



Implications of Genotype and Environment on Variation in DNA Methylation

6

Ives Y. Lim, Xinyi Lin, and Neerja Karnani

Contents

Introduction	107
Epigenetic Variants	107
DNA Methylation Is a Complex Quantitative Epigenetic Mark	107
DNA Methylation Patterns	109
Genetics of Methylome Diversity	110
SNP-meQTLs	110
SNP-ASMs	112
Environmental Influences and the Variability in the DNA Methylome	113
Alternate Gene Splicing	113
Metastable Epialleles	115
Xenobiotics	115
Combined Influences of Genotype and Environment on the Variability in the Methylomes	118
Conclusion	120
Mini-Dictionary of Terms	121
Key Facts of Epigenome-Wide Association Studies (EWAS)	121
Key Facts of Gene-Environment Interactions (GxE)	121
Key Facts of Xenobiotics	121
Summary Points	122
References	122

Abstract

The epigenome is highly plastic and reacts to the changing external conditions such as diet, lifestyle, and toxins throughout the lifespan. Epigenome-wide association studies (EWAS) provide an opportunity to identify epigenetic

I. Y. Lim · X. Lin · N. Karnani (✉)
Singapore Institute for Clinical Sciences (SICS), A*STAR, Brenner Centre for Molecular Medicine,
Singapore, Singapore
e-mail: Ives_Lim@sics.a-star.edu.sg; lin_xinyi@sics.a-star.edu.sg; neerja_karnani@sics.a-star.edu.sg

© Springer Nature Switzerland AG 2019

105

V. B. Patel, V. R. Preedy (eds.), *Handbook of Nutrition, Diet, and Epigenetics*,
https://doi.org/10.1007/978-3-319-55530-0_56

variants associated with such exposures and the associated suboptimal health outcomes. DNA methylation at 5-methylcytosine is a routinely interrogated epigenetic mark in such EWAS studies. However, depending upon the choice of biological sample, a reliable quantification of the change in methylome at a genomic locus is often confounded by cellular heterogeneity. In addition, the interpretation of cause and effect of this methylation diversity in epigenomes is further complicated by the contributions from genotype and its interaction with the environment, thereby warranting a more sophisticated approach to reliably measure and interpret EWAS findings. This chapter discusses the factors influencing the variability in DNA methylome and its implications on biological interpretations.

Keywords

DNA methylation · Epigenome-wide association studies · Cell heterogeneity · meQTL · Allele-specific methylation · Epistasis · Genotype · Environment · Metastable epialleles · Gene-environment interactions · Causality · Nutrition

List of Abbreviations

AHRR	Aryl-hydrocarbon receptor repressor
ASM	Allele-specific methylation
CGI	CpG islands
CNV	Copy number variation
CYP1A1	Cytochrome P450 family 1 subfamily A member 1
DNMT1	DNA methyltransferase 1
eQTL	Expression quantitative trait locus
EWAS	Epigenome-wide association study
FKBP5	FK506 binding protein 5
GNAS	Guanine nucleotide binding protein, Alpha stimulating
GRB10	Growth factor receptor bound protein 10
HIF1A	Hypoxia-inducible factor 1-alpha
LD	Linkage disequilibrium
MAT	Methionine adenosyltransferase
meQTL	Methylation quantitative trait locus
P-ASM	Parent-of-origin allele-specific methylation
PCR	Polymerase chain reaction
POP	Persistent organic pollutant
PTSD	Post-traumatic stress disorder
SAM	S-adenosylmethionine
S-ASM	Sequence-dependent allele-specific methylation
SNP	Single nucleotide polymorphism
TET	Ten-eleven translocation

Introduction

Epidemiological evidence suggests that nutrition and environmental exposures during prenatal phase and eventually in the life course play an important role in developing suboptimal health trajectories. Epigenetic modifications can occur in response to such environmental stimuli and diet and can change the expression of genes associated with pathological and physiological processes, including fetal development, aging, and common human diseases. Additionally, epigenetic changes due to existing pathophysiologicals can also increase susceptibility to external stimuli, thereby making the use of environment an important factor in EWAS studies. Such exposures during a life course can introduce epigenetic alterations in different cellular populations within the organism. There is also a growing body of evidence that besides the environment, the epigenome can be influenced by underlying genetic variants. Hence, understanding the interplay between epigenome, genotype, and environment is critical for a proper understanding of how epigenetic variants contribute to a phenotype. Since DNA methylation is the most routinely interrogated epigenetic mark in clinical studies, in this chapter we will discuss multiple factors influencing its quantitation and interpretation. We will first discuss the most likely genomic loci where epigenomic variation, both within and between individuals, can be found. Subsequently, we will examine how epigenomic variation can occur, by looking into how genomic, environmental, and the combination of the two can influence DNA methylation. Central to our discussions will be the biological interpretations from DNA methylation variability reported within various investigations and its implications on the eventual phenotype/physiology.

Epigenetic Variants

DNA Methylation Is a Complex Quantitative Epigenetic Mark

DNA methylation is an epigenetic mechanism that can modify the function of a gene by adding a methyl group (CH_3 -) on DNA. Covalent addition of a methyl group on the fifth carbon of a cytosine residue in a cytosine-guanine (CpG) dinucleotide pair is the most widely characterized DNA methylation process. The biological relevance of 5-methylcytosine as a major epigenetic modification in phenotype and gene expression is highly recognized and hence routinely assayed in epigenome-wide association studies (EWAS). Methylation of DNA on a CpG is a binary property, i.e., a CpG is either methylated (100% methylation) or unmethylated (0% methylation). However, somatic cells have two alleles per genome; hence, a single genomic locus within one cell can possess a trinary methylation state (0%, 50%, or 100% methylation), (Fig. 1). Adding to this complexity is the quantitation of methylation from a cell culture or a tissue sample, as the methylation status of a genomic locus from

