

Chapter 5

Epigenomics

Not all genes are expressed in all tissues at all times. While many molecular mechanisms regulating gene expression (in space and over time) are coded in the DNA sequence (e.g. enhancers, repressors, transcription factors), there is a number of so-called epigenetic mechanisms that can regulate gene expression by other means.

In this chapter, we will first review the basics of epigenetics and then describe the two most common epigenetic mechanisms, DNA methylation and histone modification. We will conclude by touching upon a few issues relevant for the integration of genomic and epigenomic information in population-based studies.

5.1 Epigenetics: Heritable, Stochastic and Environment-Induced

In the broad sense, there are three general sources of epigenetic modifications and their variations across cells, tissues and individuals: (1) *heritable* modifications, which can be inherited either from cell to daughter cell (i.e. within the life of an organism) or from a parent to a child (i.e. across generations); (2) modifications that arise from *stochastic instability* in the transfer of epigenetic markers during cell divisions; and (3) epigenetic modifications induced by the *environment*.

5.1.1 Heritable Modifications

Imprinting and X-inactivation are the two most common examples of epigenetic modifications that are inherited from cell to daughter cell (but not from parent to child). In the case of X-inactivation (in females), either the maternal or paternal X chromosome is inactivated in the progenitor cell of a particular lineage (e.g. neuron, oligodendrocyte, hepatocyte). Although the initial “choice” as to which of the two X chromosomes to inactivate is random, all cells derived (and re-derived)

subsequently from the original progenitor cell inactivate the same (maternal or paternal) X chromosome. A given X-inactivation (e.g. of the X chromosome inherited from the mother) is inherited through subsequent cell divisions (mitosis) during the life of the individual, but it is not passed on to the offspring; thus, the offspring can inherit an X chromosome that has been either active or inactive in his/her mother. Note that this mechanism creates a mosaic of cells containing either an active or inactive maternal (or paternal) X chromosome.

Similar epigenetic mechanisms, called gene imprinting, may also regulate expression of genes located on autosomal chromosomes. Thus, there are more than a hundred “mono-allelic” genes in which the allele inherited from one parent is silenced (imprinted) while the other one remains active (Henckle and Arnaud 2010). Furthermore, it appears that gene imprinting can operate in an age-related fashion: for example, maternal alleles of certain genes are expressed in the embryonic brain (while the paternal alleles of the same genes are silenced/imprinted) and vice versa in the adult brain (Gregg et al. 2010).

Trans-generational (from parent to child) transmission of epigenetic modifications has been a surprising discovery. For most of the past 150 years, Darwin’s theory of natural selection was assumed to be the only (that is to say, correct) alternative to the view formulated by Jean-Baptiste Lamarck at the beginning of the nineteenth century: that acquired traits can be passed from parent to child (Lamarck 1809). Simply put, we have taken for granted that Lamarckism was wrong and that acquired traits *cannot* be inherited. This view is now changing, however.

Although it is true that most epigenetic marks underlying the relatively stable transmission of epigenetic information from cell to daughter cell during the life of an individual (see above) are erased during gametogenesis,¹ some of these marks can be passed on to the offspring. Thus, we have trans-generational epigenetic inheritance (reviewed in Daxinger and Whitelaw 2012). An often-quoted example of this kind of inheritance is the agouti viable yellow (A^{vy}) epiallele.² In the genetically identical (i.e. isogenic) A^{vy} mice, the degree of transcription of a retrotransposon³ located upstream of the agouti (A) gene results in the various degree of the expression of this gene and, in turn, a varied amount of the agouti protein being deposited in the mouse fur. The latter affects the fur colour on a continuum, starting with the full-yellow (all-cell) coat, going into a mosaic of yellow and agouti hair and ending with the full-agouti coat. Importantly, the level of transcription of this transposable element varies in proportion to its methylation; hence, an epiallele: less methylation means more transcription and more yellow colour (Morgan et al. 1999; see Fig. 5.1). This effect appears to be related to an

¹ Gametogenesis refers to the production of gametes (eggs and sperm) in gonads, through meiosis (see Sect. 4.1.)

² Epialleles differ in their epigenetic modifications (whereas alleles differ in nucleotides).

