Science 222(4627):1034–1036
Szmulewicz MN et al (1998) Effects of Alu insertions on gene function. Electrophoresis 19 (8–9):1260–1264
Szyf M et al (1985) Cell cycle-dependent regulation of eukaryotic DNA methylase level. J Biol Chem 260(15):8653–8656
Taylor SM, Jones PA (1979) Multiple new phenotypes induced in 10T1/2 and 3T3 cells treated with 5-azacytidine. Cell 17(4):771–779
van der Ploeg LH, Flavell RA (1980) DNA methylation in the human gamma delta beta-globin locus in erythroid and nonerythroid tissues. Cell 19(4):947–958
Vanyushin BF et al (1970) Rare bases in animal DNA. Nature 225(5236):948–949
Varmuza S, Mann M (1994) Genomic imprinting—defusing the ovarian time bomb. Trends Genet 10(4):118–123
Waddington CH (1939) An introduction to modern genetics. Macmillan, New York
Waddington CH (1940) Organisers and genes. Cambridge University Press, Cambridge
Waddington CH (1942) The epigenotype. Endeavour 1:18
Waddington CH (2012) The epigenotype. 1942. Int J Epidemiol 41(1):10–13
Walker CL et al (1991) The Barr body is a looped X chromosome formed by telomere association. Proc Natl Acad Sci U S A 88(14):6191–6195
Watson JD, Crick FH (1953a) Genetical implications of the structure of deoxyribonucleic acid. Nature 171(4361):964–967
Watson JD, Crick FH (1953b) Molecular structure of nucleic acids; a structure for deoxyribose nucleic acid. Nature 171(4356):737–738
Weaver DD, Gartler SM (1975) Evidence for two active X chromosomes in a human XXY triploid. Humangenetik 28(1):39–42
Welshons WJ, Russell LB (1959) The Y-chromosome as the bearer of male determining factors in the mouse. Proc Natl Acad Sci U S A 45(4):560–566

Wheeler HL, Johnson TB (1904) 5-methylcytosine. Am Chem J 31:591
Wilkins MH et al (1953) Molecular structure of deoxypentose nucleic acids. Nature 171 (4356):738–740
Wilson EB (1911) The sex chromosomes. Archiv fu¨r mikroskopische Anatomie Pt. 2(77):249–271
Woodcock C (1973) Ultrastructure of inactive chromatin. J Cell Biol 59:A368
Woodcock C et al (1976) Structural repeating units in chromatin. I. Evidence for their general occurrence. Exp Cell Res 97:101–110
Wyatt GR (1951) The purine and pyrimidine composition of deoxypentose nucleic acids. Biochem J 48(5):584–590
Zou X et al (2012) Recognition of methylated DNA through methyl-CpG binding domain proteins. Nucleic Acids Res 40(6):2747–2758
Zubay G, Doty P (1959) The isolation and properties of deoxyribonucleoprotein particles containing single nucleic acid molecules. J Mol Biol 1:1–20

The DNA Methylation Machinery

2

The Readers, Writers and Erasers of Methylcytosine and Their Regulation

Mark Morris

Abstract

Cytosine methylation is an important epigenetic signalling mechanism that regulates gene expression. The presence of 5-methylcytosine (5-mC) at promoters is associated with gene silencing. In this chapter, we will describe our current understanding of

1. The mechanisms and molecules involved in the addition, and maintenance, of methyl groups to cytosine bases (5-mC writers)
2. The mechanism by which these 5-mC modifications are read resulting in changes in gene expression
3. The molecules and mechanisms involved in removing methyl groups from cytosine bases (5-mC erasers)

Keywords

CpG · Methylcytosine · Methyl-CpG-binding domain proteins · DNA methyltransferase · TET enzymes

2.1 DNA Methylation/Methylcytosine

In eukaryotic cells, the addition of methyl groups (CH_3) to DNA occurs almost exclusively on cytosine bases, more specifically at cytosines that are found directly 5′ to a guanine forming a CpG dinucleotide (Fig. 2.1a). Methylation occurs at the 5th atom of the 6-atom ring (Fig. 2.1b) resulting in the formation of 5-methylcytosine (5-mC). On double-stranded DNA, a CpG dinucleotide on one DNA strand will bind

M. Morris (✉)
University of Wolverhampton, Wolverhampton, UK
e-mail: M.R.Morris2@wlv.ac.uk

L. B. Hesson, A. L. Pritchard (eds.), *Clinical Epigenetics*,
https://doi.org/10.1007/978-981-13-8958-0_2

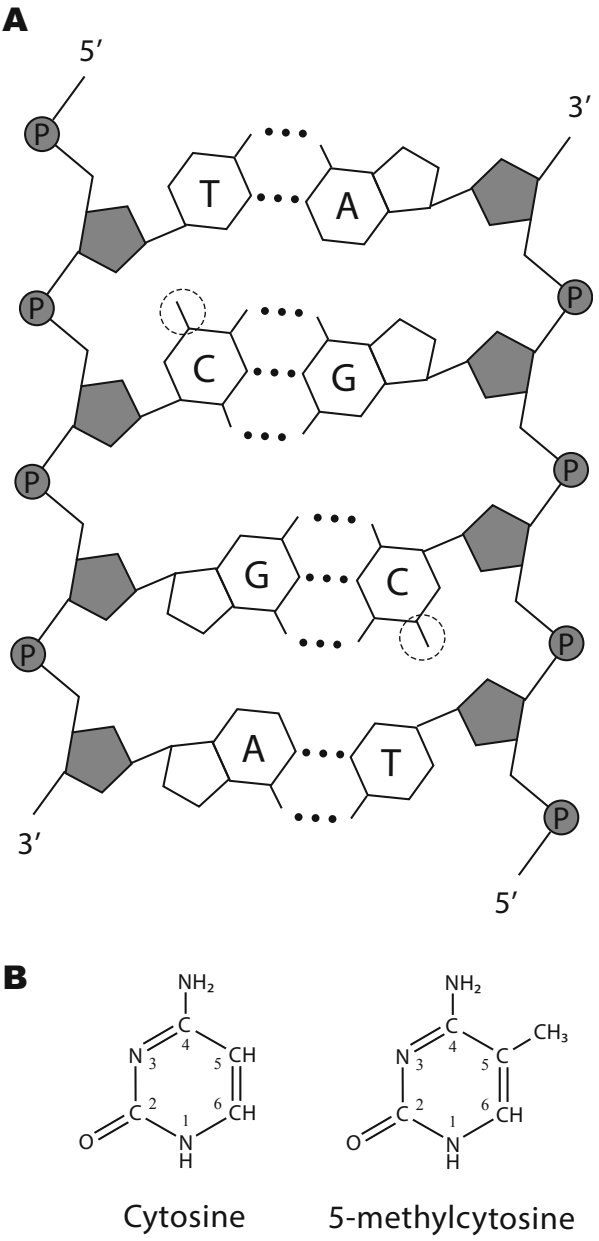

Fig. 2.1 5-methylcytosine occurs primarily within the context of a CpG dinucleotide. (**a**) Shown are two complementary (base-paired) antiparallel strands of DNA containing a CpG dinucleotide, where 'p' represents the sugar–phosphate backbone linked through phosphodiester bonds (shaded grey). In this example, the molecule contains 5-methylcytosine (dashed circled). Note that the sequence is palindromic (CG: GC), that each strand is read in the 5′ to 3′ direction and that 5-methylcytosine is present on both strands. (**b**) Methylation of cytosine occurs at the 5th atom of the 6-atom ring resulting in the formation of 5-methylcytosine

to a complementary CpG dinucleotide on the opposite DNA strand. If a cytosine on one DNA strand is methylated then its opposite counterpart will also be methylated (Fig. 2.1a; see Chap. 3 for further details). 5-mC behaves in the same way as unmethylated cytosine; it binds to guanine in double-stranded DNA (dsDNA) and

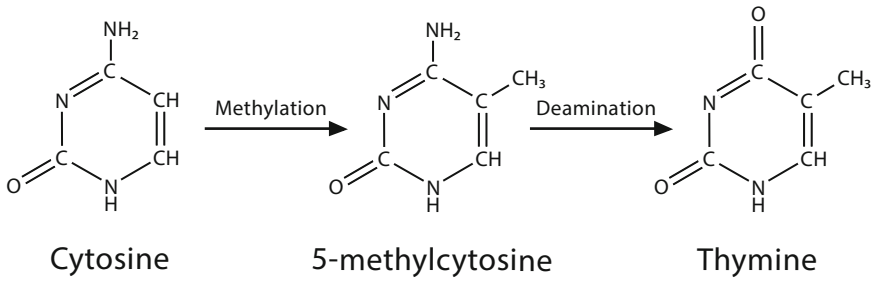

Fig. 2.2 Spontaneous deamination of 5-methylcytosine to thymine as a source of C to T transitions. The pyrimidine bases cytosine and thymine differ by virtue of a primary amine group (cytosine) or a carbonyl group (thymine) at the 4th carbon. Methylation of cytosine at the 5th atom of the 6-atom ring followed by deamination results in thymine

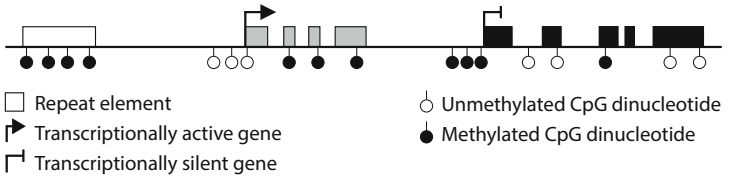

Fig. 2.3 CpG and methyl-CpG distribution. CpG dinucleotides are more abundant in regions called CpG islands, which are found in the promoter regions of a subset of genes and in non-genic sequences such as repeat elements. In normal cells, the promoter regions of expressed genes are usually unmethylated (white circles), though it should be noted that an absence of methylation does not imply the gene is actually expressed. Conversely, methylation within a gene promoter region is not conducive to gene expression. CpG dinucleotides within the body of highly expressed genes are usually methylated, whereas those in weakly or non-expressed genes are often unmethylated. CpG dinucleotides in repeat elements and transposable elements are methylated to ensure they remain transcriptionally silent

acts as the complementary base for guanine during transcription (not changing the genetic function of DNA is a defining feature of an epigenetic modification).

In humans, and other mammals, most CpG dinucleotides contain methylated cytosine (5-mC) and these methylated CpGs are distributed sparsely throughout the genome (Ziller et al. 2013). They are found at a low frequency in gene body regions and more densely at endogenous repeats and transposable elements (Bird 2002). The presence of 5-mC in gene body regions is associated with high levels of gene transcription while the more densely occurring 5-mC found at transposable elements, such as LINE-1, suppresses transcription and mobility (Schulz et al. 2006). Throughout the mammalian genome, CpG dinucleotides are relatively rare due to the spontaneous deamination of 5-mC to thymine (Lander et al. 2001) (Fig. 2.2). However, unmethylated CpG dinucleotides are more abundant in regions called CpG islands. These range in size from about 300 bp to 3 kb in length and can be found at the promoter region, often encompassing the first exon, of approximately 70% of all human genes (Fig. 2.3). There are approximately 27,000 CpG islands

found in the non-repetitive regions of the human genome (Rollins et al. 2005; Saxonov et al. 2006; Deaton and Bird 2011).

The expression pattern of genes determines the phenotype of cells, and which genes are expressed in a cell is principally determined by two factors:

1. Gene expression is regulated by the binding of appropriate (lineage-specific) transcription factors (TFs) at the promoter region. These TFs are required to recruit RNA polymerase to the start of transcribed gene regions. The presence of cell-specific transcription factors is determined during embryogenesis and is also regulated via cell signalling (Spitz and Furlong 2012).
2. For transcription factors and RNA polymerase to bind to gene promoter regions, the chromatin (DNA and associated proteins) at that region must be in an uncondensed state, that is, the number of histones in that region must be low, allowing ready access to the naked DNA.

A complex network of histone-modifying proteins regulates the level of chromatin condensation (see Chap. 4 for further details). Two key regulators are histone acetyltransferases (HATs) and histone deacetylases (HDACs). HATs add acetyl groups (CH_3CO (Ac)) while HDACs remove these from specific amino acids on histones (Verdin and Ott 2014).

The methylation status of CpGs, and in particular CpG islands, determines the acetylation status of localised histones which, in turn, regulates the condensation of that region of chromatin, the accessibility of associated promoters and thus regulates gene expression. The mechanism by which this process occurs will be discussed in the next section.

2.2 Readers of Methylcytosine: DNA Methylation and Gene Expression

The methylation status of CpGs can be read by three types of interacting proteins, these are (1) the methyl-CpG-binding domain protein (MBD) family, (2) the Kaiso proteins and Kaiso-like proteins and (3) the UHRF proteins (ubiquitin-like with PHD and RING finger domains). As the name suggests, the MBD family of proteins binds to regions containing double-stranded methylated CpG (5-mCpG) via a conserved binding domain (Nan et al. 1993), while the Kaiso and Kaiso-like proteins bind to methylated DNA via zinc finger motifs (Prokhortchouk et al. 2001; Filion et al. 2006). Both the MDB and Kaiso/Kaiso-like proteins repress the transcription of the DNA region they bind to (Hendrich and Bird 1998; Filion et al. 2006). However, the manner in which they target DNA is slightly different; Kaiso proteins bind to two consecutive methylated CpG sites whereas the MBD proteins preferentially bind to single CpG dinucleotides (Daniel et al. 2002; Sasai et al. 2010). UHRF proteins do not repress transcription; they are involved in ensuring both strands of DNA maintain 5-mCpG patterns during DNA replication; UHRF proteins bind to hemimethylated DNA (DNA where a CpG cytosine is methylated on only one strand

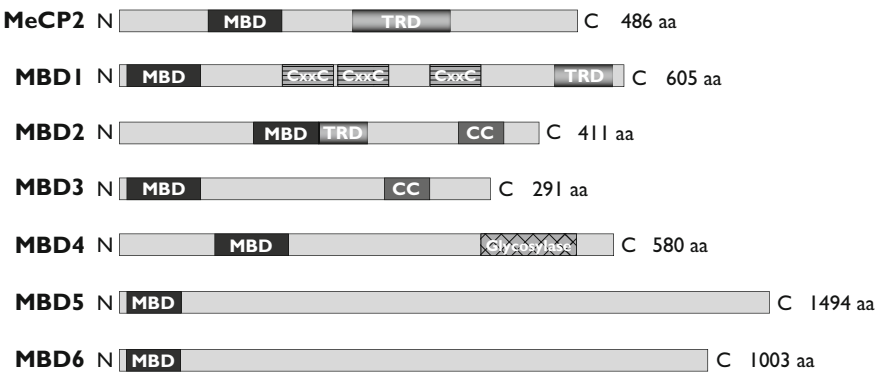

Fig. 2.4 The conserved functional domains of the methyl-binding domain proteins MeCP2, MBD1, MBD2, MBD3, MBD4, MBD5 and MBD6. All members of the methyl-binding domain family of proteins contain a methyl-CpG-binding domain (MBD) that facilitates binding to methylated DNA. MBD1 also contains CxxC domains that enables the binding of unmethylated CpG regions. MeCP2, MBD1 and MBD2 contain transcription repression domains (TDR); these regions bind to proteins that confer transcriptional repression, most notably HDAC1, 2 and 3. MBD2 and MBD3 contain coiled-coil domains (CC) that facilitate transcriptional repression through the binding of HDACs and other histone-modifying proteins such as NuRD/Mi-2. MBD4 does not contain domains involved in transcriptional repression; it possesses a glycosylase domain that facilitates mismatch repair at CpG dinucleotides. Protein size is given in amino acids (aa) and represents the size of the human homologue

of double-stranded DNA) and recruit the DNA methyltransferase DNMT1 to that location (Bostick et al. 2007; Sharif et al. 2007). See Sect. 2.3 for further details regarding the role of DNMTs.

2.2.1 Methyl-CpG-Binding Domain Proteins (MBD)

The MDB family comprises seven proteins; MeCP2 (methyl-CpG-binding protein 2), MBD1, MBD2, MBD3, MBD4, MBD5 and MBD6 (Hendrich and Bird 1998; Baymaz et al. 2014). All of these proteins bind directly to methylated DNA (via a 60–85 amino acid (aa) MBD domain). MeCP2, MBD1 and MBD2 also bind to transcriptional repressors via a transcriptional repression domain (TRD) (Wolffe et al. 1999; Boeke et al. 2000), while MBD3 interacts with transcriptional repressor proteins via a C-terminal coiled-coil domain (Hendrich and Bird 1998) (Fig. 2.4). MDB proteins can bind to different yet overlapping partner proteins via their TRD. A key family of transcriptional repressors is the histone deacetylases (HDAC1, 2 and 3). These HDACs can bind to MeCP2 and MBD1-3 (Ng et al. 1999; Kokura et al. 2001; Saito and Ishikawa 2002; Shyh-Chang and Daley 2013; Lyst et al. 2013). By bringing HDACs and other histone-modifying molecules to regions of methylated DNA, MBD proteins can direct the formation of heterochromatin. Heterochromatin consists of highly condensed chromatin that is often inaccessible to the transcriptional machinery and associated with low levels of acetylation (due to the activity of

HDACs bound to MBD proteins) (Taddei et al. 2001) at histone H3 on lysine residues 9, 14 and 18, and histone H4 on lysine residue 16 (Verdin and Ott 2014). In contrast, these regions often contain high levels of methyl groups on specific histones, such as histone H3 methylation on lysine 9 (Fuks et al. 2003). MeCP2, MBD1 and MBD2 bind to histone methyltransferase proteins (SUV39H1, SETDB1 and PRMT5) (Lunyak et al. 2002; Fujita et al. 2003; Sarraf and Stancheva 2004; Le Guezennec et al. 2006) which facilitate the methylation of local histones (for further information regarding MDBs and their binding partners, see Du et al. 2015). Heterochromatin is initiated by the juxtaposition of individual nucleosomes into dense nucleosome arrays by nucleosome remodelling complexes (see Chap. 4 for further details). Two of these complexes are NuRD/Mi-2 (a nucleosome remodelling complex with histone deacetylase activity (conferred by HDAC1 and HDAC2) (Xue et al. 1998) and the SWI/SNF ATP-dependent chromatin remodelling complex BAF (for a general review of SWI/SNF complexes, see Kadoch and Crabtree 2015).

MeCP2 binds to the BAF complex by the BRM (also known as SMARCA2) subunit (Harikrishnan et al. 2005) and MBD2 and 3 bind to the NuRD/Mi-2 complex (Ramírez et al. 2012; Baubec et al. 2013). Thus, the recognition of methylated DNA by MeCP2 and MBD1 and 3 results in modification of local histones (deacetylation and methylation). This facilitates the binding of nucleosome-remodelling protein complexes (also directed to that location by MBD proteins) and the remodelling of nucleosomes to reduce accessibility. This decrease in accessibility closes access to the local DNA for transcription factors and RNA polymerase, resulting in transcriptional silencing.

MBD4 does not play a significant role in transcription repression. Rather, it contains a glycosylase domain that facilitates mismatch repair (mC>T or C>U) at CpG dinucleotides (Hendrich et al. 1999; Petronzelli et al. 2000). Both MBD5 and 6 remain to be fully characterised; despite containing MBD domains, there is no evidence that they bind to methylated DNA (Laget et al. 2010). However, they do bind to the human polycomb deubiquitinase complex (PR-DUB) which removes ubiquitin from histone H2AK119 (Baymaz et al. 2014).

2.3 Writers of Methylcytosine: DNA Methyltransferases (DNMTs)

Humans possess five DNA methyltransferase enzymes (DNMTs); DNMT1, DNMT2, DNMT3A, DNMT3B and DNMT3L. Of these, DNMT1, DNMT3A and DNMT3B add methyl groups to cytosine bases at CpG dinucleotides within DNA (Lyko 2017) using S-adenosylmethionine as the methyl donor (Du et al. 2016). DNMT1, DNMT3A and DNMT3B contain a C-terminal catalytic domain responsible for methylating CpGs and varying N-terminal regulatory domains (Denis et al. 2011; Fig. 2.5). Although DNMT2 and DNMT3L share high levels of conservation with the other DNMTs, they do not possess catalytic domains capable of methylating CpG dinucleotides and have smaller N-terminal regulatory domains (Aapola et al. 2000; Dong et al. 2001). DNMT2 methylates a small subset of tRNAs (Legrand et al.

Fig. 2.5 The conserved functional domains of the DNA methyltransferase proteins DNMT1, DNMT2, DNMT3A, DNMT3B and DNMT3L. All members of the DNMT family share a conserved C-terminal catalytic methyltransferase domain (MTase). DNMT1 contains several conserved N-terminal regions including the DNA methyltransferase-associated protein 1-interacting domain (DMAPD) that interact with the transcriptional repressor DMAP1 and also HDAC2, the PCDNA-binding domain (PBD), the replication foci targeting domain (RFTD) and a CxxC domain that facilitates binding to unmethylated DNA. DNMT3A and DNMT3B contain Pro-Trp-Trp-Pro (PWWP) and ATRX-DNMT3-DNMT3L (ADD) domains, which interact with the N-terminal tails of histones. The ADD domain, which is conserved in DNMT3L, also facilitates binding to other regulatory proteins. Protein size is given in amino acids (aa) and represents the size of the human homologue

2017) and DNMT3L can bind to DNMT3A to form heterotetramers (2 DNMT3A molecules and 2 DNMT3L molecules). These heterotetramers have increased catalytic activity compared to DNMT3A monomers (Jia et al. 2007).

2.3.1 The Function of DNMTs: Maintenance Methylation

As described in Sect. 2.2, many of the readers of DNA methylation only recognise fully methylated double-stranded CpG dinucleotides in which the cytosine bases on both strands must be methylated. To maintain methylation during DNA replication, each newly synthesised daughter DNA strand must be methylated using the original methylated DNA strand as a template. This maintenance methylation is the principal role of DNMT1 (Bestor et al. 1988; Li et al. 1992). DNMT1 is directed to the replication fork during S phase by binding to HDAC2 and DNA methyltransferase-associated protein 1 (DMAP1) (Rountree et al. 2000). DNMT1 also binds to PCNA (proliferating cell nuclear antigen) (Chuang et al. 1997) and DNA directly at the replication fork (Suetake et al. 2006). Once at the replication fork, DNMT1 recognises hemimethylated DNA by directly interacting with the UHRF proteins (Bostick et al. 2007; Sharif et al. 2007; Berkyurek et al. 2014). Once DNMT1 is in position, its catalytic domain covers the hemimethylated DNA region and both the selection of the cytosine for methylation and the process of methylation occurs within the catalytic domain (Song et al. 2012; Bashtrykov et al. 2012). In addition to this maintenance methylation, DNMT1 (and DNMT3B) can facilitate de novo methylation of promoter regions via interactions with MBD3-NuRD/Mi-2 complexes (Cai et al. 2014).

2.3.2 The Function of DNMTs: Establishment of DNA Methylation

Appropriate DNA methylation in mammalian germ cells is essential for fertility and the viability of the resulting embryo (Messerschmidt et al. 2014). Primordial oocyte cells undergo genome-wide demethylation (Seisenberger et al. 2012), then the growing oocytes (not yet fertilised) proceed to develop complex, organised DNA methylation patterns. During this period, CpG islands associated with maternally imprinted genes are also methylated. This de novo methylation is carried out by DNMT3/DNMT3L complexes (Okano et al. 1998, 1999) and appears to be associated with histone modifications such as histone 3 lysine 36 trimethylation (H3K36me3) that are associated with transcriptional activity, suggesting that transcriptional activity is required in the non-dividing oocyte prior to gene silencing (Smallwood et al. 2011; Stewart et al. 2015). However, the precise timing and manner in which these enzymes are directed to specific regions of the oocyte genome remains to be fully clarified. Mammalian sperm cells also have distinct patterns of DNA methylation. However, unlike oocytes, as sperm progenitor cells undergo repeated mitoses at the onset of puberty, the activity of DNMT1 is required to maintain the pattern of methylation (Marques et al. 2011). Shortly after fertilisation, the embryonic genome undergoes a wave of demethylation (Sasaki and Matsui 2008) (see Sect. 2.4 for mechanistic details of DNA demethylation). The unmethylated DNA of the embryo is then re-methylated by DNMT3/DNMT3L complexes. DNMT3/DNMT3L methylates DNA in a non-selective manner; they have a preference for CpG dinucleotides but can also methylate CpH (H=A, C or T) dinucleotides (Gowher and Jeltsch 2001). Recent advances in sequencing technology have enabled the identification of frequent CpH methylation in embryonic stem cells and neurones (Lister et al. 2009; Ziller et al. 2011). Differences in cell-specific levels of CpH methylation appears to be a consequence of differential expression of DNMT3a and DNMT3b (Lee et al. 2017). The mechanism that determines which DNA sequences are de novo methylated by DNMT3/DNMT3L is unclear. However, it is known that, in addition to preferentially methylating CpG dinucleotides, this is influenced by the surrounding sequence; CpGs with purine bases on the 5′ flank and pyrimidines on the 3′ flank are preferentially methylated (Lin et al. 2002). DNMT3 binds directly to histone H3 tails (Ooi et al. 2007) and this binding allosterically activates DNMT3 resulting in the methylation of DNA local to that histone (Guo et al. 2014b; Baubec et al. 2015). The presence of large modifications on H3K4 (trimethylation or acetylation) prevents binding of DNMT3 and associated DNA methylation (Ooi et al. 2007). Genome-wide methylation studies have shown a strong correlation between the absence of H3K4me3 and DNA methylation (Weber et al. 2007; Meissner et al. 2008) suggesting that this mechanism is important in setting global DNA methylation patterns. DNMT3 also binds to histone H3 when it is trimethylated at lysine 36 (H3K36me3). This histone mark is found in gene bodies and at exon–intron boundaries and correlates to transcriptionally active regions. The presence of H3K36me3 is also inversely correlated to H3K4me3 (Vakoc et al. 2006; Barski et al. 2007; Guenther et al. 2007).

2.3.3 CpG Islands

As described previously, CpG islands are CpG-rich regions that are associated with the promoter region of approximately 70% of human genes. The majority of these CpG islands are not methylated during early embryogenesis and are associated with histone H3 modifications (H3K27me3 and H3K4me3); H3K27me3 marked regions of DNA are prone to undergo methylation during differentiation and oncogenesis (Ohm et al. 2007). The presence of transcription factors at regions of active transcription is associated with the absence of DNA methylation. This was observed more than 20 years ago (Brandeis et al. 1994; Macleod et al. 1994; Han et al. 2001) and more recently has been further validated by genome-wide association studies (GWAS) indicating that promoter region variations (and associated changes in transcription factor binding) correlated with changes in localised DNA methylation (Gutierrez-Arcelus et al. 2013). These results suggest that the cell-specific differential presence of transcription factors plays a significant role in determining the methylation status and chromatin pattern of different cell types during embryogenesis. By the blastocyst implantation stage of development, the expression of DNMT3A, B and L is reduced and the expression of DNMT1 is increased (Huntriss et al. 2004; Vassena et al. 2005; Uysal et al. 2015). At this point, cell type-specific methylation patterns have been set and are now maintained through repeated rounds of mitosis.

2.4 Erasers of Methylcytosine

There are two main processes that lead to the demethylation of DNA; the first is passive demethylation, where the absence or low expression of the maintenance methyltransferase DNMT1 results in a dilution of DNA methylation marks as the genome replicates. This results in regions of hemimethylated and, ultimately, unmethylated DNA.

As mentioned in Sect. 2.3.2, methylation marks are removed in a wave of DNA demethylation after fertilisation. Both the paternal and maternal component of the zygote genome are demethylated via passive dilution through DNA replication and an active process involving the ten–eleven translocation (TET) proteins (Guo et al. 2014a). TET proteins were first identified as being components of a fusion protein, found in acute myeloid leukaemia, due to a translocation between chromosomes ten and eleven (Lorsbach et al. 2003). The TET family of proteins (TET 1, 2 and 3) are oxygenases that catalyse the demethylation of 5mC; firstly, 5mC is converted to 5-hydroxymethylcytosine (5hmC), which is then converted to 5-formylcytosine (5fC) and finally to 5-carboxylcytosine (5caC). All three TET enzymes are capable of catalysing each of these steps (Kriaucionis and Heintz 2009; Tahiliani et al. 2009; Ito et al. 2010, 2011). The resulting 5fC or 5caC is recognised by thymine DNA glycosylase (TDG) and is restored by base excision repair (BER) (He et al. 2011). TDG excised the 5fC or 5caC resulting in an abasic site that is converted to a single-strand break. DNA polymeraseβ (Polβ) then inserts a deoxycytidine monophosphate

at the break and the double-stranded DNA is restored by DNA ligase 3 (Weber et al. 2016). This active modification–active removal (AM–AR) is independent of DNA replication (Kohli and Zhang 2013). The 5mC oxidative derivative can also remain incorporated in the genome and be converted to unmethylated cytosines by replicative dilution. This can occur in the presence of DNMT1 as it does not have a high affinity for 5hmC:C, 5fC:C and 5caC:C (Hashimoto et al. 2012; Ji et al. 2014).

In addition to demethylating genomes during embryogenesis, the TET proteins play important roles in determining pluripotency cell differentiation: Recent knockout experiments in mouse embryonic stem cells have shown that loss of TET1 or TET2 reduces the level of 5hmC and alters gene transcription. With unique targets being identified for each protein, the resulting gene expression changes have been suggested to influence differentiation potential of these stem cells (Koh et al. 2011; Huang et al. 2014).

2.4.1 Regulation of DNA Demethylation by TET Enzymes

TET protein activity requires α-KG and oxygen as substrates and Fe(II) as a cofactor. The availability of these molecules can influence the activity of the TET-catalysed reaction (Lu et al. 2015). α-KG is generated in the citric acid cycle by the activity of isocitrate dehydrogenase (IDH1, 2 and 3). Upregulation of IDH results in the production of excess α-KG and an associated increase of 5hmC throughout the genome, whereas loss-of-function mutations in IDH result in reduced α-KG production and are associated with increased levels of genomic methylation (Losman and Kaelin 2013). Similarly, fumarate and succinate can accumulate in tumour cells following disruption of the citric acid cycle and these molecules compete with α-KG, resulting in reduced TET activity (Laukka et al. 2016).

TET proteins are found to localise preferentially to CpG islands and transcriptionally active promoters; this is facilitated, in part, by DNA-binding domains that preferentially bind CpG-rich regions and, in part, due to interactions with other binding partners such as the pluripotency factor NANOG (Costa et al. 2013) and Polycomb repressive complex 2 (PRC2) (Neri et al. 2013).

In certain tissues, the levels of 5mC oxidative derivatives are relatively high, suggesting that they have functional roles beyond being intermediates for demethylation. Moreover, the balance between the activity of DNA methyltransferases and the demethylation machinery is delicate, and when an imbalance occurs, either during development or in somatic cells, disease can result.

Further details on 5mC oxidation, as well as examples of mutations in the methylation machinery that result in disease, are discussed in Chap. 3.

2.5 Conclusion

In the mammalian cell, there is a delicate balance being held between DNA-modifying enzymes, the writers and erasers of methylcytosine. During embryogenesis, the timing and molecular positioning of demethylation and then re-methylation is orchestrated with incredible precision to facilitate appropriate differentiation. Once tissue differentiation has occurred, these epigenetic marks are maintained or, in some tissue types, progressively change. In many ways, these simple markers are relatively easy to understand and molecular techniques allow us to analyse them at ever-increasing speed and throughput. However, there is still much to discover, in particular how these molecules are regulated and interact with each other and how DNA-modifying and DNA-binding molecules interact with histone-modifying and histone-binding molecules to determine chromatin structure and cell-specific gene expression. Discovering the mode of these interactions will be important for future molecular medicine.

References

Aapola U, Shibuya K, Scott HS, Ollila J, Vihinen M, Heino M, Shintani A, Kawasaki K, Minoshima S, Krohn K, Antonarakis SE, Shimizu N, Kudoh J, Peterson P (2000) Isolation and initial characterization of a novel zinc finger gene, DNMT3L, on 21q22.3, related to the cytosine-5-methyltransferase 3 gene family. Genomics 65:293–298. https://doi.org/10.1006/GENO.2000.6168

Barski A, Cuddapah S, Cui K, Roh TY, Schones DE, Wang Z, Wei G, Chepelev I, Zhao K (2007) High-resolution profiling of histone methylations in the human genome. Cell 129:823–837. https://doi.org/10.1016/j.cell.2007.05.009

Bashtrykov P, Jankevicius G, Smarandache A, Jurkowska RZ, Ragozin S, Jeltsch A (2012) Specificity of Dnmt1 for methylation of hemimethylated CpG sites resides in its catalytic domain. Chem Biol 19:572–578. https://doi.org/10.1016/J.CHEMBIOL.2012.03.010

Baubec T, Colombo DF, Wirbelauer C, Schmidt J, Burger L, Krebs AR, Akalin A, Schübeler D (2015) Genomic profiling of DNA methyltransferases reveals a role for DNMT3B in genic methylation. Nature 520:243–247. https://doi.org/10.1038/nature14176

Baubec T, Ivánek R, Lienert F, Schübeler D (2013) Methylation-dependent and -independent genomic targeting principles of the MBD protein family. Cell 153:480–492. https://doi.org/10.1016/J.CELL.2013.03.011

Baymaz HI, Fournier A, Laget S, Ji Z, Jansen PWTC, Smits AH, Ferry L, Mensinga A, Poser I, Sharrocks A, Defossez P-A, Vermeulen M (2014) MBD5 and MBD6 interact with the human PR-DUB complex through their methyl-CpG-binding domain. Proteomics 14:2179–2189. https://doi.org/10.1002/pmic.201400013

Berkyurek AC, Suetake I, Arita K, Takeshita K, Nakagawa A, Shirakawa M, Tajima S (2014) The DNA methyltransferase Dnmt1 directly interacts with the SET and RING finger-associated (SRA) domain of the multifunctional protein Uhrf1 to facilitate accession of the catalytic center to hemi-methylated DNA. J Biol Chem 289:379–386. https://doi.org/10.1074/jbc.M113.523209

Bestor T, Laudano A, Mattaliano R, Ingram V (1988) Cloning and sequencing of a cDNA encoding DNA methyltransferase of mouse cells. The carboxyl-terminal domain of the mammalian enzymes is related to bacterial restriction methyltransferases. J Mol Biol 203:971–983

Bird A (2002) DNA methylation patterns and epigenetic memory. Genes Dev 16:6–21. https://doi.org/10.1101/gad.947102

Boeke J, Ammerpohl O, Kegel S, Moehren U, Renkawitz R (2000) The minimal repression domain of MBD2b overlaps with the methyl-CpG-binding domain and binds directly to Sin3A. J Biol Chem 275:34963–34967. https://doi.org/10.1074/jbc.M005929200

Bostick M, Kim JK, Estève P-O, Clark A, Pradhan S, Jacobsen SE (2007) UHRF1 plays a role in maintaining DNA methylation in mammalian cells. Science 317:1760–1764. https://doi.org/10.1126/science.1147939

Brandeis M, Frank D, Keshet I, Siegfried Z, Mendelsohn M, Names A, Temper V, Razin A, Cedar H (1994) Spl elements protect a CpG island from de novo methylation. Nature 371:435–438. https://doi.org/10.1038/371435a0

Cai Y, Geutjes E-J, de Lint K, Roepman P, Bruurs L, Yu L-R, Wang W, van Blijswijk J, Mohammad H, de Rink I, Bernards R, Baylin SB (2014) The NuRD complex cooperates with DNMTs to maintain silencing of key colorectal tumor suppressor genes. Oncogene 33:2157–2168. https://doi.org/10.1038/onc.2013.178

Chuang LS, Ian HI, Koh TW, Ng HH, Xu G, Li BF (1997) Human DNA-(cytosine-5) methyltransferase-PCNA complex as a target for p21WAF1. Science 277:1996–2000

Costa Y, Ding J, Theunissen TW, Faiola F, Hore TA, Shliaha PV, Fidalgo M, Saunders A, Lawrence M, Dietmann S, Das S, Levasseur DN, Li Z, Xu M, Reik W, Silva JCR, Wang J (2013) NANOG-dependent function of TET1 and TET2 in establishment of pluripotency. Nature 495:370–374. https://doi.org/10.1038/nature11925

Daniel JM, Spring CM, Crawford HC, Reynolds AB, Baig A (2002) The p120(ctn)-binding partner Kaiso is a bi-modal DNA-binding protein that recognizes both a sequence-specific consensus and methylated CpG dinucleotides. Nucleic Acids Res 30:2911–2919

Deaton AM, Bird A (2011) CpG islands and the regulation of transcription. Genes Dev 25:1010–1022. https://doi.org/10.1101/gad.2037511

Denis H, Ndlovu MN, Fuks F (2011) Regulation of mammalian DNA methyltransferases: a route to new mechanisms. EMBO Rep 12:647–656. https://doi.org/10.1038/embor.2011.110

Dong A, Yoder JA, Zhang X, Zhou L, Bestor TH, Cheng X (2001) Structure of human DNMT2, an enigmatic DNA methyltransferase homolog that displays denaturant-resistant binding to DNA. Nucleic Acids Res 29:439–448. https://doi.org/10.1093/nar/29.2.439

Du Q, Luu P-L, Stirzaker C, Clark SJ (2015) Methyl-CpG-binding domain proteins: readers of the epigenome. Epigenomics 7:1051–1073. https://doi.org/10.2217/epi.15.39

Du Q, Wang Z, Schramm VL (2016) Human DNMT1 transition state structure. Proc Natl Acad Sci U S A 113:2916–2921. https://doi.org/10.1073/pnas.1522491113

Filion GJP, Zhenilo S, Salozhin S, Yamada D, Prokhortchouk E, Defossez P-A (2006) A family of human zinc finger proteins that bind methylated DNA and repress transcription. Mol Cell Biol 26:169–181. https://doi.org/10.1128/MCB.26.1.169-181.2006

Fujita N, Watanabe S, Ichimura T, Tsuruzoe S, Shinkai Y, Tachibana M, Chiba T, Nakao M (2003) Methyl-CpG binding domain 1 (MBD1) interacts with the Suv39h1-HP1 heterochromatic complex for DNA methylation-based transcriptional repression. J Biol Chem 278:24132–24138. https://doi.org/10.1074/jbc.M302283200

Fuks F, Hurd PJ, Deplus R, Kouzarides T (2003) The DNA methyltransferases associate with HP1 and the SUV39H1 histone methyltransferase. Nucleic Acids Res 31:2305–2312

Gowher H, Jeltsch A (2001) Enzymatic properties of recombinant Dnmt3a DNA methyltransferase from mouse: the enzyme modifies DNA in a non-processive manner and also methylates non-CpA sites. J Mol Biol 309:1201–1208. https://doi.org/10.1006/jmbi.2001.4710

Guenther MG, Levine SS, Boyer LA, Jaenisch R, Young RA (2007) A chromatin landmark and transcription initiation at most promoters in human cells. Cell 130:77–88. https://doi.org/10.1016/j.cell.2007.05.042

Guo F, Li X, Liang D, Li T, Zhu P, Guo H, Wu X, Wen L, Gu T-P, Hu B, Walsh CP, Li J, Tang F, Xu G-L (2014a) Active and passive demethylation of male and female pronuclear DNA in the mammalian zygote. Cell Stem Cell 15:447–459. https://doi.org/10.1016/j.stem.2014.08.003

Guo X, Wang L, Li J, Ding Z, Xiao J, Yin X, He S, Shi P, Dong L, Li G, Tian C, Wang J, Cong Y, Xu Y (2014b) Structural insight into autoinhibition and histone H3-induced activation of DNMT3A. Nature 517:640–644. https://doi.org/10.1038/nature13899

Gutierrez-Arcelus M, Lappalainen T, Montgomery SB, Buil A, Ongen H, Yurovsky A, Bryois J, Giger T, Romano L, Planchon A, Falconnet E, Bielser D, Gagnebin M, Padioleau I, Borel C, Letourneau A, Makrythanasis P, Guipponi M, Gehrig C, Antonarakis SE, Dermitzakis ET (2013) Passive and active DNA methylation and the interplay with genetic variation in gene regulation. Elife 2:e00523. https://doi.org/10.7554/eLife.00523

Han L, Lin IG, Hsieh CL (2001) Protein binding protects sites on stable episomes and in the chromosome from de novo methylation. Mol Cell Biol 21:3416–3424. https://doi.org/10.1128/MCB.21.10.3416-3424.2001

Harikrishnan KN, Chow MZ, Baker EK, Pal S, Bassal S, Brasacchio D, Wang L, Craig JM, Jones PL, Sif S, El-Osta A (2005) Brahma links the SWI/SNF chromatin-remodeling complex with MeCP2-dependent transcriptional silencing. Nat Genet 37:254–264. https://doi.org/10.1038/ng1516

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

He Y-F, Li B-Z, Li Z, Liu P, Wang Y, Tang Q, Ding J, Jia Y, Chen Z, Li L, Sun Y, Li X, Dai Q, Song C-X, Zhang K, He C, Xu G-L (2011) Tet-mediated formation of 5-carboxylcytosine and its excision by TDG in mammalian DNA. Science (80-) 333:1303–1307. https://doi.org/10.1126/science.1210944

Hendrich B, Bird A (1998) Identification and characterization of a family of mammalian methyl-CpG binding proteins. Mol Cell Biol 18:6538–6547. https://doi.org/10.1128/MCB.18.11.6538

Hendrich B, Hardeland U, Ng HH, Jiricny J, Bird A (1999) The thymine glycosylase MBD4 can bind to the product of deamination at methylated CpG sites. Nature 401:301–304. https://doi.org/10.1038/45843

Huang Y, Chavez L, Chang X, Wang X, Pastor WA, Kang J, Zepeda-Martínez JA, Pape UJ, Jacobsen SE, Peters B, Rao A (2014) Distinct roles of the methylcytosine oxidases Tet1 and Tet2 in mouse embryonic stem cells. Proc Natl Acad Sci U S A 111:1361–1366. https://doi.org/10.1073/pnas.1322921111

Huntriss J, Hinkins M, Oliver B, Harris SE, Beazley JC, Rutherford AJ, Gosden RG, Lanzendorf SE, Picton HM (2004) Expression of mRNAs for DNA methyltransferases and methyl-CpG-binding proteins in the human female germ line, preimplantation embryos, and embryonic stem cells. Mol Reprod Dev 67:323–336. https://doi.org/10.1002/mrd.20030

Ito S, D'Alessio AC, Taranova OV, Hong K, Sowers LC, Zhang Y (2010) Role of Tet proteins in 5mC to 5hmC conversion, ES-cell self-renewal and inner cell mass specification. Nature 466:1129–1133. https://doi.org/10.1038/nature09303

Ito S, Shen L, Dai Q, Wu SC, Collins LB, Swenberg JA, He C, Zhang Y (2011) Tet proteins can convert 5-methylcytosine to 5-formylcytosine and 5-carboxylcytosine. Science (80-) 333:1300–1303. https://doi.org/10.1126/science.1210597

Ji D, Lin K, Song J, Wang Y (2014) Effects of Tet-induced oxidation products of 5-methylcytosine on Dnmt1- and DNMT3a-mediated cytosine methylation. Mol Biosyst 10:1749. https://doi.org/10.1039/c4mb00150h

Jia D, Jurkowska RZ, Zhang X, Jeltsch A, Cheng X (2007) Structure of Dnmt3a bound to Dnmt3L suggests a model for de novo DNA methylation. Nature 449:248–251. https://doi.org/10.1038/nature06146

Kadoch C, Crabtree GR (2015) Mammalian SWI/SNF chromatin remodeling complexes and cancer: mechanistic insights gained from human genomics. Sci Adv 1:e1500447–e1500447. https://doi.org/10.1126/sciadv.1500447

Koh KP, Yabuuchi A, Rao S, Huang Y, Cunniff K, Nardone J, Laiho A, Tahiliani M, Sommer CA, Mostoslavsky G, Lahesmaa R, Orkin SH, Rodig SJ, Daley GQ, Rao A (2011) Tet1 and Tet2

regulate 5-hydroxymethylcytosine production and cell lineage specification in mouse embryonic stem cells. Cell Stem Cell 8:200–213. https://doi.org/10.1016/j.stem.2011.01.008

Kohli RM, Zhang Y (2013) TET enzymes, TDG and the dynamics of DNA demethylation. Nature 502:472–479. https://doi.org/10.1038/nature12750

Kokura K, Kaul SC, Wadhwa R, Nomura T, Khan MM, Shinagawa T, Yasukawa T, Colmenares C, Ishii S (2001) The Ski protein family is required for MeCP2-mediated transcriptional repression. J Biol Chem 276:34115–34121. https://doi.org/10.1074/jbc.M105747200

Kriaucionis S, Heintz N (2009) The nuclear DNA Base 5-hydroxymethylcytosine is present in Purkinje neurons and the brain. Science (80-) 324:929–930. https://doi.org/10.1126/science.1169786

Laget S, Joulie M, Le Masson F, Sasai N, Christians E, Pradhan S, Roberts RJ, Defossez P-A (2010) The human proteins MBD5 and MBD6 associate with heterochromatin but they do not bind methylated DNA. PLoS One 5:e11982. https://doi.org/10.1371/journal.pone.0011982

Lander ES, Linton LM, Birren B, Nusbaum C, Zody MC, Baldwin J, Devon K, Dewar K, Doyle M, FitzHugh W, Funke R, Gage D, Harris K, Heaford A, Howland J, Kann L, Lehoczky J, LeVine R, McEwan P, McKernan K, Meldrim J, Mesirov JP, Miranda C, Morris W, Naylor J, Raymond C, Rosetti M, Santos R, Sheridan A, Sougnez C, Stange-Thomann Y, Stojanovic N, Subramanian A, Wyman D, Rogers J, Sulston J, Ainscough R, Beck S, Bentley D, Burton J, Clee C, Carter N, Coulson A, Deadman R, Deloukas P, Dunham A, Dunham I, Durbin R, French L, Grafham D, Gregory S, Hubbard T, Humphray S, Hunt A, Jones M, Lloyd C, McMurray A, Matthews L, Mercer S, Milne S, Mullikin JC, Mungall A, Plumb R, Ross M, Shownkeen R, Sims S, Waterston RH, Wilson RK, Hillier LW, McPherson JD, Marra MA, Mardis ER, Fulton LA, Chinwalla AT, Pepin KH, Gish WR, Chissoe SL, Wendl MC, Delehaunty KD, Miner TL, Delehaunty A, Kramer JB, Cook LL, Fulton RS, Johnson DL, Minx PJ, Clifton SW, Hawkins T, Branscomb E, Predki P, Richardson P, Wenning S, Slezak T, Doggett N, Cheng JF, Olsen A, Lucas S, Elkin C, Uberbacher E, Frazier M, Gibbs RA, Muzny DM, Scherer SE, Bouck JB, Sodergren EJ, Worley KC, Rives CM, Gorrell JH, Metzker ML, Naylor SL, Kucherlapati RS, Nelson DL, Weinstock GM, Sakaki Y, Fujiyama A, Hattori M, Yada T, Toyoda A, Itoh T, Kawagoe C, Watanabe H, Totoki Y, Taylor T, Weissenbach J, Heilig R, Saurin W, Artiguenave F, Brottier P, Bruls T, Pelletier E, Robert C, Wincker P, Smith DR, Doucette-Stamm L, Rubenfield M, Weinstock K, Lee HM, Dubois J, Rosenthal A, Platzer M, Nyakatura G, Taudien S, Rump A, Yang H, Yu J, Wang J, Huang G, Gu J, Hood L, Rowen L, Madan A, Qin S, Davis RW, Federspiel NA, Abola AP, Proctor MJ, Myers RM, Schmutz J, Dickson M, Grimwood J, Cox DR, Olson MV, Kaul R, Raymond C, Shimizu N, Kawasaki K, Minoshima S, Evans GA, Athanasiou M, Schultz R, Roe BA, Chen F, Pan H, Ramser J, Lehrach H, Reinhardt R, McCombie WR, de la Bastide M, Dedhia N, Blöcker H, Hornischer K, Nordsiek G, Agarwala R, Aravind L, Bailey JA, Bateman A, Batzoglou S, Birney E, Bork P, Brown DG, Burge CB, Cerutti L, Chen HC, Church D, Clamp M, Copley RR, Doerks T, Eddy SR, Eichler EE, Furey TS, Galagan J, Gilbert JG, Harmon C, Hayashizaki Y, Haussler D, Hermjakob H, Hokamp K, Jang W, Johnson LS, Jones TA, Kasif S, Kaspryzk A, Kennedy S, Kent WJ, Kitts P, Koonin EV, Korf I, Kulp D, Lancet D, Lowe TM, McLysaght A, Mikkelsen T, Moran JV, Mulder N, Pollara VJ, Ponting CP, Schuler G, Schultz J, Slater G, Smit AF, Stupka E, Szustakowki J, Thierry-Mieg D, Thierry-Mieg J, Wagner L, Wallis J, Wheeler R, Williams A, Wolf YI, Wolfe KH, Yang SP, Yeh RF, Collins F, Guyer MS, Peterson J, Felsenfeld A, Wetterstrand KA, Patrinos A, Morgan MJ, de Jong P, Catanese JJ, Osoegawa K, Shizuya H, Choi S, Chen YJ, Szustakowki J, International Human Genome Sequencing Consortium (2001) Initial sequencing and analysis of the human genome. Nature 409:860–921. https://doi.org/10.1038/35057062

Laukka T, Mariani CJ, Ihantola T, Cao JZ, Hokkanen J, Kaelin WG, Godley LA, Koivunen P (2016) Fumarate and succinate regulate expression of hypoxia-inducible genes via TET enzymes. J Biol Chem 291:4256–4265. https://doi.org/10.1074/jbc.M115.688762

Le Guezennec X, Vermeulen M, Brinkman AB, Hoeijmakers WAM, Cohen A, Lasonder E, Stunnenberg HG (2006) MBD2/NuRD and MBD3/NuRD, two distinct complexes with

different biochemical and functional properties. Mol Cell Biol 26:843–851. https://doi.org/10.1128/MCB.26.3.843-851.2006

Lee J-H, Park S-J, Nakai K (2017) Differential landscape of non-CpG methylation in embryonic stem cells and neurons caused by DNMT3s. Sci Rep 7:11295. https://doi.org/10.1038/s41598-017-11800-1

Legrand C, Tuorto F, Hartmann M, Liebers R, Jacob D, Helm M, Lyko F (2017) Statistically robust methylation calling for whole-transcriptome bisulfite sequencing reveals distinct methylation patterns for mouse RNAs. Genome Res 27:1589–1596. https://doi.org/10.1101/gr.210666.116

Li E, Bestor TH, Jaenisch R (1992) Targeted mutation of the DNA methyltransferase gene results in embryonic lethality. Cell 69:915–926

Lin IG, Han L, Taghva A, O'Brien LE, Hsieh C-L (2002) Murine de novo methyltransferase Dnmt3a demonstrates strand asymmetry and site preference in the methylation of DNA in vitro. Mol Cell Biol 22:704–723

Lister R, Pelizzola M, Dowen RH, Hawkins RD, Hon G, Tonti-Filippini J, Nery JR, Lee L, Ye Z, Ngo Q-M, Edsall L, Antosiewicz-Bourget J, Stewart R, Ruotti V, Millar AH, Thomson JA, Ren B, Ecker JR (2009) Human DNA methylomes at base resolution show widespread epigenomic differences. Nature 462:315–322. https://doi.org/10.1038/nature08514

Lorsbach RB, Moore J, Mathew S, Raimondi SC, Mukatira ST, Downing JR (2003) TET1, a member of a novel protein family, is fused to MLL in acute myeloid leukemia containing the t(10;11)(q22;q23). Leukemia 17:637–641. https://doi.org/10.1038/sj.leu.2402834

Losman J-A, Kaelin WG (2013) What a difference a hydroxyl makes: mutant IDH, (R)-2-hydroxyglutarate, and cancer. Genes Dev 27:836–852. https://doi.org/10.1101/gad.217406.113

Lu X, Zhao BS, He C (2015) TET family proteins: oxidation activity, interacting molecules, and functions in diseases. Chem Rev 115:2225–2239. https://doi.org/10.1021/cr500470n

Lunyak VV, Burgess R, Prefontaine GG, Nelson C, Sze S-H, Chenoweth J, Schwartz P, Pevzner PA, Glass C, Mandel G, Rosenfeld MG (2002) Corepressor-dependent silencing of chromosomal regions encoding neuronal genes. Science (80-) 298:1747–1752. https://doi.org/10.1126/science.1076469

Lyko F (2017) The DNA methyltransferase family: a versatile toolkit for epigenetic regulation. Nat Rev Genet 19:81–92. https://doi.org/10.1038/nrg.2017.80

Lyst MJ, Ekiert R, Ebert DH, Merusi C, Nowak J, Selfridge J, Guy J, Kastan NR, Robinson ND, de Lima Alves F, Rappsilber J, Greenberg ME, Bird A (2013) Rett syndrome mutations abolish the interaction of MeCP2 with the NCoR/SMRT co-repressor. Nat Neurosci 16:898–902. https://doi.org/10.1038/nn.3434

Macleod D, Charlton J, Mullins J, Bird AP (1994) Sp1 sites in the mouse aprt gene promoter are required to prevent methylation of the CpG island. Genes Dev 8:2282–2292

Marques CJ, João Pinho M, Carvalho F, Bièche I, Barros A, Sousa M (2011) DNA methylation imprinting marks and DNA methyltransferase expression in human spermatogenic cell stages. Epigenetics 6:1354–1361. https://doi.org/10.4161/epi.6.11.17993

Meissner A, Mikkelsen TS, Gu H, Wernig M, Hanna J, Sivachenko A, Zhang X, Bernstein BE, Nusbaum C, Jaffe DB, Gnirke A, Jaenisch R, Lander ES (2008) Genome-scale DNA methylation maps of pluripotent and differentiated cells. Nature 454:766–770. https://doi.org/10.1038/nature07107

Messerschmidt DM, Knowles BB, Solter D (2014) DNA methylation dynamics during epigenetic reprogramming in the germline and preimplantation embryos. Genes Dev 28:812–828. https://doi.org/10.1101/gad.234294.113

Nan X, Meehan RR, Bird A (1993) Dissection of the methyl-CpG binding domain from the chromosomal protein MeCP2. Nucleic Acids Res 21:4886–4892. https://doi.org/10.1093/nar/21.21.4886

Neri F, Incarnato D, Krepelova A, Rapelli S, Pagnani A, Zecchina R, Parlato C, Oliviero S (2013) Genome-wide analysis identifies a functional association of Tet1 and Polycomb repressive complex 2 in mouse embryonic stem cells. Genome Biol 14:R91. https://doi.org/10.1186/gb-2013-14-8-r91

Ng HH, Zhang Y, Hendrich B, Johnson CA, Turner BM, Erdjument-Bromage H, Tempst P, Reinberg D, Bird A (1999) MBD2 is a transcriptional repressor belonging to the MeCP1 histone deacetylase complex. Nat Genet 23:58–61. https://doi.org/10.1038/12659

Ohm JE, McGarvey KM, Yu X, Cheng L, Schuebel KE, Cope L, Mohammad HP, Chen W, Daniel VC, Yu W, Berman DM, Jenuwein T, Pruitt K, Sharkis SJ, Watkins DN, Herman JG, Baylin SB (2007) A stem cell-like chromatin pattern may predispose tumor suppressor genes to DNA hypermethylation and heritable silencing. Nat Genet 39:237–242. https://doi.org/10.1038/ng1972

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

Okano M, Li E, Xie S (1998) Cloning and characterization of a family of novel mammalian DNA (cytosine-5) methyltransferases. Nat Genet 19:219–220. https://doi.org/10.1038/890

Ooi SKT, Qiu C, Bernstein E, Li K, Jia D, Yang Z, Erdjument-Bromage H, Tempst P, Lin S-P, Allis CD, Cheng X, Bestor TH (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

Petronzelli F, Riccio A, Markham GD, Seeholzer SH, Genuardi M, Karbowski M, Yeung AT, Matsumoto Y, Bellacosa A (2000) Investigation of the substrate spectrum of the human mismatch-specific DNA N-glycosylase MED1 (MBD4): fundamental role of the catalytic domain. J Cell Physiol 185:473–480. https://doi.org/10.1002/1097-4652(200012)185:3<473::AID-JCP19>3.0.CO;2-#

Prokhortchouk A, Hendrich B, Jørgensen H, Ruzov A, Wilm M, Georgiev G, Bird A, Prokhortchouk E (2001) The p120 catenin partner Kaiso is a DNA methylation-dependent transcriptional repressor. Genes Dev 15:1613–1618. https://doi.org/10.1101/gad.198501

Ramírez J, Dege C, Kutateladze TG, Hagman J (2012) MBD2 and multiple domains of CHD4 are required for transcriptional repression by Mi-2/NuRD complexes. Mol Cell Biol 32:5078–5088. https://doi.org/10.1128/MCB.00819-12

Rollins RA, Haghighi F, Edwards JR, Das R, Zhang MQ, Ju J, Bestor TH (2005) Large-scale structure of genomic methylation patterns. Genome Res 16:157–163. https://doi.org/10.1101/gr.4362006

Rountree MR, Bachman KE, Baylin SB (2000) DNMT1 binds HDAC2 and a new co-repressor, DMAP1, to form a complex at replication foci. Nat Genet 25:269–277. https://doi.org/10.1038/77023

Saito M, Ishikawa F (2002) The mCpG-binding domain of human MBD3 does not bind to mCpG but interacts with NuRD/Mi2 components HDAC1 and MTA2. J Biol Chem 277:35434–35439. https://doi.org/10.1074/jbc.M203455200

Sarraf SA, Stancheva I (2004) Methyl-CpG binding protein MBD1 couples histone H3 methylation at lysine 9 by SETDB1 to DNA replication and chromatin assembly. Mol Cell 15:595–605. https://doi.org/10.1016/j.molcel.2004.06.043

Sasai N, Nakao M, Defossez P-A (2010) Sequence-specific recognition of methylated DNA by human zinc-finger proteins. Nucleic Acids Res 38:5015–5022. https://doi.org/10.1093/nar/gkq280

Sasaki H, Matsui Y (2008) Epigenetic events in mammalian germ-cell development: reprogramming and beyond. Nat Rev Genet 9:129–140. https://doi.org/10.1038/nrg2295

Saxonov S, Berg P, Brutlag DL (2006) A genome-wide analysis of CpG dinucleotides in the human genome distinguishes two distinct classes of promoters. Proc Natl Acad Sci U S A 103:1412–1417. https://doi.org/10.1073/pnas.0510310103

Schulz WA, Steinhoff C, Florl AR (2006) Methylation of endogenous human retroelements in health and disease. Curr Top Microbiol Immunol 310:211–250

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

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 450:908–912. https://doi.org/10.1038/nature06397

Shyh-Chang N, Daley GQ (2013) Lin28: primal regulator of growth and metabolism in stem cells. Cell Stem Cell 12:395–406. https://doi.org/10.1016/j.stem.2013.03.005

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

Song J, Teplova M, Ishibe-Murakami S, Patel DJ (2012) Structure-based mechanistic insights into DNMT1-mediated maintenance DNA methylation. Science 335:709–712. https://doi.org/10.1126/science.1214453

Spitz F, Furlong EEM (2012) Transcription factors: from enhancer binding to developmental control. Nat Rev Genet 13:613–626. https://doi.org/10.1038/nrg3207

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 29:2449–2462. https://doi.org/10.1101/gad.271353.115

Suetake I, Hayata D, Tajima S (2006) The amino-terminus of mouse DNA methyltransferase 1 forms an independent domain and binds to DNA with the sequence involving PCNA binding motif. J Biochem 140:763–776. https://doi.org/10.1093/jb/mvj210

Taddei A, Maison C, Roche D, Almouzni G (2001) Reversible disruption of pericentric heterochromatin and centromere function by inhibiting deacetylases. Nat Cell Biol 3:114–120. https://doi.org/10.1038/35055010

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 324:930–935. https://doi.org/10.1126/science.1170116

Uysal F, Akkoyunlu G, Ozturk S (2015) Dynamic expression of DNA methyltransferases (DNMTs) in oocytes and early embryos. Biochimie 116:103–113. https://doi.org/10.1016/j.biochi.2015.06.019

Vakoc CR, Sachdeva MM, Wang H, Blobel GA (2006) Profile of histone lysine methylation across transcribed mammalian chromatin. Mol Cell Biol 26:9185–9195. https://doi.org/10.1128/MCB.01529-06

Vassena R, Dee Schramm R, Latham KE (2005) Species-dependent expression patterns of DNA methyltransferase genes in mammalian oocytes and preimplantation embryos. Mol Reprod Dev 72:430–436. https://doi.org/10.1002/mrd.20375

Verdin E, Ott M (2014) 50 years of protein acetylation: from gene regulation to epigenetics, metabolism and beyond. Nat Rev Mol Cell Biol 16:258–264. https://doi.org/10.1038/nrm3931

Weber AR, Krawczyk C, Robertson AB, Kuśnierczyk A, Vågbø CB, Schuermann D, Klungland A, Schär P (2016) Biochemical reconstitution of TET1–TDG–BER-dependent active DNA demethylation reveals a highly coordinated mechanism. Nat Commun 7:10806. https://doi.org/10.1038/ncomms10806

Weber M, Hellmann I, Stadler MB, Ramos L, Pääbo S, Rebhan M, Schübeler D (2007) Distribution, silencing potential and evolutionary impact of promoter DNA methylation in the human genome. Nat Genet 39:457–466. https://doi.org/10.1038/ng1990

Wolffe AP, Wade PA, Gegonne A, Jones PL, Ballestar E, Aubry F (1999) Mi-2 complex couples DNA methylation to chromatin remodelling and histone deacetylation. Nat Genet 23:62–66. https://doi.org/10.1038/12664

Xue Y, Wong J, Moreno GT, Young MK, Côté J, Wang W (1998) NURD, a novel complex with both ATP-dependent chromatin-remodeling and histone deacetylase activities. Mol Cell 2:851–861

Ziller MJ, Gu H, Müller F, Donaghey J, Tsai LT-Y, Kohlbacher O, De Jager PL, Rosen ED, Bennett DA, Bernstein BE, Gnirke A, Meissner A (2013) Charting a dynamic DNA methylation landscape of the human genome. Nature 500:477–481. https://doi.org/10.1038/nature12433

Ziller MJ, Müller F, Liao J, Zhang Y, Gu H, Bock C, Boyle P, Epstein CB, Bernstein BE, Lengauer T, Gnirke A, Meissner A (2011) Genomic distribution and inter-sample variation of non-CpG methylation across human cell types. PLoS Genet 7:e1002389. https://doi.org/10.1371/journal.pgen.1002389

5-Methylcytosine and Its Oxidized Derivatives

3

Gerd P. Pfeifer

Abstract

5-methylcytosine is an enzymatically produced modified cytosine base that was first identified about 70 years ago. This modification is generally understood to be incompatible with gene expression when found in regulatory regions including gene promoters and enhancers. Oxidized 5-methylcytosine bases, in the form of its hydroxyl, formyl and carboxyl derivatives, have been known for about 10 years. Their biological functions in development and in human diseases are not yet clearly defined. Human diseases including cancer and developmental and neurological disorders have been linked to mutations of enzymes that produce these modifications (e.g., DNMT3A, DNMT3B, TET2) and to proteins that recognize 5-methylcytosine (MECP2) attesting to the clinical relevance of modified cytosines and the pathways they are involved in.

Keywords

DNA methylation · 5-Methylcytosine · 5-Hydroxymethylcytosine · 5-Formylcytosine · 5-Carboxylcytosine · DNA methyltransferases · Methyl-CpG-binding proteins · 5-Methylcytosine oxidases · TET proteins · DNA demethylation

3.1 Introduction

DNA methylation is an epigenetic modification that is considered more permanent or more stable than the many posttranslational modifications occurring on chromatin proteins. DNA methyltransferases transfer methyl groups from the co-factor

G. P. Pfeifer (✉)
Center for Epigenetics, Van Andel Research Institute, Grand Rapids, MI, USA
e-mail: gerd.pfeifer@vai.org

L. B. Hesson, A. L. Pritchard (eds.), *Clinical Epigenetics*,
https://doi.org/10.1007/978-981-13-8958-0_3

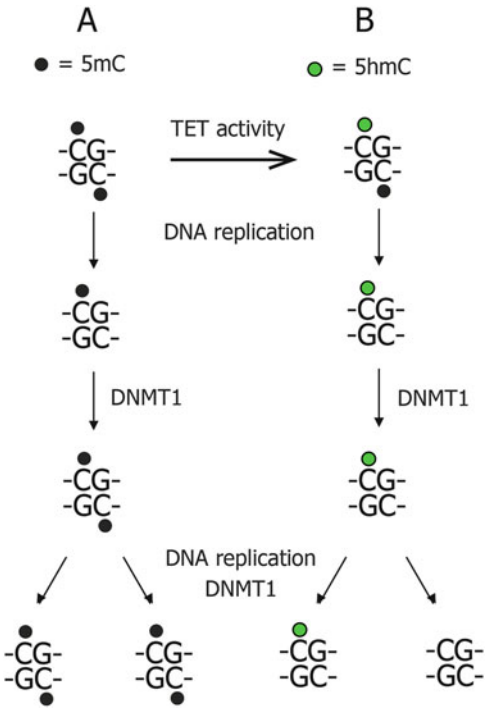

Fig. 3.1 The fate of 5-methylcytosine (5mC) and 5-hydroxymethylcytosine (5hmC) during DNA replication. (**a**) The state of 5mC modifications at CpG sites is copied by DNMT1 during replication. (**b**) DNMT1 is incapable of maintaining DNA methylation at sequences containing 5hmC

S-adenosyl-L-methionine onto carbon 5 of the cytosine ring resulting in a DNA base that has base pairing properties identical to cytosine. 5-methylcytosine (5mC) was first detected as a normal component of mammalian DNA about 70 years ago (Hotchkiss 1948; Wyatt 1950). Its biological function has been enigmatic for many decades and is still not completely understood today. In 1975, it was first proposed that 5mC plays a role in gene regulation (Holliday and Pugh 1975; Riggs 1975). Importantly, it was also recognized that this modification, which occurs preferentially at CpG palindromic dinucleotide sequences, has the capacity to be copied during DNA replication if an enzyme exists that recognizes hemimethylated DNA, a substrate which carries the methyl group on only the parental DNA strand after replication (Fig. 3.1a).

Evidence from bacterial restriction-modification systems had suggested that the cytosine modification at position 5 interferes with binding or catalysis of certain DNA-binding proteins such as restriction endonucleases (Riggs 1975). Indeed, many transcription factors are incapable of binding to methylated DNA (Watt and Molloy 1988; Tate and Bird 1993; Schubeler 2015). CpG methylation may not only negatively affect protein–DNA interactions but may also promote new interactions where the methyl group in the major groove of the DNA double helix is recognized by specific proteins. It was not until 1989, however, that the first ‘reader’ capable of recognizing CpG-methylated DNA was identified (Meehan et al. 1989). These readers initially included a small family of proteins that carry a methyl-CpG-binding

domain (MBD), a domain capable of binding to methylated CpG dinucleotides in DNA in vitro (Hendrich and Bird 1998). More recently, it has been realized that a fairly large number of sequence-specific transcription factors with recognition motifs much larger than a dinucleotide can interact with methylated DNA directly, sometimes even with a preference for the methylated state (Zhu et al. 2016; Yin et al. 2017). However, it remains unclear how common these interactions are in vivo, where CpG-methylated DNA is generally associated with inaccessible chromatin. Such additional proteins with preference for methylated CpGs include certain types of zinc finger proteins (Buck-Koehntop and Defossez 2013; Liu et al. 2013). Further, a small number of mammalian proteins carry a domain called SET and RING finger-associated (SRA) domain (Unoki et al. 2004; Johnson et al. 2007), which is also capable of interacting with methylated CpGs.

Cytosine methylation in mammals is not strictly limited to CpG sequences; it also occurs at non-CpG sites including 5′CpA, 5′CpC and 5′CpT. This non-CpG methylation is only found in certain tissues including pluripotent stem cells and certain non-dividing cells such as neurons. It is more difficult to maintain this type of methylation during cell division and it needs to be re-established after each round of replication. However, the function of non-CpG methylation is not known (Patil et al. 2014).

Long thought to be irreversible, except for iterative dilution of the modification by consecutive DNA replication cycles in the absence of DNMT1, methylated cytosines are now known to be subject to enzymatic oxidation and further processing. Remarkably, these oxidized 5mC bases were not discovered until 2009. Before that time, they were considered rare oxidation products of 5mC arising from oxidative stress-induced DNA damage. In that year, two groups described the existence of the base 5-hydroxymethylcytosine (5hmC) in brain-derived neurons and in embryonic stem cells (Kriaucionis and Heintz 2009; Tahiliani et al. 2009). One study identified a 5mC oxidase protein [ten–eleven translocation 1, or TET1, named after a translocation occurring in leukaemia (Ono et al. 2002; Lorsbach et al. 2003)] that can carry out the oxidation reaction in vitro (Tahiliani et al. 2009). There are two paralogues of TET1, TET2 and TET3, which have similar enzymatic properties (Tahiliani et al. 2009; Ito et al. 2010). These enzymes are evolutionary conserved, and a TET homologue is identifiable in other multicellular organisms, for example, it is found in the *Drosophila* genome (Dunwell et al. 2013).

In this review, emphasis will be on the potential roles of 5mC and its oxidized derivatives in human disease with a focus on mutational inactivation of writers, readers and erasers of the modifications. This is not to say that changes in expression levels or modifications of these families of proteins may be irrelevant to disease, but these alterations are not covered in this summary.

3.2 DNA Methyltransferases

Mammalian DNA methyltransferases (DNMTs) were first identified and characterized in the 1970s and 1980s (Drahovsky and Morris 1971; Turnbull and Adams 1976; Bestor and Ingram 1985; Pfeifer et al. 1985). The activities initially purified and described exhibited a substantial preference for methylation of hemimethylated DNA sites conforming to a role as maintenance DNA methyltransferases (Fig. 3.1). The first characterized DNMT protein is now referred to as DNMT1. Later, it was found that DNMT1 alone may be insufficient for complete maintenance of DNA methylation in vivo but requires the presence of an accessory factor that binds preferentially to hemimethylated CpG sites, a protein called UHRF1 (Bostick et al. 2007). Two additional genes encoding catalytically active DNMT polypeptides are found in mammals, *Dnmt3a* and *Dnmt3b*. The encoded proteins have the ability to readily methylate completely unmethylated CpG-containing substrates (Okano et al. 1999) and are considered to be the major de novo DNMTs, although DNMT1 by itself can also methylate unmethylated CpG sites. A catalytically inactive heterodimerization partner of DNMT3A and DNMT3B, known as DNMT3L, stimulates the activity of DNMT3A and DNMT3B (Chedin et al. 2002) and is highly expressed during developmental stages where de novo methylation takes place (Bourc'his et al. 2001). There is an additional DNMT-like protein, DNMT2, which has all the evolutionary conserved residues of the DNA methyltransferase catalytic domain, but which is now understood to be primarily a tRNA methyltransferase (Goll et al. 2006). Genetic ablation of *Dnmt1* in mice leads to embryonic lethality (Li et al. 1992). Deletion of *Dnmt3a* or *Dnmt3b* or both also results in severe developmental failure (Okano et al. 1999), suggesting that DNA methylation is essential in mammals.

3.3 The Function of 5-Methylcytosine

Since the loss of DNMT proteins is incompatible with mammalian development, the modified base 5mC, although very low or non-existent in certain rather complex organisms such as *Caenorhabditis elegans* or *Drosophila melanogaster*, must have an essential biological role. This role should be so important that it overrides the inherent mutability of methylated CpG sites, which has led to a 4–5-fold depletion of CpG dinucleotide sequences in mammalian genomes over evolutionary time. Mutations at CpG sites, which may occur due to spontaneous deamination of 5mC to form thymine bases, can also be promoted by carcinogens interacting preferentially with methylated CpGs (Pfeifer 2006). Differential effectiveness of DNA repair pathways operating at methylated CpGs versus unmethylated CpGs may also play a role in the preferential mutability of methylated CpG sequences. For example, uracil DNA glycosylases may be more efficient in excising uracils after cytosine deamination than thymine excision enzymes that can remove thymines from T/G mispairs formed after deamination of 5mC. There are two such proteins encoded in mammalian genomes, MBD4 and TDG, although the latter has now been shown to operate at

5-formylcytosine and 5-carboxylcytosine raising the question of what the preferred substrate of this enzyme is in vivo (Schuermann et al. 2016). As a consequence of its inherently high mutability, CpG is a common mutation site in many genetic diseases and in cancer genomes. In fact, mutations at CpG sites are so frequent that they can dominate the entire mutational spectra, for example, in colorectal or brain cancers (Pfeifer and Hainaut 2011).

CpG depletion has however eluded many CpG-rich regulatory DNA sequences. The prime examples are CpG islands, sequences at least a few hundred base pairs in length that still retain the expected CpG frequencies (roughly 1 in 16 bases) and are remarkably G + C rich. Such CpG islands are often found at gene promoter sequences where they are landmarks for active or potentially active (poised) genes. Mammalian genomes contain a set of about a dozen genes that code for proteins with a CXXC domain, a unique zinc-finger-type domain that binds to unmethylated CpG-rich sequences including CpG islands (Long et al. 2013). Several of these CXXC proteins encode epigenetic enzymes or interact with critical epigenetic modifiers. For example, 5-methylcytosine oxidase proteins (TET1, TET3), histone lysine methyltransferases (MLL1, MLL2) and histone lysine demethylases (KDM2A, KDM2B) all contain a CXXC domain. The CXXC proteins appear to have the common task of keeping CpG islands free of DNA methylation, thus ensuring the preservation of a chromatin state that allows transcriptional competence as well as keeping mutagenic processes away from CpG islands.

For many years, DNA methylation was believed to be a critical mechanism in directing development and tissue differentiation. Many studies had shown that genes that are expressed are usually unmethylated, and genes that are silenced are methylated at their 5′ ends and promoters. However, genome-wide studies did not fully support those correlations because inactive genes may also be unmethylated and are kept inactive by repressive histone modification complexes such as the Polycomb complex. One important role of 5mC in gene silencing has been ascribed to repression of endogenous retroviruses, repetitive sequences and transposable elements, which represent a large percentage of the material in mammalian genomes (Yoder et al. 1997). For these targets, the methylated regions are often co-associated with repressive histone modifications in the form of H3K9me3 (Groh and Schotta 2017).

Today, many researchers believe that DNA methylation occurs as a default process meaning that all available CpG sequences of the genome become methylated unless they are somehow protected from DNA methylation. The exact nature of these protection mechanisms, in particular those operating at CpG islands, are still not completely known. In addition to the CXXC proteins, several DNA-binding proteins as exemplified by a number of sequence-specific transcription factors have the ability to exclude CpG methylation events when they are bound to DNA (Stadler et al. 2011). Larger DNA-binding complexes, such as Polycomb proteins and their associated factors, may function in a similar way and may occupy several kilobases of sequence. It is notable that large stretches, sometimes over 10 kb in length, of Polycomb-bound genomic regions, including many homeobox genes and other developmental regulators, have very low levels of DNA methylation in normal

tissues or cell types (Xie et al. 2013). However, these unmethylated DNA methylation states break down under certain disease conditions, such as inflammation or cancer and even during the normal aging process, when Polycomb-targeted CpG-rich genes and genomic regions become frequently hypermethylated (Rauch et al. 2006; Ohm et al. 2007; Schlesinger et al. 2007; Widschwendter et al. 2007; Maegawa et al. 2010; Jung and Pfeifer 2015). Cancer-associated hypermethylation of developmental Polycomb-targeted genes is in fact a major contributing event seen prominently among the many DNA methylation changes occurring in malignant tissues. However, the mechanisms of how DNA at Polycomb-targeted genes becomes methylated, often in conjunction with loss of the Polycomb-associated H3K27me3 mark (Gal-Yam et al. 2008), is still unclear. It is also not known if methylation of Polycomb target genes has any functional role in tumour development or if it just occurs as a consequence of malignant progression. It has been argued that such methylation events may lead to irreversible silencing of certain transcription factor genes thus imposing an immature state on somatic stem cells that prevents their differentiation (Wu et al. 2010; Kalari et al. 2013; Easwaran et al. 2014). Such a DNA methylation-induced differentiation defect may contribute to tumourigenesis.

One idea that has been proposed for a long time is that 5mC plays a role in genomic noise reduction to maintain specific gene expression patterns by suppressing inappropriate transcription initiation (Bird 1995). Such spurious transcription may arise in intergenic areas but also may interfere with expression of genic regions if it occurs on the opposite DNA strand of transcribed genes, for example. One interesting finding in that regard is that 5mC appears to be enriched in gene bodies (defined as transcribed regions between the transcription start site and the transcription end site of a gene) where its level in gene bodies positively correlates with gene expression. This observation was first made for genes on the active X chromosome (Hellman and Chess 2007) but extends to almost all genes genome wide (Rauch et al. 2009). It is proposed that DNMT3B is primarily responsible for gene body DNA methylation (Baubec et al. 2015; Duymich et al. 2016). A relevant recent study used a genetic model in which *Dnmt3b* was inactivated resulting in lower gene body methylation (Neri et al. 2017). In the *Dnmt3b* mutants, anti-sense transcription and transcriptional noise was indeed increased substantially supporting the noise reduction model of 5mC function.

Methylation of CpG sequences has other unique roles in the genome. To allow differentiation between the two alleles of monoallelically expressed genes within the same nucleus of a cell, methylation marks CpG islands on the inactive X chromosome (Riggs and Pfeifer 1992) and is deposited on only one allele of imprinted genes (Mann et al. 2000; Delaval and Feil 2004). In these cases, differential methylation seems to follow or is excluded by other chromatin modification events, but may aid in long-term maintenance of the allelic expression state of an entire chromosome or of specific chromosomal regions. For example, DNA methylation of CpG islands on the inactive X chromosome in female cells probably follows the deposition of the Polycomb mark H3K27me3 on the inactive X (Heard 2004; Brockdorff 2017). In another example, during the establishment of differentially methylated regions of

imprinted genes in primordial germ cells, DNA methylation is excluded on the chromosomes of male germ cells by the histone mark H3K4me3 (Singh et al. 2013). This phenomenon occurs because the H3K4me3 modification is incompatible with DNA methyltransferase-induced de novo DNA methylation (Ooi et al. 2007).

A general model that has emerged from these studies over the past two decades is that the presence or absence of epigenetic marks on DNA or histones is initially determined by transcription factors during development. In many instances, these factors recruit co-activators such as histone acetyltransferase complexes to promote gene activity. Differential epigenetic compartmentalization of the genome then facilitates the long-term maintenance of a differentiated state. However, there are situations when repressed and DNA-methylated chromatin needs to become accessible to a transcription factor. DNA-binding proteins that can override this repression have been named 'pioneer factors'; their functional role likely involves chromatin remodelling events and, in some instances, removal of 5mC by DNA demethylation (Boller et al. 2016; Zhang et al. 2016b).

3.4 Mutations in *DNMT1*

In humans, heterozygous autosomal-dominant mutations in *DNMT1* cause adult-onset neurological phenotypes including central and peripheral neurodegeneration (Klein et al. 2011). These mutations occur in the *DNMT1* chromosomal targeting sequence and lead to protein degradation and diminished heterochromatin loading of DNMT1 protein. The syndrome has been referred to as hereditary sensory and autonomic neuropathy with dementia and hearing loss (HSAN1E). Analysis of patient samples with a *DNMT1* hotspot mutation uncovered a large number of moderate DNA methylation changes with predominantly loss of methylation (Sun et al. 2014). *DNMT1* mutations in an adjacent exon also lead to autosomal-dominant cerebellar ataxia, deafness and narcolepsy (ADCA-DN) (Winkelmann et al. 2012). However, it is still unclear how exactly these methylation changes promote these severe neurological phenotypes. Even though cancer cells often are characterized by a weakly to moderately hypomethylated DNA (Ehrlich 2009), with loss of 5mC occurring preferentially in repetitive compartments of the genome (Rauch et al. 2008), mutations in *DNMT1* have rarely been observed in human cancers.

3.5 Mutations in *DNMT3A*

Somatic mutations in the *DNMT3A* methyltransferase gene are a major driver in some cases of myelodysplastic syndrome (MDS) and acute myeloid leukaemia (AML), two types of common human haematological malignancies. The frequency of *DNMT3A* mutations is between ~20 and 25% and ~10% in AML and MDS, respectively, but these mutations are uncommon in solid tumours. A large fraction (50–60%) of DNMT3A mutations occur at amino acid position p.Arg882 (p.R882), which is within the catalytic domain of the enzyme and has a dominant negative

effect strongly reducing DNMT3A catalytic activity (Brunetti et al. 2017). A recent study used whole genome bisulfite sequencing to analyze the consequence of the DNMT3A p.Arg882 mutation and found focal DNA methylation loss even in non-leukaemic haematopoietic cells with *DNMT3A* mutation (Spencer et al. 2017). It is still unclear how these changes in methylation patterns precisely influence disease initiation and/or progression. Interestingly, this recent study also suggested that CpG island hypermethylation is a consequence of AML progression rather than a cause (Spencer et al. 2017). *DNMT3A* mutations are early events in leukaemia detected in pre-leukaemic clones and confer a poor prognosis (Brunetti et al. 2017).

De novo germline mutations in *DNMT3A* cause an overgrowth syndrome with intellectual disability, known as Tatton-Brown-Rahman syndrome (Tatton-Brown et al. 2014). In rare cases, these mutations can affect the same amino acid as seen mutated in AML ($DNMT3A^{R882H}$).

3.6 Mutations in *DNMT3B*

Mutations in both alleles of *DNMT3B* were first reported in 1999 in the human syndrome known as ICF syndrome (immunodeficiency, centromere instability and facial anomalies) (Hansen et al. 1999; Xu et al. 1999). Cells from these patients have abnormal nuclear structures in which juxtacentromeric regions undergo persistent self-association in interphase leading to centromeric instability. In *DNMT3B* mutant cells, pericentromeric satellite DNA becomes hypomethylated. In addition, hypomethylation in specific non-pericentromeric regions of chromosomes, such as subtelomeric regions and gene bodies, might also be involved in the molecular pathogenesis of ICF syndrome (Aran et al. 2011; Huang et al. 2014; Simo-Riudalbas et al. 2015).

