We are IntechOpen, the world’s leading publisher of Open Access books
Built by scientists, for scientists

3,800
Open access books available

116,000
International authors and editors

120M
Downloads

154
Countries delivered to

TOP 1%
Our authors are among the most cited scientists

12.2%
Contributors from top 500 universities

WEB OF SCIENCE™
Selection of our books indexed in the Book Citation Index in Web of Science™ Core Collection (BKCI)

Interested in publishing with us?
Contact book.department@intechopen.com

Numbers displayed above are based on latest data collected.
For more information visit www.intechopen.com
Chapter 1

Epigenetic Mechanisms of the Vascular Endothelium

Alexandra A. Majerski, Anthony C. Quinton and Philip A. Marsden

Additional information is available at the end of the chapter

http://dx.doi.org/10.5772/58569

1. Introduction

In the same way that Watson and Crick’s breakthrough in identifying the molecular structure of the DNA double helix revolutionized the last sixty years of molecular biology, the field of epigenetics is set to do the same for the next sixty (or more) years! An alphabet of four letters, and strict guidelines dictating their arrangement, comprise the genome of all known living organisms. This same system has allowed us to unravel many of the secrets to human health and to comprehend a great number of once-elusive conditions and diseases. Nevertheless, a vast number of puzzles remain to be solved, and the genetic code is but one tool used to make sense of them.

Our understanding of our genetics has grown enormously over the last several decades and the iconic image of the DNA double helix has become a cornerstone for the field. Most certainly, the information contained within this static DNA code is but a starting point. Yet, a number of seemingly straightforward questions cannot be answered without going beyond the highly conserved script. Undoubtedly, many ask: Why and how is an endothelial cell different from a brain cell when the static DNA genome is identical? To answer this and many other complex and fascinating phenomena, we must, instead, go over or above (epi-) genetics because the DNA blueprint is identical in each of the abovementioned somatic cells. A term coined in the 1940s by Conrad Waddington, epigenetics refers to chromatin-based mechanisms important in the regulation of gene expression that do not involve changes to the DNA sequence per se. Epigenetic principles, in eukaryotic organisms, are linked by an important common thread: DNA does not exist in a “naked” state. Figure 1 provides a visual representation of the added layer of chemical species, as well as proteins which themselves can be modified – all of which contribute to a layer of chemical complexity and significant implications that epigenetics strives to classify and understand.
Division and differentiation are fundamental processes for sending cells of common origin to different destinations. Logically, the timing and regulation of these processes must be dependent on more than an identical code contained in all body cells. Thus, the epigenome comprises the ever-expanding repertoire of chemical alterations that regulate the expression of genes and, hence, dictate the function of cells and the role of proteins. A number of genetic determinants, as well as lineage-specific markings, and environmental responses are used to construct the epigenome [1]. The existence of an epigenetic code is a highly contested topic [2] and one which will be explored throughout this chapter. Indeed, it is not only mistakes in the genetic code, but also deviations in the epigenome that may provide clues to understanding the onset of detrimental conditions and diseases.

2. Three mechanisms of epigenetics

2.1. Posttranslational histone modifications

Compaction of the entire 6 billion base pair genome, 3 billion in each copy of the haploid genome, into each one of 50 trillion human body cells is a truly breathtaking achievement! In each cell, 2 meters of DNA are condensed by at least 10,000 fold in order to fit into the roughly 6 µm diameter of the nucleus. In accomplishing this monumental feat, intricate folding and wrapping are the first steps in organizing the DNA. Covalent and non-covalent interactions further alter the three-dimensional arrangement of the chromatin. Chromatin refers to the state of DNA when it is spooled around the fundamental repeat unit: the nucleosome [3]. Each nucleosome consists of an octamer containing four different core histone units: histones H2A, H2B, H3 and H4. Two of each core histone unit are present within one nucleosome such that a tetramer of H3 and H4 joins two dimers of H2A and H2B [3]. At least one fifth of the amino acids in each core protein are lysine or arginine – those with highly basic side chains [4]. Approximately 147 base pairs of DNA wrap around each nucleosome. Linker histones, especially H1, interconnect each nucleosome. Structurally, each of the canonical histone core proteins has a characteristic N-terminal tail of amino acids extending out from the core, which is highly susceptible to modifications [3].

Contrary to what biologists believed for many years, kinetic experiments have shown that DNA wrapping around a nucleosome is surprisingly dynamic. Open exposure for 10 to 50 milliseconds – the length of time between rapid winding and unwinding – means that DNA is available for binding other proteins [5]. The formation of nucleosomes converts a DNA molecule into thread-like chromatin approximately one-third of its initial length [4].