³ Retrotransposon is a form of the transposable elements that first copy themselves from DNA to RNA (transcription), then back to DNA (reversed transcription), before inserting themselves into the genome in a new position. In this way, they generate insertions, deletions and translocations.

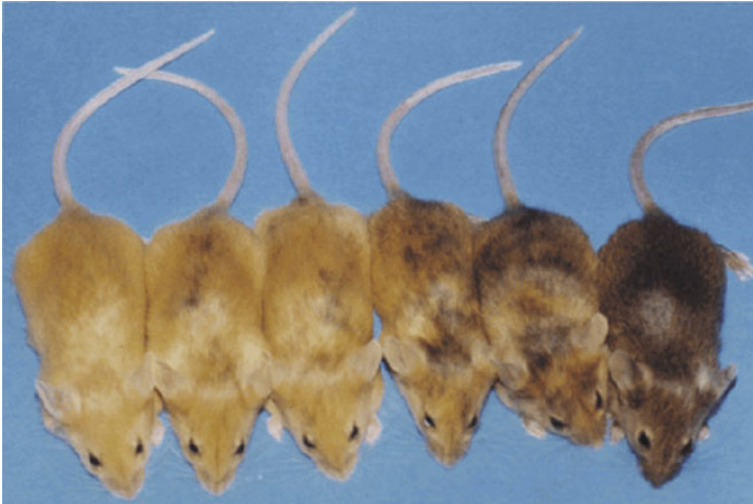


Fig. 5.1 Agouti mice. From Morgan et al. (1999)

incomplete erasure of the epigenetic mark on the retrotransposon in the maternal (but not paternal) germ line.

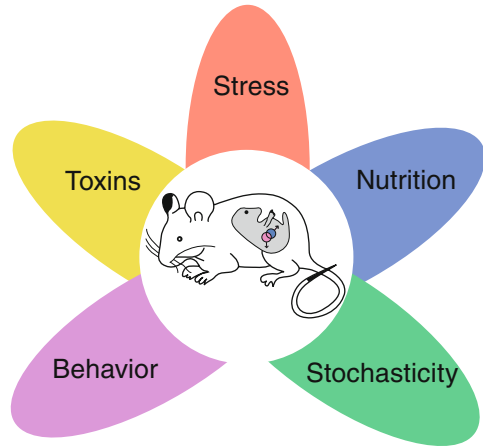
Such trans-generational epigenetic transmission is likely to vary with the number of subsequent generations. For example, certain epigenetic modifications of histones in the chromatin of *Caenorhabditis elegans* parents, known to affect their longevity, are not entirely erased during gametogenesis and, therefore, can also extend the lifespan of the offspring up to the third generation (Greer et al. 2011). In this case, this trans-generational inheritance appears to affect preferentially the epigenetic regulation (expression) of genes involved in metabolic pathways (Greer et al. 2011).

What are the sources of inter-individual variability in the epigenome? In the next two sections, we will discuss stochastic instability in the transfer of epigenetic markers during mitosis and environment-induced epigenetic modifications (see Fig. 5.2).

5.1.2 Stochastic Instability

Epigenetic marks are transmitted from cell to daughter cell during mitosis and, to a much lesser extent, from a parent to offspring through the germ line. This transfer of epigenetic information involves a variety of molecular processes (see Sect. 5.2 for details) and, not surprisingly, it is prone to errors. Thus, it has been estimated that the error rate for replicating epigenetic marks is ~ 1 in 1,000; this is much higher than the estimated error rate during the DNA replication [~ 1 in 1,000,000 bases; Hjelmeland (2011)]. These errors introduce a certain level of randomness

Fig. 5.2 Sources of epigenetic variability. Five environmental influences that affect the developing embryo and its primordial germ cells (represented by the *pink* and *blue dots*). Adapted based on Faulk and Dolinoy (2011)



into the transmission of epigenetic information. Stochastic instability then refers to the probabilistic nature of the processes that results both from a predictable action and the random element. The above example of agouti mice illustrates the range of such stochastic instability: the level of DNA methylation of the “metastable epiallele” can vary by over 80 % across the isogenic mice; a phenomenon stable during the life of an individual but stochastic across different individuals (Dolinoy et al. 2010). What are the main sources of the predictable actions and the most important time window of their influence on the epigenome?