Rare cases of facioscapulohumeral muscular dystrophy (FSHD) also have mutations in *DNMT3B* (van den Boogaard et al. 2016). This disease is characterized by hypomethylation of the regulatory regions of the *DUX4* double homeobox transcription factor gene leading to inappropriate activation of *DUX4* in skeletal muscle tissue, muscle cell toxicity and degeneration of muscle tissue (Daxinger et al. 2015). Unlike *DNMT3A*, *DNMT3B* is not commonly mutated in human cancer.

3.7 5-Methylcytosine-Binding Proteins

The search for proteins that can bind to methyl-CpG DNA initially led to the identification of a large protein complex referred to as MECP1 (Meehan et al. 1989). The first single isolated protein identified as a methyl-CpG binder was MECP2 (Lewis et al. 1992). A subsequent study found that the MECP1 complex contains the MBD2 polypeptide as the protein that directly interacts with CpG-methylated DNA (Ng et al. 1999). The MBD family consists of five members: MBD1, MBD2, MBD3, MBD4 and MECP2 (Hendrich and Bird 1998). Most studies classified MBD3 as a protein that is incapable of binding to methyl-CpG

sequences due to an amino acid change in the DNA-binding domain, although modification of the binding conditions in vitro or within the chromatin milieu may possibly lead to different outcomes. MBD4 is primarily a DNA base excision repair protein that removes thymine bases from T/G mispairs, which may arise after hydrolytic deamination of 5-methylcytosine (Hendrich and Bird 1998). The protein contains a functional DNA glycosylase domain in addition to the methyl-binding domain (MBD). MBD1, MBD2 and MECP2 are considered transcriptional repressors that are part of larger repressor complexes also containing histone deacetylase activities (Zhang et al. 1999). Thus, MBD2 is a component of the NURD complex and functions in gene silencing (Zhang et al. 1999). Mutations in MBD proteins occur rarely in human diseases with the exception of *MECP2*, which is more frequently mutated. Mutations in the X-linked gene *MECP2* in Rett syndrome were first described by Amir et al. (1999). Rett syndrome is one of the most common neurodevelopmental disorders with an incidence of about 1 in 10,000 individuals. When Rett syndrome carriers reach 6–18 months of age, they rapidly develop signs of severe autism, microcephaly, lack of speech and ataxia. Rett syndrome is described as an X-linked dominant disorder, and as males have only one copy of the X chromosome, they rarely survive. While females have two X chromosomes, the (mostly random) inactivation of one of the two chromosomes results in only a single active copy in somatic tissues. This process leads to a cellular mosaicism in disease carriers. Skewed X-inactivation can change the proportion of cells expressing the mutated MECP2 and can therefore affect disease penetrance.

Many of the mutations occur in the methyl-CpG-binding domain of the MECP2 protein. The molecular mechanism of how these mutations lead to the neurodevelopmental syndromes is still under intense investigation (Shah and Bird 2017). Mouse models have been instrumental in understanding the syndrome. Rather than affecting a small number of specific genes, the current understanding is that MECP2 acts as a genome-wide transcriptional repressor that modulates gene expression in a DNA-methylation-dependent manner. This regulation may be particularly important for long, neuron-specific genes that carry extensive CpG methylation in their gene bodies. Deficiency of MECP2 may therefore have a subtle but widespread effect on gene control in neurons, explaining the difficulty in pinning down specific molecular pathways that are disabled in Rett syndrome (Shah and Bird 2017).

Other methyl-CpG-binding proteins include the proteins UHRF1 and UHRF2, which contain the methyl-CpG-recognition domain referred to as the SRA domain. UHRF1 is a methylation maintenance factor, which is essential for maintaining cytosine methylation in mammalian cells (Jurkowska and Jeltsch 2016). Deletion of UHRF1 leads to a loss of CpG methylation (Bostick et al. 2007). The protein directly interacts with DNMT1 and promotes the maintenance function of DNMT1 at hemimethylated CpG sites in chromatin (Harrison et al. 2016; Jurkowska and Jeltsch 2016). Even though it has a similar domain architecture as UHRF1, much less is known about the biological function of UHRF2.

Certain zinc finger (ZnF) proteins have the ability to bind to CpG-methylated DNA. The first protein of this type identified was KAISO (also known as ZBTB33)

(Prokhortchouk et al. 2001). Later, other proteins of the ZnF BTB/POZ (ZBTB) family were also characterized as 5mC-binding proteins. These include ZBTB2 (Lafaye et al. 2014), ZBTB4 (Filion et al. 2006) and ZBTB38 (Filion et al. 2006). The binding of zinc finger proteins to 5mC-containing DNA generally occurs with an extended DNA sequence specificity beyond the CpG dinucleotide. These proteins may play unique functional roles, perhaps akin to transcription factors but with a repressive mode of action. Many ZBTB family proteins have so far not been characterized in detail. Based on a common recognition mode for 5-methylcytosine that involves a 5mC-Arg-G triad, it was proposed that there are probably many other ZnF proteins that can bind to methylated DNA sequences (Liu et al. 2013). Another class of transcription factors that prefer methylated DNA substrates are proteins of the extended homeodomain family (Yin et al. 2017). It still remains unclear how many of these factors can bind to nucleosomal templates.

Mutations in a few members of the ZBTB transcription factor class have been identified and at least one of them is likely involved in the control of DNA methylation patterns. Mutations in *ZBTB24* are associated with immunodeficiency, centromeric instability and facial anomalies syndrome type 2 (de Greef et al. 2011). These data suggested that ZBTB24 is somehow necessary for establishing or maintaining DNA methylation of juxtacentromeric DNA regions. However, the precise mechanistic role of ZBTB24 in this disease and in the DNA methylation pathway is still unknown.

3.8 5-Methylcytosine Oxidases, the TET Proteins

TET proteins are 5mC oxidases that act upon CpG-methylated DNA and utilize Fe^{2+}, alpha-ketoglutarate and oxygen as cofactors. Their activity can be greatly stimulated by the addition of ascorbate (vitamin C), which presumably provides a means to maintain iron in the Fe^{2+} state (Minor et al. 2013; Yin et al. 2013; Hore et al. 2016). The TET enzymes produce 5hmC as the first reaction product. In vitro, they can effectively proceed through the oxidation cascade (Fig. 3.2) to form 5-formylcytosine (5fC) and then 5-carboxylcytosine (5caC). However, in cells or tissues, the oxidation step is probably highly regulated (Song and Pfeifer 2016).

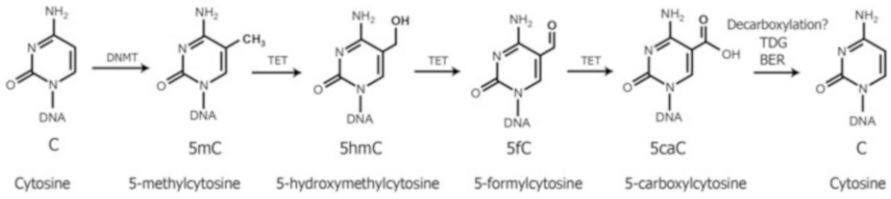

Fig. 3.2 5-methylcytosine oxidation and DNA demethylation initiated by TET proteins. Successive oxidation steps carried out by TET activities produce formylated and carboxylated cytosines that can be removed by thymine DNA glycosylase (TDG) and base excision repair (BER). The DNA decarboxylase step is hypothetical

Isotope labelling experiments have shown that 5hmC is mostly a stable DNA base (Bachman et al. 2014) and the reaction rarely proceeds to the carboxyl derivative. The further oxidation of 5hmC is likely controlled by modulation of TET activity, exemplified by interaction of TETs with other proteins such as SALL4 (Xiong et al. 2016).

Importantly, 5fC and 5caC are capable of blocking RNA polymerases (Kellinger et al. 2012; Wang et al. 2015) and DNA polymerases (unpublished data). For these reasons, these oxidation products may produce a DNA damage response and are subject to specific DNA repair processes. He et al. showed that 5fC and 5caC are substrates for thymine DNA glycosylase (TDG), which was formerly known to remove thymines from T/G mispairs. However, 5caC is likely a better substrate for TDG. After removal of the modified base, the resulting abasic sites in DNA are excised and corrected by the standard base excision repair pathways. Once this process is completed, the originally methylated cytosine is converted back to unmethylated cytosine by the TET/TDG pathway (Fig. 3.2). This chain of events is called 'active DNA demethylation' and can operate in the absence of DNA replication. The alternative process, which may be more widely used during development, is 'passive DNA demethylation'. This pathway is outlined in Fig. 3.1b and is often underappreciated. A TET enzyme introduces a 5hmC residue at a methylated CpG site. When DNA replication occurs, the hydroxymethylated cytosine on the parental DNA strand is a very poor substrate for the maintenance DNA methyltransferase DNMT1 (Valinluck and Sowers 2007; Hashimoto et al. 2012). Complete DNA demethylation ensues rapidly after several DNA replication cycles (Fig. 3.1b).

It has been difficult to demonstrate when and where exactly during organismal development the TET/TDG pathway is operative. For example, during reprogramming in fertilized oocytes, when the paternal genome preferentially undergoes 5mC oxidation (Iqbal et al. 2011; Wossidlo et al. 2011), there is little evidence for the involvement of TDG, which does not even seem to be expressed in oocytes or zygotes (Guo et al. 2014). It is also difficult to accept, in general, the concept of genome-wide 5mC oxidation involving base excision repair. This process would generate thousands or even millions of DNA strand breaks with the inherent risk of producing lethal double-strand breaks. On the other hand, although numerous searches have been performed to identify a 5caC DNA decarboxylase, including the author's laboratory, no candidate enzyme has so far been identified. Decarboxylation would avoid the occurrence of strand breaks and would appear as a more logical mechanism for completing the TET-initiated active demethylation pathway.

3.9 Biological Role of 5-Hydroxymethylcytosine in Development and Disease

Unfortunately, the biological function of 5hmC is still largely obscure. As a stable DNA base, it is unlikely to solely serve the purpose of being an intermediate in the DNA demethylation pathway (Fig. 3.2). As described below, the search for proteins

specifically recognizing 5hmC has not revealed much at this time. The highest levels of 5hmC are observed in the brain, particularly in neurons with considerably lower levels in glial cells (Münzel et al. 2010). The TET family of proteins are essential for cellular differentiation pathways. Inactivation of all three *Tet* genes in mice leads to a gastrulation defect (Dai et al. 2016). The most severe phenotype of single *Tet* gene deletion is for *Tet3* (Gu et al. 2011) with *Tet3* knockout mice rarely surviving beyond birth. TET3 and TET2 are essential for neuronal differentiation in the brain (Hahn et al. 2013). These studies tell us that TET-induced formation of 5hmC (or the complete demethylation cycle) is a crucial event, but these data still do not elucidate the mechanisms of how the oxidized 5mC bases operate in cellular physiology.

One other interesting observation is the dramatic loss of 5hmC in all types of human tumours analysed (Jin et al. 2011a). In cancer tissues, 5hmC levels are often reduced by 80–90%. While this may be useful as a biomarker for cancer, we do not exactly understand why there is such a dramatic loss of the modified base. According to data collected by The Cancer Genome Atlas (TCGA), expression levels of *TET* genes are not drastically downregulated in cancer tissue and are even increased in some types of tumours. Cell proliferation may be one reason why 5hmC is diminished in cancer. Whereas DNMT proteins can largely maintain 5mC in tumours with only a relatively small loss of 5mC occurring, the rapid division cycles may lead to a more severe 'lagging behind' of TET activity that is unable to effectively restore 5hmC modifications. One other possibility is that TET activity is impaired in tumour cells by a shift in co-factor availability, for example, a lack of ascorbic acid or by other mechanisms. Future studies are needed to analyse how 5hmC becomes depleted in tumours and will also need to dissect a possible causative role of 5hmC loss in carcinogenesis.

3.10 Mutations in TET2

TET1 and TET3 are rarely mutated in human diseases including cancer. Initially, sequencing of *TET2* in 408 haematopoietic tumour samples detected 68 somatic mutations in the *TET2* gene in myeloproliferative neoplasms (7.6%), chronic myelomonocytic leukaemia (42%) and acute myeloid leukaemia (12%) (Abdel-Wahab et al. 2009). Numerous additional sequencing studies have confirmed *TET2* as a commonly mutated gene in these diseases but the gene is rarely mutated in solid tumours. These *TET2* mutations are loss-of-function point mutations or deletions that commonly disrupt the C-terminal catalytic domain or are nonsense or frameshift mutations in the N-terminal region leading to premature truncation of protein synthesis. They occur usually on one of the two alleles and more rarely affect both alleles. Mutations in *TET2* occur during the process of clonal haematopoiesis during aging and are often acquired at the stages of haematopoietic stem or early progenitor cells (Scourzic et al. 2015; Bowman and Levine 2017). Mutations in isocitrate dehydrogenase enzymes (IDH1 and IDH2) provide an alternative path towards impeding the function of TET enzymes. The mutant IDH proteins, instead of forming the TET co-factor alpha-ketoglutarate, produce an oncogenic metabolite,

2-hydroxyglutarate (Dang et al. 2009), which is a competitive inhibitor of TET enzymes (Xu et al. 2011). Mutations in IDH1/2 and TET2 are often mutually exclusive. *Tet2*-deficient mouse models have been used to recapitulate the human haematopoietic diseases. These mice are characterized by expansion of myeloid progenitor cells leading to the accumulation of pre-leukaemic cell clones (Moran-Crusio et al. 2011). Interestingly, mutations in TET2 and in DNMT3A can co-occur in human leukaemia patients. This situation seems counterintuitive inasmuch as one enzyme is a DNA methyltransferase and the other is a DNA demethylase. However, mouse models have shown that mutations in *Dnmt3a* and *Tet2* can cooperate leading to a more severe phenotype (Zhang et al. 2016a). The result of *TET2* mutation is expected to be a disturbance of the balance of methylation and demethylation reactions with a more likely outcome of DNA hypermethylation. However, such studies with human patient material are difficult to interpret because haematological malignancies often are characterized by mutations in multiple epigenetic modifier genes. More research is needed to precisely dissect the epigenomic effects of TET2 mutations in leukaemias and myelodysplastic syndrome. If TET2 mutations indeed are linked to DNA hypermethylation, epigenetic therapies aimed at inhibiting DNA methylation may provide a fruitful therapeutic route for management of these diseases.

3.11 Proteins That Bind to Oxidized 5-Methylcytosine Derivatives

A mass spectrometry-based approach was used to identify proteins that may bind to oxidized 5mC derivatives (Iurlaro et al. 2013; Spruijt et al. 2013). Although a number of candidate proteins for binding to 5hmC, 5fC or 5caC were found, verification of their binding and any physiological role of the binding has so far been limited. Initial studies reported MBD3 and MECP2 as proteins binding to 5hmC with some degree of preference (Yildirim et al. 2011; Mellen et al. 2012). However, these studies have not been subsequently substantiated by other laboratories (Hashimoto et al. 2012; Iurlaro et al. 2013; Otani et al. 2013; Spruijt et al. 2013; Khrapunov et al. 2014). Currently, perhaps the best evidence for existence of a 5hmC-specific reader protein comes from a structural study of UHRF2 (Zhou et al. 2014). UHRF2 is a close homologue of UHRF1, a protein involved in maintenance DNA methylation. However, deletion of *Uhrf2* in mice does not lead to a major phenotype except for some relatively mild neurological defects including memory impairment (Chen et al. 2017a, b; Liu et al. 2017). On the other hand, mice lacking UHRF1 protein show developmental arrest and early embryo lethality similar to *Dnmt1*$^{-/-}$ embryos (Sharif et al. 2007).

One other possible scenario is that oxidized 5mC bases, predominantly 5hmC, are simply negative marks that block the binding of the MBD family proteins and of other proteins capable of binding to methylated CpG sequences. The lack of binding of MBD1, MBD2 and MBD4 to hydroxymethylated CpG sites has been reported (Jin et al. 2010). The functional output of 5hmC-repelled MBD repressor complexes

is expected to be the induction or maintenance of an active state of gene expression. Of interest in that regard is that 5hmC is commonly found near enhancer regions, where it signals the presence of an active or potentially active enhancer (Hon et al. 2014). Other than at enhancers, 5hmC is found at its highest levels in gene body regions of expressed, tissue-specific genes (Stroud et al. 2011; Szulwach et al. 2011; Nestor et al. 2012; Hahn et al. 2013). Its predominant presence in these genomic sites along transcribed genes suggests that TET proteins have the ability to travel with the RNA polymerase elongation complex. Genes with extensive 5hmC coverage of their intragenic regions are expressed at higher levels than genes with little 5hmC deposition (Song et al. 2011; Nestor et al. 2012; Hahn et al. 2013). The reason for this correlation is presently unclear. The correlation between gene expression levels and 5hmC is even greater than a similar correlation between gene expression and 5mC (Jin et al. 2011b). Additional work is required to determine the mechanistic connections between 5hmC in gene bodies and enhanced expression of tissue-specific (but not housekeeping) genes.

Mass spectrometry has identified a group of proteins capable of preferential interaction with 5-formylcytosine and 5-carboxylcytosine (Iurlaro et al. 2013; Spruijt et al. 2013). So far, 5fC-specific readers have not been confirmed in detail except for the repair enzyme TDG (Song and Pfeifer 2016). The initial proposal that 5fC changes the global structure of the DNA double helix (Raiber et al. 2015) has not been verified later (Hardwick et al. 2017). On the other hand, there are now several well-characterized interactions of specific proteins with 5caC. These include TDG, RNA polymerase II, which is arrested by 5caC in the template (Xue and Xu 2015), the CXXC domain of TET3 (Jin et al. 2016), and the MYC binding partner MAX, which binds to 5caC within 5′-CACGTG-3′ sequences as well as to unmodified C within these target sites (Wang et al. 2017). The low level of 5fC and 5caC observable in cells or tissues implies that these interactions are likely only transient, but it still remains as a surprise that there are such few reader proteins that selectively interact with the much more abundant DNA base 5hmC. Future investigations will likely provide additional insights into the functional roles and recognition of oxidized 5mC bases.

References

Abdel-Wahab O, Mullally A, Hedvat C, Garcia-Manero G, Patel J, Wadleigh M, Malinge S, Yao J, Kilpivaara O, Bhat R, Huberman K, Thomas S, Dolgalev I, Heguy A, Paietta E, Le Beau MM, Beran M, Tallman MS, Ebert BL, Kantarjian HM, Stone RM, Gilliland DG, Crispino JD, Levine RL (2009) Genetic characterization of TET1, TET2, and TET3 alterations in myeloid malignancies. Blood 114:144–147

Amir RE, Van den Veyver IB, Wan M, Tran CQ, Francke U, Zoghbi HY (1999) Rett syndrome is caused by mutations in X-linked MECP2, encoding methyl-CpG-binding protein 2. Nat Genet 23:185–188

Aran D, Toperoff G, Rosenberg M, Hellman A (2011) Replication timing-related and gene body-specific methylation of active human genes. Hum Mol Genet 20:670–680

Bachman M, Uribe-Lewis S, Yang X, Williams M, Murrell A, Balasubramanian S (2014) 5-hydroxymethylcytosine is a predominantly stable DNA modification. Nat Chem 6:1049–1055

Baubec T, Colombo DF, Wirbelauer C, Schmidt J, Burger L, Krebs AR, Akalin A, Schubeler D (2015) Genomic profiling of DNA methyltransferases reveals a role for DNMT3B in genic methylation. Nature 520:243–247

Bestor TH, Ingram VM (1985) Growth-dependent expression of multiple species of DNA methyltransferase in murine erythroleukemia cells. Proc Natl Acad Sci U S A 82:2674–2678

Bird AP (1995) Gene number, noise reduction and biological complexity. Trends Genet 11:94–100

Boller S, Ramamoorthy S, Akbas D, Nechanitzky R, Burger L, Murr R, Schubeler D, Grosschedl R (2016) Pioneering activity of the C-terminal domain of EBF1 shapes the chromatin landscape for B cell programming. Immunity 44:527–541

Bostick M, Kim JK, Esteve PO, Clark A, Pradhan S, Jacobsen SE (2007) UHRF1 plays a role in maintaining DNA methylation in mammalian cells. Science 317:1760–1764

Bourc'his D, Xu GL, Lin CS, Bollman B, Bestor TH (2001) Dnmt3L and the establishment of maternal genomic imprints. Science 294:2536–2539

Bowman RL, Levine RL (2017) TET2 in normal and malignant hematopoiesis. Cold Spring Harb Perspect Med 7:a026518

Brockdorff N (2017) Polycomb complexes in X chromosome inactivation. Philos Trans R Soc Lond B Biol Sci 372:20170021

Brunetti L, Gundry MC, Goodell MA (2017) DNMT3A in leukemia. Cold Spring Harb Perspect Med 7:a030320

Buck-Koehntop BA, Defossez PA (2013) On how mammalian transcription factors recognize methylated DNA. Epigenetics 8:131–137

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 99:16916–16921

Chen R, Zhang Q, Duan X, York P, Chen GD, Yin P, Zhu H, Xu M, Chen P, Wu Q, Li D, Samarut J, Xu G, Zhang P, Cao X, Li J, Wong J (2017a) The 5-hydroxymethylcytosine (5hmC) reader UHRF2 is required for normal levels of 5hmC in mouse adult brain and spatial learning and memory. J Biol Chem 292:4533–4543

Chen XR, Sun SC, Teng SW, Li L, Bie YF, Yu H, Li DL, Chen ZY, Wang Y (2017b) Uhrf2 deletion impairs the formation of hippocampus-dependent memory by changing the structure of the dentate gyrus. Brain Struct Funct 223:609–618

Dai HQ, Wang BA, Yang L, Chen JJ, Zhu GC, Sun ML, Ge H, Wang R, Chapman DL, Tang F, Sun X, Xu GL (2016) TET-mediated DNA demethylation controls gastrulation by regulating Lefty-Nodal signalling. Nature 538:528–532

Dang L, White DW, Gross S, Bennett BD, Bittinger MA, Driggers EM, Fantin VR, Jang HG, Jin S, Keenan MC, Marks KM, Prins RM, Ward PS, Yen KE, Liau LM, Rabinowitz JD, Cantley LC, Thompson CB, Vander Heiden MG, Su SM (2009) Cancer-associated IDH1 mutations produce 2-hydroxyglutarate. Nature 462:739–744

Daxinger L, Tapscott SJ, van der Maarel SM (2015) Genetic and epigenetic contributors to FSHD. Curr Opin Genet Dev 33:56–61

de Greef JC, Wang J, Balog J, den Dunnen JT, Frants RR, Straasheijm KR, Aytekin C, van der Burg M, Duprez L, Ferster A, Gennery AR, Gimelli G, Reisli I, Schuetz C, Schulz A, Smeets D, Sznajer Y, Wijmenga C, van Eggermond MC, van Ostaijen-Ten Dam MM, Lankester AC, van Tol MJD, van den Elsen PJ, Weemaes CM, van der Maarel SM (2011) Mutations in ZBTB24 are associated with immunodeficiency, centromeric instability, and facial anomalies syndrome type 2. Am J Hum Genet 88:796–804

Delaval K, Feil R (2004) Epigenetic regulation of mammalian genomic imprinting. Curr Opin Genet Dev 14:188–195

Drahovsky D, Morris NR (1971) Mechanism of action of rat liver DNA methylase. I Interaction with double-stranded methyl-acceptor DNA. J Mol Biol 57:475–489

Dunwell TL, McGuffin LJ, Dunwell JM, Pfeifer GP (2013) The mysterious presence of a 5-methylcytosine oxidase in the Drosophila genome: possible explanations. Cell Cycle 12:3357–3365

Duymich CE, Charlet J, Yang X, Jones PA, Liang G (2016) DNMT3B isoforms without catalytic activity stimulate gene body methylation as accessory proteins in somatic cells. Nat Commun 7:11453

Easwaran H, Tsai HC, Baylin SB (2014) Cancer epigenetics: tumor heterogeneity, plasticity of stem-like states, and drug resistance. Mol Cell 54:716–727

Ehrlich M (2009) DNA hypomethylation in cancer cells. Epigenomics 1:239–259

Filion GJ, Zhenilo S, Salozhin S, Yamada D, Prokhortchouk E, Defossez PA (2006) A family of human zinc finger proteins that bind methylated DNA and repress transcription. Mol Cell Biol 26:169–181

Gal-Yam EN, Egger G, Iniguez L, Holster H, Einarsson S, Zhang X, Lin JC, Liang G, Jones PA, Tanay A (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

Goll MG, Kirpekar F, Maggert KA, Yoder JA, Hsieh CL, Zhang X, Golic KG, Jacobsen SE, Bestor TH (2006) Methylation of tRNAAsp by the DNA methyltransferase homolog Dnmt2. Science 311:395–398

Groh S, Schotta G (2017) Silencing of endogenous retroviruses by heterochromatin. Cell Mol Life Sci 74:2055–2065

Gu TP, Guo F, Yang H, Wu HP, Xu GF, Liu W, Xie ZG, Shi L, He X, Jin SG, Iqbal K, Shi YG, Deng Z, Szabo PE, Pfeifer GP, Li J, Xu GL (2011) The role of Tet3 DNA dioxygenase in epigenetic reprogramming by oocytes. Nature 477:606–610

Guo F, Li X, Liang D, Li T, Zhu P, Guo H, Wu X, Wen L, Gu TP, Hu B, Walsh CP, Li J, Tang F, Xu GL (2014) Active and passive demethylation of male and female pronuclear DNA in the mammalian zygote. Cell Stem Cell 15:447–459

Hahn MA, Qiu R, Wu X, Li AX, Zhang H, Wang J, Jui J, Jin SG, Jiang Y, Pfeifer GP, Lu Q (2013) Dynamics of 5-hydroxymethylcytosine and chromatin marks in Mammalian neurogenesis. Cell Rep 3:291–300

Hansen RS, Wijmenga C, Luo P, Stanek AM, Canfield TK, Weemaes CM, Gartler SM (1999) The DNMT3B DNA methyltransferase gene is mutated in the ICF immunodeficiency syndrome. Proc Natl Acad Sci U S A 96:14412–14417

Hardwick JS, Ptchelkine D, El-Sagheer AH, Tear I, Singleton D, Phillips SEV, Lane AN, Brown T (2017) 5-Formylcytosine does not change the global structure of DNA. Nat Struct Mol Biol 24:544–552

Harrison JS, Cornett EM, Goldfarb D, DaRosa PA, Li ZM, Yan F, Dickson BM, Guo AH, Cantu DV, Kaustov L, Brown PJ, Arrowsmith CH, Erie DA, Major MB, Klevit RE, Krajewski K, Kuhlman B, Strahl BD, Rothbart SB (2016) Hemi-methylated DNA regulates DNA methylation inheritance through allosteric activation of H3 ubiquitylation by UHRF1. Elife 5:e17101

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

Heard E (2004) Recent advances in X-chromosome inactivation. Curr Opin Cell Biol 16:247–255

Hellman A, Chess A (2007) Gene body-specific methylation on the active X chromosome. Science 315:1141–1143

Hendrich B, Bird A (1998) Identification and characterization of a family of mammalian methyl-CpG binding proteins. Mol Cell Biol 18:6538–6547

Holliday R, Pugh JE (1975) DNA modification mechanisms and gene activity during development. Science 187:226–232

Hon GC, Song CX, Du T, Jin F, Selvaraj S, Lee AY, Yen CA, Ye Z, Mao SQ, Wang BA, Kuan S, Edsall LE, Zhao BS, Xu GL, He C, Ren B (2014) 5mC oxidation by Tet2 modulates enhancer activity and timing of transcriptome reprogramming during differentiation. Mol Cell 56:286–297

Hore TA, von Meyenn F, Ravichandran M, Bachman M, Ficz G, Oxley D, Santos F, Balasubramanian S, Jurkowski TP, Reik W (2016) Retinol and ascorbate drive erasure of

epigenetic memory and enhance reprogramming to naive pluripotency by complementary mechanisms. Proc Natl Acad Sci U S A 113:12202–12207
Hotchkiss RD (1948) The quantitative separation of purines, pyrimidines, and nucleosides by paper chromatography. J Biol Chem 175:315–332
Huang K, Wu Z, Liu Z, Hu G, Yu J, Chang KH, Kim KP, Le T, Faull KF, Rao N, Gennery A, Xue Z, Wang CY, Pellegrini M, Fan G (2014) Selective demethylation and altered gene expression are associated with ICF syndrome in human-induced pluripotent stem cells and mesenchymal stem cells. Hum Mol Genet 23:6448–6457
Iqbal K, Jin SG, Pfeifer GP, Szabo PE (2011) Reprogramming of the paternal genome upon fertilization involves genome-wide oxidation of 5-methylcytosine. Proc Natl Acad Sci U S A 108:3642–3647
Ito S, D'Alessio AC, Taranova OV, Hong K, Sowers LC, Zhang Y (2010) Role of Tet proteins in 5mC to 5hmC conversion, ES-cell self-renewal and inner cell mass specification. Nature 466:1129–1133
Iurlaro M, Ficz G, Oxley D, Raiber EA, Bachman M, Booth MJ, Andrews S, Balasubramanian S, Reik W (2013) A screen for hydroxymethylcytosine and formylcytosine binding proteins suggests functions in transcription and chromatin regulation. Genome Biol 14:R119
Jin SG, Kadam S, Pfeifer GP (2010) Examination of the specificity of DNA methylation profiling techniques towards 5-methylcytosine and 5-hydroxymethylcytosine. Nucleic Acids Res 38: e125
Jin SG, Jiang Y, Qiu R, Rauch TA, Wang Y, Schackert G, Krex D, Lu Q, Pfeifer GP (2011a) 5-Hydroxymethylcytosine is strongly depleted in human cancers but its levels do not correlate with IDH1 mutations. Cancer Res 71:7360–7365
Jin SG, Wu X, Li AX, Pfeifer GP (2011b) Genomic mapping of 5-hydroxymethylcytosine in the human brain. Nucleic Acids Res 39:5015–5024
Jin SG, Zhang ZM, Dunwell TL, Harter MR, Wu X, Johnson J, Li Z, Liu J, Szabo PE, Lu Q, Xu GL, Song J, Pfeifer GP (2016) Tet3 reads 5-carboxylcytosine through Its CXXC domain and is a potential guardian against neurodegeneration. Cell Rep 14:493–505
Johnson LM, Bostick M, Zhang X, Kraft E, Henderson I, Callis J, Jacobsen SE (2007) The SRA methyl-cytosine-binding domain links DNA and histone methylation. Curr Biol 17:379–384
Jung M, Pfeifer GP (2015) Aging and DNA methylation. BMC Biol 13:7
Jurkowska RZ, Jeltsch A (2016) Enzymology of mammalian DNA methyltransferases. Adv Exp Med Biol 945:87–122
Kalari S, Jung M, Kernstine KH, Takahashi T, Pfeifer GP (2013) The DNA methylation landscape of small cell lung cancer suggests a differentiation defect of neuroendocrine cells. Oncogene 32:3559–3568
Kellinger MW, Song CX, Chong J, Lu XY, He C, Wang D (2012) 5-formylcytosine and 5-carboxylcytosine reduce the rate and substrate specificity of RNA polymerase II transcription. Nat Struct Mol Biol 19:831–833
Khrapunov S, Warren C, Cheng H, Berko ER, Greally JM, Brenowitz M (2014) Unusual characteristics of the DNA binding domain of epigenetic regulatory protein MeCP2 determine its binding specificity. Biochemistry 53:3379–3391
Klein CJ, Botuyan MV, Wu Y, Ward CJ, Nicholson GA, Hammans S, Hojo K, Yamanishi H, Karpf AR, Wallace DC, Simon M, Lander C, Boardman LA, Cunningham JM, Smith GE, Litchy WJ, Boes B, Atkinson EJ, Middha S, B Dyck PJ, Parisi JE, Mer G, Smith DI, Dyck PJ (2011) Mutations in DNMT1 cause hereditary sensory neuropathy with dementia and hearing loss. Nat Genet 43:595–600
Kriaucionis S, Heintz N (2009) The nuclear DNA base 5-hydroxymethylcytosine is present in Purkinje neurons and the brain. Science 324:929–930
Lafaye C, Barbier E, Miscioscia A, Saint-Pierre C, Kraut A, Coute Y, Plo I, Gasparutto D, Ravanat JL, Breton J (2014) DNA binding of the p21 repressor ZBTB2 is inhibited by cytosine hydroxymethylation. Biochem Biophys Res Commun 446:341–346

Lewis JD, Meehan RR, Henzel WJ, Maurer-Fogy I, Jeppesen P, Klein F, Bird A (1992) Purification, sequence, and cellular localization of a novel chromosomal protein that binds to methylated DNA. Cell 69:905–914

Li E, Bestor TH, Jaenisch R (1992) Targeted mutation of the DNA methyltransferase gene results in embryonic lethality. Cell 69:915–926

Liu Y, Zhang X, Blumenthal RM, Cheng X (2013) A common mode of recognition for methylated CpG. Trends Biochem Sci 38:177–183

Liu Y, Zhang B, Meng X, Korn MJ, Parent JM, Lu LY, Yu X (2017) UHRF2 regulates local 5-methylcytosine and suppresses spontaneous seizures. Epigenetics 12:551–560

Long HK, Blackledge NP, Klose RJ (2013) ZF-CxxC domain-containing proteins, CpG islands and the chromatin connection. Biochem Soc Trans 41:727–740

Lorsbach RB, Moore J, Mathew S, Raimondi SC, Mukatira ST, Downing JR (2003) TET1, a member of a novel protein family, is fused to MLL in acute myeloid leukemia containing the t (10,11)(q22;q23). Leukemia 17:637–641

Maegawa S, Hinkal G, Kim HS, Shen L, Zhang L, Zhang J, Zhang N, Liang S, Donehower LA, Issa J-P (2010) Widespread and tissue specific age-related DNA methylation changes in mice. Genome Res 20:332–340

Mann JR, Szabo PE, Reed MR, Singer-Sam J (2000) Methylated DNA sequences in genomic imprinting. Crit Rev Eukaryot Gene Expr 10:241–257

Meehan RR, Lewis JD, McKay S, Kleiner EL, Bird AP (1989) Identification of a mammalian protein that binds specifically to DNA containing methylated CpGs. Cell 58:499–507

Mellen M, Ayata P, Dewell S, Kriaucionis S, Heintz N (2012) MeCP2 binds to 5hmC enriched within active genes and accessible chromatin in the nervous system. Cell 151:1417–1430

Minor EA, Court BL, Young JI, Wang G (2013) Ascorbate induces ten-eleven translocation (Tet) methylcytosine dioxygenase-mediated generation of 5-hydroxymethylcytosine. J Biol Chem 288:13669–13674

Moran-Crusio K, Reavie L, Shih A, Abdel-Wahab O, Ndiaye-Lobry D, Lobry C, Figueroa ME, Vasanthakumar A, Patel J, Zhao X, Perna F, Pandey S, Madzo J, Song C, Dai Q, He C, Ibrahim S, Beran M, Zavadil J, Nimer SD, Melnick A, Godley LA, Aifantis I, Levine RL (2011) Tet2 loss leads to increased hematopoietic stem cell self-renewal and myeloid transformation. Cancer Cell 20:11–24

Münzel M, Globisch D, Bruckl T, Wagner M, Welzmiller V, Michalakis S, Muller M, Biel M, Carell T (2010) Quantification of the sixth DNA base hydroxymethylcytosine in the brain. Angew Chem Int Ed Engl 49:5375–5377

Neri F, Rapelli S, Krepelova A, Incarnato D, Parlato C, Basile G, Maldotti M, Anselmi F, Oliviero S (2017) Intragenic DNA methylation prevents spurious transcription initiation. Nature 543:72–77

Nestor CE, Ottaviano R, Reddington J, Sproul D, Reinhardt D, Dunican D, Katz E, Dixon JM, Harrison DJ, Meehan RR (2012) Tissue type is a major modifier of the 5-hydroxymethylcytosine content of human genes. Genome Res 22:467–477

Ng HH, Zhang Y, Hendrich B, Johnson CA, Turner BM, Erdjument-Bromage H, Tempst P, Reinberg D, Bird A (1999) MBD2 is a transcriptional repressor belonging to the MeCP1 histone deacetylase complex. Nat Genet 23:58–61

Ohm JE, McGarvey KM, Yu X, Cheng L, Schuebel KE, Cope L, Mohammad HP, Chen W, Daniel VC, Yu W, Berman DM, Jenuwein T, Pruitt K, Sharkis SJ, Watkins DN, Herman JG, Baylin SB (2007) A stem cell-like chromatin pattern may predispose tumor suppressor genes to DNA hypermethylation and heritable silencing. Nat Genet 39:237–242

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

Ono R, Taki T, Taketani T, Taniwaki M, Kobayashi H, Hayashi Y (2002) LCX, leukemia-associated protein with a CXXC domain, is fused to MLL in acute myeloid leukemia with trilineage dysplasia having t(10;11)(q22;q23). Cancer Res 62:4075–4080

Ooi SK, Qiu C, Bernstein E, Li K, Jia D, Yang Z, Erdjument-Bromage H, Tempst P, Lin SP, Allis CD, Cheng X, Bestor TH (2007) DNMT3L connects unmethylated lysine 4 of histone H3 to de novo methylation of DNA. Nature 448:714–717

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

Patil V, Ward RL, Hesson LB (2014) The evidence for functional non-CpG methylation in mammalian cells. Epigenetics 9:823–828

Pfeifer GP (2006) Mutagenesis at methylated CpG sequences. Curr Top Microbiol Immunol 301:259–281

Pfeifer GP, Hainaut P (2011) Next-generation sequencing: emerging lessons on the origins of human cancer. Curr Opin Oncol 23:62–68

Pfeifer GP, Spiess E, Grunwald S, Boehm TL, Drahovsky D (1985) Mouse DNA-cytosine-5-methyltransferase: sequence specificity of the methylation reaction and electron microscopy of enzyme-DNA complexes. EMBO J 4:2879–2884

Prokhortchouk A, Hendrich B, Jorgensen H, Ruzov A, Wilm M, Georgiev G, Bird A, Prokhortchouk E (2001) The p120 catenin partner Kaiso is a DNA methylation-dependent transcriptional repressor. Genes Dev 15:1613–1618

Raiber EA, Murat P, Chirgadze DY, Beraldi D, Luisi BF, Balasubramanian S (2015) 5-Formylcytosine alters the structure of the DNA double helix. Nat Struct Mol Biol 22:44–49

Rauch T, Li H, Wu X, Pfeifer GP (2006) MIRA-assisted microarray analysis, a new technology for the determination of DNA methylation patterns, identifies frequent methylation of homeodomain-containing genes in lung cancer cells. Cancer Res 66:7939–7947

Rauch TA, Zhong X, Wu X, Wang M, Kernstine KH, Wang Z, Riggs AD, Pfeifer GP (2008) High-resolution mapping of DNA hypermethylation and hypomethylation in lung cancer. Proc Natl Acad Sci U S A 105:252–257

Rauch TA, Wu X, Zhong X, Riggs AD, Pfeifer GP (2009) A human B cell methylome at 100-base pair resolution. Proc Natl Acad Sci U S A 106:671–678

Riggs AD (1975) X inactivation, differentiation, and DNA methylation. Cytogenet Cell Genet 14:9–25

Riggs AD, Pfeifer GP (1992) X-chromosome inactivation and cell memory. Trends Genet 8:169–174

Schlesinger Y, Straussman R, Keshet I, Farkash S, Hecht M, Zimmerman J, Eden E, Yakhini Z, Ben-Shushan E, Reubinoff BE, Bergman Y, Simon I, Cedar H (2007) Polycomb-mediated methylation on Lys27 of histone H3 pre-marks genes for de novo methylation in cancer. Nat Genet 39:232–236

Schubeler D (2015) Function and information content of DNA methylation. Nature 517:321–326

Schuermann D, Weber AR, Schar P (2016) Active DNA demethylation by DNA repair: facts and uncertainties. DNA Repair (Amst) 44:92–102

Scourzic L, Mouly E, Bernard OA (2015) TET proteins and the control of cytosine demethylation in cancer. Genome Med 7:9

Shah RR, Bird AP (2017) MeCP2 mutations: progress towards understanding and treating Rett syndrome. Genome Med 9:17

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 450:908–912

Simo-Riudalbas L, Diaz-Lagares A, Gatto S, Gagliardi M, Crujeiras AB, Matarazzo MR, Esteller M, Sandoval J (2015) Genome-wide DNA methylation analysis identifies novel hypomethylated non-pericentromeric genes with potential clinical implications in ICF syndrome. PLoS One 10:e0132517

Singh P, Li AX, Tran DA, Oates N, Kang ER, Wu X, Szabo PE (2013) De novo DNA methylation in the male germ line occurs by default but is excluded at sites of H3K4 methylation. Cell Rep 4:205–219

Song J, Pfeifer GP (2016) Are there specific readers of oxidized 5-methylcytosine bases? Bioessays 38:1038–1047

Song CX, Szulwach KE, Fu Y, Dai Q, Yi C, Li X, Li Y, Chen CH, Zhang W, Jian X, Wang J, Zhang L, Looney TJ, Zhang B, Godley LA, Hicks LM, Lahn BT, Jin P, He C (2011) Selective chemical labeling reveals the genome-wide distribution of 5-hydroxymethylcytosine. Nat Biotechnol 29:68–72

Spencer DH, Russler-Germain DA, Ketkar S, Helton NM, Lamprecht TL, Fulton RS, Fronick CC, O'Laughlin M, Heath SE, Shinawi M, Westervelt P, Payton JE, Wartman LD, Welch JS, Wilson RK, Walter MJ, Link DC, DiPersio JF, Ley TJ (2017) CpG island hypermethylation mediated by DNMT3A is a consequence of AML progression. Cell 168:801–816 e813

Spruijt CG, Gnerlich F, Smits AH, Pfaffeneder T, Jansen PW, Bauer C, Munzel M, Wagner M, Muller M, Khan F, Eberl HC, Mensinga A, Brinkman AB, Lephikov K, Muller U, Walter J, Boelens R, van Ingen H, Leonhardt H, Carell T, Vermeulen M (2013) Dynamic readers for 5-(hydroxy)methylcytosine and its oxidized derivatives. Cell 152:1146–1159

Stadler MB, Murr R, Burger L, Ivanek R, Lienert F, Scholer A, van Nimwegen E, Wirbelauer C, Oakeley EJ, Gaidatzis D, Tiwari VK, Schubeler D (2011) DNA-binding factors shape the mouse methylome at distal regulatory regions. Nature 480:490–495

Stroud H, Feng S, Morey Kinney S, Pradhan S, Jacobsen SE (2011) 5-Hydroxymethylcytosine is associated with enhancers and gene bodies in human embryonic stem cells. Genome Biol 12: R54

Sun Z, Wu Y, Ordog T, Baheti S, Nie J, Duan X, Hojo K, Kocher JP, Dyck PJ, Klein CJ (2014) Aberrant signature methylome by DNMT1 hot spot mutation in hereditary sensory and autonomic neuropathy 1E. Epigenetics 9:1184–1193

Szulwach KE, Li X, Li Y, Song CX, Wu H, Dai Q, Irier H, Upadhyay AK, Gearing M, Levey AI, Vasanthakumar A, Godley LA, Chang Q, Cheng X, He C, Jin P (2011) 5-hmC-mediated epigenetic dynamics during postnatal neurodevelopment and aging. Nat Neurosci 14:1607–1616

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 324:930–935

Tate PH, Bird AP (1993) Effects of DNA methylation on DNA-binding proteins and gene expression. Curr Opin Genet Dev 3:226–231

Tatton-Brown K, Seal S, Ruark E, Harmer J, Ramsay E, Del Vecchio Duarte S, Zachariou A, Hanks S, O'Brien E, Aksglaede L, Baralle D, Dabir T, Gener B, Goudie D, Homfray T, Kumar A, Pilz DT, Selicorni A, Temple IK, Van Maldergem L, Yachelevich N, Childhood Overgrowth C, van Montfort R, Rahman N (2014) Mutations in the DNA methyltransferase gene DNMT3A cause an overgrowth syndrome with intellectual disability. Nat Genet 46:385–388

Turnbull JF, Adams RL (1976) DNA methylase: purification from ascites cells and the effect of various DNA substrates on its activity. Nucleic Acids Res 3:677–695

Unoki M, Nishidate T, Nakamura Y (2004) ICBP90, an E2F-1 target, recruits HDAC1 and binds to methyl-CpG through its SRA domain. Oncogene 23:7601–7610

Valinluck V, Sowers LC (2007) Endogenous cytosine damage products alter the site selectivity of human DNA maintenance methyltransferase DNMT1. Cancer Res 67:946–950

van den Boogaard ML, Lemmers R, Balog J, Wohlgemuth M, Auranen M, Mitsuhashi S, van der Vliet PJ, Straasheijm KR, van den Akker RFP, Kriek M, Laurense-Bik MEY, Raz V, van Ostaijen-Ten Dam MM, Hansson KBM, van der Kooi EL, Kiuru-Enari S, Udd B, van Tol MJD, Nishino I, Tawil R, Tapscott SJ, van Engelen BGM, van der Maarel SM (2016) Mutations in DNMT3B modify epigenetic repression of the D4Z4 repeat and the penetrance of facioscapulohumeral dystrophy. Am J Hum Genet 98:1020–1029

Wang L, Zhou Y, Xu L, Xiao R, Lu X, Chen L, Chong J, Li H, He C, Fu XD, Wang D (2015) Molecular basis for 5-carboxycytosine recognition by RNA polymerase II elongation complex. Nature 523:621–625

Wang D, Hashimoto H, Zhang X, Barwick BG, Lonial S, Boise LH, Vertino PM, Cheng X (2017) MAX is an epigenetic sensor of 5-carboxylcytosine and is altered in multiple myeloma. Nucleic Acids Res 45:2396–2407

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 2:1136–1143

Widschwendter M, Fiegl H, Egle D, Mueller-Holzner E, Spizzo G, Marth C, Weisenberger DJ, Campan M, Young J, Jacobs I, Laird PW (2007) Epigenetic stem cell signature in cancer. Nat Genet 39:157–158

Winkelmann J, Lin L, Schormair B, Kornum BR, Faraco J, Plazzi G, Melberg A, Cornelio F, Urban AE, Pizza F, Poli F, Grubert F, Wieland T, Graf E, Hallmayer J, Strom TM, Mignot E (2012) Mutations in DNMT1 cause autosomal dominant cerebellar ataxia, deafness and narcolepsy. Hum Mol Genet 21:2205–2210

Wossidlo M, Nakamura T, Lepikhov K, Marques CJ, Zakhartchenko V, Boiani M, Arand J, Nakano T, Reik W, Walter J (2011) 5-Hydroxymethylcytosine in the mammalian zygote is linked with epigenetic reprogramming. Nat Commun 2:241

Wu X, Rauch TA, Zhong X, Bennett WP, Latif F, Krex D, Pfeifer GP (2010) CpG island hypermethylation in human astrocytomas. Cancer Res 70:2718–2727

Wyatt GR (1950) Occurrence of 5-methylcytosine in nucleic acids. Nature 166:237–238

Xie W, Schultz MD, Lister R, Hou Z, Rajagopal N, Ray P, Whitaker JW, Tian S, Hawkins RD, Leung D, Yang H, Wang T, Lee AY, Swanson SA, Zhang J, Zhu Y, Kim A, Nery JR, Urich MA, Kuan S, Yen CA, Klugman S, Yu P, Suknuntha K, Propson NE, Chen H, Edsall LE, Wagner U, Li Y, Ye Z, Kulkarni A, Xuan Z, Chung WY, Chi NC, Antosiewicz-Bourget JE, Slukvin I, Stewart R, Zhang MQ, Wang W, Thomson JA, Ecker JR, Ren B (2013) Epigenomic analysis of multilineage differentiation of human embryonic stem cells. Cell 153:1134–1148

Xiong J, Zhang Z, Chen J, Huang H, Xu Y, Ding X, Zheng Y, Nishinakamura R, Xu GL, Wang H, Chen S, Gao S, Zhu B (2016) Cooperative action between SALL4A and TET proteins in stepwise oxidation of 5-methylcytosine. Mol Cell 64:913–925

Xu GL, Bestor TH, Bourc'his D, Hsieh CL, Tommerup N, Bugge M, Hulten M, Qu X, Russo JJ, Viegas-Pequignot E (1999) Chromosome instability and immunodeficiency syndrome caused by mutations in a DNA methyltransferase gene. Nature 402:187–191

Xu W, Yang H, Liu Y, Yang Y, Wang P, Kim SH, Ito S, Yang C, Wang P, Xiao MT, Liu LX, Jiang WQ, Liu J, Zhang JY, Wang B, Frye S, Zhang Y, Xu YH, Lei QY, Guan KL, Zhao SM, Xiong Y (2011) Oncometabolite 2-hydroxyglutarate is a competitive inhibitor of alpha-ketoglutarate-dependent dioxygenases. Cancer Cell 19:17–30

Xue JH, Xu GL (2015) RNA Pol II as a sensor of 5caC. Cell Res 25:1089–1090

Yildirim O, Li R, Hung JH, Chen PB, Dong X, Ee LS, Weng Z, Rando OJ, Fazzio TG (2011) Mbd3/NURD complex regulates expression of 5-hydroxymethylcytosine marked genes in embryonic stem cells. Cell 147:1498–1510

Yin R, Mao SQ, Zhao B, Chong Z, Yang Y, Zhao C, Zhang D, Huang H, Gao J, Li Z, Jiao Y, Li C, Liu S, Wu D, Gu W, Yang YG, Xu GL, Wang H (2013) Ascorbic acid enhances Tet-mediated 5-methylcytosine oxidation and promotes DNA demethylation in mammals. J Am Chem Soc 135:10396–10403

Yin Y, Morgunova E, Jolma A, Kaasinen E, Sahu B, Khund-Sayeed S, Das PK, Kivioja T, Dave K, Zhong F, Nitta KR, Taipale M, Popov A, Ginno PA, Domcke S, Yan J, Schubeler D, Vinson C, Taipale J (2017) Impact of cytosine methylation on DNA binding specificities of human transcription factors. Science 356:eaaj2239

Yoder JA, Walsh CP, Bestor TH (1997) Cytosine methylation and the ecology of intragenomic parasites. Trends Genet 13:335–340

Zhang Y, Ng HH, Erdjument-Bromage H, Tempst P, Bird A, Reinberg D (1999) Analysis of the NuRD subunits reveals a histone deacetylase core complex and a connection with DNA methylation. Genes Dev 13:1924–1935

Zhang X, Su J, Jeong M, Ko M, Huang Y, Park HJ, Guzman A, Lei Y, Huang YH, Rao A, Li W, Goodell MA (2016a) DNMT3A and TET2 compete and cooperate to repress lineage-specific transcription factors in hematopoietic stem cells. Nat Genet 48:1014–1023

Zhang Y, Zhang D, Li Q, Liang J, Sun L, Yi X, Chen Z, Yan R, Xie G, Li W, Liu S, Xu B, Li L, Yang J, He L, Shang Y (2016b) Nucleation of DNA repair factors by FOXA1 links DNA demethylation to transcriptional pioneering. Nat Genet 48:1003–1013

Zhou T, Xiong J, Wang M, Yang N, Wong J, Zhu B, Xu RM (2014) Structural basis for hydroxymethylcytosine recognition by the SRA domain of UHRF2. Mol Cell 54:879–886

Zhu H, Wang G, Qian J (2016) Transcription factors as readers and effectors of DNA methylation. Nat Rev Genet 17:551–565

The Role of Nucleosomes in Epigenetic Gene Regulation

4

Katherine A. Giles and Phillippa C. Taberlay

Abstract

Chromatin is a dynamic and highly organised structure that plays a role in all aspects of chromatin biology and controls access to DNA. The basic repeating unit of chromatin is the nucleosome, an octamer of histone proteins wrapped around twice with DNA. Transcriptional gene regulation is largely controlled by the epigenetic regulation of the nucleosome, both through the position of the nucleosome in the DNA and chemical modifications to the histone proteins within the nucleosome core. In this chapter, we discuss the role of the nucleosome in controlling gene regulation at promoters, enhancers and in gene bodies. We also discuss the role of bivalent chromatin and asymmetric nucleosomes at these elements. While the focus of this chapter is on transcriptional regulation, many of the concepts also extend to other aspects of chromatin biology including DNA repair and replication.

Keywords

Nucleosome · Gene regulation · Acetylation · Methylation · Promoter · Enhancer

K. A. Giles
Epigenetics Research Program, Genomics and Epigenetics Division, Garvan Institute of Medical Research, Darlinghurst, NSW, Australia

P. C. Taberlay (✉)
Epigenetics Research Program, Genomics and Epigenetics Division, Garvan Institute of Medical Research, Darlinghurst, NSW, Australia

School of Medicine, College of Health and Medicine, University of Tasmania, Tasmania, Australia
e-mail: Phillippa.Taberlay@utas.edu.au

L. B. Hesson, A. L. Pritchard (eds.), *Clinical Epigenetics*,
https://doi.org/10.1007/978-981-13-8958-0_4

4.1 Introduction

Chromatin is a dynamic multilayered structure that governs the accessibility of DNA and plays an essential role in gene transcription, DNA repair, cell cycle progression and DNA replication. Vast insight into the assembly of the chromatin fibre was gained when the high-resolution structure of the nucleosome, the basic repeating unit of chromatin, was solved in 1997 (Luger et al. 1997). At their core, nucleosomes contain an octamer of histone proteins—with two copies of H3 and H4 in a tetramer and two H2A–H2B dimers—that are wrapped around twice with 147 bp of DNA. The length of DNA separating two nucleosomes is called the 'linker DNA'; permissive chromatin has longer linker DNA between two nucleosomes, which can be shortened as chromatin is compacted and genes repressed. Repressed chromatin can also be bound by histone H1 located outside of the nucleosome core, which locks the DNA and nucleosome into position. Several processes in chromatin biology, including histone modifications, chromatin remodelling, histone chaperones and insertions of histone variants, coordinate chromatin structure with high precision to tightly regulate DNA processes and DNA accessibility (Fig. 4.1; Tessarz and Kouzarides 2014).

Here, we concentrate on the roles that nucleosomes play at DNA transcriptional regulatory regions. We discuss the well-studied epigenetic mechanisms of histone acetylation and methylation along with nucleosome positioning, followed by how these qualities at promoters and enhancers govern transcriptional regulation. We also discuss the links between DNA methylation and nucleosomes at these genomic features. While our focus is on transcription, many of the principles discussed are relevant for other epigenetically controlled DNA processes, such as the DNA damage response and DNA replication, due to the general influence of histone modifications on chromatin structure.

4.2 The Role of the Nucleosome

At the base level of chromatin structure, the positioning of nucleosomes and their interaction with DNA plays a large role in gene activity. Nucleosomes physically impede accessibility to DNA, and this default state must be remodelled by sliding the nucleosomes to increase the length of the linker DNA, or with nucleosome disassembly that is required for gene activity and DNA repair, both of which create a nucleosome-depleted region (NDR). An NDR may also reflect a significant increase in turnover of histone proteins, where the DNA is still transiently occupied by nucleosomes, but becomes more accessible to transcription regulatory factors (Venkatesh and Workman 2015; Zentner and Henikoff 2013), creating an artificial site of 'depletion'. The NDR creates space for transcription factors to bind, and indeed, the binding profiles of transcription factors have demonstrated that they are concentrated at open chromatin, with the highest levels of binding correlating with the most accessible chromatin (Li et al. 2011). Conversely, nucleosome reassembly is necessary for DNA replication and to re-establish the original repressed state of

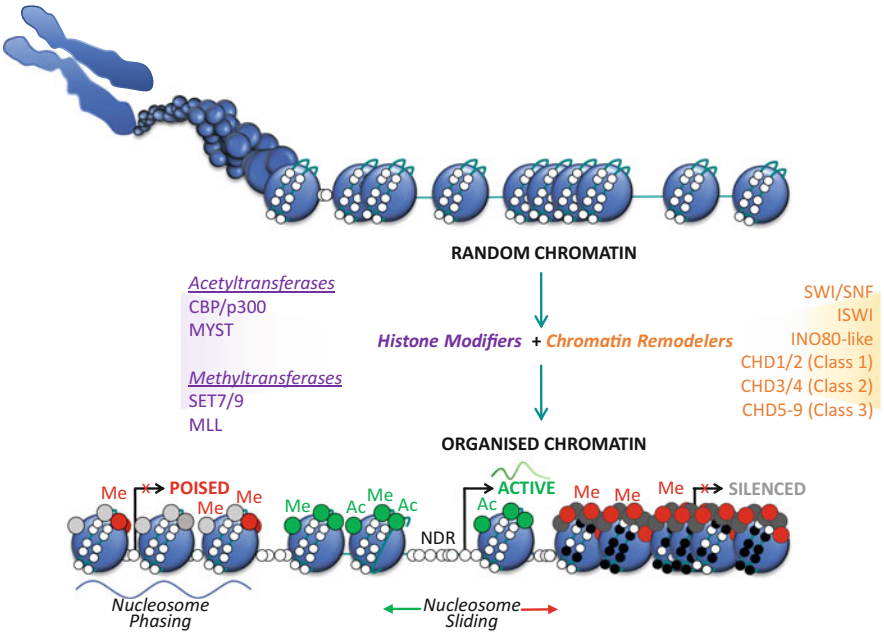

Fig. 4.1 Chromatin can be altered by post-translational modifications (PTMs) to histones or changes in nucleosome positioning. An octamer of histone proteins is randomly loaded onto DNA to form a complex called a nucleosome, which is the basic repeating unit of chromatin (top chromatin strand). Chromatin modifiers (left, in purple) catalyse histone PTMs, such as methylation (Me; red) or acetylation (Ac; green) providing a signal for chromatin 'readers'. One group of readers are the chromatin remodellers (right, in orange), which slide the nucleosomes through the DNA to create an ordered chromatin structure (bottom chromatin strand). This ordered structure can either be permissive allowing regulatory factors to bind or repressive to inhibit access to the DNA. A permissive state may comprise a nucleosome-depleted region (NDR) and increased spacing between nucleosomes. The modifications to chromatin by either PTMs or nucleosome positioning changes are dynamic and occur in response to cellular signals, such as for alterations to transcriptional patterns

genes when they are no longer needed to be expressed (Venkatesh and Workman 2015). Genomic profiling of NDRs has demonstrated that depletion generally occurs at DNA regulatory elements, which are cell type specific and reflect transcriptional patterns (Taberlay et al. 2014; Thurman et al. 2012). Furthermore, the role of the nucleosome is exemplified in studies of yeast aging, which clearly demonstrates that loss of histone proteins compromises genome stability, gene regulation and transcription (Hu et al. 2014).

The nucleosome has dual roles in the nucleus. It serves as the physical backbone for chromatin structure on a global scale, as well as the organisation of gene regulatory elements at the local level. The nucleosome therefore governs both genome-wide stability and DNA accessibility. Additionally, as it is comprised of histone proteins, a nucleosome is necessary for histone post-translational modifications and the requirement of a nucleosome for DNA methylation has also

been demonstrated. Indeed, the physical positioning of the nucleosome precedes gene silencing (You et al. 2011; Lin et al. 2007) and serves to anchor the DNA methyltransferase (DNMTs) enzymes (Sharma et al. 2010). Therefore, the nucleosome is critical for the epigenetic landscape across the genome.

The preferential positioning and organisation of nucleosomes are described as nucleosome phasing (Lohr et al. 1977; Valouev et al. 2011). At both CTCF binding sites and active gene promoters, nucleosome phasing is extremely well defined (Taberlay et al. 2014), and it is notable that at CTCF sites this is lost when CTCF does not occupy the consensus motif (Taberlay et al. 2014). Interestingly, there are at least four well-phased nucleosomes on either side of an NDR, which can still be observed even when facultative NDRs are occupied by nucleosomes across multiple cell types (Taberlay et al. 2014; Kelly et al. 2012). This suggests that nucleosome organisation is more tightly regulated than previously anticipated, occurs on a genome-wide scale and does not rely on the NDR as an orientation point.

4.2.1 Nucleosome Remodelling

ATP-dependent chromatin remodellers are responsible for the movement, assembly and ejection of nucleosomes and govern DNA accessibility. Chromatin remodellers are targeted to genomic regions by either transcription factors or ncRNA molecules as they lack sequence specificity; remodellers recognise histone post-translational modifications (PTMs) to determine the level of chromatin compaction or relaxation required (Bannister and Kouzarides 2011). There are four structural families of chromatin remodellers: SWI/SNF, ISWI, INO80-like and NuRD/CHD (Fig. 4.1; Skulte et al. 2014; Clapier et al. 2017; Johnson et al. 2005; Saha et al. 2006). These remodellers share many features such as their highly similar SNF2-like ATPase catalytic subunit and a high affinity for the nucleosome. However, they have each evolved their own specialised functions. SWI/SNF remodellers contain bromodomains recognising acetylated histones, ISWI complexes have HAND/SANT/SLIDE domains for recognising internucleosomal DNA, the NuRD/CHD family have chromodomains that recognise methylated histones and the INO80-like family have a longer peptide chain between their helicase domains that has been proposed to fit replication forks and Holliday junctions (Skulte et al. 2014; Clapier et al. 2017; Hassan et al. 2002; Morrison 2017). Chromatin remodellers can work in concert with each other at the same genomic location and the order of their binding has been suggested as important for correct regulation of gene expression (Morris et al. 2014). In addition, they can regulate nucleosomes through competative binding for the same histone modification (discussed further below). For example, the CHD1 remodeller and the BPTF (ISWI) complex both recognise the tri-methylation mark on histone 3 lysine 4 (H3K4me3) and compete to alter the phasing of nucleosomes in vivo (Lee et al. 2017a; Lin et al. 2011; Ocampo et al. 2016; Sims et al. 2005; Wysocka et al. 2006).

4.2.2 Variant Nucleosomes

The canonical histones comprise the majority of nucleosomes, but they can be replaced with histone variants for specific roles, which alter the overall structure of chromatin. For example, histone variant γH2A.X is incorporated into DNA at the site of damage after a double-strand break occurs and is recognised by several remodelling complexes including SWI/SNF and ISWI (Atsumi et al. 2015; Chang et al. 2015). Variants H2A.Z and H3.3 are incorporated into chromatin through histone exchange, largely by the INO80-like family, and are found at several active regulatory sites within the genome where they cause chromatin relaxation and are typically associated with active transcription (Morrison 2017; Valdes-Mora et al. 2012, 2017; Ku et al. 2012; Jin et al. 2009; Jin and Felsenfeld 2007).

4.3 Histone Post-Translational Modifications

The post-translational modifications (PTMs) of histone proteins serves multiple purposes, including as docking sites for chromatin binding factors, such as H3K4me3 for CHD1 and BPTF described above (Zentner and Henikoff 2013; Kouzarides 2007; Barski et al. 2007). Diverse protein complexes, known as chromatin modifiers, catalyse the covalent modification of histone tails resulting in the addition of small chemical moieties associated with either activation or repression of chromatin (Fig. 4.2). These chemical groups can include acetylation, methylation, phosphorylation, ubiquitination, sumoylation and poly-ADP-ribosylation; they act over short- or long-range genomic distances and manipulate how DNA regulatory factors interact with the nucleosome (Bert et al. 2013; Kundu et al. 2017; Wani et al.

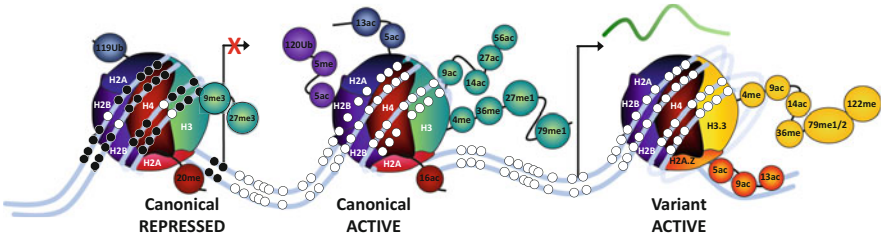

Fig. 4.2 Nucleosomes are decorated in post-translational modifications (PTMs). PTMs can alter the affinity between the histones and the DNA. Acetylation (shown as 'ac') reduces the positive charge on the histones, weakening the histone–DNA interaction and recruits factors that create a more permissive chromatin conformation. Methylation (shown as 'me') does not alter the charge of the histone, however, the location of the residue modified signals for either repression or activation of chromatin (canonical repressed or active nucleosome, left). These marks occur in combination to provide a cooperative signal. Variant nucleosomes (right) have one or more canonical histone replaced for a variant histone. For example, H3.3 (yellow) replaces H3 and H2A.Z (orange) replaces H2A, which together reduce nucleosome stability and increase mobility at DNA regulatory elements to promote active transcription

2016). Most known histone PTMs occur on the highly acidic histone tails that project out from the nucleosome; however, they can also occur on the globular domains located in the nucleosome core, which directly interact with DNA (Tropberger et al. 2013). A single histone can contain several modifications creating a synergistic signal for transcriptional change. PTMs have also been observed on variant histones (Fig. 4.2) (Atsumi et al. 2015; Chang et al. 2015; Valdes-Mora et al. 2012; Ku et al. 2012; Altaf et al. 2010; Raisner and Madhani 2006; Adhikary et al. 2016). While our knowledge has vastly increased over recent years, our understanding of histone modifications and their impact on chromatin structure is far from complete.

Histone PTMs can change the structure of chromatin by altering the affinity between the nucleosome and the surrounding DNA, as well as with adjacent nucleosomes. Acetylation and phosphorylation neutralise acidic residues of histone proteins to modify the overall charge of the nucleosome. This disrupts the interaction between histones and DNA and facilitates the 'opening' of chromatin (Steunou et al. 2014). Histone PTMs can also alter the coiling properties of chromatin by changing the way adjacent nucleosomes interact with each other. Histone 4 lysine 16 acetylation (H4K16ac) has a profound effect on the structure of the 30 nm chromatin fibre, relaxing chromatin by repelling nucleosomes away from each other without the assistance of external chromatin remodelling factors (Shogren-Knaak et al. 2006). Histone 4 lysine 20 di- and tri-methylation (H4K20me2/3) uses a similar mechanism but instead to drive chromatin compaction (Lu et al. 2008). However, to date, these are the only two modifications known to have this ability. To understand more about the function of PTMs on chromatin structure, many studies have sought to examine the effect of deleting histone tails. Interestingly, the results indicated that these deletions have no major effect on nucleosome stability but cause structural changes to the cores of H3 and H2A that result in DNA relaxation and higher nucleosome mobility (Biswas et al. 2011).

The acetylation and methylation of histone tails have been well characterised and are recognised for their roles in gene regulation. Below, we provide a brief outline of the discovery of these marks and the key players that catalyse their addition and removal, along with how chromatin readers recognise them.

4.3.1 Histone Acetylation

The first report of histone acetylation came in 1964 (Allfrey et al. 1964) and it has since been associated with active and 'open' chromatin. Acetylation occurs on lysine residues neutralising their positive charge, weakening the histone–DNA interaction and increasing the proportion of permissible chromatin available for the binding of transcriptional regulators (Steunou et al. 2014). Co-occurrence of acetylation across several residues has the potential to completely neutralise the histone and amplify the signal to be read by chromatin binding proteins. Multisite acetylation has been found at both promoter and enhancer DNA regulatory elements, positively correlating with DNase sensitivity and transcriptional activity (Wang et al. 2008; Kiefer et al. 2008).

While not as well characterised, acetylation on the histone lateral surface (globular domain) also plays an important role in chromatin biology. For example, lysine 56 acetylation (H3K56ac) occurs on the globular domain of H3 and points towards the major groove of the nucleosome, which is the entry–exit point for DNA (Tjeertes et al. 2009; Xu et al. 2005). H3K56ac is important for the binding of histone chaperones, CAF-1 and Rtt06, for the deposition of newly synthesised histones and also has known functions in DNA damage repair and transcription control (Tjeertes et al. 2009; Li et al. 2008; Stejskal et al. 2015; Vempati et al. 2010; Wurtele et al. 2012).

Two opposing groups of enzymes ultimately control histone acetylation levels: histone acetyltransferases (HATs) catalyse the addition of acetyl groups to histones, while histone deacetylases (HDACs) remove acetyl groups (Xhemalce et al. 2011). The balance of activity between these two classes of enzymes is important for proper genome function. There are two classes of HATs: type A acetylate histones within chromatin and type B acetylate cytoplasmic histones, which facilitates the loading of histones into chromatin (Parthun 2007; Hodawadekar and Marmorstein 2007). Type A HATs exhibit much more diversity and include GNAT, MYST and CBP-p300 families and have various roles in several aspects of chromatin biology (Hodawadekar and Marmorstein 2007). Type B HATs include Hat1 and Hat2 that interact with H3 and H4 with roles in DNA repair and telomere silencing (Ai and Parthun 2004; Ge et al. 2011; Kelly et al. 2000; Qin and Parthun 2006). By removal of acetyl groups, the HDACs restore the positive charge to lysines and, in doing so, trigger DNA compaction, consistent with their known repressor roles. To date, there are 18 known HDACs in humans that are grouped into four classes (Chen et al. 2015; Seto and Yoshida 2014). Class I, II and III are grouped based on the location of their catalytic domain and sequence similarities, and Class IV are the sirtuins that require NAD+ as a co-factor (Seto and Yoshida 2014). HDACs display low substrate specificity and generally exist in larger, multiprotein complexes. For example, HDAC1 is often found within the chromatin remodelling complexes NuRD, Sin3a and Co-REST (Yang and Seto 2008). Due to the number of different complexes a HDAC may be incorporated into, it can be difficult to determine which complex is responsible for a particular HDAC's activity. However, it is also known that each HDAC has distinct roles and is likely to be active in a context-dependent manner; for example, HDAC1 controls ESC differentiation in contrast to HDAC2, which does not (Dovey et al. 2010).

Histone acetylation is recognised by proteins containing a bromodomain. Bromodomain proteins are known to associate with the transcription pre-initiation complex (PIC) assembly, initiation and transition to elongation (Wang et al. 2008; De et al. 2011; Rada-Iglesias et al. 2011; Zhang et al. 2012). The remodelling complexes SWI/SNF, CECR2 (ISWI) and RSF (ISWI) all contain bromodomains and facilitate the relaxation of chromatin when they recognise an acetylation target. Indeed, almost all subunits with the SWI/SNF complex contain bromodomains, which allows it to recognise multiple acetylated residues at one time.

4.3.2 Histone Methylation

Histone methylation occurs on arginine or lysine residues and can signal for either chromatin compaction or relaxation depending on the residue modified. Unlike acetylation, methylation does not alter the charge of histones (Xhemalce et al. 2011); therefore, it is less likely to perturb chromatin structure. Lysine residues can be mono-, di- or tri-methylated, while an arginine can be mono-, symmetrically di- or asymmetrically di-methylated.

Histone lysine methylation is catalysed by histone lysine methyltransferases (HKMT) utilising S-adenosylmethionine (SAM) as the methyl donor. The known history of HKMTs is relatively short; the first HKMT was discovered in the year 2000, identifying SUV39H1 as the mediator of histone 3 lysine 9 (H3K9) methylation (Rea et al. 2000). As with acetylation, histone methylation can occur on both the N-terminal tails of histones and the globular domains. HKMTs are specific in both the methylation site they target and the level of methylation catalysed, which is controlled through a key residue within the catalytic domain (Xiao et al. 2003). For example, residue Y305 of SET7/9 can only accommodate a mono-methylated product within the active site (Xiao et al. 2003), and when this is mutated to F305, SET7/9 acts as a tri-methyltransferase (Zhang et al. 2003). Similarly, when tri-methyltransferase DIM5 has a point mutation induced at F281 to Y281, its activity changes from a tri- to mono-methyltransferase (Zhang et al. 2003). Together, these examples demonstrate the specific nature of this class of enzymes.

It was long considered that histone methylation was a stable and static mark, until the phenomenon of histone demethylation (Bannister et al. 2002) and the first lysine demethylase (lysine-specific demethylase 1; LSD1) was identified (Shi et al. 2004). As a monomer, LSD1 does not have a specific target, instead it relies on the larger multisubunit complexes it is part of, to determine the histone residue for demethylation. When part of Co-REST, LSD1 acts as a gene repressor by demethylating H3K4me1/2, while when coupled to the nuclear hormone androgen receptor (AR), it demethylates H3K9 to activate genes (Klose and Zhang 2007). Since this discovery, a series of proteins containing a jumonji (JmjC) domain have been found to demethylate histones in a highly specific fashion. For example, JMJD6 demethylates H3R3 and H4R3, and JMJD2 demethylates tri-methylated H3K9 and H3K36 (Chang et al. 2007; Tsukada et al. 2006; Whetstine et al. 2006). Interestingly, LSD1 is the only known demethylase that does not contain a JmjC domain (Chen et al. 2006).

Several protein domains can ‘read’ and physically bind to histone methylation. These include plant homeodomain (PHD) fingers and the Tudor family of domains (chromodomains, Tudor, PWWP and MBT) (Maurer-Stroh et al. 2003; Kim et al. 2006; Champagne and Kutateladze 2009). Various methylation readers can recognise the same methylation mark, but for different outcomes. For example, H3K4me3 is recognised by the PHD finger of ING proteins to ensure that genes are repressed after DNA damage, the tandem chromodomains of CHD1 for chromatin remodelling and by the Tudor and PHD domains of the JMJD2A and JHDM1D demethylases to

remove this active modification from the histone (Sims et al. 2005; Shi et al. 2006; Lee et al. 2008; Horton et al. 2010).

4.4 DNA Methylation

DNA methylation remains the best studied epigenetic modification, particularly in the context of cancer biology (Jones 2012; You and Jones 2012). The addition of a methyl group to the fifth cytosine of the pyrimidine ring is catalysed by a family of DNMTs and predominantly occurs at CpG dinucleotides across the genome. On a global scale, DNA methylation patterns are established early in development and are mitotically heritable through cell divisions. A normal somatic cell genome is methylated with the exception of CpG islands, which are defined as being greater than 500 bp, having a G+C content greater than 55% and CpG observed/expected greater than 0.65 (Takai and Jones 2002). CpG islands are normally resistant to de novo methylation and tend to be deplete of nucleosomes, although the molecular mechanisms that underpin this phenomenon remain to be identified.

The discovery of ten–eleven translocation (TET) enzymes (Ito et al. 2010; Kriaucionis and Heintz 2009; Tahiliani et al. 2009) and the DNA demethylation pathway/s has challenged the long held view that DNA methylation is a permanent marker of gene silencing. Moreover, the presence of CpH (where H is A, T or C) methylation, particularly in the brain (Ellis et al. 2017; Guo et al. 2014; Lister et al. 2009), indicates that DNA methylation is more complex than previously anticipated. Despite this, DNA methylation is still considered the mediator of long-term gene silencing and it is likely that this state is driven by nucleosome positioning and composition.

4.5 Transcription Regulation

Histone modifications and the changes to chromatin structure by chromatin remodellers act in concert to regulate gene expression and to control the binding of effector molecules to DNA. There is extensive crosstalk between these features, and in this section, we outline these mechanisms controlling gene promoter and enhancer function, as well as their role through gene bodies.

4.5.1 Promoters

Promoters are found upstream of the transcription start site (TSS) at the 5′ end of genes. Genomic arrangement of active promoters includes a well-positioned nucleosome at the +1 position relative to the TSS and bound by RNA polymerase II (RNA pol II), a clearly defined NDR at the −1 position upstream of the TSS, high enrichment of the signature promoter histone PTMs, H3K4me3 and H3K27ac, and various other histone modifications with different degrees of enrichment (Wang et al.

2008; Heintzman et al. 2007; Kundaje et al. 2012). Promoter sequences are concentrated with transcription factor binding motifs (Barski et al. 2007), and transcription factor binding along with the general transcription machinery initiates the production of RNA transcripts from the coding DNA strand.

4.5.1.1 Histone Modifications at Promoters

Histone acetylation is a prominent feature of active promoters, which tend to be hyperacetylated on several histone residues including K9, K14 and particularly K27. As described above, these acetyl groups facilitate the relaxation of chromatin to allow transcription machinery and other regulatory factors to bind. The globular domain histone mark, H3K56ac, is also found at promoters of active genes and correlates with the deposition of new H3 (Xu et al. 2005; Lo et al. 2011; Rufiange et al. 2007). H3K56ac is cell cycle dependent and marks the promoters of genes encoding the histone proteins before they are expressed in S-phase (Stejskal et al. 2015; Vempati et al. 2010), and the presence of H3K56ac is a requirement for SWI/SNF chromatin remodelling at these genes (Xu et al. 2005). The SWI/SNF complex also plays a role at other active promoters through the several bromodomains within the complex. SWI/SNF relaxes chromatin at promoters creating an ordered disassembly of nucleosomes to permit the binding of RNA pol II and the general transcription machinery (Skulte et al. 2014; Heintzman et al. 2007). Specifically, the catalytic subunit of SWI/SNF, BRG1, recognises H3K14ac, and a point mutation at this site prevents its acetylation and reduces BRG1 binding (Luebben et al. 2010; Morrison et al. 2017; Shen et al. 2007). In the absence of H3K14ac, the SWI/SNF complex is still present at promoters and histones are disassembled from the chromatin; however, the level of transcription is reduced (Church et al. 2017).

The presence of a strong H3K4me3 peak defines active promoters and is associated with Ser5-phosphorylation of RNA pol II and active transcription (Kouzarides 2007; Barski et al. 2007; Xhemalce et al. 2011; Heintzman et al. 2007; Schneider et al. 2004). Synthetic CpG-rich DNA can recruit chromatin modifiers and establish H3K4me3 marking in vitro (Thomson et al. 2010). As promoters generally have a high CpG dinucleotide content compared to other regions of the genome, it has been proposed that this sequence composition could explain why this mark exhibits a high signal at promoters but little to no signal at enhancers, which are CpG poor (Thomson et al. 2010; Calo and Wysocka 2013). Although histone methylation does not change the charge or structure of chromatin, it can prevent factors associated with silencing from binding to active promoters. H3K4me3 prevents the NuRD chromatin remodelling complex from silencing active genes and disrupts the PHD finger in DNMT3L from binding to the tail of H3 (Adams-Cioaba and Min 2009; Nishioka et al. 2002; Ooi et al. 2007; Zegerman et al. 2002).

Silenced promoters display the expected marks of epigenetic repression, which signal for compaction of chromatin. The Polycomb repressive complex 2 (PRC2) catalyses H3K27me3, a major mark of gene repression. A key example of long-range Polycomb silencing occurs during X-chromosome inactivation in females,

which silences genes for proper gene dosage (Maclary et al. 2017; Shen et al. 2008). The PRC2 complex and SWI/SNF remodelling complex have antagonistic roles at promoters; SWI/SNF evicts PRC2 in order to open chromatin at promoters while conversely, when SWI/SNF function is lost in diseases such as cancer, PRC2 remains bound to promoters and maintains the repressive gene state (Kia et al. 2008; Tamkun et al. 1992; Wilson et al. 2010). Polycomb repression allows some flexibility to gene regulation, such as at bivalent promoters that contain both active and repressive histone modifications and are expressed at low levels in embryonic stem (ES) cells (discussed in more detail below). While there is extensive research into the Polycomb silencing of genes, there are other histone mechanisms of gene silencing. Altered nucleosome positioning occurs during silencing and loss of the −2 nucleosome relative to the TSS is an early silencing event in cancer, followed by nucleosome compaction and DNA methylation (Hesson et al. 2014). Other histone modifications also lead to promoter gene silencing such as by demethylase, L3MBTL1. L3MBTL1 can facilitate the formation of compact nucleosome arrays through its three MBT domains that recognise H4K20me1/2 and H1bK26me1/2 to regulate E2F targets (Trojer et al. 2007). L3MBTL1 binding mimics the progressive accumulation of H4K20me1 during the cell cycle (Kalakonda et al. 2008). For long-term gene silencing in constitutive heterochromatin, promoters display H3K9me3, often concomitant with DNA methylation, but anti-correlated with H3K27me3 (Kundaje et al. 2012; Allan et al. 2012; Wang et al. 2016; Xia et al. 2015).

4.5.1.2 Histone Variants at Promoters

In addition to high levels of histone acetylation and specific methylation marks, histone variants H2A.Z and H3.3 are also a feature of active promoters. These variants reduce the stability of nucleosomes at promoters and are easier to displace providing transcription factors an advantage when competing with nucleosomes for access to DNA (Valdes-Mora et al. 2012; Ku et al. 2012; Jin et al. 2009; Jin and Felsenfeld 2007; Hu et al. 2013; Goldberg et al. 2010; John et al. 2008). The combination of histone variants and canonical histones at promoters affects the degree of nucleosome mobility. Canonical nucleosomes are less mobile than those containing histone variants H3.3 and H2A that in turn are less mobile than those containing H3.3 and H2A.Z, which establishes a hierarchy of stabilities for different transcriptional requirements of the cell (Jin and Felsenfeld 2007). Interestingly, nucleosomes containing H2A.Z but not H3.3 have a similar stability to the canonical nucleosome (Jin and Felsenfeld 2007). However, H2A.Z does have a pro-transcriptional role; when the SRCAP remodelling complex (from the INO80-like family) is knocked down, H2A.Z is depleted from promoters and a reduction in gene expression is observed (Hu et al. 2013; Yang et al. 2012). Interestingly, H2A.Z-containing nucleosomes shift position at promoters during the cell cycle. Genes that are silenced in mitosis shift their H2A.Z-containing +1 nucleosome upstream to occupy the TSS, which could potentially allow the rapid silencing and then reactivation across the cell cycle (Kelly et al. 2010). H2A.Z is also linked to poised Polycomb marked promoters (Taberlay et al. 2011). It occupies the promoters of key Polycomb marked developmental genes and potentially forms a specialised

chromatin state, and as with the mitotically silenced genes, it could allow for the rapid switching of expression states (Creyghton et al. 2008). As for canonical histones, histone variants are also subject to post-translational modification at promoters. H2A.Z is acetylated on several histone tail lysine residues at the promoters of active genes and is associated with increased expression and oncogene expression in cancer (Valdes-Mora et al. 2012).