Since the initial work in histone acetylation and methylation by Allfrey and colleagues [6], many other modifications have been documented, with the most frequently found to be, at steady state, lysine acetylation, mono-, di-, and tri-methylation of lysines, and the phosphorylation of serines [7]. Concomitantly, a particular collection of enzymes is responsible for each of these additions: histone acetyl transferases (HATs) and histone methyl transferases (HMTs). Importantly, each of these modifications are reversible, the opposite reactions catalyzed by a group of histone deacetylases (HDACs) and histone demethylases (HDMTs). Most often, the
binding of DNA gene regulatory proteins, or transcription factors, is a pre-requisite for the recruitment action of any of these enzymes. Such events occur at different times in the developmental history of a cell and groups of nucleosomes may be modified in a multitude of ways based on the status of the cell [4].

The mechanism by which histone posttranslational modifications (PTMs) occur can be broadly categorized into 3 groups: a) altering chromatin structure with small (acetylation, phosphorylation, methylation) or large (ubiquitylation, sumoylation) chemical groups, b) inhibiting binding factors to the chromatin, and c) attracting a particular set of proteins to the newly modified regions. In the following, we will highlight the notable features of the well-studied mechanisms and allude to the implications of them throughout the remainder of this chapter.

First, acetylation is a posttranslational histone mark that is known to be correlated with transcriptional activation. Broadly speaking, chromatin structure is loosened by the acetylation of lysine (denoted by K) residues. This modification allows for DNA binding sites to be made more accessible, at the same time that a new collection of proteins are recruited to this region of modified chromatin [8]. Positive charges at highly basic histone tails are neutralized with lysine acetylation, which reduces the binding of the amino acid to the negatively charged DNA [9]. An actual change of the physical properties of the histone tails, such as neutralization, is an example of a cis-effect. In this instance, a segmental expansion of the chromatin allows for easier access to transcriptional regulators. When other proteins are actively recruited to the chromatin and read snippets of it, a trans-effect is said to occur. The example HPTM recruiting new proteins was observed with histone tail acetylation. Bromodomain are specialized protein domains, found in HATs such as Gcn5 and CBP/p300. [10, 11].

Figure 1. Three defining epigenetic mechanisms: DNA methylation, posttranslational histone modifications, and RNA-based mechanisms. The four-lettered (adenine, cytosine, guanine, thymine) code that constitutes an individual’s genetic makeup is identical in all cell types and is highly conserved during mitotic and meiotic divisions, whereas the three epigenetic modifications differ amongst cell types and are only partly conserved during mitosis and meiosis. In the “Nurture vs. Nature” debate, genetics characterizes the latter with strong evolutionary convergence while epigenetics is influenced by the former, such that the dynamic patterns that classify it diverge rapidly through evolution.

Figure 1. Three defining epigenetic mechanisms: DNA methylation, posttranslational histone modifications, and RNA-based mechanisms. The four-lettered (adenine, cytosine, guanine, thymine) code that constitutes an individual’s genetic makeup is identical in all cell types and is highly conserved during mitotic and meiotic divisions, whereas the three epigenetic modifications differ amongst cell types and are only partly conserved during mitosis and meiosis. In the “Nurture vs. Nature” debate, genetics characterizes the latter with strong evolutionary convergence while epigenetics is influenced by the former, such that the dynamic patterns that classify it diverge rapidly through evolution.

Epigenetic Mechanisms of the Vascular Endothelium

http://dx.doi.org/10.5772/58569
Histone methylation has proven to be more perplexing than histone acetylation due to the fact that lysine can be mono-, di-, or tri-methylated and arginine can be mono- or di-methylated. Knowing that there are 24 sites of lysine and arginine methylation on the four histone core proteins, means that a tremendous number of methylated/unmethylated states are possible [12-15]. Astonishingly, lysine methylation can result in either the activation or silencing of gene expression. Generally, the methylation of lysines 4, 36, and 79 on histone H3 (H3K4me, H3K36me, H3K79me, respectively) are correlated with transcriptional activation, while the remainder have repressive activity [15]. Two of these marks, namely methylation on lysine 79 of histone H3 and 20 on histone H4, also play a vital role in DNA repair [15]. Arginine methylation is similar, in that the process has been correlated with both negative and positive transcriptional regulation in numerous contexts [16-19].

In contrast to the addition of small chemical species, other larger modifications of histones are described (acetyl, phosphate, and methyl). Ubiquitylation and sumoylation, the addition of ubiquitin (Ub) and small ubiquitin-like modifier (SUMO), respectively, increases the size of the histone by up to two-thirds [20]. Depending on the site of ubiquitylation, this process can result in either transcriptionally active or inactive segments of chromatin [21]. Currently, there are no known ubiquitin chromatin binding proteins, but it is hypothesized that ubiquitin protein is again likely involved in a more diverse array of interactions than are the smaller chemical additions. Although much less well studied, SUMO is thought to recruit deacetylases and/or block lysine substrates that could be acetylated. Both mechanisms are consistent with results that have shown sumoylation to be correlated with transcriptional repression [20].