5.1.3 Environment-Induced Epigenetic Modifications

During gametogenesis, epigenetic marks are first (largely) erased and then re-established *de novo*. Thus, *in utero*, the re-establishment of epigenetic marks can be influenced by a variety of *environmental influences* acting on the pregnant mother (F0 generation) and both the somatic and germ lines of the embryo (F1 generation). In the case of the former, such somatic “epimutations” will be transmitted from cell to daughter cell throughout the prenatal and post-natal life of the exposed (F1) offspring; these effects are likely to be present in all tissues. In the case of the latter, the germ-line epimutations have the potential for being transmitted to the subsequent (F2) generation, and beyond (see Fig. 1.3). On the other hand, if the environment acts *post-natally*, its effect is more likely to be seen in specific tissues; thus, tobacco smoke is more likely to affect lung tissue, and diet (together with local bacteria; see the paragraph on “body environment” in Sect. 3.1) is more likely to affect the gut tissue. Figure 5.2 illustrates the most common exposures that have been investigated in the context of environment-induced changes in epigenome; these vary from the effects of specific diets, toxins, stress and behaviour (reviewed in Faulk and Dolinoy 2011). One of the most exiting models of behaviour-induced epigenetic modifications has been introduced by Michael Meaney and his colleague; in a series of

experiments, they showed that high (vs. low) levels of licking and grooming (by the dam) during the first week of pups' lives is associated with lower levels of methylation of the promoter of the glucocorticoid receptor (and, in turn, its higher expression) in the hippocampus, as well as lower response of the hypothalamic-pituitary-adrenal axis to stress (reviewed in Zhang et al. 2013; see Fig. 5.3).

As pointed out above, epigenetic modifications involve a variety of mechanisms, including DNA methylation, modifications of histone proteins and post-transcriptional modifications of non-coding RNA, such as microRNAs, which regulate translation of mRNA into polypeptides. I will now describe two of these mechanisms in some detail.

5.2 DNA Methylation and Histone Modifications

The methylation of cytosine is the most common epigenetic modification of DNA. Given that the diploid human genome contains 6×10^9 nucleotides (A, C, G and T), there are about 150,000,000 cytosines that—in theory—can exist, in either a methylated or unmethylated state. But, in fact, methylation takes place most often

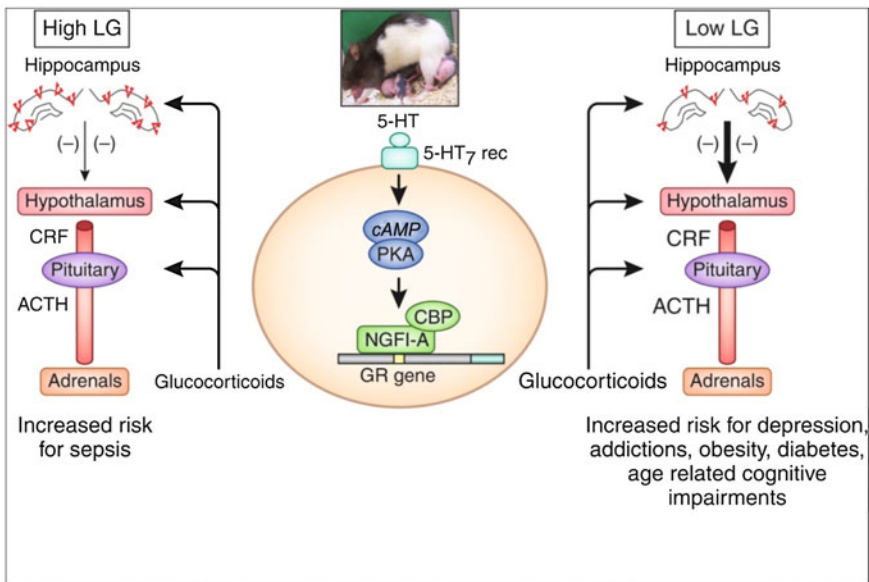


Fig. 5.3 Associations between maternal behaviour (licking and grooming), expression of the glucocorticoid receptor in the hippocampus, regulation of the hypothalamus-pituitary-adrenal axis and psychopathology (*right side*). *LG* licking and grooming, *ACTH* adrenocorticotropicin, *CRF* corticotropin releasing factor, *5-HT* serotonin, *cAMP* cyclic adenosine monophosphate, *PKA* protein kinase A, *NGFI-A* nerve growth factor-inducible factor A, *CBP* CREB-binding protein, *GR* glucocorticoid receptor. From Zhang et al. (2013)