4.5.1.3 Nucleosome Positioning at Promoters

As described above, nucleosomes are well positioned at promoters. In particular, the well-characterised nucleosomes located at the +1 and −1 positions relative to the TSS establish a nucleosome-depleted region at promoters (Kubik et al. 2015). In yeast, the chromatin structure remodelling (RSC) complex has a well-established role in maintaining these nucleosome positions (Kubik et al. 2015; Nagai et al. 2017), and both BRG1 and SNF5 of the SWI/SNF complex are necessary for maintaining the −1 and +1 nucleosome positions at repressed gene promoters in mammalian cells (Tolstorukov et al. 2013). Loss of either BRG1 or SNF5 proteins results in nucleosome depletion at the TSS and a shorter DNA linker distance from 184 to 174 bp between nucleosomes adjacent to the TSS. This suggests that nucleosomes are remodelled outwards from the TSS creating open DNA for transcription factor binding and gene expression (Tolstorukov et al. 2013). This is contrary to the typical role for BRG1, which is to keep chromatin active and open (Morris et al. 2014; Bultman et al. 2005; Hu et al. 2011; Laurette et al. 2015; Li et al. 2013). While the positioning of nucleosomes at the TSS intuitively appears repressive and would block access of RNA pol II and the general transcription machinery binding to promoters, they are in fact required for transcription. Nagai and colleagues demonstrated that the presence of the +1 nucleosome, while not stimulating transcription, greatly increases transcriptional levels compared to naked DNA, suggesting that chromatin itself is a requirement of transcription (Nagai et al. 2017). The positioning of nucleosomes around most transcription factor binding sites at promoters has found to be asymmetrical, and while the complete function of this is unknown, it is predicted to be linked with the direction of transcription (Kundaje et al. 2012).

4.5.1.4 DNA Methylation at Promoters

While nucleosome positioning is not sequence specific, the DNA sequence can still play a role in the formation of chromatin. The sequence of DNA has been computationally predicted to intrinsically encode approximately 50% of nucleosome positioning and has a preference for AT-rich DNA (Caserta et al. 2009; Peckham et al. 2007; Segal et al. 2006). It is likely that chromatin remodellers then fine-tune this positioning in a cell-type-specific context. As described above, at promoter CpG islands, DNA tends to be unmethylated and deplete of nucleosomes in normal somatic cells. The high CpG content is protective of an active epigenetic signature and prevents nucleosomes from becoming densely compacted resulting in active transcription (Lister et al. 2009). DNA sequence and nucleosome positioning are also linked through DNA methylation. The positioning of nucleosomes is required

prior to de novo DNA methylation during differentiation (You et al. 2011), and the linker DNA between the nucleosomes acts as a site for 'seeding' DNA methylation (Hinshelwood et al. 2009).

4.5.1.5 Promoter Nucleosomes and Cancer

During cancer development, changes to histone modifications and nucleosome positioning occur at promoters, which is also linked with alterations in DNA methylation. In the early stages of cancer development, Polycomb marked promoters lose their H3K27me3 signature and become DNA hypermethylated (Gal-Yam et al. 2008; Widschwendter et al. 2007). While these genes are already suppressed, this epigenetic switching has been proposed to reduce the flexibility of their regulation (Gal-Yam et al. 2008). There can also be cooperation between these silencing mechanisms, the Polycomb protein EZH2 is known to target DNMTs to promoters and overexpression of EZH2 increases the level of DNA methylation (Vire et al. 2006). Indeed, in cholangiocarcinoma cells, miR-34A is silenced by EZH2 and DNA methylation, resulting in increased growth through activation of Notch signalling (Kwon et al. 2017). Additionally, EZH2 target genes can be reactivated by 5-aza-2′-deoxycytidine (Vire et al. 2006). DNA hypomethylation in cancer has also been linked to redistribution of the Polycomb mark H3K27me3 (Reddington et al. 2013). The expression of chromatin remodellers, and in turn nucleosome positioning, has also been linked to promoter DNA hypermethylation in cancer. For example, both CD44 and E-cadherin are silenced by promoter hypermethylation in C33A cervical cancer cells, which are also deplete of remodeller proteins BRG1 and BRM. This promoter hypermethylation can be reversed by treatment with 5-aza-2-′-deoxycytidine, a DNA methylation inhibitor, or re-expression BRG1 or BRM, further demonstrating the link between nucleosome positioning and DNA methylation (Banine et al. 2005) and the complexities in epigenetic remodelling as a cause and consequence of cancer onset and progression.

4.5.2 Enhancers

Enhancers are distal regulatory elements that control cell-type-specific gene expression patterns. The distance between an enhancer and its cognate promoter is not fixed, and varies from within an intron of the target gene to hundreds of kilobases away (Bulger and Groudine 2011; Stroud et al. 2011; Ong and Corces 2011, 2012) and from distinct chromosomes (Bateman et al. 2012; Geyer et al. 1990; Muller and Schaffner 1990). Enhancers function to increase transcription by either facilitating the formation of the PIC or to assist the transition from initiation to elongation at promoters. The number of enhancers far exceeds the number of promoters; enhancers may regulate more than a single promoter and promoters may be regulated by multiple enhancers, revealing that there is vast complexity in the regulation of gene expression over genome-scale distances. The central core of an enhancer is commonly 200–500 bp and densely packed with transcription factor motifs (Spitz and Furlong 2012), though the entire enhancer can span up to 4 kb in length. As with

promoters, active and poised enhancers are associated with NDRs and DNase hypersensitivity. Their epigenetic features then distinguish them, with poised enhancers enriched for H3K4me1 and active enhancers annotated by both H3K27ac and H3K4me1 (Heintzman et al. 2007; Creyghton et al. 2010; Hon et al. 2009; Zentner et al. 2011). However, enhancers can also contain other PTMs such as H3K18ac and histone variants, which can also contain PTMs such as H2A.Zac (Valdes-Mora et al. 2017; Creyghton et al. 2010). These epigenetic features have been used to predict the location of enhancers across multiple cell types and model systems (Rada-Iglesias et al. 2011; Arnold et al. 2013; Blow et al. 2010; Bogdanovic et al. 2012; Bonn et al. 2012; Mercer et al. 2011; May et al. 2011; Visel et al. 2009); however, this is insufficient to determine the target gene promoter and to validate activity in individual cell types.

4.5.2.1 Histone Modifications at Enhancers

H3K4me1 was the first histone modification linked to enhancers and identifies both poised and active states. H3K4me1 is not unique to enhancers, occurring at promoters as well; however, it is a high ratio of H3K4me1 to H3K4me3 that broadly distinguishes enhancers from promoters (Heintzman et al. 2007; Koch and Andrau 2011; Pekowska et al. 2011). H3K4me1 is recognised by the chromodomain on acetyltransferase Tip60, a member of the p400 chromatin remodelling complex. The p400 complex is largely responsible for dimer exchange of canonical H2A-H2B for H2A.Z-H2B, and it has been proposed that H3K4me1 is a marker for H2A.Z placement at enhancers (Altaf et al. 2010; Calo and Wysocka 2013; Yang et al. 2012; Jeong et al. 2011). H3K4me1 is one of the early events in enhancer patterning, preceding nucleosome depletion; furthermore, H3K27ac is almost exclusively acquired at enhancers pre-marked by H3K4me1 (Taberlay et al. 2014; Bonn et al. 2012) and is important for separating active (H3K27ac present) from poised (H3K27ac absent) enhancers (Heintzman et al. 2007; Creyghton et al. 2010). CBP/p300 catalyses the acetylation of H3K27, and chromatin immunoprecipitation mapping of this enzyme has also been used to identify active enhancers (Rada-Iglesias et al. 2011; Visel et al. 2009; Pasini et al. 2010; Kim et al. 2010; Jin et al. 2011; Ghisletti et al. 2010). However, as CBP/p300 can acetylate other histone residues, such as H3K18ac, it is not an exclusive marker of distal enhancers (Jin et al. 2011). Other acetyl marks found at enhancers include H3K9ac, H3K18ac and on histone variant H2A.Zac (Valdes-Mora et al. 2017; Zentner et al. 2011; Krebs et al. 2011; Dalvai et al. 2013).

4.5.2.2 Nucleosome Positioning at Enhancers

Nucleosome positioning and turnover is important for enhancer function. There is a high turnover of nucleosomes at transcription factor binding sites, flanked by less mobile nucleosomes containing histone modifications (He et al. 2010). Indeed, it has been shown that AR-responsive enhancers have well-positioned nucleosomes at the edges that become more stable and densely packed, while there is a higher turnover of nucleosomes in the AR binding site when the enhancers are activated (He et al. 2010). Polycomb-repressed promoters are often controlled by poised enhancers that

are nucleosome depleted, marked by H3K4me1 and deplete of H3K27ac, and upon forced reactivation of these genes, the rate of histone turnover and nucleosome depletion increases and the acquisition of H3K27ac occurs concomitant with Polycomb eviction and gene expression (Taberlay et al. 2011). Chromatin remodellers have a significant effect on nucleosome movement and positioning at enhancers. The SWI/SNF remodeller BRG1 is significantly more enriched at enhancers than promoters (Bao et al. 2015), and AR-responsive enhancers are dependent on BRG1 for nucleosome turnover in prostate cells (Hu et al. 2011; Andreu-Vieyra et al. 2011). CHD7 is also known to preferentially associate with enhancers compared to promoters (Schnetz et al. 2009).

4.5.2.3 DNA Methylation at Enhancers

Enhancers are typically CpG poor and do not have the protection from natural chromatin condensation and nucleosome packing observed at CpG island promoters. CpG sites within enhancers tend to be hypomethylated but exhibit high levels of 5-hydroxymethylcytosine (Lister et al. 2009; Stadler et al. 2011). Approximately 50% of CpG sites at enhancers contain 5-hydroxymethylcytosine, compared to only 15% of promoter CpGs (Skvortsova et al. 2017). Moreover, there is a high concordance of 5-hydroxymethylcytosine with H3K4me1 and H3K27ac, unlike DNA methylation (Stroud et al. 2011; Pastor et al. 2011; Ficz et al. 2011). It remains unknown whether the high level of 5- hydroxymethylcytosine at enhancers is due to continual activity by the TET enzymes within the DNA demethylation pathway (Dai et al. 2016; He et al. 2011), or if it represents a discrete epigenetic mark with a unique regulatory function (Shen and Zhang 2013; Thomson et al. 2013). It is possible that 5-hydroxymethylcytosine serves as a protective mechanism from DNA methylation to prevent long-term silencing of temporal gene expression at enhancers, but this has yet to be proven experimentally.

4.5.3 Gene Bodies

4.5.3.1 Histone Modifications at Gene Bodies

The most prominent histone modification observed at exons in gene bodies of actively transcribed genes is H3K36me3 (Andersson et al. 2009; Dhami et al. 2010; Edmunds et al. 2008; Kolasinska-Zwierz et al. 2009), a pattern that was first observed in *Caenorhabditis elegans* and is evolutionally conserved in mice and humans (Kolasinska-Zwierz et al. 2009). H3K36me3 is catalysed by SET domain containing 2 (SETD2) (Edmunds et al. 2008; Sun et al. 2005), which will associate only with the phosphorylated form of RNA pol II further supporting the link with positive transcription (Sun et al. 2005). Speckle-type POZ protein (SPOP) is an E3 ubiquitin ligase that targets SETD2 for degradation, and functional work has shown that knockdown of SPOP increases H3K36me3 while knockdown of SETD2 decreases H3K36me3 levels, with approximately 80% of total active genes affected (Zhu et al. 2017). Loss of SPOP and SETD2 results in over 1000 alternate splicing events, linking H3K36me3 with the control of exon inclusion (Zhu et al. 2017).

Indeed, included exons have higher levels of H3K36me3 compared to those that are spliced out, and it is speculated that H3K36me3 makes exons more 'visible' to the splicing machinery (Andersson et al. 2009; Dhami et al. 2010). H3K36me3 is recognised by several 'reader' proteins including the Tudor domain of PHF1 that stabilises the nucleosome in a more 'open' conformation; the PWWP domain of BS69 is thought to facilitate the recruitment of chromatin modifiers EZH2, HDAC1 and BRG1; and the PWWP domains of DNMT3A/B that methylate DNA in gene bodies (Musselman et al. 2013; Rondelet et al. 2016; Wang et al. 2014).

While it is the most prominent, H3K36me3 is not the only modification enriched at gene bodies. H3K36me3, H3K79me1, H3K27me1 and H2BK5me1 are enriched at exons and correlated with high levels of exon expression (Andersson et al. 2009), as opposed to H3K36me1 and H3K27me3 that are enriched at exons expressed at low frequencies (Dhami et al. 2010). Interestingly, H3K36me1 and H3K27me3 are located at genes with low overall expression, and these PTMs are removed to allow for H3K36me3 and H3K27me1 as expression increases. This is thought to provide a 'priming' mechanism ready to easily switch when signals for increased expression are present (Dhami et al. 2010).

4.5.3.2 Nucleosome Positioning at Gene Bodies

Nucleosome positioning in gene bodies has a preference for constitutive exons (Andersson et al. 2009; Dhami et al. 2010; Huang et al. 2012). Nucleosome positioning in these exons has been associated with splice sites, where there is higher nucleosome enrichment at constitutive sites over skipped sites and at distal alternative splice sites compared to proximal alternative splice sites (Huang et al. 2012). The well-positioned nucleosomes of exons are independent of overall gene transcription levels (Andersson et al. 2009). Studies in yeast have demonstrated that CHD1 and ISWI chromatin remodelling complexes remodel nucleosomes associated with gene bodies. Both of these complexes are found bound to promoter NDRs but are active in gene bodies (Zentner et al. 2013). They are also highly associated with nucleosome turnover and the rate of transcription elongation, suggesting that unwrapping and re-wrapping of nucleosomes during transcription is transient (Zentner et al. 2013). Further work focusing on CHD1 has shown that its loss alters H3K36me3 and H3K4me3 patterns within 1 kb of the TSS and CHD1 has been suggested to establish the boundaries of these modifications (Lee et al. 2017a). Furthermore, intron-containing genes in yeast display significantly less intron retention with loss of CHD1, suggesting that CHD1 is not only important for elongation but also for correct splicing (Lee et al. 2017a).

4.5.3.3 DNA Methylation at Gene Bodies

In contrast to its known role in gene silencing when positioned at promoters, DNA methylation is associated with active genes when positioned at the exons of gene bodies. It has been suggested that this prevents aberrant transcripts from starting within the gene (Teissandier and Bourc'his 2017) and promotes the rapid incorporation of nucleosomes after transcription to protect DNA from damage (Choi 2010; Gelfman et al. 2013; Ndlovu et al. 2011). Knockdown of SETD2 and

the reduction of H3K36me3 in gene bodies does not affect the level of DNA methylation, and conversely in the HCT116 colorectal cancer cell line, where DNA methyltransferases DNMT1 and DNMT3B are knocked out, transcription and H3K36me3 enrichment are not affected by the reduction in methylation, suggesting that these two features are established independently of each other (Hahn et al. 2011). Interestingly, in gene bodies with low CpG density, H3K9me3 is thought to play a similar role to DNA methylation. At these genes, 90% of which are zinc-finger proteins, both H3K36me3 and H3K9me3 are present in the gene bodies and there are low levels of transcription (Hahn et al. 2011). Although not extensively studied to date, 5-hydroxymethylcytosine has also been reported at gene bodies of ES cells (Stroud et al. 2011).

4.5.4 Bivalent Chromatin

DNA regulatory features that contain both active and repressive histone PTMs are considered bivalent regulatory regions (Bernstein et al. 2006; Azuara et al. 2006; Shema et al. 2016). Bivalency maintains a poised or 'paused' state, with a low level of expression detected (Voigt et al. 2013), but with potential to be rapidly activated or silenced upon signals received by the cell (Voigt et al. 2013). Bivalent chromatin is largely associated with ES cells and genes involved in differentiation (Taberlay et al. 2011; Voigt et al. 2013; Bernhart et al. 2016; Zheng et al. 2016; Zhou et al. 2018), with twice as many bivalent regions being identified in ES and induced pluripotent stem (iPS) cells compared to other normal somatic and cancer cells (Bernhart et al. 2016).

4.5.4.1 Bivalent Promoters

Bivalent promoters feature the active promoter mark H3K4me3 and the repressive mark H3K27me3, and are bound by the PRC2 complex and RNA pol II (Bernstein et al. 2006; Bernhart et al. 2016; Gaertner et al. 2012; Margaritis and Holstege 2008; Mikkelsen et al. 2007; Sen et al. 2016). While H3K4me3 and H3K27me3 overlap in broad zones upstream of the TSS, it is the enrichment over the TSS and +1 nucleosomes that are important for maintaining the bivalent state (Sen et al. 2016). There are conflicting reports of the effect of DNA methylation on the bivalent state of genes, where DNA methylation has been associated with both repression and activation of expression (Bernhart et al. 2016; Lee et al. 2017b; Kretzmer et al. 2015). In ES cells, cell-type-specific methylation was found to occur at CpG Islands of bivalent genes and associated with a loss of both H3K4me3 and H3K27me3 during differentiation, suggesting that there is a role for this during lineage commitment (Lee et al. 2017b). DNA hypermethylation at bivalent promoters has also recently been reported in TET1/2/3 triple knockout (TKO) ES cells (Verma et al. 2018), where the TET proteins were found to preserve 5hmC at bivalent promoters. Loss of the TET proteins allows DNMT3B to establish DNA methylation at bivalent promoters and prevents neuronal development genes from being activated and normal differentiation from occurring (Verma et al. 2018). However, in cancer

cells, hypermethylation at bivalent chromatin regions has been linked to increased expression and a decrease in the flexibility of expression of these genes (Bernhart et al. 2016). Together, this suggests a complex role for the interplay between bivalent promoters and DNA methylation, and further work is required to fully understand the role of DNA methylation on bivalent chromatin.

4.5.4.2 Bivalent Enhancers

Bivalent enhancers are marked by H3K4me1 and H3K27me3 and, to date, have only been identified in ES cells (Rada-Iglesias et al. 2011; Zentner et al. 2011). Bivalent enhancers are considered in a poised state due to the absence of H3K27ac. As ES cells differentiate, H3K27me3 is reduced as H3K27ac is gained, suggesting that these regions are pre-marked to become active enhancers in early development (Rada-Iglesias et al. 2011). Interestingly, a subclass of these enhancers also contains the repressive histone mark H3K9me3 along with H3K27me3, suggesting that Polycomb is not the only mechanism for defining this enhancer group (Zentner et al. 2011). Bivalent enhancers tend to target genes with bivalent promoters demonstrating a connection between epigenetic bivalency at DNA regulatory elements.

4.5.5 Nucleosome Asymmetry

While it is known that, as a whole, genes are asymmetric for nucleosome modifications—when comparing the 5′, middle and 3′ ends of a gene—recent research has demonstrated that individual asymmetrical nucleosomes exist in yeast and human chromatin at DNA regulatory elements (Kundaje et al. 2012; Calo and Wysocka 2013; Levendosky et al. 2016; Ramachandran et al. 2015). Nucleosome asymmetry occurs where there are different PTMs on each copy of a particular histone within a single nucleosome core. Nucleosome asymmetry has the potential to affect how a protein binds to chromatin (Ichikawa et al. 2017). If a protein or multisubunit complex requires both histone tails to bind, the binding affinity will be reduced on asymmetrical nucleosomes; however, if each tail recruits proteins separately, asymmetry has the potential to have a range of results depending on the modifications present.

At promoters, nucleosome asymmetry orientates the direction of transcription by guiding the positioning of RNA pol II at the +1 nucleosome and facilitating its passage into gene bodies (Levendosky et al. 2016; Ramachandran et al. 2015; Rhee et al. 2014). In yeast, H3K9ac and H2BK123ub are found on the proximal half of the +1 nucleosome, while H2A.Z is enriched on the distal half, and it is the highly unstable nature of H2A.Z that is expected to facilitate the movement of RNA pol II into genes (Rhee et al. 2014). In ES cells, nucleosome asymmetry has been identified at bivalent promoters, where the active and repressive histone PTMs, H3K4me3 and H3K27me3, exist on different copies of H3 tails in vivo (Voigt et al. 2012). Interestingly, the catalysis of H3K27me3 is inhibited when either H3K4me3 or

H3K36me3 is present in a symmetrical pattern at promoters, suggesting that this is a protective mechanism against gene repression (Voigt et al. 2012).

Nucleosome asymmetry is also important for the function of chromatin remodellers and adds another layer of complexity to transcriptional regulation signals. The remodeller, CHD1, can sense both entry and exit points (for DNA) of the nucleosome, but it is ubiquitination of H2B on the entry side that stimulates CHD1 to slide nucleosomes, moving them unidirectionally towards the entry side in vitro. It has been speculated that this mechanism contributes to the sliding of nucleosomes upstream of the +1 nucleosome towards the transcription start site in vivo (Voigt et al. 2012). A study on yeast chromatin has observed that asymmetric nucleosomes are bound by remodelling complex, RSC, which is required for maintaining their asymmetry, but that the SWI/SNF complex was depleted from asymmetric nucleosomes (Ramachandran et al. 2015). This suggests that asymmetric nucleosomes are specifically targeted by remodeller proteins and it is this capability that fine-tunes their mechanism of action.

4.6 Future Directions

There is increasing evidence for the role of ncRNA chromatin biology and nucleosome movement (Moazed 2009; Rinn et al. 2007; Tsai et al. 2010; Cajigas et al. 2015; Han et al. 2014; Kawaguchi et al. 2015; Leisegang et al. 2017; Onorati et al. 2011), and in 2011, the first direct interaction between an ATP-dependent chromatin remodeller and a long ncRNA was discovered (Onorati et al. 2011). The biology and significance of these ncRNAs on chromatin dynamics and gene transcription is only beginning to be realised. ncRNAs are ideal to act on chromatin due to the specificity from sequence homology for targeting chromatin modifying and remodelling enzymes to the genome, and understanding their mechanisms of action is an exciting field for future research.

Aberrant histone modifications and the mutation of epigenetic proteins is a feature of many cancer types with the potential to alter the epigenome, and there is a large interest in how this contributes to disease. Aberrant histone modifications are linked to cancer initiation, progression and metastasis through alteration of gene expression profiles and integrity of chromosome segregation (Venkatesh and Workman 2015). At promoters, epigenetic silencing requires the loss of an NDR, H3K4me3 and CTCF, and the number of aberrantly silenced tumour suppressor genes in cancer would suggest that it is easier to silence an already active region than it is to activate a silenced region (Taberlay et al. 2014). It would be particularly interesting to characterise the existence of any skew towards atypical activation of transcriptional repressors as a result of epigenetic mutations, in particular those that modulate nucleosome positioning to silence genes or disrupt the cell cycle.

Mutations and disruptions to chromatin modifiers and remodellers themselves also have the potential to disrupt transcription. As outlined above, a point mutation that switches Phe to Tyr or vice versa is sufficient to alter the activity of histone demethylases. ATP-dependent chromatin remodellers are highly mutated in cancer

(Skulte et al. 2014; You and Jones 2012; Reisman et al. 2009). Improper function of ATP-dependent remodellers can interfere with how histone modifications are read and result in atypical nucleosome positioning. Chromatin remodellers generally have inactivating mutations highlighting a tumour suppressor role, but there are also reports of their overexpression where they function as an oncogene silencing other tumour suppressors (Skulte et al. 2014). An additional layer of complexity with respect to aberrant gene regulation in cancer arises due to the mutations of cofactors that interact with chromatin remodellers. It is likely that these mutations interfere with the targeting of chromatin remodellers to their correct target genes or, potentially, inhibit their recruitment to chromatin at all.

4.7 Concluding Remarks

Chromatin is a dynamic structure with a complex role in many DNA processes including transcription, repair and replication. A faithful copy of the epigenome is required to maintain these processes in daughter cells, as well as prevent disease and cellular disruption. How the epigenome is faithfully inherited to daughter cells after division is still unclear, and work is currently underway to understand more about this process (Berger et al. 2009) and the cooperative nature of different epigenetic features, including during transcription regulation.

References

Adams-Cioaba MA, Min J (2009) Structure and function of histone methylation binding proteins. Biochem Cell Biol 87(1):93–105

Adhikary S, Sanyal S, Basu M, Sengupta I, Sen S, Srivastava DK et al (2016) Selective recognition of H3.1K36 dimethylation/H4K16 acetylation facilitates the regulation of all-trans-retinoic Acid (ATRA)-responsive genes by putative chromatin reader ZMYND8. J Biol Chem 291 (6):2664–2681

Ai X, Parthun MR (2004) The nuclear Hat1p/Hat2p complex: a molecular link between type B histone acetyltransferases and chromatin assembly. Mol Cell 14(2):195–205

Allan RS, Zueva E, Cammas F, Schreiber HA, Masson V, Belz GT et al (2012) An epigenetic silencing pathway controlling T helper 2 cell lineage commitment. Nature 487(7406):249–253

Allfrey VG, Faulkner R, Mirsky AE (1964) Acetylation and methylation of histones and their possible role in the regulation of RNA synthesis. Proc Natl Acad Sci U S A 51:786–794

Altaf M, Auger A, Monnet-Saksouk J, Brodeur J, Piquet S, Cramet M et al (2010) NuA4-dependent acetylation of nucleosomal histones H4 and H2A directly stimulates incorporation of H2A.Z by the SWR1 complex. J Biol Chem 285(21):15966–15977

Andersson R, Enroth S, Rada-Iglesias A, Wadelius C, Komorowski J (2009) Nucleosomes are well positioned in exons and carry characteristic histone modifications. Genome Res 19 (10):1732–1741

Andreu-Vieyra C, Lai J, Berman BP, Frenkel B, Jia L, Jones PA et al (2011) Dynamic nucleosome-depleted regions at androgen receptor enhancers in the absence of ligand in prostate cancer cells. Mol Cell Biol 31(23):4648–4662

Arnold CD, Gerlach D, Stelzer C, Boryn LM, Rath M, Stark A (2013) Genome-wide quantitative enhancer activity maps identified by STARR-seq. Science 339(6123):1074–1077

Atsumi Y, Minakawa Y, Ono M, Dobashi S, Shinohe K, Shinohara A et al (2015) ATM and SIRT6/SNF2H mediate transient H2AX stabilization when DSBs form by blocking HUWE1 to allow efficient gammaH2AX foci formation. Cell Rep 13(12):2728–2740

Azuara V, Perry P, Sauer S, Spivakov M, Jorgensen HF, John RM et al (2006) Chromatin signatures of pluripotent cell lines. Nat Cell Biol 8(5):532–538

Banine F, Bartlett C, Gunawardena R, Muchardt C, Yaniv M, Knudsen ES et al (2005) SWI/SNF chromatin-remodeling factors induce changes in DNA methylation to promote transcriptional activation. Cancer Res 65(9):3542–3547

Bannister AJ, Kouzarides T (2011) Regulation of chromatin by histone modifications. Cell Res 21 (3):381–395

Bannister AJ, Schneider R, Kouzarides T (2002) Histone methylation: dynamic or static? Cell 109 (7):801–806

Bao X, Rubin AJ, Qu K, Zhang J, Giresi PG, Chang HY et al (2015) A novel ATAC-seq approach reveals lineage-specific reinforcement of the open chromatin landscape via cooperation between BAF and p63. Genome Biol 16:284

Barski A, Cuddapah S, Cui K, Roh TY, Schones DE, Wang Z et al (2007) High-resolution profiling of histone methylations in the human genome. Cell 129(4):823–837

Bateman JR, Johnson JE, Locke MN (2012) Comparing enhancer action in cis and in trans. Genetics 191(4):1143–1155

Berger SL, Kouzarides T, Shiekhattar R, Shilatifard A (2009) An operational definition of epigenetics. Genes Dev 23(7):781–783

Bernhart SH, Kretzmer H, Holdt LM, Juhling F, Ammerpohl O, Bergmann AK et al (2016) Changes of bivalent chromatin coincide with increased expression of developmental genes in cancer. Sci Rep 6:37393

Bernstein BE, Mikkelsen TS, Xie X, Kamal M, Huebert DJ, Cuff J et al (2006) A bivalent chromatin structure marks key developmental genes in embryonic stem cells. Cell 125 (2):315–326

Bert SA, Robinson MD, Strbenac D, Statham AL, Song JZ, Hulf T et al (2013) Regional activation of the cancer genome by long-range epigenetic remodeling. Cancer Cell 23(1):9–22

Biswas M, Voltz K, Smith JC, Langowski J (2011) Role of histone tails in structural stability of the nucleosome. PLoS Comput Biol 7(12):e1002279

Blow MJ, McCulley DJ, Li Z, Zhang T, Akiyama JA, Holt A et al (2010) ChIP-Seq identification of weakly conserved heart enhancers. Nat Genet 42(9):806–810

Bogdanovic O, Fernandez-Minan A, Tena JJ, de la Calle-Mustienes E, Hidalgo C, van Kruysbergen I et al (2012) Dynamics of enhancer chromatin signatures mark the transition from pluripotency to cell specification during embryogenesis. Genome Res 22(10):2043–2053

Bonn S, Zinzen RP, Girardot C, Gustafson EH, Perez-Gonzalez A, Delhomme N et al (2012) Tissue-specific analysis of chromatin state identifies temporal signatures of enhancer activity during embryonic development. Nat Genet 44(2):148–156

Bulger M, Groudine M (2011) Functional and mechanistic diversity of distal transcription enhancers. Cell 144(3):327–339

Bultman SJ, Gebuhr TC, Magnuson T (2005) A Brg1 mutation that uncouples ATPase activity from chromatin remodeling reveals an essential role for SWI/SNF-related complexes in beta-globin expression and erythroid development. Genes Dev 19(23):2849–2861

Cajigas I, Leib DE, Cochrane J, Luo H, Swyter KR, Chen S et al (2015) Evf2 lncRNA/BRG1/DLX1 interactions reveal RNA-dependent inhibition of chromatin remodeling. Development 142(15):2641–2652

Calo E, Wysocka J (2013) Modification of enhancer chromatin: what, how, and why? Mol Cell 49 (5):825–837

Caserta M, Agricola E, Churcher M, Hiriart E, Verdone L, Di Mauro E et al (2009) A translational signature for nucleosome positioning in vivo. Nucleic Acids Res 37(16):5309–5321

Champagne KS, Kutateladze TG (2009) Structural insight into histone recognition by the ING PHD fingers. Curr Drug Targets 10(5):432–441

Chang B, Chen Y, Zhao Y, Bruick RK (2007) JMJD6 is a histone arginine demethylase. Science 318(5849):444–447

Chang FT, Chan FL, JD RM, Udugama M, Mayne L, Collas P et al (2015) CHK1-driven histone H3.3 serine 31 phosphorylation is important for chromatin maintenance and cell survival in human ALT cancer cells. Nucleic Acids Res 43(5):2603–2614

Chen Y, Yang Y, Wang F, Wan K, Yamane K, Zhang Y et al (2006) Crystal structure of human histone lysine-specific demethylase 1 (LSD1). Proc Natl Acad Sci U S A 103(38):13956–13961

Chen HP, Zhao YT, Zhao TC (2015) Histone deacetylases and mechanisms of regulation of gene expression. Crit Rev Oncog 20(1-2):35–47

Choi JK (2010) Contrasting chromatin organization of CpG islands and exons in the human genome. Genome Biol 11(7):R70

Church M, Smith KC, Alhussain MM, Pennings S, Fleming AB (2017) Sas3 and Ada2(Gcn5)-dependent histone H3 acetylation is required for transcription elongation at the de-repressed FLO1 gene. Nucleic Acids Res 45(8):4413–4430

Clapier CR, Iwasa J, Cairns BR, Peterson CL (2017) Mechanisms of action and regulation of ATP-dependent chromatin-remodelling complexes. Nat Rev Mol Cell Biol 18(7):407–422

Creyghton MP, Markoulaki S, Levine SS, Hanna J, Lodato MA, Sha K et al (2008) H2AZ is enriched at polycomb complex target genes in ES cells and is necessary for lineage commitment. Cell 135(4):649–661

Creyghton MP, Cheng AW, Welstead GG, Kooistra T, Carey BW, Steine EJ et al (2010) Histone H3K27ac separates active from poised enhancers and predicts developmental state. Proc Natl Acad Sci U S A 107(50):21931–21936

Dai HQ, Wang BA, Yang L, Chen JJ, Zhu GC, Sun ML et al (2016) TET-mediated DNA demethylation controls gastrulation by regulating Lefty-Nodal signalling. Nature 538 (7626):528–532

Dalvai M, Bellucci L, Fleury L, Lavigne AC, Moutahir F, Bystricky K (2013) H2A.Z-dependent crosstalk between enhancer and promoter regulates cyclin D1 expression. Oncogene 32 (36):4243–4251

De S, Wurster AL, Precht P, Wood WH 3rd, Becker KG, Pazin MJ (2011) Dynamic BRG1 recruitment during T helper differentiation and activation reveals distal regulatory elements. Mol Cell Biol 31(7):1512–1527

Dhami P, Saffrey P, Bruce AW, Dillon SC, Chiang K, Bonhoure N et al (2010) Complex exon-intron marking by histone modifications is not determined solely by nucleosome distribution. PLoS One 5(8):e12339

Dovey OM, Foster CT, Cowley SM (2010) Histone deacetylase 1 (HDAC1), but not HDAC2, controls embryonic stem cell differentiation. Proc Natl Acad Sci U S A 107(18):8242–8247

Edmunds JW, Mahadevan LC, Clayton AL (2008) Dynamic histone H3 methylation during gene induction: HYPB/Setd2 mediates all H3K36 trimethylation. EMBO J 27(2):406–420

Ellis SE, Gupta S, Moes A, West AB, Arking DE (2017) Exaggerated CpH methylation in the autism-affected brain. Molecular Autism 8:6

Ficz G, Branco MR, Seisenberger S, Santos F, Krueger F, Hore TA et al (2011) Dynamic regulation of 5-hydroxymethylcytosine in mouse ES cells and during differentiation. Nature 473 (7347):398–402

Gaertner B, Johnston J, Chen K, Wallaschek N, Paulson A, Garruss AS et al (2012) Poised RNA polymerase II changes over developmental time and prepares genes for future expression. Cell Rep 2(6):1670–1683

Gal-Yam EN, Egger G, Iniguez L, Holster H, Einarsson S, Zhang X 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(35):12979–12984

Ge Z, Wang H, Parthun MR (2011) Nuclear Hat1p complex (NuB4) components participate in DNA repair-linked chromatin reassembly. J Biol Chem 286(19):16790–16799

Gelfman S, Cohen N, Yearim A, Ast G (2013) DNA-methylation effect on cotranscriptional splicing is dependent on GC architecture of the exon-intron structure. Genome Res 23 (5):789–799

Geyer PK, Green MM, Corces VG (1990) Tissue-specific transcriptional enhancers may act in trans on the gene located in the homologous chromosome: the molecular basis of transvection in Drosophila. EMBO J 9(7):2247–2256

Ghisletti S, Barozzi I, Mietton F, Polletti S, De Santa F, Venturini E et al (2010) Identification and characterization of enhancers controlling the inflammatory gene expression program in macrophages. Immunity 32(3):317–328

Goldberg AD, Banaszynski LA, Noh KM, Lewis PW, Elsaesser SJ, Stadler S et al (2010) Distinct factors control histone variant H3.3 localization at specific genomic regions. Cell 140 (5):678–691

Guo JU, Su Y, Shin JH, Shin J, Li H, Xie B et al (2014) Distribution, recognition and regulation of non-CpG methylation in the adult mammalian brain. Nat Neurosci 17(2):215–222

Hahn MA, Wu X, Li AX, Hahn T, Pfeifer GP (2011) Relationship between gene body DNA methylation and intragenic H3K9me3 and H3K36me3 chromatin marks. PLoS One 6(4):e18844

Han P, Li W, Lin CH, Yang J, Shang C, Nuernberg ST et al (2014) A long noncoding RNA protects the heart from pathological hypertrophy. Nature 514(7520):102–106

Hassan AH, Prochasson P, Neely KE, Galasinski SC, Chandy M, Carrozza MJ et al (2002) Function and selectivity of bromodomains in anchoring chromatin-modifying complexes to promoter nucleosomes. Cell 111(3):369–379

He HH, Meyer CA, Shin H, Bailey ST, Wei G, Wang Q et al (2010) Nucleosome dynamics define transcriptional enhancers. Nat Genet 42(4):343–347

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

Heintzman ND, Stuart RK, Hon G, Fu Y, Ching CW, Hawkins RD et al (2007) Distinct and predictive chromatin signatures of transcriptional promoters and enhancers in the human genome. Nat Genet 39(3):311–318

Hesson LB, Sloane MA, Wong JW, Nunez AC, Srivastava S, Ng B et al (2014) Altered promoter nucleosome positioning is an early event in gene silencing. Epigenetics 9(10):1422–1430

Hinshelwood RA, Melki JR, Huschtscha LI, Paul C, Song JZ, Stirzaker C et al (2009) Aberrant de novo methylation of the p16INK4A CpG island is initiated post gene silencing in association with chromatin remodelling and mimics nucleosome positioning. Hum Mol Genet 18 (16):3098–3109

Hodawadekar SC, Marmorstein R (2007) Chemistry of acetyl transfer by histone modifying enzymes: structure, mechanism and implications for effector design. Oncogene 26 (37):5528–5540

Hon GC, Hawkins RD, Ren B (2009) Predictive chromatin signatures in the mammalian genome. Hum Mol Genet 18(R2):R195–R201

Horton JR, Upadhyay AK, Qi HH, Zhang X, Shi Y, Cheng X (2010) Enzymatic and structural insights for substrate specificity of a family of jumonji histone lysine demethylases. Nat Struct Mol Biol 17(1):38–43

Hu G, Schones DE, Cui K, Ybarra R, Northrup D, Tang Q et al (2011) Regulation of nucleosome landscape and transcription factor targeting at tissue-specific enhancers by BRG1. Genome Res 21(10):1650–1658

Hu G, Cui K, Northrup D, Liu C, Wang C, Tang Q et al (2013) H2A.Z facilitates access of active and repressive complexes to chromatin in embryonic stem cell self-renewal and differentiation. Cell Stem Cell 12(2):180–192

Hu Z, Chen K, Xia Z, Chavez M, Pal S, Seol JH et al (2014) Nucleosome loss leads to global transcriptional up-regulation and genomic instability during yeast aging. Genes Dev 28 (4):396–408

Huang H, Yu S, Liu H, Sun X (2012) Nucleosome organization in sequences of alternative events in human genome. Biosystems 109(2):214–219
Ichikawa Y, Connelly CF, Appleboim A, Miller TC, Jacobi H, Abshiru NA et al (2017) A synthetic biology approach to probing nucleosome symmetry. Elife 6:e28836
Ito S, D'Alessio AC, Taranova OV, Hong K, Sowers LC, Zhang Y (2010) Role of Tet proteins in 5mC to 5hmC conversion, ES-cell self-renewal and inner cell mass specification. Nature 466 (7310):1129–1133
Jeong KW, Kim K, Situ AJ, Ulmer TS, An W, Stallcup MR (2011) Recognition of enhancer element-specific histone methylation by TIP60 in transcriptional activation. Nat Struct Mol Biol 18(12):1358–1365
Jin C, Felsenfeld G (2007) Nucleosome stability mediated by histone variants H3.3 and H2A.Z. Genes Dev 21(12):1519–1529
Jin C, Zang C, Wei G, Cui K, Peng W, Zhao K et al (2009) H3.3/H2A.Z double variant-containing nucleosomes mark 'nucleosome-free regions' of active promoters and other regulatory regions. Nat Genet 41(8):941–945
Jin Q, Yu LR, Wang L, Zhang Z, Kasper LH, Lee JE et al (2011) Distinct roles of GCN5/PCAF-mediated H3K9ac and CBP/p300-mediated H3K18/27ac in nuclear receptor transactivation. EMBO J 30(2):249–262
John S, Sabo PJ, Johnson TA, Sung MH, Biddie SC, Lightman SL et al (2008) Interaction of the glucocorticoid receptor with the chromatin landscape. Mol Cell 29(5):611–624
Johnson CN, Adkins NL, Georgel P (2005) Chromatin remodeling complexes: ATP-dependent machines in action. Biochem Cell Biol 83(4):405–417
Jones PA (2012) Functions of DNA methylation: islands, start sites, gene bodies and beyond. Nat Rev Genet 13(7):484–492
Kalakonda N, Fischle W, Boccuni P, Gurvich N, Hoya-Arias R, Zhao X et al (2008) Histone H4 lysine 20 monomethylation promotes transcriptional repression by L3MBTL1. Oncogene 27 (31):4293–4304
Kawaguchi T, Tanigawa A, Naganuma T, Ohkawa Y, Souquere S, Pierron G et al (2015) SWI/SNF chromatin-remodeling complexes function in noncoding RNA-dependent assembly of nuclear bodies. Proc Natl Acad Sci U S A 112(14):4304–4309
Kelly TJ, Qin S, Gottschling DE, Parthun MR (2000) Type B histone acetyltransferase Hat1p participates in telomeric silencing. Mol Cell Biol 20(19):7051–7058
Kelly TK, Miranda TB, Liang G, Berman BP, Lin JC, Tanay A et al (2010) H2A.Z maintenance during mitosis reveals nucleosome shifting on mitotically silenced genes. Mol Cell 39 (6):901–911
Kelly TK, Liu Y, Lay FD, Liang G, Berman BP, Jones PA (2012) Genome-wide mapping of nucleosome positioning and DNA methylation within individual DNA molecules. Genome Res 22(12):2497–2506
Kia SK, Gorski MM, Giannakopoulos S, Verrijzer CP (2008) SWI/SNF mediates polycomb eviction and epigenetic reprogramming of the INK4b-ARF-INK4a locus. Mol Cell Biol 28 (10):3457–3464
Kiefer CM, Hou C, Little JA, Dean A (2008) Epigenetics of beta-globin gene regulation. Mutat Res 647(1-2):68–76
Kim J, Daniel J, Espejo A, Lake A, Krishna M, Xia L et al (2006) Tudor, MBT and chromo domains gauge the degree of lysine methylation. EMBO Rep 7(4):397–403
Kim TK, Hemberg M, Gray JM, Costa AM, Bear DM, Wu J et al (2010) Widespread transcription at neuronal activity-regulated enhancers. Nature 465(7295):182–187
Klose RJ, Zhang Y (2007) Regulation of histone methylation by demethylimination and demethylation. Nat Rev Mol Cell Biol 8(4):307–318
Koch F, Andrau JC (2011) Initiating RNA polymerase II and TIPs as hallmarks of enhancer activity and tissue-specificity. Transcription 2(6):263–268
Kolasinska-Zwierz P, Down T, Latorre I, Liu T, Liu XS, Ahringer J (2009) Differential chromatin marking of introns and expressed exons by H3K36me3. Nat Genet 41(3):376–381

Kouzarides T (2007) Chromatin modifications and their function. Cell 128(4):693–705
Krebs AR, Karmodiya K, Lindahl-Allen M, Struhl K, Tora L (2011) SAGA and ATAC histone acetyl transferase complexes regulate distinct sets of genes and ATAC defines a class of p300-independent enhancers. Mol Cell 44(3):410–423
Kretzmer H, Bernhart SH, Wang W, Haake A, Weniger MA, Bergmann AK et al (2015) DNA methylome analysis in Burkitt and follicular lymphomas identifies differentially methylated regions linked to somatic mutation and transcriptional control. Nat Genet 47(11):1316–1325
Kriaucionis S, Heintz N (2009) The nuclear DNA base 5-hydroxymethylcytosine is present in Purkinje neurons and the brain. Science 324(5929):929–930
Ku M, Jaffe JD, Koche RP, Rheinbay E, Endoh M, Koseki H et al (2012) H2A.Z landscapes and dual modifications in pluripotent and multipotent stem cells underlie complex genome regulatory functions. Genome Biol 13(10):R85
Kubik S, Bruzzone MJ, Jacquet P, Falcone JL, Rougemont J, Shore D (2015) Nucleosome stability distinguishes two different promoter types at all protein-coding genes in yeast. Mol Cell 60 (3):422–434
Kundaje A, Kyriazopoulou-Panagiotopoulou S, Libbrecht M, Smith CL, Raha D, Winters EE et al (2012) Ubiquitous heterogeneity and asymmetry of the chromatin environment at regulatory elements. Genome Res 22(9):1735–1747
Kundu S, Ji F, Sunwoo H, Jain G, Lee JT, Sadreyev RI et al (2017) Polycomb repressive complex 1 generates discrete compacted domains that change during differentiation. Mol Cell 65 (3):432–446. e5
Kwon H, Song K, Han C, Zhang J, Lu L, Chen W et al (2017) Epigenetic silencing of miRNA-34a in human cholangiocarcinoma via EZH2 and DNA methylation: impact on regulation of notch pathway. Am J Pathol 187(10):2288–2299
Laurette P, Strub T, Koludrovic D, Keime C, Le Gras S, Seberg H et al (2015) Transcription factor MITF and remodeller BRG1 define chromatin organisation at regulatory elements in melanoma cells. Elife 4:e06857
Lee J, Thompson JR, Botuyan MV, Mer G (2008) Distinct binding modes specify the recognition of methylated histones H3K4 and H4K20 by JMJD2A-tudor. Nat Struct Mol Biol 15(1):109–111
Lee Y, Park D, Iyer VR (2017a) The ATP-dependent chromatin remodeler Chd1 is recruited by transcription elongation factors and maintains H3K4me3/H3K36me3 domains at actively transcribed and spliced genes. Nucleic Acids Res 45(14):8646
Lee SM, Lee J, Noh KM, Choi WY, Jeon S, Oh GT et al (2017b) Intragenic CpG islands play important roles in bivalent chromatin assembly of developmental genes. Proc Natl Acad Sci U S A 114(10):E1885–E1e94
Leisegang MS, Fork C, Josipovic I, Richter FM, Preussner J, Hu J et al (2017) Long noncoding RNA MANTIS facilitates endothelial angiogenic function. Circulation 136(1):65–79
Levendosky RF, Sabantsev A, Deindl S, Bowman GD (2016) The Chd1 chromatin remodeler shifts hexasomes unidirectionally. Elife 5:e21356
Li Q, Zhou H, Wurtele H, Davies B, Horazdovsky B, Verreault A et al (2008) Acetylation of histone H3 lysine 56 regulates replication-coupled nucleosome assembly. Cell 134(2):244–255
Li XY, Thomas S, Sabo PJ, Eisen MB, Stamatoyannopoulos JA, Biggin MD (2011) The role of chromatin accessibility in directing the widespread, overlapping patterns of Drosophila transcription factor binding. Genome Biol 12(4):R34
Li L, Liu D, Bu D, Chen S, Wu J, Tang C et al (2013) Brg1-dependent epigenetic control of vascular smooth muscle cell proliferation by hydrogen sulfide. Biochim Biophys Acta 1833 (6):1347–1355
Lin JC, Jeong S, Liang G, Takai D, Fatemi M, Tsai YC et al (2007) Role of nucleosomal occupancy in the epigenetic silencing of the MLH1 CpG island. Cancer Cell 12(5):432–444
Lin JJ, Lehmann LW, Bonora G, Sridharan R, Vashisht AA, Tran N et al (2011) Mediator coordinates PIC assembly with recruitment of CHD1. Genes Dev 25(20):2198–2209

Lister R, Pelizzola M, Dowen RH, Hawkins RD, Hon G, Tonti-Filippini J et al (2009) Human DNA methylomes at base resolution show widespread epigenomic differences. Nature 462 (7271):315–322

Lo KA, Bauchmann MK, Baumann AP, Donahue CJ, Thiede MA, Hayes LS et al (2011) Genome-wide profiling of H3K56 acetylation and transcription factor binding sites in human adipocytes. PLoS One 6(6):e19778

Lohr D, Tatchell K, Van Holde KE (1977) On the occurrence of nucleosome phasing in chromatin. Cell 12(3):829–836

Lu X, Simon MD, Chodaparambil JV, Hansen JC, Shokat KM, Luger K (2008) The effect of H3K79 dimethylation and H4K20 trimethylation on nucleosome and chromatin structure. Nat Struct Mol Biol 15(10):1122–1124

Luebben WR, Sharma N, Nyborg JK (2010) Nucleosome eviction and activated transcription require p300 acetylation of histone H3 lysine 14. Proc Natl Acad Sci U S A 107 (45):19254–19259

Luger K, Mader 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

Maclary E, Hinten M, Harris C, Sethuraman S, Gayen S, Kalantry S (2017) PRC2 represses transcribed genes on the imprinted inactive X chromosome in mice. Genome Biol 18(1):82

Margaritis T, Holstege FC (2008) Poised RNA polymerase II gives pause for thought. Cell 133 (4):581–584

Maurer-Stroh S, Dickens NJ, Hughes-Davies L, Kouzarides T, Eisenhaber F, Ponting CP (2003) The Tudor domain 'Royal Family': Tudor, plant Agenet, Chromo, PWWP and MBT domains. Trends Biochem Sci 28(2):69–74

May D, Blow MJ, Kaplan T, McCulley DJ, Jensen BC, Akiyama JA et al (2011) Large-scale discovery of enhancers from human heart tissue. Nat Genet 44(1):89–93

Mercer EM, Lin YC, Benner C, Jhunjhunwala S, Dutkowski J, Flores M et al (2011) Multilineage priming of enhancer repertoires precedes commitment to the B and myeloid cell lineages in hematopoietic progenitors. Immunity 35(3):413–425

Mikkelsen TS, Ku M, Jaffe DB, Issac B, Lieberman E, Giannoukos G et al (2007) Genome-wide maps of chromatin state in pluripotent and lineage-committed cells. Nature 448(7153):553–560

Moazed D (2009) Small RNAs in transcriptional gene silencing and genome defence. Nature 457 (7228):413–420

Morris SA, Baek S, Sung MH, John S, Wiench M, Johnson TA et al (2014) Overlapping chromatin-remodeling systems collaborate genome wide at dynamic chromatin transitions. Nat Struct Mol Biol 21(1):73–81

Morrison AJ (2017) Genome maintenance functions of the INO80 chromatin remodeller. Philos Trans R Soc Lond B Biol Sci 372(1731):20160289

Morrison EA, Sanchez JC, Ronan JL, Farrell DP, Varzavand K, Johnson JK et al (2017) DNA binding drives the association of BRG1/hBRM bromodomains with nucleosomes. Nat Commun 8:16080

Muller HP, Schaffner W (1990) Transcriptional enhancers can act in trans. Trends Genet 6 (9):300–304

Musselman CA, Gibson MD, Hartwick EW, North JA, Gatchalian J, Poirier MG et al (2013) Binding of PHF1 Tudor to H3K36me3 enhances nucleosome accessibility. Nat Commun 4:2969

Nagai S, Davis RE, Mattei PJ, Eagen KP, Kornberg RD (2017) Chromatin potentiates transcription. Proc Natl Acad Sci U S A 114(7):1536–1541

Ndlovu MN, Denis H, Fuks F (2011) Exposing the DNA methylome iceberg. Trends Biochem Sci 36(7):381–387

Nishioka K, Chuikov S, Sarma K, Erdjument-Bromage H, Allis CD, Tempst P et al (2002) Set9, a novel histone H3 methyltransferase that facilitates transcription by precluding histone tail modifications required for heterochromatin formation. Genes Dev 16(4):479–489

Ocampo J, Chereji RV, Eriksson PR, Clark DJ (2016) The ISW1 and CHD1 ATP-dependent chromatin remodelers compete to set nucleosome spacing in vivo. Nucleic Acids Res 44 (10):4625–4635

Ong CT, Corces VG (2011) Enhancer function: new insights into the regulation of tissue-specific gene expression. Nat Rev Genet 12(4):283–293

Ong CT, Corces VG (2012) Enhancers: emerging roles in cell fate specification. EMBO Rep 13 (5):423–430

Onorati MC, Lazzaro S, Mallik M, Ingrassia AM, Carreca AP, Singh AK et al (2011) The ISWI chromatin remodeler organizes the hsromega ncRNA-containing omega speckle nuclear compartments. PLoS Genet 7(5):e1002096

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

Parthun MR (2007) Hat1: the emerging cellular roles of a type B histone acetyltransferase. Oncogene 26(37):5319–5328

Pasini D, Malatesta M, Jung HR, Walfridsson J, Willer A, Olsson L et al (2010) Characterization of an antagonistic switch between histone H3 lysine 27 methylation and acetylation in the transcriptional regulation of Polycomb group target genes. Nucleic Acids Res 38 (15):4958–4969

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

Peckham HE, Thurman RE, Fu Y, Stamatoyannopoulos JA, Noble WS, Struhl K et al (2007) Nucleosome positioning signals in genomic DNA. Genome Res 17(8):1170–1177

Pekowska A, Benoukraf T, Zacarias-Cabeza J, Belhocine M, Koch F, Holota H et al (2011) H3K4 tri-methylation provides an epigenetic signature of active enhancers. EMBO J 30 (20):4198–4210

Qin S, Parthun MR (2006) Recruitment of the type B histone acetyltransferase Hat1p to chromatin is linked to DNA double-strand breaks. Mol Cell Biol 26(9):3649–3658

Rada-Iglesias A, Bajpai R, Swigut T, Brugmann SA, Flynn RA, Wysocka J (2011) A unique chromatin signature uncovers early developmental enhancers in humans. Nature 470 (7333):279–283

Raisner RM, Madhani HD (2006) Patterning chromatin: form and function for H2A.Z variant nucleosomes. Curr Opin Genet Dev 16(2):119–124

Ramachandran S, Zentner GE, Henikoff S (2015) Asymmetric nucleosomes flank promoters in the budding yeast genome. Genome Res 25(3):381–390

Rea S, Eisenhaber F, O'Carroll D, Strahl BD, Sun ZW, Schmid M et al (2000) Regulation of chromatin structure by site-specific histone H3 methyltransferases. Nature 406(6796):593–599

Reddington JP, Perricone SM, Nestor CE, Reichmann J, Youngson NA, Suzuki M et al (2013) Redistribution of H3K27me3 upon DNA hypomethylation results in de-repression of Polycomb target genes. Genome Biol 14(3):R25

Reisman D, Glaros S, Thompson EA (2009) The SWI/SNF complex and cancer. Oncogene 28 (14):1653–1668

Rhee HS, Bataille AR, Zhang L, Pugh BF (2014) Subnucleosomal structures and nucleosome asymmetry across a genome. Cell 159(6):1377–1388

Rinn JL, Kertesz M, Wang JK, Squazzo SL, Xu X, Brugmann SA et al (2007) Functional demarcation of active and silent chromatin domains in human HOX loci by noncoding RNAs. Cell 129(7):1311–1323

Rondelet G, Dal Maso T, Willems L, Wouters J (2016) Structural basis for recognition of histone H3K36me3 nucleosome by human de novo DNA methyltransferases 3A and 3B. J Struct Biol 194(3):357–367

Rufiange A, Jacques PE, Bhat W, Robert F, Nourani A (2007) Genome-wide replication-independent histone H3 exchange occurs predominantly at promoters and implicates H3 K56 acetylation and Asf1. Mol Cell 27(3):393–405

Saha A, Wittmeyer J, Cairns BR (2006) Mechanisms for nucleosome movement by ATP-dependent chromatin remodeling complexes. Results Probl Cell Differ 41:127–148

Schneider R, Bannister AJ, Myers FA, Thorne AW, Crane-Robinson C, Kouzarides T (2004) Histone H3 lysine 4 methylation patterns in higher eukaryotic genes. Nat Cell Biol 6(1):73–77

Schnetz MP, Bartels CF, Shastri K, Balasubramanian D, Zentner GE, Balaji R et al (2009) Genomic distribution of CHD7 on chromatin tracks H3K4 methylation patterns. Genome Res 19 (4):590–601

Segal E, Fondufe-Mittendorf Y, Chen L, Thastrom A, Field Y, Moore IK et al (2006) A genomic code for nucleosome positioning. Nature 442(7104):772–778

Sen S, Block KF, Pasini A, Baylin SB, Easwaran H (2016) Genome-wide positioning of bivalent mononucleosomes. BMC Med Genomics 9(1):60

Seto E, Yoshida M (2014) Erasers of histone acetylation: the histone deacetylase enzymes. Cold Spring Harb Perspect Biol 6(4):a018713

Sharma S, Kelly TK, Jones PA (2010) Epigenetics in cancer. Carcinogenesis 31(1):27–36

Shema E, Jones D, Shoresh N, Donohue L, Ram O, Bernstein BE (2016) Single-molecule decoding of combinatorially modified nucleosomes. Science 352(6286):717–721

Shen L, Zhang Y (2013) 5-Hydroxymethylcytosine: generation, fate, and genomic distribution. Curr Opin Cell Biol 25(3):289–296

Shen W, Xu C, Huang W, Zhang J, Carlson JE, Tu X et al (2007) Solution structure of human Brg1 bromodomain and its specific binding to acetylated histone tails. Biochemistry 46 (8):2100–2110

Shen Y, Matsuno Y, Fouse SD, Rao N, Root S, Xu R et al (2008) X-inactivation in female human embryonic stem cells is in a nonrandom pattern and prone to epigenetic alterations. Proc Natl Acad Sci U S A 105(12):4709–4714

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

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

Shogren-Knaak M, Ishii H, Sun JM, Pazin MJ, Davie JR, Peterson CL (2006) Histone H4-K16 acetylation controls chromatin structure and protein interactions. Science 311(5762):844–847

Sims RJ 3rd, Chen CF, Santos-Rosa H, Kouzarides T, Patel SS, Reinberg D (2005) Human but not yeast CHD1 binds directly and selectively to histone H3 methylated at lysine 4 via its tandem chromodomains. J Biol Chem 280(51):41789–41792

Skulte KA, Phan L, Clark SJ, Taberlay PC (2014) Chromatin remodeler mutations in human cancers: epigenetic implications. Epigenomics 6(4):397–414

Skvortsova K, Zotenko E, Luu PL, Gould CM, Nair SS, Clark SJ et al (2017) Comprehensive evaluation of genome-wide 5-hydroxymethylcytosine profiling approaches in human DNA. Epigenetics Chromatin 10:16

Spitz F, Furlong EE (2012) Transcription factors: from enhancer binding to developmental control. Nat Rev Genet 13(9):613–626

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(7378):490–495

Stejskal S, Stepka K, Tesarova L, Stejskal K, Matejkova M, Simara P et al (2015) Cell cycle-dependent changes in H3K56ac in human cells. Cell Cycle 14(24):3851–3863

Steunou A-L, Rossetto D, Côté J (2014) Regulating chromatin by histone acetylation. In: Workman JL, Abmayr SM (eds) Fundamentals of chromatin. Springer, New York, NY, pp 147–212

Stroud H, Feng S, Morey Kinney S, Pradhan S, Jacobsen SE (2011) 5-Hydroxymethylcytosine is associated with enhancers and gene bodies in human embryonic stem cells. Genome Biol 12(6): R54

Sun XJ, Wei J, Wu XY, Hu M, Wang L, Wang HH et al (2005) Identification and characterization of a novel human histone H3 lysine 36-specific methyltransferase. J Biol Chem 280 (42):35261–35271

Taberlay PC, Kelly TK, Liu CC, You JS, De Carvalho DD, Miranda TB et al (2011) Polycomb-repressed genes have permissive enhancers that initiate reprogramming. Cell 147(6):1283–1294

Taberlay PC, Statham AL, Kelly TK, Clark SJ, Jones PA (2014) Reconfiguration of nucleosome-depleted regions at distal regulatory elements accompanies DNA methylation of enhancers and insulators in cancer. Genome Res 24(9):1421–1432

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

Takai D, Jones PA (2002) Comprehensive analysis of CpG islands in human chromosomes 21 and 22. Proc Natl Acad Sci U S A 99(6):3740–3745

Tamkun JW, Deuring R, Scott MP, Kissinger M, Pattatucci AM, Kaufman TC et al (1992) brahma: a regulator of Drosophila homeotic genes structurally related to the yeast transcriptional activator SNF2/SWI2. Cell 68(3):561–572

Teissandier A, Bourc'his D (2017) Gene body DNA methylation conspires with H3K36me3 to preclude aberrant transcription. EMBO J 36(11):1471–1473

Tessarz P, Kouzarides T (2014) Histone core modifications regulating nucleosome structure and dynamics. Nat Rev Mol Cell Biol 15(11):703–708

Thomson JP, Skene PJ, Selfridge J, Clouaire T, Guy J, Webb S et al (2010) CpG islands influence chromatin structure via the CpG-binding protein Cfp1. Nature 464(7291):1082–1086

Thomson JP, Hunter JM, Lempiainen H, Muller A, Terranova R, Moggs JG et al (2013) Dynamic changes in 5-hydroxymethylation signatures underpin early and late events in drug exposed liver. Nucleic Acids Res 41(11):5639–5654

Thurman RE, Rynes E, Humbert R, Vierstra J, Maurano MT, Haugen E et al (2012) The accessible chromatin landscape of the human genome. Nature 489(7414):75–82

Tjeertes JV, Miller KM, Jackson SP (2009) Screen for DNA-damage-responsive histone modifications identifies H3K9Ac and H3K56Ac in human cells. EMBO J 28(13):1878–1889

Tolstorukov MY, Sansam CG, Lu P, Koellhoffer EC, Helming KC, Alver BH et al (2013) Swi/Snf chromatin remodeling/tumor suppressor complex establishes nucleosome occupancy at target promoters. Proc Natl Acad Sci U S A 110(25):10165–10170

Trojer P, Li G, Sims RJ 3rd, Vaquero A, Kalakonda N, Boccuni P et al (2007) L3MBTL1, a histone-methylation-dependent chromatin lock. Cell 129(5):915–928

Tropberger P, Pott S, Keller C, Kamieniarz-Gdula K, Caron M, Richter F et al (2013) Regulation of transcription through acetylation of H3K122 on the lateral surface of the histone octamer. Cell 152(4):859–872

Tsai MC, Manor O, Wan Y, Mosammaparast N, Wang JK, Lan F et al (2010) Long noncoding RNA as modular scaffold of histone modification complexes. Science 329(5992):689–693

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

Valdes-Mora F, Song JZ, Statham AL, Strbenac D, Robinson MD, Nair SS et al (2012) Acetylation of H2A.Z is a key epigenetic modification associated with gene deregulation and epigenetic remodeling in cancer. Genome Res 22(2):307–321

Valdes-Mora F, Gould CM, Colino-Sanguino Y, Qu W, Song JZ, Taylor KM et al (2017) Acetylated histone variant H2A.Z is involved in the activation of neo-enhancers in prostate cancer. Nat Commun 8(1):1346

Valouev A, Johnson SM, Boyd SD, Smith CL, Fire AZ, Sidow A (2011) Determinants of nucleosome organization in primary human cells. Nature 474(7352):516–520

Vempati RK, Jayani RS, Notani D, Sengupta A, Galande S, Haldar D (2010) p300-mediated acetylation of histone H3 lysine 56 functions in DNA damage response in mammals. J Biol Chem 285(37):28553–28564

Venkatesh S, Workman JL (2015) Histone exchange, chromatin structure and the regulation of transcription. Nat Rev Mol Cell Biol 16(3):178–189

Verma N, Pan H, Dore LC, Shukla A, Li QV, Pelham-Webb B et al (2018) TET proteins safeguard bivalent promoters from de novo methylation in human embryonic stem cells. Nat Genet 50 (1):83–95

Vire 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

Visel A, Blow MJ, Li Z, Zhang T, Akiyama JA, Holt A et al (2009) ChIP-seq accurately predicts tissue-specific activity of enhancers. Nature 457(7231):854–858

Voigt P, LeRoy G, Drury WJ 3rd, Zee BM, Son J, Beck DB et al (2012) Asymmetrically modified nucleosomes. Cell 151(1):181–193

Voigt P, Tee WW, Reinberg D (2013) A double take on bivalent promoters. Genes Dev 27 (12):1318–1338

Wang Z, Zang C, Rosenfeld JA, Schones DE, Barski A, Cuddapah S et al (2008) Combinatorial patterns of histone acetylations and methylations in the human genome. Nat Genet 40 (7):897–903

Wang J, Qin S, Li F, Li S, Zhang W, Peng J et al (2014) Crystal structure of human BS69 Bromo-ZnF-PWWP reveals its role in H3K36me3 nucleosome binding. Cell Res 24(7):890–893

Wang DY, An SH, Liu L, Bai SS, Wu KX, Zhu R et al (2016) Hepatitis B virus X protein influences enrichment profiles of H3K9me3 on promoter regions in human hepatoma cell lines. Oncotarget 7(51):84883–84892

Wani AH, Boettiger AN, Schorderet P, Ergun A, Munger C, Sadreyev RI et al (2016) Chromatin topology is coupled to Polycomb group protein subnuclear organization. Nat Commun 7:10291

Whetstine JR, Nottke A, Lan F, Huarte M, Smolikov S, Chen Z et al (2006) Reversal of histone lysine trimethylation by the JMJD2 family of histone demethylases. Cell 125(3):467–481

Widschwendter M, Fiegl H, Egle D, Mueller-Holzner E, Spizzo G, Marth C et al (2007) Epigenetic stem cell signature in cancer. Nat Genet 39(2):157–158

Wilson BG, Wang X, Shen X, McKenna ES, Lemieux ME, Cho YJ et al (2010) Epigenetic antagonism between polycomb and SWI/SNF complexes during oncogenic transformation. Cancer Cell 18(4):316–328

Wurtele H, Kaiser GS, Bacal J, St-Hilaire E, Lee EH, Tsao S et al (2012) Histone H3 lysine 56 acetylation and the response to DNA replication fork damage. Mol Cell Biol 32(1):154–172

Wysocka J, Swigut T, Xiao H, Milne TA, Kwon SY, Landry J et al (2006) A PHD finger of NURF couples histone H3 lysine 4 trimethylation with chromatin remodelling. Nature 442 (7098):86–90

Xhemalce B, Dawson MA, Bannister AJ (2011) Histone modifications. Reviews in cell biology and molecular medicine. Wiley-VCH Verlag GmbH & Co. KGaA, Weinheim

Xia RH, Wang Z, Zhang CY, Hu YH, Zhou RR, Wang LZ et al (2015) Low expression of endothelin receptor B (EDNRB) is related to H3K9me3 binding with the EDNRB promoter region and is associated with the clinical T tumor stage in salivary adenoid cystic carcinoma. Oral Surg Oral Med Oral Pathol Oral Radiol 120(2):258–268

Xiao B, Jing C, Wilson JR, Walker PA, Vasisht N, Kelly G et al (2003) Structure and catalytic mechanism of the human histone methyltransferase SET7/9. Nature 421(6923):652–656

Xu F, Zhang K, Grunstein M (2005) Acetylation in histone H3 globular domain regulates gene expression in yeast. Cell 121(3):375–385

Yang XJ, Seto E (2008) The Rpd3/Hda1 family of lysine deacetylases: from bacteria and yeast to mice and men. Nat Rev Mol Cell Biol 9(3):206–218

Yang X, Noushmehr H, Han H, Andreu-Vieyra C, Liang G, Jones PA (2012) Gene reactivation by 5-aza-2'-deoxycytidine-induced demethylation requires SRCAP-mediated H2A.Z insertion to establish nucleosome depleted regions. PLoS Genet 8(3):e1002604

You JS, Jones PA (2012) Cancer genetics and epigenetics: two sides of the same coin? Cancer Cell 22(1):9–20

You JS, Kelly TK, De Carvalho DD, Taberlay PC, Liang G, Jones PA (2011) OCT4 establishes and maintains nucleosome-depleted regions that provide additional layers of epigenetic regulation of its target genes. Proc Natl Acad Sci U S A 108(35):14497–14502

Zegerman P, Canas B, Pappin D, Kouzarides T (2002) Histone H3 lysine 4 methylation disrupts binding of nucleosome remodeling and deacetylase (NuRD) repressor complex. J Biol Chem 277(14):11621–11624

Zentner GE, Henikoff S (2013) Regulation of nucleosome dynamics by histone modifications. Nat Struct Mol Biol 20(3):259–266

Zentner GE, Tesar PJ, Scacheri PC (2011) Epigenetic signatures distinguish multiple classes of enhancers with distinct cellular functions. Genome Res 21(8):1273–1283

Zentner GE, Tsukiyama T, Henikoff S (2013) ISWI and CHD chromatin remodelers bind promoters but act in gene bodies. PLoS Genet 9(2):e1003317

Zhang X, Yang Z, Khan SI, Horton JR, Tamaru H, Selker EU et al (2003) Structural basis for the product specificity of histone lysine methyltransferases. Mol Cell 12(1):177–185

Zhang W, Prakash C, Sum C, Gong Y, Li Y, Kwok JJ et al (2012) Bromodomain-containing protein 4 (BRD4) regulates RNA polymerase II serine 2 phosphorylation in human CD4+ T cells. J Biol Chem 287(51):43137–43155

Zheng H, Huang B, Zhang B, Xiang Y, Du Z, Xu Q et al (2016) Resetting epigenetic memory by reprogramming of histone modifications in mammals. Mol Cell 63(6):1066–1079

Zhou Y, Zheng L, Li F, Wan M, Fan Y, Zhou X et al (2018) Bivalent histone codes on WNT5A during odontogenic differentiation. J Dent Res 97:99–107

Zhu K, Lei PJ, Ju LG, Wang X, Huang K, Yang B et al (2017) SPOP-containing complex regulates SETD2 stability and H3K36me3-coupled alternative splicing. Nucleic Acids Res 45(1):92–105

Circular RNAs in Human Health and Disease

5

Liam Grouse, Ashton Curry-Hyde, Bei Jun Chen, and Michael Janitz

Abstract

Non-coding RNA molecules have become increasingly more important in the realm of understanding the transcriptome. The functional role of the non-coding transcriptome has only relatively recently become an area of expansive research, largely through the methods of next-generation sequencing. Circular RNAs (circRNAs), a subtype of mostly non-coding RNAs, have recently been shown to have considerable abundance and cell-type- and development-specific expression patterns, in particular in human tissues. CircRNAs also show a distinct expression pattern in complex human diseases such as neurodegenerative diseases and cancer. This chapter focuses on the role of circRNAs as a biomarker in cancer using hepatocellular carcinoma (HCC) as an example. CircRNAs, as potential biomarkers of cancer, could not only play a significant role in diagnosis and monitoring of malignant growth but also might serve as primary targets for therapeutic intervention.

Keywords

Biomarkers · Circular RNAs · Hepatocellular carcinoma (HCC) · RNA-seq · Transcriptome

5.1 Introduction

Advancements in high-throughput RNA sequencing (RNA-seq) technologies and related bioinformatic tools have revolutionised our understanding of transcriptome. This has helped define molecular features of malignancies without the limitations of

L. Grouse · A. Curry-Hyde · B. J. Chen · M. Janitz (✉)
School of Biotechnology and Biomolecular Sciences, University of New South Wales, Sydney, NSW, Australia
e-mail: m.janitz@unsw.edu.au

L. B. Hesson, A. L. Pritchard (eds.), *Clinical Epigenetics*,
https://doi.org/10.1007/978-981-13-8958-0_5

preceding transcriptomic tools such as microarrays, which are limited by their inability to detect transcripts de novo, or Sanger sequencing of cDNA/expressed sequence tag libraries, which is limited by its high costs and low throughput. In order to understand the principals behind this revolution in transcriptomics characterisation, it is important to understand the different types of RNA produced by human cells and the methodology of linear RNA-seq.

5.2 Types of RNA

There are different types of RNA that can be produced by human cells, which have distinct functions. Approximately 3% of RNA produced encodes a protein template (messenger RNA), while the remaining 97% is non-coding RNA (ncRNA). The most abundant ncRNAs are transfer RNA (tRNA) and ribosomal RNA (rRNA), which are both vital to the process of mRNA translation. This chapter focuses on a type of ncRNA called circular RNAs (circRNAs), its role within the cell and how it can be involved in human disease (Palazzo and Lee 2015).

5.2.1 Circular RNAs

CircRNAs are single-stranded, generally non-coding RNA molecules formed by backsplicing: the process in which the 5′ and 3′ ends of linear precursor mRNA (pre-mRNA) are covalently linked to create a circular structure (Salzman et al. 2012). The site of circRNA linkage, forming a phosphodiester bond, is termed the backsplice junction. In addition to their shape, circRNAs are distinguished from linear RNAs by the absence of a poly(A) tail and a 5′ cap (Szabo and Salzman 2016). Due to their circular structure, exons within circRNAs appear out of order from their genomic template and linear RNA counterparts (Nigro et al. 1991). CircRNAs can exist as exonic circRNAs—circRNAs that contain single or multiple exons; intronic circRNAs/RNA lariats—circRNAs that contain intronic regions only; and exon–intron circRNAs—circRNAs that contain both exons and introns (Jeck and Sharpless 2014; Memczak et al. 2013).

The size of circRNAs can range from smaller than 100 nucleotides to larger than 4 kb, although in humans they are commonly composed of a few hundred nucleotides (Lasda and Parker 2014). CircRNAs have been identified for thousands of human genes, accounting for 20–30% of the human transcriptome (You et al. 2015). Increasing circRNA diversity is the ability for multiple circRNA isoforms to arise from the same gene, a phenomenon known as alternative circularisation (Zhang et al. 2014). Although alternative circularisation can produce tens of distinct circular products from a single gene, most genes only produce one or two distinct circRNAs (Szabo and Salzman 2016). CircRNAs are considerably more stable than linear RNAs. For mammalian cells, the median half-life of circRNAs (18.8–23.7 h) is at least 2.5 times longer than their linear counterparts (4.0–7.4 h). Nevertheless, the half-life of circRNAs can vary considerably and can be greater than 48 h (Enuka et al. 2016; Jeck et al. 2013).

Sequence analysis coupled with splice inhibitor assays has shown that both linear mRNAs and circRNAs are derived from pre-mRNA and its interactions with the canonical spliceosome (Starke et al. 2015; Guo et al. 2014). The co-expression of circRNAs and linear mRNAs could be competitive, leading to a negative correlation in expressions between the two RNA subtypes (Ashwal-Fluss et al. 2014); however, several studies also observed independent expressions of circular and linear RNA (Gao et al. 2015; Salzman et al. 2013; Veno et al. 2015). Although for most genes expressing circRNAs the abundance of circles is roughly 5–10% of linear counterparts (Salzman et al. 2013), for many other genes, however, circular transcripts are the primary RNA species transcribed from the host locus (Salzman et al. 2012; Veno et al. 2015; Rybak-Wolf et al. 2015). Moreover, some genes express circular transcripts preferentially in certain tissues. For example, *CDR1as* expression is exclusively circular in the brain (Memczak et al. 2013).

A good body of evidence has uncovered specific expression of many circRNAs across different cell types and developmental stages (Memczak et al. 2013; Salzman et al. 2013; Veno et al. 2015; Rybak-Wolf et al. 2015). For example, circRNA hsa_circRNA_2149 is specific to human leukocyte cell line $CD19^+$ and is absent in $CD34^+$ leukocytes, neutrophils or the embryonic kidney cell line HEK293 (Memczak et al. 2013). These observations are not exclusive to humans, with a study in mice reporting the exclusive presence of circRNA derived from gene *Rims2* in the granular layer of the cerebellum with an expression ~20-fold higher than its linear RNA counterpart (Rybak-Wolf et al. 2015).

5.3 Circular RNAs in Human Disease

Many circRNAs exhibit differential expressions in a diverse range of human disease such as neurodegenerative diseases and cancers (Chen et al. 2018, 2016; Li et al. 2015a; Zheng et al. 2016). For example, (1) in brains with multiple system atrophy (MSA), five circRNAs: *IQCK*, *MAP 4K3*, *EFCAB11*, *DTNA* and *MCTP1* have specific overexpression in the frontal cortex (Chen et al. 2016); (2) exosomal circRNAs in peripheral blood of colon cancer patients have unique expression patterns compared to those found in controls (Li et al. 2015a); (3) transcriptome analysis of endometrial cancer data reported 120 differentially expressed circRNA between tumour and healthy tissue samples (Chen et al. 2018). These examples of differential expression of circRNAs in diseases suggest regulatory roles of circRNAs and additionally make them potential biomarkers for non-invasive diagnosis of diseases, especially those of the brain.

5.4 Linear RNA-seq

In order to understand the recent advances in the circRNA research field, it is important to comprehend the RNA-seq technique in the context of linear RNA. A typical high-throughput (also known as next-generation sequencing) linear RNA-seq

workflow can be divided into three core stages: (1) library preparation, (2) sequencing and (3) data analysis.

5.4.1 Library Preparation

Library preparation is conducted pre-sequencing and involves selecting the desired RNA species, fragmenting it into 200–500 nucleotides, and attaching adaptors to its cDNA product. Selecting the desired RNA species often includes the removal of rRNA, which forms a high percentage of total RNA, which would otherwise consuming sequencing reads and reduce sequencing quality (Costa et al. 2010). rRNA removal can be achieved with poly(A) selection or rRNA depletion. RNA species can also be selected via gel electrophoresis that distinguishes based on RNA size and shape. Fragmentation of a given sequence is a necessary step during library preparation. Phred scores determine the estimated error of probability of bases called during downstream processing by assigning a quality score value to each base called (Cock et al. 2010). The higher the Phred score, the higher the base call accuracy (Endrullat et al. 2016). Fragmentation can occur using several methods such as nebulisation, sonification and enzymatic DNA digestion (Knierim et al. 2011). For small RNAs such as miRNAs and piRNAs, fragmentation is not necessary. A limitation of RNA-seq is that standard cDNA conversion does not maintain strand orientation. In order to maintain strand of origin information for each transcript, alternative library preparation is required but this is both time consuming and costly (Levin et al. 2010).

5.4.2 Sequencing

Over the years, a variety of high-throughput sequencing platforms have been developed. Despite sharing similar basic principles, Illumina platforms have dominated the sequencing industry (https://sapac.illumina.com/?langsel=/au/). Samples loaded into Illumina platforms first undergo cluster generation before sequencing can begin—each cDNA molecule is amplified in a separate well located on a slide, known as a flow cell. Clusters are generated via bridge amplification, which involves forming a cDNA bridge structure, synthesising the complementary strand with polymerase and denaturing the bridge to leave twice the amount of cDNA. In order to facilitate bridge formation, cDNA adaptors are hybridised to complementary oligos lining the surface of the flow cell. Once these clusters are generated, sequencing-by-synthesis (SBS) can begin. During SBS, RNA strands from each cluster have their complementary strand synthesised by polymerase. However, the nucleotides incorporated into the complementary strand are fluorescently tagged and contain a reversible terminator group. The reversible terminator group prevents more than one nucleotide from incorporating at a time to allow for the detection of each fluorescent signal. After the addition of a nucleotide, the fluorescent tag and terminator group are cleaved, allowing the next nucleotide to

be incorporated. A single read is generated when polymerase finishes incorporating all nucleotides into a complementary strand.

5.4.3 Data Analysis

Data analysis involves several steps, broadly grouped into assembly and expression analyses. Assembly involves aligning (mapping) sequencing reads to a reference genome and transcript assembly, while expression analysis can involve differential expression assessment and pathway analysis. Alignment tools primarily determine the genomic origin of each read based on sequence similarity. Selected alignment parameters should be flexible enough to account for sequencing and reference genome errors, while stringent enough to align without losing specificity. Reputable alignment tools include TopHat, GSNAP and STAR. By utilising different algorithms and pre-existing knowledge about isoform structure, transcript assembly can be achieved by tools such as Cufflinks, SLIDE and StringTie. Counting the number of reads corresponding to each transcript also allows assembly tools to quantify expression levels. There are typically three units by which expression is calculated: (1) RPKM (reads per kilobase million), (2) FPKM (fragments per kilobase million) and (3) TPM (transcripts per kilobase million). Further analysis can include a determination of differential expression between samples or comparison of transcript expression profiles within a sample, for which raw read counts must be normalised, then statistically determined to be significant. Normalisation corrects for bases which can arise from variation in gene length (Oshlack and Wakefield 2009), library fragment size and sequencing read depth (Mortazavi et al. 2008). Bias arising from read depth can be observed from sequencing identical samples at different depths; samples sequenced at a greater depth will generate more reads, and wrongly appear to have increased expression. After differential expression analysis, pathways analysis is a further assessment, which can be conducted to determine potential biological pathways affected by the differentially expressed genes.

5.5 CircRNA Detection Tools, Challenges and Solutions

CircRNAs can be identified by various biochemical tools and bioinformatic pipelines following RNA-seq. These analytical pipelines require two bioinformatic tools: a standard alignment tool and a circRNA detection tool. Unless RNase R treatment is applied (see below), reads sequenced and aligned in circRNA detection are derived from fragmented linear RNAs and circRNAs and then converted to cDNA. However, only canonically spliced RNA reads are aligned by standard alignment tools, while circRNA reads remain aside due to their out-of-order exon arrangement. To identify and characterise circRNAs, these separate reads are aligned by circRNA detection tools (Fig. 5.1).

Detection tools can be broadly classified as candidate-based (also known as pseudo-reference-based) or segmented-read-based (also known as fragmented-

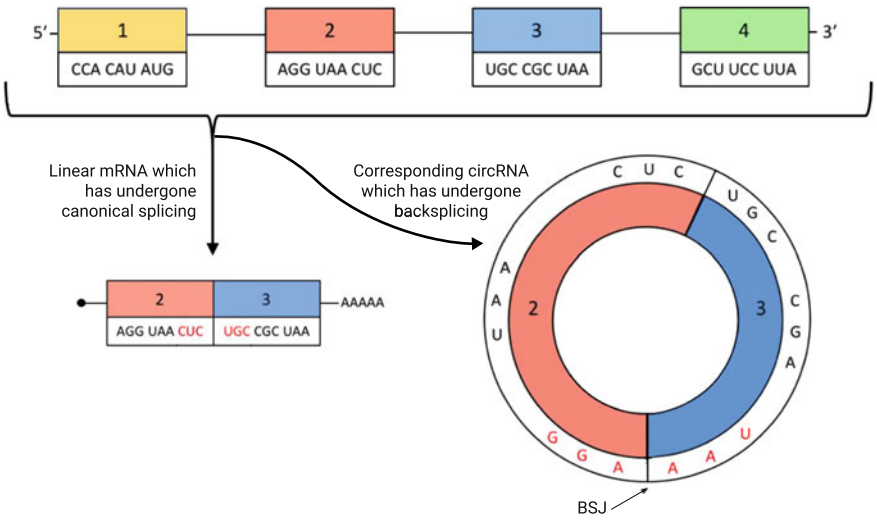

Fig. 5.1 Schematic illustration of the generation of a linear mRNA (left) via canonical splicing and a circular RNA (right) via backsplicing in which the splice donor of exon 3 splices to the acceptor of exon 2, forming a circlular RNA molecule without 5′–3′polarity and a poly(A) tail, and forms back-spliced junction (BSJ)

based) (Chen et al. 2015). For candidate-based tools such as NCLScan and PTESFinder, reference genome annotation is required as exon boundaries must be known. Annotation allows circRNA reads to be compared to different combinations of out-of-order exons derived from a reference genome. A downside of candidate-based tools is that they cannot identify circRNAs transcribed from unannotated genomic origins (KNIFE is an exception). Segmented-read-based tools do not require gene annotation to identify circRNAs.