As a final note on PTMs, chromatin remodelling complexes that rely on the hydrolysis of ATP are known to temporarily alter the structure of nucleosomes by reducing the attraction between DNA and its neighbouring histone. This is commonly referred to as “nucleosome sliding” and the action briefly makes the DNA more susceptible to the effects of other proteins [4]. This sequence of events likely allows for a greater range of interactions of which we are not yet fully aware. It is important to stress that chromatin structure is highly regulated by nucleosome eviction and alternative histone core protein usage. This is a new area of study.

2.2. DNA methylation

The second major epigenetic pathway is DNA methylation - the addition of a methyl group (-CH₃) to the five position of the nitrogenous base, cytosine (Figure 1). Often referred to as the “fifth base pair,” the product of this reaction, 5-methylcytosine (5mC), is the most common mutagenic base in the mammalian genome. Mutation of methyl CpG to uracil pG then TpG results in CpG depletion and TpG enrichment in the mammalian genome. DNA methylation is predominantly observed at CpG sites in the mammalian genome, with the methyl modification at non-CpG dinucleotides unusual, except perhaps in stem cells and mature neurons. Not to be confused with cytosine complementarily binding with guanine across the DNA double helix, CpG sites refer to locations where cytosine and guanine are separated by a phosphodiester bond, adjacent to one another. The vast majority, approximately 70-80%, of all CpG sites in the mammalian genome are, indeed, methylated [22]. Distribution of methylation
across the genome shows enrichment at centromeric heterochromatin, repetitive element (transposon), and non-coding regions [23].

Unlike the great variability of histone modifications, DNA methylation is well known to be a mark of transcriptional repression [24]. Having been the first epigenetic mechanism identified, DNA methylation has been studied for the longest time and its consequences are well entrenched in a variety of essential biochemical processes, abnormal conditions, and diseases. X chromosome inactivation and imprinting are two such phenomena where normal DNA methylation in early development plays a vital role in setting the stage for monoallelic expression from chromosomes during cellular maturation [25]. Deregulation of the methylation mechanism can lead to Immunodeficiency Centromeric instability, and Facial abnormalities syndrome (ICF) as well as being a contributing factor to cancer progression [26].

There are two main hypothesized mechanisms by which DNA methylation is proposed to repress gene expression. The first of these is a physical interference with the binding of regular transcription factors. The methyl group itself may physically project into the major groove of the DNA double helix, whereby it interferes with the sequence-specific recruitment of DNA proteins [27]. In support of this, a number of transcription factors have been found to recognize GC rich regions of the genome – the very same regions that are most susceptible to CpG methylation. Subsequently, these factors are unable to bind when methylcytosine takes the place of cytosine. The second proposed mechanism is in opposition to the first, as proteins are not repelled by the presence of 5mC, but are instead attracted to it. After the purification and cloning of an individual methyl-CpG binding protein, MeCP2 [28], four more similar proteins were identified and are now referred to as methyl-CpG binding domain proteins (MBDs). The attraction of these MBDs has been shown to block the binding of activating transcription factors and to actually recruit to the chromatin, histone-modifying enzymes, chromatin remodelers, and RNA molecules [29-31]. Several such conglomerates have been proposed to co-operatively silence gene expression [32].

2.2.1. DNA methyltransferases

The reaction that yields 5mC as its product is catalyzed by a family of three DNA methyltransferases (DNMTs): DNMT1, DNMT3A and DNMT3B [33]. DNMT1, the first of three family members, is a maintenance methyltransferase as it is responsible for transmitting methylation patterns from one cellular generation to the next. Replication of DNA demands that the two strands first be unwound, and it is during this time that DNMT1 is recruited to the replication fork. Binding of proliferating cell nuclear antigen (PCNA) to both strands and UHRF1 (ubiquitin-like with PHD and RING finger domains 1) protein via its characteristic SET-and RING-associated (SRA) domain, are pre-requisites for the recruitment of DNMT1 [34-36]. This enzyme exhibits methylation preference for hemi-methylated DNA, that is, double stranded DNA where only one strand exhibits methylation [37, 38]. Following replication, the parent strand displays the necessary methylation patterns and is subsequently used as a template for the DNMT1-mediated methyl addition to the nascent, naked strand of DNA. Such sites are accurately methylated due to the palindromic nature of the CpG sequences and the fact that methylation patterns between the two strands are symmetrical. Two other DNMTs include
DNMT3A and DNMT3B, both of which have been shown to function as de novo methyltransferases, displaying preferential methyl addition to unmethylated DNA [39]. These DNMTs play a crucial role in embryonic development, during which time the methylation patterns to be propagated in somatic cells are established [40].

The catalytic domain of the DNA methyltransferases is conserved amongst all of the DNMTs. Despite this, very little similarity is seen at their N-terminal domain. The regulatory domain of DNMT1 is composed of at least five motifs that are not present on the other enzymes, whereas DNMT3A and 3B exhibit two unique motifs. DNMT1 motifs include: a PCNA-interacting domain, a nuclear localization signal, a replication foci-targeting domain, a DNA-binding CXXC region, and a protein-protein interaction motif called the bromo-adjacent homology domain [26]. Those of DNMT3A and 3B include a motif essential for heterochromatin association, PWWP [41, 42] or “pro-trp-trp-pro,” as well as a motif for protein-protein interactions: the ATRX-related cysteine-rich region containing a zinc finger and an atypical PHD domain [26].