(but not exclusively) when cytosine sits next to guanine—that is, if it is present as CpG⁴ dinucleotides; when CpG dinucleotides cluster together, we talk about CpG islands. Note that about 60 % of human gene promoters are associated with CpG islands, thus providing a powerful means for DNA methylation to influence gene expression (Portela and Esteller 2010). The so-called CpG shores (areas flanking the CpG islands) show higher variation in DNA methylation, even though the CpG density is lower in these regions as compared with the CpG islands (Rakyan et al. 2011). As shown in Fig. 5.4, DNA methylation can occur not only in CpG islands and shores but also in the gene body, and on repetitive sequences, including those associated with transposable elements (see above for the agouti mice). In general,

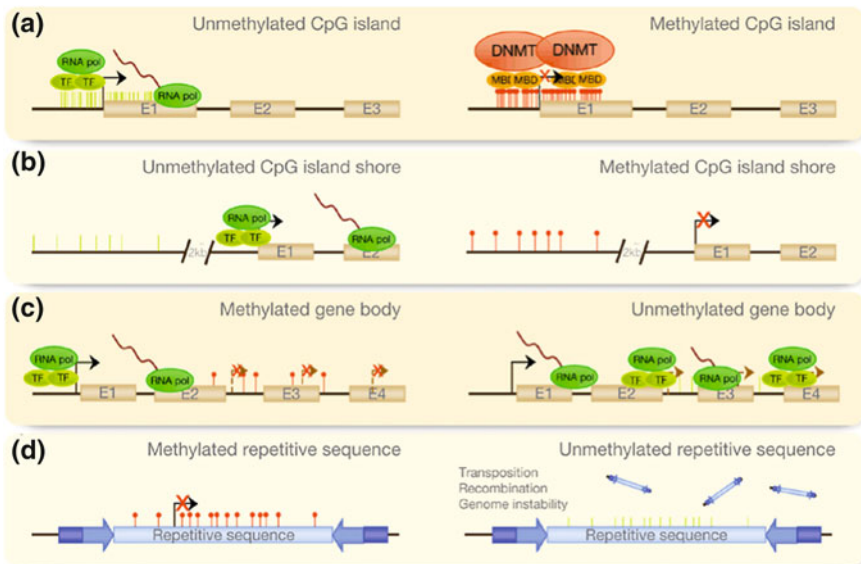


Fig. 5.4 DNA methylation. “DNA methylation can occur in different regions of the genome. The alteration of these patterns leads to disease in the cells. The normal scenario is depicted in the *left column* and alterations of this pattern are shown on the *right*. **a** CpG islands at promoters of genes are normally unmethylated, allowing transcription. Aberrant hypermethylation leads to transcriptional inactivation. **b** The same pattern is observed when studying island shores, which are located up to 2 kb upstream of the CpG island. **c** However, when methylation occurs at the gene body, it facilitates transcription, preventing spurious transcription initiations. In disease, the gene body tends to demethylate, allowing transcription to be initiated at several incorrect sites. **d** Finally, repetitive sequences appear to be hypermethylated, preventing chromosomal instability, translocations and gene disruption through the reactivation of endoparasitic sequences. This pattern is also altered in disease.” E Exon, TF transcription factor, RNA pol RNA polymerase, DNMT DNA methyltransferase, MBD methyl-CpG-binding-domain proteins. From Portela and Esteller (2010)

⁴ Note the nomenclature: CG refers to cytosine located on one DNA strand and guanine on the other (complementary) strand. On the other hand, CpG refers to the two nucleotides being located side by side on the same DNA strand.

methylation of the CpG islands at promoters of genes is associated with a reduced rate of DNA transcription. On the other hand, methylation in the gene body facilitates gene expression by preventing spurious initiations of transcription (Portela and Esteller 2010).

How does DNA methylation regulate gene expression? One of the mechanisms involves recruitment of special proteins⁵ that go on to recruit histone-modifying and chromatin-remodelling complexes to the methylated site. As explained below, the DNA-chromatin complex is a powerful regulator of gene expression by virtue of “opening” and “closing” DNA for transcription (see Sect. 4.1). Another mechanism involves direct inhibition of transcription by preventing the binding of relevant transcription factors to the promoter (e.g. Kuroda et al. 2009).