For detection tools such as find_circ, PTESFinder and UROBORUS, whole circRNA reads are not aligned to the reference genomes. Instead, only short nucleotide sequences from each read are mapped (known as anchors). To reduce the number of false-positive circRNAs, detection tools also possess read filtering strategies which can be modified by the user. Filtering can include selecting for flanking GU/AG motifs, setting a minimum number of supporting reads, removing reads mapping to the mitochondrial genome and setting a minimum circRNA size. Detection tools can be also classified on read compatibility. Some detection tools are restricted to paired-end (PE) data (reads are generated from sequencing cDNA molecules in both directions) or single-end (SE) data (reads are generated from sequencing cDNA molecules in one direction), but most are compatible with both. For SE data, if only one read is generated from a circRNA, then that read must cross the backsplice junction for circRNA identification. As a result, the read corresponds to two distinct genomic origins. However, if two or more SE reads are generated from the same circRNA, then those reads can come from any sequence within the

circRNA. Similarly, for PE data, each read can come from any sequence within the circRNA.

A major challenge of circRNA detection is that there is no gold-standard detection tool. Different detection tools implement different detection strategies, each with their own bias, strengths and weaknesses. For example, KNIFE can identify circRNAs at both annotated and un-annotated exon boundaries, but it is restricted in detecting circRNAs in regions of genomic variation. For each experiment, the selected detection tool should be based on research aims and available computational capacity. Computational capacity includes random access memory, physical disk space and processor speed. Research aims should encompass the trade-offs between each detection tool's accuracy and sensitivity. As detection accuracy increases, sensitivity tends to decrease. Even for highly accurate detection tools such as MapSplice, a significant proportion (11%) of circRNAs are likely false positives. Accuracy is further decreased for circRNAs detected by a single detection tool. These circRNAs are known as exotic circRNAs and are detected with a false-positive rate as low as 20% (by MapSplice) and as high as 79% (by find_circ). Nonetheless, updated detection tools such as CIRI2 have improved overall circRNA detection (Hansen 2018). If further accuracy is required, a multi-bioinformatic approach can be implemented—circRNAs common between numerous detection tools are likely to be *bona fide* (Hansen et al. 2016).

The relatively low abundance of circRNAs in total RNA reduces the quality of sequencing data. To overcome this challenge, library preparation must include circRNA enrichment through ribosomal RNA (rRNA) depletion. Poly(A) selection is not suitable for circRNAs due to the absence of a poly(A) tail. Still, a few circRNAs will be present in poly(A) selection because the process is not completely efficient and because some circRNAs contain A-rich sequences. Further enrichment can be achieved with RNase R, although its use is debatable (Jeck et al. 2013). As a 3′ to 5′ exonuclease, RNase R degrades linear RNA molecules, but it may also degrade particular circRNAs (Jeck et al. 2013). Suspected circRNAs can be validated with probes for Northern blotting and fluorescent in situ hybridisation, as well as inverse or outward-facing primers for reverse transcriptase quantitative PCR (qPCR) (Capel et al. 1993; Barrett and Salzman 2016).

5.6 Mechanisms of CircRNA Action

The most established circRNA function to date is as a microRNA (miRNA) sponge. Many circRNAs, including zinc finger (*ZNF*) gene family of circRNAs and mouse circular *Sry*, are enriched with target sites for a number of miRNA families (Guo et al. 2014; Hansen et al. 2013). MiRNAs are ~21-nucleotide-long non-coding RNA transcripts that function in RNA silencing and post-translational regulation by guiding the effector protein Argonaute (AGO) to coding mRNAs, consequently repressing the latter's protein production (Ambros 2004; Bartel 2009). As a well-studied case, a human circRNA transcribed antisense to the *CDR1* gene harbours >60 conserved binding sites for miR-7 (Memczak et al. 2013; Guo et al. 2014).

miR-7 targets many important players implicated in various pathways and diseases such as several oncogenic factors in cancer-associated signalling pathways, including EGFR, IRS-1 and IRS-2 (Kefas et al. 2008), and α-synuclein in Parkinson's disease (Junn et al. 2009). It has been proposed that as a miRNA sponge, the *CDR1* antisense transcript (*CDR1as*) binds to miR-7, sequestering it away from its targets, resulting in an upregulation of miR-7-targeted mRNA. Indeed, PAR-CLIP (photoactivatable-ribonucleoside-enhanced crosslinking and immunoprecipitation) experiments for human AGO showed that *CDR1as* is densely bound by the miRNA effector protein AGO in the cytoplasm. Further, both miR-7 loss-of-function study and ectopic delivery of *CDR1as* caused significant reduction in brain sizes in zebrafish and other animals (Memczak et al. 2013). These results strongly support the notion that *CDR1as* inhibits miR-7 via sponging. Another example from a recent study found that the circRNA *HIPK3* contains 18 potential binding sites that were observed to sponge to nine miRNAs. More specifically, circular *HIPK3* binds to miR-124 and inhibits its activity; silencing of circular *HIPK3* significantly inhibits cell growth in humans (Zheng et al. 2016).

Some other studies suggested that circRNAs can also directly bind to other RNA-binding proteins (RBP), and consequently sequester and transport target RBPs (Hentze and Preiss 2013; Wilusz and Sharp 2013). However, it is still unclear if miRNA sponges or protein decoys are common functions of circRNA. The fact that only few circRNAs share properties similar to well-defined miRNA sponges and that circRNAs exist in species that lack RNA interference (RNAi) pathways strongly suggests other mechanisms of circRNAs action that have yet to be discovered (Guo et al. 2014; van Rossum et al. 2016).

5.7 Current Understanding of CircRNAs as Biomarkers

Emerging evidence has implicated a role for circRNAs in disease states, including neurodegenerative diseases and cancers. The high stability and enrichment in exosomes mean that circRNAs have the potential to serve as biomarkers for disease detection, diagnosis and progress monitoring (Li et al. 2015a). In addition to circRNAs in MSA, colon and endometrial cancers discussed earlier, several other groups have investigated circRNA expression for use as biomarkers in various cancer types. For example, in colorectal and ovarian cancer, the circular-to-linear ratio of RNA transcripts is lower in tumour tissues and there is an inverse relationship between this ratio and tumour cell growth rate (Bachmayr-Heyda et al. 2015). Other studies have reported differentially expressed individual circRNAs in laryngeal squamous cell cancer and gastric cancer tissues (Li et al. 2015b; Xuan et al. 2016). Taken together, there is mounting evidence that circRNA represents a potential model for development of novel biomarkers.

To gain an in-depth and holistic understanding of cancer, metastatic phenotypes must be observed beyond what is morphologically available. Characterising RNA expression levels can allow interpretation of the consequences of the genetic and epigenetic changes driving metastasis, thus enabling the links between different

phenotypes and molecular characteristics to be identified. Compared to the healthy cells in an individual, those with tumourigenic transformation display a clear difference in transcriptome profiles. Here, we will be using hepatocellular carcinoma (HCC) as an example. HCC is an aggressive primary liver cancer (PLC) which responds poorly to treatment. The difficulties in treating HCC are reflected by a high mortality-to-incidence ratio and low survival rate, particularly for patients beyond early-stage diagnosis (White et al. 2016). As a largely preventable disease, recent years have seen considerable progress in eliminating HCC risk factors, including obesity, diet, diabetes, alcohol and tobacco use, in addition to hepatitis B and C (White et al. 2016), whereas the development of diagnostic tools allowing early identification of the cancer has not been a focus of attention. It has been recently suggested that a number of different non-coding RNA species including circular RNAs (circRNAs), microRNAs (miRNAs) (Thurnherr et al. 2016), P-element Induced WImpy testis (PIWI)–interacting RNAs (piRNAs) (Rizzo et al. 2016) and small interfering RNAs (siRNAs) (Farra et al. 2015) are differentially expressed in HCC and are involved in the molecular pathology. Given the high stability of circRNAs, this differential expression in tumourigenic cells may be detectable in tissues such as saliva, blood and exosomes as a result of shedding of the tumour cells (Meng et al. 2017). Unlike linear RNAs, which are degraded by exonucleases and virtually absent in peripheral whole blood, circRNAs are enriched to levels comparable to that in the brain, >15-fold higher expression than non-neuronal organs like the liver (Memczak et al. 2015). This makes circRNA biomarkers of HCC easier to detect in whole blood, where peripheral cellular expression levels are minimal (Memczak et al. 2015).

This theory has come to fruition in recent years, with potential circRNA biomarkers of HCC having been discovered based on their abnormal expression levels in tumours. Yu et al. demonstrated that cSMARCA5 (hsa_circ_0001445), a circRNA transcribed from the SMARCA5 gene, is significantly downregulated in HCC resulting in a worsened prognosis, whereas mRNA and protein levels of SMARCA5 are upregulated (Yu et al. 2018). Han et al. demonstrated a similar pattern of findings based on the downregulation of circMTO1 in HCC tissue resulting in a worsened prognosis (hsa_circRNA_0007874/hsa_circRNA_104135) and suggested the possibility of circMTO1 upregulation resulting in the inflation of mRNA and protein expression levels of tumour suppressor gene, p21 (Han et al. 2017). For use as biomarkers, however, circRNA upregulation is considered to be more useful than downregulation, since expression is easier to quantify with detection tools that lack sensitivity and it can be more difficult to identify, interpret and distinguish the clinical significance of low-expressed circRNAs. Upregulation of circRNAs has been observed in individuals with HCC and appears more frequently than downregulation. Huang et al. demonstrated the upregulation of circRNA_10038 (hsa_circRNA_100338) and circRNA_104075 (hsa_circRNA_104075), both of which were positively correlated with metastatic properties of HCC cell lines. Further study by Huang et al. looked at circRNA_10038 and the association with a specific miRNA, miR-141-3p, and found cancer-inhibiting antagonistic properties associated with miR-141-3p directed

at circRNA_10038 that were positively correlated with the regulation of liver cancer cell metastasis (Huang et al. 2017).

CircRNAs are not by-products of alternative splicing. Hansen et al. demonstrated that circRNAs can act as miRNA sponges; circRNAs bind miRNA, preventing their repression of mRNA and resulting in further control of gene expression, as determined by their results with circRNA ciRS-7 acting as a sponge for miRNA miR-138 (Hansen et al. 2013).

Beyond interacting with miRNA, circRNAs have also been shown to interact with proteins and spliceosomal machinery, further influencing expression levels and cell proliferation (Ashwal-Fluss et al. 2014; Yu et al. 2018; Han et al. 2017; Du et al. 2016; Dudekula et al. 2016; Yu et al. 2016). The role of circRNAs as miRNA sponges, and the accumulating evidence to suggest circRNAs are efficient biomarkers in the disease profile for HCC is increasingly propagating the notion behind circRNA involvement in HCC disease pathology, and ultimately as significant molecular components of the disease profile.

Additionally, as more research is undertaken, it is likely that additional functions of circRNAs will be discovered. Where the relationship between circRNAs and a disease is due to a direct role in driving the pathogenic process, the circRNAs are a distinct feature of disease, meaning their use as a biomarker does not rely on a correlation between factors. In addition to their use as an early diagnostic tool, such a relationship allows the potential use of circRNAs as biomarkers for further clinical parameters, such as predicting prognosis and assessing the effectiveness of therapeutic agents.

5.8 Future Requirements

CircRNAs show tremendous potential as clinical biomarkers of disease diagnosis, prognosis and treatment assessment. Encouraging circRNA biomarker usage is the development of new circRNA detection tools to identify previously undiscovered circRNA types. One promising detection tool is accurate circRNA_finder suite (acfs), the first circRNA detection tool to identify fusion circRNAs (You and Conrad 2016). To further progress the use of circRNA biomarkers, the establishment of reputable pipelines with high sensitivity and precision is essential. For such pipelines, performance standards should be comparable to current pipelines available in linear RNA-seq analysis. In the meantime, multi-bioinformatic approaches should be adopted. Also, spurring the use of circRNA biomarkers is the introduction of third-generation sequencing, which enables long-read sequencing and direct RNA-seq. Implementing third-generation sequencing removes artefacts and bias which would arise from cDNA conversion and simplifies computational analysis by generating full-length reads. What's more is that direct RNA-seq technologies provides real time, massive parallel sequencing, with portability and affordability. Taken together with data produced to date, it is highly likely that the discovery of more disease-driving circRNAs will occur in the future, which opens the door to

their routine clinical use as early detection biomarkers, prognostic indicators and therapeutic efficacy indicators.

References

Ambros V (2004) The functions of animal microRNAs. Nature 431(7006):350–355. https://doi.org/10.1038/nature02871

Ashwal-Fluss R, Meyer M, Pamudurti NR, Ivanov A, Bartok O, Hanan M, Evantal N, Memczak S, Rajewsky N, Kadener S (2014) circRNA biogenesis competes with pre-mRNA splicing. Mol Cell 56(1):55–66. https://doi.org/10.1016/j.molcel.2014.08.019

Bachmayr-Heyda A, Reiner AT, Auer K, Sukhbaatar N, Aust S, Bachleitner-Hofmann T, Mesteri I, Grunt TW, Zeillinger R, Pils D (2015) Correlation of circular RNA abundance with proliferation—exemplified with colorectal and ovarian cancer, idiopathic lung fibrosis, and normal human tissues. Sci Rep 5:8057. https://doi.org/10.1038/srep08057

Barrett SP, Salzman J (2016) Circular RNAs: analysis, expression and potential functions. Development 143(11):1838–1847. https://doi.org/10.1242/dev.128074

Bartel DP (2009) MicroRNAs: target recognition and regulatory functions. Cell 136(2):215–233. https://doi.org/10.1016/j.cell.2009.01.002

Capel B, Swain A, Nicolis S, Hacker A, Walter M, Koopman P, Goodfellow P, Lovell-Badge R (1993) Circular transcripts of the testis-determining gene Sry in adult mouse testis. Cell 73 (5):1019–1030

Chen I, Chen CY, Chuang TJ (2015) Biogenesis, identification, and function of exonic circular RNAs. Wiley Interdiscip Rev RNA 6(5):563–579. https://doi.org/10.1002/wrna.1294

Chen BJ, Mills JD, Takenaka K, Bliim N, Halliday GM, Janitz M (2016) Characterization of circular RNAs landscape in multiple system atrophy brain. J Neurochem 139(3):485–496. https://doi.org/10.1111/jnc.13752

Chen BJ, Byrne FL, Takenaka K, Modesitt SC, Olzomer EM, Mills JD, Farrell R, Hoehn KL, Janitz M (2018) Analysis of the circular RNA transcriptome in endometrial cancer. Oncotarget 9 (5):5786–5796. https://doi.org/10.18632/oncotarget.23534

Cock PJ, Fields CJ, Goto N, Heuer ML, Rice PM (2010) The Sanger FASTQ file format for sequences with quality scores, and the Solexa/Illumina FASTQ variants. Nucleic Acids Res 38 (6):1767–1771. https://doi.org/10.1093/nar/gkp1137

Costa V, Angelini C, De Feis I, Ciccodicola A (2010) Uncovering the complexity of transcriptomes with RNA-Seq. J Biomed Biotechnol 2010:853916. https://doi.org/10.1155/2010/853916

Du WW, Yang W, Liu E, Yang Z, Dhaliwal P, Yang BB (2016) Foxo3 circular RNA retards cell cycle progression via forming ternary complexes with p21 and CDK2. Nucleic Acids Res 44 (6):2846–2858. https://doi.org/10.1093/nar/gkw027

Dudekula DB, Panda AC, Grammatikakis I, De S, Abdelmohsen K, Gorospe M (2016) CircInteractome: a web tool for exploring circular RNAs and their interacting proteins and microRNAs. RNA Biol 13(1):34–42. https://doi.org/10.1080/15476286.2015.1128065

Endrullat C, Glokler J, Franke P, Frohme M (2016) Standardization and quality management in next-generation sequencing. Appl Transl Genom 10:2–9. https://doi.org/10.1016/j.atg.2016.06.001

Enuka Y, Lauriola M, Feldman ME, Sas-Chen A, Ulitsky I, Yarden Y (2016) Circular RNAs are long-lived and display only minimal early alterations in response to a growth factor. Nucleic Acids Res 44(3):1370–1383. https://doi.org/10.1093/nar/gkv1367

Farra R, Grassi M, Grassi G, Dapas B (2015) Therapeutic potential of small interfering RNAs/micro interfering RNA in hepatocellular carcinoma. World J Gastroenterol 21(30):8994–9001. https://doi.org/10.3748/wjg.v21.i30.8994

Gao Y, Wang J, Zhao F (2015) CIRI: an efficient and unbiased algorithm for de novo circular RNA identification. Genome Biol 16:4. https://doi.org/10.1186/s13059-014-0571-3

Guo JU, Agarwal V, Guo H, Bartel DP (2014) Expanded identification and characterization of mammalian circular RNAs. Genome Biol 15(7):409. https://doi.org/10.1186/s13059-014-0409-z

Han D, Li J, Wang H, Su X, Hou J, Gu Y, Qian C, Lin Y, Liu X, Huang M, Li N, Zhou W, Yu Y, Cao X (2017) Circular RNA circMTO1 acts as the sponge of microRNA-9 to suppress hepatocellular carcinoma progression. Hepatology 66(4):1151–1164. https://doi.org/10.1002/hep.29270

Hansen TB (2018) Improved circRNA Identification by combining prediction algorithms. Front Cell Dev Biol 6:20. https://doi.org/10.3389/fcell.2018.00020

Hansen TB, Jensen TI, Clausen BH, Bramsen JB, Finsen B, Damgaard CK, Kjems J (2013) Natural RNA circles function as efficient microRNA sponges. Nature 495(7441):384–388. https://doi.org/10.1038/nature11993

Hansen TB, Veno MT, Damgaard CK, Kjems J (2016) Comparison of circular RNA prediction tools. Nucleic Acids Res 44(6):e58. https://doi.org/10.1093/nar/gkv1458

Hentze MW, Preiss T (2013) Circular RNAs: splicing's enigma variations. EMBO J 32 (7):923–925. https://doi.org/10.1038/emboj.2013.53

Huang XY, Huang ZL, Xu YH, Zheng Q, Chen Z, Song W, Zhou J, Tang ZY, Huang XY (2017) Comprehensive circular RNA profiling reveals the regulatory role of the circRNA-100338/miR-141-3p pathway in hepatitis B-related hepatocellular carcinoma. Sci Rep 7(1):5428. https://doi.org/10.1038/s41598-017-05432-8

Jeck WR, Sharpless NE (2014) Detecting and characterizing circular RNAs. Nat Biotechnol 32 (5):453–461. https://doi.org/10.1038/nbt.2890

Jeck WR, Sorrentino JA, Wang K, Slevin MK, Burd CE, Liu J, Marzluff WF, Sharpless NE (2013) Circular RNAs are abundant, conserved, and associated with ALU repeats. RNA 19 (2):141–157. https://doi.org/10.1261/rna.035667.112

Junn E, Lee KW, Jeong BS, Chan TW, Im JY, Mouradian MM (2009) Repression of alpha-synuclein expression and toxicity by microRNA-7. Proc Natl Acad Sci U S A 106 (31):13052–13057. https://doi.org/10.1073/pnas.0906277106

Kefas B, Godlewski J, Comeau L, Li Y, Abounader R, Hawkinson M, Lee J, Fine H, Chiocca EA, Lawler S, Purow B (2008) microRNA-7 inhibits the epidermal growth factor receptor and the Akt pathway and is down-regulated in glioblastoma. Cancer Res 68(10):3566–3572. https://doi.org/10.1158/0008-5472.CAN-07-6639

Knierim E, Lucke B, Schwarz JM, Schuelke M, Seelow D (2011) Systematic comparison of three methods for fragmentation of long-range PCR products for next generation sequencing. PLoS One 6(11):e28240. https://doi.org/10.1371/journal.pone.0028240

Lasda E, Parker R (2014) Circular RNAs: diversity of form and function. RNA 20(12):1829–1842. https://doi.org/10.1261/rna.047126.114

Levin JZ, Yassour M, Adiconis X, Nusbaum C, Thompson DA, Friedman N, Gnirke A, Regev A (2010) Comprehensive comparative analysis of strand-specific RNA sequencing methods. Nat Methods 7(9):709–715. https://doi.org/10.1038/nmeth.1491

Li Y, Zheng Q, Bao C, Li S, Guo W, Zhao J, Chen D, Gu J, He X, Huang S (2015a) Circular RNA is enriched and stable in exosomes: a promising biomarker for cancer diagnosis. Cell Res 25 (8):981–984. https://doi.org/10.1038/cr.2015.82

Li PF, Chen SC, Chen HL, Mo XY, Li TW, Shao YF, Xiao BX, Guo JM (2015b) Using circular RNA as a novel type of biomarker in the screening of gastric cancer. Clinica Chimica Acta 444:132–136. https://doi.org/10.1016/j.cca.2015.02.018

Memczak S, Jens M, Elefsinioti A, Torti F, Krueger J, Rybak A, Maier L, Mackowiak SD, Gregersen LH, Munschauer M, Loewer A, Ziebold U, Landthaler M, Kocks C, le Noble F, Rajewsky N (2013) Circular RNAs are a large class of animal RNAs with regulatory potency. Nature 495(7441):333–338. https://doi.org/10.1038/nature11928

Memczak S, Papavasileiou P, Peters O, Rajewsky N (2015) Identification and characterization of circular RNAs as a new class of putative biomarkers in human blood. Plos One 10(10): e0141214. https://doi.org/10.1371/journal.pone.0141214

Meng S, Zhou H, Feng Z, Xu Z, Tang Y, Li P, Wu M (2017) CircRNA: functions and properties of a novel potential biomarker for cancer. Mol Cancer 16(1):94. https://doi.org/10.1186/s12943-017-0663-2

Mortazavi A, Williams BA, McCue K, Schaeffer L, Wold B (2008) Mapping and quantifying mammalian transcriptomes by RNA-Seq. Nat Methods 5(7):621–628. https://doi.org/10.1038/nmeth.1226

Nigro JM, Cho KR, Fearon ER, Kern SE, Ruppert JM, Oliner JD, Kinzler KW, Vogelstein B (1991) Scrambled exons. Cell 64(3):607–613. https://doi.org/10.1016/0092-8674(91)90244-s

Oshlack A, Wakefield MJ (2009) Transcript length bias in RNA-seq data confounds systems biology. Biol Direct 4(1):14. https://doi.org/10.1186/1745-6150-4-14

Palazzo AF, Lee ES (2015) Non-coding RNA: what is functional and what is junk? Front Genet 6:2. https://doi.org/10.3389/fgene.2015.00002

Rizzo F, Rinaldi A, Marchese G, Coviello E, Sellitto A, Cordella A, Giurato G, Nassa G, Ravo M, Tarallo R, Milanesi L, Destro A, Torzilli G, Roncalli M, Di Tommaso L, Weisz A (2016) Specific patterns of PIWI-interacting small noncoding RNA expression in dysplastic liver nodules and hepatocellular carcinoma. Oncotarget 7(34):54650–54661. https://doi.org/10.18632/oncotarget.10567

Rybak-Wolf A, Stottmeister C, Glazar P, Jens M, Pino N, Giusti S, Hanan M, Behm M, Bartok O, Ashwal-Fluss R, Herzog M, Schreyer L, Papavasileiou P, Ivanov A, Ohman M, Refojo D, Kadener S, Rajewsky N (2015) Circular RNAs in the mammalian brain are highly abundant, conserved, and dynamically expressed. Mol Cell 58(5):870–885. https://doi.org/10.1016/j.molcel.2015.03.027

Salzman J, Gawad C, Wang PL, Lacayo N, Brown PO (2012) Circular RNAs are the predominant transcript isoform from hundreds of human genes in diverse cell types. PLoS One 7(2):e30733. https://doi.org/10.1371/journal.pone.0030733

Salzman J, Chen RE, Olsen MN, Wang PL, Brown PO (2013) Cell-type specific features of circular RNA expression. PLoS Genet 9(9):e1003777. https://doi.org/10.1371/journal.pgen.1003777

Starke S, Jost I, Rossbach O, Schneider T, Schreiner S, Hung LH, Bindereif A (2015) Exon circularization requires canonical splice signals. Cell Rep 10(1):103–111. https://doi.org/10.1016/j.celrep.2014.12.002

Szabo L, Salzman J (2016) Detecting circular RNAs: bioinformatic and experimental challenges. Nat Rev Genet 17(11):679–692. https://doi.org/10.1038/nrg.2016.114

Thurnherr T, Mah WC, Lei Z, Jin Y, Rozen SG, Lee CG (2016) Differentially expressed miRNAs in hepatocellular carcinoma target genes in the genetic information processing and metabolism pathways. Sci Rep 6:20065. https://doi.org/10.1038/srep20065

van Rossum D, Verheijen BM, Pasterkamp RJ (2016) Circular RNAs: novel regulators of neuronal development. Front Mol Neurosci 9:74. https://doi.org/10.3389/fnmol.2016.00074

Veno MT, Hansen TB, Veno ST, Clausen BH, Grebing M, Finsen B, Holm IE, Kjems J (2015) Spatio-temporal regulation of circular RNA expression during porcine embryonic brain development. Genome Biol 16:245. https://doi.org/10.1186/s13059-015-0801-3

White DL, Kanwal F, Jiao L, El-Serag HB (2016) Epidemiology of hepatocellular carcinoma. In: Carr BI (ed) Hepatocellular carcinoma: diagnosis and treatment. Springer, Cham, pp 3–24. https://doi.org/10.1007/978-3-319-34214-6_1

Wilusz JE, Sharp PA (2013) Molecular biology. A circuitous route to noncoding RNA. Science 340 (6131):440–441. https://doi.org/10.1126/science.1238522

Xuan LJ, Qu LM, Zhou H, Wang P, Yu HY, Wu TY, Wang X, Li QY, Tian LL, Liu M, Sun YN (2016) Circular RNA: a novel biomarker for progressive laryngeal cancer. Am J Transl Res 8 (2):932–939

You X, Conrad TO (2016) Acfs: accurate circRNA identification and quantification from RNA-Seq data. Sci Rep 6:38820. https://doi.org/10.1038/srep38820

You X, Vlatkovic I, Babic A, Will T, Epstein I, Tushev G, Akbalik G, Wang M, Glock C, Quedenau C, Wang X, Hou J, Liu H, Sun W, Sambandan S, Chen T, Schuman EM, Chen W

(2015) Neural circular RNAs are derived from synaptic genes and regulated by development and plasticity. Nat Neurosci 18(4):603–610. https://doi.org/10.1038/nn.3975

Yu L, Gong X, Sun L, Zhou Q, Lu B, Zhu L (2016) The circular RNA Cdr1as act as an oncogene in hepatocellular carcinoma through targeting miR-7 expression. PLoS One 11(7):e0158347. https://doi.org/10.1371/journal.pone.0158347

Yu J, Xu QG, Wang ZG, Yang Y, Zhang L, Ma JZ, Sun SH, Yang F, Zhou WP (2018) Circular RNA cSMARCA5 inhibits growth and metastasis in hepatocellular carcinoma. J Hepatol 68 (6):1214–1227. https://doi.org/10.1016/j.jhep.2018.01.012

Zhang XO, Wang HB, Zhang Y, Lu X, Chen LL, Yang L (2014) Complementary sequence-mediated exon circularization. Cell 159(1):134–147. https://doi.org/10.1016/j.cell.2014.09.001

Zheng Q, Bao C, Guo W, Li S, Chen J, Chen B, Luo Y, Lyu D, Li Y, Shi G, Liang L, Gu J, He X, Huang S (2016) Circular RNA profiling reveals an abundant circHIPK3 that regulates cell growth by sponging multiple miRNAs. Nat Commun 7:11215. https://doi.org/10.1038/ncomms11215

The Role of Histone Variants in Cancer

6

Antonia L. Pritchard

Abstract

In eukaryotes, chromosomal DNA is packaged around histone proteins into tightly coiled structures, called nucleosomes, which then form further repeating condensed elements. This fundamental repeating unit, chromatin, ensures that the large eukaryotic genome can pack into the nucleus while still allowing gene expression. Specific mechanisms have evolved to regulate access to DNA by the proteins that control gene expression, such as transcription factors and polymerases, and to control the looping structure of the chromatin to allow interaction between distant DNA locations. One of the major mechanisms by which this is achieved is by insertion of variant histones, which provide different degrees of open chromatin structure compared to canonical proteins. A further mechanism that can alter control of gene expression is the post-translational modification of histone proteins, which not only physically alter chromatin structure but also recruit remodelling enzymes that reposition nucleosomes. This chapter describes the various histone variants and modifications that control chromatin structure and highlight how these modifications can alter transcription. These histone variants can be altered in disease states; alterations that occur in cancer will then be briefly outlined.

Keywords

Histone variants · Nucleosome · Chromatin · Histones · Post-translational modification · Cancer

A. L. Pritchard (✉)
An Lochran, University of the Highlands and Islands, Inverness, UK
e-mail: Antonia.Pritchard@uhi.ac.uk

L. B. Hesson, A. L. Pritchard (eds.), *Clinical Epigenetics*,
https://doi.org/10.1007/978-981-13-8958-0_6

6.1 Nucleosome Organisation

Since the landmark X-ray diffraction study of Finch and Klug (Luger et al. 1997; Finch et al. 1981) showed that chromatin was organised into 30nm-wide solenoidal coils, the nature of the structure of chromosomes has been a source of intense research (Bian and Belmont 2012). Nucleosomes comprise approximately 147 bp of DNA wrapped around a histone octamer. This complex consists of two copies of each of the canonical histones (H2A, H2B, H3 and H4), which are 100–140 amino acids long, consisting of a histone-fold domain (HFD) (Arents and Moudrianakis 1995), flanked by N- and C- terminal regions that can be chemically modified (Fig. 6.1). The nucleosome complex is assembled first into tetramers of H3 and H4 due to a strong four-helix bundle interaction between the two H3 proteins (Luger et al. 1997); this then complexes with two H2A/H2B dimers, which dock at the DNA entry and exit sites through the H2A C-terminus docking domain. The nucleosomes are linked to each other via variable linker DNA regions, which range from 10 to100 bp in length (Noll 1977) and interact with the 200–400 amino acid long H1 linker histone proteins. H1 histones contact the DNA exit/entry into the nucleosome and are able to control the modulation of chromatin fibre folding and nucleosome complex binding to DNA regions (Fig. 6.1).

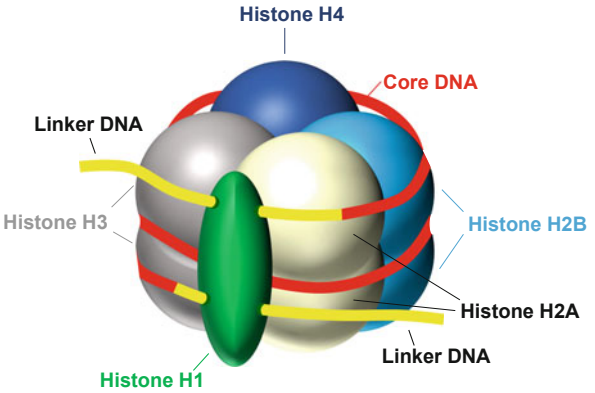

Fig. 6.1 Nucleosome organisation. Two copies of each canonical histone (H2A, H2B, H3 and H4) form the nucleosome complex. It is first assembled into tetramers of H3 and H4 then complexes with two H2A/H2B dimers. Each nucleosome is linked via variable linker DNA regions, which also interact with the H1 linker histone proteins. H1 histones can therefore control modulation of chromatin fibre folding

6.2 Canonical and Variant Histones

While the canonical histones are the most abundant histones in any genome, variants of these histones also exist, which have been driven by evolution and can confer locus-specific properties. In humans, variants of the canonical histones H2A, H2B and H3 have been identified. While no variants of H4 have been found in higher eukaryotes (Talbert et al. 2012), they been identified in lower eukaryotes [such as trypanosomes (Siegel et al. 2009) and some urochordates (Moosmann et al. 2011)].

When DNA is replicating during the S phase of the cell cycle, chromatin organisation needs to be maintained, meaning there is a high demand for histone synthesis during this phase. In order to facilitate this, the genes encoding the canonical histones consist of a single exon and use a stem-loop structure at the 3′ end of the mRNA to allow rapid transcription and translation. In contrast, histone variants are produced in smaller quantities from distinct genes, which are translated in a controlled, cell-specific manner in order to replace canonical histones during chromatin metabolism. Given this requirement, the genes tend to have introns and produce polyadenylated mRNA, allowing histone variants of each subtype to be individually controlled in terms of transcription, RNA processing (including alternative splicing) and timing of deposition upon DNA during the cell cycle.

6.2.1 Histone Chaperones

The control of spatial and temporal histone deposition/eviction is carried out by chaperone and ATP-dependent remodelling complexes. Histone chaperones have dual functions, such that they bind histones immediately after synthesis in order to control stability or degradation (Cook et al. 2011) and also actively participate in nucleosome formation (Burgess and Zhang 2013). Additionally, chaperones have individual roles, including (a) aiding the trafficking of histones from the cytoplasm to the nucleus, via regulating interaction with importins (Campos et al. 2010), (b) facilitating the promotion of nucleosome formation (Campos et al. 2010; Elsasser 2013) and (c) controlling the post-translational modifications (PTMs) of histones, through binding histones to the responsible enzyme (Cakmakci et al. 2008). The ATP-dependent chromatin remodelling complexes organise the structure of chromatin by sliding or ejecting assembled nucleosomes, allowing exchange of canonical dimers with dimers of variant histone (Narlikar et al. 2013).

6.2.2 Nomenclature of Histone Variants

Historically, the names assigned to homologous histone proteins in different species has resulted in confusion over the nomenclature of histone variants, including inconsistent naming styles and multiple synonyms; however, in 2012, a unified method of naming these proteins was proposed (Talbert et al. 2012). This new nomenclature builds on the historical naming and incorporates phylogenetic

relationships, providing an indication of structure and function. This nomenclature labels the variants with organism-specific paralogue number suffixes, with the letter suffixes indicating the structurally distinct clades of variants and was designed to make the relationship between variant subtypes explicit. The following section details the human variant histones and their known functions, which are summarised in Table 6.1.

6.2.3 Histone Variants of H2A

In humans, eight variants of histone H2A exist: H2A.X, H2A.Z.1, H2A.Z.2.1, H2A.Z.2.2, H2A Barr body deficient (H2A.Bbd; also known as H2A.B), macroH2A1.1, macroH2A1.2 and macroH2A2. These variant histones differ from canonical H2A around the C-terminus, which is the site of DNA entry/exit. These alterations influence nucleosome stability and dynamics, resulting in functional diversification of the nucleosomes in which they are incorporated, including changes in binding to DNA and/or the H1 linker histone. Additionally, the L1 region in the histone fold, where the dimerisation interaction between the two H2A histone variants takes place, can be altered in the variant histone, further increasing the differences in functionality compared to the canonical histone.

6.2.3.1 Function of H2A Histone Variants

H2A.X

H2A.X is the most commonly incorporated H2A variant and is the best characterised in terms of function. This variant differs from the canonical H2A by a functionally important C-terminal phosphorylation motif 'SQ(E/D)Φ' (where Φ represents a hydrophobic residue). The phosphorylation of H2A.X is important in DNA damage response, chromatin remodelling and X-inactivation, regulating DNA repair, suppression of anti-sense RNA and the recruitment of RNA polymerase II. In particular, the phosphorylation of serine at position 139, located near the DNA entry/exit point on the nucleosome, produces the ϒH2AX variant and controls many of these functions. Additionally, there is a one amino acid insertion in L1-loop in H2A.X compared to H2A, which is the region controlling the H2A-H2B dimer interaction, and a one amino acid deletion in the docking domain relative to canonical H2A that is also important in these functions.

H2A.Z.1, H2A.Z.2.1 and H2A.Z.2.2

H2A.Z variant histone includes the substitution of an amino acid resulting in a negatively charged sequence motif 'DEELD' (as opposed to DEELN), which causes the nucleosome to become more compact. Histone variants H2A.Z.1 and H2A.Z.2 are encoded by two genes, *H2AFZ* and *H2AFV*, respectively, and the H2A.Z.2 variants H2A.Z.2.1 and H2A.Z.2.2 are generated by alternative splicing (Bonisch et al. 2012). H2A.Z.1, H2A.Z.2.1 and H2A.Z.2.2 differ by three crucial amino acids, p.T15A, p.S38T and p.V128A (Eirin-Lopez et al. 2009). The p.S38T substitution is

Table 6.1 Summary of the human histone variants, chaperone and remodellers and known functions

Variant histone	Chaperone/ remodeller	Function
H2A variants		
H2A.X	FACT (chaperone)	• DNA repair • Suppression of anti-sense RNA • Recruitment of RNA polymerase II • Chromatin remodelling
H2A.Z.1	p400 (remodeller) SRCAP (remodeller) INO80 (remodeller) ANP32E (chaperone)	Promotion of gene expression, including Pol II recruitment, transcription regulation, DNA repair and suppression of antisense RNA
H2A.Z.2.1	p400 (remodeller) SRCAP (remodeller)	Promotion of gene expression, including Pol II recruitment, transcription regulation, DNA repair and suppression of antisense RNA
H2A.Z.2.2	p400 (remodeller) SRCAP (remodeller) TIP60 complex (chaperone) p400 NuA4 TIP60 (KAT5)	Incorporation leads to unstable nucleosomes Promotion of gene expression, including Pol II recruitment, transcription regulation, DNA repair and suppression of antisense RNA
H2A.Bbd	NAP1 (chaperone)	Promotion of gene expression
macroH2A1.1	HIRA Complex (chaperone): HIRA CABIN1 UBN1 Asf1	• Repression of gene expression • X-inactivation • Promotion of gene expression
macroH2A1.2	ATRX (remodeller) HIRA complex (chaperone): HIRA CABIN1 UBN1 Asf1 (chaperone)	• Repression of gene expression • X-inactivation • Promotion of gene expression

(continued)

Table 6.1 (continued)

Variant histone	Chaperone/ remodeller	Function
macroH2A2	HIRA complex (chaperone): HIRA CABIN1 UBN1 Asf1 (chaperone)	Repression of gene expression
H2B variants		
H2BFWT	SWI-SNF (remodeller)	• Spermatogenesis • Telomere-associated function of sperm
TSH2B	Unknown	• Expressed in testes • Expressed in sperm cells • Chromatin- to-nucleoprotamine transition
H3 variants		
H3.1	CAF1 Complex (chaperone): p150 (CHAF1A) p60 (CHAF1B) RbAp48 (RBBP4)	• Deposited in a DNA synthesis-dependent manner • The CAF1 complex deposits during cell cycle
H3.2	CAF1 complex (chaperone): p150 (CHAF1A) p60 (CHAF1B) RbAp48 (RBBP4)	• Deposited in a DNA synthesis-dependent manner • The CAF1 complex deposits during cell cycle
CENPA	HJURP (chaperone) DAXX (chaperone)	• Highly enriched at centromeres • Epigenetic determinant of centromeres • DAXX deposits at non-centromeric sites in CENPA overexpressing cells
H3.3	HIRA complex (chaperone): HIRA CABIN1 UBN1 DAXX (chaperone) ATRX (chaperone)	• Deposited in a DNA synthesis independent manner by the HIRA complex • Deposited in a DNA synthesis independent manner at telomeres, rDNA and pericentric heterochromatin by DAXX and ATRX
H3.4	Unknown	• Replacement histone variant • Testis-specific • Incorporated genome-wide at late meiosis or during early spermiogenesis
H3.5	Unknown	• Replacement histone variant • Testis specific
H3.Y.1	Unknown	Replacement histone variant
H3.Y.2	Unknown	Replacement histone variant

within the histone fold at the end of the α1-helix, located before the L1-loop, and likely alters the mobility of the variants in cells compared to the canonical histone (Horikoshi et al. 2013).

H2A.Z.1 is enriched in nucleosomes at transcriptional start sites and is involved in the control of transcriptional regulation (Barski et al. 2007). Furthermore, when found in heterochromatin (condensed chromatin), it participates in the formation of pericentric and centric chromatin, while its presence in euchromatin (lightly packed chromatin) is usually associated with promoters of active genes.

The histone variants H2A.Z.2.1 and H2A.Z.2.2 originate from alternative splicing; compared to H2A.Z.2.1, H2A.Z.2.2 is 14 amino acids shorter and contains 6 alternative amino acids in its C-terminus. The shortened C-terminus of H2A.Z.2.2 means that when it is incorporated into chromatin, it results in severely destabilised nucleosomes due to a loosened binding to H3 (Bonisch et al. 2012). The shortened C-terminal sequence also results in loss of ubiquitination sites at positions p.K120 and p.K121 (Sarcinella et al. 2007) and part of the H3/H4 docking domain (Suto et al. 2000) in H2A.Z.2.2. The majority of H2A.Z.2.2 is unbound in the nucleus, with only a minority stably incorporated into chromatin (Bonisch et al. 2012).

H2A.Bbd

Histone variant H2A.Bbd is the most divergent H2A variant histone, with its sequence homology only being 38.1% compared to canonical H2A (Buschbeck and Hake 2017). H2A.Bbd lacks an acidic patch in the carboxy-terminus resulting in it only being able to organise 118–130 bp of DNA, which results in the formation of less compact chromatin, facilitating active DNA transcription (Bao et al. 2004; Doyen et al. 2006; Zhou et al. 2007). Specifically, H2A.Bbd is associated with the elongating form of RNA polymerase, components of the spliceosome and mRNA processing, providing a vital role in two critical aspects of the expression of genes (Tolstorukov et al. 2012).

macroH2A1.1, macroH2A1.2 and macroH2A2

The macroH2A variant histone contains an HFD and an extra, long C-terminal macro domain which can bind poly-ADP-ribose. The crystal structures of both domains have been solved, but the inter-domain linker is too flexible to be crystallized. Histone variants macroH2A1.1 and macroH2A1.2 are encoded by alternative splicing of the *H2AFY* gene, and macroH2A2 is encoded by the *H2AFY2* gene (Rasmussen et al. 1999). The macroH2A variants are associated with the inactivated X-chromosome in females, suggesting a possible role in gene silencing (Tanasijevic and Rasmussen 2011; Costanzi and Pehrson 1998; Zhang et al. 2005), and, indeed, these variant histones are depleted at the transcription start site of active genes (Buschbeck et al. 2009; Gamble et al. 2010). Additionally, macroH2A plays a role during early embryonic development and are detectable between the 8- and 16-cell stages; however, whether there are differences in the expression of the macroH2A isoforms in humans remains to be determined, but differential expression is observed in mice and zebrafish (Buschbeck et al. 2009; Creppe et al. 2012). The chromatin remodeller ATRX negatively regulates the

amount of macroH2A present at genes, such as observed at the α-globin gene locus. Furthermore, recent evidence, currently in pre-print with bioRxiv (not peer reviewed; doi: 10.1101/333468), shows that macroH2A can function as a regulator of epigenetic cell memory, passed from parent to daughter cells.

6.2.4 Summary of Histone H2A Variants

The distribution of H2A variant histones across the genome is indicative of the specific, localised regulation of gene expression. Active chromatin is enriched with H2A.Z.1, H2A.Z.2.1 and H2A.Z.2.2 at the promoters and with H2A.Bbd in the gene bodies, both increasing gene expression. Genes with low/repressed expression do not contain these histone variants but are associated with macroH2A1.1, macroH2A1.2 and macroH2A2; indeed, H2A.Bbd and the macroH2A variants generally mutually exclusively incorporated along the genome, with some exceptions (Tolstorukov et al. 2012).

6.2.5 Histone Variants of H2B

In humans, only two variants of H2B have been identified to date, histone H2B type WT (H2BFWT; also known as H2B.W) and histone H2B1A (also known as TSH2B or TH2B).

6.2.5.1 Function of H2B Histone Variants

The above H2B variant histones are both testes specific. H2BFWT does not affect the overall structure or stability of the nucleosome when incorporated; however, the remodeller SWI/SNF specifically mobilises nucleosomes with H2BFWT incorporated. Additionally, this variant histone is unable to recruit chromosomal condensation factors nor take part in the assembly of mitotic chromosomes during the cell cycle, which has been shown to be due to the extended N-terminal domain it possesses compared to canonical H2B (Boulard et al. 2006). H2BFWT is involved with spermatogenesis telomere-associated function in sperm and is found in spermatogenic cells. The histone variant H2B1A is specifically distributed in the basal area of nuclei in sperm but were only found in a subpopulation of the cells (Zalensky et al. 2002).

6.2.6 Histone Variants of H3

In humans, six variants of canonical H3 exist: H3.3, histone H3-like centromeric protein A (known as CENPA), H3.4 (also known as H3.1T), H3.5, H3.Y.1 (also known as H3.Y) and H3.Y.2 (also known as H3.X). Except for CENPA, the H3 variants are highly conserved and differ by only a few amino acids. Broadly, the H3 variants fall into two subgroups, based on their function. The first are replicative

histones, which peak in expression during S phase of the cell cycle and are important histones for DNA replication. They are deposited in a DNA synthesis-coupled manner and in humans are exemplified by the canonical histones H3.1 and H3.2. These histones are produced in S-phase cells and incorporate into chromatin during DNA replication (Tagami et al. 2004). The second type are replacement histone variants, which are generally expressed independently of S phase of the cell cycle and are incorporated independently of DNA synthesis; in humans, these are exemplified by variant histones H3.3 (Ahmad and Henikoff 2002) and CENPA (Palmer et al. 1987).

6.2.6.1 Function of H3 Histone Variants

Replacement Histone Variants: H3.3, CENPA

Histone H3.3 is encoded by two genes, *H3F3A* and *H3F3B*, a genetic occurrence unique to histones in humans, where two conventional intron-containing genes encode an identical protein (Tang et al. 2015). H3.3 plays an important role in the maintenance of genome integrity during mammalian development and is particularly abundant during early cell division in embryonic development, where it is primarily involved in transcriptional regulation (Jang et al. 2015). The H3.3 variant histone is considered to be a 'replacement' histone, where a nucleosome is lost independent of replication and the canonical H3/H4 tetramer is replaced by the H3.3 variant and its H4 partner. The incorporation of the H3.3 histone variant is often opportunistic, occurring when DNA is exposed at dynamic regions such as gene promoters, the body of active genes, and at *cis*-regulatory elements (Mito et al. 2007; Ray-Gallet et al. 2011; Schneiderman et al. 2012). The function of H3.3 is controlled by the complex formed with the chaperone proteins. The HIRA chaperone complex promotes H3.3 deposition at transient nucleosome-free regions, whereas DAXX/ATRX are important for H3.3 enrichment at pericentric heterochromatin and telomeres (Table 6.1). One of the marks of an active promoter, H3 lysine 4 dimethylation (H3K4me2), is often accompanied by other features of accessible chromatin, including the H3.3 variant histone (Goldberg et al. 2010). Additionally, the function of H3.3 can be altered by the deposition complex used to chaperone the histone variant into position. When HIRA chaperones H3.3 to naked DNA (e.g. in regions of gene activation), it has been shown to protect chromatin integrity (Ray-Gallet et al. 2011), whereas H3.3 incorporated via DAXX-ATRX contributes to the establishment of heterochromatin (Goldberg et al. 2010; Drane et al. 2010). The variant histone H3.3 also contributes to the maintenance of telomeres in conjunction with ATRX (Wong et al. 2010) (Table 6.1), further demonstrating the diverse roles of its incorporation.

CENPA is also a 'replacement' histone variant, which is vital for chromosomal segregation. CENPA is controlled by the chaperon protein HJURP (*H*olliday *ju*nction *r*ecognition *p*rotein) to specifically target centromeric deposition during the cell cycle (Dunleavy et al. 2009; Foltz et al. 2009) (Table 6.1); CENPA therefore marks centromeres. CENPA acts to nucleate the kinetochore formation and is essential for proper chromosome segregation during mitosis (as reviewed in Verdaasdonk and Bloom 2011).

Function of H3.4

The histone variant H3.4 is encoded by *HIST3H3* in humans and is believed to be restricted in expression to the testes (Witt et al. 1996). H3.4 shares similar histone chaperone motifs with H3.1 and H3.2 (Table 6.1) and is incorporated into the genome in a replication-coupled manner (Maehara et al. 2015). H3.4 lacks the p.A111 residue, which is conserved between H3.1, H3.2 and H3.3 and is essential for NAP1-mediated nucleosome assembly, indicating this mechanism is not used for H3.4 incorporation (Tachiwana et al. 2008). It is the human orthologue of the mouse H3t gene, which has been more extensively studied. In murine models, H3t has been demonstrated to be essential for normal spermatogenesis. H3t replaces canonical H3 histones at the spermatogonia stage of spermatogenesis and, when knocked down, results in azoospermia due to loss of haploid germ cells and inability to enter into meiosis (Ueda et al. 2017).

Function of H3.5

H3.5 is conserved among great apes and Neanderthals, but not in non-hominid primates, and is highly expressed in the human testes (Schenk et al. 2011). The H3.5 nucleosome is less stable than the H3.3 nucleosome, potentially due to the H3.5-specific p.L103 residue, which corresponds to the H3.3 p.F104 residue, reducing the hydrophobic interaction with histone H4 compared to H3.3. Given the accumulation of H3.5 around the promoters of genes, it has been hypothesised that the unstable H3.5 nucleosome may function to alter the chromatin dynamics around transcription start sites during spermatogenesis (Urahama et al. 2016).

Function of H3.Y.1 and H3.Y.2

Histone H3.Y is a primate-specific histone variant evolutionarily derived from H3.3 expressed in several normal tissues, including testes and certain brain regions, such as the cerebral cortex, hippocampus and cerebellum (Wiedemann et al. 2010). H3.Y preferentially accumulates in euchromatin, and knockdown of H3.Y results in transcriptional suppression, suggesting a positive role in stimulation of gene expression (Wiedemann et al. 2010). Examination of the dynamics of H3.Y inclusion demonstrated that H3.Y-specific residues are often located on the nucleosome entry/exit sites. Together with biochemical evidence that the DNA ends of the H3.Y nucleosomes were more flexible than those of the H3.3 nucleosome, that the linker H1 histone binds to the H3.Y nucleosome less efficiently and that H3.Y accumulates around the transcription start site, it is suggested that H3.Y-containing nucleosomes form relaxed chromatin allowing access to transcription factors and gene expression (Kujirai et al. 2016).

6.3 Further Complexity of Variant Histone Chromatin Control

Histone variants can further control the structure of chromatin via different combinations of canonical and variant histones and then through the modification of protein structure via post-translational modifications (PTM).

6.3.1 Heterotypic Nucleosomes and Homotypic Variant Nucleosomes

Heterotypic nucleosomes contain either one canonical histone and a variant histone, or two different variant histones, instead of two identical histones; there can be combinations for each of the dimers that make up the core octamer (also known as nucleosome asymmetry). Each of these different combinations can further alter the features of the nucleosome in which they are incorporated compared to homotypic nucleosomes, adding significant complexity to the system. These combinations allow for variability in nucleosome formation, altering the stability and structure of chromatin, allowing for enhancer and active promoter regions to be more or less accessible by transcription factors and regulatory complexes (Jin et al. 2009). Examples include: (a) the combination of H2A.Z and H2A, which when resolved by crystal structure revealed that instead of instability being conferred by steric clashing of the L1 loop regions, the H2A L1 loop was not changed, but the H2A.Z L1 loop was drastically altered, which resulted in a more stable nucleosome than when the homotypic H2A.Z is incorporated (Horikoshi et al. 2016); (b) the heterotypic H3.3 and H3.Y nucleosome (Wiedemann et al. 2010) confers more flexibility and reduced H1 linker histone binding compared to the homotypic H3.3 nucleosomes, likely making chromatin more relaxed and allowing access to the transcriptional machinery (Kujirai et al. 2016).

6.3.2 Post-translational Modifications

Histone variants can be modified by different types of PTMs (McKittrick et al. 2004; Hake et al. 2006). The differences in sequences between the canonical and variant histones provide diverse opportunities for levels and types of the PTMs. These include phosphorylation, acetylation, ubiquitination and methylation (Singh et al. 2012; Draker et al. 2012; Wen et al. 2014; Campos and Reinberg 2009). These chemical changes to individual amino acids alter the ability of regulatory factors to bind DNA, leading to alterations in gene expression. PTMs can subtly or dramatically alter the properties of histones either individually or in combination, adding yet more complexity to the system of chromatin access/nucleosome packaging. PTMs are recognised by specific variant 'reader' molecules that recruit additional chromatin remodelling complexes to the site of histone deposition. For example, H2A.X C-terminal phosphorylation (termed ϒH2A.X) occurs as a result of DNA damage, which is specifically recognised by the early DNA damage repair protein, microcephalin (MCPH1), and acts as an early sensor of DNA damage (Singh et al. 2012).

6.4 Histone Variants in Disease

The potential for histones and variant histones to be altered in disease states is very broad, but one of the most intensively studied areas of research of abnormal inclusion or modification of histone variants has been in cancer.

6.4.1 Histone Variants in Cancer

There have been extensive and systematic reviews on the role of histone variants in the tumourigenic process. Recommended reviews include: Buschbeck and Hake (2017) and Vardabasso et al. (2014). Only some specific examples will be provided below, to illustrate how variant histones can be subverted in cancer.

6.4.1.1 H2A.Z.2 in Melanoma

Variant H2A.Z.2 has been linked with the progression of malignant melanoma. This has been shown to be via the promotion of cellular proliferation by the recruitment of the transcription regulators BRD2 (Gallagher et al. 2014) and E2F1 to E2F target genes, facilitating their transcription (Vardabasso et al. 2016). BRD2 binds to histone-acetylated lysine residues, acting as scaffolds to recruit chromatin-modifying factors, and is over-expressed in melanoma; when depleted by the epigenetic regulator I-BET151, an isoxazole class pan-BET family inhibitor, it causes G1/S cell cycle arrest in melanoma cells (Gallagher et al. 2014). Furthermore, high H2A.Z.2 expression correlates with poor survival in patients, and its depletion sensitises cells to chemotherapy and targeted therapies (Vardabasso et al. 2016).

6.4.1.2 CENPA in Cancer

Over-expression of CENPA has been observed in many cancer types, including colorectal cancer (Tomonaga et al. 2003), lung adenocarcinoma (Wu et al. 2012; Sun et al. 2016), oestrogen receptor-positive breast cancer (Sun et al. 2016; McGovern et al. 2012), epithelial ovarian cancer (Qiu et al. 2013) and 16 other types (Sun et al. 2016). Over-expression of CENPA has been associated with genome instability in a cancer genomic background (Amato et al. 2009), a condition that can initiate cancer and promote disease progression. Furthermore, the over-expression of CENPA can result in aneuploidy. This occurs due to incorrect incorporation of CENPA in a chromosomal location away from the centromere resulting in the formation of ectopic neocentromeres and kinetochores in chromosome arms, instead of at the centromere (Heun et al. 2006; Olszak et al. 2011). These ectopic structures disturb the normal segregation of sister chromatids during mitosis, leading to aneuploidy (Heun et al. 2006). Therefore, CENPA can be both a marker of cancerous cells as well as a lynchpin in the tumourigenic process.

6.4.1.3 Chaperone Proteins in Cancer

There have been a large number of reports of chaperone proteins aberrations in cancer compared to normal tissue. Some examples are included below; more

comprehensive reviews are available from Vardabasso et al. (2014) and Burgess and Zhang (2013).

The chaperone HJURP, which functions to control deposition of CENPA (Table 6.1), has been reported to be over-expressed in a number of cancer types, including lung (Kato et al. 2007) and breast cancer (Hu et al. 2010), glioma (de Tayrac et al. 2013) and astrocytoma (Valente et al. 2013); this has been shown to correlate with patient survival (Hu et al. 2010; Valente et al. 2013; Montes de Oca et al. 2015). As previously discussed, over-expression of CENPA itself is associated with genomic instability and aneuploidy due to incorrect positional insertion of this variant histone; therefore, it might be expected that concurrent or independent over-expression of its chaperone protein functions in this same manner.

ATRX is a chaperone protein for macroH2A1 (Ratnakumar et al. 2012) and H3.3 (Goldberg et al. 2010; Lewis et al. 2010). The interaction of H3.3 with ATRX is in conjunction with DAXX to deposit variant histones specifically at telomeres. MacroH2A1, however, only interacts with ATRX (and not DAXX), which has been shown to be due to the two histone variants interacting with unique ATRX complexes (Ratnakumar et al. 2012). Furthermore, the interaction between macroH2A1 with ATRX is a negative regulation, such that loss of ATRX results in increased macroH2A1. This is opposite to the effect of interaction between ATRX and DAXX with H3.3, which promotes deposition. There are a number of cancers with aberration of ATRX and DAXX, including paediatric glioblastoma (Schwartzentruber et al. 2012) and pancreatic neuroendocrine tumours (Heaphy et al. 2011; Jiao et al. 2011). Furthermore, other cancers have specific loss of function events (homozygous deletion, loss of heterozygosity, nonsense and frame-shift variants) in ATRX alone, including sarcoma, breast cancer, colorectal cancer, endometrial cancer, prostate cancer, glioblastoma, pancreatic neuroendocrine tumours and kidney cancer (TCGA, pan-cancer analysis, cBioportal). There are mutually exclusive observations of H3.3 (*H3F3A* and *H3F3B*) amplification and loss of function mutations in ATRX (TCGA, pan-cancer analysis, cBioportal). Loss of ATRX and DAXX is associated with activation of the alternative lengthening of telomeres (ALT) pathway of telomere regulation, which leads to longer telomeres; indeed loss of ATRX and DAXX has been used as a surrogate marker of ALT in cancers such as pancreatic endocrine tumours (Amorim et al. 2016).

6.5 Conclusion

Histones and their variants play a significant role in controlling chromatin complexity by creating specialised nucleosomes. Each type of histone has a different effect on chromatin structure, meaning that the various possible combinations of canonical and variant histones influence nucleosome stability, complexity and plasticity. Furthermore, the PTM of histones and variant histones provides additional complexity to the system, allowing subtle or more dramatic alterations to chromatin access or providing markers for binding of DNA altering complexes. Given the intricacies of this system, the potential for histones to be altered in disease, either as a cause of the

issue or as a consequence, is broad. It is highly likely that with advances in technology and ability to more rapidly and accurately assess each of the potential modifications to histones within nucleosomes, the nature of these disease-causing changes will continue to be elucidated and incorporated into clinical testing screens.

References

Ahmad K, Henikoff S (2002) The histone variant H3.3 marks active chromatin by replication-independent nucleosome assembly. Mol Cell 9(6):1191–1200

Amato A, Schillaci T, Lentini L, Di Leonardo A (2009) CENPA overexpression promotes genome instability in pRb-depleted human cells. Mol Cancer 8:119. https://doi.org/10.1186/1476-4598-8-119

Amorim JP, Santos G, Vinagre J, Soares P (2016) The role of ATRX in the alternative lengthening of telomeres (ALT) phenotype. Genes (Basel) 7(9):66. https://doi.org/10.3390/genes7090066

Arents G, Moudrianakis EN (1995) The histone fold: a ubiquitous architectural motif utilized in DNA compaction and protein dimerization. Proc Natl Acad Sci U S A 92(24):11170–11174

Bao Y, Konesky K, Park YJ, Rosu S, Dyer PN, Rangasamy D, Tremethick DJ, Laybourn PJ, Luger K (2004) Nucleosomes containing the histone variant H2A.Bbd organize only 118 base pairs of DNA. The EMBO J 23(16):3314–3324. https://doi.org/10.1038/sj.emboj.7600316

Barski A, Cuddapah S, Cui K, Roh TY, Schones DE, Wang Z, Wei G, Chepelev I, Zhao K (2007) High-resolution profiling of histone methylations in the human genome. Cell 129(4):823–837. https://doi.org/10.1016/j.cell.2007.05.009

Bian Q, Belmont AS (2012) Revisiting higher-order and large-scale chromatin organization. Curr Opin Cell Biol 24(3):359–366. https://doi.org/10.1016/j.ceb.2012.03.003

Bonisch C, Schneider K, Punzeler S, Wiedemann SM, Bielmeier C, Bocola M, Eberl HC, Kuegel W, Neumann J, Kremmer E, Leonhardt H, Mann M, Michaelis J, Schermelleh L, Hake SB (2012) H2A.Z.2.2 is an alternatively spliced histone H2A.Z variant that causes severe nucleosome destabilization. Nucleic Acids Res 40(13):5951–5964. https://doi.org/10.1093/nar/gks267

Boulard M, Gautier T, Mbele GO, Gerson V, Hamiche A, Angelov D, Bouvet P, Dimitrov S (2006) The NH2 tail of the novel histone variant H2BFWT exhibits properties distinct from conventional H2B with respect to the assembly of mitotic chromosomes. Mol Cell Biol 26 (4):1518–1526. https://doi.org/10.1128/MCB.26.4.1518-1526.2006

Burgess RJ, Zhang Z (2013) Histone chaperones in nucleosome assembly and human disease. Nat Struct Mol Biol 20(1):14–22. https://doi.org/10.1038/nsmb.2461

Buschbeck M, Hake SB (2017) Variants of core histones and their roles in cell fate decisions, development and cancer. Nat Rev Mol Cell Biol 18(5):299–314. https://doi.org/10.1038/nrm.2016.166

Buschbeck M, Uribesalgo I, Wibowo I, Rue P, Martin D, Gutierrez A, Morey L, Guigo R, Lopez-Schier H, Di Croce L (2009) The histone variant macroH2A is an epigenetic regulator of key developmental genes. Nat Struct Mol Biol 16(10):1074–1079. https://doi.org/10.1038/nsmb.1665

Cakmakci NG, Lerner RS, Wagner EJ, Zheng L, Marzluff WF (2008) SLIP1, a factor required for activation of histone mRNA translation by the stem-loop binding protein. Mol Cell Biol 28 (3):1182–1194. https://doi.org/10.1128/MCB.01500-07

Campos EI, Reinberg D (2009) Histones: annotating chromatin. Annu Rev Genet 43:559–599. https://doi.org/10.1146/annurev.genet.032608.103928

Campos EI, Fillingham J, Li G, Zheng H, Voigt P, Kuo WH, Seepany H, Gao Z, Day LA, Greenblatt JF, Reinberg D (2010) The program for processing newly synthesized histones H3.1 and H4. Nat Struct Mol Biol 17(11):1343–1351. https://doi.org/10.1038/nsmb.1911

Cook AJ, Gurard-Levin ZA, Vassias I, Almouzni G (2011) A specific function for the histone chaperone NASP to fine-tune a reservoir of soluble H3-H4 in the histone supply chain. Mol Cell 44(6):918–927. https://doi.org/10.1016/j.molcel.2011.11.021

Costanzi C, Pehrson JR (1998) Histone macroH2A1 is concentrated in the inactive X chromosome of female mammals. Nature 393(6685):599–601. https://doi.org/10.1038/31275

Creppe C, Janich P, Cantarino N, Noguera M, Valero V, Musulen E, Douet J, Posavec M, Martin-Caballero J, Sumoy L, Di Croce L, Benitah SA, Buschbeck M (2012) MacroH2A1 regulates the balance between self-renewal and differentiation commitment in embryonic and adult stem cells. Mol Cell Biol 32(8):1442–1452. https://doi.org/10.1128/MCB.06323-11

de Tayrac M, Saikali S, Aubry M, Bellaud P, Boniface R, Quillien V, Mosser J (2013) Prognostic significance of EDN/RB, HJURP, p60/CAF-1 and PDLI4, four new markers in high-grade gliomas. PLoS One 8(9):e73332. https://doi.org/10.1371/journal.pone.0073332

Doyen CM, Montel F, Gautier T, Menoni H, Claudet C, Delacour-Larose M, Angelov D, Hamiche A, Bednar J, Faivre-Moskalenko C, Bouvet P, Dimitrov S (2006) Dissection of the unusual structural and functional properties of the variant H2A.Bbd nucleosome. EMBO J 25 (18):4234–4244. https://doi.org/10.1038/sj.emboj.7601310

Draker R, Ng MK, Sarcinella E, Ignatchenko V, Kislinger T, Cheung P (2012) A combination of H2A.Z and H4 acetylation recruits Brd2 to chromatin during transcriptional activation. PLoS Genet 8(11):e1003047. https://doi.org/10.1371/journal.pgen.1003047

Drane P, Ouararhni K, Depaux A, Shuaib M, Hamiche A (2010) The death-associated protein DAXX is a novel histone chaperone involved in the replication-independent deposition of H3.3. Genes Dev 24(12):1253–1265. https://doi.org/10.1101/gad.566910

Dunleavy EM, Roche D, Tagami H, Lacoste N, Ray-Gallet D, Nakamura Y, Daigo Y, Nakatani Y, Almouzni-Pettinotti G (2009) HJURP is a cell-cycle-dependent maintenance and deposition factor of CENP-A at centromeres. Cell 137(3):485–497. https://doi.org/10.1016/j.cell.2009.02.040

Eirin-Lopez JM, Gonzalez-Romero R, Dryhurst D, Ishibashi T, Ausio J (2009) The evolutionary differentiation of two histone H2A.Z variants in chordates (H2A.Z-1 and H2A.Z-2) is mediated by a stepwise mutation process that affects three amino acid residues. BMC Evol Biol 9:31. https://doi.org/10.1186/1471-2148-9-31

Elsasser SJ (2013) A common structural theme in histone chaperones mimics interhistone contacts. Trends Biochem Sci 38(7):333–336. https://doi.org/10.1016/j.tibs.2013.04.002

Finch JT, Brown RS, Richmond T, Rushton B, Lutter LC, Klug A (1981) X-ray diffraction study of a new crystal form of the nucleosome core showing higher resolution. J Mol Biol 145 (4):757–769

Foltz DR, Jansen LE, Bailey AO, Yates JR 3rd, Bassett EA, Wood S, Black BE, Cleveland DW (2009) Centromere-specific assembly of CENP-a nucleosomes is mediated by HJURP. Cell 137 (3):472–484. https://doi.org/10.1016/j.cell.2009.02.039

Gallagher SJ, Mijatov B, Gunatilake D, Tiffen JC, Gowrishankar K, Jin L, Pupo GM, Cullinane C, Prinjha RK, Smithers N, McArthur GA, Rizos H, Hersey P (2014) The epigenetic regulator I-BET151 induces BIM-dependent apoptosis and cell cycle arrest of human melanoma cells. J Invest Dermatol 134(11):2795–2805. https://doi.org/10.1038/jid.2014.243

Gamble MJ, Frizzell KM, Yang C, Krishnakumar R, Kraus WL (2010) The histone variant macroH2A1 marks repressed autosomal chromatin, but protects a subset of its target genes from silencing. Genes Dev 24(1):21–32. https://doi.org/10.1101/gad.1876110

Goldberg AD, Banaszynski LA, Noh KM, Lewis PW, Elsaesser SJ, Stadler S, Dewell S, Law M, Guo X, Li X, Wen D, Chapgier A, DeKelver RC, Miller JC, Lee YL, Boydston EA, Holmes MC, Gregory PD, Greally JM, Rafii S, Yang C, Scambler PJ, Garrick D, Gibbons RJ, Higgs DR, Cristea IM, Urnov FD, Zheng D, Allis CD (2010) Distinct factors control histone variant H3.3 localization at specific genomic regions. Cell 140(5):678–691. https://doi.org/10.1016/j.cell.2010.01.003

Hake SB, Garcia BA, Duncan EM, Kauer M, Dellaire G, Shabanowitz J, Bazett-Jones DP, Allis CD, Hunt DF (2006) Expression patterns and post-translational modifications associated with

mammalian histone H3 variants. J Biol Chem 281(1):559–568. https://doi.org/10.1074/jbc.M509266200

Heaphy CM, de Wilde RF, Jiao Y, Klein AP, Edil BH, Shi C, Bettegowda C, Rodriguez FJ, Eberhart CG, Hebbar S, Offerhaus GJ, McLendon R, Rasheed BA, He Y, Yan H, Bigner DD, Oba-Shinjo SM, Marie SK, Riggins GJ, Kinzler KW, Vogelstein B, Hruban RH, Maitra A, Papadopoulos N, Meeker AK (2011) Altered telomeres in tumors with ATRX and DAXX mutations. Science 333(6041):425. https://doi.org/10.1126/science.1207313

Heun P, Erhardt S, Blower MD, Weiss S, Skora AD, Karpen GH (2006) Mislocalization of the Drosophila centromere-specific histone CID promotes formation of functional ectopic kinetochores. Dev Cell 10(3):303–315. https://doi.org/10.1016/j.devcel.2006.01.014

Horikoshi N, Sato K, Shimada K, Arimura Y, Osakabe A, Tachiwana H, Hayashi-Takanaka Y, Iwasaki W, Kagawa W, Harata M, Kimura H, Kurumizaka H (2013) Structural polymorphism in the L1 loop regions of human H2A.Z.1 and H2A.Z.2. Acta Crystallogr D Biol Crystallogr 69 (Pt 12):2431–2439. https://doi.org/10.1107/S090744491302252X

Horikoshi N, Arimura Y, Taguchi H, Kurumizaka H (2016) Crystal structures of heterotypic nucleosomes containing histones H2A.Z and H2A. Open Biol 6(6):160127. https://doi.org/10.1098/rsob.160127

Hu Z, Huang G, Sadanandam A, Gu S, Lenburg ME, Pai M, Bayani N, Blakely EA, Gray JW, Mao JH (2010) The expression level of HJURP has an independent prognostic impact and predicts the sensitivity to radiotherapy in breast cancer. Breast Cancer Res 12(2):R18. https://doi.org/10.1186/bcr2487

Jang CW, Shibata Y, Starmer J, Yee D, Magnuson T (2015) Histone H3.3 maintains genome integrity during mammalian development. Genes Dev 29(13):1377–1392. https://doi.org/10.1101/gad.264150.115

Jiao Y, Shi C, Edil BH, de Wilde RF, Klimstra DS, Maitra A, Schulick RD, Tang LH, Wolfgang CL, Choti MA, Velculescu VE, Diaz LA Jr, Vogelstein B, Kinzler KW, Hruban RH, Papadopoulos N (2011) DAXX/ATRX, MEN1, and mTOR pathway genes are frequently altered in pancreatic neuroendocrine tumors. Science 331(6021):1199–1203. https://doi.org/10.1126/science.1200609

Jin C, Zang C, Wei G, Cui K, Peng W, Zhao K, Felsenfeld G (2009) H3.3/H2A.Z double variant-containing nucleosomes mark 'nucleosome-free regions' of active promoters and other regulatory regions. Nat Genet 41(8):941–945. https://doi.org/10.1038/ng.409

Kato T, Sato N, Hayama S, Yamabuki T, Ito T, Miyamoto M, Kondo S, Nakamura Y, Daigo Y (2007) Activation of Holliday junction recognizing protein involved in the chromosomal stability and immortality of cancer cells. Cancer Res 67(18):8544–8553. https://doi.org/10.1158/0008-5472.CAN-07-1307

Kujirai T, Horikoshi N, Sato K, Maehara K, Machida S, Osakabe A, Kimura H, Ohkawa Y, Kurumizaka H (2016) Structure and function of human histone H3.Y nucleosome. Nucleic Acids Res 44(13):6127–6141. https://doi.org/10.1093/nar/gkw202

Lewis PW, Elsaesser SJ, Noh KM, Stadler SC, Allis CD (2010) Daxx is an H3.3-specific histone chaperone and cooperates with ATRX in replication-independent chromatin assembly at telomeres. Proc Natl Acad Sci U S A 107(32):14075–14080. https://doi.org/10.1073/pnas.1008850107

Luger K, Mader 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

Maehara K, Harada A, Sato Y, Matsumoto M, Nakayama KI, Kimura H, Ohkawa Y (2015) Tissue-specific expression of histone H3 variants diversified after species separation. Epigenetics Chromatin 8:35. https://doi.org/10.1186/s13072-015-0027-3

McGovern SL, Qi Y, Pusztai L, Symmans WF, Buchholz TA (2012) Centromere protein-A, an essential centromere protein, is a prognostic marker for relapse in estrogen receptor-positive breast cancer. Breast Cancer Res 14(3):R72. https://doi.org/10.1186/bcr3181

McKittrick E, Gafken PR, Ahmad K, Henikoff S (2004) Histone H3.3 is enriched in covalent modifications associated with active chromatin. Proc Natl Acad Sci U S A 101(6):1525–1530. https://doi.org/10.1073/pnas.0308092100

Mito Y, Henikoff JG, Henikoff S (2007) Histone replacement marks the boundaries of cis-regulatory domains. Science 315(5817):1408–1411. https://doi.org/10.1126/science.1134004

Montes de Oca R, Gurard-Levin ZA, Berger F, Rehman H, Martel E, Corpet A, de Koning L, Vassias I, Wilson LO, Meseure D, Reyal F, Savignoni A, Asselain B, Sastre-Garau X, Almouzni G (2015) The histone chaperone HJURP is a new independent prognostic marker for luminal A breast carcinoma. Mol Oncol 9(3):657–674. https://doi.org/10.1016/j.molonc.2014.11.002

Moosmann A, Campsteijn C, Jansen PW, Nasrallah C, Raasholm M, Stunnenberg HG, Thompson EM (2011) Histone variant innovation in a rapidly evolving chordate lineage. BMC Evol Biol 11:208. https://doi.org/10.1186/1471-2148-11-208

Narlikar GJ, Sundaramoorthy R, Owen-Hughes T (2013) Mechanisms and functions of ATP-dependent chromatin-remodeling enzymes. Cell 154(3):490–503. https://doi.org/10.1016/j.cell.2013.07.011

Noll M (1977) DNA folding in the nucleosome. J Mol Biol 116(1):49–71

Olszak AM, van Essen D, Pereira AJ, Diehl S, Manke T, Maiato H, Saccani S, Heun P (2011) Heterochromatin boundaries are hotspots for de novo kinetochore formation. Nat Cell Biol 13(7):799–808. https://doi.org/10.1038/ncb2272

Palmer DK, O'Day K, Wener MH, Andrews BS, Margolis RL (1987) A 17-kD centromere protein (CENP-A) copurifies with nucleosome core particles and with histones. J Cell Biol 104(4):805–815

Qiu JJ, Guo JJ, Lv TJ, Jin HY, Ding JX, Feng WW, Zhang Y, Hua KQ (2013) Prognostic value of centromere protein-A expression in patients with epithelial ovarian cancer. Tumour Biol 34(5):2971–2975. https://doi.org/10.1007/s13277-013-0860-6

Rasmussen TP, Huang T, Mastrangelo MA, Loring J, Panning B, Jaenisch R (1999) Messenger RNAs encoding mouse histone macroH2A1 isoforms are expressed at similar levels in male and female cells and result from alternative splicing. Nucleic Acids Res 27(18):3685–3689

Ratnakumar K, Duarte LF, LeRoy G, Hasson D, Smeets D, Vardabasso C, Bonisch C, Zeng T, Xiang B, Zhang DY, Li H, Wang X, Hake SB, Schermelleh L, Garcia BA, Bernstein E (2012) ATRX-mediated chromatin association of histone variant macroH2A1 regulates alpha-globin expression. Genes Dev 26(5):433–438. https://doi.org/10.1101/gad.179416.111

Ray-Gallet D, Woolfe A, Vassias I, Pellentz C, Lacoste N, Puri A, Schultz DC, Pchelintsev NA, Adams PD, Jansen LE, Almouzni G (2011) Dynamics of histone H3 deposition in vivo reveal a nucleosome gap-filling mechanism for H3.3 to maintain chromatin integrity. Mol Cell 44(6):928–941. https://doi.org/10.1016/j.molcel.2011.12.006

Sarcinella E, Zuzarte PC, Lau PN, Draker R, Cheung P (2007) Monoubiquitylation of H2A.Z distinguishes its association with euchromatin or facultative heterochromatin. Mol Cell Biol 27(18):6457–6468. https://doi.org/10.1128/MCB.00241-07

Schenk R, Jenke A, Zilbauer M, Wirth S, Postberg J (2011) H3.5 is a novel hominid-specific histone H3 variant that is specifically expressed in the seminiferous tubules of human testes. Chromosoma 120(3):275–285. https://doi.org/10.1007/s00412-011-0310-4

Schneiderman JI, Orsi GA, Hughes KT, Loppin B, Ahmad K (2012) Nucleosome-depleted chromatin gaps recruit assembly factors for the H3.3 histone variant. Proc Natl Acad Sci U S A 109(48):19721–19726. https://doi.org/10.1073/pnas.1206629109

Schwartzentruber J, Korshunov A, Liu XY, Jones DT, Pfaff E, Jacob K, Sturm D, Fontebasso AM, Quang DA, Tonjes M, Hovestadt V, Albrecht S, Kool M, Nantel A, Konermann C, Lindroth A, Jager N, Rausch T, Ryzhova M, Korbel JO, Hielscher T, Hauser P, Garami M, Klekner A, Bognar L, Ebinger M, Schuhmann MU, Scheurlen W, Pekrun A, Fruhwald MC, Roggendorf W, Kramm C, Durken M, Atkinson J, Lepage P, Montpetit A, Zakrzewska M, Zakrzewski K, Liberski PP, Dong Z, Siegel P, Kulozik AE, Zapatka M, Guha A, Malkin D, Felsberg J, Reifenberger G, von Deimling A, Ichimura K, Collins VP, Witt H, Milde T, Witt O,

Zhang C, Castelo-Branco P, Lichter P, Faury D, Tabori U, Plass C, Majewski J, Pfister SM, Jabado N (2012) Driver mutations in histone H3.3 and chromatin remodelling genes in paediatric glioblastoma. Nature 482(7384):226–231. https://doi.org/10.1038/nature10833

Siegel TN, Hekstra DR, Kemp LE, Figueiredo LM, Lowell JE, Fenyo D, Wang X, Dewell S, Cross GA (2009) Four histone variants mark the boundaries of polycistronic transcription units in Trypanosoma brucei. Genes Dev 23(9):1063–1076. https://doi.org/10.1101/gad.1790409

Singh N, Basnet H, Wiltshire TD, Mohammad DH, Thompson JR, Heroux A, Botuyan MV, Yaffe MB, Couch FJ, Rosenfeld MG, Mer G (2012) Dual recognition of phosphoserine and phosphotyrosine in histone variant H2A.X by DNA damage response protein MCPH1. Proc Natl Acad Sci U S A 109(36):14381–14386. https://doi.org/10.1073/pnas.1212366109

Sun X, Clermont PL, Jiao W, Helgason CD, Gout PW, Wang Y, Qu S (2016) Elevated expression of the centromere protein-A(CENP-A)-encoding gene as a prognostic and predictive biomarker in human cancers. Int J Cancer 139(4):899–907. https://doi.org/10.1002/ijc.30133

Suto RK, Clarkson MJ, Tremethick DJ, Luger K (2000) Crystal structure of a nucleosome core particle containing the variant histone H2A.Z. Nat Struct Biol 7(12):1121–1124. https://doi.org/10.1038/81971

Tachiwana H, Osakabe A, Kimura H, Kurumizaka H (2008) Nucleosome formation with the testis-specific histone H3 variant, H3t, by human nucleosome assembly proteins in vitro. Nucleic Acids Res 36(7):2208–2218. https://doi.org/10.1093/nar/gkn060

Tagami H, Ray-Gallet D, Almouzni G, Nakatani Y (2004) Histone H3.1 and H3.3 complexes mediate nucleosome assembly pathways dependent or independent of DNA synthesis. Cell 116 (1):51–61

Talbert PB, Ahmad K, Almouzni G, Ausio J, Berger F, Bhalla PL, Bonner WM, Cande WZ, Chadwick BP, Chan SW, Cross GA, Cui L, Dimitrov SI, Doenecke D, Eirin-Lopez JM, Gorovsky MA, Hake SB, Hamkalo BA, Holec S, Jacobsen SE, Kamieniarz K, Khochbin S, Ladurner AG, Landsman D, Latham JA, Loppin B, Malik HS, Marzluff WF, Pehrson JR, Postberg J, Schneider R, Singh MB, Smith MM, Thompson E, Torres-Padilla ME, Tremethick DJ, Turner BM, Waterborg JH, Wollmann H, Yelagandula R, Zhu B, Henikoff S (2012) A unified phylogeny-based nomenclature for histone variants. Epigenetics Chromatin 5:7. https://doi.org/10.1186/1756-8935-5-7

Tanasijevic B, Rasmussen TP (2011) X chromosome inactivation and differentiation occur readily in ES cells doubly-deficient for macroH2A1 and macroH2A2. PLoS One 6(6):e21512. https://doi.org/10.1371/journal.pone.0021512

Tang MC, Jacobs SA, Mattiske DM, Soh YM, Graham AN, Tran A, Lim SL, Hudson DF, Kalitsis P, O'Bryan MK, Wong LH, Mann JR (2015) Contribution of the two genes encoding histone variant h3.3 to viability and fertility in mice. PLoS Genet 11(2):e1004964. https://doi.org/10.1371/journal.pgen.1004964

Tolstorukov MY, Goldman JA, Gilbert C, Ogryzko V, Kingston RE, Park PJ (2012) Histone variant H2A.Bbd is associated with active transcription and mRNA processing in human cells. Mol Cell 47(4):596–607. https://doi.org/10.1016/j.molcel.2012.06.011

Tomonaga T, Matsushita K, Yamaguchi S, Oohashi T, Shimada H, Ochiai T, Yoda K, Nomura F (2003) Overexpression and mistargeting of centromere protein-A in human primary colorectal cancer. Cancer Res 63(13):3511–3516

Ueda J, Harada A, Urahama T, Machida S, Maehara K, Hada M, Makino Y, Nogami J, Horikoshi N, Osakabe A, Taguchi H, Tanaka H, Tachiwana H, Yao T, Yamada M, Iwamoto T, Isotani A, Ikawa M, Tachibana T, Okada Y, Kimura H, Ohkawa Y, Kurumizaka H, Yamagata K (2017) Testis-specific histone variant H3t gene is essential for entry into spermatogenesis. Cell Rep 18(3):593–600. https://doi.org/10.1016/j.celrep.2016.12.065

Urahama T, Harada A, Maehara K, Horikoshi N, Sato K, Sato Y, Shiraishi K, Sugino N, Osakabe A, Tachiwana H, Kagawa W, Kimura H, Ohkawa Y, Kurumizaka H (2016) Histone H3.5 forms an unstable nucleosome and accumulates around transcription start sites in human testis. Epigenetics Chromatin 9:2. https://doi.org/10.1186/s13072-016-0051-y

Valente V, Serafim RB, de Oliveira LC, Adorni FS, Torrieri R, Tirapelli DP, Espreafico EM, Oba-Shinjo SM, Marie SK, Paco-Larson ML, Carlotti CG Jr (2013) Modulation of HJURP (Holliday junction-recognizing protein) levels is correlated with glioblastoma cells survival. PLoS One 8(4):e62200. https://doi.org/10.1371/journal.pone.0062200

Vardabasso C, Hasson D, Ratnakumar K, Chung CY, Duarte LF, Bernstein E (2014) Histone variants: emerging players in cancer biology. Cell Mol Life Sci 71(3):379–404. https://doi.org/10.1007/s00018-013-1343-z

Vardabasso C, Hake SB, Bernstein E (2016) Histone variant H2A.Z.2: A novel driver of melanoma progression. Mol Cell Oncol 3(2):e1073417. https://doi.org/10.1080/23723556.2015.1073417

Verdaasdonk JS, Bloom K (2011) Centromeres: unique chromatin structures that drive chromosome segregation. Nat Rev Mol Cell Biol 12(5):320–332. https://doi.org/10.1038/nrm3107

Wen H, Li Y, Xi Y, Jiang S, Stratton S, Peng D, Tanaka K, Ren Y, Xia Z, Wu J, Li B, Barton MC, Li W, Li H, Shi X (2014) ZMYND11 links histone H3.3K36me3 to transcription elongation and tumour suppression. Nature 508(7495):263–268. https://doi.org/10.1038/nature13045

Wiedemann SM, Mildner SN, Bonisch C, Israel L, Maiser A, Matheisl S, Straub T, Merkl R, Leonhardt H, Kremmer E, Schermelleh L, Hake SB (2010) Identification and characterization of two novel primate-specific histone H3 variants, H3.X and H3.Y. J Cell Biol 190(5):777–791. https://doi.org/10.1083/jcb.201002043

Witt O, Albig W, Doenecke D (1996) Testis-specific expression of a novel human H3 histone gene. Exp Cell Res 229(2):301–306. https://doi.org/10.1006/excr.1996.0375

Wong LH, McGhie JD, Sim M, Anderson MA, Ahn S, Hannan RD, George AJ, Morgan KA, Mann JR, Choo KH (2010) ATRX interacts with H3.3 in maintaining telomere structural integrity in pluripotent embryonic stem cells. Genome Res 20(3):351–360. https://doi.org/10.1101/gr.101477.109

Wu Q, Qian YM, Zhao XL, Wang SM, Feng XJ, Chen XF, Zhang SH (2012) Expression and prognostic significance of centromere protein A in human lung adenocarcinoma. Lung Cancer 77(2):407–414. https://doi.org/10.1016/j.lungcan.2012.04.007

Zalensky AO, Siino JS, Gineitis AA, Zalenskaya IA, Tomilin NV, Yau P, Bradbury EM (2002) Human testis/sperm-specific histone H2B (hTSH2B). Molecular cloning and characterization. J Biol Chem 277(45):43474–43480. https://doi.org/10.1074/jbc.M206065200

Zhang R, Poustovoitov MV, Ye X, Santos HA, Chen W, Daganzo SM, Erzberger JP, Serebriiskii IG, Canutescu AA, Dunbrack RL, Pehrson JR, Berger JM, Kaufman PD, Adams PD (2005) Formation of macroH2A-containing senescence-associated heterochromatin foci and senescence driven by ASF1a and HIRA. Dev Cell 8(1):19–30. https://doi.org/10.1016/j.devcel.2004.10.019

Zhou J, Fan JY, Rangasamy D, Tremethick DJ (2007) The nucleosome surface regulates chromatin compaction and couples it with transcriptional repression. Nat Struct Mol Biol 14(11):1070–1076. https://doi.org/10.1038/nsmb1323

DNA Methylation and Carcinogenesis: Current and Future Perspectives

7

Vibha Patil and Zdenko Herceg

Abstract

Alterations in DNA methylation occur frequently in virtually all types of cancer. In light of the strong cumulative evidence of the role of epigenetic modifications in carcinogenesis, the epigenome is now considered an attractive candidate for therapeutically targeting cancer. This chapter highlights the role of DNA methylation and its contribution to cancer development and progression, as well as the current and future perspectives of utilising DNA methylation for improving clinical outcomes for cancer patients.