2.2.2. CpG Islands

The fact that 5mC is the most common mutagenic base of the mammalian genome can be explained by methylcytosine’s susceptibility to deamination [43, 44]. Upon deamination, the modified nitrogenous base is replaced by thymine and a T:G mismatch is generated. Rather than the rapid repair mechanism activated by cytosine deamination to uracil, T:G mismatches tend to accumulate and are not readily removed. Consequently, the genome becomes CpG depleted and TpG enriched. As indicated in Table 1, CpG islands (CGIs) are regions of the genome that have been spared this phenomenon; the number of CpG sites observed (O) is equivalent to the number that would be expected (E) based on C or G nucleotide abundance. As a result, the ideal situation would be an O/E of 1.

The definition of a CpG island has undergone a great deal of revision since the earliest studies dealing with these genomic regions [45] and are now rather arbitrarily defined as being > 500 bp [46], but averaging around 1000 bp. The O/E range determines the strength of a CpG island with the strongest having O/E > 0.75, a weak one with an O/E between 0.45 and 0.75, and a poor CGI having an O/E < 0.45. These regions are characterized by their elevated C+G content with the %C+G content > 55% [46]. These segments are especially fascinating due to their unusually low levels of DNA methylation marks. In essence, these regions are very often observed to be unmethylated. Due to the lack of repressive flags, genes that are downstream of CGI containing promoters are often in an active state of expression. It has been speculated that practically all CGIs are transcription start sites (TSS), largely of housekeeping and developmentally important genes [47, 48]. Approximately 70% of all gene promoters are CGIs [49].

The relationship between structure and function is seen most vividly in many of the compositional characteristics of CGIs. For instance, a recent study found that the CGI at the promoters of a particular group of genes was relatively nucleosome-deficient [50]. That is, the underlying DNA was highly accessible to activating transcription factors because less of it was spun around nucleosomes. Evidence from additional studies has shown that this nucleosome
deficiency is a feature of CGI promoters, in general [51-53]. It is not yet known whether this phenomenon is a result of intrinsic chromatin instability or nucleosome exclusion due to the presence of the transcription initiation complex. Interestingly, it was found early in the study of CpG islands that, when found at promoter regions, these segments lack the canonical TATA box that so many eukaryotic promoters possess [54, 55]. Constitutive binding of RNA polymerase II is a frequently documented feature of CGIs. Presence of RNA polymerase II would evidently be associated with transcriptional activation as it is the major player involved in regular transcription.

Several histone chromatin features are also known to typify CpG islands, one of which is the depletion of histone 3 lysine residue 36 dimethylation, H3K36me2 [56]. Even though no conclusion has been reached as to why it is the absence of this dimethylation that characterizes most CpG islands, the modification has been reported, in yeast, to inhibit transcriptional initiation via histone deacetylase attraction. [57-59]. This absence makes way for the transcriptionally permissive state of CGIs. Furthermore, CGIs are marked by activating histone acetylation; both H3 and H4 acetylation are frequently observed [51, 60]. Notably, Cfp1 (CXXC finger protein 1) is a well characterized and fundamental component of the Setd1 H3K4 methyltransferase complex [61] and is drawn towards the overwhelming majority of CGIs in the mouse genome [62]. In the context of CpG islands, this protein directs H3K4 trimethylation, the activity of which has been shown to prevent DNA methylation of the underlying code [63, 64]. Hence, the DNA is maintained in a potentially active transcriptional state.

2.2.3. 5-hydroxymethylcytosine: the 6th base pair

The long-standing study of DNA methylation took a significant turn in 2009 with the rediscovery of what is now commonly referred to as the “6th base pair”: 5-hydroxymethylcytosine (5hmC) [65]. This base is the product of an enzymatic reaction that adds a hydroxyl (-OH) group to the methyl of 5-methylcytosine as shown in Figure 2. 5hmC was initially detected in 1972 in mammalian DNA [66], but has since been found primarily, and abundantly, in Purkinje neurons [67], the central nervous system (in the brainstem, spinal cord, and especially the cerebellum [67] and cerebral cortex) [68], and in embryonic stem (ES) cells [65, 69]. The highest tissue levels of 5hmC have been consistently found in the brain [69-71]. The striking abundance of 5hmC in ES cells and PGCs has led scientists to delve into the possibility that 5hmC is linked with the removal of 5mC in such settings, in vivo. There is evidence to suggest that, unlike 5mC, 5hmC may be correlated with transcriptional activation, after having been found at euchromatic regions in mice [72, 73]. In essence, while 5mC is known to be a mark of transcriptional repression, 5hmC may be associated with turning genes back on. Since 2009, a large quantity of research has been generated, dedicated to studying the potential role of 5hmC in gene expression regulation. Despite this, we have only just scratched the surface of discovery as the role of 5hmC in the genome continues to remain highly elusive.