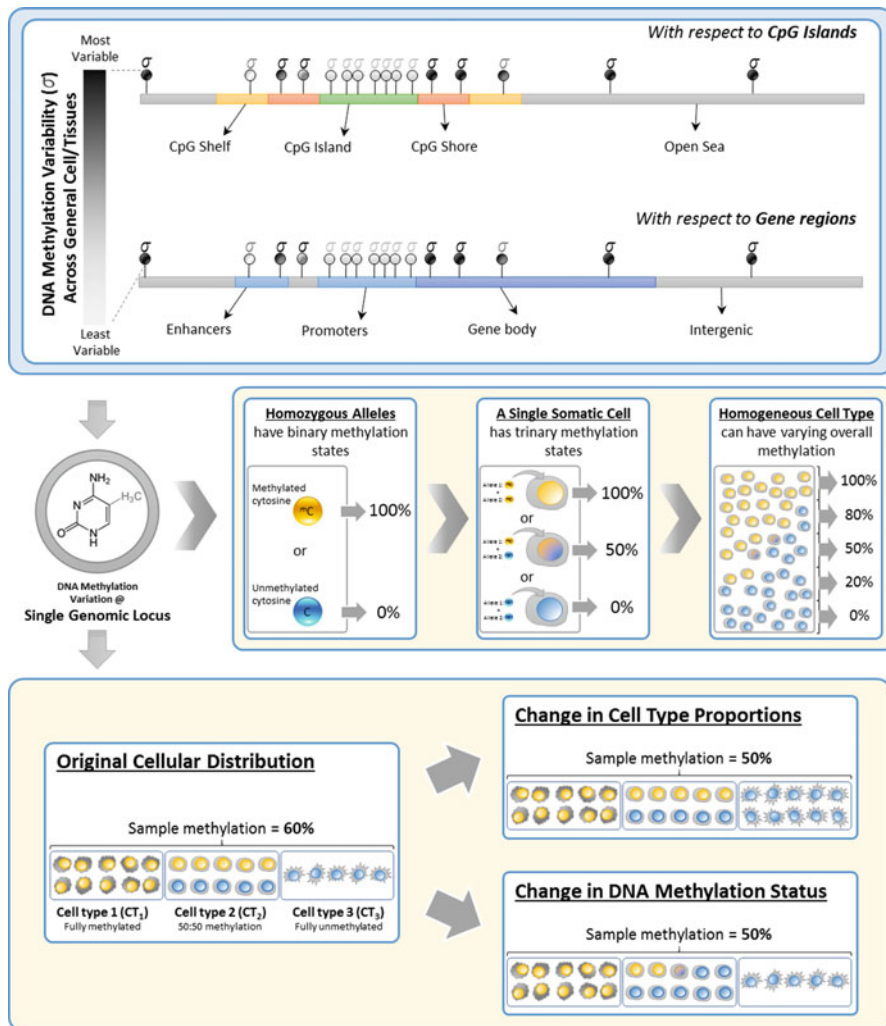


Fig. 1 Distribution and detection of variable CpGs in the genome. *Top panel* provides a locus centric annotation of variable CpGs typically found in the genome. Each CpG is represented by a ball and a stick symbol; *white to black* gradient inside the ball represents the variability in DNA methylation. The darker the color of the ball, the higher is the CpG variability (see the *vertical scale bar* for reference). For example, it is more likely to find variable CpGs in CpG shores and open seas than CpG islands. The *middle panel* shows that within a single genomic locus, homozygous alleles can only possess one of the two binary DNA methylation states. Since somatic cells have two alleles, each cell can possess one of the three DNA methylation states. Permutations of DNA methylation states within each cell can lead to variability in DNA methylation at that CpG within a biological sample. The *bottom panel* illustrates how two major biological events drive DNA methylation changes within the same biological samples. First, as most biological tissues consist of multiple cell types, a change in cellular proportions (either by increasing or decreasing select cellular subpopulations) can lead to overall DNA methylation changes. Second, one or more cell types within a biological sample could have acquired changes within its DNA methylation status, leading to an overall DNA methylation change

such samples is the combined measure from multiple cells, thereby losing the single cell resolution of DNA methylation states (Fig. 1). Biological samples containing a mixture of cell types can affect the detection of a “true” epigenetic variation. An example of this is the use of blood samples to identify epigenetic variants associated with aging. Recent findings show that cell-type proportions in blood change with age and can hence confound epigenetic variant analysis and lead to false positives (Jaffe and Irizarry 2014). The power to detect cell-specific associations from mixed cell-type samples depends on the cell-type proportions, effect size in specific cells, and also the sample size. To account for cellular heterogeneity in epigenetic association studies, a commonly used approach is to use cell-type reference sets as fixed effects in regression models, but such reference sets need not be available for all tissue types, and hence methods like surrogate variable analysis (SVA) are typically used. In summary, a reliable biological interpretation of the variation in DNA methylation in any sample has to equate to different proportions of subcellular populations, each possessing their unique trinary methylation states, such that the summation of these subpopulations reflects the overall DNA methylation state measured. Moreover, comparing homogenous cell types or accounting for cellular heterogeneity reduces noise, thereby increasing the statistical power and confidence of uncovering a true biological phenomenon.

DNA Methylation Patterns

The functional placement of this epigenetic mark in the genome (Fig. 1) is essential for maintaining the optimal transcriptional landscape and also the response to developmental and environmental cues. Low DNA methylation levels in promoters and high DNA methylation in gene bodies typically associate with increased gene expression (Ball et al. 2009), whereas high DNA methylation levels in promoters, intergenic regions, and repeat elements generally associate with gene repression (Brenet et al. 2011). However, this directionality is not exclusive (Mehta et al. 2013; Teh et al. 2014), as majority of the autosomal CpGs that are dynamically regulated during development are located distal to the TSS (Ziller et al. 2013). Recently, disease-associated epigenetic variation has also been found to be enriched in the super enhancers, the regulatory regions critical for cell identity and function (Heyn et al. 2016).

The human genome also comprises of CG-rich regions (Fig. 1), termed CpG islands (CGIs), which are typically ≥ 200 bp long, with $\geq 50\%$ CG content, and an observed/expected CpG frequency of ≥ 0.6 (Gardiner-Garden and Frommer 1987). These CGIs (Gardiner-Garden and Frommer 1987) possess largely conserved methylation patterns, which are mostly unmethylated at all stages of development, in all tissue types (Antequera and Bird 1993), and less susceptible to environment. Seventy percent of gene promoters contain a CGI (Saxonov et al. 2006) and exhibit tissue-specific methylation patterns (Mohn et al. 2008). Flanking traditional CGIs, and up to 2 kb away, are CpG shores (Irizarry et al. 2009). The regulatory effects of CpG shores on gene expression resemble promoters, as hypermethylation of such regions downregulates transcription (Irizarry et al. 2009). CpG shores also exhibit

tissue-specific methylation patterns, but unlike CGIs, they are highly susceptible to environment and disease and hence constitute the most variable methylation sites of the genome (Mehta et al. 2013). There have also been various attempts at defining other “biologically interesting genomic regions” including CpG valleys (Xie et al. 2013), canyons (Jeong et al. 2014), and ravines (Edgar et al. 2014). However, the latter CpG classifications are much more obscure and underrepresented in literature as they encompass biological observations such as consistently hypomethylated genomic loci, which may potentially be cell or tissue specific, hence making practical applications of these definitions challenging and less generic.

Genetics of Methylome Diversity

Differences in genomic composition and structure can influence both DNA methylation and phenotypic outcomes (Fig. 2). Such DNA modifications are not just limited to single nucleotide polymorphisms (SNPs) and indels but can include large-scale structural variants such as copy number variations (CNVs), chromosomal translocations, and aneuploidy. In this section, we discuss various mechanisms by which genetic variants can lead to epigenetic variation.

When discussing genetic influence, individual genetic variants may contribute independently toward an underlying phenomenon. The summation of these contributions would be considered the combined additive genetic influence. However, some genetic variants may elicit influence over other variants, such that the combined influence is less/more than the summation of the two; such nonadditive genetic interactions are known as epistatic effects.

SNP-meQTLs

Irrespective of the additive or epistatic effects, genetic influences can manifest its effect through indirect influences on a CpG, or via direct disruption of a CpG site. SNPs that are associated with quantitative changes in methylation level are known as meQTLs. Though meQTLs significantly associate with the methylation status of their respective CpGs, a single meQTL can significantly associate with many CpGs, both in *cis* and in *trans*. *Cis*-meQTLs typically associate with neighboring CpGs, usually along a linkage disequilibrium (LD) block, while *trans*-meQTLs mostly associate with very few or single CpGs on a different chromosome (Gaunt et al. 2016). Only a small proportion of all surveyed CpGs are known to be associated with meQTL (0.12–15%), but these meQTLs have a considerable impact on the methylation level (10–97%). Over 90% of *cis*-meQTLs affect their target sites within several kilobases and are underrepresented in CGIs (Taudt et al. 2016).