Keywords

DNA methylation · Epigenetic therapy · CRISPR · Decitabine · Biomarkers · Hypermethylation · Hypomethylation

7.1 Epigenetic Alterations in Cancer

7.1.1 DNA Methylation

The main avenue through which the epigenome is found to be altered in cancer is via DNA methylation, including global hypomethylation and site-specific hypermethylation, as well as direct mutations within cytosine methylation-containing sequences (Sawan et al. 2008). The intricate ways in which these changes interact to promote cancer are yet to be fully understood (Sinčić and Herceg 2011). This section provides an overview of the types of cancer-related DNA methylation

V. Patil · Z. Herceg (✉)
Epigenetics Group, International Agency for Research on Cancer, World Health Organization, Lyon CEDEX 08, France
e-mail: PatilV@fellows.iarc.fr; HercegZ@iarc.fr

L. B. Hesson, A. L. Pritchard (eds.), *Clinical Epigenetics*,
https://doi.org/10.1007/978-981-13-8958-0_7

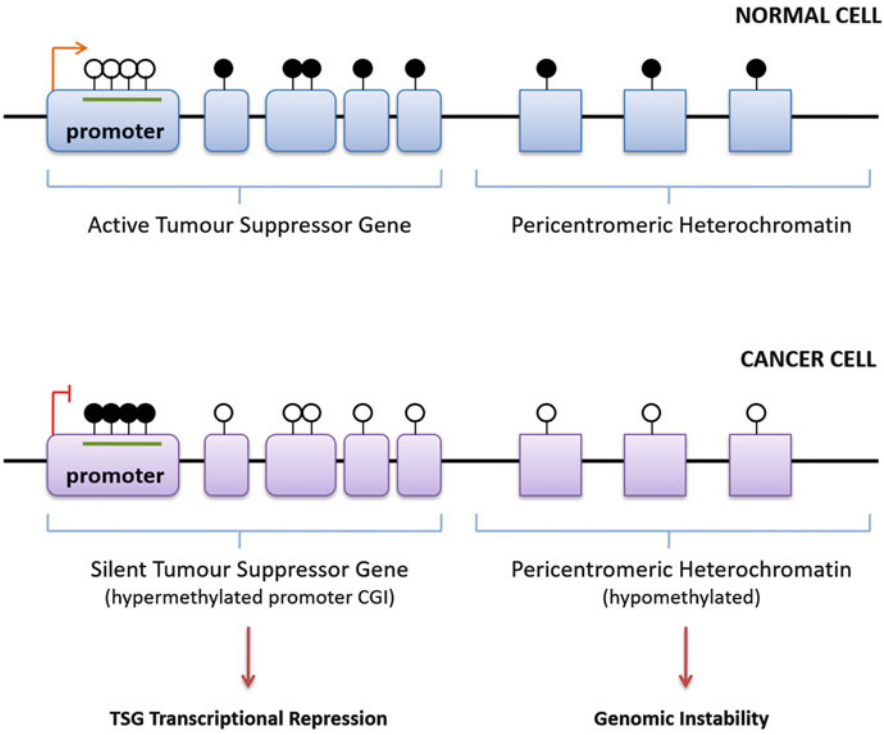

Fig. 7.1 DNA methylation changes in cancer. In normal cells, transcriptionally active tumour suppressor genes contain an unmethylated CpG island promoter. Methylation of repeat elements is important for ensuring that repeat rich regions remain transcriptionally silent, whilst methylation of pericentromeric heterochromatin is important for maintaining heterochromatin structure. In cancer cells, some tumour suppressor genes become hypermethylated within the CpG island promoter, resulting in transcriptional silencing of the linked gene. Additionally, DNA hypomethylation of pericentromeric heterochromatin may contribute to genomic instability due to, for example, increased mitotic recombination, whilst hypomethylation of repeat elements can result in their aberrant expression, including the expression of transposable elements

alterations that occur frequently in virtually all cancer types. A large body of research has shown that these altered methylation patterns are an essential component of tumour development and progression.

7.1.2 Hypomethylation

Global hypomethylation is the overall reduction of total cytosine methylation across regions of the genome where a much higher quantity of methylation would normally be expected. Thanks to multiple studies analysing normal tissue versus cancer tissue, we now know that hypomethylation is frequently detected in the majority of cancer types including prostate cancer, ovarian cancer, hepatocellular carcinoma and glioblastomas (Bedford and van Helden 1987; Cadieux et al. 2006; Chen et al.

1998; Kim et al. 1994; Lin et al. 2001). As highlighted in Fig. 7.1, 80% of CpG sites outside of CpG-rich regions (CpG Islands) are methylated. In cancer cells, however, the proportion of CpG sites outside of CpG islands that are methylated decreases to around 40–60%. Global hypomethylation contributes to one of the hallmark characteristics of cancer, that is, genomic instability (Baylin and Jones 2016). Evidence from DNA methyltransferase 1 (*DNMT1*) knockout studies shows that this reduction in cytosine methylation leads to increased chromosome frailty indicated by increased mutation rates and aneuploidy (Chen et al. 1998; Gaudet et al. 2003). For example, one study determined this by using mice containing a single copy of the *Tp53* and *Nf1* genes and a hypomorphic *Dnmt1* allele (an allele with partial loss of *Dnmt1* function, $Dnmt1^{-}/^{+}$), which showed that chromosome 11 had a higher frequency of deletions concomitant with centromeric hypomethylation and instability (Eden et al. 2003).

The largest contributor of global hypomethylation observed in cancer is demethylation of repeat regions, such as LINE-1 and Alu sequences, which collectively comprise approximately half of the genome (Ehrlich 2009; Park et al. 2014; Xiao-Jie et al. 2016). This type of global hypomethylation has been associated with tumour progression and degree of malignancy in some cancers. For example, methylation analysis of LINE-1 retrotransposon regions in 113 prostate carcinoma samples using southern blot hybridisation showed hypomethylation in 49% of prostate carcinoma samples, and this hypomethylation was significantly associated with high tumour grade (p value $= 0.04$) (Florl et al. 2004). Another study used bisulphite-specific PCR and DNA sequencing to assess LINE-1 methylation status in 71 HCC cases. It was shown that 83.7% of HCC patients had LINE-1 hypomethylation. LINE-1 hypomethylation was associated with a poor overall survival of 35 months compared to approximately 40 months for HCC patients with hypermethylation in LINE-1 (Gao et al. 2014). Alternatively, site-specific DNA hypomethylation can also be found within transcriptional regulatory regions such as gene promoters and enhancers. This site-specific hypomethylation is involved in activating expression of oncogenes and other target genes that would normally remain inactive (Pakneshan et al. 2004). For example, protease urokinase (PLAU/uPA) is found to be hypomethylated and over-expressed in breast cancer and correlates with cancer stage and progression. uPA mRNA was detected only in highly invasive tumour cells where the promoter was completely demethylated, whereas no uPA mRNA was detected in normal breast epithelium or low-invasive breast cancer cells, which contained methylated uPA promoters (Pakneshan et al. 2004). Other features of promoter sequences (such as CpG density) are important modulators of the impact that DNA methylation may have on the transcriptional activity of the gene (Weber et al. 2007). Other studies, using microarray expression analysis of gastric cancer tissues, have also shown DNA hypomethylation can lead to over-expression of oncogenes such as *R-RAS*, *FRAT2* and *RHO6* in gastric cancer (Nishigaki et al. 2005; Wilson et al. 2007). Cancer-associated site-specific hypomethylation and overall global demethylation are interlinked processes (Kaneda et al. 2004) whose frequent occurrence across a wide variety of cancer types contributes to genomic instability, inappropriate gene activation as well as the degree of malignancy.

7.1.3 Hypermethylation

In the human genome, 72% of all gene promoter regions are encompassed by a CpG island (CGI) (Saxonov et al. 2006). In cancer cells, hypermethylation is often observed at promoter CGIs of key genes involved in processes such as cell cycle regulation, DNA repair, apoptosis and differentiation (Fig. 7.1). The *RASSF1*, *CDKN2A*, *MLH1, PCDH18* and *BRCA1* are examples of genes that are inactivated by promoter hypermethylation in cancer (Baylin 2005; Kanwal and Gupta 2010; Zhou et al. 2017). Promoter hypermethylation of CGIs correlates with transcriptional inactivity which leads to reduced expression of proteins that are essential for downstream cellular pathways such as DNA damage repair and enables cells to accumulate further genetic lesions. For example, the DNA mismatch repair (MMR) system constitutes a set of genes which protects DNA integrity by repairing errors that can arise during DNA replication. These errors can occur as a result of slippage of the DNA polymerases along repetitive sequences, commonly known as microsatellite instability (MSI). Microsatellite loci contain a large portion of repetitive sequences, making them vulnerable to this type of damage, and MSI can arise from defects in the MMR system. Hence, an impaired MMR system can increase frame shift mutations especially if the microsatellites are located within a gene. *MLH1* is an important MMR gene, and epigenetic inactivation via promoter hypermethylation is closely associated with many cancers including hereditary colon cancer, sporadic colon cancer, non-small cell lung cancer, acute myeloid leukaemia and ovarian cancer. Loss of *MLH1* expression in sporadic colorectal cancer most commonly occurs by biallelic promoter hypermethylation, resulting in the loss of MLH1 protein and consequently its binding partner PMS2. These proteins form a heterodimer that is recruited by the MSH2-MSH6 heterodimer, which is involved in the recognition of single nucleotide mismatches and insertion-deletion loops. Repair involves a process of nucleotide excision and resynthesis, which is facilitated by other MMR proteins including exonuclease 1 (EXO1), the proliferating cell nuclear antigen (PCNA), replication factor C (RFC) and DNA polymerases δ or ε (Pluciennik et al. 2010). Loss of MMR ability results in a marked accumulation of genetic lesions across the genome and a characteristic mutation signature (Alexandrov et al. 2013). In colorectal cancer, a causal link between MSI and *MLH1* promoter hypermethylation is highlighted by the fact that methylation of *MLH1* gives rise to microsatellite instability (MSI+) cancers, which account for about 15 % of all colorectal cancers (Esteller et al. 2001).

Another example is promoter hypermethylation of *BRCA1*, which occurs in a subset of sporadic breast cancers. BRCA1 is an essential component of maintaining DNA stability and contributes to the regulation of DNA double-strand break repair, cell cycle check point regulation, ubiquitination and transcriptional regulation (Yoshida and Miki 2004). Hence, reduced expression or inactivation of this gene via promoter hypermethylation is thought to contribute to accumulations of DNA mutations in neoplastic cells (Catteau and Morris 2002).

Epigenetic inactivation of the death-associated protein kinase (*DAPK1*) gene via promoter hypermethylation is also a frequent event in a variety of tumours including

Table 7.1 Classes of hypermethylated genes in cancer

Class of hypermethylated gene	Examples
Known TSG	• *VHL* • *CDH1* • *CDKN2A* • *MLH1* • *APC* • *STK4*
Candidate TSG	• *FHIT* • *RASSF1A* • *MGMT* • *GST* • *GATA4/5* • *DAP*
Genes discovered through screens for hypermethylated genes	• *HIC-1* • *SFRP1, SFRP2, SFRP4, SFRP5* • *BMP-3* • *SLC5A8* • *SSI1*

Taken from Baylin SB, Jones PA (2016) Epigenetic determinants of cancer. *Cold Spring Harb Perspect Biol* 8(9)

lung, bladder, renal, breast and head and neck cancer. The prevalence of hypermethylation of this gene in non-small cell lung carcinoma may be as high as 23–44% (Pulling et al. 2004). Strikingly, one study observed *DAPK1* hypermethylation in 100% of Burkett's lymphoma samples and 84% of B-cell non-Hodgkin's lymphomas (Katzenellenbogen et al. 1999), suggesting a particularly important role in some haematological malignancies. DAP-kinase is an enzyme whose expression is required for gamma interferon-induced apoptosis triggered by detachment of cells from the extracellular matrix (Pulling et al. 2004). Since loss of DAP-kinase can allow cells to evade apoptosis, it has been suggested that epigenetic inactivation of the *DAPK1* gene at the early stages of cancer formation provides a selective growth advantage, whilst loss at a later stage of cancer formation allows for metastasis (Pulling et al. 2004).

The causal mechanism responsible for aberrant DNA methylation in cancer is not yet completely understood (Vaissière et al. 2008); however, the advent of genome-wide sequencing technologies has produced an ever-increasing list of genes affected by aberrant promoter hypermethylation in virtually all cancer types. Currently, hypermethylated genes can be classified into three main categories, namely, classic tumour suppressor genes (TSGs), candidate TSGs and hypermethylated genes with an as yet unknown functional consequence, as summarised by Baylin and Jones in Table 7.1 (Baylin and Jones 2016). Classic TSGs constitute those genes that have a proven causal role in the tumorigenic process and show pathogenic germline mutations in families with hereditary cancer syndromes. An example of a classic TSG inactivated by hypermethylation is *MLH1* in colorectal cancer, which as previously described, has been linked with processes important in cancer

development, including DNA mismatch repair, and shows pathogenic germline mutations in approximately 40% of individuals with Lynch syndrome. Candidate TSGs are genes that encode proteins with functions consistent with tumour suppression but do not show pathogenic germline mutations in hereditary cancer syndromes. As a result, the causal role of these genes in tumorigenesis is unclear. An example would be *RASSF1A,* which shows hypermethylation in many tumour types including lung, kidney and breast cancers. *RASSF1A* hypermethylation disrupts apoptotic pathways, as well as microtubule dynamics during mitotic progression (Hesson et al. 2007), but as yet has not been found to be mutated in hereditary cancer syndromes. Finally, the third list of genes, which include *SLC5A8* and *SFRP1*, are identified as hypermethylated in tumours but have no known functional role in cancer. These genes are important to investigate as they may be key in further dissecting the causal mechanisms of cancer (Baylin and Jones 2016). The challenge now remains to identify which of this third class of hypermethylated genes are essential in driving the tumorigenic process and which are likely to be passenger events that arise as a consequence of cancer, since this has interesting clinical implications for the future (Jones et al. 2016).

7.1.4 Somatic Mutations and DNA Methylation

Another method of disrupting methylation within the genome is via somatic mutation. These mutations may fall within DNA regions that would normally contain methylated cytosine. Missense mutations can arise from DNA damage via deamination where cytosine and methylated cytosine spontaneously deaminate to thymine. If this type of damage is not repaired, it results in a C:G>A:T transition. DNA repair is less efficient at repairing deamination-induced mismatches and can leave DNA more susceptible to further damage. If this occurs within a transcriptionally active region, this can affect the expression and/or function of important genes. It is now known that a significantly large proportion of such mutations occur within the context of CpG dinucleotides, which, depending on their genomic context, are often methylated (Greenblatt et al. 1994). A study that collectively analysed 2567 *TP53* mutations that were reported across different cancer types found that 24% of all *TP53* mutations were detected within a CpG context. From this analysis, 'hotspots' could also be identified based on frequency of mutation at any particular site (Greenblatt et al. 1994). For example, approximately 50% of all colon cancers contained mutations within three specific sites of CpG dinucleotides within the *TP53* gene that are normally methylated (Denissenko et al. 1996; Greenblatt et al. 1994). It has been shown that carcinogen exposure results in carcinogenic adduct formation to occur at preferential binding sites within the *TP53* gene (Denissenko et al. 1996). By using plasmid DNA containing genomic *TP53* sequences and exposing them to the carcinogen, investigators confirmed that methylated CpG sites attracted carcinogenic adduct formation with a higher propensity than non-methylated sites. Hence, the presence of CpG methylation modification itself may be promoting aberrant somatic mutations, which then alter the expected

methylation patterns and propagate carcinogens. In this scenario, aberrant methylation may affect gene expression indirectly, through predisposing the gene sequence to mutational events.

7.1.5 Mutations in Methylation Modifiers

Another important consideration is the mutation of genes that have a direct impact on DNA methylation patterns; this concept extends to histone modification patterns as well as chromatin structure via nucleosome positioning, discussed in further detail below. For example, IDH1 and IDH2 are isocitrate dehydrogenase enzymes that are involved in the production of α-ketoglutarate. Mutations in these genes can result in an alternative product, 2-hydroxyglutarate, which has been associated with hypermethylation in cancers such as leukaemia and brain (glioma) and colon cancer. *IDH1* and *IDH2* mutations are found in tumours with a higher frequency of promoter CpG island hypermethylation. This is particularly relevant since α-ketoglutarate is a crucial cofactor for the TET group of enzymes, which actively demethylate DNA. Mutations in the IDH1/2 enzymes can result in an imbalance of cytosine methylation levels in the genome. Furthermore, mouse models in which a common human *IDH1* mutation was inserted into the endogenous murine *Idh1* locus, and expressed in all cells of myeloid lineage, confirmed these mutations are involved in promoting the patterns of DNA methylation alterations often observed in human acute myeloid leukaemia (Sasaki et al. 2012). Together these studies show that mutations in a specific subset of genes can cause aberrant DNA methylation patterns that ultimately contribute to tumorigenesis.

7.2 The Clinical Relevance of DNA Methylation in Cancer

7.2.1 Epigenetic Therapies

In light of the strong cumulative evidence of the role of epigenetic modifications in carcinogenesis, the epigenome is an attractive candidate for therapeutic targeting in cancer. An epigenetic therapy can be defined as a treatment that aims to target epigenetic aberrations in cancer in order to gain clinical benefits (Sharma et al. 2010). This approach aims to restore the normal epigenetic state, via direct or indirect alternative mechanisms (Jones et al. 2016). Currently there are multiple epigenetic drugs being developed and tested which target a variety of different epigenetic modifications and modifiers with the aim of inhibiting histone methylation or altering nucleosome occupancy as well as inhibiting DNA methylation. This section outlines evidence showcasing how epigenetic therapies can benefit cancer patients by improving clinical outcomes with a focus on DNA methylation inhibitors (Fig. 7.2).

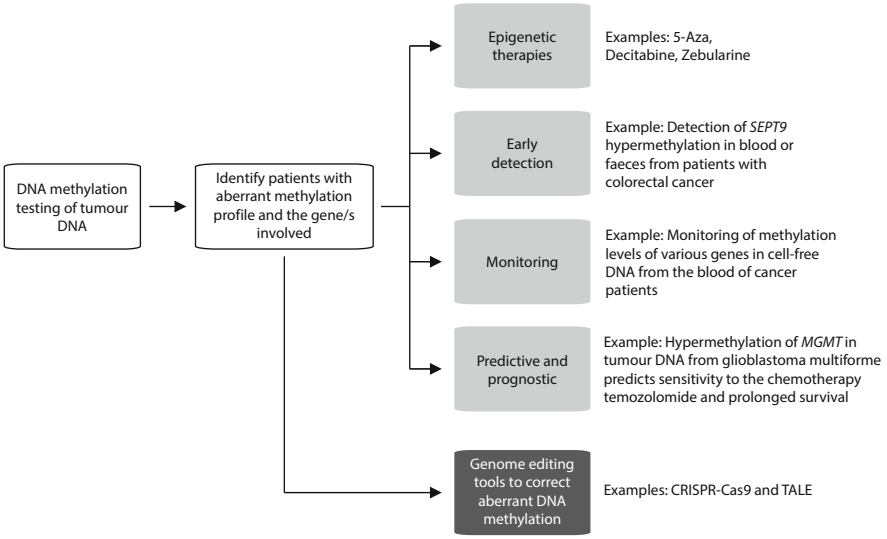

Fig. 7.2 A summarised schematic of the clinical relevance of DNA methylation testing in cancer indicating current (light grey) and future (dark grey) perspectives

7.2.2 Reversibility of Aberrant DNA Methylation

DNA methyltransferase (DNMT) enzymes are regulated by oncogenic and tumour suppressor pathways. Over-expression and high activity of DNMTs are observed in many types of cancer. These observations provide a rational for inhibiting DNMTs as a potential therapeutic option.

The DNMT inhibitors 5-aza-2′-deoxycytidine (5-Aza-dC or 5-Aza-dC) and 5-azacytidine (5-Aza) are examples of epigenetic drugs that alter levels of DNA methylation. 5-Aza-dC is a later deoxy derivative of 5-Aza; they are both nucleoside analogues that act, in part, by reducing levels of DNA methylation. In doing so, they reactivate the expression of several tumour suppressor genes that are silenced in cancer cells by promoter hypermethylation (Christman 2002), though recent work has identified other mechanisms of action of these drugs (Chiappinelli et al. 2015; Strick et al. 2015).

First produced in the mid-1960s by Piskala and Sorm, 5-Aza and 5-Aza-dC were soon recognised as potent inhibitors of DNA methylation (Futscher 2012; Sorm et al. 1964). 5-Aza-dC enters the cell and is incorporated into replicating DNA in the place of cytosine. Consequently, in the presence of 5-Aza or 5-Aza-dC, DNMT becomes trapped onto DNA via a covalent bond where it is proteolytically degraded. The overall result is DNMT1 depletion and DNA hypomethylation (Egger et al. 2004; Santini et al. 2001). Both 5-Aza and 5-Aza-dC were approved by the US Food and Drug Administration (FDA) in 2004 and 2006, respectively, for use as first-line therapies in the treatment of myelodysplastic syndrome (MDS) due to their overall

beneficial outcomes for patients (Issa et al. 2005). In the phase III clinical trial that formed the basis of the current FDA approval standards, de novo and secondary MDS patients were randomly assigned to supportive care only or 15 mg/m^2 of 5-Aza-dC every 8 h for 3 days every 6 weeks with supportive care. A complete response (CR) versus partial response (PR) in this study was defined by set criteria for parameters including bone marrow, haemoglobin, neutrophil and platelet levels. The overall response rate (CR + PR) in patients exposed to the drug was 17% compared to 0% in the group assigned with supportive care only. Furthermore, in the group of patients exposed to the drug, those with pathologically confirmed MDS showed a 21% overall response rate. Progression to AML or death in the 5-Aza-dC exposed group was on average 4.5 months later than those in the supportive care only group (Saba and Wijermans 2005). On average, 50% of MDS patients treated with 5-Aza-dC showed haematologic improvement and an improved quality of life, whilst 20% of MDS patients showed no response (Karahoca and Momparler 2013).

7.2.3 Epigenetic Therapies as Useful Adjuncts to Traditional Therapies

Combining DNA methylation inhibitors with traditional chemotherapy options has been trialled under the rationale that reversing cancer-associated hypermethylation may enhance therapeutic outcomes (Fig. 7.2). For example, some success in phase I/II clinical trials of 5-Aza combined with entinostat (a novel HDAC inhibitor undergoing pre-clinical testing) in non-small cell lung cancer (NSCLC) has been reported (Juergens et al. 2011). In this clinical trial, a median 6.4-month survival rate in the patient cohort treated with 5-Aza and entinostat was observed. This compares favourably with the single existing FDA-approved drug for this patient population, erlotinib, which has a median survival of 6.7 months (vs. 4.7 months for placebo). Improved survival was attributed to the demethylation and reactivation of key genes epigenetically silenced in NSCLC (Juergens et al. 2011). Appropriate dosage is a critical consideration in regard to achieving optimal therapeutic outcome. More recent clinical trials have shown low doses of 5-Aza-dC are much more efficacious. For example, a dosage schedule of 20 mg/m^2 every hour for 5 days resulted in a 39% CR rate (Juergens et al. 2011). Lower doses also had lower risk of previously encountered side effects such as thrombocytopenia [reviewed by (Saba 2007)].

The ability to reactivate key TSGs via demethylation is at the crux of the promising results these drugs have produced in both in vitro assays and in clinical settings. A number of studies to date have shown administering 5-Aza-dC to cancer cells can induce re-expression of key TSGs such as *MLH1* (Khamas et al. 2012; Mossman et al. 2010). In fact, the sensitisation of cancer cells to chemotherapies caused by epigenetic drugs is thought to be due to the re-expression of pro-apoptotic genes (Plumb et al. 2004). In vitro studies of human ovarian cancer cell lines have shown *BRCA1* promoter hypermethylation may contribute to the development of drug resistance, which often impedes therapeutic outcome normally achieved with platinum-based therapies such as cisplatin. Hence, concomitant treatment with a

DNA methylation inhibitor was suggested to improve patient outcome (Wang et al. 2010). In a phase II clinical trial, seventeen patients with platinum-resistant ovarian cancer were exposed to low dose 5-Aza-dC treatment (10 mg/m^2) administered daily for 5 days followed by carboplatin, which was administered on Day 8. This treatment cycle was continued until disease progression or intolerable toxicity. Through global and gene-specific methylation analyses, it was concluded that demethylation of important tumour suppressor genes such as *MLH1, RASSF1A* and *HOXA10* positively correlates with progression-free survival. By altering cancer pathways, via gene reactivation and restoring carboplatin sensitivity, 5-Aza-dC pretreatment resulted in a much higher response rate in this patient group (Matei et al. 2012). Gene reactivation is considered a better predictor of response to therapy than DNA hypomethylation (Blum et al. 2007; Kantarjian et al. 2007), which highlights the importance of sustaining gene re-expression for the purposes of achieving long-term clinical response.

In conclusion, using epigenetic drugs in combination with existing chemotherapeutic options has the potential to be a successful approach for cancer treatment. Further details regarding epigenetic therapies and their mechanisms of action are described in Chapter 10, "Novel Epigenetic Therapies".

7.2.4 Personalised Treatment Options

Personalised treatment options are an avenue of cancer therapy that provides tailored treatment to patients based on their personal genomic profile. The notion that tumour heterogeneity may theoretically be manipulated or controlled via epigenetic intervention has highlighted epigenetic drugs as potential candidates for personalised treatment options. This rationale centres on the fact that methylation profiles contribute heavily to phenotypic differences observed between tumours especially in the context of drug resistance. An example would be the case of breast cancer and endocrine therapy. More than two thirds of breast cancers express the oestrogen receptor gene (*ESR1*) and are categorised as ER+ breast cancer. Once ER+ status has been confirmed, normally through testing of surgically removed breast tumour tissue, inhibiting the body from producing hormones is a common approach to slowing or stopping the growth of hormone-dependent tumours. Response to endocrine therapies, such as tamoxifen, has been shown to be directly proportional to the degree of endocrine receptor expression (Lustberg and Ramaswamy 2011). One of the major mechanisms of resistance to endocrine therapy is a gradual loss of *ESR1* expression, which can be due to epigenetic modification, including methylation of the CpG island in the *ESR1* promoter. Breast cancer patients that have acquired resistance over time can be resensitised to endocrine therapy via exposure to hypomethylating agents such as Decitabine. The resensitisation is a direct result of *ESR1* promoter demethylation which allows *ESR1* mRNA and functional proteins to be re-expressed (Lustberg and Ramaswamy 2011). Similarly, loss of DNA mismatch repair occurring via *MLH1* promoter hypermethylation has been shown to be a key instigator in producing resistance toward many anti-cancer drugs in various tumours

(Yan et al. 2016). Epigenetic intervention has also been shown to be of use in this circumstance. For example, drug-resistant ovarian and colon tumour xenografts were shown to become resensitised to cisplatin and carboplatin after exposure to Decitabine, which was attributed to *MLH1* re-expression and restoration of mismatch repair functionality (Plumb et al. 2000). These examples show that tailored therapeutic modalities customised to the methylation profile of specific genes within treatment-resistant tumours may provide further options when conventional therapies fail.

7.3 Clinical Implications of DNA Methylation in Cancer Prognosis

DNA methylation is potentially important in cancer staging, in diagnosis and in predicting treatment response (Fig. 7.2). Through the advent of technologies that now allow for genome-wide methylation profiling of tumours, it has become apparent that promoter hypermethylation of specific genes in certain cancers is a recurrent event. This section highlights some of the evidence from multiple studies that have allowed for the identification of certain genes which, when hypermethylated in specific cancers, have implications for cancer staging, diagnosis and treatment.

7.3.1 DNA Methylation as a Potential Biomarker in Cancer Diagnosis

The utility of DNA methylation analysis has recently been considered in cancer diagnosis. If a biomarker is defined as a region that shows consistent difference between cancerous and non-cancerous cells, then in the context of DNA methylation, a biomarker can either be aberrantly hypermethylated or hypomethylated. This consistent change can be considered a reflection of abnormally functioning oncogenes or tumour suppressor genes. Hence, there is a large number of potential biomarkers that may be considered for development as cancer biomarkers (Levenson and Melnikov 2012). The diagnosis of some cancer types is less complicated, such as skin melanoma, where the location of the tumour is highly accessible for biopsy and visual assessment is also possible. This is also somewhat true in breast cancer, where palpable lesions can be biopsied. In other circumstances, however, such as ovarian cancer, the problem of diagnosis becomes much more complicated since ovarian cancer has to be differentiated from other conditions such as inflammation, cyst formation and benign lesions (Levenson 2010), and because obtaining a biopsy is a necessarily highly invasive procedure. The inefficiency of diagnosing cancer early, prior to metastatic spread, is one of the leading causes of mortality from cancer. DNA methylation cancer biomarkers have been identified in a number of studies in DNA extracted from blood samples (as circulating tumour DNA), or directly from tumour samples, and either approach has its own benefits and challenges. For example, analysis of 27,000 CpG sites in peripheral blood DNA from 148 healthy

individuals and 113 age-matched pretreatment ovarian cancers identified specific patterns of DNA methylation in cancer cases (Teschendorff et al. 2009). This included 100 differentially methylated CpG sites that could be used as potential predictive markers of ovarian cancer. These 100 CpG sites were then validated in a blind test cohort of 58 healthy controls and 43 cancer cases using ROC curves. The predictive CpG sites identified diseased cases from healthy control cases with an 'area under the curve' value of 0.82 and a 95% confidence interval of 0.8. Hence, this study provided evidence for the use of DNA methylation profiling as a predictive marker of cancer diagnosis (Teschendorff et al. 2009). The fact that the differential methylation was identified in peripheral blood cells makes it an attractive option for further development as a convenient blood-based indicator to identify individuals requiring further screening for ovarian cancer. In another such example, methylation of the *SEPT9* gene has been developed as a tool to assist in the diagnosis of colorectal cancer (Church et al. 2010). Nearly 8000 asymptomatic patients from 32 clinical sites in the United States and Germany undergoing routine screening colonoscopy were recruited for a large-scale study. Blood samples from participants were used to analyse *SEPT9* methylation and compared to colonoscopy screening for colorectal cancer detection. This study concluded the use of the *SEPT9* test was able to detect 67% of colorectal cancers with an 11% false-positive rate (Church et al. 2010). Recently, further development of the *SEPT9* test has increased overall sensitivity of colorectal cancer detection to 90 with 88% specificity and a 12% false-positive rate (Warren et al. 2011). These results indicate that the *SEPT9* test performs with a better sensitivity and specificity than the widely used faecal occult blood test and could therefore triage patients for further examination, including colonoscopy, which would significantly reduce the need for this expensive procedure.

Growing evidence suggests that DNA methylation profiling has the potential to become an integral part of cancer diagnosis as a biomarker. The advent of large consortiums containing genome-wide methylation analysis of a plethora of tumour types has made investigation into finding alternative tools, such as the *SEPT9* test, a real possibility.

7.3.2 Cancer Stratification and Treatment Implications

DNA methylation profiles have also proven useful in predicting outcomes of cancer patients by effectively identifying cancer subtypes at diagnosis or during treatment that stratify with survival. An example of the identification of tumour subtypes is in glioblastoma, where genome-wide promoter DNA methylation profiling of 272 tumours from TCGA revealed a subset of CpG island that present with hypermethylation, termed the glioma-CpG island methylator phenotype (gCIMP) (Noushmehr et al. 2010). gCIMP DNA methylation profiles were found to be more predominant (30%) in the proneural subgroup of low-grade gliomas and were associated with a distinct pattern of gene expression, as well as strong correlation with *IDH1* mutations that affected amino acid 132 of *IDH1* (Noushmehr et al. 2010; Yan et al. 2009). Subsequently, the same *IDH1* mutation was experimentally shown

to constitute the molecular basis for CIMP phenotype in glioma (Turcan et al. 2012). Hence, this study showed the significant utility of DNA methylation profiling in stratifying distinct subtypes of cancer.

As altered DNA methylation states can impact gene expression, it is intuitive that DNA methylation patterns could be predictive of response to specific drugs and drug combinations. An example of a well-established marker predicting response to treatment is hypermethylation of the *MGMT* promoter of patients with glioblastoma multiforme. Alkylating agents such as temozolomide with concurrent radiotherapy is often the first-line therapy for glioblastoma. Such alkylating agents cause DNA damage by the transfer of alkyl groups at several sites within DNA, including the O^6 position of guanine. MGMT is normally responsible for maintaining genomic stability via the recruitment of DNA repair enzymes that remove alkylating agents from DNA (Dunn et al. 2009). Promoter hypermethylation and decreased gene expression of *MGMT* compromises the capacity of cells to repair this type of DNA damage and correlates with overall increased survival of glioblastoma patients treated with alkylating agents. In one study, *MGMT* promoter hypermethylation, found in 45% of the 206 glioblastoma cases studied, was found to be an independent, favourable prognostic factor irrespective of treatment. Among the patients that contained *MGMT* promoter hypermethylation, a median survival of 21.7 months was achieved via temozolomide and radiotherapy treatment compared to 15.3 months for those who were treated with radiotherapy alone. No significant survival benefit was observed between the two treatment strategies (temozolomide and radiotherapy vs radiotherapy alone) in the group of patients that did not have *MGMT* promoter hypermethylation. Hence, it was concluded that *MGMT* promoter hypermethylation can subtype glioblastoma patients into groups that may benefit with adjuvant temozolomide treatment (Hegi et al. 2005). In another study of 109 glioblastoma cases, *MGMT* methylation was associated with a median overall survival of 12.4 months with a 2-year survival of 17.9%, compared to 11.1 months and 0% 2 year survival in patients with unmethylated *MGMT* (Dunn et al. 2009). Additional analysis from this study revealed the degree of promoter methylation (>35%) correlated with the longest survival (overall survival was 26.2 months, 2-year survival of 59.7%) (Dunn et al. 2009). Hence, this study also concluded *MGMT* promoter methylation status was an independent prognostic variable for overall survival. Methylations of other genes that have been linked with a predictive potential have been identified in various other cancer contexts. For example, *BRCA1* promoter methylation has been linked with increased chemo-sensitivity towards platinum-based therapy in epithelial ovarian cancer (Chaudhry et al. 2009).

DNA methylation profiles may have the potential to be used as an assessment of drug response and drug resistance, as well as screening for suitability of selected drug treatments. Although further work is needed to assess whether DNA methylation profiling could be incorporated into clinical decision-making, several early studies have provided a promising start (Jones et al. 2016; Levenson 2010).

7.4 Future Perspectives of DNA Methylation in Cancer Treatment

Cumulative evidence has now made it apparent that a variety of epigenetic molecular pathways are involved in the pathogenesis of cancer. Current work on developing targeted approaches for manipulating the epigenome has opened exciting new potential trajectories for the future utility of epigenetic therapy in the clinical setting (Fig. 7.2). Genome editing using technologies such as clustered regularly interspaced short palindromic repeats (CRISPR), as well as transcription activator-like effector nucleases (TALEs), are now emerging tools for DNA methylation manipulation. The CRISPR-Cas9 genome-editing system has been discussed with much enthusiasm in the epigenetic community. There are two main components to this system, the targeting domain, which is the Cas9 endonuclease, and a functional domain, which is the catalytic domain. CRISPR-Cas9 is a versatile tool for recognising specific genomic sequences and inducing double-strand breaks. These double-strand breaks are then repaired via endogenous DNA repair mechanisms. This genome-editing system has recently been repurposed for use in manipulating DNA methylation.

In one of the first attempts of utilising genome editing for DNA methylation, an in vitro study fused the Cas9 targeting domain to the catalytic domain of DNMT3A (Vojta et al. 2016). The group was able to show successful targeting of DNA methylation to regions of DNA of 35 bp in size. Use of multiple guide RNAs concomitant with the Cas9-DNMT3A fusion allowed for targeted DNA methylation of wider regions. This was demonstrated for the *IL6ST* and *BACH2* promoters in human kidney cells where the promoters are known to be unmethylated. Targeted methylation was confirmed via pyrosequencing assays showing the Cas9-DNMT3A fusion was able to methylate the promoters to a sufficient degree (between 25 and 55% for the selected CpGs) to reduce expression of the target loci. It was also demonstrated that this targeted methylation alteration was heritable across mitotic divisions (Vojta et al. 2016). Other independent studies have also demonstrated similar findings in vivo. For example, recently the technique was successfully utilised in vivo with a Cas9-TET1 fusion, which elicited targeted demethylation of the *Bdnf* promoter to a sufficient degree to activate *Bdnf* expression in mice neurons (Liu et al. 2016).

Similarly, the TALE system has been adopted for epigenome editing (Yamazaki et al. 2017). In one study, the TALE recognition sequence, which targeted repeated sequences in pericentromeres, was successfully fused with the bacterial CpG methyltransferase (MSssI). ChIP-qPCR assays demonstrated successful hypermethylation of the repeat sequences in pericentromeres, which were unmethylated prior to the TALE-MSssI fusion. This study was able to demonstrate a new way of targeting methylation without the aid of other binding partners, as is the case for CRISPR-Cas9 (Yamazaki et al. 2017).

At present, there is much excitement regarding genome-editing tools and their potential application for alteration of the epigenetic landscape. Whilst DNA inhibitors have made it to phase II clinical trials, with some gaining FDA approval,

these strategies will likely be limited by their lack of specificity. Although at the very early stages of research, theoretically, tools such as the CRISPR-Cas9 and TALE systems would allow for targeted alterations of region-specific DNA methylation at single dinucleotide resolution. If adequately developed further, the ability to switch the transcriptional status of key genes using genome-editing tools may very well be an important part of future epigenetic therapy and holds promise for vast improvements regarding tailored therapeutic outcomes for cancer.

References

Alexandrov LB, Nik-Zainal S, Wedge DC, Aparicio SA, Behjati S, Biankin AV, Bignell GR, Bolli N, Borg A, Børresen-Dale AL, Boyault S, Burkhardt B, Butler AP, Caldas C, Davies HR, Desmedt C, Eils R, Eyfjörd JE, Foekens JA, Greaves M, Hosoda F, Hutter B, Ilicic T, Imbeaud S, Imielinski M, Imielinsk M, Jäger N, Jones DT, Jones D, Knappskog S, Kool M, Lakhani SR, López-Otín C, Martin S, Munshi NC, Nakamura H, Northcott PA, Pajic M, Papaemmanuil E, Paradiso A, Pearson JV, Puente XS, Raine K, Ramakrishna M, Richardson AL, Richter J, Rosenstiel P, Schlesner M, Schumacher TN, Span PN, Teague JW, Totoki Y, Tutt AN, Valdés-Mas R, van Buuren MM, van 't Veer L, Vincent-Salomon A, Waddell N, Yates LR, Zucman-Rossi J, Futreal PA, McDermott U, Lichter P, Meyerson M, Grimmond SM, Siebert R, Campo E, Shibata T, Pfister SM, Campbell PJ, Stratton MR, Initiative APCG, Consortium IBC, Consortium IM-S, PedBrain I (2013) Signatures of mutational processes in human cancer. Nature 500(7463):415–421

Baylin SB (2005) DNA methylation and gene silencing in cancer. Nat Clin Pract Oncol 2:S4–S11

Baylin SB, Jones PA (2016) Epigenetic determinants of cancer. Cold Spring Harb Perspect Biol 8 (9):a019505

Bedford MT, van Helden PD (1987) Hypomethylation of DNA in pathological conditions of the human prostate. Cancer Res 47(20):5274–5276

Blum W, Klisovic RB, Hackanson B, Liu Z, Liu S, Devine H, Vukosavljevic T, Huynh L, Lozanski G, Kefauver C, Plass C, Devine SM, Heerema NA, Murgo A, Chan KK, Grever MR, Byrd JC, Marcucci G (2007) Phase I study of decitabine alone or in combination with valproic acid in acute myeloid leukemia. J Clin Oncol 25(25):3884–3891

Cadieux B, Ching TT, VandenBerg SR, Costello JF (2006) Genome-wide hypomethylation in human glioblastomas associated with specific copy number alteration, methylenetetrahydrofolate reductase allele status, and increased proliferation. Cancer Res 66 (17):8469–8476

Catteau A, Morris JR (2002) BRCA1 methylation: a significant role in tumour development? Semin Cancer Biol 12(5):359–371

Chaudhry P, Srinivasan R, Patel FD (2009) Utility of gene promoter methylation in prediction of response to platinum-based chemotherapy in epithelial ovarian cancer (EOC). Cancer Invest 27 (8):877–884

Chen RZ, Pettersson U, Beard C, Jackson-Grusby L, Jaenisch R (1998) DNA hypomethylation leads to elevated mutation rates. Nature 395(6697):89–93

Chiappinelli KB, Strissel PL, Desrichard A, Li H, Henke C, Akman B, Hein A, Rote NS, Cope LM, Snyder A, Makarov V, Budhu S, Slamon DJ, Wolchok JD, Pardoll DM, Beckmann MW, Zahnow CA, Merghoub T, Chan TA, Baylin SB, Strick R (2015) Inhibiting DNA methylation causes an interferon response in cancer via dsRNA including endogenous retroviruses. Cell 162 (5):974–986

Christman JK (2002) 5-Azacytidine and 5-aza-2'-deoxycytidine as inhibitors of DNA methylation: mechanistic studies and their implications for cancer therapy. Oncogene 21(35):5483–5495

Church TR, Wandell M, Lofton-Day C, Mongin S, Bluminstein BA, Allen J, Roesch T, Snover D, Day R, Ransohof DA (2010) Prospective clinical validation of an assay for methylated SEPT9 DNA in human plasma as a colorectal cancer screening tool in average risk men and women $\geq$50 years. Gastroenterology 139(1):e18

Denissenko MF, Pao A, Tang M, Pfeifer GP (1996) Preferential formation of benzo[a]pyrene adducts at lung cancer mutational hotspots in P53. Science 274(5286):430–432

Dunn J, Baborie A, Alam F, Joyce K, Moxham M, Sibson R, Crooks D, Husband D, Shenoy A, Brodbelt A, Wong H, Liloglou T, Haylock B, Walker C (2009) Extent of MGMT promoter methylation correlates with outcome in glioblastomas given temozolomide and radiotherapy. Br J Cancer 101(1):124–131

Eden A, Gaudet F, Waghmare A, Jaenisch R (2003) Chromosomal instability and tumors promoted by DNA hypomethylation. Science 300(5618):455

Egger G, Liang G, Aparicio A, Jones PA (2004) Epigenetics in human disease and prospects for epigenetic therapy. Nature 429(6990):457–463

Ehrlich M (2009) DNA hypomethylation in cancer cells. Epigenomics 1(2):239–259

Esteller M, Corn PG, Baylin SB, Herman JG (2001) A gene hypermethylation profile of human cancer. Cancer Res 61(8):3225–3229

Florl AR, Steinhoff C, Müller M, Seifert HH, Hader C, Engers R, Ackermann R, Schulz WA (2004) Coordinate hypermethylation at specific genes in prostate carcinoma precedes LINE-1 hypomethylation. Br J Cancer 91(5):985–994

Futscher B (2012) 5-aza-2'deoxycytidine. In: Schwab M (ed) Encyclopedia of Cancer. Springer, Berlin Heidelberg, pp 325–328

Gao XD, Qu JH, Chang XJ, Lu YY, Bai WL, Wang H, Xu ZX, An LJ, Wang CP, Zeng Z, Yang YP (2014) Hypomethylation of long interspersed nuclear element-1 promoter is associated with poor outcomes for curative resected hepatocellular carcinoma. Liver Int 34(1):136–146

Gaudet F, Hodgson JG, Eden A, Jackson-Grusby L, Dausman J, Gray JW, Leonhardt H, Jaenisch R (2003) Induction of tumors in mice by genomic hypomethylation. Science 300(5618):489–492

Greenblatt MS, Bennett WP, Hollstein M, Harris CC (1994) Mutations in the p53 tumor suppressor gene: clues to cancer etiology and molecular pathogenesis. Cancer Res 54(18):4855–4878

Hegi ME, Diserens AC, Gorlia T, Hamou MF, de Tribolet N, Weller M, Kros JM, Hainfellner JA, Mason W, Mariani L, Bromberg JE, Hau P, Mirimanoff RO, Cairncross JG, Janzer RC, Stupp R (2005) MGMT gene silencing and benefit from temozolomide in glioblastoma. N Engl J Med 352(10):997–1003

Hesson LB, Cooper WN, Latif F (2007) The role of RASSF1A methylation in cancer. Dis Markers 23(1-2):73–87

Issa JP, Kantarjian HM, Kirkpatrick P (2005) Azacitidine. Nat Rev Drug Discov 4(4):275–276

Jones PA, Issa JP, Baylin S (2016) Targeting the cancer epigenome for therapy. Nat Rev Genet 17 (10):630–641

Juergens RA, Wrangle J, Vendetti FP, Murphy SC, Zhao M, Coleman B, Sebree R, Rodgers K, Hooker CM, Franco N, Lee B, Tsai S, Delgado IE, Rudek MA, Belinsky SA, Herman JG, Baylin SB, Brock MV, Rudin CM (2011) Combination epigenetic therapy has efficacy in patients with refractory advanced non-small cell lung cancer. Cancer Discov 1(7):598–607

Kaneda A, Tsukamoto T, Takamura-Enya T, Watanabe N, Kaminishi M, Sugimura T, Tatematsu M, Toshikazu U (2004) Frequent hypomethylation in multiple promoter CpG islands is associated with global hypomethylation, but not with frequent promoter hypermethylation. Cancer Sci 95(1):58–64

Kantarjian H, Oki Y, Garcia-Manero G, Huang X, O'Brien S, Cortes J, Faderl S, Bueso-Ramos C, Ravandi F, Estrov Z, Ferrajoli A, Wierda W, Shan J, Davis J, Giles F, Saba HI, Issa JP (2007) Results of a randomized study of 3 schedules of low-dose decitabine in higher-risk myelodysplastic syndrome and chronic myelomonocytic leukemia. Blood 109(1):52–57

Kanwal R, Gupta S (2010) Epigenetics and cancer. J Appl Physiol (1985) 109(2):598–605

Karahoca M, Momparler RL (2013) Pharmacokinetic and pharmacodynamic analysis of 5-aza-2'--deoxycytidine (decitabine) in the design of its dose-schedule for cancer therapy. Clin Epigenetics 5(1):3

Katzenellenbogen RA, Baylin SB, Herman JG (1999) Hypermethylation of the DAP-kinase CpG island is a common alteration in B-cell malignancies. Blood 93(12):4347–4353

Khamas A, Ishikawa T, Mogushi K, Iida S, Ishiguro M, Tanaka H, Uetake H, Sugihara K (2012) Genome-wide screening for methylation-silenced genes in colorectal cancer. Int J Oncol 41 (2):490–496

Kim YI, Giuliano A, Hatch KD, Schneider A, Nour MA, Dallal GE, Selhub J, Mason JB (1994) Global DNA hypomethylation increases progressively in cervical dysplasia and carcinoma. Cancer 74(3):893–899

Levenson VV (2010) DNA methylation as a universal biomarker. Expert Rev Mol Diagn 10 (4):481–488

Levenson VV, Melnikov AA (2012) DNA methylation as clinically useful biomarkers-light at the end of the tunnel. Pharmaceuticals (Basel) 5(1):94–113

Lin CH, Hsieh SY, Sheen IS, Lee WC, Chen TC, Shyu WC, Liaw YF (2001) Genome-wide hypomethylation in hepatocellular carcinogenesis. Cancer Res 61(10):4238–4243

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.e17

Lustberg MB, Ramaswamy B (2011) Epigenetic therapy in breast cancer. Curr Breast Cancer Rep 3 (1):34–43

Matei D, Fang F, Shen C, Schilder J, Arnold A, Zeng Y, Berry WA, Huang T, Nephew KP (2012) Epigenetic resensitization to platinum in ovarian cancer. Cancer Res 72(9):2197–2205

Mossman D, Kim KT, Scott RJ (2010) Demethylation by 5-aza-2'-deoxycytidine in colorectal cancer cells targets genomic DNA whilst promoter CpG island methylation persists. BMC Cancer 10:366

Nishigaki M, Aoyagi K, Danjoh I, Fukaya M, Yanagihara K, Sakamoto H, Yoshida T, Sasaki H (2005) Discovery of aberrant expression of R-RAS by cancer-linked DNA hypomethylation in gastric cancer using microarrays. Cancer Res 65(6):2115–2124

Noushmehr H, Weisenberger DJ, Diefes K, Phillips HS, Pujara K, Berman BP, Pan F, Pelloski CE, Sulman EP, Bhat KP, Verhaak RG, Hoadley KA, Hayes DN, Perou CM, Schmidt HK, Ding L, Wilson RK, Van Den Berg D, Shen H, Bengtsson H, Neuvial P, Cope LM, Buckley J, Herman JG, Baylin SB, Laird PW, Aldape K, Network CGAR (2010) Identification of a CpG island methylator phenotype that defines a distinct subgroup of glioma. Cancer Cell 17(5):510–522

Pakneshan P, Têtu B, Rabbani SA (2004) Demethylation of urokinase promoter as a prognostic marker in patients with breast carcinoma. Clin Cancer Res 10(9):3035–3041

Park SY, Seo AN, Jung HY, Gwak JM, Jung N, Cho NY, Kang GH (2014) Alu and LINE-1 hypomethylation is associated with HER2 enriched subtype of breast cancer. PLoS One 9(6): e100429

Pluciennik A, Dzantiev L, Iyer RR, Constantin N, Kadyrov FA, Modrich P (2010) PCNA function in the activation and strand direction of MutLα endonuclease in mismatch repair. Proc Natl Acad Sci 107(37):16066–16071

Plumb JA, Strathdee G, Sludden J, Kaye SB, Brown R (2000) Reversal of drug resistance in human tumor xenografts by 2'-deoxy-5-azacytidine-induced demethylation of the hMLH1 gene promoter. Cancer Res 60(21):6039–6044

Plumb JA, Steele N, Finn PW, Brown R (2004) Epigenetic approaches to cancer therapy. Biochem Soc Trans 32(Pt 6):1095–1097

Pulling LC, Vuillemenot BR, Hutt JA, Devereux TR, Belinsky SA (2004) Aberrant promoter hypermethylation of the death-associated protein kinase gene is early and frequent in murine lung tumors induced by cigarette smoke and tobacco carcinogens. Cancer Res 64 (11):3844–3848

Saba HI (2007) Decitabine in the treatment of myelodysplastic syndromes. Ther Clin Risk Manag 3 (5):807–817

Saba HI, Wijermans PW (2005) Decitabine in Myelodysplastic Syndromes. Semin Hematol 42: S23–S31

Santini V, Kantarjian HM, Issa JP (2001) Changes in DNA methylation in neoplasia: pathophysiology and therapeutic implications. Ann Intern Med 134(7):573–586

Sasaki M, Knobbe CB, Munger JC, Lind EF, Brenner D, Brustle A, Harris IS, Holmes R, Wakeham A, Haight J, You-Ten A, Li WY, Schalm S, Su SM, Virtanen C, Reifenberger G, Ohashi PS, Barber DL, Figueroa ME, Melnick A, Zuniga-Pflucker J-C, Mak TW (2012) IDH1 (R132H) mutation increases murine haematopoietic progenitors and alters epigenetics. Nature 488(7413):656–659

Sawan C, Vaissière T, Murr R, Herceg Z (2008) Epigenetic drivers and genetic passengers on the road to cancer. Mutat Res 642(1–2):1–13

Saxonov S, Berg P, Brutlag DL (2006) A genome-wide analysis of CpG dinucleotides in the human genome distinguishes two distinct classes of promoters. Proc Natl Acad Sci U S A 103 (5):1412–1417

Sharma S, Kelly TK, Jones PA (2010) Epigenetics in cancer. Carcinogenesis 31(1):27–36

Sinčić N, Herceg Z (2011) DNA methylation and cancer: ghosts and angels above the genes. Curr Opin Oncol 23(1):69–76

Sorm F, Piskala A, Cihak A, Vesely J (1964) 5-Azacytidine, a new, highly effective cancerostatic. Experientia 20(4):202–203

Strick R, Strissel PL, Baylin SB, Chiappinelli KB (2015) Unraveling the molecular pathways of DNA-methylation inhibitors: human endogenous retroviruses induce the innate immune response in tumors. Oncoimmunology 5(5):e1122160

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(12):e8274

Turcan S, Rohle D, Goenka A, Walsh LA, Fang F, Yilmaz E, Campos C, Fabius AW, Lu C, Ward PS, Thompson CB, Kaufman A, Guryanova O, Levine R, Heguy A, Viale A, Morris LG, Huse JT, Mellinghoff IK, Chan TA (2012) IDH1 mutation is sufficient to establish the glioma hypermethylator phenotype. Nature 483(7390):479–483

Vaissière T, Sawan C, Herceg Z (2008) Epigenetic interplay between histone modifications and DNA methylation in gene silencing. Mutat Res 659(1–2):40–48

Vojta A, Dobrinić P, Tadić V, Bočkor L, Korać P, Julg B, Klasić M, Zoldoš V (2016) Repurposing the CRISPR-Cas9 system for targeted DNA methylation. Nucleic Acids Res 44(12):5615–5628

Wang YQ, Zhang JR, Li SD, He YY, Yang YX, Liu XL, Wan XP (2010) Aberrant methylation of breast and ovarian cancer susceptibility gene 1 in chemosensitive human ovarian cancer cells does not involve the phosphatidylinositol 3'-kinase-Akt pathway. Cancer Sci 101 (7):1618–1623

Warren JD, Xiong W, Bunker AM, Vaughn CP, Furtado LV, Roberts WL, Fang JC, Samowitz WS, Heichman KA (2011) Septin 9 methylated DNA is a sensitive and specific blood test for colorectal cancer. BMC Med 9:133

Weber M, Hellmann I, Stadler MB, Ramos L, Pääbo S, Rebhan M, Schübeler D (2007) Distribution, silencing potential and evolutionary impact of promoter DNA methylation in the human genome. Nat Genet 39(4):457–466

Wilson AS, Power BE, Molloy PL (2007) DNA hypomethylation and human diseases. Biochim Biophys Acta 1775(1):138–162

Xiao-Jie L, Hui-Ying X, Qi X, Jiang X, Shi-Jie M (2016) LINE-1 in cancer: multifaceted functions and potential clinical implications. Genet Med 18(5):431–439

Yamazaki T, Hatano Y, Handa T, Kato S, Hoida K, Yamamura R, Fukuyama T, Uematsu T, Kobayashi N, Kimura H, Yamagata K (2017) Targeted DNA methylation in pericentromeres with genome editing-based artificial DNA methyltransferase. PLoS One 12(5):e0177764

Yan H, Parsons DW, Jin G, McLendon R, Rasheed BA, Yuan W, Kos I, Batinic-Haberle I, Jones S, Riggins GJ, Friedman H, Friedman A, Reardon D, Herndon J, Kinzler KW, Velculescu VE,

Vogelstein B, Bigner DD (2009) IDH1 and IDH2 mutations in gliomas. N Engl J Med 360 (8):765–773

Yan W, Herman JG, Guo M (2016) Epigenome-based personalized medicine in human cancer. Epigenomics 8(1):119–133

Yoshida K, Miki Y (2004) Role of BRCA1 and BRCA2 as regulators of DNA repair, transcription, and cell cycle in response to DNA damage. Cancer Sci 95(11):866–871

Zhou D, Tang W, Su G, Cai M, An HX, Zhang Y (2017) PCDH18 is frequently inactivated by promoter methylation in colorectal cancer. Sci Rep 7(1):2819

Dysregulation of *Cis*-Regulatory Elements in Cancer

8

Jayne A. Barbour and Jason W. H. Wong

Abstract

Cis-regulatory elements (CREs) are DNA sequences in the genome that regulate gene expression through their interaction with transcription factors and the transcription pre-initiation complex. These elements control the expression of genes that define the identity and function of each individual cell. Precisely coordinated changes in the *cis*-regulation of gene expression are now known to play a crucial role in normal organismal development. Changes in *cis*-regulation have now also been implicated in many human diseases, particularly in cancer. The aim of this chapter is to highlight the clinical potential of recent research that has identified specific roles of the dysregulation of CREs in cancer. This chapter will begin by giving an overview of the function of key CREs while providing examples of how dysregulation of these elements can lead to cancer development. As somatic mutations are a hallmark of cancers, we will focus on the role of somatic changes in genomic DNA that lead to alterations in the control of expression in key oncogenes. Finally, this chapter will highlight some potential clinical utility of recent research in the field and emerging therapies that can be used to target dysregulation in CREs.

J. A. Barbour · J. W. H. Wong (✉)
Prince of Wales Clinical School, UNSW, Sydney, NSW, Australia

Lowy Cancer Research Centre, UNSW, Sydney, NSW, Australia

School of Biomedical Sciences, Li Ka Shing Faculty of Medicine, The University of Hong Kong, Pok Fu Lam, Hong Kong
e-mail: jwhwong@hku.hk

L. B. Hesson, A. L. Pritchard (eds.), *Clinical Epigenetics*,
https://doi.org/10.1007/978-981-13-8958-0_8

Keywords
Cis-regulatory elements · Cancer · Somatic mutations · Promoter · Enhancer · Insulator

8.1 Introduction

Cancer is a heterogeneous group of diseases characterised by loss of cellular growth control and immortality. An important feature of cancer is genomic instability, including high levels of somatic alterations to the DNA code (Vogelstein et al. 2013). The reduction in cost of next-generation sequencing (NGS), which enables rapid sequencing of the genome, has led to the improved feasibility of undertaking large-scale whole genome sequencing (WGS) studies of cancer genomes. Global sequencing initiatives led by the International Cancer Genome Consortium (ICGC) and The Cancer Genomic Atlas (TCGA) were undertaken in recent years to describe the landscape of somatic mutations in cancer (Hudson et al. 2010). The resulting sequence data for 10,952 exomes and 1048 whole genomes across 40 different cancer types (as of December 2017) has been curated and stored in the Catalogue of Somatic Mutations in Cancer (COSMIC) database (Forbes et al. 2017). This has generated a wealth of information of the mutations in cancer, although analysis was mostly confined to protein coding regions, meaning that the characterisation of translocations or non-coding mutations is less complete.

Meanwhile, results of the Encyclopedia of DNA Elements (ENCODE) project in 2012 revealed that despite only 1% of the genome being protein coding, up to 80% of the genome is functional (The ENCODE Project Consortium 2012). While these data are likely to be inflated due to the methods used, these findings highlight the importance of previously unexplored non-coding regions of the genome, particularly in complex regulation of gene expression, including epigenetic changes. Functionality of non-coding regulatory elements and epigenetic regulation can be measured by a range of experimental techniques coupled with NGS presented in Table 8.1. These provide insight into epigenetic mechanisms such as transcription factor binding (measured by chromatin immunoprecipitation (ChIP) coupled with sequencing), DNA methylation (assessed by bisulphite sequencing) and DNA accessibility (assessed by techniques that generate DNA fragments at accessible regions for sequencing and three-dimensional DNA conformation through cross-linking and sequencing DNA to map DNA-DNA interactions). Ultimately, functionality of regulatory elements occurs when these elements interact with *trans* factors or other regulatory elements (Maston et al. 2006). These CREs include promoters, enhancers, insulators, silencers and transcription factor binding sites (Maston et al. 2006).

CREs and the epigenome are known to be dysregulated in cancer (discussed in the next section) and, interestingly, somatic mutations form hotspots at CREs in some cancers (Perera et al. 2016; Sabarinathan et al. 2016). In 2013, the first recurrent somatic mutation in a CRE in cancer was discovered (Horn et al. 2013;

Table 8.1 Description of sequencing techniques to characterise functional epigenome and functional regulatory elements

Measurement	Method	Description
DNA-protein interactions	ChIP-seq	DNA is cross-linked with protein by formaldehyde and nuclei are isolated and fragmented. Immunoprecipitation (IP) is performed and fragmented DNA is sequenced (Barski et al. 2007; Johnson et al. 2007)
	ChIP-exo	As above but exonuclease digestion is included after IP and ligation to digest excess sequence up to the antibody to improve resolution (Rhee and Pugh 2011)
DNA methylation	Bisulphite sequencing	DNA is treated with bisulphite which converts cytosines to uracil and hence unmethylated cytosines are read as uracil. Methylated cytosines are unable to be converted and hence read as a cytosine (Clark et al. 1994)
	Methyl probe capture seq	Targeted capture of regions of known CpG sites is performed on bisulphite-converted DNA which is less expensive than genome-wide bisulphite sequencing (Ivanov et al. 2013)
DNA accessibility or nucleosome occupancy	DNase-seq	Formaldehyde is used to cross-link proteins to DNA and DNase I enzyme digestion is performed. Only accessible regions are able to be digested by the enzyme as inaccessible regions are protected from digestion by chromatin and sequenced (Hesselberth et al. 2009)
	FAIRE-seq	Formaldehyde cross-linked protein-DNA is sonicated and phenol-chloroform DNA extraction is performed and genomic regions immediately upstream of genes will be enriched in the aqueous phase (Giresi and Lieb 2009)
	MNase-seq	Micrococcal nuclease digestion of native chromatin followed by sequencing reveals genomic loci of high nucleosome occupancy as nucleosome-associated DNA is insensitive to MNase (Barski et al. 2007)
	ATAC-seq	Hyperactive transposase enzyme Tn5 cleaves exposed DNA of native chromatin leaving open chromatin for sequencing (Buenrostro et al. 2013)
	NOMe-seq	Nucleosome occupancy and methylation are simultaneously assessed. Native chromatin is subject to GpC methyltransferase M.CviPI treatment which can methylate GpC dinucleotides that are not bound by nucleosomes. Subsequent bisulphite conversion and sequencing are performed to characterise nucleosome-free CpG methylation (Kelly et al. 2012)
DNA-DNA interaction or chromosome conformation	3C	Assesses one specific DNA-DNA interaction DNA is cross-linked with protein and DNA is digested using a restriction enzyme and DNA

(continued)

Table 8.1 (continued)

Measurement	Method	Description
		fragments that are proximal are ligated together (chromatin proximity ligation). Primers are designed for two loci of interest that interact directed towards the ligated restriction site (Dekker et al. 2002)
	4C	Assesses all potential interactions with one DNA locus of interest 3C template described above is subject to another round of restriction digest and ligation. Interactions of a region of interest are characterised by inverse PCR and sequencing (Simonis et al. 2006; Zhao et al. 2006)
	5C Many-to-many	Assesses many DNA interactions Chromatin proximity ligation is performed and then universal primers are used for sequencing many interactions (Dostie et al. 2006)
	Hi-C All-to-all	Assesses all DNA interactions Chromatin proximity ligation is performed followed by marking of ends with biotin so that the fragments can be pulled down and sequenced capturing all interactions (Lieberman-Aiden et al. 2009)
	ChIA-PET	Assesses all DNA interactions associated bound by a specific *trans* factor ChIP is coupled with chromatin proximity ligation and paired end sequencing (Fullwood and Ruan 2009)

Huang et al. 2013). There has since been significant research interest into characterising cancer-causing mutations in previously unexplored non-coding regions of the genome. This literature, as well as the potential for these alterations to be used clinically for diagnosis, prognosis and personalised medicine, will be discussed in this chapter.

8.2 Function of CREs and Dysregulation in Cancer

Every cell in the body has the same genetic code despite vast phenotypic variability, highlighting the exquisite control of gene expression regulation. Since oncogenic transformation requires a shift in cellular identity, it is not surprising that perturbed regulation of gene expression underpins many aspects of cancer pathology (Hanahan and Weinberg 2011). Notably, genes that are required for cell growth are important in development but are usually silenced in mature adult cells; if activated in mature adult cells, these genes may drive cancer development and are known as ‘oncogenes’ (Hanahan and Weinberg 2011; Lee and Muller 2010). Conversely, tumour

suppressor genes usually function to control progression through the cell cycle in adult cells, but can become aberrantly silenced, causing cancer (Hanahan and Weinberg 2011; Lee and Muller 2010). Gene expression occurs when RNA polymerases are recruited to transcriptional start sites (TSS) enabling RNA synthesis from the DNA template (Ptashne and Gann 1997). The underlying higher-order regulation of this process is achieved through binding of DNA binding proteins, notably transcription factors (or *trans* factors) to CREs (Maston et al. 2006). CREs are functional non-coding DNA sequences including promoters, enhancers, super-enhancers, silencers and insulators (Maston et al. 2006). Regulation of binding to CREs occurs at the element itself by processes that include reduced binding affinity of *trans* factors through DNA methylation (Yin et al. 2017) and through accessibility to that element by chromatin states (Felsenfeld et al. 1996; Voss and Hager 2014). Here we will now give some examples of how CREs may be involved in cancer.

8.2.1 Accessibility to CREs

As binding of *trans* factors to CREs is required for regulation of gene expression, change to their accessibility is a form of gene expression regulation (Felsenfeld et al. 1996; Voss and Hager 2014). The basic unit of chromatin is the nucleosome, which consists of an octamer of histone proteins around which approximately 146 bp of DNA is wrapped (Kornberg and Lorch 1999). One major way by which accessibility to CREs is regulated is the positioning of nucleosomes relative to those sequences. In cancer, nucleosome depletion occurs in CREs including promoters, enhancers and insulators (Taberlay et al. 2014). Another way accessibility to CREs is controlled is chromatin density, which is regulated by post-translational modification of histone proteins (Kornberg and Lorch 1999). Different histone marks can be associated with active and repressed promoter and enhancer elements (The ENCODE Project Consortium 2012; Rada-Iglesias et al. 2011; Shlyueva et al. 2014), and these histone marks are globally dysregulated in cancer (Baylin and Jones 2011). As a result, access to CREs is altered, leading to cancer cell-specific changes in gene expression.

8.2.2 Promoters

A promoter is a non-coding sequence of roughly 100–1000 bases that is usually proximal to the gene that it regulates (Maston et al. 2006). Promoters always consist of a core promoter sequence, but may also contain proximal promoter sequences (Maston et al. 2006; Riethoven 2010). The core promoter is necessary for initiation of transcription and contains the TSS, an RNA polymerase binding site and generic transcription factor binding sites such as a TATA box (Maston et al. 2006; Riethoven 2010). The proximal promoter is nearby and contains additional sequence-specific transcription factor binding sites that fine-tune gene expression. Promoters activate gene expression when transcription factors bind to the core promoter region, attracting RNA polymerases (Ptashne and Gann 1997). Cytosine methylation at

CpG islands in promoters blocks binding of transcription factors, resulting in gene silencing (Yin et al. 2017). Promoter elements are dysregulated in cancer through hypermethylation resulting to silencing of tumour suppressor genes, which is important in cancer development (Baylin 2005).

8.2.3 Enhancers and Super-Enhancers

Enhancers are 50–500-bp non-coding genetic elements that physically interact with promoter regions up to 1 Mb away and regulate their activity (Shlyueva et al. 2014). Like promoters, they can contain multiple sequence-specific transcription factor binding sites but uniquely contain binding sites for RNA polymerase II, p300 and CBP (Shlyueva et al. 2014). Binding of transcription factors to enhancers can also be regulated by cytosine methylation. Additionally, enhancer activity is regulated by histone marks that modulate accessibility to the enhancer, as discussed above. Dysregulation of enhancers that control oncogene or tumour suppressor gene expression can be important in cancer development (Sur and Taipale 2016). Experimental deletion of cancer-associated enhancers has a phenotypic effect on proliferation in several in vitro and in vivo models (Sur and Taipale 2016). Another class of CRE is the super-enhancer. This is a region of DNA characterised by many enhancers that are grouped together (within 12.5 kb) with unusually high levels of activity (Pott and Lieb 2015). Super-enhancers can control transcription of multiple genes hence playing an important role in specific gene expression signatures and cellular identity (Hnisz et al. 2013). This critical role in the control of transcriptional networks implicates super-enhancers in cancer development (Ko 2017).

8.2.4 Silencers and Insulators

Silencers are specific DNA sequences that prevent transcription through binding to transcriptional repressors. Disruption of silencers that control oncogenes is a common occurrence in cancer (Kazanets et al. 2016). For example, expression of the tumour suppressor *RB1* (retinoblastoma) gene is silenced through deregulation of epigenetic silencers Pipsqueak and Lola promoting tumourigenesis in *Drosophila* eye (Ferres-Marco et al. 2006). One of the most ubiquitous subsets of silencers in the human genome contains binding sites for CCCTC-binding factor (CTCF). There are over 50,000 CTCF binding sites in the genome and CTCF binding can be regulated by DNA methylation of CpG dinucleotides within the recognition sequence. Disruption of CTCF silencing function is linked with cancer (Fiorentino and Giordano 2012; Oh et al. 2017).

Additionally, when CTCF binding sites occur in topologically associated domain (TAD) boundaries, they act to form insulators. Insulators are non-coding boundary elements that enable DNA loop formation. These DNA loops can contain many genes that can be co-ordinately regulated to either positively or negatively regulate expression in what is referred to as an ‘insulated neighbourhood’. Unlike silencers

that are specific to a gene, insulated neighbourhoods can enable the silencing or activation of many genes within the loop. Insulators can regulate the expression of genes over long distances including genes involved in the regulation of cellular identity (Dowen et al. 2014). It has been noted in the literature that TAD borders can become disrupted in cancer leading to oncogenesis through loss of insulation of tumour suppressor genes (Valton and Dekker 2016). For example, hypermethylation of CTCF binding sites occurs in IDH mutant gliomas due to oncometabolite-mediated disruption of 5′-methylcytosine hydroxylase TET enzymes (Flavahan et al. 2016). This hypermethylation of CTCF binding sites results in loss of insulation of genes, which can lead to increases in the expression of oncogenes including the receptor tyrosine kinase gene *PDGFRA,* a prominent oncogene involved in the pathogenesis of glioma (Flavahan et al. 2016). Another example of the disruption of an insulated neighbourhood causes altered expression in the *IGF2-H19* locus. This region is imprinted with the paternal allele for *IGF2* and the maternal allele for *H19* being expressed in normal cells. The expression of these genes is linked to the methylation of a CTCF binding site in the *H19* promoter element (Steenman et al. 1994). In a mouse model with a disrupted CTCF site in the *H19* promoter region, there are loss of imprinting of *Igf2* and increased *Igf2* expression from the maternal allele (Damaschke et al. 2017). These mice displayed spontaneous neoplasia suggesting this event is involved in tumour initiation (Damaschke et al. 2017). Further, allele-specific hypermethylation of the canonical CTCF binding site in humans results in loss of imprinting and biallelic expression of *IGF2* in colorectal cancer samples (Nakagawa et al. 2001).

8.3 Acquisition of Somatic Mutations and Structural Variants in Cancer Genomes

A mutation is a permanent alteration in the DNA code of a cell. These can be inherited through the germline, where the mutation is present in every cell of the body. In contrast, somatic mutations are an acquired error in a specific cell. A cell with a somatic mutation is known as a ‘clone’ and can proliferate to generate more cells that harbour the same mutation; this process is known as ‘clonal expansion’. Cancer is characterised by a higher burden of somatic mutations that provide the cell with a proliferative advantage, allowing it to undergo clonal expansion to form a tumour (Vogelstein et al. 2013).

Mutations can be broadly grouped as single nucleotide variants (SNVs) and structural variants including insertions, deletions, copy number variants and larger-scale structural variants such as translocations and inversions (Table 8.2). SNVs are the most frequently occurring mutation where one base is substituted with another and is generally the result of a DNA replication error or defect in DNA repair (Chatterjee and Walker 2017). Insertions and deletions (indels) are the second most common variant and describe mutations whereby small nucleotide regions (typically 2–16 and up to 50 bp; Mullaney et al. 2010) are gained or lost from a genomic region, respectively (Mills et al. 2006). A copy number variant is where

Table 8.2 Features of somatic alterations and their presence in cancer

Alteration	Definition	Proposed causes	Presence in cancer
Single nucleotide variants (SNVs)	One base is substituted with another	DNA replication error DNA repair deficiencies (Chatterjee and Walker 2017)	Observed in all cancer types
Insertions	A small insertion of (2–16 bases) at a site in the genome (Mills et al. 2006)	Polymerase strand slippage (Garcia-Diaz and Kunkel 2006)	Observed in all cancer types but most pronounced in microsatellite instable cancers
Deletions	A small deletion of (2–16 bases) at a site in the genome (Mills et al. 2006)	Polymerase strand slippage (Garcia-Diaz and Kunkel 2006)	As above
Copy number variant	Loss or gains of copies of a gene (Freeman et al. 2006)	Non-allelic homologous recombination and microhomology-mediated end joining (Hastings et al. 2009)	Observed in all cancer types
Chromosomal translocations and inversions	Translocations are rearrangements of non-homologous chromosomes and inversions are segments of chromosomes that are inverted in the same position (Nambiar et al. 2008)	Improperly repaired dsDNA breaks (Currall et al. 2013)	Expected to be found in all cancer types with higher levels observed in some leukaemias, lymphomas and sarcomas (Rowley 2001)

regions of the genome encoding a gene are repeated or lost (Freeman et al. 2006) and is caused by non-allelic homologous recombination and microhomology-mediated end joining (Hastings et al. 2009). Another form of somatic alteration is large structural variants involving large inversions or translocations whereby distal genomic regions are fused (Currall et al. 2013). These are caused by errors in double-strand break repair and can result in profound genomic abnormalities and can result in substantial changes in gene expression (Chiang et al. 2017; Harewood and Fraser 2014). Collectively, these different types of alterations can arise from endogenous processes including errors during DNA replication, spontaneous deamination of bases, intrinsic enzymes and intrinsic DNA breaks or from exogenous damage to DNA from sources such as reactive oxygen species (ROS), ultraviolet light (UV), ionising radiation and alkylating agents (Chatterjee and Walker 2017). These alterations are observed in all cancer types (Table 8.2).

Somatic mutations observed in cancer play a critical role in the development and molecular phenotype of cancer. Individuals with germline mutations in double-strand break repair (Fackenthal and Olopade 2007), mismatch repair (Lynch et al. 2015) and nucleotide excision repair (Qiu et al. 2008) have substantially increased risk of certain cancers suggesting somatic mutations are important in tumour

initiation. These have been extensively studied in coding regions of cancer genomes, and the remainder of the chapter will focus on recent literature highlighting their roles in non-coding regulatory elements.

8.4 Somatic Alterations to *Cis*-Regulatory Regions in Cancer

It is clear from the above sections that dysregulation of CREs and the presence of somatic alterations are both crucial factors in cancer pathology. To date, the majority of research in cancer genomics has focused on coding genes. Results of the ENCODE project coupled with falling costs of NGS have enabled WGS of many tumours and given rise to a significant amount of non-coding somatic mutation data, revealing a role for mutations in CREs.

8.4.1 Large Surveys of Non-coding Cancer Mutations Reveal Frequent Mutations in CREs

The first highly recurrent somatic mutation in a CRE was discovered in the promoter region of the telomerase reverse transcriptase (*TERT*) gene in 2013 (Horn et al. 2013; Huang et al. 2013). One of two different mutations in the *TERT* promoter occurred in 71% of a cohort of melanoma and 16% of cancer cell lines from a panel of 150, especially melanoma, bladder and liver cancer cell lines (Huang et al. 2013). An important finding of this study was that the mutations generated new binding sites for the E26 transformation-specific (ETS) family transcription factor GA-binding protein (GABP) and increased transcription two- to fourfold in *TERT*, which is likely to promote cellular immortality (Huang et al. 2013). A recent study counter-intuitively found that putatively activating *TERT* promoter mutations are associated with slight reduction in telomere length suggesting that there may be a more complex relationship between *TERT* gene expression and cellular immortality (Hayward et al. 2017). Discovery of recurrent, functional *TERT* promoter mutations alerted the research community to the fact that previously unexplored regions of the genome could contain functionally important somatic mutations. Further, it was estimated that of all somatic mutations, less than 0.06% are in coding exons, compared with ~35% in regulatory regions of the genome (Melton et al. 2015); it is an intuitive finding, given that regulatory elements make up a significant portion of the genome (The ENCODE Project Consortium 2012).

This inspired several research groups to perform comprehensive pan-cancer surveys of mutations in non-coding regions of the genome (Fredriksson et al. 2014; Melton et al. 2015; Weinhold et al. 2014). Each of these studies identified TERT, thereby validating their methods, but also found several recurrent *cis*-regulatory mutations proximal to cancer-related genes suggesting that they could be cancer drivers. An interesting observation that came out of these large-scale studies was that non-coding mutations were especially prevalent in melanoma compared with other cancer types (Weinhold et al. 2014). Understanding the differences

between melanoma and other cancers was an opportunity to shed light on the mechanisms of mutagenesis in non-coding mutations. It was found that transcription factor binding blocks DNA nucleotide excision repair causing promoter elements to acquire mutations, and this is most common in cutaneous melanoma due to the dependence on this pathway in ultraviolet radiation-induced DNA damage repair (Perera et al. 2016; Sabarinathan et al. 2016).