Recall that DNA is packaged with protein complexes to form chromatin. Heterochromatin contains tightly packed and inactive DNA, whereas euchromatin contains a stretched out and active DNA molecule ready for transcription (Sect. 4.1). Histone proteins are the main actors in the transformation of heterochromatin to euchromatin and back (see Fig. 4.1). Gene expression requires uncoiling of chromatin fibres, a process guided by H1 histone and its bonds with DNA molecules. Once uncoiled, two turns of DNA molecule are wrapped around an octamer of core histones (H2, H3A, H3B and H4) in the individual nucleosomes; the N-terminal tails of the core histones protrude out of a nucleosome (see Fig. 4.1b). Importantly, various chemical processes⁶ can modify amino acids in these histone tails leading, in turn, to the binding of different proteins, affecting local condensation of chromatin and, as such, the level of transcriptional activity at this particular stretch of DNA molecule.⁷

5.3 Bringing Together Genome and Epigenome

Advancements made in mapping DNA variations in the human genome have enabled a GWAS-based search for genetic variants associated with complex traits. In the same manner, we are now in a position to carry out genome-wide scans for epigenetic marks; at present, the relevant technology enables us to do so only for DNA methylations (Rakyan et al. 2011). For example, the Infinium HumanMethylation450 BeadChip Kit allows one to assess methylation state at 485,000 methylation sites across the entire genome, with an average of 17 CpG sites per

⁵ Methyl-CpG binding-domain proteins.

⁶ Acetylation, methylation, phosphorylation, sumoylation and ubiquitylation.

⁷ For example, trimethylation of lysine 27 in H3 histone (H3K27me3) is associated with gene silencing (e.g. Soshnikova and Duboule 2008). As described above, deficiencies in another of the histone-trimethylation complexes (H3K4me3) influence the lifespan of *C. elegans* and that epigenetic modifications (demethylation) of this complex, when located in the vicinity of certain genes, are transmitted across four generations, together with the phenotype. In other words, it influences expression of these genes and longevity (Greer et al. 2011).

gene region and coverage of 96 % of CpG islands. In principle, this information can be interrogated in three ways that reflect the combination of time of exposure to the presumed methylation event (prenatal vs. post-natal) and type of cells (somatic or germ lines) affected by the event. This is in addition to the genetic control of methylation events.

In the case of *prenatal events*, such as exposure to cigarette smoking or a diet low on folic acid, we would expect to find epimutations in both somatic and germ lines of the offspring.

In case of *somatic-cell lines*, epimutations should be present across a variety of tissues. Given the limited availability of tissue in human population-based studies, this assumption can be tested—for example—by comparing the epigenome of DNA extracted from blood cells, buccal (epithelial) cells, bulge cells of the hair follicles or epithelial cells found in urine and faeces (Rakyan et al. 2011). Naturally, the global genome-wide rate of epimutations (defined as a deviation from the reference sample) may serve as a useful phenotype and, perhaps, a more precise proxy of the actual level of exposure to the environment under study. Furthermore, the presence of epimutations in a particular gene region provides an important parameter to be used in evaluating interactions between a particular environment (e.g. cigarette smoking during pregnancy) and specific genetic variations (Text Box 5.1).

Text Box 5.1. Smoking during pregnancy and DNA methylation

In one of our studies, we have observed an interaction between the *BDNF* genotype and prenatal exposure to maternal smoking with regard to the relationship between brain and behaviour; but only non-exposed individuals showed the effect of genotype (Lotfipour et al. 2009). We speculated that this gene might have been “silenced” in the exposed offspring, thus rendering DNA variations in the gene irrelevant in this group. We followed up this idea and showed that the two groups (exposed and non-exposed adolescents) differed in the methylation rate in the CpG island of one of the promoters of the *BDNF* gene (Toledo-Rodriguez et al. 2010). Obviously, the methylation rate was assessed here using DNA extracted from blood cells (and not the brain); as pointed out above, however, the exposure presumably associated with the higher methylation rate occurred, in this case, prenatally and, as such, was likely to affect all tissues to a more similar extent than one would expect for post-natal exposure to the same environment.