With respect to biological plausibility, it is important to note that differences exist between studies investigating tissue samples as compared to those using cell lines. In tissues, *cis*-meQTLs appear to position very near to the associated CpG (45 bp in the brain) and the *cis*-eQTLs lie farther away (68 kb in the brain) (Gibbs et al. 2010); Note

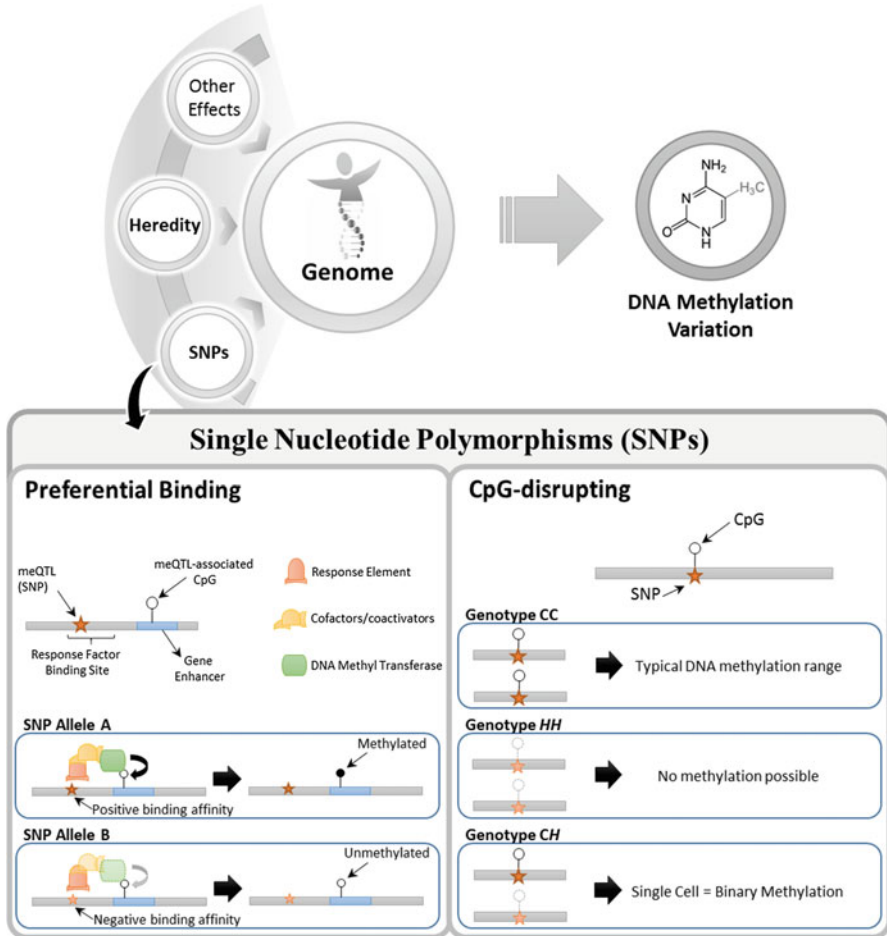


Fig. 2 Genetic factors contributing to the variability in methylome. Underlying DNA sequences can influence the extent of DNA methylation variation. While heredity is usually taken as a surrogate for combined genetic influence, individual components such as single nucleotide polymorphisms (SNPs) also contribute to this variation. In the case of SNPs, the *bottom left panel* illustrates how variation can occur due to preferential binding of DNA-binding proteins, as well as the presence/absence in select cell/tissue populations, leading to cell-/tissue-specific DNA methylation states. The *bottom right panel* illustrates how a SNP located in a CpG site can disrupt DNA methylation

that these distances represent the peak enrichment, or mode. Most other papers report the mean or median. Within a biological system, where binding interactions are sensitive to spatial proximity (such as in this situation), the peak enrichment would provide the most meaningful biological information. Additionally, most meQTLs do not overlap eQTLs, as the SNPs influencing local CpG methylation levels are different from those that influence local gene expression. However in cell line studies, meQTLs appear to be enriched in eQTLs (Bell et al. 2011). Despite various cell-/tissue-type

corrections used in studies examining tissue samples, the inability to fully tease out specific cellular subpopulations could lead to DNA methylation variation being attributed to both inter-cell-type differences as well as intra-cell subpopulation differences, as opposed to the relatively more homogenous nature of cell lines.

SNP-ASMs

Sequence-dependent allele-specific methylation (S-ASM) is distinct from the parent-of-origin-dependent ASMs (P-ASM), as the latter represents the imprinted regions of the genome that are preferentially expressed from either the paternal or maternal allele, without the dependence on genotype. S-ASM can be CpG disrupting or can also involve a SNP affecting the CpG function in *cis* (He et al. 2015). CpG-disrupting SNPs represent a unique form of ASM, where the CpG-disrupting allelic variant leads to a complete loss of methylation on one allele, while its corresponding CpG polymorph on the other allele retains normal methylation status (Kerkel et al. 2008). Though both imprinted and most sequence-based ASMs not involving CpG-disrupting SNPs reflect 50% methylation status at CpGs, the DNA methylation profile of ASMs with CpG-disrupting SNPs is a complicated scenario. The homozygous genotypes where the CpG site is not disrupted reflect a normal DNA methylation status. However, in genotypes possessing the CpG-disrupting SNP on either one or both the alleles, DNA methylation levels would depend on the methylation analysis platform. DNA methylation is typically measured after bisulfite conversion followed by polymerase chain reaction (PCR) amplification. An unmethylated cytosine is converted to uracil after bisulfite conversion and then to thymine during PCR, whereas a methylated cytosine does not undergo a change in the bisulfite step and is retained as cytosine during PCR. Subsequent comparisons between the ratios of thymines and cytosines provide the level of DNA methylation at the genomic loci. All other nucleotides (adenine, thymine, and guanine) are not affected by the bisulfite conversion step. As such, depending on the SNP variant as well as surveying technology, a SNP-disrupted CpG site (which is no longer a CpG) could be misread as methylated, unmethylated, missing, or mismatched. Notably, differences in detection methodologies, and quality control procedures, make such systematic errors potentially more prevalent in microarray-based than sequencing-based technologies.

Differentiation between the methylation states of heterogeneous alleles, thereby identifying ASM loci, can be achieved through the investigation of heterozygous SNPs in neighboring CpGs, usually along the same linkage disequilibrium block (Kerkel et al. 2008). However, such investigations typically result in very small representations of the entire genome. Nevertheless, it is believed that such regions would appear to possess intermediate DNA methylation levels (30–70%) within a biological sample, encompassing up to 18% of intermediate DNA methylation levels observed in the human genome (Elliott et al. 2015). However, since the systematic errors caused by CpG disrupting ASMs are difficult to identify, most DNA methylation studies prefer to discard DNA methylation measurements at such CpG dinucleotide positions.

ASM should not be confused with hemi-methylation, where different strands of the same DNA molecule have different methylation status. In hemi-methylated DNA (rare in non-oncogenic cells), although the average DNA methylation would be similar to ASM-CpGs (approximately 50%), the DNA methylation distribution of a hemi-methylated loci would be a single peak at 50%, whereas the corresponding distribution of an ASM-CpG would be multimodal and, depending on the type of ASM, not necessarily centered at 0%, 50%, and 100%.