However, the functional significance of *cis*-regulatory mutations was challenged by the observation that regulatory mutations were not associated with the expected changes in gene expression with exception of *TERT*; in fact, most somatic mutations did not alter gene expression levels (Fredriksson et al. 2014). The current consensus is therefore that if mutations in non-coding regions do not cause changes in gene expression, they are likely to not be important for driving gene expression (Fredriksson et al. 2014). This sparked careful consideration regarding the biological significance of these mutations and the likelihood that they can drive cancer. The functional impact of mutations in coding regions is usually assessed in vitro or in vivo by whether the mutation affects the function of the protein that it encodes and whether that protein can function to promote cancer development or progression. Coding mutations are classed as non-synonymous or synonymous if they cause or do not cause a change in the amino acid sequence of the protein, respectively. In general, a non-synonymous mutation is more likely to cause a change in protein function when compared with a synonymous mutation, due to the effect of amino acid alteration. It is however less straightforward to predict the functional consequences of the mutations that occur in non-coding regions of the genome. In non-coding regions, the prediction of whether mutations are functionally important is via the assessment of whether they lie in a consensus motif for a specific *trans* factor and can increase or decrease binding affinity of those factors; there must also be a change in function caused by altered binding, such as a change to the expression of a linked gene. In silico analysis of whether a non-coding mutation is recurrent across cohorts of tumours is also indicative that the mutation may play a functional role in driving tumourigenesis. Further, since accessibility to CREs and expression of appropriate *trans* factors are cell-type specific, it is also important that the functionality of that CRE is validated in the tumour by methods such as ChIP-seq or overlap with chromatin features that mark functional CREs. The predicted functional effect then needs to be demonstrated using in vitro models to show if they cause a meaningful change in gene expression of target genes.

An investigation of the biological significance of recurrent somatic mutations in promoter regions using a luciferase reporter assay rather than mRNA levels has found that approximately 20% of mutations affect promoter activity (Poulos et al. 2015). This highlights that somatic mutations in promoter elements beyond *TERT* may indeed be important in the pathogenesis of cancer. Further, a recent study found that more statistical power is required to find somatic driver mutations in CREs (Rheinbay et al. 2017). In this study, a cohort of 360 primary breast cancer samples were sequenced to a high depth of coverage and revealed nine somatic substitutions that were recurrent, three of which were experimentally validated to be functionally important (Rheinbay et al. 2017). Importantly, power calculations indicate that less

than half of recurrent *cis*-regulatory mutations have been found (Rheinbay et al. 2017). Although there are few examples of *cis*-regulatory mutations as striking as TERT, it is highly likely there are lower-frequency *cis*-regulatory mutations that are not discovered yet.

8.4.2 Mutations in Promoter Elements and Transcription Factor Binding Sites that Are Important in Cancer

Promoter regions acquire mutations at an accelerated rate (Perera et al. 2016; Sabarinathan et al. 2016; Smith et al. 2015) so it is important to consider the contribution of these to driving cancer. *TERT* promoter mutations were first identified in melanoma in 2013 (Horn et al. 2013; Huang et al. 2013) and later shown to be recurrent and biologically functional in mantle cell lymphoma (Panero et al. 2016) and hepatocellular carcinoma (Pezzuto et al. 2017). Genomic rearrangements proximal to *TERT* have also been reported in high-risk neuroblastoma (Peifer et al. 2015). As well as being recurrent and functional, *TERT* promoter mutations can also be informative for patient prognosis. *TERT* promoter mutations are associated with poor prognosis in glioblastoma (Simon et al. 2015), glioma (Gao et al. 2016), non-small cell lung cancer (Jung et al. 2017) and renal cancer (Casuscelli et al. 2017). In addition to prognosis, *TERT* promoter mutations are associated with a clinically distinct subgroup of brain cancer suggesting it may even be important in molecular diagnosis (Killela et al. 2014). Interestingly, in bladder cancer, *TERT* promoter mutations are associated with poor prognosis in people without a common germline single-nucleotide polymorphism in the Ets2 binding site of the *TERT* promoter but not in people with it (Rachakonda et al. 2013), which highlights the importance of integrating somatic with germline mutations in the clinic. The fact that *TERT* promoter mutations predict poor prognosis in several distinct tumour types justifies them as a bona fide somatic regulatory 'driver' mutations and importantly means that these mutations could be useful in the clinic.

To date there are few examples of recurrent, somatic promoter mutations that are as clearly clinically significant as *TERT*. While there are reports of promoter mutations in cancer-associated genes that affect expression (Mathelier et al. 2015; Melton et al. 2015), these are not known to be associated with altered prognosis or experimentally shown to have a cancer driver role. A promoter mutation for subunit D of the succinate dehydrogenase complex (*SDHD*) was found to reduce *SDHD* gene expression and was associated with poor prognosis in melanoma and thus potentially important in a clinical setting (Weinhold et al. 2014). More recently, analysis of a large cohort of breast cancers identified three recurrent somatic mutations in regulatory regions that showed functional effects on gene expression (Rheinbay et al. 2017). This included somatic mutations in the promoter of *FOXOA1*, which is an established oncogene in hormone-positive cancer (Rheinbay et al. 2017). These mutations were experimentally demonstrated to reduce the growth of MCF-7 breast cancer cells treated with the anti-oestrogen therapy fulvestrant in vitro (Rheinbay et al. 2017). This suggests that tumours with these

mutations may be resistant to anti-oestrogen therapy posing as an example of how *cis*-regulatory mutations may be clinically important.

8.4.3 Somatic Alterations Can Create and Remodel Enhancers and Super-Enhancers by 'Enhancer Hijacking' to Drive Oncogene Expression

Non-coding somatic alteration of enhancer elements is also a frequent event in cancer. In 2014, it was demonstrated that somatic insertions 2–18 bp created a transcription factor binding site for MYB in a specific location that caused creation of a super-enhancer upstream of the *TAL1* oncogene in a subset of T-ALL cells (Mansour et al. 2014). *TAL1* is an oncogene very important to the pathogenesis of T-ALL and hence this is an example of enhancer 'driver'. This extraordinary example was followed by work assessing insertions at enhancers across 102 cancer cell types which highlighted that insertions in enhancer regions near oncogenes are a frequent event (Abraham et al. 2017). The functional significance of one such example that drives expression of the *LMO2* oncogene is proof of concept that these can have cancer driver roles (Abraham et al. 2017). Therefore, it is possible that there are still more somatic insertions in enhancer regions that are critical cancer drivers.

Larger translocations are also involved in enhancer hijacking and subsequent oncogene expression. In leukaemia, a somatic chromosomal rearrangement repositions the *GATA2* enhancer to be proximal to the cancer driver *EVI1* hence activating its transcription (Groschel et al. 2014). Removal of this repositioned enhancer by genomic editing silenced *EVI1* and reduced colony formation (Groschel et al. 2014) highlighting that this somatic rearrangement may indeed be a cancer driver. The *MYC* locus is frequently rearranged with different alternative loci, and these rearrangements can cause super-enhancer hijacking in multiple myeloma (Affer et al. 2014). More recently, copy number alterations have been shown to cause enhancer hijacking (Weischenfeldt et al. 2017).

The mechanism explaining why enhancers are vulnerable to these alterations is unclear. However, there is evidence that endogenous mutagenic enzymes apolipoprotein B mRNA editing enzyme (APOBEC) and activation-induced cytidine deaminase (AID) target enhancer/super-enhancer regions which may promote hijacking. APOBEC-associated mutations can cause enhancers that drive *LMO1* expression (Li et al. 2017). Many super-enhancers are AID targets, especially in B cells, and thus AID recruitment may provide a mechanism of how DNA breaks are introduced (Qian et al. 2014), since AID is known to induce DNA breaks (Hasham et al. 2012). Interestingly, the Myc enhancer is an AID target of MEFs and B cells suggesting that it may be a common AID target across different cell types (Qian et al. 2014). This may help explain why the Myc super-enhancer region is commonly rearranged.

8.4.4 Somatic Disruption of the Silencer CTCF and Insulators Is a Frequent Event in Cancer

While promoters and enhancers represent a large portion of the literature, there is also evidence for a role of somatic disruption to silencers and insulators in cancer. As discussed earlier, CTCF binding sites are ubiquitous in the mammalian genome and can occur at TAD boundaries to cause DNA loop formation or insulated neighbourhoods and disruption of these can promote oncogene expression (Damaschke et al. 2017; Flavahan et al. 2016; Nakagawa et al. 2001; Valton and Dekker 2016). Interestingly, SNVs and structural variants occur in CTCF binding sites, especially at TAD boundaries (Canela et al. 2017; Hnisz et al. 2016; Katainen et al. 2015; Poulos et al. 2016; Guo et al. 2018). However, whether these somatic alterations can alter the expression of cancer-related genes remains an open question.

A study that sequenced 213 colorectal cancer genomes found SNV hotspots in occupied CTCF binding sites in microsatellite stable tumours with A:T>C:G transitions being a major contributor to these SNVs (Katainen et al. 2015). Other cancer types analysed including oesophageal adenocarcinoma, liver cancer, pancreatic cancer, malignant lymphoma and ovarian cancer also displayed enrichment of mutations within CTCF sites (Katainen et al. 2015). Enrichment of mutation rate in CTCF binding sites was also displayed in an independent publication on oesophageal adenocarcinoma and liver cancer (Hnisz et al. 2016). These CTCF binding site mutation hotspots occur more specifically in the CTCF binding sites of TAD boundaries suggesting possibility of functional disruption of insulators (Hnisz et al. 2016). Differential nucleotide excision repair may explain a proportion of these mutations (Poulos et al. 2016); however the mechanism for A:T>C:G transitions in those hotspots remains unclear. In addition to SNVs, frequent deletions of insulators have been observed in T-cell acute lymphoblastic leukaemia (T-ALL) (Hnisz et al. 2016). A non-malignant T-cell line can undergo malignant transformation and oncogene expression when TAD boundaries are perturbed experimentally by CRISPR/Cas9 editing (Hnisz et al. 2016). This proof of concept provides evidence that alterations in these hotspots are possibly functionally important (Hnisz et al. 2016), albeit in a non-physiological setting. A study interrogating the functional impact of CTCF binding site mutations in melanoma revealed that these mutations reduce CTCF binding (Poulos et al. 2016). The effect that these mutations had on 3D chromatin structure was not assessed; however analysis of chromatin loops harbouring mutations revealed no enrichment for cancer-associated genes when compared with loops without mutations (Poulos et al. 2016). With over 50,000 CTCF binding sites in the human genome, identifying CTCF site mutations that are functionally important remains a major analytical challenge. Recent work has highlighted that TAD boundaries are also vulnerable to structural variation due to topoisomerase 2B (TOP2B)-mediated dsDNA breaks that are made to alleviate topological stress at these sites (Canela et al. 2017). To summarise, CTCF binding sites at TAD borders are frequently mutated in cancer, but more functional evidence is required to conclude whether these are cancer driver events.

8.5 Clinical Uses of Somatic Alterations to CREs in Cancer

Whether somatic mutations and structural variants in CREs have clinical utility is still currently being investigated. While it has long been known that cancer is a heterogeneous disease, the extent of this heterogeneity is now being appreciated with the advent of NGS technologies. It is now known that one cancer type can be subdivided based on distinct molecular signatures. Further, advances in single cell WGS have illustrated the substantial intratumoural heterogeneity that exists (Navin 2015). Molecular characterisation and stratification of cancer with NGS technologies can now be used to help select a therapy targeted towards one or more genetic changes present in the tumour or to predict drug resistance ('precision oncology'; Cummings et al. 2016).

The field of non-coding somatic mutations is less developed than the field of coding mutations, and this is reflected in the clinical progress of non-coding mutations. While coding somatic mutations for known drivers are beginning to be routinely used in clinical oncology, there is currently only one example of a CRE mutation being included on NGS cancer panels and commercially available for testing, the *TERT* promoter mutation. *TERT* promoter mutations have appeared on an NGS gene panel for glioma (Zacher et al. 2017) and a multiplex PCR panel for non-muscle bladder cancer (Ward et al. 2016). There are several examples of recurrent, non-coding regulatory 'drivers' associated with altered prognosis as discussed in the previous section, which could be used in the future.

In addition to diagnostic and prognostic use, *cis*-regulatory somatic mutations may also help guide choice of therapy. There are a number of clinical trials underway investigating the efficacy of *TERT* targeted therapy for tumours overexpressing *TERT*; examples of *TERT* targeted therapies in clinical trials include imetelstat, various immunotherapies and GRNVAC1 (reviewed in Buseman et al. 2012). In leukaemia, tumours with the somatic rearrangement in enhancer region that leads to activation of *EVI1* (described above) are sensitive to inhibitors of the bromodomain and extraterminal protein family (BET) in vitro (Groschel et al. 2014). These drugs prevent BET proteins from interacting with the acetylated histone proteins (Braun and Gardin 2017), and there is one published phase I clinical trial (Berthon et al.2016) and several ongoing phase I and I/II trials investigating the efficacy of these drugs for the treatment of leukaemia (reviewed in Braun and Gardin 2017). Preclinical work suggests that leukaemias harbouring this somatic rearrangement may respond well to BET inhibitors. However, since clinical trials are still at an early stage, it is not known whether BET inhibitors will be a targeted therapy for leukaemias with these specific somatic mutations. As mentioned earlier, *FOXOA1* promoter mutations reduce efficacy of anti-oestrogen therapy in vitro (Rheinbay et al. 2017) suggesting that these mutations may confer resistance to oestrogen receptor therapies. However, there is no clinical evidence to date that suggests that breast cancers with these mutations should be treated differently.

While the above are encouraging examples of drug targeting based on *cis*-regulatory mutations, there are specific challenges with *cis*-regulatory mutations that are not shared with coding mutations. For instance, *cis*-regulatory mutations

may be more difficult to therapeutically target than protein-coding mutations. While promoter mutations that cause overexpression of a single protein may be an appropriate therapeutic target, mutations in enhancers or insulators could be associated with more widespread transcriptional changes that are difficult to target therapeutically. However, if the presence of somatic mutations in some enhancers can be used to predict widespread transcriptional changes, then sequencing of a small handful of enhancers may ultimately prove useful in predicting the transcriptional profile of a tumour without the need for sequencing the entire transcriptome. Another challenge to implementing the testing of *cis*-regulatory somatic mutations in tumours is their relatively lower frequency when compared with coding mutations (Rheinbay et al. 2017), which may have health and economic implications when considering non-coding mutations for gene panels. Overall, further research is required to delineate whether these non-coding mutations are clinically actionable before they become part of routine testing in the clinic.

8.6 Conclusion

Research investigating the role of non-coding mutations in cancer is still in its infancy compared with the heavily explored coding mutations. However, the discovery of recurrent *TERT* promoter mutations in 2013 has provided a precedent for the value of characterising recurrent non-coding somatic mutations in cancer. Somatic mutations in CREs in cancer are more frequent than other non-coding regions of the genome, and while there has been some controversy over the proportion of these mutations causing biologically relevant functional consequences, there is increasing evidence that these non-coding mutations regulate expression of oncogenes and may have clinical utility. The identification of selected somatic mutations in tumours is currently used to inform clinical management and guide therapeutic choice. However, with exception of *TERT* promoter mutations, non-coding mutations are not currently considered clinically actionable, largely because few examples have been identified and functionally characterised. Other reasons may be the relative infrequency with which non-coding mutations occur, lack of knowledge regarding their interactions with other mutations, environmental factors and germline variants. Research that addresses these areas will lead to a deeper understanding of how non-coding mutations in CREs may promote cancer and how they can be used in the clinical management of patients.

References

Abraham BJ et al (2017) Small genomic insertions form enhancers that misregulate oncogenes. Nat Commun 8:14385. https://doi.org/10.1038/ncomms14385

Affer M et al (2014) Promiscuous MYC locus rearrangements hijack enhancers but mostly super-enhancers to dysregulate MYC expression in multiple myeloma. Leukemia 28:1725–1735. https://doi.org/10.1038/leu.2014.70

Barski A et al (2007) High-resolution profiling of histone methylations in the human genome. Cell 129:823–837. https://doi.org/10.1016/j.cell.2007.05.009
Baylin SB (2005) DNA methylation and gene silencing in cancer. Nat Clin Pract Oncol 2(Suppl 1): S4–S11. https://doi.org/10.1038/ncponc0354
Baylin SB, Jones PA (2011) A decade of exploring the cancer epigenome—biological and translational implications. Nat Rev Cancer 11:726–734. https://doi.org/10.1038/nrc3130
Berthon C et al (2016) Bromodomain inhibitor OTX015 in patients with acute leukaemia: a dose-escalation, phase 1 study. Lancet Haematol 3:e186–e195. https://doi.org/10.1016/S2352-3026(15)00247-1
Braun T, Gardin C (2017) Investigational BET bromodomain protein inhibitors in early stage clinical trials for acute myelogenous leukemia (AML). Expert Opin Investig Drugs 26:803–811. https://doi.org/10.1080/13543784.2017.1335711
Buenrostro JD, Giresi PG, Zaba LC, Chang HY, Greenleaf WJ (2013) Transposition of native chromatin for multimodal regulatory analysis and personal epigenomics. Nat Methods 10:1213–1218. https://doi.org/10.1038/nmeth.2688
Buseman CM, Wright WE, Shay JW (2012) Is telomerase a viable target in cancer? Mutat Res 730:90–97. https://doi.org/10.1016/j.mrfmmm.2011.07.006
Canela A et al (2017) Genome organization drives chromosome. Fragility Cell 170:507–521.e518. https://doi.org/10.1016/j.cell.2017.06.034
Casuscelli J et al (2017) Characterization and impact of TERT promoter region mutations on clinical outcome in renal cell carcinoma. Eur Urol Focus S2405-4569(17)30212-2. https://doi.org/10.1016/j.euf.2017.09.008
Chatterjee N, Walker GC (2017) Mechanisms of DNA damage, repair, and mutagenesis. Environ Mol Mutagen 58:235–263. https://doi.org/10.1002/em.22087
Chiang C et al (2017) The impact of structural variation on human gene expression. Nat Genet 49:692–699. https://doi.org/10.1038/ng.3834
Clark SJ, Harrison J, Paul CL, Frommer M (1994) High sensitivity mapping of methylated cytosines. Nucleic Acids Res 22:2990–2997
Cummings CA, Peters E, Lacroix L, Andre F, Lackner MR (2016) The role of next-generation sequencing in enabling personalized oncology therapy. Clin Transl Sci 9:283–292. https://doi.org/10.1111/cts.12429
Currall BB, Chiang C, Talkowski ME, Morton CC (2013) Mechanisms for structural variation in the human genome. Curr Genet Med Rep 1:81–90. https://doi.org/10.1007/s40142-013-0012-8
Damaschke NA et al (2017) Loss of Igf2 gene imprinting in murine prostate promotes widespread neoplastic. Growth Cancer Res 77:5236–5247. https://doi.org/10.1158/0008-5472.can-16-3089
Dekker J et al (2002) Capturing chromosome conformation. Science 295(5558):1306–1311. https://doi.org/10.1126/science.1067799
Dostie J et al (2006) Chromosome conformation capture carbon copy (5C): a massively parallel solution for mapping interactions between genomic elements. Genome Res 16:1299–1309. https://doi.org/10.1101/gr.5571506
Dowen JM et al (2014) Control of cell identity genes occurs in insulated neighborhoods in mammalian chromosomes. Cell 159:374–387. https://doi.org/10.1016/j.cell.2014.09.030
Fackenthal JD, Olopade OI (2007) Breast cancer risk associated with BRCA1 and BRCA2 in diverse populations. Nat Rev Cancer 7:937–948. https://doi.org/10.1038/nrc2054
Felsenfeld G, Boyes J, Chung J, Clark D, Studitsky V (1996) Chromatin structure and gene expression. Proc Natl Acad Sci U S A 93:9384–9388
Ferres-Marco D, Gutierrez-Garcia I, Vallejo DM, Bolivar J, Gutierrez-Avino FJ, Dominguez M (2006) Epigenetic silencers and Notch collaborate to promote malignant tumours by Rb silencing. Nature 439:430–436. https://doi.org/10.1038/nature04376
Fiorentino FP, Giordano A (2012) The tumor suppressor role of CTCF. J Cell Physiol 227:479–492. https://doi.org/10.1002/jcp.22780
Flavahan WA et al (2016) Insulator dysfunction and oncogene activation in IDH mutant gliomas. Nature 529:110–114. https://doi.org/10.1038/nature16490

Forbes SA et al (2017) COSMIC: somatic cancer genetics at high-resolution. Nucleic Acids Res 45: D777–d783. https://doi.org/10.1093/nar/gkw1121

Fredriksson NJ, Ny L, Nilsson JA, Larsson E (2014) Systematic analysis of noncoding somatic mutations and gene expression alterations across 14 tumor types. Nat Genet 46:1258–1263. https://doi.org/10.1038/ng.3141

Freeman JL et al (2006) Copy number variation: new insights in genome diversity. Genome Res 16:949–961. https://doi.org/10.1101/gr.3677206

Fullwood MJ, Ruan Y (2009) ChIP-based methods for the identification of long-range chromatin interactions. J Cell Biochem 107:30–39. https://doi.org/10.1002/jcb.22116

Gao K et al (2016) TERT promoter mutations and long telomere length predict poor survival and radiotherapy resistance in gliomas. Oncotarget 7:8712–8725. https://doi.org/10.18632/oncotarget.6007

Garcia-Diaz M, Kunkel TA (2006) Mechanism of a genetic glissando: structural biology of indel mutations. Trends Biochem Sci 31:206–214. https://doi.org/10.1016/j.tibs.2006.02.004

Giresi PG, Lieb JD (2009) Isolation of active regulatory elements from eukaryotic chromatin using FAIRE (Formaldehyde assisted isolation of regulatory elements). Methods 48:233–239. https://doi.org/10.1016/j.ymeth.2009.03.003

Groschel S et al (2014) A single oncogenic enhancer rearrangement causes concomitant EVI1 and GATA2 deregulation in leukemia. Cell 157:369–381. https://doi.org/10.1016/j.cell.2014.02.019

Guo YA et al (2018) Mutation hotspots at CTCF binding sites coupled to chromosomal instability in gastrointestinal cancers. Nat Commun 9:1520. https://doi.org/10.1038/s41467-018-03828-2

Hanahan D, Weinberg RA (2011) Hallmarks of cancer: the next generation. Cell 144:646–674. https://doi.org/10.1016/j.cell.2011.02.013

Harewood L, Fraser P (2014) The impact of chromosomal rearrangements on regulation of gene expression. Hum Mol Genet 23:R76–R82. https://doi.org/10.1093/hmg/ddu278

Hasham MG et al (2012) AID-initiated off-target DNA breaks are detected and resolved during S-phase. J Immunol 189(5):2374–2382. https://doi.org/10.4049/jimmunol.1200414

Hastings P, Lupski JR, Rosenberg SM, Ira G (2009) Mechanisms of change in gene copy number. Nat Rev Genet 10:551–564. https://doi.org/10.1038/nrg2593

Hayward NK et al (2017) Whole-genome landscapes of major melanoma subtypes. Nature 545 (7653):175–180. https://doi.org/10.1038/nature22071

Hesselberth JR et al (2009) Global mapping of protein-DNA interactions in vivo by digital genomic footprinting. Nat Methods 6:283–289. https://doi.org/10.1038/nmeth.1313

Hnisz D et al (2013) Super-enhancers in the control of cell identity and disease. Cell 155:934–947. https://doi.org/10.1016/j.cell.2013.09.053

Hnisz D et al (2016) Activation of proto-oncogenes by disruption of chromosome neighborhoods. Science 351:1454–1458. https://doi.org/10.1126/science.aad9024

Horn S et al (2013) TERT promoter mutations in familial and sporadic melanoma. Science 339:959–961. https://doi.org/10.1126/science.1230062

Huang FW, Hodis E, Xu MJ, Kryukov GV, Chin L, Garraway LA (2013) Highly recurrent TERT promoter mutations in human melanoma. Science 339:957–959. https://doi.org/10.1126/science.1229259

Hudson TJ et al (2010) International network of cancer genome projects. Nature 464:993–998. https://doi.org/10.1038/nature08987

Ivanov M, Kals M, Kacevska M, Metspalu A, Ingelman-Sundberg M, Milani L (2013) In-solution hybrid capture of bisulfite-converted DNA for targeted bisulfite sequencing of 174 ADME genes. Nucleic Acids Res 41:e72. https://doi.org/10.1093/nar/gks1467

Johnson DS, Mortazavi A, Myers RM, Wold B (2007) Genome-wide mapping of in vivo protein-DNA interactions. Science 316:1497–1502. https://doi.org/10.1126/science.1141319

Jung SJ, Kim DS, Park WJ, Lee H, Choi IJ, Park JY, Lee JH (2017) Mutation of the TERT promoter leads to poor prognosis of patients with non-small cell lung cancer. Oncol Lett 14:1609–1614. https://doi.org/10.3892/ol.2017.6284

Katainen R et al (2015) CTCF/cohesin-binding sites are frequently mutated in cancer. Nat Genet 47:818–821. https://doi.org/10.1038/ng.3335
Kazanets A, Shorstova T, Hilmi K, Marques M, Witcher M (2016) Epigenetic silencing of tumor suppressor genes: paradigms, puzzles, and potential. Biochim Biophys Acta 1865:275–288. https://doi.org/10.1016/j.bbcan.2016.04.001
Kelly TK et al (2012) Genome-wide mapping of nucleosome positioning and DNA methylation within individual DNA molecules. Genome Res 22(12):2497–2506. https://doi.org/10.1101/gr.143008.112
Killela PJ et al (2014) Mutations in IDH1, IDH2, and in the TERT promoter define clinically distinct subgroups of adult malignant gliomas. Oncotarget 5:1515–1525. https://doi.org/10.18632/oncotarget.1765
Ko JY (2017) Functional enhancers as master regulators of tissue-specific gene regulation and cancer development. Mol Cells 40:169–177. https://doi.org/10.14348/molcells.2017.0033
Kornberg RD, Lorch Y (1999) Twenty-five years of the nucleosome, fundamental particle of the eukaryote chromosome. Cell 98:285–294
Lee E, Muller WJ (2010) Oncogenes and tumor suppressor genes. Cold Spring Harb Perspect Biol 2:a003236. https://doi.org/10.1101/cshperspect.a003236
Li Z et al (2017) APOBEC signature mutation generates an oncogenic enhancer that drives LMO1 expression in T-ALL. Leukemia 31:2057–2064. https://doi.org/10.1038/leu.2017.75
Lieberman-Aiden E et al (2009) Comprehensive mapping of long-range interactions reveals folding principles of the human genome. Science 326:289–293. https://doi.org/10.1126/science.1181369
Lynch HT, Snyder CL, Shaw TG, Heinen CD, Hitchins MP (2015) Milestones of Lynch syndrome: 1895–2015. Nat Rev Cancer 15:181–194. https://doi.org/10.1038/nrc3878
Mansour MR et al (2014) Oncogene regulation. An oncogenic super-enhancer formed through somatic mutation of a noncoding intergenic element. Science 346:1373–1377. https://doi.org/10.1126/science.1259037
Maston GA, Evans SK, Green MR (2006) Transcriptional regulatory elements in the human genome. Annu Rev Genomics Hum Genet 7:29–59. https://doi.org/10.1146/annurev.genom.7.080505.115623
Mathelier A, Lefebvre C, Zhang AW, Arenillas DJ, Ding J, Wasserman WW, Shah SP (2015) Cis-regulatory somatic mutations and gene-expression alteration in B-cell lymphomas. Genome Biol 16:84. https://doi.org/10.1186/s13059-015-0648-7
Melton C, Reuter JA, Spacek DV, Snyder M (2015) Recurrent somatic mutations in regulatory regions of human cancer genomes. Nat Genet 47:710–716. https://doi.org/10.1038/ng.3332
Mills RE, Luttig CT, Larkins CE, Beauchamp A, Tsui C, Pittard WS, Devine SE (2006) An initial map of insertion and deletion (INDEL) variation in the human genome. Genome Res 16:1182–1190. https://doi.org/10.1101/gr.4565806
Mullaney JM, Mills RE, Pittard WS, Devine SE (2010) Small insertions and deletions (INDELs) in human genomes. Hum Mol Genet 19:R131–R136. https://doi.org/10.1093/hmg/ddq400
Nakagawa H, Chadwick RB, Peltomaki P, Plass C, Nakamura Y, de La Chapelle A (2001) Loss of imprinting of the insulin-like growth factor II gene occurs by biallelic methylation in a core region of H19-associated CTCF-binding sites in colorectal cancer. Proc Natl Acad Sci U S A 98:591–596. https://doi.org/10.1073/pnas.011528698
Nambiar M et al (2008) Chromosomal translocations in cancer. Biochim Biophys Acta 1786(2):139–152. https://doi.org/10.1016/j.bbcan.2008.07.005
Navin NE (2015) The first five years of single-cell cancer genomics and beyond. Genome Res 25:1499–1507. https://doi.org/10.1101/gr.191098.115
Oh S, Oh C, Yoo KH (2017) Functional roles of CTCF in breast cancer. BMB Rep 50:445–453
Panero J et al (2016) Acquired TERT promoter mutations stimulate TERT transcription in mantle cell lymphoma. Am J Hematol 91:481–485. https://doi.org/10.1002/ajh.24324
Peifer M et al (2015) Telomerase activation by genomic rearrangements in high-risk neuroblastoma. Nature 526:700–704. https://doi.org/10.1038/nature14980

Perera D, Poulos RC, Shah A, Beck D, Pimanda JE, Wong JW (2016) Differential DNA repair underlies mutation hotspots at active promoters in cancer genomes. Nature 532:259–263. https://doi.org/10.1038/nature17437

Pezzuto F, Buonaguro L, Buonaguro FM, Tornesello ML (2017) Frequency and geographic distribution of TERT promoter mutations in primary hepatocellular carcinoma. Infect Agent Cancer 12:27. https://doi.org/10.1186/s13027-017-0138-5

Pott S, Lieb JD (2015) What are super-enhancers? Nat Genet 47:8–12. https://doi.org/10.1038/ng.3167

Poulos RC, Thoms JA, Shah A, Beck D, Pimanda JE, Wong JW (2015) Systematic screening of promoter regions pinpoints functional cis-regulatory mutations in a cutaneous melanoma genome. Mol Cancer Res 13:1218–1226. https://doi.org/10.1158/1541-7786.mcr-15-0146

Poulos RC, Thoms JAI, Guan YF, Unnikrishnan A, Pimanda JE, Wong JWH (2016) Functional mutations form at CTCF-cohesin binding sites in melanoma due to uneven nucleotide excision repair across the motif. Cell Rep 17:2865–2872. https://doi.org/10.1016/j.celrep.2016.11.055

Ptashne M, Gann A (1997) Transcriptional activation by recruitment. Nature 386:569–577. https://doi.org/10.1038/386569a0

Qian J et al (2014) B cell super-enhancers and regulatory clusters recruit AID tumorigenic activity. Cell 159:1524–1537. https://doi.org/10.1016/j.cell.2014.11.013

Qiu L, Wang Z, Shi X, Wang Z (2008) Associations between XPC polymorphisms and risk of cancers: a meta-analysis. Eur J Cancer 44:2241–2253. https://doi.org/10.1016/j.ejca.2008.06.024

Rachakonda PS et al (2013) TERT promoter mutations in bladder cancer affect patient survival and disease recurrence through modification by a common polymorphism. Proc Natl Acad Sci U S A 110:17426–17431. https://doi.org/10.1073/pnas.1310522110

Rada-Iglesias A, Bajpai R, Swigut T, Brugmann SA, Flynn RA, Wysocka J (2011) A unique chromatin signature uncovers early developmental enhancers in humans. Nature 470:279–283. https://doi.org/10.1038/nature09692

Rhee HS, Pugh BF (2011) Comprehensive genome-wide protein-DNA interactions detected at single-nucleotide resolution. Cell 147:1408–1419. https://doi.org/10.1016/j.cell.2011.11.013

Rheinbay E et al (2017) Recurrent and functional regulatory mutations in breast cancer. Nature 547(7661):55–60. https://doi.org/10.1038/nature22992

Riethoven JJ (2010) Regulatory regions in DNA: promoters, enhancers, silencers, and insulators. Methods Mol Biol 674:33–42. https://doi.org/10.1007/978-1-60761-854-6_3

Rowley JD (2001) Chromosome translocations: dangerous liaisons revisited. Nat Rev Cancer 1:245–250. https://doi.org/10.1038/35106108

Sabarinathan R, Mularoni L, Deu-Pons J, Gonzalez-Perez A, Lopez-Bigas N (2016) Nucleotide excision repair is impaired by binding of transcription factors to DNA. Nature 532:264–267. https://doi.org/10.1038/nature17661

Shlyueva D, Stampfel G, Stark A (2014) Transcriptional enhancers: from properties to genome-wide predictions. Nat Rev Genet 15:272–286. https://doi.org/10.1038/nrg3682

Simon M et al (2015) TERT promoter mutations: a novel independent prognostic factor in primary glioblastomas. Neuro Oncol 17:45–52. https://doi.org/10.1093/neuonc/nou158

Simonis M et al (2006) Nuclear organization of active and inactive chromatin domains uncovered by chromosome conformation capture-on-chip (4C). Nat Genet 38:1348–1354. https://doi.org/10.1038/ng1896

Smith KS, Yadav VK, Pedersen BS, Shaknovich R, Geraci MW, Pollard KS, De S (2015) Signatures of accelerated somatic evolution in gene promoters in multiple cancer types. Nucleic Acids Res 43:5307–5317. https://doi.org/10.1093/nar/gkv419

Steenman MJ et al (1994) Loss of imprinting of IGF2 is linked to reduced expression and abnormal methylation of H19 in Wilms' tumour. Nat Genet 7(3):433–439. https://doi.org/10.1038/ng0794-433

Sur I, Taipale J (2016) The role of enhancers in cancer. Nat Rev Cancer 16:483–493. https://doi.org/10.1038/nrc.2016.62

Taberlay PC, Statham AL, Kelly TK, Clark SJ, Jones PA (2014) Reconfiguration of nucleosome-depleted regions at distal regulatory elements accompanies DNA methylation of enhancers and insulators in cancer. Genome Res 24:1421–1432. https://doi.org/10.1101/gr.163485.113

The ENCODE Project Consortium (2012) An integrated encyclopedia of DNA elements in the human genome. Nature 489:57–74. https://doi.org/10.1038/nature11247

Valton AL, Dekker J (2016) TAD disruption as oncogenic driver. Curr Opin Genet Dev 36:34–40. https://doi.org/10.1016/j.gde.2016.03.008

Vogelstein B, Papadopoulos N, Velculescu VE, Zhou S, Diaz LA Jr, Kinzler KW (2013) Cancer genome landscapes. Science 339:1546–1558. https://doi.org/10.1126/science.1235122

Voss TC, Hager GL (2014) Dynamic regulation of transcriptional states by chromatin and transcription factors. Nat Rev Genet 15:69–81. https://doi.org/10.1038/nrg3623

Ward DG et al (2016) Multiplex PCR and next generation sequencing for the non-invasive detection of bladder cancer. PLoS One 11:e0149756. https://doi.org/10.1371/journal.pone.0149756

Weinhold N, Jacobsen A, Schultz N, Sander C, Lee W (2014) Genome-wide analysis of noncoding regulatory mutations in cancer. Nat Genet 46:1160–1165. https://doi.org/10.1038/ng.3101

Weischenfeldt J et al (2017) Pan-cancer analysis of somatic copy-number alterations implicates IRS4 and IGF2 in enhancer hijacking. Nat Genet 49:65–74. https://doi.org/10.1038/ng.3722

Yin Y et al (2017) Impact of cytosine methylation on DNA binding specificities of human transcription factors. Science 356:eaaj2239. https://doi.org/10.1126/science.aaj2239

Zacher A et al (2017) Molecular diagnostics of gliomas using next generation sequencing of a glioma-tailored gene panel. Brain Pathol 27:146–159. https://doi.org/10.1111/bpa.12367

Zhao Z et al (2006) Circular chromosome conformation capture (4C) uncovers extensive networks of epigenetically regulated intra- and interchromosomal interactions. Nat Genet 38:1341–1347. https://doi.org/10.1038/ng1891

Germline Epigenetic Testing of Imprinting Disorders in a Diagnostic Setting

9

Elizabeth Algar

Abstract

Imprinting disorders are complex and ideally require diagnostic testing in a specialized laboratory. The testing laboratory must have a thorough understanding of the disease biology and testing platform limitations. As the biological basis for these rare disorders is revealed, testing approaches must encompass both genomic and epigenomic investigations to ensure that affected families are fully informed of transmission and recurrence risks.

Keywords

Genomic imprinting · Methylation · Methylation-sensitive multiplex ligation-dependent probe amplification · Clinical testing

Abbreviations

SNV	Single nucleotide variation
MLPA	Multiplex ligation-dependent probe amplification
MS-MLPA	Methylation-sensitive multiplex ligation-dependent probe amplification
LOM	Loss of methylation
GOM	Gain of methylation
DMR	Differentially methylated region
FTT	Failure to thrive
IUGR	Intrauterine growth restriction

E. Algar (✉)
Centre for Cancer Research, Hudson Institute of Medical Research, Clayton, VIC, Australia

Department of Translational Medicine, Monash University, Clayton, VIC, Australia
e-mail: elizabeth.algar@monash.edu

L. B. Hesson, A. L. Pritchard (eds.), *Clinical Epigenetics*,
https://doi.org/10.1007/978-981-13-8958-0_9

9.1 Introduction

One of the main applications of germline epigenetic testing is for the diagnosis of human imprinting disorders, comparatively rare developmental conditions with a range of phenotypic features and severity. Recognition of imprinting disorders has increased in the last decade with improved phenotypic and genomic classification. Human imprinting disorders are caused either by the abnormal expression or inactivation of imprinted genes. Imprinted genes are monoallelically expressed genes where expression is determined by parental origin. The inactivation of a single expressed allele is sufficient to cause an imprinting disorder, mimicking a dominant genetic condition. Common causation is the silencing or removal of an active allele by methylation, uniparental isodisomy, deletion or mutation, or the activation of a normally silent allele by demethylation or by disruption to imprinting control in *cis*. In some circumstances abnormal dosage of the active allele may also be caused by trisomy, uniparental isodisomy and tandem duplications. Although the mutation spectrum in each imprinting disorder shows considerable variation, disturbance to the epigenetic modification of an imprinted gene (or genes), resulting in altered gene dosage in early development, is often a common feature.

9.2 Genomic Imprinting

Imprinted genes are genes that are exclusively monoallelically expressed from only one of the parental chromosomes. This is in contrast to genes exhibiting a tendency towards allele-specific expression that is independent of parental origin and regulated in *cis* by the presence of SNVs (Tycko 2010). The direction of imprinting, i.e. the parent of origin of either the expressed or silenced allele, is highly conserved across all eutherian mammals. Correct parental gene dosage is essential for normal human development. This was first demonstrated by early studies on mice showing that androgenetic (paternally derived chromosomes only) and parthenogenetic (maternally derived chromosomes only) embryos did not survive to term (Surani and Barton 1983; McGrath and Solter 1984). Tetraploid mammalian embryos were also not viable demonstrating the importance of gene dosage for normal mammalian development. Finely tuned mechanisms of genomic imprinting have thus evolved to maintain a delicate balance between gene expression from the maternal and paternal genomes.

Various theories have been postulated as to the evolution of imprinting in mammals including the parental conflict theory that proposed the paternal genome was epigenetically modified to promote the expression of genes for increasing offspring growth in the next generation, thereby increasing the genetic fitness of fathers. In contrast, the maternal genome is hypothesized as epigenetically modified to restrict resources to offspring in a litter to conserve future maternal reproductive potential (Moore and Haig 1991; Moore and Reik 1996). Other theories propose that genomic imprinting mechanisms originally evolved as a defensive mechanism in mammals to silence viral genomes (McDonald et al. 2005).

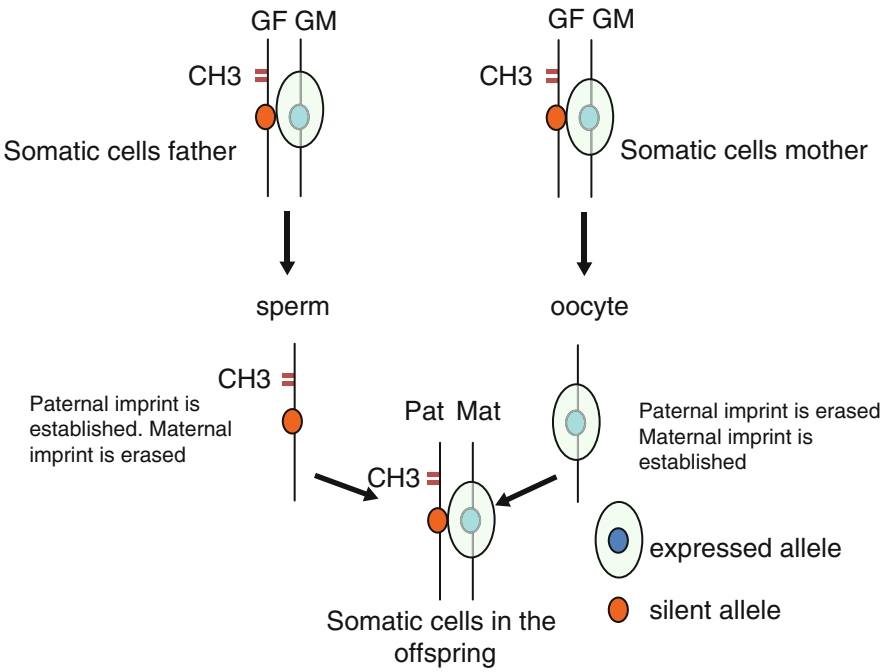

Fig. 9.1 Resetting of the parental imprints in the gametes. The diagram is a simplistic illustration of how correctly imprinted chromosomes derived from the grandparental generation end up in the zygote. GF and GM depict the parental chromosomes inherited from the grandparental generation where GF is a grandfather and GM is a grandmother belonging to the zygote. During meiosis in each parent, the gametes acquire one of the grandparental chromosomes (or region thereof), and the imprint is erased and reset to reflect its new parental origin. When the correctly imprinted gametes join at fertilization, development in the zygote is normal. Pat and Mat refer to the paternally and maternally inherited chromosomes in the zygote, respectively

Genomic imprints are set in the primordial germ cells of the gametes according to their parent of origin. When the gametes come together in the zygote, each genome carries the appropriate parental imprint to ensure correct dosage of the parental genome. Imprinting in the zygote must then be maintained with each cell division for normal development. Factors that affect the resetting of parental imprints in the primordial germ cells or factors that affect imprinting maintenance in the zygote have the potential to predispose to imprinting errors in the foetus (see Fig. 9.1).

There are more than 100 imprinted genes in humans (www.geneimprint.com), and many of these imprinted genes are clustered in imprinting domains. Imprinted loci are usually controlled by a region of differential methylation, or DMR (differentially methylated region), within the imprinting domain. Imprinted expression within imprinting domains or imprinting centres can be regulated by several mechanisms and many of these remain to be fully elucidated. Some involve antisense transcription of long non-coding RNA, others the formation of boundary elements via binding of CTCF and interactions in *cis* between long-range enhancers

and imprinted gene promoters. Differential methylation is however a common theme in the most well-characterized imprinting centres. Examples are discussed in the sections following.

While imprinted genes themselves can carry dominant mutations affecting DNA sequence, more commonly the imprinting process itself is affected either by epigenetic disruption to imprinting control, by insertions or deletions disrupting the imprinting control region (ICR) or by whole chromosome or region rearrangements including uniparental disomies and chromosomal duplications and deletions. Considerable variation however exists in the distribution of these mutation classes in human imprinting disorders, and for some of the more rare conditions, the relative distribution of genomic deletions, uniparental disomies and epimutation is not definitively known. Imprinting disorders can be inherited when a genetic disruption occurs via deletions or insertions within an ICR. While these occurrences are rare, it is important to be able to detect them for accurate genetic counselling in affected families. An example is shown in Fig. 9.2 of a hypothetical mutation in a family with Prader-Willi syndrome where a mutation affects imprinting establishment in the gametes and prevents correct imprinting of the paternal SNURF/SNRPN locus. This mutation remains silent in the family until it is paternally transmitted and produces an incorrectly imprinted paternal chromosome in a grandson.

Depending on the timing of the acquisition of imprinting defects, either in the gametes or in early embryogenesis, the imprinting defect may be complete and present in all somatic tissues, or partial, and hence mosaic in affected somatic tissues. Thus methylation, or other epigenetic signatures affecting imprinting, when measured, may show variation dependent on the developmental timing at which they arose. Co-incident with this, considerable phenotypic variability may be observed with imprinting defects that are acquired later in embryogenesis and these tend to be associated with milder phenotypes. Where methylation defects or chromosomal defects such as uniparental disomies are acquired post-zygotically, they manifest as mosaic in peripheral blood or other tissues (buccal cells, skin fibroblasts) that are accessible for testing. From a practical perspective, reference ranges for methylation at imprinted loci in normal reference specimens need to be carefully established for reliable diagnosis of imprinting disorders, and the detection limit for the diagnostic method must be known. This is most important for reliable diagnosis of mosaic forms of Russell-Silver syndrome, Beckwith-Wiedemann syndrome and hemihypertrophy so that the possibility of a false-negative diagnosis can be estimated with some degree of accuracy.

9.3 Human Imprinting Disorders

While it is likely that disruption to imprinting at each of the 100 or more known human imprinted loci has potential to affect normal development, only seven imprinted domains (IDs) have been definitively associated with described or classified imprinting disorders. These IDs are located on chromosome regions 11p15.5, 15q11–13, 7p12.2, 7q32.2, 6q24, 20q13 and 14q32. Imprinting disruption to these

regions causes the syndromes outlined in Table 9.1. Collectively these disorders are associated with developmental abnormalities and either growth restriction or overgrowth, as summarized in Table 9.1. Together they are classified as rare disorders; however, they have a wide range of incidence in the population. Beckwith-Wiedemann syndrome and its clinical variant hemihypertrophy have a combined population frequency of 1 in 10,000; however, this figure may be an underestimate due to the fact that milder presentations may be overlooked (Mussa et al. 2013). Russell-Silver syndrome has a population frequency of between 1 in 3000 and 1 in 100,000 (Abu-Amero et al. 2008). Bilateral Wilms tumour has a population frequency of between 1 in 50,000 and 1 in 80,000 (Charlton et al. 2017). Prader-Willi syndrome has a population frequency of 1 in 25,000 (Smith et al. 2003), and Angelman syndrome has a population frequency of between 1 in 12,000 and 1 in 24,000 (Mertz et al. 2013). The incidence of multi-locus imprinting disruption (MLID) is unknown; however, it is predicted to affect approximately 1 in 50,000 [based on the assumption that approximately 1 in 5 Beckwith-Wiedemann presentations may have MLID (Azzi et al. 2009)]. Pseudohypoparathyroidism 1b affects less than 1 in 150,000 (0.79/100,000, Orphanet Report Series, November 2011). Temple syndrome and Kagami-Ogata syndrome are rare disorders with a

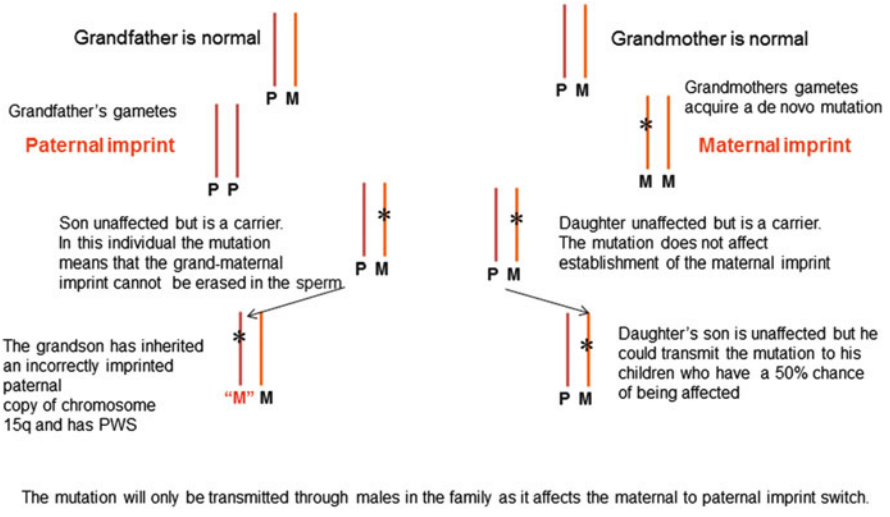

Fig. 9.2 Transmission of a hypothetical SNURF/SNRPN imprinting centre mutation in a family with a grandson affected by Prader-Willi syndrome. M and P refer to the maternal and paternal imprint on 15q11–13, respectively. The gametes of the grandmother carry an imprinting centre mutation denoted by an asterisk that affects the setting of a correct paternal imprint in the gametes of her son. This imprinting centre mutation does not affect the establishment of a maternal imprint in either her daughter or her son. The son's offspring have a 50% chance of inheriting the incorrectly imprinted chromosome 15q11–13 which still carries a maternal, rather than a paternal, imprint. In this scenario Prader-Willi syndrome will be the result as there are two chromosomes with a maternal imprint present in affected offspring. The daughter will be a carrier of the mutation and has a 50% chance of transmitting it to future generations, where paternal transmission will cause Prader-Willi syndrome

Table 9.1 Human imprinting disorders in which the epigenetic defect has been characterized

Syndrome	Imprinted domain	Epimutation/ genomic Mut[n]	Phenotypic features
Russell-Silver syndrome (OMIM 180860)	11p15.5 H19/IGF2 IG-DMR	LOM H19/IGF2 IG-DMR (60%)	IUGR, short stature, FTT, triangular facies, hemi-atrophy, clinodactyly
Russell-Silver syndrome (OMIM 180860)	7p12.2 and 7q32.1	7p12.2 GRB10 GOM -matUPD7 7q32.1 PEG1/ MEST GOM -matUPD7 (10% overall)	As above
Beckwith-Wiedemann syndrome (OMIM 130650)	11p15.5 H19/IGF2 IG-DMR	GOM H19/IGF2 IG -DMR (10%)	Macrosomia, high birth weight, transient neonatal hypoglycaemia, macroglossia, exomphalos and umbilical hernia, hemihypertrophy, cancer, ear creases and pits
Beckwith-Wiedemann syndrome (OMIM 130650)	11p15.5 H19/IGF2 IG -DMR 11p15.5 KCNQ1OT1: TSS-DMR	patUPD11p15.5 (15–20%)	As above
Beckwith-Wiedemann syndrome (OMIM 130650)	11p15.5 KCNQ1OT1: TSS-DMR	LOM KCNQ1OT1: TSS-DMR (50%)	As above
Wilms tumour (OMIM 194071)	11p15.5. H19/IGF2 IG-DMR.	GOM H19/IGF2 IG-DMR (<5%)	Wilms tumour.
Wilms tumour (OMIM 194071)	11p15.5 H19/IGF2 IG-DMR 11p15.5 KCNQ1OT1: TSS-DMR	patUPD11p15.5 (<5%)	Wilms tumour
Hemihypertrophy	11p15.5. H19/IGF2 IG -DMR and 11p15.5 KCNQ1OT1: TSS-DMR	patUPD11p15.5 (15%)	Macrosomia, asymmetry-hemihypertrophy. Wilms tumour Hepatoblastoma
Hypomethylation at imprinted loci (HIL) or multi-locus imprinting disruption (MLID)	Multiple including 6q24, 11p15.5, 20q13, 7p12.2, 19q13	LOM at 11p15.5 KCNQ1OT1: TSS-DMR plus other IDs in varying combinations (50%)	BWS and atypical features including IUGR and TNDM

(continued)

Table 9.1 (continued)

Syndrome	Imprinted domain	Epimutation/ genomic Mut[n]	Phenotypic features
Transient neonatal diabetes 1 (OMIM 601410)	6q24	LOM mat TND DMR. ZAC/PLAGL1 disruption	Transient hyperglycaemia as neonate. Macroglossia, exomphalos and umbilical hernia, IUGR
Prader-Willi syndrome (OMIM 176270)	15q11–13	GOM pat. SNRPN DMR (1%) Paternal 15q11 del or matUPD15q11 (>95%)	Hypotonia, obesity, FTT, short stature, hypogonadism, facial anomalies
Angelman syndrome (OMIM 105830)	15q11–13	LOM mat. SNRPN DMR (2–3%) Maternal 15q11del or patUPD 15q11 removing UBE3A expression (80%)	Mental retardation. Absence of speech, inappropriate laughter, ataxia, seizures, tongue protrusion
Pseudohypoparathyroidism 1b (OMIM 603233)	20q13.2	LOM GNAS1 DMR exon 1A. STX16 deletion	Parathyroid hormone resistance as for PHP1a but without Albright hereditary osteodystrophy (AHO)
Temple syndrome (OMIM 616222)	14q32	LOM IG-DMR DLK1/GTL2 MatUPD14, paternal del 14q32	Short stature, hypotonia, motor delay, feeding problems, small hands and feet
Kagami-Ogata syndrome (OMIM 608149)	14q32	GOM IG-DMR DLK1/GTL2 PatUPD14, maternal del 14q32	Skeletal abnormalities, joint contractures, dysmorphism, DD

Diagnostic testing is now available for these disorders using a variety of methods of which the most informative is MS-MLPA. The table does not show a comprehensive list of all mutations in these disorders and shows only the imprinting defect or the genomic defect that causes disruption to imprinted gene expression, along with mutation frequency where it is known. Genes carrying point mutations or microdeletions have not been included but are discussed in the text. *LOM* loss of methylation, *GOM* gain of methylation, *DMR* differentially methylated region, *IUGR* intrauterine growth restriction, *MLID* multi-locus imprinting disturbance, *TNDM* transient neonatal diabetes mellitus, *TSS* transcription start site, *ID* imprinting domain, *FTT* failure to thrive, *DD* developmental delay

presently unknown population frequency, and transient neonatal diabetes 1 has a population frequency between 1 in 300,000 and 1 in 400,000 (Shield 2000).

Several of these disorders can be associated with single gene mutations; however, in general these are less frequent than an epimutation. An exception to this is Angelman syndrome where UBE3A mutations are more prevalent than epimutations and are present in 10% of cases, whereas disruption to genomic imprinting by an altered pattern of methylation, as opposed to a chromosome deletion or UPD, is comparatively rare in this condition affecting 2–3% of cases (reviewed in Sadikovic et al. 2014). Rare inactivating mutations affecting IGF2 and activating mutations in CDKN1C have been reported in Russell-Silver syndrome (Liu et al. 2017; Brioude et al. 2013), and inactivating mutations in CDKN1C occur in 5% of sporadic cases with Beckwith-Wiedemann syndrome (Romanelli et al. 2010). Transient neonatal diabetes 1 can be caused by ZFP57 mutations in up to 50% of cases (Bak et al. 2016). ZFP57 mutations can also occur in association with widespread loss of methylation defects involving PLAGL and other loci (Mackay et al. 2008). NLRP5 mutations have a maternal effect on imprinting in the zygote when present in the mother and have been associated with MLID (Docherty et al. 2015). Hence, single gene mutations can cause an epimutation, as in the case of ZFP57 and NLRP5, or lead to the inactivation of an expressed imprinted gene equivalent to a dominant mutation as in the case of CDKN1C mutation in BWS, IGF2 mutation in RSS and UBE3A mutation in Angelman syndrome. In pseudohypoparathyroidism type 1b (PHP 1b), deletion of the STX16 gene can occur in association with loss of methylation at GNAS exon 1A, and it is thought to play a role in regulating imprinting of the GNAS locus (reviewed in Turan and Bastepe 2015).

In Wilms tumour, imprinting disruption to the 11p15.5 H19/IGF2 IG-DMR occurs as a germline event in up to 12% of bilateral cases and overall in <5% of cases (Scott et al. 2008; Cardoso et al. 2012). Mosaic paternal isodisomy for chromosome region 11p15 (patUPD11p15), leading to disruption of the expression of several imprinted genes including IGF2 and CDKN1C, occurs as a germline event in a further subset of cases predominantly those with some additional features characteristic of Beckwith-Wiedemann syndrome including hemihypertrophy (Cardoso et al. 2012).

While epimutation in Beckwith-Wiedemann syndrome at 11p15.5 KCNQ1OT1: TSS-DMR is more common than paternal isodisomy for chromosome 11p15 (50% of cases versus 15–20% of cases), in Prader-Willi and Angelman syndromes, deletions of the paternal and maternal copies of 15q11–13, respectively (with or without isodisomy), are the most frequent genomic abnormalities encountered leading to abnormal gene dosage from the parental chromosomes in greater than 90% of cases. Imprinting defects affecting the SNRPN locus in both Prader-Willi and Angelman syndromes are comparatively far less common with approximately 1% affected in Prader-Willi syndrome and 2–3% in Angelman syndrome. MAGEL2 mutations have also been reported in Prader-Willi syndrome; however, their role is currently unclear (Buiting et al. 2014; Schaaf et al. 2013).

Epimutation involving gain of methylation within the H19/IGF2 IG-DMR is present in 10% of cases with Beckwith-Wiedemann syndrome, and 20% of these cases may have an underlying mutation or deletion within the DMR that is heritable. These include deletions that disrupt the formation of boundary elements within the imprinting centre on the maternal chromosome and disruption to OCT4 binding sites that prevent the resetting of the maternal imprints either in the oocytes or in the preimplantation embryo (Riccio et al. 2009; Poole et al. 2012; Berland et al. 2013). Duplication of the H19/IGF2 IG-DMR has also been reported in familial Wilms tumour (Algar et al. 2007), and deletions within the maternal KCNQ1OT1: TSS-DMR can also occur in familial Beckwith-Wiedemann syndrome (Algar et al. 2011; Niemitz et al. 2004).

Temple syndrome is a rare syndrome predominantly caused by matUPD14 and paternal deletions of chromosome 14 affecting 14q32. A subgroup of patients with biparental inheritance for 14q32 have an epimutation with a maternal-only methylation pattern at the IG-DMR DLK1/GTL2 (Buiting et al. 2008; Ioannides et al. 2014). Kagami-Ogata syndrome is equally rare and is caused by patUPD14 and maternal deletions of chromosome 14 affecting 14q32 (Kagami et al. 2005).

The mechanics of allelic silencing and activation associated with these disorders is not well understood; however, the analysis of methylation patterns within each DMR provides information as to the parental origin of the alleles that are present. Methylation testing for these disorders can provide a diagnosis of the condition but does not in all cases reveal the underlying mechanism. Where a maternal uniparental isodisomy is present for example, any imprinted locus within the affected region will carry a maternal pattern of methylation and a normal chromosome copy number. Conversely where a paternal uniparental isodisomy is present, imprinted loci within the affected region will have a methylation pattern exclusively of paternal origin and a normal chromosome copy number. Ancillary methods to identify or exclude allele copy number changes, microdeletions or insertions within the imprinting centre, or to confirm a maternal or paternal isodisomy, are almost always required and are extremely important to rule out heritable mutations. These usually include genotyping studies and sequencing.

Some of the comparatively well-characterized imprinting centres are those on chromosome regions 11p15.5 and 15q11–13. Illustrated examples are provided in Figs. 9.3, 9.4 and 9.5 of how imprinting in each domain is maintained.

Mutations that affect the expressed alleles of imprinted genes are dominant mutations. These mutations may however be silent for generations dependent on the transmitting parent. For example, a mutation affecting a paternally expressed imprinted gene will only affect the phenotype when it is paternally transmitted. Maternal transmission of the same mutation will not produce a phenotype as the affected allele remains silent in the next generation. Conversely, a maternally expressed imprinted gene will only affect the phenotype when the mutation is maternally transmitted. A hypothetical example is shown in Fig. 9.6. Consequently, complete molecular diagnosis of imprinting disorders usually involves both genomic and epigenomic testing approaches.

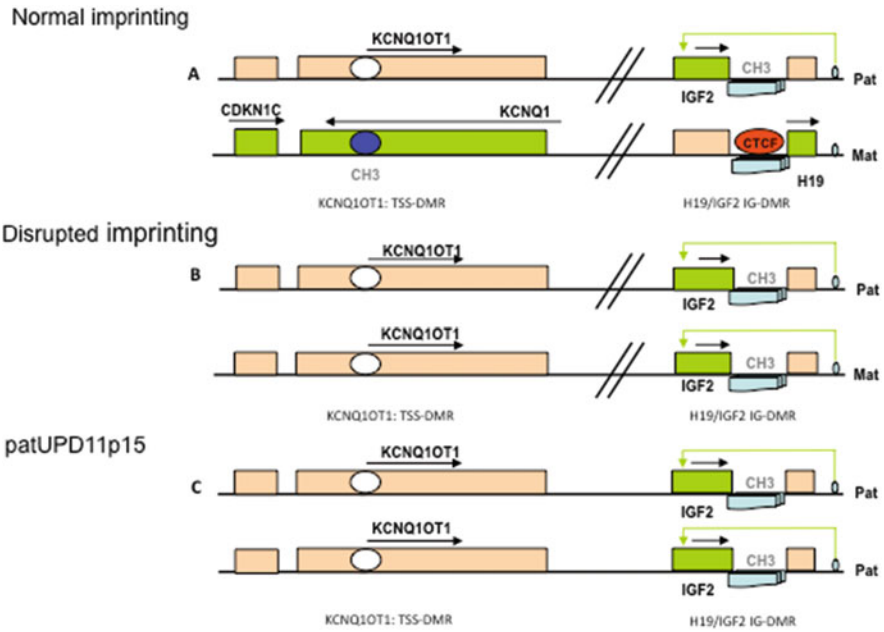

Fig. 9.3 Genomic imprinting disruption in Beckwith-Wiedemann syndrome (BWS). (**a**) Normal imprinting at 11p15.5 is maintained at H19/IGF2 IG-DMR and KCNQ1OT1: TSS-DMR by differential methylation of the parental chromosomes. In the H19/IGF2 IG-DMR, binding of CTCF to the unmethylated DMR forms a boundary element preventing the interaction of distal enhancers (small light blue circles) with fetal *IGF2* promoters. The mechanisms of imprinting maintenance at KCNQ1OT1: TSS-DMR have not yet been definitively determined. Expressed alleles are depicted by green boxes and denoted by gene name. (**b**) Disruption to imprinting affects either H19/IGF2 IG-DMR or KCNQ1OT1: TSS-DMR independently and leads to overlapping phenotypes in BWS. Patients have either a gain of methylation at *H19* (10% of cases) or a loss of methylation at *KCNQ1OT1* (50% of cases). Gain of methylation at *H19* leads to biallelic expression of imprinted *IGF2*, and loss of methylation at *KCNQ1OT1* leads to biallelic silencing of imprinted *CDKN1C*. (**c**) Paternal isodisomy (patUPD11p15) affects 15–20% of BWS cases, and a paternal imprint is predominant in these individuals at both H19/IGF2 IG-DMR and KCNQ1OT1: TSS-DMR leading to abnormal expression of both *IGF2* and *CDKN1C*. Paternal isodisomy in BWS is typically mosaic

9.4 Methods to Detect Germline Methylation Abnormalities in a Clinical Setting

9.4.1 Methylation-Sensitive Multiplex Ligation-Dependent Probe Amplification (MS-MLPA)

Methylation-sensitive multiplex ligation-dependent probe amplification (MS-MLPA) has become the mainstay and method of choice for the determination of allelic methylation in human imprinting disorders in a clinical setting (Eggermann

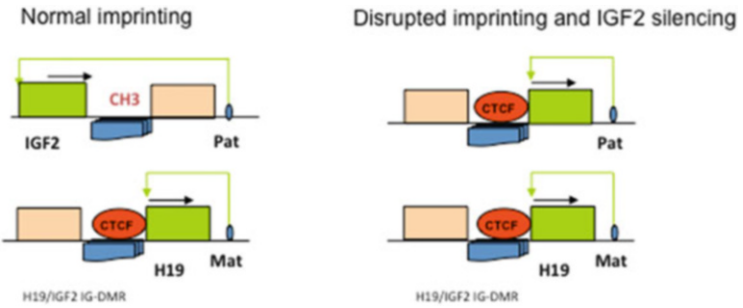

Fig. 9.4 Genomic imprinting disruption in Russell-Silver syndrome (RSS). The predominant imprinting defect is a loss of methylation at the paternal H19/IGF2 IG-DMR leading to CTCF binding preventing access to the paternal *IGF2* promoters (up to 60% of cases). This results in biallelic silencing of *IGF2* and the growth restriction associated with the phenotype, which is characterized by severe intrauterine and post-natal growth restriction. Mosaicism is a common occurrence and may be associated with hemiatrophy

et al. 2016; Ramsden et al. 2010; Grafodatskaya et al. 2017). The principle of the method is similar to that of conventional MLPA. In conventional MLPA a series of test and control probes are designed to hybridize to genomic regions of interest. Control probes are designed to target regions of the genome that are typically not affected by copy number variations. Following the ligation of paired probe oligonucleotides flanking the target region, a PCR amplification step is performed with 5′ and 3′ universal primers in which both test and control probes are amplified. Regions in which the right and left oligonucleotide sequences of any probe have not ligated do not amplify. Products amplified following probe ligation are subjected to capillary gel electrophoresis where fluorophores attached to universal primers, recognizing the 5′ and 3′ oligonucleotide sequences of the right and left hand probes, respectively, enable fragments to be identified and quantified.

The fluorescent signal for each target probe amplified is normalized by comparison to the population of control probes and then quantified by comparison to a normal reference specimen run in parallel. For an invariant genomic region, a normalized test probe has a value of close to 1.0, normalized to a reference specimen. If a heterozygous deletion is present that affects probe ligation, then only the probe targeting the non-deleted allele will amplify, and a normalized ratio of 0.50 will be obtained. If a heterozygous insertion is present, a normalized ratio for that probe of 1.50 will be obtained.

Test probes are designed to the target region of interest and at least two test probes are usually required for each coding exon of the gene under interrogation. Each probe comprises a paired forward and reverse oligonucleotide designed to target a genomic region. MLPA is designed to generate amplicons in the size range from 130 to 480 nucleotides in length. MLPA is performed in a single tube and is a

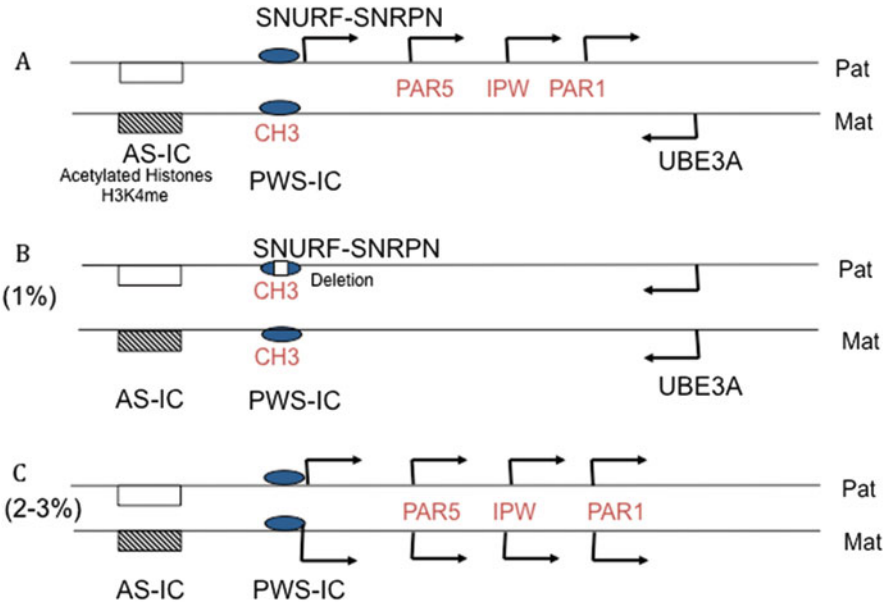

Fig. 9.5 Genomic imprinting disruption in Prader-Willi and Angelman syndromes. (**a**) Imprinting on 15q11.2 is maintained by differential methylation within the *SNURF-SNRPN* imprinting centre. This is associated with paternal expression of a number of transcripts including *PAR5*, *IPW*, *PAR1*, *ZNF127* and *NDN* and with maternal expression of *UBE3A*. (**b**) Disruption to imprinting can be caused by a gain of methylation on the paternal copy of *SNURF-SNRPN* leading to impaired expression of paternal transcripts in individuals affected by Prader-Willi syndrome (PWS) (1% of cases) or more rarely by deletions, shown as a white gap within the imprinting centre. (**c**) Angelman syndrome can be caused by a loss of methylation on the maternal chromosome leading to silencing of the *UBE3A* gene that is imprinted in the brain (2–3% of cases) and biallelic expression of paternally expressed transcripts. An Angelman syndrome imprinting centre lies 35 kb distal to *SNURF-SNRPN*, and deletions affecting the IC are also associated with loss of methylation and biallelic expression of paternally expressed transcripts. *AS-IC* Angelman syndrome imprinting centre; *PWS-IC* Prader-Willi syndrome imprinting centre, *CH3* methylation, *Pat* paternal allele, *Mat* maternal allele

relatively straightforward technique to establish in a diagnostic laboratory. All that is required is a PCR machine and access to capillary gel electrophoresis. MLPA tests can be designed in-house (Zhi and Hatchwell 2008) although a large range of commercial MLPA and MS-MLPA kits are now available.

MS-MLPA is a variation on MLPA and is used for DNA methylation analysis at specific sites. The ligation of the MLPA probe oligonucleotides is combined with digestion of the genomic DNA-probe hybrid complexes with methylation-sensitive restriction enzymes, such as *HhaI* (Nygren et al. 2005). Where the target site is methylated, *HhaI* digestion cannot occur, and the region is amplified. If the region is not methylated, and a *HhaI* site is located within the probe binding region, the ligated oligonucleotide sequence is continuously digested with the net result that the flanking oligonucleotide sequence is unable to be amplified by PCR. For any

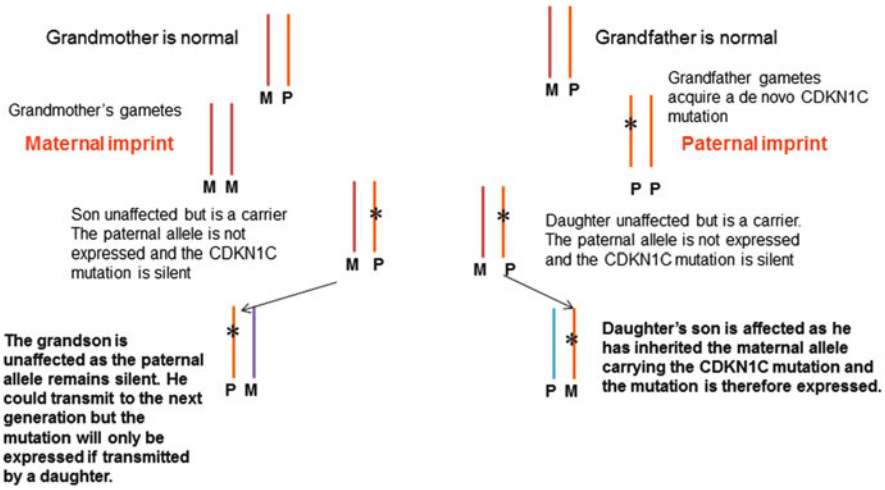

Fig. 9.6 Transmission of mutations in imprinted genes. A phenotype is present only when the expressed allele carries a mutation. Mutations in silent alleles can be transmitted for generations without causing a phenotype. Example shown is relevant to the transmission of a mutation in the *CDKN1C* gene in Beckwith-Wiedemann syndrome. The mutation will only be transmitted through females in the family as it is the maternally inherited allele that is expressed. A carrier female has a 50% chance of transmitting the mutation to her offspring

given DNA specimen, MLPA and MS-MLPA reactions are run in parallel. Typically the probe hybridization is performed in a single tube following which the specimen is divided into two separate reaction tubes for the ligation step (MLPA) and ligation and digestion step (MS-MLPA).

When both alleles are methylated at a target site, neither allele is digested by the methylation-sensitive restriction enzyme, and the comparison of signals obtained from the MS-MLPA and conventional MLPA yields a theoretical ratio of 1.0. If only one allele is methylated at a target site, then the unmethylated allele will be digested and a theoretical ratio of 0.50 is obtained. If neither allele is methylated, then complete digestion of both alleles with *HhaI* will prevent ligation and amplification at that site and a theoretical value for methylation of 0.0 will be obtained. Hence, if the sequence and methylation pattern within an imprinting centre or site of differential methylation are known, then MS-MLPA can be used to interrogate that region and determine whether methylation throughout the region is normal or abnormal. At the same time the MLPA reaction performed in parallel generates copy number data for the same region so one can readily determine whether an abnormal methylation result is explainable by a deletion or duplication of one of the parental chromosomes, or whether the abnormal methylation is due to an imprinting defect or a uniparental isodisomy. However MS-MLPA will not distinguish between a uniparental isodisomy and an imprinting defect and follow-up genotyping studies are required. MS-MLPA probe mixes are now commercially available as research-only use kits for the diagnosis of all of the imprinting disorders that are outlined in Table 9.1.

Representative MS-MLPA data for different imprinting disorders is shown in Fig. 9.7.

MS-MLPA is a rapid and informative technique in a diagnostic setting, and because of its capacity to generate regional copy number data simultaneously, including single exon deletions and insertions, it is more informative than techniques such as methylation-specific PCR (MS-PCR), methylation-sensitive high-resolution

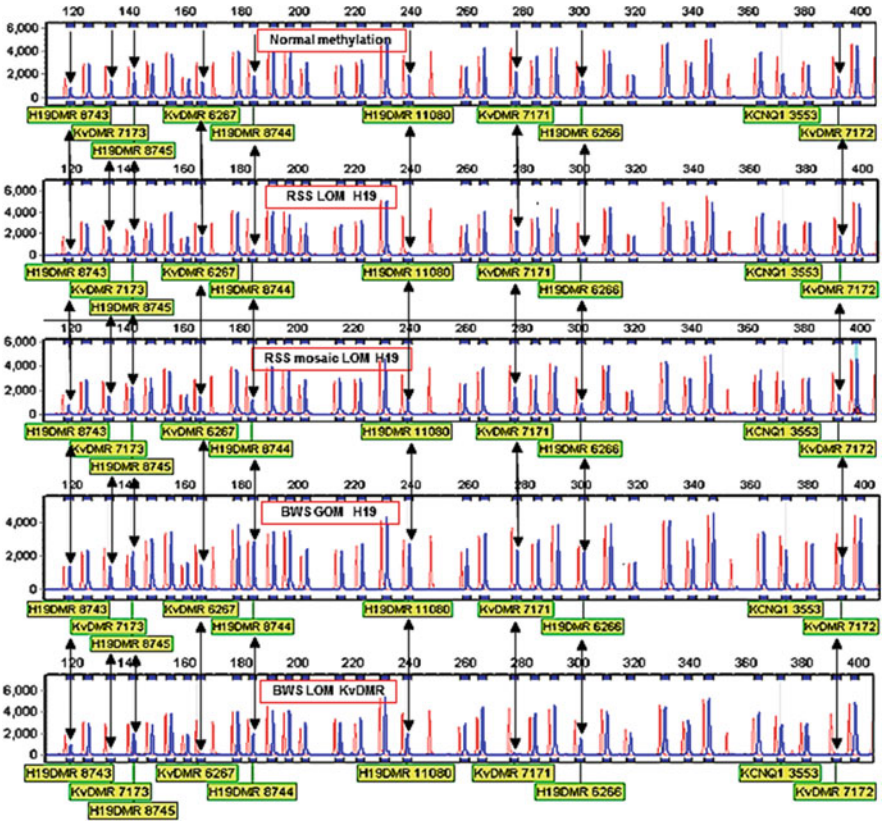

Fig. 9.7 Example data from MS-MLPA performed on patients diagnosed with methylation abnormalities consistent with Beckwith-Wiedemann syndrome and Russell-Silver syndrome. Data were obtained using the MS-MLPA kit BWS/RSS ME030B1 from MRC Holland. The blue peaks represent probes detecting fragments after methylation-sensitive restriction enzyme digestion with HhaI, and the red peaks represent the reference trace. Blue peaks disappear with complete loss of methylation where the probe is within an imprinting centre (either H19/IGF2 IG-DMR or KCNQ1OT1: TSS-DMR) and are equivalent to the red peaks when both alleles are methylated. In a normal specimen, probes targeting the differentially methylated imprinting centres are close to 50% of the height of the paired undigested specimen. Probes falling between these extremes may represent mosaic forms of these disorders, and normal reference ranges must be established in unaffected individuals. *LOM* loss of methylation, *GOM* gain of methylation, *BWS* Beckwith-Wiedemann syndrome, *RSS* Russell-Silver syndrome

melting (MS-HRM) or sequencing bisulphite-modified DNA, where copy number data is not measured and mechanisms cannot be ascertained.

The laboratory using MS-MLPA for the diagnosis of imprinting disorders must have a thorough understanding of the genetics of these disorders before embarking on testing. Positive and negative control specimens and normal reference specimens must be run in any MS-MLPA reaction. The uncertainty of measurement (population variation and technical variation) for both copy number and methylation probes must also be calculated as part of the test validation pipeline. This is of crucial importance for the diagnosis of imprinting disorders such as Beckwith-Wiedemann syndrome, Russell-Silver syndrome and hemihypertrophy where methylation abnormalities are frequently mosaic. Furthermore DNA isolated from different tissues must be independently validated, as some probes display small methylation differences depending on the source of the DNA and the method of extraction. It is our experience that DNA isolated from buccal cells and amniocytes tends to show a wider variation in methylation within the 11p15.5 imprinting centres at some probes when compared with DNA isolated from blood. In addition, cultured lymphocytes consistently show variable methylation at probes within the H19/IGF2 IG-DMR (Algar unpublished observation).

9.4.2 Methylation-Sensitive High-Resolution Melting Analysis

Methylation-sensitive high-resolution melting (MS-HRM) is a cost-effective platform for looking at methylation abnormalities in imprinting disorders and is based on the principle that sequence differences in bisulphite-modified DNA, attributed to methylation status, change the melting characteristic of the amplicon. During the melting process, an intercalating dye is released from the amplicon according to the rate at which it melts. The release of dye is captured as a decreasing fluorescence signal. However MS-HRM has limitations in that the specimen must be converted with bisulphite modification, only one imprinting centre can be examined in each test, and standards must be carefully defined with known levels of methylation so that mosaic methylation cases can be correctly scored (White et al. 2007; Alders et al. 2009; Wojdacz et al. 2008). The read-out of melt curve analysis is a combination of the melting profile for each nucleotide in the amplicon and hence provides a snap-shot of methylation throughout an amplicon. Evidence suggests however that SNVs do not significantly alter the assessment of methylation using this technique (Alders et al.2009). The method is relatively fast (compared to more traditional methods such as southern blotting), robust and sensitive; however, borderline abnormal methylation cases require follow-up with alternative methods (to exclude copy number variations including duplications) and the method is not strictly quantitative. Through the use of known standards, one can however identify the methylation range into which a specimen would fall. The method has been used for the diagnosis of methylation abnormalities within the imprinting centres in Beckwith-Wiedemann, Russell-Silver, Prader-Willi and Angelman syndromes.

9.4.3 Taqman Allele-Specific Methylated Multiplex Quantitative PCR

This technology utilizes Taqman probes targeting bisulphite-modified DNA in a real-time PCR assay to quantify unmethylated and methylated sequence at each imprinting centre. Probes targeting methylated and unmethylated bisulphite-converted DNA are labelled with distinct fluorophores so that the relative amounts of methylated versus unmethylated alleles can be precisely quantified. Advantages of this technique are that only very small amounts of DNA are required (~10 ng) and that the data generated are quantitative (Azzi et al. 2011). The method was developed for testing on Beckwith-Wiedemann and Russell-Silver syndrome cases; however, it may be adaptable for the analysis of additional human imprinting disorders. While this approach is viable for the detection of methylation differences, further assays are required to investigate the underlying aetiology of the methylation defect, for example, whether it is attributable to uniparental isodisomy and copy number variations or to an imprinting defect.

9.4.4 Bisulphite Pyrosequencing

Pyrosequencing is a real-time DNA sequencing method based on a "sequencing through synthesis" modality where nucleotide incorporation utilizes a DNA polymerase generating inorganic pyrophosphate (PPi) in quantitative proportion to nucleotide incorporation. The release of PPi causes the conversion of luciferin to oxyluciferin via the generation of ATP (Nyren 1987). A charge-coupled device (CCD) is used to detect light generated in proportion to the number of complementary nucleotides incorporated during strand synthesis. Four nucleotides are used for strand synthesis including deoxycytidine triphosphate (dCTP), deoxyguanosine triphosphate (dGTP), deoxythymidine triphosphate (dTTP) and deoxyadenosine alpha-thiotriphosphate (dATPαS). One of the key advantages of pyrosequencing is that it not only provides real-time sequence information, but data generated is a quantitative measure of each nucleotide thereby making it a powerful technique for measuring the relative amounts of different alleles. This property also enables the quantification of DNA methylation, heterozygosity, ploidy, hematopoietic chimerism and mixed genotypes in heterogeneous samples (e.g. tumour and normal cells). Pyrosequencing involves the initial generation of a single-stranded biotinylated PCR product that is attached to streptavidin-coated sepharose beads. This single-stranded product is released into the sequencing reaction mixture after binding of the sequencing primer, following which the reaction commences.

For allelic methylation analysis, the DNA is bisulphite-modified prior to PCR and sequencing. For an imprinted locus, or site of differential allelic methylation, the ratio of C:T or G:A should approach 1:1 assuming 50% methylation at CpGs. For application in diagnostic testing, a normal methylation range must be established in a population of normal specimens of the same type (at least ten specimens would be considered a minimum number). The limit of detection at each CpG should also be

ideally verified by testing demethylated control specimens and 100% methylated DNA. Demethylated DNA specimens can be generated by treating human cell lines with demethylating agents such as 5-aza-2′deoxycytidine and used to identify lower limits of methylation; however, imprinted DMRs are not equally sensitive to this treatment. Methylated DNA is available commercially. Repeat testing of reference specimens should be performed to derive a coefficient of variation or uncertainty of measurement at each CpG site. Advancements in pyrosequencing now allow up to 15 different CpG dinucleotides within 80bp of sequence to be analysed making it a useful technology for interrogating both methylation and SNVs within imprinting centres.

Bisulphite pyrosequencing for the analysis of imprinting disorders has been more typically used to date in research rather than in diagnostic settings (Murphy et al. 2012; Mackay et al. 2008; Woodfine et al. 2011). The advantage of pyrosequencing technology is that only small amounts of DNA are required and assays can be multiplexed making it relatively cost effective for examining large numbers of specimens from a variety of conditions; however, specialized instrumentation and analysis software are required for implementation.