A vital question regarding 5hmC is whether it functions as an intermediate in enzymatic reactions or as a stable entity unto itself. In this chapter, considerable space will be dedicated to discussing the possible demethylation pathways in which 5hmC is speculated to be involved. In this and many contexts, demethylation is defined to be the removal of 5mC and
In the majority of these pathways where 5hmC is required, it is often thought to act as an intermediate, rapidly replaced by another chemical species. Levels of 5hmC are substantially lower than 5mC in the genome, at approximately 10% of 5mC and 0.4% of all cytosines [74]. These minute levels are thought to be consistent with the short-lived species hypothesis as elevated levels would indicate a longer duration. If 5hmC generation is merely a pre-requisite for subsequent demethylation pathways regardless of what form they take, then 5hmC would be considered a transient species.

Evidence is also emerging which supports the notion of 5hmC as an entirely new landmark in the epigenetic gene regulatory landscape. In fact, the modified nitrogenous base may even be responsible for attracting a unique panel of chromatin and transcriptional modifiers. Numerous mechanisms have been proposed based on preliminary observations. Methyl-CpG binding domain protein 1 (MBD1) specifically recognizes methylated DNA and attracts histone deacetylases and H3 lysine 9 methyltransferases (also repressive in action) [75-79]. 5hmC has

<table>
<thead>
<tr>
<th>Distinguished by</th>
<th>Unusually high levels of CpG sites, within a GC rich region, in comparison to the rest of the genome</th>
</tr>
</thead>
<tbody>
<tr>
<td>GC percentage</td>
<td>% C + G &gt; 55%</td>
</tr>
<tr>
<td>Length</td>
<td>Arbitrarily defined as &gt; 500 bp</td>
</tr>
<tr>
<td></td>
<td>Average 1000 bp</td>
</tr>
<tr>
<td>Location</td>
<td>Most often associated with transcription start sites (TSSs)</td>
</tr>
<tr>
<td></td>
<td>Roughly 70% of mammalian gene promoters contain CGIs</td>
</tr>
<tr>
<td>Classification of CGI</td>
<td></td>
</tr>
<tr>
<td>Strong CGI</td>
<td>O/E &gt; 0.75</td>
</tr>
<tr>
<td>Weak CGI</td>
<td>0.45 &lt; O/E &lt; 0.75</td>
</tr>
<tr>
<td>Poor CGI</td>
<td>O/E &lt; 0.45</td>
</tr>
<tr>
<td>Epigenetic Patterns</td>
<td></td>
</tr>
<tr>
<td>Generally unmethylated</td>
<td></td>
</tr>
<tr>
<td>H3K36me2 depletion</td>
<td></td>
</tr>
<tr>
<td>Nucleosome deficiency</td>
<td></td>
</tr>
<tr>
<td>Histone acetylation (H3/H4 Ac)</td>
<td></td>
</tr>
<tr>
<td>H3K4me3 enrichment</td>
<td></td>
</tr>
<tr>
<td>Cfp1 binding</td>
<td></td>
</tr>
<tr>
<td>Constitutive binding of RNA polymerase II</td>
<td></td>
</tr>
</tbody>
</table>

Table 1. Characteristics of CpG Islands (CGIs)

CpG islands are regions of the genome that have remained CpG, rather than TpG rich, the latter being the product of CpG deamination that has taken place over evolutionary time. O/E indicates the ratio of observed CpG sites: expected number of CpG sites. Strong CGIs tend to have a high O/E meaning that the number of observed CpG sites is approaching the expected number of CpG sites, making for a maximum O/E of 1. Often unmethylated, these regions of the genome are associated with transcription start sites. A number of epigenetic patterns have been found to characterize CGIs (H3K36me2=dimethylation of lysine 36 on histone H3 and H3K4me3=trimethylation of lysine 4 on histone H3).
been shown to greatly inhibit the binding of these MBDs to DNA [80-82]. On the same note, DNMT1 is also well known to recruit H3K9 methyltransferases, but the enzyme is blocked by the very presence of 5hmC, so the likelihood of H3K9 methyltransferase binding is also reduced. Methylation has also been implicated in nucleosome compaction, making DNA less accessible to transcription factors [83]. Conversely, the effect of 5hmC on nucleosome compaction has yet to be determined, but the nuclei of Purkinje neurons have been reported to exhibit marked decondensation [84].

Figure 2. Methyl modification to cytosine and hydroxyl addition to 5-methylcytosine. The addition of a methyl group at the 5-position of cytosine is mediated by the DNA methyltransferase (DNMT) family, where S-adenosylmethionine serves as the methyl donor. Notes: The ten-eleven translocation (TET) family of enzymes is able to oxidize 5-methylcytosine to 5-hydroxymethylcytosine in an oxygen-dependent reaction requiring adenosine triphosphate and 2-oxoglutarate. Cytosine can then be generated from the action of putative demethylases on 5-hydroxymethylcytosine, but has not yet been fully described.