Like meQTLs, ASM regions can also be tissue specific (Schultz et al. 2015). *GNAS* is maternally expressed in the anterior pituitary (Hayward et al. 2001), thyroid (Germain-Lee et al. 2002), and ovaries (Mantovani et al. 2002), but biallelically expressed in the bone and visceral adipose (Mantovani et al. 2004); *GRB10* is paternally expressed in the brain but maternally expressed in most other peripheral tissues (Arnaud et al. 2003). Notably, while this tissue specificity may have arisen due to the preferential binding of cofactors to certain genetic sequences, or the presence/absence of select cofactors/coactivators leading to ASM, allele-specific expression (following ASM) has yet to be conclusively identified due to technological limitations (Ladd-Acosta & Fallin 2015).

Environmental Influences and the Variability in the DNA Methylome

As molecular epidemiologists attempted to quantify the contribution of mean genome-wide genetic influences on CpG site-specific variation in DNA methylation, it became clear that most tissues reflected no more than 20% heritability (Gordon et al. 2012), and other influences such as the environment played a much larger role in the determination of methylation diversity. Similar to genetic interactions, the epigenome can mediate and also modify the phenotypic response toward an external stimulus. Such stimuli can manifest via a cornucopia of factors such as diet, toxin exposure, physical activity, parenting, and even ambient temperature and altitude. Here, we discuss a few examples of the biological mechanisms of how environmental influences have been shown to be associated with both differential DNA methylation and phenotypic changes.

Alternate Gene Splicing

An example of a strong deterministic environmental outcome linked to DNA methylation differences involves the phenotypic dimorphism of honey bees (*Apis mellifera*), where queen bees are much larger, fertile, and live longer than female worker counterparts, despite possessing identical DNA (Fig. 3). Female bee larvae within the first three instars can develop either into sterile workers or queens (Shuel and Dixon 1959), and this phenotypic dimorphism is determined by who is fed the “royal jelly,” that develops them into queens (Haydak 1943). DNA methylation

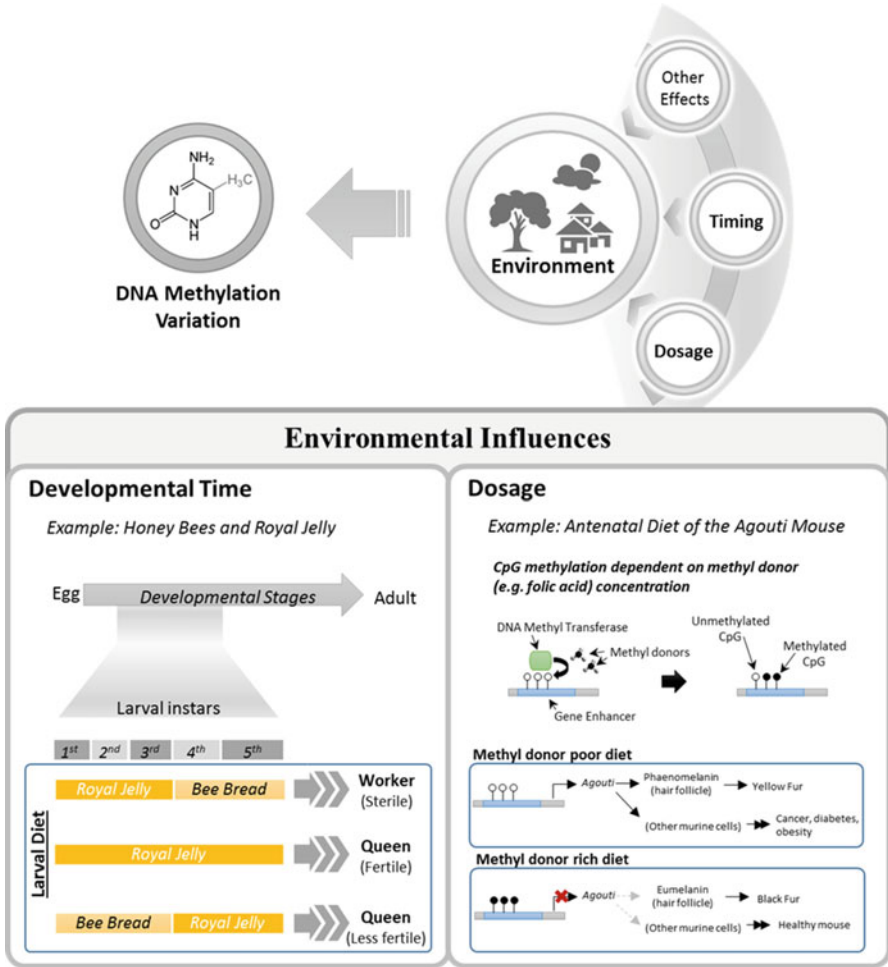


Fig. 3 Scenarios in which environmental influences can modify the methylome. Different environmental exposures can also lead to DNA methylation variation. The extent of such influences is associated with both the timing of exposure in relation to the developmental stage of the organism. For example, the *bottom left panel* illustrates how honey bee (*Apis mellifera*) larvae fed a diet of royal jelly develop into queen bees, whereas those fed a diet of honey and pollen (also known as bee bread) past the third instar develop into sterile worker bees. Dosage of exposure is also a determinant of DNA methylation variation. For example, the *bottom right panel* illustrates how the dietary methyl donor content of pregnant Agouti mice determines the extent of methylation within a gene enhancer region. Methyl donor-poor diet results in an unmethylated state, which manifests as yellow-furred mice offspring. Methyl donor-rich diet results in a methylated state, which manifests as black-furred mice offspring. Intermediate levels result in a heterogeneous methylation state, which manifests as varying degrees of brown-furred mice offspring

differences between the queen and workers potentially support splicing mechanisms, altering chromatin to modulate differential gene expression (Wittkopp 2007).

Metastable Epialleles

Dietary influence can also lead to differential phenotypic outcomes, and a classic example of this is the dietary folate levels of antenatal Agouti mice that influences the degree of methylation of a retrotransposon in the promoter of the Agouti gene of developing mice embryos, leading to differential birth phenotypes (Waterland and Jirtle 2003) (Fig. 3). Unfortunately for these mice, ectopic *Agouti* expression leads to increased predispositions toward cancer, diabetes, and obesity (Morgan et al. 1999). The *Agouti* retrotransposon is interesting as specific genomic loci within it have been identified as metastable epialleles A^{vy} (Duhl et al. 1994), $Axin^{Fu}$ (Vasicek et al. 1997), and $Cabp^{IAP}$ (Druker et al. 2004). Metastable epialleles are identical alleles exhibiting differential expression due to stably maintained epigenetic modifications, established typically during embryonic development (Rakyan et al. 2002). These metastable epialleles are DNA methylation dependent, as increasing methylation leads to increased suppression of promoter activity and, consequently, darker colored mice. This DNA methylation appears to be stochastic but highly correlated with antenatal dietary folate consumption; hence, more dietary folate associates with greater methylation at these epiallelic sites (Waterland et al. 2006). In humans, a study on naturally occurring seasonal differences in the diet of rural Gambian women showed that the seasonal variations in methyl donor nutrient intake of mothers at conception influence the methylation status of the metastable epialleles in offspring (Dominguez-Salas et al. 2014).