9.4.5 Methylation Arrays and Genome-Wide Methods

Methylation arrays use bead chip technology enabling the analysis of thousands of probes targeting CpG sites in the genome. At time of writing, this technology can interrogate 850,000 sites, whereas previous iterations enabled the examination of 450,000 and 27,000 sites. DNA is bisulphite-modified before hybridization to the bead chip and probes on the chip are designed to recognize either methylated or unmethylated versions of the allele. Sophisticated bioinformatics is required for analysis after normalization of the raw data, and quality controls including methods to control for complete conversion of the DNA during bisulphite modification are required. Data output of signal intensity is converted into methylated and unmethylated enabling the derivation of beta values for methylation (Dedeurwaerder et al. 2011). This technology has the power to profile multiple imprinted DMRs. While methylation arrays have immense power to identify the underlying epigenetics of potentially new imprinted gene disorders, it is not the most practical or cost-effective approach to routine genetic diagnosis of an imprinting disorder where the clinical presentation suggests more targeted testing may be informative. Used in a research setting, it can reveal new subgroups of imprinting disorders, convergence and divergence of the epigenome in known imprinting disorders and the epigenomic consequences of mutations in *trans*-acting imprinting factors (Docherty et al. 2014).

Methyl-CpG binding domain protein sequencing (MBD Cap-seq) utilizes fragmented native DNA precipitated by an MBD2 capture protein to enrich methylated DNA regions for sequencing. High sequencing coverage of CpGs can be achieved (Lan et al. 2011) and it has been used to examine the effects on the epigenome of ZFP57 mutations in transient neonatal diabetes type 1 patients who

have hypomethylation at multiple imprinted loci (Bak et al. 2016). Adapters are ligated to the isolated methylated DNA, which is then amplified by PCR using primers to the adapter sequences. PCR libraries are sequenced to identify the methylated regions and quantify the level of methylation. A similar approach is used in MBD-ChIP where the enriched methylated DNA is profiled using microarrays (Brinkman et al. 2010).

MeDIP utilizes a methyl CpG cytosine antibody to immunoprecipitate DNA with methylated CpG sites. The enriched methylated DNA fractions can then be analysed by tiling arrays or by sequencing (MeDIP-seq). With all enrichment methods, bias can be introduced by CpG density and copy number variations, and computational tools to normalize CpG content across a range of densities are required. MeDIP-seq and MBD-cap seq differ in their coverage of genomic regions and DMRs with MBD-cap seq detecting twice as many DMRs at comparable sequencing depths. Both methods have low statistical power to detect DMRs in CpG-poor genomic regions (reviewed in Yong et al. 2016).

9.4.6 Next-Generation Sequencing

It is foreseeable that whole genome sequencing of bisulphite-modified DNA (BS-seq) or other sequencing technologies detecting methylation at multiple loci will become the preferred analysis platform for both research interrogation and diagnostic testing of DNA methylation in disease conditions. It is also likely that, for the diagnosis of imprinting disorders, a whole genome or targeted sequencing approach may also be ultimately preferred as the relationships between both *cis*- and *trans*-acting factors affecting imprinting are revealed.

It is important to identify and distinguish heritable imprinting disorders from those arising sporadically. As an example, SNVs affecting OCT4 binding sites in the 11p15.5 *H19/IGF2* IG-DMR are associated with heritable forms of Beckwith-Wiedemann syndrome (Poole et al. 2012; Berland et al. 2013), and a haplotype affecting *KCNQ1*, when maternally inherited, is associated with BWS with loss of methylation at *KCNQ1OT1:* TSS-DMR (Demars et al. 2014). A technology that combines both SNV and methylation analysis in a single assay would be immensely powerful in this respect. Insertions and deletions within or *cis*-acting on imprinting centres have also been described in several imprinting disorders including Beckwith-Wiedemann syndrome, Russell-Silver syndrome, Prader-Willi syndrome and transient neonatal diabetes, as well as single gene mutations in *trans*-acting genes that underlie the imprinting defect (Rezwan et al. 2015; Bak et al. 2016). The capacity to detect all imprinting abnormalities and any associated SNVs has the potential to lead to new classifications within these disorders.

Several challenges have been noted as potential obstacles to obtaining and analysing reliable data from BS-seq (Zhang and Jeltsch 2010; Krueger and Andrews 2011; Yong et al. 2016). Bisulphite conversion of the DNA leads to decreased sequence complexity and increased redundancy as the representation of cytosine and thymidine in the converted genome changes following bisulphite treatment. In

addition the sequence of the target DNA after conversion cannot always be predicted, complicating the library design for DNA capture. Bisulphite-converted DNA is also unstable. However several design modifications to library preparation have yielded improved outcomes that have compared favourably against other genome-wide methods for methylation analysis, although overall coverage remains low (Ball et al. 2009; Ivanov et al. 2013; Deng et al. 2009; Hodges et al. 2009; Komori et al. 2011).

One of the challenging aspects of measuring methylation in disease conditions is its variability across tissue types and susceptibility to environmental conditions. However, these challenges are unlikely to be as significant for the investigation of genomic imprinting disorders. Genomic imprints are usually remarkably stable within tissues in which the gene in question is imprinted. Difficulties in measuring disrupted imprinting could arise however where the disorder is mosaic as in Beckwith-Wiedemann syndrome, idiopathic hemihypertrophy and Russell-Silver syndrome; however, the enhanced sensitivity of a sequencing approach with high read-depth could potentially lead to improved diagnostic yields for these conditions and is likely in the future to out-perform technologies in current use. As with all diagnostic approaches for analysis of imprinting disorders, a normal reference range must be rigorously established in reference samples for appropriate interpretation.

9.5 Summary and Concluding Remarks

Human imprinting disorders are complex disorders characterized by defined genomic and epigenomic mutations; however, only a handful of these disorders has been sufficiently well characterized to enable reliable diagnostic testing to be performed. New imprinting disorders are likely to be identified from the integration of RNA-seq and genome-wide methylation data in defined clinical conditions. With these developments, new testing approaches will emerge, and the molecular ascertainment of these conditions will lead to new insights into their underlying aetiology.

References

Abu-Amero S, Monk D, Frost J, Preece M, Stanier P, Moore GE (2008) The genetic aetiology of Silver-Russell syndrome. J Med Genet 45(4):193–199

Alders M, Bliek J, vd Lip K, vd Bogaard R, Mannens M (2009) Determination of KCNQ1OT1 and H19 methylation levels in BWS and SRS patients using methylation-sensitive high-resolution melting analysis. Eur J Hum Genet 17(4):467–473

Algar E, Dagar V, Sebaj M, Pachter N (2011) An 11p15 imprinting centre region 2 deletion in a family with Beckwith-Wiedemann syndrome provides insights into imprinting control at CDKN1C. PLoS One 6(12):e29034

Algar EM, St Heaps L, Darmanian A, Dagar V, Prawitt D, Peters GB, Collins F (2007) Paternally inherited submicroscopic duplication at 11p15.5 implicates insulin-like growth factor II in overgrowth and Wilms' tumorigenesis. Cancer Res 67(5):2360–2365

Azzi S, Rossignol S, Steunou V, Sas T, Thibaud N, Danton F, Le Jule M, Heinrichs C, Cabrol S, Gicquel C, Le Bouc Y, Netchine I (2009) Multilocus methylation analysis in a large cohort of

11p15-related foetal growth disorders (Russell-Silver and Beckwith-Wiedemann syndromes) reveals simultaneous loss of methylation at paternal and maternal imprinted loci. Hum Mol Genet 18:4724–4733

Azzi S, Steunou V, Rousseau A, Rossignol S, Thibaud N, Danton F, Le Jule M, Gicquel C, Le Bouc Y, Netchine I (2011) Allele-specific methylated multiplex real-time quantitative PCR (ASMM RTQ-PCR), a powerful method for diagnosing loss of imprinting of the 11p15 region in Russell-Silver and Beckwith-Wiedemann syndromes. Hum Mutat 32(2):249–258

Bak M, Boonen SE, Dahl C, Hahnemann JM, Mackay DJDG, Tumer Z, Gronskov K, Temple IK, Guldberg P, Tommerup N (2016) Genome-wide DNA methylation analysis of transient neonatal diabetes type 1 patients with mutations in ZFP57. BMC Med Genet 17:29

Ball MP, Li JB, Gao Y, Lee JH, LeProust EM, Park IH, Xie B, Daley GQ, Church GM (2009) Targeted and genome-scale strategies reveal gene-body methylation signatures in human cells. Nat Biotechnol 27(4):361–368

Berland S, Appelback M, Bruland O, Beygo J, Buiting K, Mackay DJ, Karen Temple I, Houge G (2013) Evidence for anticipation in Beckwith-Wiedemann syndrome. Eur J Hum Genet 21 (12):1344–1348

Brinkman AB, Simmer F, Ma K, Kaan A, Zhu J, Stunnenberg HG (2010) Whole-genome DNA methylation profiling using MethylCap-seq. Methods 52(3):232–236

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

Buiting K, Di Donato N, Beygo J, Bens S, von der Hagen M, Hackmann K, Horsthemke B (2014) Clinical phenotypes of MAGEL2 mutations and deletions. Orphanet J Rare Dis 9:40. https://doi.org/10.1186/1750-1172-9-40

Buiting K, Kanber D, Martin-Subero JI, Lieb W, Terhal P, Albrecht B, Purmann S, Gross S, Lich C, Siebert R, Horsthemke B, Gillessen-Kaesbach G (2008) Clinical features of maternal uniparental disomy 14 in patients with an epimutation and a deletion of the imprinted DLK1/GTL2 gene cluster. Hum Mutat 29(9):1141–1146

Cardoso LC, Tenorio Castano JA, Pereira HS, Lima MA, Dos Santos AC, de Faria PS, Ferman S, Seuanez HN, Nevado JB, de Almeida JC, Lapunzina P, Vargas FR (2012) Constitutional and somatic methylation status of DMRH19 and KvDMR in Wilms tumor patients. Genet Mol Biol 35(4):714–724

Charlton J, Irtan S, Bergeron C, Pritchard-Jones K (2017) Bilateral Wilms tumour: a review of clinical and molecular features. Expert Rev Mol Med 19:e8

Dedeurwaerder S, Defrance M, Calonne E, Denis H, Sotiriou C, Fuks F (2011) Evaluation of the infinium methylation 450K technology. Epigenomics 3(6):771–784

Demars J, Shmela ME, Khan AW, Lee KS, Azzi S, Dehais P, Netchine I, Rossignol S, Le Bouc Y, El-Osta A, Gicquel C (2014) Genetic variants within the second intron of the KCNQ1 gene affect CTCF binding and confer a risk of Beckwith-Wiedemann syndrome upon maternal transmission. J Med Genet 51(8):502–511

Deng J, Shoemaker R, Xie B, Gore A, LeProust EM, Antosiewicz-Bourget J, Egli D, Maherali N, Park IH, Yu J, Daley GQ, Eggan K, Hochedlinger K, Thomson J, Wang W, Gao Y, Zhang K (2009) Targeted bisulfite sequencing reveals changes in DNA methylation associated with nuclear reprogramming. Nat Biotechnol 27(4):353–360

Docherty LE, Rezwan FI, Poole RL, Jagoe H, Lake H, Lockett GA, Arshad H, Wilson DI, Holloway JW, Temple IK, Mackay DJG (2014) Genome-wide DNA methylation analysis of patients with imprinting disorders identifies differentially methylated regions associated with novel candidate imprinted genes. J Med Genet 51(4):229–238

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

Eggermann K, Bliek J, Brioude F, Algar E, Buiting K, Russo S, Tumer Z, Monk D, Moore G, Antoniadi T, Macdonald F, Netchine I, Lombardi P, Soellner L, Begemann M, Prawitt D, Maher ER, Mannens M, Riccio A, Weksberg R, Lapunzina P, Gronskov K, Mackay DJ, Eggermann T (2016) EMQN best practice guidelines for the molecular genetic testing and reporting of chromosome 11p15 imprinting disorders: Silver-Russell and Beckwith-Wiedemann syndrome. Eur J Hum Genet 24(10):1377–1387

Grafodatskaya D, Choufani S, Basran R, Weksberg R (2017) An update on molecular diagnostic testing of human imprinting disorders. J Pediatr Genet 6(1):3–17

Hodges E, Smith AD, Kendall J, Xuan Z, Ravi K, Rooks M, Zhang MQ, Ye K, Bhattacharjee A, Brizuela L, McCombie WR, Wigler M, Hannon GJ, Hicks JB (2009) High definition profiling of mammalian DNA methylation by array capture and single molecule bisulfite sequencing. Genome Res 19(9):1593–1605

Ioannides Y, Lokulo-Sodipe K, Mackay DJ, Davies JH, Temple IK (2014) Temple syndrome: improving the recognition of an underdiagnosed chromosome 14 imprinting disorder: an analysis of 51 published cases. J Med Genet 51(8):495–501

Ivanov M, Kals M, Kacevska M, Metspalu A, Ingelman-Sundberg M, Milani L (2013) In-solution hybrid capture of bisulfite-converted DNA for targeted bisulfite sequencing of 174 ADME genes. Nucleic Acids Res 41(6):e72

McGrath J, Solter D (1984) Completion of mouse embryogenesis requires both the maternal and paternal genomes. Cell 37:179–183

Kagami M, Nishimura G, Okuyama T, Hayashidani M, Takeuchi T, Tanaka S, Ishino F, Kurosawa K, Ogata T (2005) Segmental and full paternal isodisomy for chromosome 14 in three patients: narrowing the critical region and implication for the clinical features. Am J Med Genet A 138a(2):127–132

Komori HK, LaMere SA, Torkamani A, Hart GT, Kotsopoulos S, Warner J, Samuels ML, Olson J, Head SR, Ordoukhanian P, Lee PL, Link DR, Salomon DR (2011) Application of microdroplet PCR for large-scale targeted bisulfite sequencing. Genome Res 21(10):1738–1745

Krueger F, Andrews SR (2011) Bismark: a flexible aligner and methylation caller for Bisulfite-Seq applications. Bioinformatics 27(11):1571–1572

Lan X, Adams C, Landers M, Dudas M, Krissinger D, Marnellos G, Bonneville R, Xu M, Wang J, Huang THM, Meredith G, Jin VX (2011) High resolution detection and analysis of CpG dinucleotides methylation using MBD-Seq technology. PLoS One 6(7):e22226

Liu D, Wang Y, Yang XA, Liu D (2017) De novo mutation of paternal IGF2 gene causing Silver-Russell Syndrome in a sporadic patient. Front Genet 8:105

Surani MA, Barton SC (1983) Development of gynogenetic eggs in the mouse: implications for parthenogenetic embryos. Science 222:1034–1036

Mackay DJ, Callaway JL, Marks SM, White HE, Acerini CL, Boonen SE, Dayanikli P, Firth HV, Goodship JA, Haemers AP (2008) Hypomethylation of multiple imprinted loci in individuals with transient neonatal diabetes is associated with mutations in ZFP57. Nat Genet 40:949–951

McDonald JF, Matzke MA, Matzke AJ (2005) Host defenses to transposable elements and the evolution of genomic imprinting. Cytogenet Genome Res 110(1-4):242–249

Mertz LG, Christensen R, Vogel I, Hertz JM, Nielsen KB, Gronskov K, Ostergaard JR (2013) Angelman syndrome in Denmark. birth incidence, genetic findings, and age at diagnosis. Am J Med Genet A 161a(9):2197–2203

Moore T, Haig D (1991) Genomic imprinting in mammalian development: a parental tug-of-war. Trends Genet 7(2):45–49

Moore T, Reik W (1996) Genetic conflict in early development: parental imprinting in normal and abnormal growth. Rev Reprod 1(2):73–77

Murphy SK, Huang Z, Hoyo C (2012) Differentially methylated regions of imprinted genes in prenatal, perinatal and postnatal human tissues. PLoS One 7(7):e40924

Mussa A, Russo S, De Crescenzo A, Chiesa N, Molinatto C, Selicorni A, Richiardi L, Larizza L, Silengo MC, Riccio A, Ferrero GB (2013) Prevalence of Beckwith-Wiedemann syndrome in North West of Italy. Am J Med Genet A 161a(10):2481–2486

Niemitz EL, DeBaun MR, Fallon J, Murakami K, Kugoh H, Oshimura M, Feinberg AP (2004) Microdeletion of LIT1 in familial Beckwith-Wiedemann syndrome. Am J Hum Genet 75 (5):844–849

Nygren AO, Ameziane N, Duarte HM, Vijzelaar RN, Waisfisz Q, Hess CJ, Schouten JP, Errami A (2005) Methylation-specific MLPA (MS-MLPA): simultaneous detection of CpG methylation and copy number changes of up to 40 sequences. Nucleic Acids Res 33(14):e128

Nyren P (1987) Enzymatic method for continuous monitoring of DNA polymerase activity. Anal Biochem 167(2):235–238

Poole RL, Leith DJ, Docherty LE, Shmela ME, Gicquel C, Splitt M, Temple IK, Mackay DJ (2012) Beckwith-Wiedemann syndrome caused by maternally inherited mutation of an OCT-binding motif in the IGF2/H19-imprinting control region, ICR1. Eur J Hum Genet 20(2):240–243

Ramsden SC, Clayton-Smith J, Birch R, Buiting K (2010) Practice guidelines for the molecular analysis of Prader-Willi and Angelman syndromes. BMC Med Genet 11:70

Rezwan FI, Poole RL, Prescott T, Walker JM, Karen Temple I, Mackay DJG (2015) Very small deletions within the NESP55 gene in pseudohypoparathyroidism type 1b. Eur J Hum Genet 23 (4):494–499

Riccio A, Sparago A, Verde G, De Crescenzo A, Citro V, Cubellis MV, Ferrero GB, Silengo MC, Russo S, Larizza L, Cerrato F (2009) Inherited and sporadic epimutations at the IGF2-H19 locus in Beckwith-Wiedemann syndrome and Wilms' tumor. Endocr Dev 14:1–9

Romanelli V, Belinchon A, Benito-Sanz S, Martinez-Glez V, Gracia-Bouthelier R, Heath KE, Campos-Barros A, Garcia-Minaur S, Fernandez L, Meneses H, Lopez-Siguero JP, Guillen-Navarro E, Gomez-Puertas P, Wesselink JJ, Mercado G, Esteban-Marfil V, Palomo R, Mena R, Sanchez A, Del Campo M, Lapunzina P (2010) CDKN1C (p57(Kip2)) analysis in Beckwith-Wiedemann syndrome (BWS) patients: genotype-phenotype correlations, novel mutations, and polymorphisms. Am J Med Genet A 152A(6):1390–1397

Sadikovic B, Fernandes P, Zhang VW, Ward PA, Miloslavskaya I, Rhead W, Rosenbaum R, Gin R, Roa B, Fang P (2014) Mutation update for UBE3A variants in Angelman syndrome. Hum Mutat 35(12):1407–1417

Schaaf CP, Gonzalez-Garay M, Xia F, Potocki L, Gripp KW, Zhang B, Peters BA, McElwain MA, Drmanac R, Beaudet AL, Caskey CT, Yang Y (2013) Truncating mutations of MAGEL2 cause Prader-Willi phenotypes and autism. Nat Genet 45:1405–1408. https://doi.org/10.1038/ng.2776

Scott RH, Douglas J, Baskcomb L, Huxter N, Barker K, Hanks S, Craft A, Gerrard M, Kohler JA, Levitt GA, Picton S, Pizer B, Ronghe MD, Williams D, Cook JA, Pujol P, Maher ER, Birch JM, Stiller CA, Pritchard-Jones K, Rahman N (2008) Constitutional 11p15 abnormalities, including heritable imprinting center mutations, cause nonsyndromic Wilms tumor. Nat Genet 40 (11):1329–1334

Shield JP (2000) Neonatal diabetes: new insights into aetiology and implications. Horm Res 53 (Suppl 1):7–11

Smith A, Egan J, Ridley G, Haan E, Montgomery P, Williams K, Elliott E (2003) Birth prevalence of Prader-Willi syndrome in Australia. Arch Dis Child 88(3):263–264

Turan S, Bastepe M (2015) GNAS spectrum of disorders. Curr Osteoporos Rep 13(3):146–158

Tycko B (2010) Allele-specific DNA methylation: beyond imprinting. Hum Mol Genet 19(R2): R210–R220

White HE, Hall VJ, Cross NC (2007) Methylation-sensitive high-resolution melting-curve analysis of the SNRPN gene as a diagnostic screen for Prader-Willi and Angelman syndromes. Clin Chem 53(11):1960–1962

Wojdacz TK, Dobrovic A, Algar EM (2008) Rapid detection of methylation change at H19 in human imprinting disorders using methylation-sensitive high-resolution melting. Hum Mutat 29 (10):1255–1260

Woodfine K, Huddleston JE, Murrell A (2011) Quantitative analysis of DNA methylation at all human imprinted regions reveals preservation of epigenetic stability in adult somatic tissue. Epigenetics Chromatin 4(1):1

Yong WS, Hsu FM, Chen PY (2016) Profiling genome-wide DNA methylation. Epigenetics Chromatin 9:26

Zhang Y, Jeltsch A (2010) The application of next generation sequencing in DNA methylation analysis. Genes (Basel) 1(1):85–101

Zhi J, Hatchwell E (2008) Human MLPA Probe Design (H-MAPD): a probe design tool for both electrophoresis-based and bead-coupled human multiplex ligation-dependent probe amplification assays. BMC Genomics 9:407

Cancer Methylation Biomarkers in Circulating Cell-Free DNA

10

Ruth Pidsley and Clare Stirzaker

Abstract

Screening and diagnosing tumours through circulating cell-free DNA represents an important *paradigm shift* in precision medicine. Molecular profiling of cell-free DNA in human blood serves as a 'liquid biopsy' which, in contrast to tissue biopsies, provides a minimally invasive method for predictive and prognostic marker detection, as well as early and serial assessment of metastatic disease, including follow-up during remission, characterizing treatment response, and monitoring minimal residual disease. Changes in DNA methylation are one of the earliest, most stable and frequent alterations in cancer genomes. Therefore, DNA methylation signatures have great potential as molecular biomarkers to guide clinical management of disease in many tumour types. The detection of DNA methylation signatures in cell-free DNA is an important advance in the clinical utility of liquid biopsy for diagnosis, prognosis, monitoring treatment response and tumour burden.

Keywords

DNA methylation · Epigenetics · Biomarkers · Circulating DNA · cfDNA · Cancer · Early detection · Diagnosis

R. Pidsley · C. Stirzaker (✉)
Epigenetics Research Laboratory, Garvan Institute of Medical Research, Sydney, NSW, Australia

St Vincent's Clinical School, UNSW Sydney, Sydney, NSW, Australia
e-mail: c.stirzaker@garvan.org.au

L. B. Hesson, A. L. Pritchard (eds.), *Clinical Epigenetics*,
https://doi.org/10.1007/978-981-13-8958-0_10

10.1 Introduction

Plasma and serum contain small fragments of DNA released into the circulating blood, which is not associated with any blood cell type. This circulating cell-free DNA (cfDNA) is generally shed from normal cells, including leukocytes; however, in patients with cancer, a proportion of cfDNA is derived from tumour cells, i.e. circulating tumour DNA (ctDNA). cfDNA (which includes ctDNA) can be extracted from both plasma and serum (Fig. 10.1). However, serum typically yields higher amounts of DNA, and there is evidence that this is derived from the lysis of leukocytes during serum processing in the laboratory rather than reflecting an actual greater amount of cfDNA (Warton et al. 2014; Lee et al. 2001). In contrast, plasma is less impacted by the lysis of leukocytes and sample processing artefacts, and therefore, blood plasma is considered a superior source of cfDNA. cfDNA isolated from plasma is usually <170–500 base-pairs (bp), mostly corresponding to <170 bp mono-nucleosomal and <300 bp di-nucleosomal DNA fragments. Levels of cfDNA are generally very low in healthy individuals, ranging between <0 and 50 ng/ml. While increasing concentrations of cfDNA are frequently present in cancer patients due to increased ctDNA, the quantification of cfDNA concentrations *alone* is not currently viewed as a useful diagnostic method, owing to the wide range of cfDNA

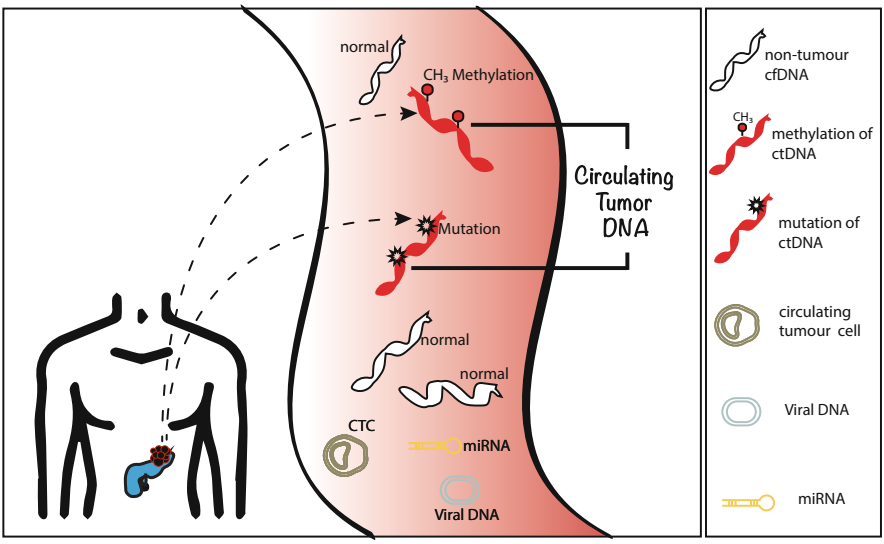

Fig. 10.1 Circulating cell-free nucleic acids in the blood. Plasma and serum contain small fragments of DNA that are released into the blood. This circulating-free DNA (cfDNA) is generally shed from normal cells by apoptosis or necrosis; however, in patients with cancer, a proportion of cfDNA is derived from tumour cells, known as circulating tumour DNA (ctDNA). Tumour-specific alterations, such as genetic mutations, or epigenetic abnormalities, including DNA methylation, can be detected in ctDNA. Viral DNA, circulating tumour cells (CTCs), and miRNA may also be detected in blood. (Illustrated by Madhavi Maddugoda)

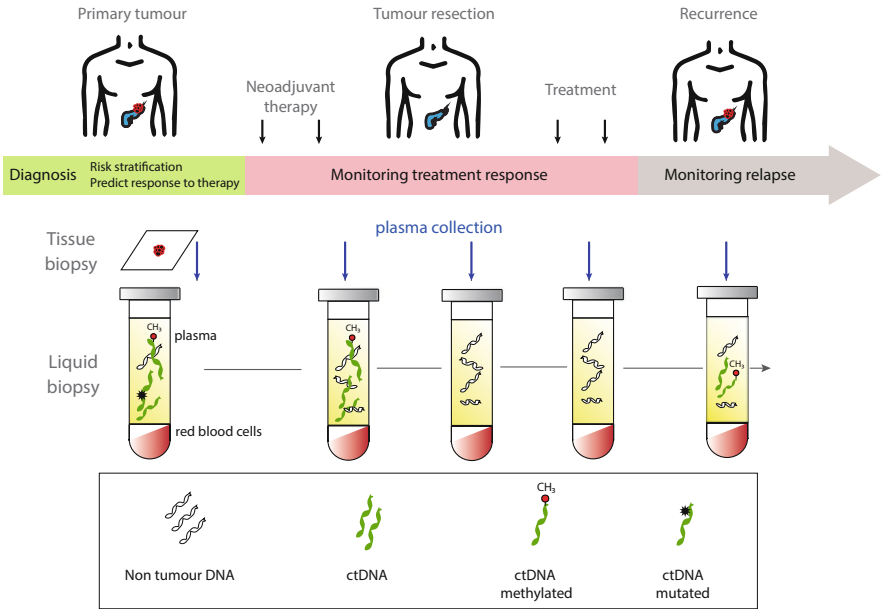

Fig. 10.2 Clinical utility of cell-free tumour DNA. Cancer-specific DNA methylation biomarkers can be assessed by tissue or liquid biopsy for diagnosis. Methylation biomarkers may provide useful clinical information for prognosis of risk or for predicting response to therapy. Following tumour resection, the presence of methylated cancer-specific DNA methylated biomarkers can be assessed to indicate the presence of residual tumour cells. During the course of treatment, tumour-specific cfDNA methylation can continue to be assessed by liquid biopsy to monitor response to therapy and the early detection of disease relapse and resistance. (lllustrated by Madhavi Maddugoda)

concentrations that are found in healthy individuals, which can overlap those in patients with malignant disease.

Molecular profiling of tumours is central to the clinical management of many cancer patients. However, obtaining tumour material requires invasive intervention, and, moreover, a tumour biopsy only provides a snapshot of any molecular aberrations present and may not reflect the complete heterogeneous molecular profile of the tumour. In addition to tumour heterogeneity, tumours evolve over the course of the disease and treatment, and continuous monitoring of tumour-specific changes by tissue biopsies is not practical. ctDNA not only contains the same mutations as primary tumour cells but also reflects the same tumour-specific DNA methylation patterns (Schwarzenbach et al. 2011), representing the entire tumour genome. In the last few years, it has been shown that the molecular landscape of a tumour and its metastases is reflected in ctDNA and can inform the efficacy and relevance of a chosen treatment specific for a molecular target or indicate the emergence of treatment resistance, before it is clinically obvious (Murtaza et al. 2015). There is therefore widespread interest in the benefit of cfDNA analyses, also known as 'liquid biopsies', for the diagnosis and monitoring of disease for enhanced clinical management of patients (Krishnamurthy et al. 2017; Dawson et al. 2013) (Fig. 10.2).

10.2 DNA Methylation Biomarkers

DNA methylation is one of the earliest, most stable and frequent alterations in cancer genomes. Importantly, atypical DNA methylation can have a similar function to genetic mutations or deletions. During the initiation and progression of cancer, genome-wide changes in DNA methylation patterns are typically observed, characterized by DNA hypermethylation of gene regulatory regions and CpG islands, associated primarily with the silencing of tumour suppressor genes. In parallel, DNA hypomethylation of adjacent genic and CpG-poor intergenic regions leads to the activation and expression of oncogenes (Jones and Baylin 2007). The advancement of genome-wide DNA methylation sequencing technologies has allowed for comprehensive mapping of cancer methylomes, providing greater insight into the underlying mechanisms and location of cancer-specific methylation changes (Stirzaker et al. 2014). Since DNA methylation changes are common in cancer, DNA methylation signatures are being reviewed as potential molecular biomarkers of disease. Moreover, DNA methylation provides a stable mark, and the tools to detect methylated DNA in clinical samples are well developed.

In tumour cells, molecular alterations such as DNA mutations, copy number aberrations, deletions and insertions are reflected in ctDNA released from the tumour into the blood. Mutations driving tumour development and progression, present in a wide range of oncogenes and tumour suppressor genes, can potentially be spread across any part of a commonly mutated gene, making evaluation difficult. In contrast to gene mutations, aberrant DNA methylation associated with the cancer phenotype tends to occur in specific regulatory regions of the DNA (e.g. gene promoters and enhancers) that can be easily targeted for measurement. Assessment of epigenetic alterations such as DNA methylation in cfDNA will also be useful for diagnosing or monitoring those diseases that do not have reliable genetic biomarkers. Like genetic alterations, DNA methylation patterns are retained in the released ctDNA, making detection of tumour-specific DNA methylation in patient plasma a highly feasible approach for the development of a blood-based test.

In this chapter, we review the advances in DNA methylation ctDNA biomarkers in common cancers, including colon, breast, prostate, lung, brain, pancreatic and ovarian cancers. We examine recent studies that suggest that methylation signatures in cfDNA can determine the tissue-of-origin of tumours (Lehmann et al. 2011; Snyder et al. 2016). We discuss the clinical utility of methylation ctDNA biomarkers. Finally, we discuss the challenges and future perspectives in implementing these assays into a clinical setting.

10.3 Methylated DNA Blood-Based Biomarkers

10.3.1 Colorectal Cancer

Colorectal cancer (CRC) is the third most commonly diagnosed cancer and cause of cancer-related death in both men and women (Siegel et al. 2014). Tests for the presence of colorectal cancer in the form of colonoscopy and the faecal occult blood test (FOBT) have existed for over 50 years and are an illustrative example of the enormous benefit of screening for early detection in reducing cancer mortality (Shaukat et al. 2013; Nishihara et al. 2013). However, the sensitivity of a single stool test using FOBT is as low as 10–30%; therefore three tests are recommended to increase sensitivity to acceptable levels (>90%). Furthermore, the effectiveness of colonoscopies can be variable and is dependent on patient compliance, the quality of bowel preparation, the sophistication of equipment used and the expertise of gastroendoscopists in recognizing more unusual precursor lesions, including flat adenomas. It is thought that these variables may contribute to the incidence of interval cancers (those that arise between colonoscopies). The above limitations in current screening strategies highlight the need for new screening options for improved sensitivity and specificity as alternatives to faecal screening and colonoscopy to encourage further uptake (Scholefield et al. 2012). To this end, DNA methylation biomarkers are promising tools that offer an alternative approach for early cancer detection.

Numerous studies have been performed to identify and assess methylation-based biomarkers that differentiate control blood from patients with colorectal cancer (CRC), through whole-genome screening (Lange et al. 2012; Pedersen et al. 2014; Takane et al. 2014), as well as candidate gene testing approaches (Cassinotti et al. 2012; Lee et al. 2009; Pack et al. 2013). These studies have identified methylated promoters of genes with a well-established role in CRC tumour progression, such as *RASSF1A, RB1, APC*, *E-cadherin, MGMT* and *RASSF2A* (Cassinotti et al. 2012; Lee et al. 2009; Pack et al. 2013; Hesson et al. 2005) as well as novel methylated genes such as *THBD* and *C9orf50* (Lange et al. 2012), *CAHM*, a long-non-coding RNA gene (Pedersen et al. 2014), and *USP44,* a candidate tumour suppressor gene (Sloane et al. 2014). Many studies using candidate methylated genes have shown excellent sensitivity and specificity in blood plasma; for example, methylation of *PPP1R3C* and *EFHD1* in a cohort of 120 CRC plasma samples and 96 control patients showed a high sensitivity of 90% and specificity of 96% (Takane et al. 2014). Similarly, promoter methylation of *THBD* in pre-therapeutic plasma and serum samples was able to differentiate 107 CRC patients and 98 normal controls with a sensitivity of 71% and a specificity of 80% (Lange et al. 2012).

The most rigorously studied epigenetic mark to date is the methylated *SEPT9* gene promoter, first identified as being differentially methylated in a discovery project comparing CRC tissue and non-pathological tissue (Lofton-Day et al. 2008). In order to specifically develop a blood-based test, there was a stringent requirement at the discovery phase that *SEPT9* showed minimal methylation in leukocytes (Lofton-Day et al. 2008). Following the initial discovery of *SEPT9*

methylation as a potential biomarker, evaluation was performed in a number of retrospective trials comparing plasma cfDNA methylation in healthy controls and colorectal cancer patients (Grutzmann et al. 2008; deVos et al. 2009; Warren et al. 2011). These results were promising with sensitivity ranging between 72 and 90% and specificity of 88–90%. A large prospective study, the PRESEPT trial, enrolled over 7900 participants (Church et al. 2014); this showed a sensitivity of 63.9% and a specificity of 88.4% (Warren et al. 2011; Church et al. 2014). Overall, these studies were significant in demonstrating the potential utility of a methylated ctDNA biomarker, from initial discovery into clinical testing for CRC screening.

10.3.2 Breast Cancer

Detection of breast cancer at an early stage, predicting outcome, monitoring response to therapy and detecting disease relapse are all key to improving the outcomes for breast cancer patients. Numerous DNA methylation studies on cfDNA have shown the potential of methylated biomarkers in early diagnosis and in the follow-up of breast cancer patients; however, a reproducible blood-based test has yet to be incorporated into routine clinical testing. One of the confounding factors is that the detection rates of DNA methylation markers in blood are low, even in advanced disease.

For early-stage breast cancer detection, Shan et al. reported a six-gene epigenetic panel (*SFN*, *CDKN2A* (p16INK4A), *MLH1*, *HOXD13*, *PCDHGB7* and *RASSF1A*); using a MethyLight assay, they examined serum from 749 breast cancer cases. The six-panel test achieved 79.6 and 82.4% sensitivity with a specificity of 72.4 and 78.1% in diagnosis of breast cancer compared with healthy and benign disease controls, respectively (Shan et al. 2016). To enhance the sensitivity of detecting methylated DNA for early detection, Uehiro et al. used droplet digital methylation-specific PCR (ddMSP) and showed that a four-gene methylation marker panel (*RASGRF1*, *CPXM1*, *HOXA10* and *DACH1*) could detect Stage 0–I breast cancer with a sensitivity of 84.6% and specificity of 82.7%, comparable with that of mammography screening (Uehiro et al. 2016).

In Stage IV metastatic breast cancer patients, Fackler et al. used a ten-gene panel (*AKR1B1, COL6A2*, *GPX7*, *HIST1H3C*, *HOXB4*, *RASGRF2*, *TM6SF1*, *ARHGEF7*, *TMEFF2* and *RASSF1*) and a quantitative multiplex methylation-specific PCR assay (cMethDNA) and showed that methylated DNA could be detected in a test (28 normal, 24 cancer) and training cohort of patient samples (27 normal, 33 cancer) with a sensitivity of 91% and a specificity of 96% (Fackler et al. 2014). In a further pilot study, the cMethDNA assay also reflected patient response to chemotherapy ($n = 29$), demonstrating the utility of the cMethDNA assay in cfDNA in metastatic breast cancer to monitor tumour burden and treatment response (Fackler et al. 2014). In a larger prospective study (141 breast cancer patients), Visvanathan et al. reported that six genes from the above ten-gene panel (*AKR1B1, HOXB4, RASGRF2, RASSF1, HIST1H3C* and *TM6SF1*) were a strong predictor of survival outcomes in metastatic breast cancer and useful in disease stratification and disease monitoring

(Visvanathan et al. 2017). In another study, methylation of *RASSF1* and *APC* in pretreatment sera of breast cancer patients was associated with a worse outcome and was more powerful than standard prognostic parameters (Muller et al. 2003). Other genes that have been shown to be useful for monitoring response to therapy include *BRCA1* (Sharma et al. 2012), *STRATIFIN* (Zurita et al. 2010), *RASSF1A* and *NEUROD1* (Fiegl et al. 2008, 2005). Numerous other methylated genes including *CDKN2A* (p16INK4A), *CDH1*, *DAPK11*, *HIC1*, *RARB*, *CDH13*, *ESR1* and *GSTP1* [reviewed in (Shivapurkar and Gazdar 2010)] and *BRCA1*, *ATM*, *IGF2*, *CDH1*, *ESR1*, *SYK*, *TIMP3*, *RARB*, *APC* and *RASSF1A* [reviewed in (Tang et al. 2016)] have been evaluated as methylation biomarkers in serum. Thus, to date, there have been a number of studies supporting the utility of ctDNA methylated biomarkers for both early detection and disease monitoring in breast cancer but using relatively small sample numbers. Promising candidate biomarkers now need to be validated in large screening populations by high-quality studies to allow translation of methylated biomarkers into a clinical setting.

10.3.3 Prostate Cancer

Prostate cancer is the second most common cancer and fifth most common cause of cancer death in men worldwide (Ferlay et al. 2015). Currently, diagnosis is achieved through assessment of blood prostate-specific antigen (PSA) levels, digital rectal examination (clinical T-stage) and histological examination of needle biopsies (Gleason grade). However, there is controversy over serum PSA testing as it has a poor specificity for prostate cancer (Roddam et al. 2005), which can lead to unnecessary biopsies and treatment. Needle biopsies also carry a risk of infection (Nam et al. 2010) or can miss the site of the tumour (You et al. 2016). Thus blood-based DNA methylation biomarkers are required to improve the accuracy and safety of prostate cancer diagnosis.

GSTP1 promoter hypermethylation is one of the earliest molecular alterations in the development of prostate cancer and is evident in the majority (>90%) of prostate tumour specimens (Jeronimo et al. 2002). It can also be detected in the bodily fluids of prostate cancer patients such as urine, semen and blood serum and plasma samples (Wu et al. 2011). Indeed, a recent meta-analysis of *GSTP1* hypermethylation in plasma and serum samples revealed a specificity of 96% (90%) and sensitivity of 42% (36%) for prostate cancer detection for MSP and (non-MSP) methods (Wu et al. 2011).

Several studies have shown the prognostic utility of *GSTP1* methylation in circulating cfDNA (Bastian et al. 2005; Mahon et al. 2014; Reibenwein et al. 2007). For example, a recent study from our group demonstrated that *GSTP1* methylation in cfDNA from plasma was associated with overall survival and response to chemotherapy in men with advanced prostate cancer (Mahon et al. 2014). Indeed, *GSTP1* methylation levels, both prior to and after one chemotherapy cycle, were stronger predictors of overall survival than changes in PSA levels at three months post-chemotherapy.

Other genes have also shown promise as cfDNA methylation biomarkers in prostate cancer (Horning et al. 2015; Okegawa et al. 2010; Brait et al. 2017; Deng et al. 2016; Wang et al. 2014). Horning et al. examined genes involved in the androgen biosynthesis pathway in patients who had their prostates removed due to high-risk prostate cancer (radical prostatectomy, RP). They found an association between promoter hypermethylation of *SRD5A2* and *CYP11A1* in cfDNA with biochemical recurrence (PSA level rise) after RP (Horning et al. 2015). In another study, cfDNA methylation at the promoter or regulatory regions of the *APC*, *GSTP1*, *RASSF1A*, *MDR1* and *PTGS2* genes was associated with overall survival time in men with advanced prostate cancer (Okegawa et al. 2010). Interestingly this association was strongest in the patients with detectable circulating tumour cells whose lysis is thought to be one of the main sources of ctDNA. Finally, a study of early-stage prostate cancer patients showed that a methylation biomarker panel combining *MCAM*, *ESR1* (ERα) and *ESR2* (ERβ) was able to detect prostate cancer from serum cfDNA with 75% sensitivity and 70% sensitivity (Brait et al. 2017).

10.3.4 Ovarian Cancer

Ovarian cancer is the most lethal gynaecological cancer and the fifth leading cause of cancer death in women. To date, the most well-studied and long-utilized biomarker for ovarian cancer is CA-125, a large glycoprotein that is over-expressed in ovarian cancer cells and secreted into the bloodstream of patients (Bast Jr. et al. 1981; Yin and Lloyd 2001). However, CA-125 alone is not suitable as a diagnostic marker due to its unacceptably low specificity (Duffy et al. 2005).

A number of studies have reported aberrant DNA methylation changes in ovarian cancer (Barton et al. 2008; Earp and Cunningham 2015). Furthermore, many of these tumour-specific methylation changes have been detected in the blood DNA of ovarian cancer patients (Collins et al. 2006; Gifford et al. 2004; Ibanez de Caceres et al. 2004; Liggett et al. 2011). For example, Ibanez De Caceres et al. found hypermethylation of one or both *BRCA1* and *RASSF1A* gene promoters in DNA from 34/50 (68%) tumour tissues, and methylation of *APC*, *CDKN2A* (both p14ARF and p16INK4A) and *DAPK1* genes in DNA from the remaining 16 tumours, increased diagnostic coverage to 100%. Importantly, hypermethylation of the same genes was observed in 41 of 50 (82%) matched plasma and serum DNA samples, including 13/17 of early Stage I disease. No hypermethylation was observed in serum from 40 control women (100% specificity), indicating that this tumour-specific hypermethylation in serum DNA could be used for early detection of ovarian cancer (Ibanez de Caceres et al. 2004). A number of other tumour suppressor genes have been reported to be hypermethylated in cfDNA of epithelial ovarian cancer; these include *MLH1* (Gifford et al. 2004), *HOXA9* (Wu et al. 2007), *SPARC* (Socha et al. 2009), *HIC1* (Tam et al. 2007), *DAPK* (Collins et al. 2006), *SLIT2* (Dong et al. 2012), *OPCML* (Zhou et al. 2014) and *RASSF2A* (Wu et al. 2014).

Small cohort studies have further demonstrated the potential utility of DNA methylation biomarkers in cfDNA, including additional support for *RASSF1A* and

BRCA1. For example, Liggett et al. showed that differential methylation of three promoters (*RASSF1A*, *CALCA* and *EP300*) differentiated ovarian carcinoma and healthy controls with a sensitivity of 90.0% and a specificity of 86.7%. Three different promoters (*BRCA1*, *CALCA* and *CDKN1C*) were able to discriminate benign ovarian disease from healthy controls with a sensitivity of 90.0% and a specificity of 76.7%; and two promoters (*RASSF1A* and *PGR-PROX*) were able to distinguish ovarian cancer and benign ovarian disease, with a sensitivity of 80.0% and a specificity of 73.3% (Liggett et al. 2011). In another study of 87 ovarian cancer patient serum samples, 53 patients with benign ovarian tumours and 62 healthy controls, Zhang et al. used a multiplex MSP assay to show that a seven-gene panel (*APC*, *RASSF1A*, *CDH1*, *RUNX3*, *TFP12*, SFRP5 and *OPCML*) achieved a sensitivity of 85.3% and a specificity of 90.5%, with strikingly higher rates compared with a single CA-125 assay (sensitivity of 56.1% at specificity of 64.1% $p = 0.0036$) (Zhang et al. 2013). In another study of 33 plasma samples from ovarian cancer patients, a five-gene panel, comprised of *BRCA1*, *IC1*, *PAX5*, *PGR-PROX* and *THBS1,* had an 85% sensitivity and 61% specificity for cancer detection (Melnikov et al. 2009a). In a more recent study, Wang et al. used a multiplex-nested MSP-PCR assay with a three-gene panel, *RUNX*, *TFP12* and *OPCML*, on 114 serum cfDNA samples and detected DNA methylation with a sensitivity of 90.1% and a specificity of 91.1% (Wang et al. 2015a). It should be noted that while a number of promising candidate ovarian methylation cfDNA biomarker studies have been reported, these have generally been performed on relatively small numbers of ovarian cancer patient samples (Giannopoulou et al. 2017), and further validation is required.

10.3.5 Lung Cancer

Lung cancer is currently the second most common cancer and a leading cause of cancer-related deaths (Siegel et al. 2014). There is widespread interest in exploring non-invasive biomarker assays as taking tissue from late-stage patients can be dangerous and painful. Early detection represents one of the most promising approaches to reduce the growing burden of this cancer worldwide. While several circulating biomarkers are being investigated for screening of lung cancer, currently the accuracy of detection is not sufficient to eliminate biopsy and histopathology for the final diagnosis (Eggert et al. 2017).

Early studies identified *CDKN2A* promoter methylation in the plasma of lung patients (Bearzatto et al. 2002), and it was the focus of a search for an early diagnostic biomarker. However, many of these studies only had small numbers of healthy controls (Bearzatto et al. 2002) or did not include a cancer-free control group in their analysis (Kurakawa et al. 2001; Ng et al. 2002; An et al. 2002). Subsequent studies showed *CDKN2A* promoter methylation in the plasma of cancer patients with a sensitivity of 25/110 (22%), 21/55 (38%) and 61%, respectively (Wang et al. 2006; Hsu et al. 2007; Zhang et al. 2011). These studies indicate that *CDKN2A* promoter methylation may be useful for lung cancer detection but potentially as part of a biomarker panel rather than a single gene.

Other studies have identified numerous lung cancer-specific differentially methylated gene promoters: for example, Zhang et al. showed that a five-gene panel (*APC*, *RASSF1A*, *CDH13*, *KLK10* and *DLEC1*) achieved a sensitivity of 83.6% and a specificity of 74.0% for cancer diagnosis in 110 Stage I/II non-small-cell lung cancer and 50 cancer-free plasma controls (Zhang et al. 2011); Rykova et al. investigated a five-gene panel (*APC*, *RASSF1A*, *RARB*, *CDH1* and *CDH13*) and showed hypermethylation in cancer blood specimens as a promising marker for early cancer detection (Rykova et al. 2004); Grote et al. identified *CDKN2A* (p16INK4A), *RARB2* and *SEMA3B* as candidate biomarkers for lung cancer (Grote et al. 2005). In a larger, well-powered study of 411 cancer and control samples, *SHOX2* methylation in plasma was shown as a potential biomarker in the diagnosis of lung cancer (Kneip et al. 2011); here DNA methylation could distinguish malignant lung disease and controls at a sensitivity of 60% (95% confidence interval, 53–67%) and a specificity of 90% (95% confidence interval, 84–94%). Interestingly, cancer in patients with Stages II (72%), III (55%) and IV (83%) was detected at a higher sensitivity when compared with Stage I patients. Weiss et al. showed DNA methylation of a two-gene panel, *SHOX2* and *PTGER4*, in three independent case-control studies examining 330 plasma samples, lung cancer and differentiated nonmalignant disease (Weiss et al. 2017). In another study, *HOXD10*, *PAX9*, *PTPRN2* and *STAG3* were identified as the top four methylation markers for lung cancer diagnosis, yielding an AUC of 0.85 (95% CI, 0.72–0.95) (Wielscher et al. 2015). Finally, *SEPT9* methylation was interrogated in lung cancer and was shown to correctly identify 31/70 (44%) of lung cancer samples while only giving a positive signal in 4/100 controls (Powrozek et al. 2014). Thus, many studies have revealed various methylated gene promoters and combinations of genes to be effective in discriminating lung cancer patients from non-cancer controls [reviewed in (Lu et al. 2017)].

One of the confounding factors in the development of a methylated lung cancer biomarker is choosing the appropriate control group. The majority of lung cancer cases are linked to smoking, and smoking itself leads to DNA methylation changes (Ostrow et al. 2013). Some of the cfDNA biomarkers identified have been able to show differences in specificity and correctly identify negative samples, compared to both smoker and non-smoker controls (Ostrow et al. 2010). Long-term follow-up of these control subjects will reveal if positive methylation signals in the smoker control group are false positives or early methylation changes in the initiation of lung cancer.

Methylated ctDNA biomarkers are also being interrogated for their utility in prognosis, monitoring treatment response and early detection of recurrence. Schmidt et al. showed that the levels of methylated *SHOX2* correlated with a response to therapy; there was also a strong relationship between survival and plasma methylated *SHOX2* (*mSHOX2*) value ($p < 0.001$) (hazard ratio 11.08), suggesting m*SHOX2* can be used as a predictive biomarker (Schmidt et al. 2015). Ponomaryova et al. used a quantitative measure of *RASSF1A* and *RARB2* methylation to demonstrate that methylation could differentiate lung cancer patients and healthy subjects with 87% sensitivity and 75% specificity. Importantly they also found that it was

indicative of treatment response, as an increase in methylation during the follow-up period was associated with disease relapse (Ponomaryova et al. 2013). In another study, methylation of the doublecortin-like kinase (*DCLK1*) gene promoter region correlated with a lower survival ($p = <0.001$, HR = 4.235), suggesting its prognostic value in lung cancer patients (Powrozek et al. 2016). Other studies have shown that breast cancer metastasis suppressor-1 promoter methylation (*BRMS1*) (Balgkouranidou et al. 2014) and *SOX17* methylation in ctDNA both provide important prognostic information (Balgkouranidou et al. 2016). In assessing the utility of methylated ctDNA for predicting response to chemotherapy, Salazar et al. showed that methylation of the *CHFR* gene (checkpoint with forkhead and ring finger domains) had a negative impact on survival for patients treated with second-line EGFR tyrosine kinase inhibitors (TKIs) compared to chemotherapy in a cohort of 366 patients (Salazar et al. 2011); Wang et al. showed that methylation of the *APC/RASSF1A* genes was predictive of improved tumour response to chemotherapy in a study of 316 patients with advanced lung cancer (Wang et al. 2015b), while Ramirez et al. demonstrated that *SFN* (14-3-3 Sigma) methylation was prognostic of survival (Ramirez et al. 2005). Interestingly, in a recent study, Guo et al. developed a metric that combines information about the pattern of DNA methylation on single molecules together with average DNA methylation levels across the same region (Guo et al. 2017). Using this metric they were able to diagnose lung cancer in cfDNA with a sensitivity of 93.1% and a specificity of 90.6%, and demonstrated better diagnostic performance than using average DNA methylation levels alone.

10.3.6 Brain Cancer

There are more than 100 different types of brain and central nervous system (CNS) tumours, and these can be broadly categorized as either benign or malignant (Louis et al. 2007). Gliomas, arising from glial cells, represent 70% of the primary malignant CNS tumours (Chen et al. 2016). Patients with aggressive gliomas, such as glioblastoma multiforme (GBM), have very poor outcomes, with a median survival time of just 12 months (Johnson et al. 2012) and account for more than half of all deaths from brain tumours. These poor outcomes are partly due to the fact that brain tumours are difficult to diagnose and assess due to their location. Current diagnostic tools include imaging and tissue biopsies, but imaging techniques have limited success for diagnosis and are particularly poor at prognosis and monitoring of recurrence after treatment (Langen et al. 2017). There is thus a need for early and accurate diagnosis with a non-invasive technique, making detection through cfDNA particularly appealing for this tumour type.

Detection of cfDNA from brain tumours is complicated by the location of the tumour inside the cranial vault (Lavon et al. 2010). This means that cfDNA released into the extracellular space will not immediately enter the bloodstream. Instead the cfDNA enters the cerebral spinal fluid (CSF) and circulates within the ventricular system, before eventually entering the bloodstream through blood vessels over the brain surface. However, because of this journey through the CSF circulatory

pathway, brain tumour ctDNA in the blood is likely to be more dilute than ctDNA released from other systemic tumour types with ready access to the bloodstream, as shown by Bettergowda et al. (2014).

The first studies of brain tumour ctDNA simply sought to quantify and characterize cfDNA levels in the CSF and bloodstream (Shi et al. 2012). In a study of 70 glioma patients and 22 healthy controls, Shi et al. amplified Alu elements to show that cfDNA concentration and integrity were increased in the CSF of glioma patients but found no such change to cfDNA in the bloodstream. In contrast, Lavon et al. profiled serum cfDNA of $n = 70$ brain tumour patients and found concordance with molecular biomarkers in the matched tumour samples, including methylation level of the *MGMT* gene and loss of heterozygosity of chromosome 10q (Lavon et al. 2010). It should be noted that these biomarkers were detected in serum cfDNA in ~50% of cases; but the study provided important proof of principle that brain tumour cfDNA could enter and be detected in the bloodstream even months after brain surgery (so not just as a result of the necessary, temporary insult to the brain during surgery).

The *MGMT* gene has been a focus in GBM as it is frequently hypermethylated in the tumour tissue itself, and its inactivation enhances the action of alkylating drugs such as dacarbazine and temozolomide, which are the first-line treatment for GBM (Barault et al. 2015). Thus, detection of *MGMT* hypermethylation in ctDNA in the bloodstream has immediate utility as a predictive marker of treatment response. Fiano et al. investigated this in 58 patients with glioma, all undergoing surgical resection and temozolomide treatment (Fiano et al. 2014). They reported *MGMT* hypermethylation in 60.4% of tumour tissue and 41.4% in plasma, with concordance between tumour and plasma methylation levels. Patients with hypermethylated *MGMT* in the plasma and tumour tissue had increased survival and decreased mortality compared to those without. However, the authors note that the effect was strongest in tissue and urge caution in using plasma as a proxy for tumour tissue methylation at the present time.

As an alternative to interrogating individual candidate genes, Chen et al. considered the diagnostic and prognostic utility of methylation of Alu elements in cfDNA (Chen et al. 2016). Alu is a highly abundant, short interspersed element (SINE), which comprises more than 10% of the human genome. Its methylation status is routinely used as a proxy for whole-genome global DNA methylation. Hypomethylation of Alu elements in tumours is a known source of genomic instability. Accordingly, Alu is frequently reported as hypomethylated in tumours from many different cancer types. In this study, the researchers used a liquid chip system to isolate and measure Alu methylation in cfDNA from $n = 109$ glioma patients, 56 patients with benign intracranial tumours and 50 healthy individuals, as well as matched tumour tissue from $n = 36/109$ glioma patients. The results showed that glioma patients had lower Alu methylation than controls, providing a diagnostic sensitivity of 84.4% and specificity of 80.2%. Within glioma patients, Alu methylation was lower in high-grade versus low-grade patients and negatively associated with survival. Alu methylation levels were highly concordant between tumour and serum, indicating that cfDNA may be a useful substitute for tissue biopsies.

However, other tumour types may also release hypomethylated Alu DNA molecules, and so disease specificity would need to be improved by incorporating other glioma-specific mutations or methylation changes into the cfDNA assay.

Indeed this was the focus of a follow-up paper from the same institute (Gong et al. 2017). In this study, Gong et al. used bisulphite sequencing to profile DNA methylation at Alu, along with *MGMT*, *CDKN2A* (p16INK4a), and *RASSF1A* in 124 glioma patients and 58 healthy controls. There was a significant difference in DNA methylation levels between glioma patients and healthy controls at all markers, except for *RASSF1A*. They reported a diagnostic accuracy of 0.904 for Alu and 0.962 for *MGMT* methylation, and crucially, each marker was able to identify additional patients missed by the other. They also observed a difference in DNA methylation levels between patients with low- and high-grade glioma: Alu methylation was higher in low-grade patients versus high-grade patients; and *MGMT* was lower in low-grade patients versus higher-grade patients. For both markers, high levels of methylation were associated with increased survival. This is thought to be because *MGMT* hypermethylation and gene inactivation improve response (and therefore survival) to radiotherapy and chemotherapy with alkylating agents. Together the results of these brain tumour studies show that DNA methylation profiling in cfDNA holds real potential for diagnostic and prognostic value and that an immediate focus should be to fully evaluate the utility of combining information about Alu and *MGMT* methylation.

10.3.7 Pancreatic Cancer

Pancreatic cancer is a highly aggressive disease with a 5-year survival rate of just 1–14%, depending on disease stage (American Cancer Society 2018). The malignancy of the disease is exacerbated by the difficulties of diagnosis and treatment, as there are few symptoms at early stages of the disease, and, once diagnosed, patients show a poor response to radio- or chemotherapy (Henriksen et al. 2015). Currently, the best chance of survival is early detection, followed by complete resection of the tumour, which can increase the 5-year survival rate to 54.5% (Hartwig et al. 2011). However, there are no effective diagnostic markers for early-stage pancreatic cancer (Kim and Ahuja 2015). Once symptoms do arise, at later stages of the disease, patients are diagnosed through a combination of techniques such as ultrasounds and endoscopies, some of which are invasive and carry a risk of complications. A serum protein, CA19-9, is elevated in patients with advanced disease and used for treatment monitoring; however it is not an effective tool for diagnostic screening as it is found in other cancer types, and a subset of the population cannot express it (Ballehaninna and Chamberlain 2012). Tests based on cfDNA would provide a much-needed non-invasive diagnostic method for early stages of the disease. One of the few known risk factors for pancreatic cancer is chronic inflammation of the pancreas (pancreatitis), so cfDNA tests that can identify pancreatitis, or distinguish it from pancreatic cancer, would also have clinical utility.

One of the first epigenetic studies of cfDNA methylation in pancreatic cancer patients was performed by Melnikov et al. (Melnikov et al. 2009b). The authors used a 56-gene panel, originally derived for a breast cancer study, to test for methylation differences in plasma from $n = 30$ pancreatic cancer patients and $n = 30$ healthy controls. They identified five-gene promoters associated with pancreatic cancer *(CCND2*, *SOCS1*, *THBS1*, *PLAU* and *VHL)*. A subsequent study from the same laboratory applied the same 56-gene panel to subjects with pancreatic cancer, chronic pancreatitis and healthy controls ($n = 30$ of each) (Liggett et al. 2010). Their analysis identified methylation in 14 genes that distinguished pancreatic cancer patients from those with chronic pancreatitis, and 8 genes that distinguished patients with chronic pancreatitis from healthy controls. Of these genes, a subset was unique to each comparison, suggesting that the methylation profiles of pancreatitis and pancreatic cancer can also be distinguished from one another through methylation profiling. Conversely the genes that were common to each comparison, such as *CCND2*, may reflect the inflammation component in pancreatic cancer. The authors validated their findings using a statistical technique of cross-validation, giving high sensitivity and specificity. However, testing in an independent cohort would be required to fully assess the validity of these markers. These studies by Melnikov and Liggett (Melnikov et al. 2009b; Liggett et al. 2010) do not provide the necessary details to determine whether there was overlap in the patients across the two studies, making validation even more essential.

Yi et al. conducted a comprehensive study to identify a novel methylation biomarker panel in pancreatic cancer (Yi et al. 2013). They began by using a pharmacological approach to identify genes that were activated in four pancreatic cancer cell lines after treatment with a demethylating agent. Comparison with other expression datasets led to eight genes that were considered uniquely hypermethylated in pancreatic cancer. These were then validated in pancreatic tissue from $n = 123$ pancreatic cancer patients. The genes *BNC1* and *ADAMST1* were found to be the most frequently hypermethylated. As a next step, they profiled the methylation of these two genes in serum cfDNA from $n = 42$ pancreatic cancer patients and $n = 26$ healthy controls. Both genes were able to distinguish between the two groups, with 79 and 48% sensitivity and 89% and 92% specificity, for *BNC1* and *ADAMTS1*, respectively. A sub-analysis on the ten patients with Stage I disease showed that pancreatic cancer could be detected with 90% sensitivity, which if validated would have huge clinical utility for early detection. However, sample numbers were low; therefore, validation in additional cohorts is required.

Other studies began with a more focused candidate gene approach to assess methylation. Park et al. performed methylation-specific PCR of six candidate genes, which were selected due to reported hypermethylation in primary pancreatic cancer tissues (*UCHL1*, *NPTX2*, *SARP2*, *PENK*, *CDKN2A* (p16INK4a) and *RASSF1A*) (Park et al. 2012a). In a preliminary study, they profiled plasma cfDNA in subjects with pancreatic cancer ($n = 16$), chronic pancreatitis ($n = 13$) and healthy controls ($n = 29$). Promoter hypermethylation was evident in the pancreatic and chronic pancreatitis group, with an average of 1.6 and 1.2 hypermethylated genes per subject, respectively. The difference between the two groups was not statistically

significant. In comparison, just 1/29 of the controls exhibited hypermethylation, in *CDKN2A* (p16INK4a) only. *NPTX2* was the most frequently methylated gene in pancreatic cancer (6/16 patients), so a follow-up study from the same lab examined this gene in a larger cohort with pancreatic cancer ($n = 104$), chronic pancreatitis ($n = 60$) and benign gallstone disease ($n = 5$) (Park et al. 2012b). Plasma cfDNA was extracted from blood collected prior to treatment, ensuring that any methylation events were not a consequence of treatment. Here they found that *NPTX2* hypermethylation could distinguish pancreatic cancer patients from chronic pancreatitis with 80% sensitivity and 76% specificity. Furthermore, the levels of DNA methylation corresponded to cancer stage and so may have prognostic potential.

More recently, Henriksen et al. published a suite of studies assessing plasma cfDNA promoter hypermethylation in a cohort of 95 pancreatic cancer patients (Henriksen et al. 2016, 2017a, b). In all of these studies, they used methylation-specific PCR with a panel of 28 genes selected from the literature as being hypermethylated in plasma or serum cfDNA, pancreatic juice or tumour tissue (including the studies detailed above). In the first study, they assessed the diagnostic utility of these genes through comparison with three control groups: (1) subjects with chronic pancreatitis ($n = 97$), (2) subjects with acute pancreatitis ($n = 59$) and (3) healthy controls ($n = 27$). They observed that the number of hypermethylated genes was much higher in the patients with pancreatic cancer, compared to the other three control groups, and was correlated with the level of cell-free DNA in the plasma. They used this data to build a diagnostic prediction model that distinguished between pancreatic cancer and those with chronic pancreatitis and healthy controls with AUC of 86% (sensitivity = 76%, specificity = 83%). The model included age and methylation status of *APC*, *BMP3*, *BNC1*, *MEST*, *RASSF1A*, *SFRP1*, *SFRP2* and *TFPI2*. The model showed the same accuracy whether applied to all patients (Stages I–IV) or just patients with Stage I and II disease, indicating that the methylation of these genes may have utility as a cfDNA biomarker of early-stage disease.

In a follow-up study, the authors considered the difference between disease stages (Henriksen et al. 2017b) and found that Stage I, II or III patients had an equivalent mean number of hypermethylated genes (~7/28), whereas those with Stage IV disease had significantly more (~10/28). A prediction model based on this data distinguished Stage I–III from Stage IV disease with AUC of 0.87 (sensitivity = 74% and specificity = 87%) and included a distinct set of genes from their previous diagnostic model: *SEPT9*, *SST*, *ALX4*, *CDKN2B*, *HIC1*, *MLH1*, *NEUROG1* and *BNC1*. In a third study, the authors built models to predict survival of the $n = 95$ pancreatic cancer patients. Survival time was lowest in patients with >10 hypermethylated genes. Stage-specific models were able to stratify patients into risk groups according to survival and showed that the methylated genes had the potential to be used as prognostic biomarkers. The findings from these three related papers hold promise for pancreatic cancer diagnosis and prognosis; however it should be noted that all three papers used data from the same, single cohort, so again, validation with larger cohorts is required. Indeed, a limitation of these types of studies is that when findings are applied to individual cases, the specificity is too low

to make clinically meaningful decisions, highlighting the need for extensive validation.

10.4 Cancer Tissue-of-Origin Detection Using cfDNA Methylation Biomarkers

It is well known that each cell type in the body carries unique tissue-specific gene expression and DNA methylation profile. Researchers have harnessed these epigenetic signatures to determine the tissue origins of cfDNA (Snyder et al. 2016; Guo et al. 2017; Lehmann-Werman et al. 2016; Sun et al. 2015). As cfDNA is released from dying cells, tissues with higher than normal cell death such as organs with tumours can be identified through their defining epigenetic patterns.

Sun et al. applied this theory to show that patients with hepatocellular carcinoma (HCC) had higher levels of cfDNA from the liver than healthy individuals (Sun et al. 2015). They first identified single CpG loci that showed tissue-specific methylation in a range of normal human tissues (including the liver) and used this to develop a deconvolution algorithm for distinguishing heterogeneous DNA from different mixtures of tissues. They then applied this algorithm to whole-genome bisulphite sequencing data from the plasma DNA of 29 HCC patients and 32 controls and found that the median percentage of liver cfDNA was 24% for HCC patients versus 10.7% for controls ($P < 0.001$).

Lehmann-Werman et al. used a similar approach to identify patients with pancreatic cancer (Lehmann-Werman et al. 2016). They performed DNA methylation microarray analysis on pancreatic duct and acinar cells (thought to be the cell of origin of pancreatic ductal adenocarcinoma) and identified two CpG sites, near the *REG1A* and *CUX2* genes, that showed distinct hypomethylation compared to other tissues. They measured methylation levels at these two CpG sites (and immediately adjacent CpGs) in cfDNA from 42 subjects with pancreatic cancer and 47 healthy subjects. Their results showed that healthy subjects had very low amounts of unmethylated DNA at these marker regions, whereas 20/42 of the pancreatic patients had amounts above background levels, particularly those patients with advanced disease. Interestingly, unmethylated DNA molecules at these regions were also detectable in patients with pancreatitis, showing their wider utility for detecting damage to the pancreas in a range of conditions. The unmethylated signal from *REG1A* was stronger in patients with pancreatitis, whereas at *CUX2*, the signal was stronger in patients with pancreatic cancer. This likely reflects the different identities of cells dying in each condition as the authors note that *REG1A* was identified from acinar and ductal cells, whereas *CUX2* was identified primarily from ductal cells.

A more recent study by Guo et al. identified genomic regions or 'blocks' showing highly coordinated methylation that could distinguish different tissue types (Guo et al. 2017). They were able to deconvolute the DNA methylation profile of cfDNA using a metric that, for each block, *combined* average DNA methylation levels with the pattern of methylation on single molecules. This metric showed better performance than using average methylation or methylation at individual CpG site. This

revealed that in healthy samples, as expected, white blood cell DNA made up the majority of cfDNA and that DNA from a range of other normal tissues was represented. However, in cancer patients, the tissue of origin of the tumour cells was the most abundant fraction of the (non-white blood cell) cfDNA. They were thus able to use tissue-of-origin signatures to predict colorectal cancer and lung cancer from with an average prediction accuracy of 82.8 and 88.5%, respectively. Additional results from their study indicate that accurate diagnosis from cfDNA may be best achieved through a method that considers both cancer *and* tissue-of-origin signatures. Thus, methylation haplotyping in plasma has potential for early detection of a tumour and identifying the primary tissue of origin, as well as monitoring tumour progression and metastasis to other organs.

10.5 Analytical Methods for Detecting Methylated ctDNA

The focus on DNA methylation as a biomarker in cancer has led to the development of many techniques that can exquisitely and sensitively detect DNA methylation in clinical samples. Key for any approach is the ability to distinguish small amounts of methylated DNA in one or more genomic regions within an excess of cfDNA originating from lysis of leukocytes or other normal tissues. Sensitive detection of DNA methylation at very low abundance requires a high signal-to-noise ratio. Current methodological approaches for biomarker discovery are separated into two major categories: (1) targeted methods to assay specific methylated CpG sites or genomic regions and (2) untargeted whole-genome sequencing of plasma DNA which requires the quantity of ctDNA to be around 5–10% of the total DNA fraction (Chang et al. 2017). In a clinical setting, targeted approaches interrogating clinically relevant methylated genes are preferred.

Numerous experimental methods have been used to assess and validate candidate DNA methylation biomarkers. These include bisulphite pyrosequencing (Tost and Gut 2007a, b), combined bisulphite restriction analysis (COBRA) (Brena et al. 2006), EpiTYPER (Ehrich et al. 2005), MethyLight (Eads et al. 2000), MSP (Herman et al. 1996), Headloop-MSP (Montavon et al. 2012; Devaney et al. 2011; Rand et al. 2005), digital bisulphite genomic sequencing (Weisenberger et al. 2008), digital MethyLight (Weisenberger et al. 2008), methylation-sensitive high-resolution melting (Wong and Dobrovic 2011) and targeted multiplex bisulphite amplicon sequencing (Korbie et al. 2015).

While MSP can identify one methylated allele in 1000 unmethylated alleles (Herman et al. 1996), MethyLight was shown to be at least tenfold more sensitive, detecting one methylated allele in the presence of 10,000-fold excess of unmethylated alleles (Eads et al. 2000). Digital methylation-sensitive high-resolution melting (HS-HRM) has been used to detect *BRAC*1 methylation in peripheral blood, sensitive to a level of 0.1% methylated molecules (Snell et al. 2008). Using digital MethyLight, Weisenberger et al. could identify single-molecule, cancer-specific DNA methylation in the CpG islands of *RUNX3*, *CLDN5* and *FOXE1* on bisulphite-converted DNA isolated from 100 μL of plasma from breast cancer

patients, identifying methylation against a large background of unmethylated molecules and providing a quantitative PCR result (Weisenberger et al. 2008). Similarly, Lange et al. used multiplex digital MethylLight PCR assay to detect methylation of *THBD* and *C9orf50* genes in plasma and serum samples (Lange et al. 2012). Recently, Liu et al. report the inclusion of a methylation-sensitive enrichment step, using a double-stranded DNA-specific nuclease, to enrich methylated alleles (Methylation Specific Nuclease-assisted Minor-allele Enrichment (MS-NaME)), combined with digital MethyLight, to detect methylated sequences down to 0.1% in plasma DNA (Liu et al. 2017).

Next-generation sequencing technologies are revolutionizing the ability to detect methylated tumour DNA in cfDNA, where ctDNA may be $<1\%$ of the total cfDNA, allowing detection of single methylated DNA molecules against a large background of unmethylated molecules and providing a quantitative PCR result. For example, Lehmann-Werman et al. showed that bisulphite treatment of cfDNA, PCR amplification and MiSeq amplicon sequencing enabled detection of unique tissue-specific epigenetic signatures from cfDNA with as little as 0.1–0.5% methylated DNA (Lehmann-Werman et al. 2016). A recent comparison of DNA methylation assays for biomarker development and clinical applications showed that targeted amplicon bisulphite sequencing and pyrosequencing showed the best all-round performance and could be considered technologies ready for widespread use in biomarker development and clinical applications (BLUEPRINT consortium 2016). Key features for any method for the development of methylation biomarkers for ctDNA include accuracy, sensitivity, robustness of the assay, high reproducibility, low sample input and practical feasibility. As with any methylation-based assay, there is also the need to define threshold levels of methylation and consider whether this is defined by one methylated CpG site per molecule or a number of CpG sites across the region.

10.6 Clinical Utility of Methylated ctDNA

There is a widespread interest in the clinical utility of circulating-free tumour (ct) DNA methylation as a 'liquid biopsy'. In this section, we discuss the potential stages at which it could be applied to improve patient management (Murtaza et al. 2015), illustrated with key examples from the preceding cancer-specific sections.

ctDNA Methylation as a Diagnostic Marker Early screening programs and early diagnosis have an important impact in improving disease-free survival and reducing mortality in cancer patients. The use of methylated ctDNA biomarkers for early diagnosis is being intensely investigated as it provides a non-invasive approach for early diagnosis, which fosters patient compliance and can therefore be included in future screening programs. As discussed above, a number of potential methylated ctDNA biomarkers for diagnosis are being proposed, such as *GSTP1* methylation in prostate cancer (Wu et al. 2011; Devaney et al. 2011), *SEPT9* in colorectal cancer, *RASSF1A* in breast cancer (Fackler et al. 2014), *RASSF1A* in ovarian cancer patients

(Ibanez de Caceres et al. 2004), *MGMT* in brain cancer (Lavon et al. 2010) and numerous gene panels for lung-specific diagnosis (Weiss et al. 2017; Wielscher et al. 2015). Methylated ctDNA biomarkers also offer the prospect of classifying cancers into biologically and clinically distinct disease subtypes at diagnosis for targeted treatments.

ctDNA as a Prognostic Biomarker Methylated ctDNA biomarkers are being assessed as biomarkers for risk stratification and disease recurrence. For example, in ovarian cancer, *BRCA1* promoter methylation was shown to be associated with poor patient outcome, significantly shorter progression-free and overall survival, compared with patients with mutated or wild-type *BRCA1* (Chiang et al. 2006). In breast cancer, hypermethylation of *RASSF1* and *APC* in the sera of breast cancer patients was associated with a worse outcome and performed better than current standard prognostic measures (Muller et al. 2003), while *GSTP1* methylation in prostate cancer cfDNA was associated with overall survival (Mahon et al. 2014). In pancreatic cancer, the levels of DNA methylation of *NPTX2* corresponded to cancer stage and are therefore of prognostic potential.

ctDNA as a Predictive Biomarker Importantly, methylated DNA has also been shown to be useful in the identification of determinants for targeted therapy. *BRCA1* methylation is associated with sensitivity to platinum treatment in a pilot study of 35 epithelial ovarian cancers (Chaudhry et al. 2009), suggesting it may serve as a biomarker for aggressive ovarian cancer and predict response to therapy. In glioblastoma multiforme (GBM), methylation of *MGMT* gene enhances the action of alkylating drugs, such as dacarbazine and temozolomide, which are the first-line treatment, informing the predictive utility of *MGMT* hypermethylation (Barault et al. 2015).

Monitoring Treatment Efficiency, Minimal Residual Disease and Early Disease Relapse Finally, monitoring treatment response is critical in all phases of cancer management to avoid continuing ineffective therapies, to prevent unnecessary side effects and to inform the necessity of different therapeutic regimes as early as possible. A number of genes have been shown to be useful for monitoring response to therapy; for example, in breast cancer, *BRCA1* (Sharma et al. 2012), *SFN* (Zurita et al. 2010), *RASSF1A* and *NEUROD1* (Fiegl et al. 2008; Fiegl et al. 2005) have been used to monitor treatment efficiency. In prostate cancer, *GSTP1* methylation in cfDNA was associated with response to chemotherapy (Mahon et al. 2014).

The development of resistance to chemotherapeutic and targeted agents is a major problem faced in the care of cancer patients. Thus, monitoring of minimal residual disease and detecting the evolution of resistance in real time in a non-invasive way represent an area where methylated ctDNA biomarkers offer enormous clinical potential.

10.7 Methylated cfDNA Biomarkers Currently in the Clinic

Even though tremendous efforts have been made over the past few decades to discover cancer biomarkers, there remains a striking discrepancy between the effort directed toward biomarker discovery and the number of markers in clinical practice. One of the confounding issues in translating a novel biomarker discovery into clinical practice is the analytical, diagnostic and regulatory requirements for a clinical assay (Fuzery et al. 2013). Nevertheless, a number of DNA methylated biomarkers are now entering clinical practice.

Colvera™ is a clinically validated laboratory-developed blood test designed to detect CRC from cfDNA. It is a PCR-based blood test for detecting hypermethylation of *BCAT1* and *IKZF1* genes. The product has taken over 10 years of research, development and clinical testing in over 3800 volunteers. Recent data shows that Colvera™ is sensitive in detecting colorectal cancer recurrence; it is being tested in conjunction with the carcinoembryonic antigen (CEA) test to identify recurrence before symptoms appear in the hope that earlier detection may lead to improved patient survival.

Epi proColon® (Epigenomics AG, Berlin, Germany) is a commercial FDA-approved blood test that detects methylated *SEPT9* DNA for colorectal cancer screening. This is the first blood test approved for cancer screening and represents an important milestone. Epigenomics® (Berlin, Germany) has also announced CE-IVD marking (EU approval for selling an in vitro diagnostic device) for its blood-based lung cancer test, Epi proLung®, which is based on the DNA methylation biomarkers *SHOX2* and *PTGER4* and provides additional testing options to detect lung cancer.

10.8 Challenges and Limitations

Despite the extensive clinical utility of cfDNA, there are still challenges in its study in the laboratory and in its widespread clinical implementation (Zarzour et al. 2013). Notably, cfDNA is frequently found in plasma at very low concentrations, and therefore, its isolation and quantitation are not straightforward and simple tasks. Furthermore, ctDNA with genetic and epigenetic alterations often comprises only a small portion of the total cfDNA in circulating plasma and can be confounded by cfDNA contributed by normal cells. In fact, ctDNA can range from 0.01% to more than 90% of the cfDNA (Bettegowda et al. 2014; Thierry et al. 2014; Heitzer et al. 2013; Leary et al. 2012), posing additional limitations in interpreting results.

While improvements in technology now allow the analysis of extremely rare alleles, there is a lack of consensus on how to best use the available methods; it is critical to establish benchmarks for the analysis of ctDNA, for implementation in a clinical setting. Technical issues, such as the method of cfDNA extraction, DNA methylation profiling and methylation thresholds, need to be standardized, and there needs to be consensus on reporting results. More studies are needed to provide clinical standards for the use of liquid biopsies as a clinical specimen. This will require well-designed and sufficiently powered multi-centre studies with large cohorts of cancer patients and with well-defined clinical staging and outcomes.

There is a need to use suitable control groups, e.g. patients with other cancer types to show specificity of the methylation change to the particular cancer of interest. Appropriate healthy controls should also be considered; e.g. for lung cancer, smokers without lung cancer need to be assessed as a control group to distinguish between the effect of smoking on methylation and the association between methylation and cancer. Thus, for the methylated biomarkers to be included in routine clinical care, the biomarker needs to pass the highest standards of analytical validity, clinical validity and clinical utility (Febbo et al. 2011).

Studies have shown that the clinical utility of methylated biomarkers may best lie in their use in combinatorial approach with standard immunohistochemical and/or tumour pathological features. For example, combining oestrogen receptor status, and lymph node status, in a multivariate model, with a panel of four methylated marker genes, *PITX2*, *BMP4*, *FGF4* and *FAM110A*, added significant information to increase the outcome prediction performance to anthracycline-based chemotherapy treatment in breast cancer (Hartmann et al. 2009). The quantification of cfDNA concentrations *alone* does not seem to be useful in the diagnostic setting, owing to the overlapping DNA concentrations that are found in healthy individuals with those in patients with benign and malignant disease, and elevated cfDNA concentration may be caused by non-cancerous events such as inflammation or a benign cell-proliferative lesion. However, the assessment of cfDNA concentration in *combination* with other blood tumour biomarkers has been shown to be useful in increasing the clinical predictive importance of these tests (Holm et al. 2010).

10.9 Future Perspectives

The clinical utility of cfDNA presents a promising non-invasive strategy for the early detection and monitoring of cancer with the capability of detection before the onset of clinical symptoms or before progression to more advanced stages, where the disease burden is high and treatments options are limited. Currently, efficient management of cancer patients relies on early diagnosis, precise tumour staging and monitoring of treatment. Histological evaluation of tumour tissues obtained from biopsies, as well as blood samples, is the 'gold standard' of diagnosis, and most patients are evaluated only once. However, we know that during a patient's cancer journey, the genomic, epigenomic and transcriptomic characteristics of their tumour(s) will vary with natural evolution and in response to treatment. ctDNA assays allow repeated monitoring and represent an efficient and more accurate means of assessing these dynamic changes. This approach also circumvents the limitation of using the original biopsies of the primary tumour (if available), which are often years old and can be quite different to a patient's recurrent or metastasized tumours. As cancer patient's genomic and epigenomic tumour profiles are determined, the use of cfDNA tests will be exploited for personalized therapy and monitoring response to treatment. Minimally invasive blood analyses of cfDNA have the potential to complement and enhance the existing cancer tissue and blood biomarkers in the future.

Acknowledgements We thank Madhavi Maddugoda for the preparation of figures and reviewing the manuscript. This work is supported by National Health and Medical Research Council (NHMRC) project grants (CS:APP1128912 and RP:APP1106870), Cancer Council New South Wales (CCNSW) project grant (RP: RG 18-09), National Breast Cancer Foundation IIRS Grant (CS) and Cancer Institute of New South Wales Fellowship (CINSW) to RP (14/ECF/1-23). The contents of the published material are solely the responsibility of the administering institution and individual authors and do not reflect the views of the NHMRC.