In case of *germ lines*, epimutations can be transmitted from parent to child. But note that—in the case of prenatal exposure—this process can be a mix of the direct effect of exposure *and* the trans-generational transmission of epigenetic marks. Thus, a given prenatal exposure can be transmitted from the first generation (F0 = smoking mother) to the grandchild (F2) simply by exposing the germ-line cells of the child (F1) and not erasing completely this epigenetic modification at the time of fertilization of the “exposed” (F1) gametes (i.e. oocytes) by

non-exposed ones (i.e. sperm). The pure exposure-independent trans-generational transmission of the epigenetic information thus takes place only in subsequent generations (F2–F3, F3–F4, etc.). Thus, we need to be cautious when interpreting findings that are based on three generations only (i.e. children of mothers who were exposed prenatally to a given environment).⁸

In summary, we need to consider the following four elements of trans-generational transmission: (1) epimutations of somatic-cell lines of the F1 embryo/foetus; (2) epimutations of germ lines of the F1 embryo; (3) incomplete erasure of epigenetic marks during fertilization, giving rise to the F2 generation; and (4) transmission of epigenetic marks from F2 to F3 (and subsequent) generations.

Finally, in case of post-natal exposures, most of such epigenetic effects are likely to affect only tissues that either are in direct physical contact with the agent (e.g. cigarette smoke) or represent a target tissue in a biochemical cascade initiated by the agent (e.g. stress-induced stimulation of glucocorticoid receptors of the hippocampus). For the human brain, such tissue-specific epigenetic modifications can be assessed only via surgical removals or biopsies (e.g. a repository for gliomas [<http://cainegrator.nci.nih.gov/rembrandt/>]; Riddick and Fine 2011) or in post-mortem tissue (e.g. McGowan et al. 2009).





	Key advantage	Key disadvantage
Case versus control (singletons) 	Many cohorts exist	Cannot easily control for environmental and genetic confounders
Families 	Can study potential inheritance	Few large cohorts of this type exist
Disease-discordant monozygotic twins 	Can control for genetics	Few large cohorts of this type exist
Prospectively sampled, longitudinal 	Can establish causality	Slow and difficult to establish

Fig. 5.5 Design of genome-wide epigenetic studies. From Rakyan et al. (2011)

⁸ The often-quoted examples of a trans-generational effect of food availability on reproduction, health and mortality, such as the Dutch Hunger Winter (Roseboom et al. 2011) and the Overkalix cohort in Northern Sweden (Pembrey et al. 2006), involve only three generations.

Finally, we can ask what are the possible designs of large, genome-wide epigenetic studies. As illustrated in Fig. 5.5, these studies fall into two different classes. In the case of disease-based phenotypes known at the time of the study, one can use either case-control cohorts of unrelated individuals or monozygotic twins discordant for the presence of a given disease. The latter design has the obvious advantage of effectively removing genetic variations as the possible “cause” of the disease through monozygosity. If studying quantitative traits rather than diseases, one can resort either to a family-based design or a longitudinal study. In both cases, one may be able to distinguish between inherited and *de novo* epimutations, especially if biospecimens are available in multiple generations, and evaluate their associations with the system-level phenotype (and genotype) of interest.