Xenobiotics

A notable observation in twin studies is the variability of genetic influence on DNA methylation variation over time (estimated though heritability) (Wong et al. 2010; Fraga et al. 2005). Concomitantly, this implies differential influences from either environmental or stochastic sources across a person's life course. In the previous examples of environmental influence on DNA methylation variation, we have seen that the environment can drastically influence phenotypic outcomes especially during critical "windows" of development during the life course of the organism. In those examples, environmental exposure outside the critical "window" of development would have little to no long-term impact on the DNA methylome. However, there are environmental influences which can influence DNA methylation variation throughout the life course of the organism. In this section, we describe such environmental influences from a xenobiotic perspective. Xenobiotics are essentially foreign chemical substances not expected to be found in the body and may influence DNA methylation variation through a variety of mechanisms (Fig. 4). Here, we

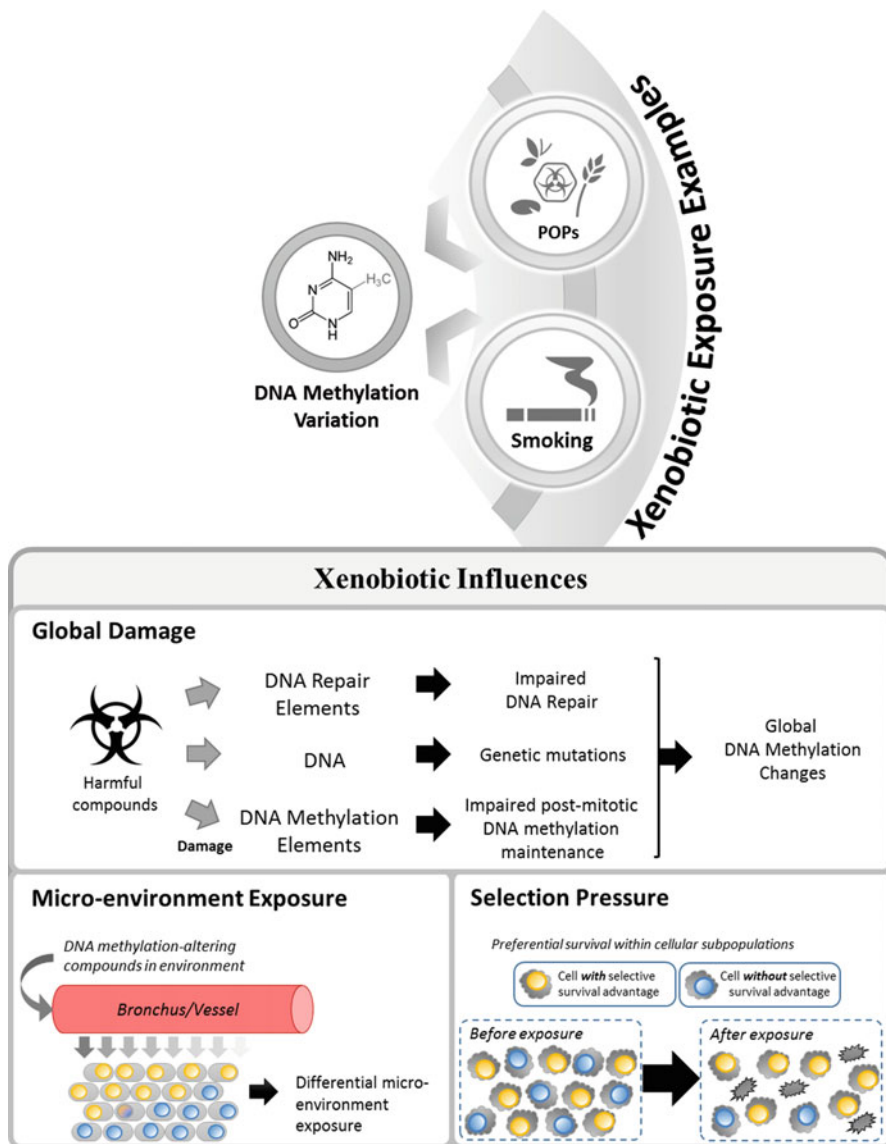


Fig. 4 Exposure to xenobiotics and differential DNA methylation. Exposure to cigarette smoke as well as persistent organic pollutants (POPs) are used to illustrate various potential mechanisms which could lead to DNA methylation variation in a biological system. The *top panel* labelled “Global damage” illustrates how xenobiotics may potentially damage various components of biological systems including DNA repair mechanisms, DNA methylation mechanisms, and the underlying DNA sequence as well. Damaged DNA repair elements result in impaired capacity to repair mutated DNA, which may lead to altered genetic influences on DNA methylation variation such as preferential loss of CpG sites through 5-methylcytosine to thymine mutations. Furthermore, damaged DNA methylation mechanisms could lead to impaired postmitotic DNA methylation

highlight the biological inferences from two examples in particular: cigarette smoke exposure and persistent organic pollutants (POPs).

Cigarette smoke exposure depicts how xenobiotic exposure from a singular environmental exposure source can exhibit adverse effects on health (Sherman 1991). Smoking is a powerful modifier of DNA methylation (Breitling et al. 2011; Lee and Pausova 2013) and can influence the epigenome in multiple tissues. Maternal smoking has been associated with altered DNA methylation in not only the placenta but also the offspring tissues, and examples of gene loci affected include xenobiotic-detoxifying genes, *AHRR* and *CYP1A1* (Joubert et al. 2012).

Persistent organic pollutants (POPs) represent another class of xenobiotics. POPs are resistant to natural degradation and can build up within a biological system with time, a process known as bioaccumulation. POPs preferentially bioaccumulate in adipose tissues due to their mostly lipophilic nature, which also drives DNA methylation differences (van den Dungen et al. 2017).

Besides the global, random influences of xenobiotics, certain xenobiotics can induce DNA methylation variation through more targeted pathways. For example, cigarette smoke exposure is closely tied to hypoxia. Hypoxia can influence variation in DNA methylation by alteration of the methyl donors available in the body. It triggers the upregulation of hypoxia-inducible factor 1- α (HIF1A), which in turn drives the upregulation of methionine adenosyltransferase (MAT). MAT is an enzyme responsible for the production of S-adenosylmethionine (SAM), a crucial methyl donor necessary for the formation of 5-methylcytosines (Finkelstein 2000). Hypoxia also leads to a reduction of TET enzymatic activity – an important enzyme essential for active demethylation of 5mCs (Thienpont et al. 2016). Both mechanisms drive global hypermethylation of actively dividing cells. In addition, xenobiotics may induce DNA damage, introducing novel mutations (Huang et al. 2012) which may potentially influence DNA methylation variation via genetic effects. For example, polycyclic aromatic hydrocarbons, including POPs, increase the mutability of 5mC through the formation of guanine adducts, resulting in 5mC to T transformations post-DNA replication (Pfeifer 2006). Differential xenobiotic microenvironmental exposure may also lead to preferential cell death in specific cell populations, resulting in differential cellular proportions postexposure (Huang et al. 2012), which may also lead to altered sample DNA methylation profiles.