2.2.4. Ten eleven-translocation (TET) enzymes

The 2009 seminal paper by Tahiliani et al. identified the TET (ten-eleven translocation) enzymes as being those responsible for the generation of 5hmC by the hydroxylation of 5mC [65]. Adding yet another layer of complexity to uncovering the role of 5hmC in the genome, TET enzymes are also responsible for the further oxidation of 5hmC into new products: 5-formylcytosine (5fC) and 5-carboxycytosine (5caC) [70, 85, 86]. TET enzymes are present throughout the Metazoa and the three forms, TET1, TET2, and TET3, are thought to have arisen as a result of a triplication event that occurred in jawed vertebrates [87]. All three murine forms have been seen to oxidize 5mC to 5hmC in vitro and in vivo [88, 89]. Moreover, the presence of 5mC consistently appears to be a pre-requisite for the generation of 5hmC in vivo [69, 72], suggesting that TET-mediated oxidation is the only means of formation. All of the TET proteins are Fe$^{2+}$ and 2-oxoglutarate dependent dioxygenases [65, 88, 90]. The term, dioxygenase, refers to an enzyme that catalyzes the addition of both oxygen atoms from molecular oxygen to another organic substrate. The iron ion is utilized as a co-factor, whereas 2-oxoglutarate is a co-substrate.
The mammalian TET enzymes all exhibit a high degree of homology amongst their C-terminal domains. It is the Cys-rich and double stranded-helix DSBH region (CD domain) [87] belonging to the Cupin-like dioxygenase superfamily, that contains the catalytic domain for 2-oxoglutarate and iron (II) activity [65, 88]. Further oxidation of 5hmC into 5fC and 5caC also takes place at this same active site [85]. This CD domain also contains the largest identifiable portion of the enzyme: the spacer region. For the most part, the function of this region is unknown, although some leads for future research do exist. For instance, Upabhyay et al. noted remarkable sequence overlap between the spacer region of TET1 and the C-terminal domain of RNA polymerase II of *S. cerevisiae* [91]. Likely of most significance are the conserved residues that are key sites for posttranslational modifications such as histone methylation and phosphorylation [92]. Located on the N terminus (~60 amino acids) of TET1 and TET3 (not TET2) enzymes is the CXXC zinc finger domain [65, 87]. Even though its function has not been conclusively determined, the CXXC domain in TET1 has been shown to bind to unmodified cytosine, as well as 5mC and 5hmC [93, 94]. Speculation has tended towards this domain serving as the functional unit responsible for directing the enzymes to particular genomic regions, almost always CpG sites, so they can execute their oxidative action [92]. A similar CXXC domain on DNMT1, the enzyme on which it was first discovered, possesses comparable DNA binding properties [95], initially having suggested an equivalent role for the domain in TETs.

### 2.3. Long non-coding RNA

Contrary to the central dogma of biology, in which RNA plays a passive role in transmitting genetic information to be translated into proteins, recent advances have shown RNA to play a much more active role in cellular regulation. Long non-coding RNAs (lncRNAs) are the most recent and least well characterized of functional “non-coding” RNAs. The distinction between lncRNAs and miRNAs is arbitrary, but has been defined as greater than 200 nucleotides for an lncRNA [96]. Unlike miRNAs, however, which seem to function through similar pathways, lncRNAs have been shown to play a diverse assortment of regulatory roles, including posttranscriptional repression, RNA splicing, and RNA degradation [97]. Despite this variety, perhaps the most interesting and well-studied mechanism is the ability of many lncRNAs to act as epigenetic modulators, interacting with chromatin remodeling complexes and other epigenetic machinery [98].

*Xist* was one of the first lncRNAs to be discovered [99] and, consequently, is the most well described. The *Xist* gene is found in the X chromosome inactivating center (XIC) where it plays a major role in the inactivation of one of the two X chromosomes in females [100], a process which equalizes X chromosome dosage to that of a male karyotype. *Xist* acts in cis to repress the transcription of the chromosome on which it is transcribed [101]. *Xist’s* function as an epigenetic regulator was confirmed with the discovery of an interacting factor, the polycomb repressive complex 2 (PRC2) [102], a complex that mediates the repressive histone trimethylation mark H3K27me3. PRC2 is recruited to the inactive X chromosome (Xi) by a short repetitive sequence, Repeat A (RepA), conserved on the *Xist* transcript [102]. Tethering of *Xist*
to the chromosome from which it was transcribed allows for allele specific recruitment of the PRC2 complex and subsequent inactivation [101].