References

American Cancer Society. Pancreatic cancer survival rates, by stage. https://www.cancer.org/cancer/pancreatic-cancer/detection-diagnosis-staging/survival-rates.html Accessed Jan 2018

An Q, Liu Y, Gao Y, Huang J, Fong X, Li L et al (2002) Detection of p16 hypermethylation in circulating plasma DNA of non-small cell lung cancer patients. Cancer Lett 188(1-2):109–114

Balgkouranidou I, Chimonidou M, Milaki G, Tsarouxa EG, Kakolyris S, Welch DR et al (2014) Breast cancer metastasis suppressor-1 promoter methylation in cell-free DNA provides prognostic information in non-small cell lung cancer. Br J Cancer 110(8):2054–2062

Balgkouranidou I, Chimonidou M, Milaki G, Tsaroucha E, Kakolyris S, Georgoulias V et al (2016) SOX17 promoter methylation in plasma circulating tumor DNA of patients with non-small cell lung cancer. Clin Chem Lab Med 54(8):1385–1393

Ballehaninna UK, Chamberlain RS (2012) The clinical utility of serum CA 19-9 in the diagnosis, prognosis and management of pancreatic adenocarcinoma: an evidence based appraisal. J Gastrointest Oncol 3(2):105–119

Barault L, Amatu A, Bleeker FE, Moutinho C, Falcomata C, Fiano V et al (2015) Digital PCR quantification of MGMT methylation refines prediction of clinical benefit from alkylating agents in glioblastoma and metastatic colorectal cancer. Ann Oncol 26(9):1994–1999

Barton CA, Hacker NF, Clark SJ, O'Brien PM (2008) DNA methylation changes in ovarian cancer: implications for early diagnosis, prognosis and treatment. Gynecol Oncol 109(1):129–139

Bast RC Jr, Feeney M, Lazarus H, Nadler LM, Colvin RB, Knapp RC (1981) Reactivity of a monoclonal antibody with human ovarian carcinoma. J Clin Invest 68(5):1331–1337

Bastian PJ, Palapattu GS, Lin X, Yegnasubramanian S, Mangold LA, Trock B et al (2005) Preoperative serum DNA GSTP1 CpG island hypermethylation and the risk of early prostate-specific antigen recurrence following radical prostatectomy. Clin Cancer Res 11(11):4037–4043

Bearzatto A, Conte D, Frattini M, Zaffaroni N, Andriani F, Balestra D et al (2002) p16(INK4A) hypermethylation detected by fluorescent methylation-specific PCR in plasmas from non-small cell lung cancer. Clin Cancer Res 8(12):3782–3787

Bettegowda C, Sausen M, Leary RJ, Kinde I, Wang Y, Agrawal N et al (2014) Detection of circulating tumor DNA in early- and late-stage human malignancies. Sci Transl Med 6 (224):224ra24

BLUEPRINT consortium (2016) Quantitative comparison of DNA methylation assays for biomarker development and clinical applications. Nat Biotechnol 34(7):726–737

Brait M, Banerjee M, Maldonado L, Ooki A, Loyo M, Guida E et al (2017) Promoter methylation of MCAM, ERalpha and ERbeta in serum of early stage prostate cancer patients. Oncotarget 8 (9):15431–15440

Brena RM, Auer H, Kornacker K, Plass C (2006) Quantification of DNA methylation in electrofluidics chips (Bio-COBRA). Nat Protoc 1(1):52–58

Cassinotti E, Melson J, Liggett T, Melnikov A, Yi Q, Replogle C et al (2012) DNA methylation patterns in blood of patients with colorectal cancer and adenomatous colorectal polyps. Int J Cancer 131(5):1153–1157

Chang Y, Tolani B, Nie X, Zhi X, Hu M, He B (2017) Review of the clinical applications and technological advances of circulating tumor DNA in cancer monitoring. Ther Clin Risk Manag 13:1363–1374

Chaudhry P, Srinivasan R, Patel FD (2009) Utility of gene promoter methylation in prediction of response to platinum-based chemotherapy in epithelial ovarian cancer (EOC). Cancer Invest 27 (8):877–884

Chen J, Huan W, Zuo H, Zhao L, Huang C, Liu X et al (2016) Alu methylation serves as a biomarker for non-invasive diagnosis of glioma. Oncotarget 7(18):26099–26106

Chiang JW, Karlan BY, Cass L, Baldwin RL (2006) BRCA1 promoter methylation predicts adverse ovarian cancer prognosis. Gynecol Oncol 101(3):403–410

Church TR, Wandell M, Lofton-Day C, Mongin SJ, Burger M, Payne SR et al (2014) Prospective evaluation of methylated SEPT9 in plasma for detection of asymptomatic colorectal cancer. Gut 63(2):317–325

Collins Y, Diciocco R, Keitz B, Lele S, Odunsi K (2006) Methylation of death-associated protein kinase in ovarian carcinomas. Int J Gynecol Cancer 16(Suppl 1):195–199

Dawson SJ, Rosenfeld N, Caldas C (2013) Circulating tumor DNA to monitor metastatic breast cancer. N Engl J Med 369(1):93–94

Deng QK, Lei YG, Lin YL, Ma JG, Li WP (2016) Prognostic value of protocadherin10 (PCDH10) methylation in serum of prostate cancer patients. Med Sci Monit 22:516–521

Devaney J, Stirzaker C, Qu W, Song JZ, Statham AL, Patterson KI et al (2011) Epigenetic deregulation across chromosome 2q14.2 differentiates normal from prostate cancer and provides a regional panel of novel DNA methylation cancer biomarkers. Cancer Epidemiol Biomarkers Prev 20(1):148–159

deVos T, Tetzner R, Model F, Weiss G, Schuster M, Distler J et al (2009) Circulating methylated SEPT9 DNA in plasma is a biomarker for colorectal cancer. Clin Chem 55(7):1337–1346

Dong R, Yu J, Pu H, Zhang Z, Xu X (2012) Frequent SLIT2 promoter methylation in the serum of patients with ovarian cancer. J Int Med Res 40(2):681–686

Duffy MJ, Bonfrer JM, Kulpa J, Rustin GJ, Soletormos G, Torre GC et al (2005) CA125 in ovarian cancer: European Group on Tumor Markers guidelines for clinical use. Int J Gynecol Cancer 15 (5):679–691

Eads CA, Danenberg KD, Kawakami K, Saltz LB, Blake C, Shibata D et al (2000) MethyLight: a high-throughput assay to measure DNA methylation. Nucleic Acids Res 28(8):E32

Earp MA, Cunningham JM (2015) DNA methylation changes in epithelial ovarian cancer histotypes. Genomics 106(6):311–321

Eggert JA, Palavanzadeh M, Blanton A (2017) Screening and early detection of lung cancer. Semin Oncol Nurs 33(2):129–140

Ehrich M, Nelson MR, Stanssens P, Zabeau M, Liloglou T, Xinarianos G et al (2005) Quantitative high-throughput analysis of DNA methylation patterns by base-specific cleavage and mass spectrometry. Proc Natl Acad Sci U S A 102(44):15785–15790

Fackler MJ, Lopez Bujanda Z, Umbricht C, Teo WW, Cho S, Zhang Z et al (2014) Novel methylated biomarkers and a robust assay to detect circulating tumor DNA in metastatic breast cancer. Cancer Res 74(8):2160–2170

Febbo PG, Ladanyi M, Aldape KD, De Marzo AM, Hammond ME, Hayes DF et al (2011) NCCN Task Force report: evaluating the clinical utility of tumor markers in oncology. J Natl Compr Canc Netw 9(Suppl 5):S1–S32. quiz S3

Ferlay J, Soerjomataram I, Dikshit R, Eser S, Mathers C, Rebelo M et al (2015) Cancer incidence and mortality worldwide: sources, methods and major patterns in GLOBOCAN 2012. Int J Cancer 136(5):E359–E386

Fiano V, Trevisan M, Trevisan E, Senetta R, Castiglione A, Sacerdote C et al (2014) MGMT promoter methylation in plasma of glioma patients receiving temozolomide. J Neurooncol 117 (2):347–357

Fiegl H, Millinger S, Mueller-Holzner E, Marth C, Ensinger C, Berger A et al (2005) Circulating tumor-specific DNA: a marker for monitoring efficacy of adjuvant therapy in cancer patients. Cancer Res 65(4):1141–1145

Fiegl H, Jones A, Hauser-Kronberger C, Hutarew G, Reitsamer R, Jones RL et al (2008) Methylated NEUROD1 promoter is a marker for chemosensitivity in breast cancer. Clin Cancer Res 14 (11):3494–3502

Fuzery AK, Levin J, Chan MM, Chan DW (2013) Translation of proteomic biomarkers into FDA approved cancer diagnostics: issues and challenges. Clin Proteomics 10(1):13

Giannopoulou L, Kasimir-Bauer S, Lianidou ES (2017) Liquid biopsy in ovarian cancer: recent advances on circulating tumor cells and circulating tumor DNA. Clin Chem Lab Med 56:186–197

Gifford G, Paul J, Vasey PA, Kaye SB, Brown R (2004) The acquisition of hMLH1 methylation in plasma DNA after chemotherapy predicts poor survival for ovarian cancer patients. Clin Cancer Res 10(13):4420–4426

Gong M, Shi W, Qi J, Shao G, Shi Z, Wang J et al (2017) Alu hypomethylation and MGMT hypermethylation in serum as biomarkers of glioma. Oncotarget 8(44):76797–76806

Grote HJ, Schmiemann V, Geddert H, Rohr UP, Kappes R, Gabbert HE et al (2005) Aberrant promoter methylation of p16(INK4a), RARB2 and SEMA3B in bronchial aspirates from patients with suspected lung cancer. Int J Cancer 116(5):720–725

Grutzmann R, Molnar B, Pilarsky C, Habermann JK, Schlag PM, Saeger HD et al (2008) Sensitive detection of colorectal cancer in peripheral blood by septin 9 DNA methylation assay. PLoS One 3(11):e3759

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(4):635–642

Hartmann O, Spyratos F, Harbeck N, Dietrich D, Fassbender A, Schmitt M et al (2009) DNA methylation markers predict outcome in node-positive, estrogen receptor-positive breast cancer with adjuvant anthracycline-based chemotherapy. Clin Cancer Res 15(1):315–323

Hartwig W, Hackert T, Hinz U, Gluth A, Bergmann F, Strobel O et al (2011) Pancreatic cancer surgery in the new millennium: better prediction of outcome. Ann Surg 254(2):311–319

Heitzer E, Auer M, Hoffmann EM, Pichler M, Gasch C, Ulz P et al (2013) Establishment of tumor-specific copy number alterations from plasma DNA of patients with cancer. Int J Cancer 133 (2):346–356

Henriksen SD, Madsen PH, Krarup H, Thorlacius-Ussing O (2015) DNA hypermethylation as a blood-based marker for pancreatic cancer: a literature review. Pancreas 44(7):1036–1045

Henriksen SD, Madsen PH, Larsen AC, Johansen MB, Drewes AM, Pedersen IS et al (2016) Cell-free DNA promoter hypermethylation in plasma as a diagnostic marker for pancreatic adenocarcinoma. Clin Epigenetics 8:117

Henriksen SD, Madsen PH, Larsen AC, Johansen MB, Pedersen IS, Krarup H et al (2017a) Cell-free DNA promoter hypermethylation in plasma as a predictive marker for survival of patients with pancreatic adenocarcinoma. Oncotarget 8(55):93942–93956

Henriksen SD, Madsen PH, Larsen AC, Johansen MB, Pedersen IS, Krarup H et al (2017b) Promoter hypermethylation in plasma-derived cell-free DNA as a prognostic marker for pancreatic adenocarcinoma staging. Int J Cancer 141(12):2489–2497

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

Hesson LB, Wilson R, Morton D, Adams C, Walker M, Maher ER et al (2005) CpG island promoter hypermethylation of a novel Ras-effector gene RASSF2A is an early event in colon carcinogenesis and correlates inversely with K-ras mutations. Oncogene 24(24):3987–3994

Holm K, Hegardt C, Staaf J, Vallon-Christersson J, Jonsson G, Olsson H et al (2010) Molecular subtypes of breast cancer are associated with characteristic DNA methylation patterns. Breast Cancer Res 12(3):R36

Horning AM, Awe JA, Wang CM, Liu J, Lai Z, Wang VY et al (2015) DNA methylation screening of primary prostate tumors identifies SRD5A2 and CYP11A1 as candidate markers for assessing risk of biochemical recurrence. Prostate 75(15):1790–1801

Hsu HS, Chen TP, Hung CH, Wen CK, Lin RK, Lee HC et al (2007) Characterization of a multiple epigenetic marker panel for lung cancer detection and risk assessment in plasma. Cancer 110 (9):2019–2026

Ibanez de Caceres I, Battagli C, Esteller M, Herman JG, Dulaimi E, Edelson MI et al (2004) Tumor cell-specific BRCA1 and RASSF1A hypermethylation in serum, plasma, and peritoneal fluid from ovarian cancer patients. Cancer Res 64(18):6476–6481

Jeronimo C, Usadel H, Henrique R, Silva C, Oliveira J, Lopes C et al (2002) Quantitative GSTP1 hypermethylation in bodily fluids of patients with prostate cancer. Urology 60(6):1131–1135

Johnson DR, Ma DJ, Buckner JC, Hammack JE (2012) Conditional probability of long-term survival in glioblastoma: a population-based analysis. Cancer 118(22):5608–5613

Jones PA, Baylin SB (2007) The epigenomics of cancer. Cell 128(4):683–692

Kim VM, Ahuja N (2015) Early detection of pancreatic cancer. Chin J Cancer Res 27(4):321–331

Kneip C, Schmidt B, Seegebarth A, Weickmann S, Fleischhacker M, Liebenberg V et al (2011) SHOX2 DNA methylation is a biomarker for the diagnosis of lung cancer in plasma. J Thorac Oncol 6(10):1632–1638

Korbie D, Lin E, Wall D, Nair SS, Stirzaker C, Clark SJ et al (2015) Multiplex bisulfite PCR resequencing of clinical FFPE DNA. Clin Epigenetics 7(1):28

Krishnamurthy N, Spencer E, Torkamani A, Nicholson L (2017) Liquid biopsies for cancer: Coming to a patient near you. J Clin Med 6(1):3

Kurakawa E, Shimamoto T, Utsumi K, Hirano T, Kato H, Ohyashiki K (2001) Hypermethylation of p16(INK4a) and p15(INK4b) genes in non-small cell lung cancer. Int J Oncol 19(2):277–281

Lange CP, Campan M, Hinoue T, Schmitz RF, van der Meulen-de Jong AE, Slingerland H et al (2012) Genome-scale discovery of DNA-methylation biomarkers for blood-based detection of colorectal cancer. PLoS One 7(11):e50266

Langen KJ, Galldiks N, Hattingen E, Shah NJ (2017) Advances in neuro-oncology imaging. Nat Rev Neurol 13(5):279–289

Lavon I, Refael M, Zelikovitch B, Shalom E, Siegal T (2010) Serum DNA can define tumor-specific genetic and epigenetic markers in gliomas of various grades. Neuro Oncol 12(2):173–180

Leary RJ, Sausen M, Kinde I, Papadopoulos N, Carpten JD, Craig D et al (2012) Detection of chromosomal alterations in the circulation of cancer patients with whole-genome sequencing. Sci Transl Med 4(162):162ra54

Lee TH, Montalvo L, Chrebtow V, Busch MP (2001) Quantitation of genomic DNA in plasma and serum samples: higher concentrations of genomic DNA found in serum than in plasma. Transfusion 41(2):276–282

Lee BB, Lee EJ, Jung EH, Chun HK, Chang DK, Song SY et al (2009) Aberrant methylation of APC, MGMT, RASSF2A, and Wif-1 genes in plasma as a biomarker for early detection of colorectal cancer. Clin Cancer Res 15(19):6185–6191

Lehmann BD, Bauer JA, Chen X, Sanders ME, Chakravarthy AB, Shyr Y et al (2011) Identification of human triple-negative breast cancer subtypes and preclinical models for selection of targeted therapies. J Clin Invest 121(7):2750–2767

Lehmann-Werman R, Neiman D, Zemmour H, Moss J, Magenheim J, Vaknin-Dembinsky A et al (2016) Identification of tissue-specific cell death using methylation patterns of circulating DNA. Proc Natl Acad Sci U S A 113(13):E1826–E1834

Liggett T, Melnikov A, Yi QL, Replogle C, Brand R, Kaul K et al (2010) Differential methylation of cell-free circulating DNA among patients with pancreatic cancer versus chronic pancreatitis. Cancer 116(7):1674–1680

Liggett TE, Melnikov A, Yi Q, Replogle C, Hu W, Rotmensch J et al (2011) Distinctive DNA methylation patterns of cell-free plasma DNA in women with malignant ovarian tumors. Gynecol Oncol 120(1):113–120

Liu Y, Song C, Ladas I, Fitarelli-Kiehl M, Makrigiorgos GM (2017) Methylation-sensitive enrichment of minor DNA alleles using a double-strand DNA-specific nuclease. Nucleic Acids Res 45 (6):e39

Lofton-Day C, Model F, Devos T, Tetzner R, Distler J, Schuster M et al (2008) DNA methylation biomarkers for blood-based colorectal cancer screening. Clin Chem 54(2):414–423

Louis DN, Ohgaki H, Wiestler OD, Cavenee WK, Burger PC, Jouvet A et al (2007) The 2007 WHO classification of tumours of the central nervous system. Acta Neuropathol 114(2):97–109

Lu Y, Li S, Zhu S, Gong Y, Shi J, Xu L (2017) Methylated DNA/RNA in body fluids as biomarkers for lung cancer. Biol Proced Online 19:2

Mahon KL, Qu W, Devaney J, Paul C, Castillo L, Wykes RJ et al (2014) Methylated glutathione S-transferase 1 (mGSTP1) is a potential plasma free DNA epigenetic marker of prognosis and response to chemotherapy in castrate-resistant prostate cancer. Br J Cancer 111(9):1802–1809

Melnikov A, Scholtens D, Godwin A, Levenson V (2009a) Differential methylation profile of ovarian cancer in tissues and plasma. J Mol Diagn 11(1):60–65

Melnikov AA, Scholtens D, Talamonti MS, Bentrem DJ, Levenson VV (2009b) Methylation profile of circulating plasma DNA in patients with pancreatic cancer. J Surg Oncol 99 (2):119–122

Montavon C, Gloss BS, Warton K, Barton CA, Statham AL, Scurry JP et al (2012) Prognostic and diagnostic significance of DNA methylation patterns in high grade serous ovarian cancer. Gynecol Oncol 124(3):582–588

Muller HM, Widschwendter A, Fiegl H, Ivarsson L, Goebel G, Perkmann E et al (2003) DNA methylation in serum of breast cancer patients: an independent prognostic marker. Cancer Res 63(22):7641–7645

Murtaza M, Dawson SJ, Pogrebniak K, Rueda OM, Provenzano E, Grant J et al (2015) Multifocal clonal evolution characterized using circulating tumour DNA in a case of metastatic breast cancer. Nat Commun 6:8760

Nam RK, Saskin R, Lee Y, Liu Y, Law C, Klotz LH et al (2010) Increasing hospital admission rates for urological complications after transrectal ultrasound guided prostate biopsy. J Urol 183 (3):963–968

Ng CS, Zhang J, Wan S, Lee TW, Arifi AA, Mok T et al (2002) Tumor p16M is a possible marker of advanced stage in non-small cell lung cancer. J Surg Oncol 79(2):101–106

Nishihara R, Wu K, Lochhead P, Morikawa T, Liao X, Qian ZR et al (2013) Long-term colorectal-cancer incidence and mortality after lower endoscopy. N Engl J Med 369(12):1095–1105

Okegawa T, Nutahara K, Higashihara E (2010) Association of circulating tumor cells with tumor-related methylated DNA in patients with hormone-refractory prostate cancer. Int J Urol 17 (5):466–475

Ostrow KL, Hoque MO, Loyo M, Brait M, Greenberg A, Siegfried JM et al (2010) Molecular analysis of plasma DNA for the early detection of lung cancer by quantitative methylation-specific PCR. Clin Cancer Res 16(13):3463–3472

Ostrow KL, Michailidi C, Guerrero-Preston R, Hoque MO, Greenberg A, Rom W et al (2013) Cigarette smoke induces methylation of the tumor suppressor gene NISCH. Epigenetics 8 (4):383–388

Pack SC, Kim HR, Lim SW, Kim HY, Ko JY, Lee KS et al (2013) Usefulness of plasma epigenetic changes of five major genes involved in the pathogenesis of colorectal cancer. Int J Colorectal Dis 28(1):139–147

Park JW, Baek IH, Kim YT (2012a) Preliminary study analyzing the methylated genes in the plasma of patients with pancreatic cancer. Scand J Surg 101(1):38–44

Park JK, Ryu JK, Yoon WJ, Lee SH, Lee GY, Jeong KS et al (2012b) The role of quantitative NPTX2 hypermethylation as a novel serum diagnostic marker in pancreatic cancer. Pancreas 41 (1):95–101

Pedersen SK, Mitchell SM, Graham LD, McEvoy A, Thomas ML, Baker RT et al (2014) CAHM, a long non-coding RNA gene hypermethylated in colorectal neoplasia. Epigenetics 9 (8):1071–1082

Ponomaryova AA, Rykova EY, Cherdyntseva NV, Skvortsova TE, Dobrodeev AY, Zav'yalov AA et al (2013) Potentialities of aberrantly methylated circulating DNA for diagnostics and post-treatment follow-up of lung cancer patients. Lung Cancer 81(3):397–403

Powrozek T, Krawczyk P, Kucharczyk T, Milanowski J (2014) Septin 9 promoter region methylation in free circulating DNA-potential role in noninvasive diagnosis of lung cancer: preliminary report. Med Oncol 31(4):917

Powrozek T, Krawczyk P, Nicos M, Kuznar-Kaminska B, Batura-Gabryel H, Milanowski J (2016) Methylation of the DCLK1 promoter region in circulating free DNA and its prognostic value in lung cancer patients. Clin Transl Oncol 18(4):398–404

Ramirez JL, Rosell R, Taron M, Sanchez-Ronco M, Alberola V, de Las Penas R et al (2005) 14-3-3sigma methylation in pretreatment serum circulating DNA of cisplatin-plus-gemcitabine-treated advanced non-small-cell lung cancer patients predicts survival: The Spanish Lung Cancer Group. J Clin Oncol Off J Am Soc Clin Oncol 23(36):9105–9112

Rand KN, Ho T, Qu W, Mitchell SM, White R, Clark SJ et al (2005) Headloop suppression PCR and its application to selective amplification of methylated DNA sequences. Nucleic Acids Res 33(14):e127

Reibenwein J, Pils D, Horak P, Tomicek B, Goldner G, Worel N et al (2007) Promoter hypermethylation of GSTP1, AR, and 14-3-3sigma in serum of prostate cancer patients and its clinical relevance. Prostate 67(4):427–432

Roddam AW, Duffy MJ, Hamdy FC, Ward AM, Patnick J, Price CP et al (2005) Use of prostate-specific antigen (PSA) isoforms for the detection of prostate cancer in men with a PSA level of 2–10 ng/ml: systematic review and meta-analysis. Eur Urol 48(3):386–399. discussion 98–9

Rykova EY, Skvortsova TE, Laktionov PP, Tamkovich SN, Bryzgunova OE, Starikov AV et al (2004) Investigation of tumor-derived extracellular DNA in blood of cancer patients by methylation-specific PCR. Nucleosides Nucleotides Nucleic Acids 23(6-7):855–859

Salazar F, Molina MA, Sanchez-Ronco M, Moran T, Ramirez JL, Sanchez JM et al (2011) First-line therapy and methylation status of CHFR in serum influence outcome to chemotherapy versus EGFR tyrosine kinase inhibitors as second-line therapy in stage IV non-small-cell lung cancer patients. Lung Cancer 72(1):84–91

Schmidt B, Beyer J, Dietrich D, Bork I, Liebenberg V, Fleischhacker M (2015) Quantification of cell-free mSHOX2 Plasma DNA for therapy monitoring in advanced stage non-small cell (NSCLC) and small-cell lung cancer (SCLC) patients. PLoS One 10(2):e0118195

Scholefield JH, Moss SM, Mangham CM, Whynes DK, Hardcastle JD (2012) Nottingham trial of faecal occult blood testing for colorectal cancer: a 20-year follow-up. Gut 61(7):1036–1040

Schwarzenbach H, Hoon DS, Pantel K (2011) Cell-free nucleic acids as biomarkers in cancer patients. Nat Rev Cancer 11(6):426–437

Shan M, Yin H, Li J, Li X, Wang D, Su Y et al (2016) Detection of aberrant methylation of a six-gene panel in serum DNA for diagnosis of breast cancer. Oncotarget 7(14):18485–18494

Sharma G, Mirza S, Parshad R, Gupta SD, Ralhan R (2012) DNA methylation of circulating DNA: a marker for monitoring efficacy of neoadjuvant chemotherapy in breast cancer patients. Tumour Biol 33(6):1837–1843

Shaukat A, Mongin SJ, Geisser MS, Lederle FA, Bond JH, Mandel JS et al (2013) Long-term mortality after screening for colorectal cancer. N Engl J Med 369(12):1106–1114

Shi W, Lv C, Qi J, Zhao W, Wu X, Jing R et al (2012) Prognostic value of free DNA quantification in serum and cerebrospinal fluid in glioma patients. J Mol Neurosci 46(3):470–475

Shivapurkar N, Gazdar AF (2010) DNA methylation based biomarkers in non-invasive cancer screening. Curr Mol Med 10(2):123–132

Siegel R, Ma J, Zou Z, Jemal A (2014) Cancer statistics, 2014. CA Cancer J Clin 64(1):9–29

Sloane MA, Wong JW, Perera D, Nunez AC, Pimanda JE, Hawkins NJ et al (2014) Epigenetic inactivation of the candidate tumor suppressor USP44 is a frequent and early event in colorectal neoplasia. Epigenetics 9(8):1092–1100

Snell C, Krypuy M, Wong EM, kConFab i LMB, Dobrovic A (2008) BRCA1 promoter methylation in peripheral blood DNA of mutation negative familial breast cancer patients with a BRCA1 tumour phenotype. Breast Cancer Res 10(1):R12

Snyder MW, Kircher M, Hill AJ, Daza RM, Shendure J (2016) Cell-free DNA comprises an in vivo nucleosome footprint that informs its tissues-of-origin. Cell 164(1-2):57–68

Socha MJ, Said N, Dai Y, Kwong J, Ramalingam P, Trieu V et al (2009) Aberrant promoter methylation of SPARC in ovarian cancer. Neoplasia 11(2):126–135

Stirzaker C, Taberlay PC, Statham AL, Clark SJ (2014) Mining cancer methylomes: prospects and challenges. Trends Genet 30(2):75–84

Sun K, Jiang P, Chan KC, Wong J, Cheng YK, Liang RH 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(40):E5503–E5512

Takane K, Midorikawa Y, Yagi K, Sakai A, Aburatani H, Takayama T et al (2014) Aberrant promoter methylation of PPP1R3C and EFHD1 in plasma of colorectal cancer patients. Cancer Med 3(5):1235–1245

Tam KF, Liu VW, Liu SS, Tsang PC, Cheung AN, Yip AM et al (2007) Methylation profile in benign, borderline and malignant ovarian tumors. J Cancer Res Clin Oncol 133(5):331–341

Tang Q, Cheng J, Cao X, Surowy H, Burwinkel B (2016) Blood-based DNA methylation as biomarker for breast cancer: a systematic review. Clin Epigenetics 8:115

Thierry AR, Mouliere F, El Messaoudi S, Mollevi C, Lopez-Crapez E, Rolet F et al (2014) Clinical validation of the detection of KRAS and BRAF mutations from circulating tumor DNA. Nat Med 20(4):430–435

Tost J, Gut IG (2007a) DNA methylation analysis by pyrosequencing. Nat Protoc 2(9):2265–2275

Tost J, Gut IG (2007b) Analysis of gene-specific DNA methylation patterns by pyrosequencing technology. Methods Mol Biol 373:89–102

Uehiro N, Sato F, Pu F, Tanaka S, Kawashima M, Kawaguchi K et al (2016) Circulating cell-free DNA-based epigenetic assay can detect early breast cancer. Breast Cancer Res 18(1):129

Visvanathan K, Fackler MS, Zhang Z, Lopez-Bujanda ZA, Jeter SC, Sokoll LJ et al (2017) Monitoring of serum DNA methylation as an early independent marker of response and survival in metastatic breast cancer: TBCRC 005 prospective biomarker study. J Clin Oncol 35:751–758

Wang YC, Hsu HS, Chen TP, Chen JT (2006) Molecular diagnostic markers for lung cancer in sputum and plasma. Ann N Y Acad Sci 1075:179–184

Wang L, Lin YL, Li B, Wang YZ, Li WP, Ma JG (2014) Aberrant promoter methylation of the cadherin 13 gene in serum and its relationship with clinicopathological features of prostate cancer. J Int Med Res 42(5):1085–1092

Wang B, Yu L, Yang GZ, Luo X, Huang L (2015a) Application of multiplex nested methylated specific PCR in early diagnosis of epithelial ovarian cancer. Asian Pac J Cancer Prev 16 (7):3003–3007

Wang H, Zhang B, Chen D, Xia W, Zhang J, Wang F et al (2015b) Real-time monitoring efficiency and toxicity of chemotherapy in patients with advanced lung cancer. Clin Epigenetics 7:119

Warren JD, Xiong W, Bunker AM, Vaughn CP, Furtado LV, Roberts WL et al (2011) Septin 9 methylated DNA is a sensitive and specific blood test for colorectal cancer. BMC Med 9:133

Warton K, Lin V, Navin T, Armstrong NJ, Kaplan W, Ying K et al (2014) Methylation-capture and next-generation sequencing of free circulating DNA from human plasma. BMC Genomics 15:476

Weisenberger DJ, Trinh BN, Campan M, Sharma S, Long TI, Ananthnarayan S et al (2008) DNA methylation analysis by digital bisulfite genomic sequencing and digital MethyLight. Nucleic Acids Res 36(14):4689–4698

Weiss G, Schlegel A, Kottwitz D, Konig T, Tetzner R (2017) Validation of the SHOX2/PTGER4 DNA methylation marker panel for plasma-based discrimination between patients with malignant and nonmalignant lung disease. J Thorac Oncol 12(1):77–84

Wielscher M, Vierlinger K, Kegler U, Ziesche R, Gsur A, Weinhausel A (2015) Diagnostic performance of plasma DNA methylation profiles in lung cancer, pulmonary fibrosis and COPD. EBioMedicine 2(8):929–936

Wong EM, Dobrovic A (2011) Assessing gene-specific methylation using HRM-based analysis. Methods Mol Biol 687:207–217

Wu Q, Lothe RA, Ahlquist T, Silins I, Trope CG, Micci F et al (2007) DNA methylation profiling of ovarian carcinomas and their in vitro models identifies HOXA9, HOXB5, SCGB3A1, and CRABP1 as novel targets. Mol Cancer 6:45

Wu T, Giovannucci E, Welge J, Mallick P, Tang WY, Ho SM (2011) Measurement of GSTP1 promoter methylation in body fluids may complement PSA screening: a meta-analysis. Br J Cancer 105(1):65–73

Wu Y, Zhang X, Lin L, Ma XP, Ma YC, Liu PS (2014) Aberrant methylation of RASSF2A in tumors and plasma of patients with epithelial ovarian cancer. Asian Pac J Cancer Prev 15 (3):1171–1176

Yi JM, Guzzetta AA, Bailey VJ, Downing SR, Van Neste L, Chiappinelli KB et al (2013) Novel methylation biomarker panel for the early detection of pancreatic cancer. Clin Cancer Res 19 (23):6544–6555

Yin BW, Lloyd KO (2001) Molecular cloning of the CA125 ovarian cancer antigen: identification as a new mucin, MUC16. J Biol Chem 276(29):27371–27375

You MW, Kim MH, Kim JK, Cho KS (2016) The characteristics and spatial distributions of initially missed and rebiopsy-detected prostate cancers. Ultrasonography 35(3):226–233

Zarzour P, Hesson LB, Ward RL (2013) Establishing the clinical utility of epigenetic markers in cancer: many challenges ahead. Epigenomics 5(5):513–523

Zhang Y, Wang R, Song H, Huang G, Yi J, Zheng Y et al (2011) Methylation of multiple genes as a candidate biomarker in non-small cell lung cancer. Cancer Lett 303(1):21–28

Zhang Q, Hu G, Yang Q, Dong R, Xie X, Ma D et al (2013) A multiplex methylation-specific PCR assay for the detection of early-stage ovarian cancer using cell-free serum DNA. Gynecol Oncol 130(1):132–139

Zhou F, Ma M, Tao G, Chen X, Xie W, Wang Y et al (2014) Detection of circulating methylated opioid binding protein/cell adhesion molecule-like gene as a biomarker for ovarian carcinoma. Clin Lab 60(5):759–765

Zurita M, Lara PC, del Moral R, Torres B, Linares-Fernandez JL, Arrabal SR et al (2010) Hypermethylated 14-3-3-sigma and ESR1 gene promoters in serum as candidate biomarkers for the diagnosis and treatment efficacy of breast cancer metastasis. BMC Cancer 10:217

The Clinical Utility of Epigenetics: A Case Study

11

Luke B. Hesson, Mathew A. Sloane, and Antonia L. Pritchard

Abstract

In a clinical pathology setting, epigenetic testing is rare and underutilised. In this chapter, we describe a patient with early-onset colorectal cancer caused by a constitutional epimutation. This case study emphasises the importance of epigenetic testing when standard genetic testing has failed to identify a cause for disease predisposition.

Keywords

Clinical testing · Epigenetic testing · Constitutional epimutation · Colorectal cancer · Lynch syndrome · *MLH1*

11.1 Introduction

Familial cancer can be caused by pathogenic germline variants in cancer predisposition genes. Often, however, no pathogenic germline variant can be found, despite a clinical history that is indicative of familial cancer. A possible cause of cancer

L. B. Hesson (✉)
Prince of Wales Clinical School, UNSW, Sydney, NSW, Australia

Garvan Institute for Medical Research, 384 Victoria Street, Darlinghurst, NSW, Australia
e-mail: l.hesson@unsw.edu.au

M. A. Sloane
Prince of Wales Clinical School, UNSW, Sydney, NSW, Australia

A. L. Pritchard (✉)
Genetics and Immunology Research Group, Centre for Health Science, University of the Highlands and Islands, Inverness, UK
e-mail: Antonia.Pritchard@uhi.ac.uk

L. B. Hesson, A. L. Pritchard (eds.), *Clinical Epigenetics*,
https://doi.org/10.1007/978-981-13-8958-0_11

predisposition in these people is the inheritance of a constitutional epimutation in a cancer predisposition gene.

The term epimutation was proposed to describe epigenetic changes in gene activity based on DNA methylation. This distinguishes epimutations from classical mutations, which are due to changes in DNA sequence (Holliday 1985; Jeggo and Holliday 1986). More specifically, the term is used to describe an epigenetic aberration that results in the transcriptional silencing of a gene that is normally active or the expression of a gene that is normally inactive (Hesson et al. 2010). Epimutations that are widely distributed in normal tissues of an individual and predispose to disease are known as constitutional epimutations (Hesson et al. 2010; Hitchins 2015). Constitutional epimutations have been described in a wide variety of genes and diseases (Table 11.1). The criteria for identifying a constitutional epimutation have been described previously (Sloane et al. 2016).

In this chapter, we use the case study of a patient who presented with early-onset colorectal cancer and underwent standard genetic testing of known cancer predisposition genes with negative results, to demonstrate the clinical importance of testing for epimutations. We describe the genetic and epigenetic testing that was performed, which showed that the cause of cancer predisposition in this patient was a constitutional epimutation of the *MLH1* gene. This case has been previously reported (Sloane et al. 2015). Embedded within the text are important questions to encourage consideration of the reasons for the clinical and laboratory decisions made during this process; answers are provided at the end of this chapter.

11.2 Clinical Background

The 29-year-old male patient presented with a moderately differentiated colonic adenocarcinoma of the ascending colon (individual III-7 in family pedigree, Fig. 11.1). Colonoscopic evaluation at presentation showed no evidence of polyposis. Surgery was performed to remove the tumour. Clinical history showed his brother (III-6, Fig. 11.1) had died of acute myeloid leukaemia at age 17 years and his sister (III-8, Fig. 11.1) reported hyperplastic polyps and a low-grade tubular adenoma at age 24 years. This adenoma was located in the ascending colon and showed normal staining of the DNA mismatch repair (MMR) proteins. The patient's maternal grandfather (I-1, Fig. 11.1) had microsatellite stable colorectal cancer and prostate cancer at age 67 years. No other family history of cancer was reported.

Immunohistochemical (IHC) staining of the patient's tumour tissue showed loss of staining of the MLH1 and PMS2 proteins (but not MSH2 or MSH6). Microsatellite instability (unstable at the microsatellite markers BAT25, BAT26, BAT40 and D17S250) was detected with no mutation in the *BRAF* gene.

The patient's young age at onset of colorectal cancer, the molecular features described above and the absence of polyposis lead his doctor to suspect Lynch

Table 11.1 Examples of constitutional epimutations that predispose to disease

Disease	Gene	Genetic variation	Comment	References
Lynch syndrome	MLH1	Not determined	Primary epimutations. Mechanism unknown	Suter et al. (2004), Hitchins et al. (2007), Pineda et al. (2012), Goel et al. (2011), Sloane et al. (2015)
		LRG216t1:c.-27C>A	Segregates faithfully with a sequence variant in the 5' UTR of the *MLH1* gene. Mechanism unknown	Hitchins et al. (2011), Kwok et al. (2014)
		Duplication of a 280–375 kb region that encompassed *MLH1*, three flanking genes *TRANK1*, *EPM2AIP1* and *LRRFIP2*, and the 5' part of *GOLGA4*	Results in mosaic constitutional methylation of *MLH1*	Morak et al. (2011)
	MSH2	Deletions upstream of *MSH2*	Deletion of the transcription termination signal of the upstream *EPCAM* gene results in transcriptional read-through and hypermethylation of the *MSH2* promoter	Ligtenberg et al. (2009), Tutlewska et al. (2013)
Early-onset colorectal cancer	*PTPRJ*	Partial tandem duplication of *PTPRJ*	Germline epigenetic silencing of *PTPRJ* is induced by transcriptional read-through of the promoter of the juxtaposed intact copy of the gene	Venkatachalam et al. (2010)
Early-onset breast cancer	*BRCA1*	LRG292t1:c.-107A>T	Secondary epimutation that segregates faithfully with a sequence variant in the 5' UTR of the *BRCA1* gene. Mechanism unknown	Evans et al. (2018)

(continued)

Table 11.1 (continued)

Disease	Gene	Genetic variation	Comment	References
Fragile X syndrome	*FMR1*	Expansion of a trinucleotide repeat within the 5′ UTR of the *FMR1* gene. Example nomenclature for a repeat of 1000 units is LRG762t1: c.-128GGC(1000), where '()' is used to indicate measurement uncertainty in the precise length of the repeat	Expansion of the trinucleotide repeat to more than ~200 copies results in hypermethylation of the *FMR1* gene promoter	Oberle et al. (1991), Biancalana et al. (2015)
Alpha thalassaemia	*HBA2*	Deletion downstream of *HBA2*	Deletion juxtaposes part of the downstream *LUC7L* gene with the HBA2 gene. This deletion includes the transcription termination signal of *LUC7L* resulting in antisense transcriptional read-through and hypermethylation of the *HBA2* promoter	Tufarelli et al. (2003)
X-linked disease	*XIST*	-43C>A/G relative to XIST RNA transcript (NR_001464)	Abolishes (-43C>A) or enhances (-43C>G) CTCF binding, thereby altering XIST expression and skewing X-inactivation	Tomkins et al. (2002), Pugacheva et al. (2005), Plenge et al., (1997)
B-cell chronic lymphocytic leukaemia (CLL)	*DAPK1*	Point mutation more than 6 kb upstream of the *DAPK1* promoter (c.-6531A>G)	Higher-binding affinity of the HOXB7 transcriptional repressor in the mutant allele results in a reduction in DAPK1 expression to 25% of the wild-type allele and DAPK1 promoter hypermethylation	Raval et al. (2007)

Facioscapulohumeral muscular dystrophy	*DUX4* within the D4Z4 macrosatellite repeat	Contraction of the D4Z4 macrosatellite repeat array at chromosome 4q35 to 10 or fewer copies	The D4Z4 macrosatellite contains multiple copies of a 3.3 kb repeat, with each repeat containing a copy of the *DUX4* gene. Contraction results in hypomethylation and expression of multiple copies of *DUX4*, which is pathogenic	Calandra et al. (2016)
		Pathogenic (inactivating) variants in the *SMCHD1* gene	The product of the *SMCHD1* gene regulates DNA methylation. Pathogenic variants cause hypomethylation of the D4Z4 region in *trans*. Pathogenic variants in *SMCHD1* also cause Bosma arhinia microphthalmia syndrome (BAMS); however it is unclear whether the mechanism of disease involves constitutional epimutations, though differences in the expression of a range of genes involved in craniofacial patterning have been observed. The penetrance of FSHD can also be modified by inactivating variants in the *DNMT3B* gene, which regulates DNA methylation genome wide	Gordon et al. (2017), Shaw et al., (2017), van den Boogaard et al. (2016)

Disorders of genomic imprinting, which are also examples of constitutional epimutations, have been omitted from this table and are described in Chap. 9. Sequence variants are provided using Human Genome Variation Society (HGVS) nomenclature. Where known, the mechanism leading to epimutation is described. Most epimutations are secondary to sequence or structural changes, some within the promoter region of the affected gene. See Liu et al. (2016) for a review of the pathogenicity of mismatch repair gene promoter variants in Lynch syndrome. *UTR* untranslated region, *LRG* Locus Reference Genomic record

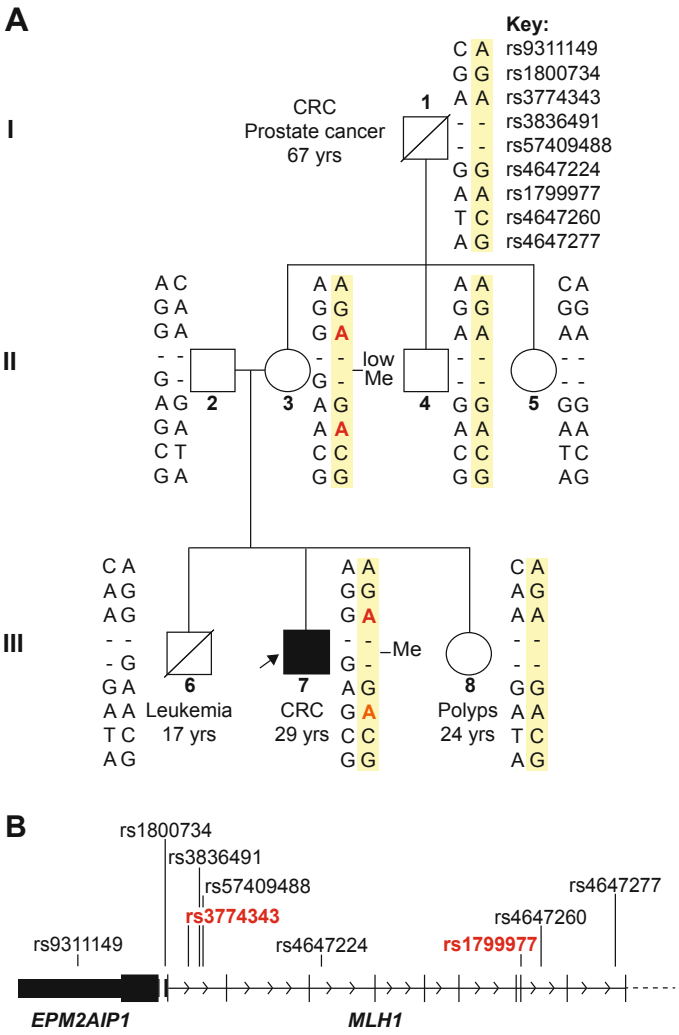

Fig. 11.1 Family pedigree. (**a**) Family pedigree with haplotypes shown adjacent to each family member. The haplotype that harboured the epimutation in the patient (III-7, arrowhead) and his mother is indicated with yellow highlight. The boldface A in the patient's and mother's haplotype indicates the genotype of the methylated allele at rs3774343 and rs1799977. CRC indicates colorectal cancer; dashes, single nucleotide deletion at this location; and Me, methylation. (**b**) The location of single nucleotide polymorphisms (SNPs) used in haplotype analysis. Indicated in bold red text are the informative rs3774343 and rs1799977 SNPs

syndrome as the cause of his colorectal cancer. Based on eviQ guidelines[1], genetic testing of germline DNA (obtained from blood) for pathogenic variants in the MMR genes that demonstrate loss of staining in the tumour was requested.

Question 1: *Why is the presence or absence of polyposis relevant?*
Question 2: *Based on the clinical background above, which MMR proteins should be tested using IHC?*
Question 3: *Why was BRAF testing of the tumour requested?*
Question 4: *Why also test for microsatellite instability?*
Question 5: *For MMR gene testing, what molecular tests should be requested?*
Question 6: *Why do both MLH1 and PMS2 show loss of immunohistochemical staining?*

11.3 Results of Genetic Testing

No germline variants in exons or splice sites or copy number variants were detected in *MLH1* or *PMS2*.

Question 7: *Do the results of genetic testing provide a diagnosis of Lynch syndrome?*
Question 8: *Could the patient have Lynch syndrome, and if so, what types of molecular alterations would not have been detected by the tests performed above?*

11.4 Laboratory Tests Performed to Determine the Presence or Absence of a Constitutional Epimutation

Further tests for a possible constitutional epimutation were requested. Fresh samples of peripheral blood, hair and buccal mucosa were obtained from the patient and his mother and saliva only from other living relatives. Archived bone marrow aspirates (taken to evaluate candidacy for bone marrow transplantation for the treatment of leukaemia) were also available from the patient (age 13 years) and his brother (age 16 years; III-6, Fig. 11.1).

Testing included haplotyping across the *MLH1* locus (Fig. 11.1b), DNA methylation analysis of the *MLH1* promoter (Fig. 11.2) and *MLH1* expression studies. *MLH1* promoter methylation testing was performed using methylation-specific multiplex ligation-dependent probe amplification (MS-MLPA), quantitative methylation-specific PCR (qMSP), pyrosequencing and allelic bisulphite sequencing. Expression studies were performed using pyrosequencing (as described

[1]Refer to eviQ guidelines (https://www.eviq.org.au).

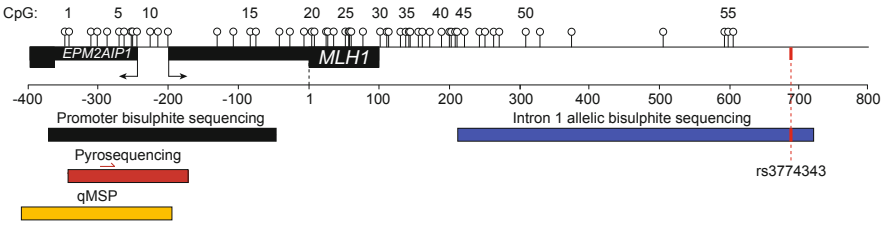

Fig. 11.2 A map of the *MLH1* promoter CpG island and the region investigated with each assay. The small circles indicate the location of CpG dinucleotides. Numbered are the CpG dinucleotides (top) and the nucleotide position relative to the adenine within the ATG translation start codon of *MLH1*:LRG216t1. Small red arrow indicates the location of the nested pyrosequencing primer, which provides a measurement of methylation levels of CpG dinucleotides 5–9. Indicated is the SNP rs3774343, which was used to distinguish alleles. qMSP, quantitative methylation-specific polymerase chain reaction

previously (Kwok and Hitchins 2015)), and haplotypes across the *MLH1* locus were obtained using Sanger sequencing.

Question 9: *Why are multiple tissue types requested to test for a constitutional epimutation?*

Question 10: *What is the value of haplotyping in this scenario?*

11.5 Results of Testing to Determine the Presence or Absence of a Constitutional Epimutation and Interpretation

MS-MLPA and pyrosequencing detected *MLH1* promoter hypermethylation (27%–56%) in normal tissues from all three germ layers, as well as in tumour tissue (Fig. 11.3a). Bisulphite sequencing of the *MLH1* promoter in blood, saliva, hair follicles and bone marrow showed dense methylation limited to the maternally inherited allele (Fig. 11.3b and c) that was also transcriptionally silent (Fig. 11.3d).

Pyrosequencing (Fig. 11.4a) and qMSP (Fig. 11.4b) showed the patient's mother had approximately 2–5% methylation in tissues derived from all germ layers. No methylation was detected in four other family members (Fig. 11.4b), including the patient's sister, who inherited the same maternal allele, and his brother, who inherited the other maternal allele. Both of the patient's siblings had biallelic expression of *MLH1* (Fig. 11.4c). Confirming these observations, the patient's mother's samples showed dense methylation of approximately 5% of *MLH1* promoter molecules that was confined to the same allele transmitted to the patient (Fig. 11.5).

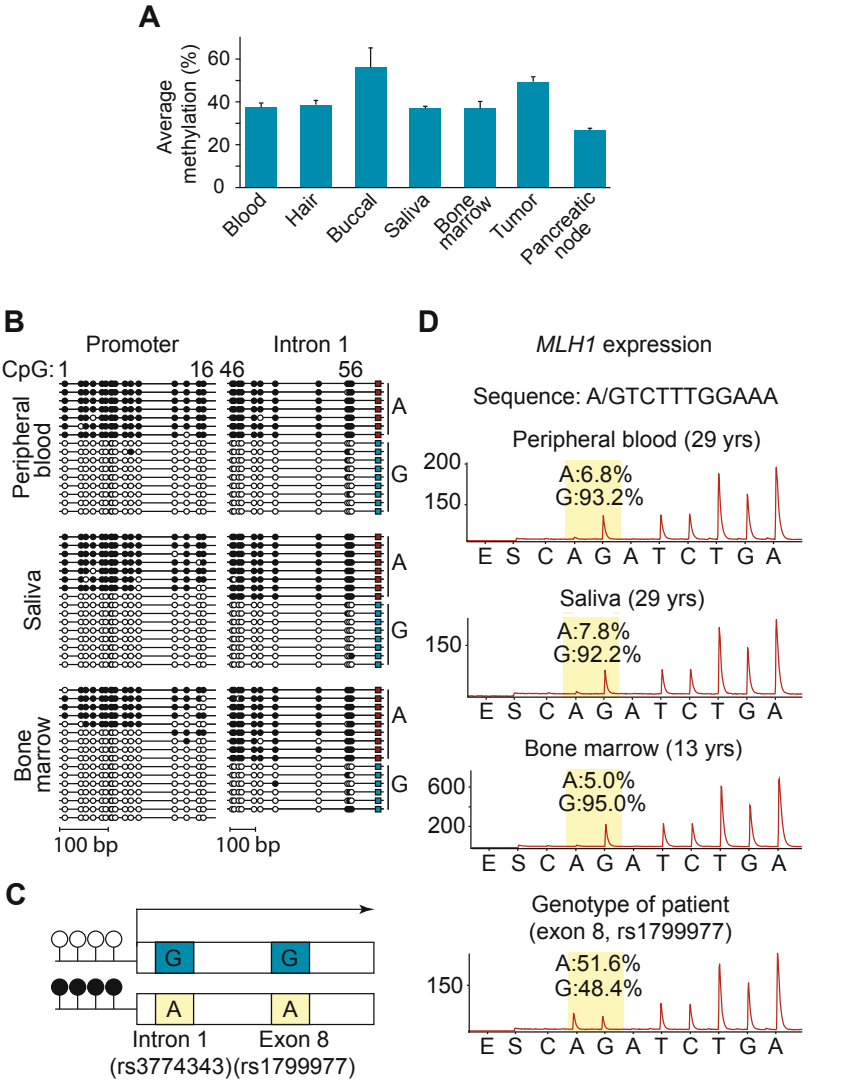

Fig. 11.3 *MLH1* methylation and expression in the patient. (**a**) Mean methylation levels detected in constitutional and tumour DNA of the patient by means of pyrosequencing. (**b**) Allelic bisulphite sequencing of the *MLH1* promoter and intron 1 in the patient. Each line represents an individual molecule, with white and black circles denoting unmethylated and methylated CpG sites, respectively. The red (A) and blue (G) squares represent the rs3774343 SNP within intron 1 used to distinguish alleles (in the patient, the genotype A is on the maternal allele). (**c**) A schematic summarising the *MLH1* methylation and expression data for the patient. (**d**) Pyrograms showing the relative levels of *MLH1* expression from each allele in the patient. The heterozygous germline genotype at the rs1799977 SNP is shown in the bottom pyrogram. The reference sequence for the pyrosequencing assay is shown at the top

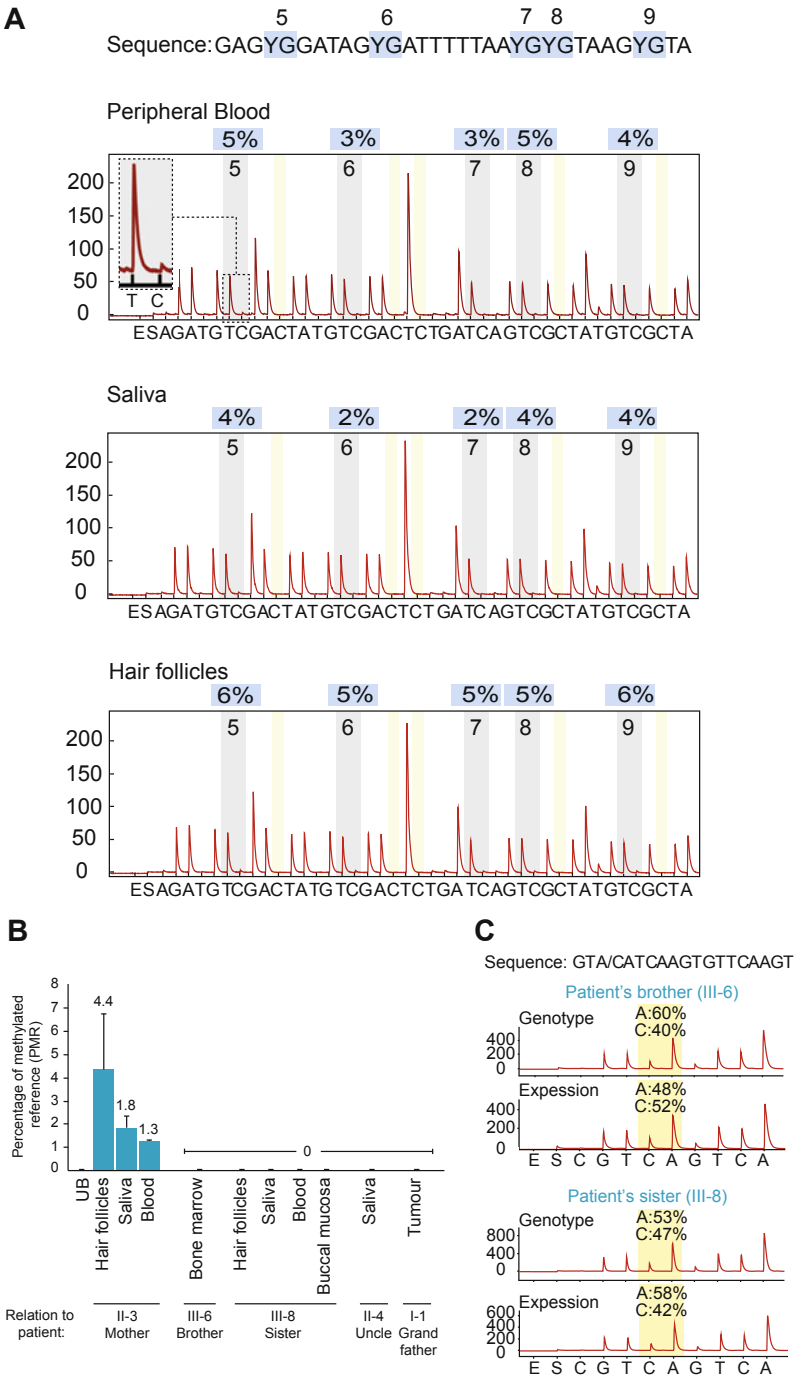

Fig. 11.4 *MLH1* methylation and expression in family members. (**a**) Pyrograms generated by pyrosequencing of bisulphite-modified DNA from the patient's mother. The percentages shown at the top of each grey bar represent the quantified DNA methylation level at five CpG sites (CpGs 5–9) in the promoter. *Y*-axis, relative light units. (**b**) Levels of *MLH1* CpG island methylation in various family

Question 11: *What is a possible cause of false-positive methylation using MS-MLPA?*

Question 12: *Do the results of MLH1 methylation and expression studies confirm the presence of a constitutional epimutation?*

Question 13: *Why was methylation at the MLH1 promoter and intron 1 investigated?*

Question 14: *Does the patient's mother have a constitutional epimutation of MLH1 and is she predisposed to cancer?*

11.6 Comments

The findings of molecular testing show that the patient's cancer predisposition was caused by a constitutional *MLH1* epimutation on his maternally inherited allele. Genetic and epigenetic testing of the patient's mother is consistent with her being a gonosomal mosaic (Fig. 11.6). Estimates suggest that *MLH1* epimutations account for 1–10% of patients who meet the clinical diagnostic criteria for Lynch syndrome with loss of MLH1 expression in their tumours by IHC but in whom standard genetic testing has failed to detect a pathogenic germline MMR gene variant (Hitchins and Lynch 2014).

This case study was chosen as it highlights several features that require consideration when testing for the presence of a constitutional epimutation.

Non-Mendelian Inheritance The patient inherited a constitutional epimutation on the maternal allele and developed cancer. The patient's sister inherited the same maternal allele; however the absence of a constitutional epimutation in *MLH1* in her shows she is not at high risk of developing cancer; the colonic polyps in her are unrelated to Lynch syndrome. This non-Mendelian inheritance was due to mosaicism in the mother rather than erasure of the epimutation in the germline, as described in other cases (see Hitchins et al. (2007) for an example). As demonstrated by this case study, an asymptomatic mosaic individual can have a non-mosaic affected child. Indeed, the majority of primary constitutional epimutations (see below) in Lynch syndrome occur in the absence of a family history of Lynch syndrome-associated cancers.

Example of a Primary Constitutional Epimutation Epimutations are identified as either primary (those not associated with a *cis* genetic variant) or secondary (those

Fig. 11.4 (continued) members measured with qMSP. UB indicates unmethylated DNA from the peripheral blood of a healthy control. (**c**) Pyrograms showing the heterozygous germline genotype and expression of both *MLH1* alleles at the rs9311149 SNP in the patient's brother and sister. *Y*-axis, relative light units. The reference sequence for the pyrosequencing assay is shown at the top

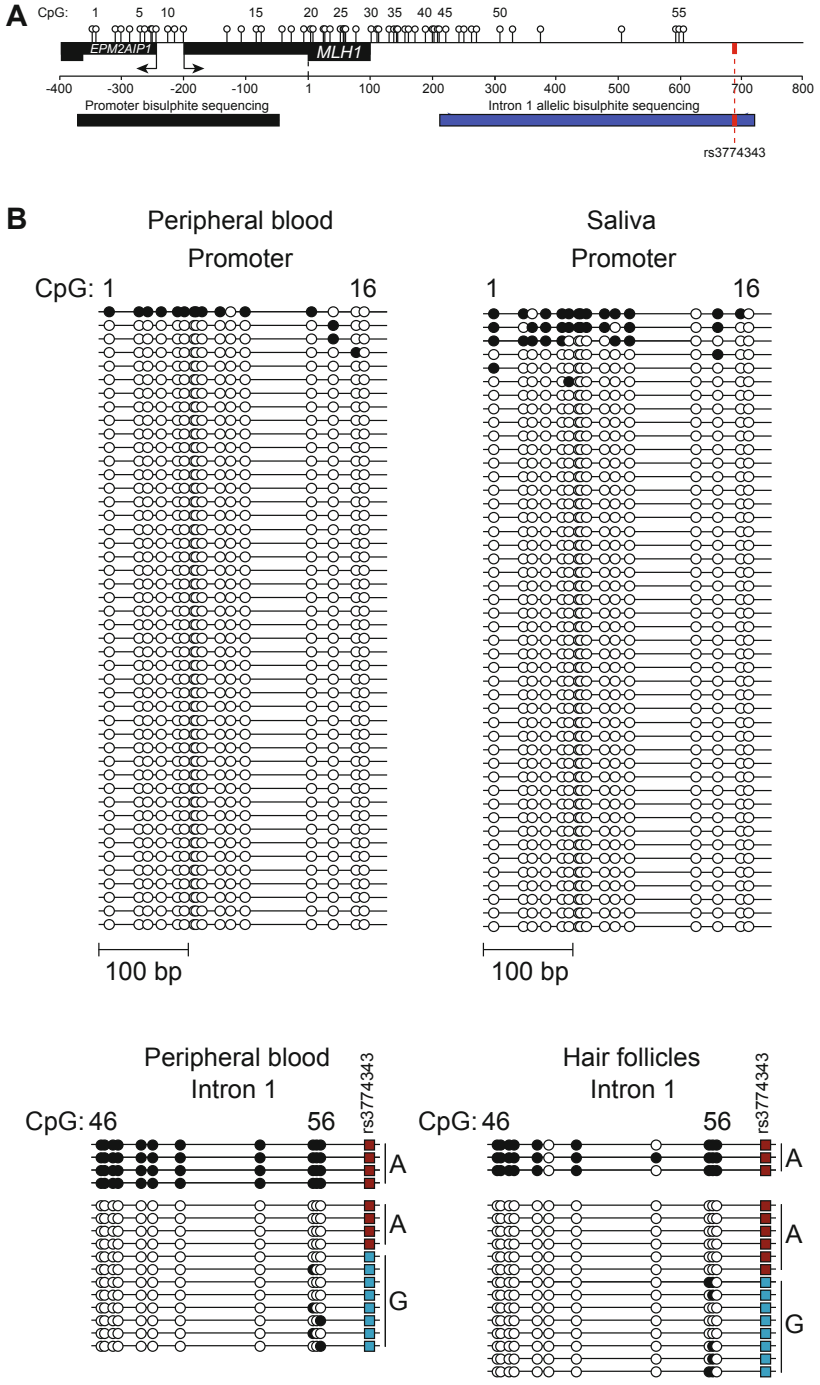

Fig. 11.5 Mosaic *MLH1* methylation in the patient's mother. (**a**) Map of the *MLH1* CpG island promoter and the regions investigated in Panel B. (**b**) Allelic bisulphite sequencing of the *MLH1* promoter and intron 1 in the patient's mother. Each line represents an individual molecule, with

that are associated with a *cis*-acting sequence or structural variant) epimutations. In this case study, the absence of any potentially pathogenic sequence or structural alterations across the *MLH1* gene identifies it as a primary constitutional epimutation. This is an important distinction to make, because primary epimutations can only be detected by epigenetic analyses. Furthermore, secondary epimutations often segregate faithfully with the *cis*-acting genetic alteration and show Mendelian inheritance.

Stability of Constitutional Epimutations By analysing several tissue types taken from the proband over a 16-year period, this case study demonstrates that epimutations are stable over the lifetime of an individual.

Specimen Types One of the criteria for identifying an epimutation as constitutional is demonstrating its presence in all three germ layers of an individual (Sloane et al. 2016). The analysis of DNA from peripheral blood lymphocytes, hair follicles and buccal cells (representative of the mesoderm, ectoderm and endoderm, respectively) helps to determine whether the epimutation is soma-wide, which would suggest it arose early in embryonic development.

Importance of Haplotype Analysis Haplotype analysis of individuals within a family is important to identify which allele harbours the epimutation and is relevant for identifying parental origin, pattern of inheritance and other family members that may benefit from epigenetic testing for risk assessment.

In conclusion, this case study highlights the potential importance of epimutations in familial disorders. It also demonstrates that the inheritance of an epimutation can be complex and requires careful interpretation using several sources of DNA and RNA, as well as the analysis of family members in conjunction with their medical history. Specific for Lynch syndrome, there are some useful resources for clinical management that can be consulted for further information, including genetic testing, which can be found at the InSiGHT database (Thompson et al. 2014) and eviQ, an Australian government, freely available online resource of cancer treatment protocols developed by multidisciplinary teams of cancer specialists.

11.7 Answers to the Questions Embedded in the Text

Answer 1 The presence of polyposis, in which the colonic mucosa contains hundreds to thousands of polyps, helps to distinguish familial adenomatous polyposis (FAP) from other syndromes that predispose to colorectal cancer. FAP

Fig. 11.5 (continued) white and black circles denoting unmethylated and methylated CpG sites, respectively. The red (A) and blue (G) squares represent the rs3774343 SNP within intron 1 used to distinguish alleles

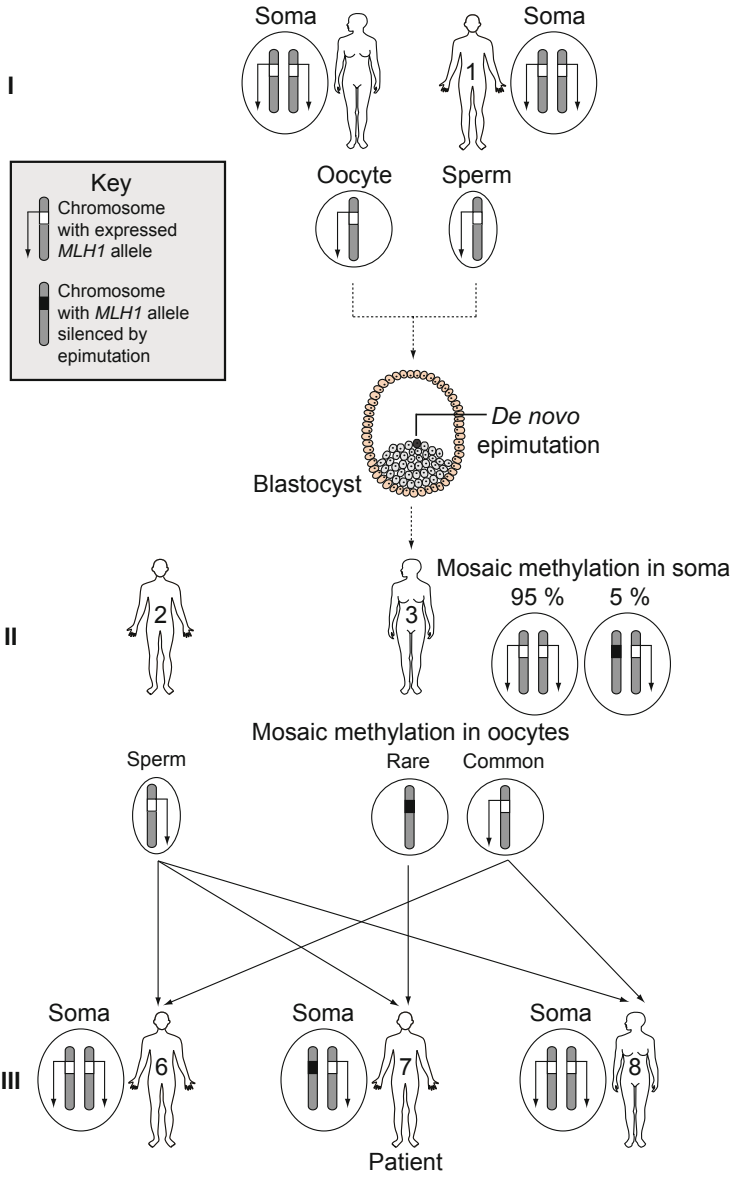

Fig. 11.6 A model to explain the non-Mendelian inheritance pattern. The non-Mendelian pattern of inheritance is explained by an epigenetic mosaic model

is caused by the inheritance of pathogenic sequence variants in the gene *APC*. The absence of polyposis indicates FAP is highly unlikely to be the cause of the patient's cancer.

Answer 2 Pathogenic germline variants of *MLH1*, *MSH2*, *MSH6* and *PMS2* occur in 35–40%, 44–48%, 8–10% and 2–8% of cases of Lynch syndrome, respectively (Schofield et al. 2014; Steinke et al. 2014). Given that loss of either is a possible cause of early-onset cancer, staining for all four of the MMR proteins (MLH1, PMS2, MSH2 and MSH6) should be performed. The pattern of immunohistochemical loss of MMR protein should be used to guide which MMR genes are tested (see Answer 5).

Answer 3 Mutations in *BRAF* can distinguish sporadic colorectal cancers with microsatellite instability from familial colorectal cancers with microsatellite instability. Loss of MMR protein staining in a tumour without *BRAF* mutation would be strong evidence of Lynch syndrome rather than sporadic cancer.

Answer 4 MSI testing is complementary to MMR protein staining and will provide a molecular assessment of mismatch repair proficiency in tumour tissue. MSI testing will also identify cases of false-positive MMR protein staining or partial loss of MMR protein levels that are difficult to detect by IHC.

Answer 5 Full gene sequencing and copy number analysis of MMR genes showing loss of staining by IHC should be performed[2]. This is because pathogenic sequence variants and structural alterations have been described throughout the *MLH1*, *PMS2*, *MSH2* and *MSH6* genes. Testing of MMR genes that do not show loss of staining by IHC is not required. The types of genetic alterations that can cause Lynch syndrome include single nucleotide variants (SNVs), indels, complete or partial gene deletion, partial gene duplication or rearrangements, including inversions (Liu et al. 2016). Sanger sequencing or massively parallel sequencing (MPS) would be appropriate for SNV and indel detection. Copy number alterations can be detected using multiplex ligation-dependent probe amplification (MLPA), though these may also be detected by appropriate MPS approaches.

Answer 6 Loss of both MLH1 and PMS2 protein staining typically occurs together. These proteins bind to form a heterodimer in the cell, and the loss of MLH1 results in the degradation of its binding partner PMS2. Likewise, MSH2 and MSH6 also form a heterodimer, and the loss of either MSH2 or MSH6 will lead to degradation of the binding partner.

Answer 7 No, these findings are not diagnostic for Lynch syndrome because a pathogenic variant has not yet been identified.

Answer 8 Yes, the patient could have Lynch syndrome, but conclusive diagnosis of Lynch syndrome requires specific criteria (see Answer 12). It is important to note

[2]Refer to eviQ guidelines for a flow chart describing IHC-guided genetic testing (https://www.eviq.org.au/additional-clinical-information-cancer-genetics/3185-ihc-guided-genetic-testing).

that standard molecular testing cannot identify a genetic cause in around 30% of cases with a strong clinical suspicion of Lynch syndrome. The types of molecular alteration that may not be detected by standard molecular testing (such as that described above) include:

- Non-coding variants, such as a promoter variant that affects gene expression or an intronic variant that creates or abolishes a splice site (sometimes referred to as cryptic splicing variants). The promoter and intronic DNA of MMR genes may not have been sequenced, or a variant detected within these regions may have not been reported due to difficulty in interpretation (i.e. a variant of uncertain significance [VUS] was identified). VUSs may be pathogenic or benign, but there is insufficient functional evidence to classify as either at the time of reporting.
- Structural rearrangements, such as an inversion, that does not affect exon copy number or sequence.
- Constitutional epimutations that cause soma-wide transcriptional inactivation of one allele of a gene[3].

Finally, the causes of false-negative results depend on the laboratory method used in testing. For example, allele dropout may occur in methods dependent on PCR amplification, such as Sanger sequencing or some MPS-based methods. This can be caused by a polymorphism at the binding site of a PCR primer that prevents amplification of one allele. Whole genome sequencing could be performed to identify potentially pathogenic non-coding sequence or structural alterations. Or epigenetic testing could be performed to detect the presence of a constitutional epimutation in *MLH1*.

Answer 9 A constitutional epimutation of a gene is defined as the soma-wide hypermethylation and transcriptional silencing of one allele. The most definitive way of assessing this is by testing samples taken from all three germ layers from the patient. Peripheral blood, hair follicles and saliva and buccal mucosa are representative derivatives of the mesoderm, ectoderm and endoderm, respectively. Constitutional epimutations are usually present in all three germ layers; however examples of mosaic constitutional epimutations in *MLH1* and *MSH2* have been described (Hitchins et al. 2011; Ligtenberg et al. 2009).

Answer 10 Haplotyping can indicate which allele harbours the epimutation and identify family members that have inherited the same haplotype. This is important in determining which other family members could be tested for the presence of an

[3]Constitutional epimutations of *MLH1,* but not *PMS2,* have been described in Lynch syndrome. Constitutional epimutations of *PMS2* are theoretically possible but may not occur or may not have been detected due to the presence of multiple *PMS2* pseudogenes that make methylation analysis of this gene technically challenging.

epimutation to aid in the assessment of their cancer risk. It is important to note that whilst the majority of *MLH1* epimutations are inherited on the maternal allele, inheritance on the paternal allele is also possible (see Goel et al. (2011) for an example).

Answer 11 MS-MLPA uses a methylation-sensitive restriction enzyme known as HhaI that will not cut DNA if cytosine in the recognition sequence (GCGC) is methylated. The most likely cause of false-positive methylation is the presence of a sequence variant that abolishes a HhaI recognition sequence. This will prevent HhaI from cutting at that site on one allele, thereby mimicking methylation. This type of false-positive methylation is usually distinguishable from genuine methylation by the absence of detectable methylation at other HhaI sites within the same gene promoter. For example, the *MLH1* promoter contains 7 HhaI sites, and the presence of the SNV *MLH1*[NM_000249.3]:c.-7C>T abolishes a single HhaI site. This results in apparent methylation of only one HhaI site within the *MLH1* promoter, with no methylation detected at six other HhaI sites (see Hesson et al. (2015) for an example).

Answer 12 Yes, the patient has a constitutional epimutation of *MLH1*, which by definition inactivates one copy of *MLH1* throughout all normal tissues. According to the International Society for Gastrointestinal Hereditary Tumours (InSiGHT) criteria (Thompson et al. 2014), a regulatory aberration that leads to defective transcription is sufficient to classify the aberration as pathogenic. However, a diagnosis of Lynch syndrome will require further evidence such as co-segregation of the epimutation with disease or the diagnosis of two or more tumours with a Lynch syndrome molecular phenotype. However, co-segregation of the epimutation with disease is a criterion that is often not observed due to the fact that constitutional epimutations in *MLH1* can show non-Mendelian inheritance.

Answer 13 Methylation at both the *MLH1* promoter and intron 1 was investigated in order to determine which allele (maternal or paternal) was methylated. This can only be done if there are sequence differences between the two alleles in the region investigated for methylation. In this patient, the sequence of the promoter region was identical on both alleles (i.e. no SNPs were present that could be used to distinguish them). However, genotyping across the *MLH1* gene using Sanger sequencing (see Fig. 11.1b) identified sequence differences between the two alleles, including the rs3774343 SNP in intron 1. This region was also hypermethylated because it flanks the CpG island promoter. The data shows that methylation in intron 1 was confined to the maternally inherited allele and, by inference, that the methylation seen on ~50% of the promoter molecules sequenced is also on the maternally inherited allele.

Answer 14 The patient's mother is mosaic for an *MLH1* epimutation. The evidence presented in this case study is consistent with a scenario in which this arose as a de novo event in an embryonic stem cell as she developed in utero. That abnormal embryonic stem cell gave rise to the dense methylation seen in approximately 5% of

the cells in her body, whereas her remaining cells are derived from unaffected embryonic stem cells. We know from the results of molecular studies that the *MLH1* epimutation must also be present in a proportion of her germ cells because the patient has inherited the epimutation. This defines the patient's mother as a "gonosomal mosaic" and suggests that the patient arose from the fertilisation of a rare oocyte with the epimutation (Fig. 11.6). His sister, who inherited the same *MLH1* allele, must have been derived from an oocyte without the epimutation (Fig. 11.6). The risk of disease in individuals with mosaic epimutations is variable, difficult to quantify and dependent on the gene involved, the pattern of mosaicism and the proportion of affected cells. Low-level constitutional methylation can occur in other cancer predisposition genes including *BRCA1* (Hansmann et al. 2012) and *CDNK2A* (Hyland et al. 2014); however these cases show negligible increased cancer risk.

References

Biancalana V et al (2015) EMQN best practice guidelines for the molecular genetic testing and reporting of fragile X syndrome and other fragile X-associated disorders. Eur J Hum Genet 23 (4):417–425

Calandra P et al (2016) Allele-specific DNA hypomethylation characterises FSHD1 and FSHD2. J Med Genet 53(5):348–355

Evans DGR et al (2018) A dominantly inherited 5' UTR variant causing methylation-associated silencing of BRCA1 as a cause of breast and ovarian cancer. Am J Hum Genet 103(2):213–220

Goel A et al (2011) De novo constitutional MLH1 epimutations confer early-onset colorectal cancer in two new sporadic Lynch syndrome cases, with derivation of the epimutation on the paternal allele in one. Int J Cancer 128(4):869–878

Gordon CT et al (2017) De novo mutations in SMCHD1 cause Bosma arhinia microphthalmia syndrome and abrogate nasal development. Nat Genet 49(2):249–255

Hansmann T et al (2012) Constitutive promoter methylation of BRCA1 and RAD51C in patients with familial ovarian cancer and early-onset sporadic breast cancer. Hum Mol Genet 21 (21):4669–4679

Hesson LB et al (2010) Epimutations and cancer predisposition: importance and mechanisms. Curr Opin Genet Dev 20(3):290–298

Hesson LB et al (2015) Lynch syndrome associated with two MLH1 promoter variants and allelic imbalance of MLH1 expression. Hum Mutat 36(6):622–630

Hitchins MP (2015) Constitutional epimutation as a mechanism for cancer causality and heritability? Nat Rev Cancer 15(10):625–634

Hitchins MP, Lynch HT (2014) Dawning of the epigenetic era in hereditary cancer. Clin Genet 85 (5):413–416

Hitchins MP et al (2007) Inheritance of a cancer-associated MLH1 germ-line epimutation. N Engl J Med 356(7):697–705

Hitchins MP et al (2011) Dominantly inherited constitutional epigenetic silencing of MLH1 in a cancer-affected family is linked to a single nucleotide variant within the 5'UTR. Cancer Cell 20 (2):200–213

Holliday R (1985) The Genetical Society abstracts of papers presented at the Two Hundred and Second Meeting of the Society on the 1st, 2nd and 3rd April 1985 at the University of Edinburgh. Heredity 55:277–287

Hyland PL et al (2014) Constitutional promoter methylation and risk of familial melanoma. Epigenetics 9(5):685–692

Jeggo PA, Holliday R (1986) Azacytidine-induced reactivation of a DNA repair gene in Chinese hamster ovary cells. Mol Cell Biol 6(8):2944–2949

Kwok CT, Hitchins MP (2015) Allele quantification pyrosequencing(R) at designated SNP sites to detect allelic expression imbalance and loss-of-heterozygosity. Methods Mol Biol 1315:153–171

Kwok CT et al (2014) The MLH1 c.-27C>A and c.85G>T variants are linked to dominantly inherited MLH1 epimutation and are borne on a European ancestral haplotype. Eur J Hum Genet 22(5):617–624

Ligtenberg MJ et al (2009) Heritable somatic methylation and inactivation of MSH2 in families with Lynch syndrome due to deletion of the 3' exons of TACSTD1. Nat Genet 41(1):112–117

Liu Q et al (2016) Understanding the pathogenicity of noncoding mismatch repair gene promoter variants in Lynch syndrome. Hum Mutat 37(5):417–426

Morak M et al (2011) Biallelic MLH1 SNP cDNA expression or constitutional promoter methylation can hide genomic rearrangements causing Lynch syndrome. J Med Genet 48(8):513–519

Oberle I et al (1991) Instability of a 550-base pair DNA segment and abnormal methylation in fragile X syndrome. Science 252(5009):1097–1102

Pineda M et al (2012) MLH1 methylation screening is effective in identifying epimutation carriers. Eur J Hum Genet 20(12):1256–1264

Plenge RM et al (1997) A promoter mutation in the XIST gene in two unrelated families with skewed X-chromosome inactivation. Nat Genet 17(3):353–356

Pugacheva EM et al (2005) Familial cases of point mutations in the XIST promoter reveal a correlation between CTCF binding and pre-emptive choices of X chromosome inactivation. Hum Mol Genet 14(7):953–965

Raval A et al (2007) Downregulation of death-associated protein kinase 1 (DAPK1) in chronic lymphocytic leukemia. Cell 129(5):879–890

Schofield L et al (2014) Population-based screening for Lynch syndrome in Western Australia. Int J Cancer 135(5):1085–1091

Shaw ND et al (2017) SMCHD1 mutations associated with a rare muscular dystrophy can also cause isolated arhinia and Bosma arhinia microphthalmia syndrome. Nat Genet 49(2):238–248

Sloane MA et al (2015) Mosaic epigenetic inheritance as a cause of early-onset colorectal cancer. JAMA Oncol 1(7):953–957

Sloane MA et al (2016) Defining the criteria for identifying constitutional epimutations. Clin Epigenetics 8:39

Steinke V et al (2014) Evaluating the performance of clinical criteria for predicting mismatch repair gene mutations in Lynch syndrome: a comprehensive analysis of 3,671 families. Int J Cancer 135(1):69–77

Suter CM et al (2004) Germline epimutation of MLH1 in individuals with multiple cancers. Nat Genet 36(5):497–501

Thompson BA et al (2014) Application of a 5-tiered scheme for standardized classification of 2,360 unique mismatch repair gene variants in the InSiGHT locus-specific database. Nat Genet 46 (2):107–115

Tomkins DJ et al (2002) Lack of expression of XIST from a small ring X chromosome containing the XIST locus in a girl with short stature, facial dysmorphism and developmental delay. Eur J Hum Genet 10(1):44–51

Tufarelli C et al (2003) Transcription of antisense RNA leading to gene silencing and methylation as a novel cause of human genetic disease. Nat Genet 34(2):157–165

Tutlewska K et al (2013) Germline deletions in the EPCAM gene as a cause of Lynch syndrome – literature review. Hered Cancer Clin Pract 11(1):9

van den Boogaard ML et al (2016) Mutations in DNMT3B modify epigenetic repression of the D4Z4 repeat and the penetrance of facioscapulohumeral dystrophy. Am J Hum Genet 98 (5):1020–1029

Venkatachalam R et al (2010) Germline epigenetic silencing of the tumor suppressor gene PTPRJ in early-onset familial colorectal cancer. Gastroenterology 139(6):2221–2224