Apart from influencing molecular mechanisms directly tied to DNA methylation, DNA methylation variation can also be achieved through modulation of DNA-



Fig. 4 (continued) maintenance. The *bottom left panel* illustrates how differential microenvironmental exposure within tissues can lead to differential environmental dosages. DNA methylation responses toward such differential exposure may be dependent on such differential chemical dosage. Finally, the *bottom right panel* illustrates how cells may contain elements conferring selective survival advantages toward toxic chemical exposures such as increased antioxidant production either stochastically or due to different cellular function. Such differential survival advantages may lead to overall DNA methylation differences based on a change in cellular proportions

binding factors, which in turn alter the DNA methylation patterns. Moreover, xenobiotic exposure may elicit differential cellular responses, potentially resulting in increased intraindividual DNA methylation variation. For example, nicotine decreases DNMT1 expression depending on the number of specific nicotinic receptors in murine neural cells (Satta et al. 2008).

Combined Influences of Genotype and Environment on the Variability in the Methylomes

Until now we have discussed the independent influences of genotype and environment (G+E) on the methylation diversity in the population. However, the true biological influence on epigenetic variation is likely to be contributed by a complex interplay of both the factors. Therefore, numerous epidemiology studies have begun to examine their joint influence on epigenetic variation (Fig. 5). There could be two scenarios: (1) Genetic variants could direct the DNA methylation status of specific genomic loci, which in turn determine the response toward a particular environment, or (2) an environmental exposure could direct the DNA methylation status of specific genomic loci, which in turn modifies the genomic transcription of particular alleles, changing the observed phenotype. In both these scenarios, a proportion of DNA methylation variation can be attributed to the interaction between genetic and environmental influences (G x E).

Unfortunately, modelling G x E in the context of its influence on DNA methylation alone is challenging due to a myriad of reasons similar to modelling G and E influences independently. This is not only limited to nonadditive allelic influences, nonlinear biological effects (Aliev et al. 2014), heteroscedasticity (White 1980), and confounding interference from other covariates (Keller 2014) but also shared heritability between covariates and genotype of interest (Dudbridge and Fletcher 2014). Hence, it is important to address the underlying hypothesis and apply the appropriate statistical model for an accurate biological interpretation of data (see Duncan and Keller 2011 for critique). Furthermore, interrogating G x E requires larger sample sizes.

Nevertheless, successful G x E association studies can have immensely beneficial clinical outcomes. Most obviously, the identification of significant G x E effects highlights the genetic subpopulations most vulnerable to the environmental exposure interrogated. One of the few examples of G x E influences on DNA methylation which ultimately manifest in a quantifiable phenotype is the exposure to childhood abuse and trauma that significantly associates with hypomethylation only in select CpGs in subjects carrying the rs1360780 A/T risk allele (Klengel et al. 2013). This SNP lies in the functional enhancer of the FK506 binding protein 5 (FKBP5) gene and was previously shown to moderate the risk for post-traumatic stress disorder (PTSD) (Binder et al. 2004), enhancing distal glucocorticoid response, thereby upregulating FKBP5 expression (Klengel et al. 2013). Early trauma-induced hypomethylation of select CpGs (in intron 7 of FKBP5) further amplifies these allele-specific events, leading to differential G x E-dependent glucocorticoid receptor

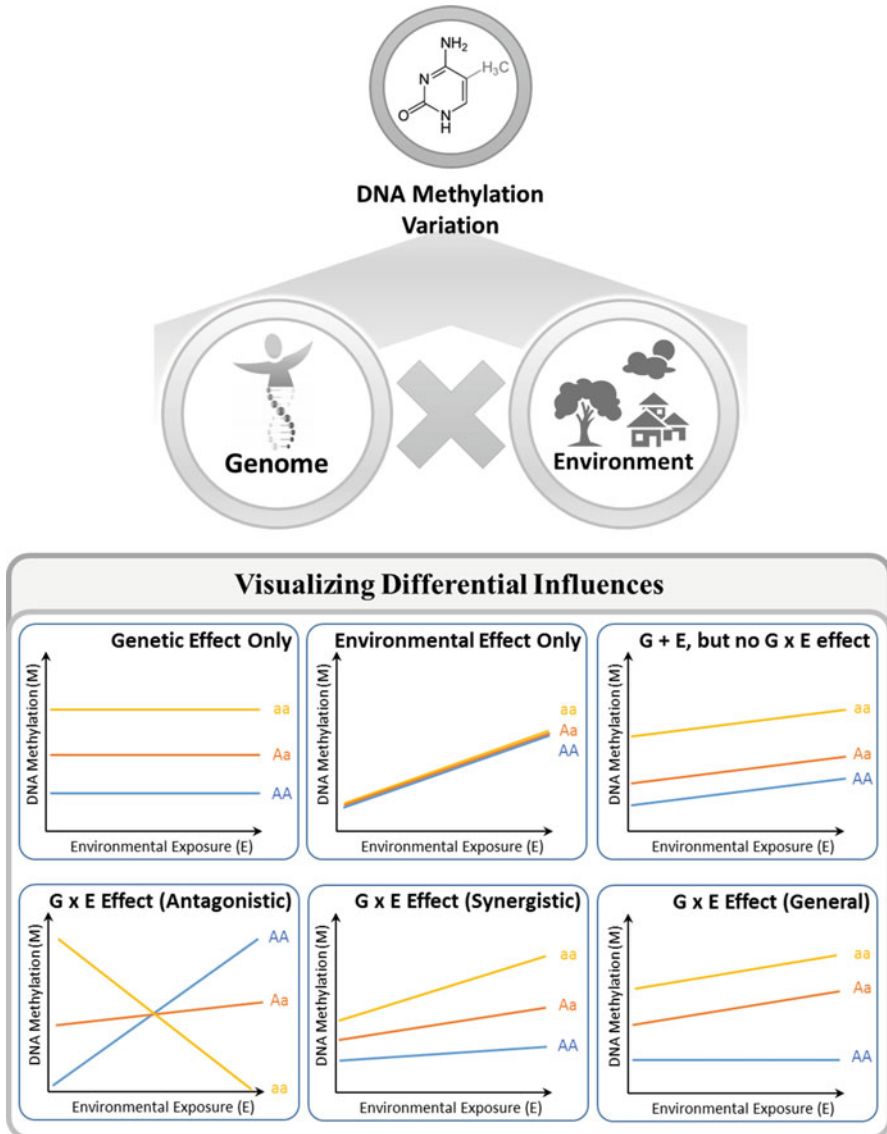


Fig. 5 Differential influence of genetic and environmental factors on DNA methylation variation. In a simplistic model depicting the relationship between genes, environment, and DNA methylation, differences between the trends reflected by individual genotypes may reveal the nature of gene-environment influences on DNA methylation variation. Here, we represent the DNA methylation trends with respect to increasing environmental exposure with lines; different colors represent different genotypic variants, with *blue* representing the homozygous major alleles AA, *orange* representing the heterozygous Aa, and *yellow* representing the homozygous minor alleles aa. For simplicity, we assume that all cofounders are accounted for, and the DNA methylation values adjusted accordingly within the model. Within such a paradigm, a purely genetic influence (independent of environmental exposure) on DNA methylation variation would reflect three parallel

sensitivity (Klengel et al. 2013). In another related example, Yousefi et al. (2013) uncovered 2 CpGs (cg03050981 and cg11807188) whose DNA methylation levels were significantly influenced by the rs12059300 genotype, as well as in utero smoking exposure. Their plotting of differential allelic variant risk, DNA methylation, and serum leptin concentration also nicely demonstrates G x M effects on serum leptin concentrations.