Another important IncRNA in X chromosome inactivation is the Xist antisense RNA, Tsix. Tsix is expressed on the active chromosome and blocks Xist expression, thereby preventing its inactivation [103]. The action of Tsix illustrates another mechanism of epigenetic regulation in which IncRNA can take part: DNA methylation. Recently, evidence has suggested that Tsix mediates DNA methylation at the Xist promoter, silencing the Xist gene [104]. Tsix may activate DNA methyltransferase 3A (DNMT3A), one of the de novo methyltransferases, and guide it to the Xist promoter where it is involved in maintaining the silencing of the Xist gene after initial repression.

X chromosome inactivation is not the only cellular process in which lncRNA is involved, and there are numerous other examples of lncRNAs as epigenetic regulators. For example, HOTAIR is an IncRNA located in the homeobox C (HOXC) cluster of genes [105]. The HOX genes are a family of genes that serve as developmental cues in the embryo [106]. The expression pattern of different HOX genes corresponds to specific tissue development and ensures the correct relative positions of body parts. The 39 HOX genes are present in 4 gene clusters, HOXA-D, each present on a separate chromosome [107]. HOTAIR is expressed in the HOXC cluster where it acts in trans to suppress expression on the HOXD cluster. In this way, HOTAIR is required to maintain the specific expression pattern of HOX genes throughout embryonic development. Like the Xist RNA, HOTAIR achieves this repression through the recruitment of the PRC2 complex, which then catalyzes the repressive H3K27 trimethylation mark. Indeed, knockdown of HOTAIR has shown loss of H3K27 trimethylation in the HOXD gene cluster, as well as an absence of the PRC2 complex. This suggests that ncRNA may be a global requirement for activation of PRC2 H3K27me3.

3. DNA methylation and development

Most methylation patterns are stably inherited, remaining static throughout the lifetime of an organism. Specific cellular processes as well as cell-lineage specificity require activation or elimination of these patterns. Fluctuations in 5mC and 5hmC levels are most vividly seen in the cases of pre-implantation development and primordial germ cell formation (Figure 3). In order to understand the significance of the rapid changes in modified cytosine levels, a number of recent and vital observations must be taken into account.

a. Pre-implantation embryo

Throughout the duration of development, especially cellular differentiation, somatic-cell lineages acquire specialized DNA methylation and histone modification patterns. Even in the earliest stages, both sperm and oocyte are known to contain appreciable levels of DNA methylation. Many of the methylated regions are shared between oocyte and sperm, such as intracisternal A particles (IAPs) [108], although there are differentially methylated regions (DMRs) which is oocyte-or sperm-specific. Active and repressive histone modifications are
present in the oocyte at this time [109], and it is maternally inherited proteins that will direct
the events of early cleavage divisions. On the other hand, sperm arrive with temporary histone
substitutes, protamines, which play a similarly functional role in packaging [110].

Upon fertilization, the sperm genome is the earliest to undergo major modifications, the first
of which is protamine removal and substitution with histones [111]. Prior to the fusion of the
maternal and paternal pronuclei, a vast genome-wide demethylation event occurs in the
paternal pronucleus [112-115], whereby its methylation patterns are erased almost in their
entirety. This event can occur in as few as six hours following fertilization [112, 113, 116]. This
demethylation has been assessed repeatedly via immunostaining with antibodies specific for
5mC [112, 113, 116] as well as bisulfite sequencing which has clearly shown that one of two
genomes becomes demethylated. A recently discovered and intriguing fact is that this genome-
wide demethylation is actually a mass oxidation event [117-119]. 5mC specific antibodies do
not recognize 5hmC or any other oxidized cytosine, yet it has been repeatedly and conclusively
demonstrated that the male pronucleus does exhibit a marked increase in 5hmC levels at this
time [117-119]. Complimentary antibody staining has taken place in order to deduce this.
Regions unaffected by this event include paternal DMRs in imprinted genes [120, 121] and
IAPs [108]. It is now well known that it is the maternal TET3 enzyme that executes the oxidative
action on the paternal pronucleus [119]. In contrast to the other TET enzymes, TET3 knock‐
down has been demonstrated to result in the near deficit of 5hmC.

The presence of histone modifications unique to the maternal pronucleus is linked to its spared
erasure at this time of otherwise mass oxidation. Developmental pluripotency-associated 3
protein (DPPA3, also known as stella), has been identified as the key protector of the maternal
pronucleus as well as that of a few paternally imprinted genes [122]. As a matter of fact, the
protection offered by this protein is almost indisputable, as experiments have shown that an
absence of DPPA3 results in the hydroxymethylation of both paternal and maternal genomes
[117, 122]. This protein was originally identified as a result of the upregulation of its encoding
gene during initial PGC development [123]. Dimethylation of H3 histone at lysine 9, a mark
that is restricted to these sites [122], has been reported to be responsible for the attraction of
DPPA3 to these regions.