References

- Daxinger L, Whitelaw E (2012) Understanding transgenerational epigenetic inheritance via the gametes in mammals. *Nat Rev Genet* 13(3):153–162
- Dolinoy DC, Weinhouse C, Jones TR, Rozek LS, Jirtle RL (2010) Variable histone modifications at the A(vy) metastable epiallele. *Epigenetics* 5(7):637–644
- Faulk C, Dolinoy DC (2011) Timing is everything: the when and how of environmentally induced changes in the epigenome of animals. *Epigenetics* 6(7):791–797
- Greer EL, Maures TJ, Ucar D, Hauswirth AG, Mancini E, Lim JP, Benayoun BA, Shi Y, Brunet A (2011) Transgenerational epigenetic inheritance of longevity in *Caenorhabditis elegans*. *Nat* 479(7373):365–371. doi:10.1038/nature10572
- Gregg C, Zhang J, Weissbourd B, Luo S, Schroth GP, Haig D, Dulac C (2010) High-resolution analysis of parent-of-origin allelic expression in the mouse brain. *Sci* 329(5992):643–648. doi:10.1126/science.1190830
- Henckel A, Arnaud P (2010) Genome-wide identification of new imprinted genes. *Brief Funct Genomics* 9(4):304–314. doi:10.1093/bfgp/elq016
- Hjelmeland LM (2011) Dark matters in AMD genetics: epigenetics and stochasticity. *Invest Ophthalmol Vis Sci* 52(3):1622–1631. doi:10.1167/iovs.10-6765
- Kuroda A, Rauch TA, Todorov I, Ku HT, Al-Abdullah IH, Kandeel F, Mullen Y, Pfeifer GP, Ferreri K (2009) Insulin gene expression is regulated by DNA methylation. *PLoS One* 4(9):e6953. doi:10.1371/journal.pone.0006953
- Lamarck JB (1809) Philosophie zoologique: ou Exposition des considérations relative à l’histoire naturelle des animaux. In: Dentu et L’Auteur, Paris
- Lotfipour S, Ferguson E, Leonard G, Perron M, Pike B, Richer L, Seguin JR, Toro R, Veillette S, Pausova Z, Paus T (2009) Orbitofrontal cortex and drug use during adolescence: role of prenatal exposure to maternal smoking and BDNF genotype. *Arch Gen Psychiatry* 66(11):1244–1252
- McGowan PO, Sasaki A, D’Alessio AC, Dymov S, Labonte B, Szyf M, Turecki G, Meaney MJ (2009) Epigenetic regulation of the glucocorticoid receptor in human brain associates with childhood abuse. *Nat Neurosci* 12(3):342–348. doi:10.1038/nn.2270
- Morgan HD, Sutherland HG, Martin DI, Whitelaw E (1999) Epigenetic inheritance at the agouti locus in the mouse. *Nat Genet* 23(3):314–318. doi:10.1038/15490
- Pembrey ME, Bygren LO, Kaati G, Edvinsson S, Northstone K, Sjöström M, Golding J (2006) Sex-specific, male-line transgenerational responses in humans. *Eur J Hum Genet* 14(2):159–166. doi:5201538 [pii] 10.1038/sj.ejhg.5201538

- Portela A, Esteller M (2010) Epigenetic modifications and human disease. *Nat Biotechnol* 28(10):1057–1068. doi:[10.1038/nbt.1685](https://doi.org/10.1038/nbt.1685)
- Rakyan VK, Down TA, Balding DJ, Beck S (2011) Epigenome-wide association studies for common human diseases. *Nat Rev Genet* 12(8):529–541. doi:[10.1038/nrg3000](https://doi.org/10.1038/nrg3000)
- Riddick G, Fine HA (2011) Integration and analysis of genome-scale data from gliomas. *Nat rev Neurol* 7(8):439–450. doi:[10.1038/nrneurol.2011.100](https://doi.org/10.1038/nrneurol.2011.100)
- Roseboom TJ, Painter RC, van Abeelen AF, Veenendaal MV, de Rooij SR (2011) Hungry in the womb: what are the consequences? Lessons from the Dutch famine. *Maturitas* 70(2):141–145. doi:[10.1016/j.maturitas.2011.06.017](https://doi.org/10.1016/j.maturitas.2011.06.017)
- Soshnikova N, Duboule D (2008) Epigenetic regulation of Hox gene activation: the waltz of methyls. *BioEssays News Rev Mol Cell Dev Bio* 30(3):199–202. doi:[10.1002/bies.20724](https://doi.org/10.1002/bies.20724)
- Toledo-Rodriguez M, Lotfipour S, Leonard G, Perron M, Richer L, Veillette S, Pausova Z, Paus T (2010) Maternal smoking during pregnancy is associated with epigenetic modifications of the brain-derived neurotrophic factor-6 exon in adolescent offspring. *Am J Med Genet B Neuro psychiatr Genet*
- Zhang TY, Labonte B, Wen XL, Turecki G, Meaney MJ (2013) Epigenetic mechanisms for the early environmental regulation of hippocampal glucocorticoid receptor gene expression in rodents and humans. *Neuro Psycho Pharmacol* 38(1):111–123. doi:[10.1038/npp.2012.149](https://doi.org/10.1038/npp.2012.149)