Conclusion

While DNA methylation is one of the most routinely interrogated epigenetic phenomena in clinical studies, several concepts usually lie hidden behind the one-dimensional assessment of this mark. This chapter illustrates how different factors can affect population epigenome variation analysis and outcomes, and hence their consideration in refining future EWAS studies. Due to the multifactorial and polygenic nature of complex human diseases, complementing epigenetic findings with genetics will be more useful in identifying the molecular causes more reliably. A comprehensive mapping of active versus passive DNA methylation patterns in different tissues will be the key to determine the causative function of methylation in altering gene expression versus being a biomarker of transcriptional status. This analysis will be critical for distinguishing diagnostic versus intervention approaches as a follow-up of these findings. Likewise, an intricate assessment of the environmental cues and their cross talk with the genotype and epigenome will provide a better understanding of causality in suboptimal health outcomes and disease onset.



Fig. 5 (continued) trend lines with a zero gradient as illustrated in the *top left panel* labelled “Genetic Effect Only.” Similarly, a purely environmental influence (independent of genetic influence) on DNA methylation variation would reflect a trend line common between the genetic variants, as illustrated in the *top middle panel* labelled “Environmental Effect Only.” If the genetic and environmental influence, when added together, can explain the total variance in DNA methylation, this is known as a G + E effect. In other words, DNA methylation variation can be explained by the additive influence of both genetic and environmental influences independently. This would be reflected by three *parallel lines* with a nonzero gradient – as illustrated in the *top right panel* labelled “G + E, but no G x E effect.” Here, the distance between two trend lines reflects the genetic effect, while the gradient itself represents the environmental effect. Note that it is also possible to have two lines instead of three, when two genotypic variants share similar influences on DNA methylation variation. The *bottom three panels* illustrate differing scenarios where there is a gene-environment interaction (*GxE*) effect on DNA methylation variation. Notably while many combinations of G x E relationships are possible, this figure only illustrates three for simplicity. If a CpG-disrupting SNP is present (e.g., in the A allele), no methylation would be observed – potentially reflecting a relationship illustrated in the *bottom right panel*. If there was incomplete dominance of antagonistic relationships, a trend reflected by the *bottom left panel* could arise. If the genetic variants “A” and “a” have synergistic relationships in relation to DNA methylation, a trend reflected by the *bottom middle panel* could arise

Mini-Dictionary of Terms

- **Epistasis** – An interaction between genes, where an allele of one gene modifies the phenotype of another.
- **Allele-specific methylation (ASM)** – Presence of DNA methylation on only one of the two alleles. This could be due to genetic effects, imprinting, or random methylation of one allele.
- **Metastable epialleles** – Genetically identical individuals can have variably expressed alleles due to epigenetic modifications established during early development. Metastable alleles are thought to be highly vulnerable to environmental influences.
- **meQTL** – SNPs associated with quantitative changes in methylation level.
- **GxE** – Gene-environment interactions where the two different genotypes respond to environmental change in different ways.

Key Facts of Epigenome-Wide Association Studies (EWAS)

- EWAS is a genome-wide interrogation of epigenetic variation in a population and its association with the phenotype of interest.
- Epigenetic variations can cause disease but can also arise as a consequence of disease.
- EWAS generally involve DNA methylation, as studying this epigenetic mark requires less clinical samples than that required for investigating histone modifications.
- Epigenetic patterns are distinct between cells/tissues, as well as differing pathologies.

Key Facts of Gene-Environment Interactions (GxE)

- GxE is when two different genotypes respond to the variation in the environment in different ways.
- Because of GxE some people are more susceptible or resistant to a certain disease in a particular environment. Likewise, an individual's response to medication or nutrition can vary.
- Gene-environment interactions can modulate the epigenome.

Key Facts of Xenobiotics

- Xenobiotics are chemical compounds that are foreign to the organism, as they are neither produced nor consumed as part of the diet, for example, cigarette smoke and POPs.

- Xenobiotics are studied for their potential effects on the environment and human health. It is important to know if they are biodegradable or will remain in the environment and cause harm.
- POPs can be bioaccumulated and biomagnified. They are lipid soluble and are hence stored in the adipose tissue of animals.
- Exposure to POPs in pregnancy can have lifelong implications on the offspring, including its epigenome.

Summary Points

- DNA methylation is least variable in CGIs and promoters and most variable in CpG shores, open seas, and intergenic regions.
- Reliable quantitation of DNA methylation from clinical samples is confounded by cellular heterogeneity and cell proportions.
- Interindividual differences in DNA methylation can be influenced by underlying genetic variations.
- Genetic variants can influence DNA methylation variation by either affecting the binding of DNA-binding factors or by disrupting the CpG sites.
- meQTLs can influence CpGs in *cis* and *trans* and are underrepresented in CGIs.
- There can also be allele-specific influences of the genotype on the variability in the methylome.
- Early environmental exposures, especially those during periods of rapid development and maturation, have more drastic influences on DNA methylation variability.
- Gene-environment interactions result in differing genotypic susceptibilities to DNA methylation at different levels of environmental exposure.
- Longitudinal study designs, environmental exposures, as well as the genetic information are critical for appropriate causation and mediation analyses in EWAS studies.

References

- Aliev F et al (2014) Testing for measured gene-environment interaction: problems with the use of cross-product terms and a regression model reparameterization solution. *Behav Genet* 44(2):165–181
- Antequera F, Bird A (1993) Number of CpG islands and genes in human and mouse. *Proc Natl Acad Sci U S A* 90(24):11995–11999
- Arnaud P et al (2003) Conserved methylation imprints in the human and mouse GRB10 genes with divergent allelic expression suggests differential reading of the same mark. *Hum Mol Genet* 12(9):1005–1019
- Ball MP et al (2009) Targeted and genome-scale strategies reveal gene-body methylation signatures in human cells. *Nat Biotechnol* 27(4):361–368. Available at: <https://doi.org/10.1038/nbt.1533>
- Bell JT et al (2011) DNA methylation patterns associate with genetic and gene expression variation in HapMap cell lines. *Genome Biol* 12(1):R10
- Binder EB et al (2004) Polymorphisms in FKBP5 are associated with increased recurrence of depressive episodes and rapid response to antidepressant treatment. *Nat Genet* 36(12):1319–1325

- Breitling LP et al (2011) Tobacco-smoking-related differential DNA methylation: 27K discovery and replication. *Am J Hum Genet* 88:450–457
- Brenet F et al (2011) DNA methylation of the first exon is tightly linked to transcriptional silencing. *PLoS One* 6(1):e14524
- Dominguez-Salas P et al (2014) Maternal nutrition at conception modulates DNA methylation of human metastable epialleles. *Nat Commun* 5:3746–3752
- Druker R et al (2004) Complex patterns of transcription at the insertion site of a retrotransposon in the mouse. *Nucleic Acids Res* 32(19):5800–5808
- Dudbridge F, Fletcher O (2014) Gene-environment dependence creates spurious gene-environment interaction. *Am J Hum Genet* 95(3):301–307
- Duhl DM et al (1994) Neomorphic agouti mutations in obese yellow mice. *Nat Genet* 8(1):59–65
- Duncan LE, Keller MC (2011) A critical review of the first 10 years of candidate gene-by-environment interaction research in psychiatry. *Am J Psychiatr* 168(10):1041–1049
- Edgar R et al (2014) Meta-analysis of human methylomes reveals stably methylated sequences surrounding CpG islands associated with high gene expression. *Epigenetics Chromatin* 7(1):28
- Elliott G et al (2015) Intermediate DNA methylation is a conserved signature of genome regulation. *Nat Commun* 6:6363
- Finkelstein JD (2000) Pathways and regulation of homocysteine metabolism in mammals. *Semin Thromb Hemost* 26(3):219–225
- Fraga MF et al (2005) Epigenetic differences arise during the lifetime of monozygotic twins. *Proc Natl Acad Sci U S A* 102(30):10604–10609
- Gardiner-Garden M, Frommer M (1987) CpG Islands in vertebrate genomes. *J Mol Biol* 196(2):261–282
- Gaunt TR et al (2016) Systematic identification of genetic influences on methylation across the human life course. *Genome Biol* 17(1):61
- Germain-Lee EL et al (2002) Paternal imprinting of *Gas* in the human thyroid as the basis of TSH resistance in pseudohypoparathyroidism type 1a. *Biochem Biophys Res Commun* 296(1):67–72
- Gibbs JR et al (2010) Abundant quantitative trait loci exist for DNA methylation and gene expression in human brain. *PLoS Genet* 6(5):29
- Gordon L et al (2012) Neonatal DNA methylation profile in human twins is specified by a complex interplay between intrauterine environmental and genetic factors, subject to tissue-specific influence. *Genome Res* 22(8):1395–1406
- Haydak MH (1943) Larval food and development of castes in the honeybee. *J Econ Ent* 36:778–792
- Hayward BE et al (2001) Imprinting of the *Gsα* gene *GNAS1* in the pathogenesis of acromegaly. *J Clin Investig* 107(6)
- He J et al (2015) Characterization and machine learning prediction of allele-specific DNA methylation. *Genomics* 106(6):331–339
- Heyn H et al (2016) Epigenomic analysis detects aberrant super-enhancer DNA methylation in human cancer. *Genome Biol* 17:11–26
- Huang J et al (2012) Telomere shortening and DNA damage of embryonic stem cells induced by cigarette smoke. *Reprod Toxicol* 35:89–95
- Irizarry RA et al (2009) The human colon cancer methylome shows similar hypo- and hypermethylation at conserved tissue-specific CpG island shores. *Nat Genet* 41(2):178–186
- Jaffe AE, Irizarry RA (2014) Accounting for cellular heterogeneity is critical in epigenome-wide association studies. *Genome Biol* 15(2):R31
- Jeong M et al (2014) Large conserved domains of low DNA methylation maintained by *Dnmt3a*. *Nat Genet* 46(1):17–23
- Joubert BR et al (2012) 450K epigenome-wide scan identifies differential DNA methylation in newborns related to maternal smoking during pregnancy. *Environ Health Perspect* 120:1425–1431
- Keller MC (2014) Gene × environment interaction studies have not properly controlled for potential confounders: the problem and the (simple) solution. *Biol Psychiatry* 75(1):18–24
- Kerkel K et al (2008) Genomic surveys by methylation-sensitive SNP analysis identify sequence-dependent allele-specific DNA methylation. *Nat Genet* 40(7):904–908
- Klengel T et al (2013) Allele-specific FKBP5 DNA demethylation mediates gene-childhood trauma interactions. *Nat Neurosci* 16(1):33–41

- Ladd-Acosta C, Fallin MD (2015) The role of epigenetics in genetic and environmental epidemiology. *Epigenomics* 8:epi.15.102
- Lee KWK, Pausova Z (2013) Cigarette smoking and DNA methylation. *Front Genet* 4(JUL):132
- Mantovani G et al (2002) The Gs?? Gene: predominant maternal origin of transcription in human thyroid gland and gonads. *J Clin Endocrinol Metab* 87(10):4736–4740
- Mantovani G et al (2004) Biallelic expression of the Galpha gene in human bone and adipose tissue. *J Clin Endocrinol Metab* 89(12):6316–6319
- Mehta D et al (2013) Childhood maltreatment is associated with distinct genomic and epigenetic profiles in posttraumatic stress disorder. *Proc Natl Acad Sci* 110(20):8302–8307
- Mohn F et al (2008) Lineage-specific Polycomb targets and De Novo DNA methylation define restriction and potential of neuronal progenitors. *Mol Cell* 30(6):755–766
- Morgan HD et al (1999) Epigenetic inheritance at the agouti locus in the mouse. *Nat Genet* 23(3):314–318
- Pfeifer GP (2006) Mutagenesis at methylated CpG sequences. *Curr Top Microbiol Immunol* 301:259–281
- Rakyan VK et al (2002) Metastable epialleles in mammals. *Trends Genet* 18(7):348–351
- Satta R et al (2008) Nicotine decreases DNA methyltransferase 1 expression and glutamic acid decarboxylase 67 promoter methylation in GABAergic interneurons. *Proc Natl Acad Sci* 105(42):16356–16361
- 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
- Schultz MD et al (2015) Human body epigenome maps reveal noncanonical DNA methylation variation. *Nature* 523(7559):212–216
- Sherman CB (1991) Health effects of cigarette smoking. *Clin Chest Med* 12(4):643–658
- Shuel RW, Dixon S (1959) Studies in the mode of action of royal jelly in honeybee development. *J Can Zool* 37(4):803–813
- Taudt A et al (2016) Genetic sources of population epigenomic variation. *Nature Rev Genet* 17(6):319–332
- Teh AL et al (2014) The effect of genotype and in utero environment on interindividual variation in neonate DNA methylomes. *Genome Res* 24(7):1064–1074
- Thienpont B et al (2016) Tumour hypoxia causes DNA hypermethylation by reducing TET activity. *Nature* 537(7618):63–68
- van de Dungen MW et al (2017) Persistent organic pollutants alter DNA methylation during human adipocyte differentiation. *Toxicol In Vitro* 40:79–87
- Vasicek TJ et al (1997) Two dominant mutations in the mouse fused gene are the result of transposon insertions. *Genetics* 147(2):777–786
- Waterland RA, Jirtle RL (2003) Transposable elements: targets for early nutritional effects on epigenetic gene regulation. *Mol Cell Biol* 23(15):5293–5300
- Waterland RA et al (2006) Maternal methyl supplements increase offspring DNA methylation at Axin fused. *Genesis* 44(9):401–406
- White H (1980) A heteroskedasticity-consistent covariance-matrix estimator and a direct test for heteroskedasticity. *Econometrica* 48(4):817–838
- Wittkopp PJ (2007) Variable gene expression in eukaryotes: a network perspective. *J Exp Biol* 210(Pt 9):1567–1575
- Wong CCY et al (2010) A longitudinal study of epigenetic variation in twins. *Epigenetics* 5(6):516–526
- Xie W et al (2013) Epigenomic analysis of multilineage differentiation of human embryonic stem cells. *Cell* 153(5):1134–1148
- Yousefi M et al (2013) The methylation of the LEPR/LEPROT genotype at the promoter and body regions influence concentrations of leptin in girls and BMI at age 18 years if their mother smoked during pregnancy. *Int J Mol Epidemiol Genet* 4(2):86–100
- Ziller MJ et al (2013) Charting a dynamic DNA methylation landscape of the human genome. *Nature* 500(7463):477–481