As indicated in Figure 3, the 2-16 cell stage exhibits a substantial decrease in both maternal-
ly dominant 5mC and paternally specific 5hmC levels. It is a well known fact that the
maternal genome also loses all methylation marks during early embryogenesis. This,
however, has not been attributed to a mass oxidation event, hence, no generation of 5hmC
takes place at this time. The cytoplasmic localization of DNMT1o, a splice variant of
DNMT1, is attributed to the passive replication-dependent dilution of all oxidized cyto-
sines, [112, 116, 124] accounting also for maternal genome methylation erasure. That is, with
each division, 50% of the genomic methylation and hydroxymethylation patterns are lost
due to the absence of maintenance methylation [125], until virtually no modified cyto-
sines remain by the sixteen cell stage [112, 116].

Another generally accepted fact is that a wave of de novo methylation takes place at approxi-
mately the time when the blastula implants into the uterus [116, 120]. This activity occurs very
rapidly and 5mC levels plateau once all of the necessary methylation patterns have been
instilled (Figure 3). That is, no more demethylation is known to take place because patterns placed onto the genome at this stage are the same as those that will be found in somatic cells of later development and maturation. This de novo methylation is now known to be catalyzed by the de novo methyltransferases 3A and 3B [116]. In fact, inactivation of both DNMT3A and 3B has been shown to result in early embryonic lethality while mutations in the encoding genes have also resulted in postnatal or embryonic lethality [126, 127]. Establishing the patterns of the genome at this stage allows for the predictable mitotic transmission of methylation marks throughout the lifetime of an organism [40].

As opposed to the undoubtedly critical role of TET3 following fertilization in mice, TET1 and TET2 are thought to have modulatory, yet not essential roles in development, ES cell endurance, and pluripotency [84]. At the blastula stage, the function of TET1 and TET2 proteins in the inner cell mass is likely to repress lineage-specific genes, counteracting the effect of methylation, so as to permit activation of certain genes later in development [128, 129]. Consequently, they are thought to be fine tuning methylation patterns at this point [84]. Consistent experimental results indicate that there is, however, much still to be confirmed about the relative role of TET proteins in development. As a point of discrepancy, TET1 deficient mouse ES cells have been seen to maintain their self-renewal capabilities in vitro and develop ordinarily in vivo [130]. Furthermore, numerous groups have demonstrated that the knockout of both TET1 and TET2 results in the normal development and healthy adulthood of fertile mice [130-134]. It is hypothesized that such discrepancies may be a result of the alternate TET proteins compensating for the depletion of one another [92]. Although the biological roles of TET1 and TET2 in zygotic development are still highly disputable, that of TET3 appears to be more comprehensive. One study, in particular, highlighted the importance of this particular enzyme in embryogenesis by introducing a homozygous mutation in the TET3 encoding gene, subsequently observing neonatal lethality [119]. The same study came to the conclusion that failure to demethylate the male pronucleus is a direct result of TET3 depletion.

Evidently, epigenetic modifications play an unquestionably enormous role in early development. Despite remaining relatively constant in the mitotic transmission of later development, methylation patterns undergo rapid, yet precisely timed fluctuations in zygotic development. 5hmC abundance points to a vital role of the TET enzymes, as no other pathway has been conclusively verified for the generation of hydroxymethylcytosine from methylcytosine. In light of the recent re-discovery of 5hmC and already substantial amount of scientific inquiry that has gone into deciphering its role in development, we can hope for a great deal more to come in the near future!

b. Primordial Germ Cells

Primordial germ cells (PGCs) are the cells that ultimately give rise to the germ line, the oocyte and sperm. Similar to what has been observed in zygotic development, mouse PGCs, derived from epiblast cells, are known to experience a correspondingly fast decline in 5mC levels [135] between embryonic day (E) 9.5 and 13.5 [124, 136, 137]. Very recently, Hackett et al. decisively demonstrated that the demethylation that takes place in PGCs is conceptually equivalent to
the mass oxidation event that has been observed during embryonic development. In this case, TET1 and TET2 are responsible for the generation of 5hmC [124, 138].

Thus far, the acquisition of 5hmC appears to take place between E9.5 and E11.5 [124], but seems to be lost at a rate that is typical of replication-dependent pattern dilution [124]. Such circumstances closely parallel those seen in zygotic development. In the same way that the DNMT1o splice variant has been seen to be excluded from the nucleus in zygotic development, UHRF1 consistently appears to be down-regulated in PGC formation [84]. Such a situation prevents DNMT1 from performing its prescribed action. Evidence indicates that TET-mediated oxidation in PGCs is likely made possible by passive, rather than active demethylation [124]. This observation is further substantiated by the absence of both 5fC and 5caC following PGC demethylation, as shown in immunocytochemistry studies [124], whereas their presence in zygotic development indicates an increased likelihood for an active demethylation mechanism at work.

By definition, methylation patterns at CpG islands will never be erased during normal development. Yet, CpG island methylation in the DMR of imprinted genes must undergo
erasure so as to ensure that gender-specific methylation associated with imprinting can be put into place during the remainder of germ cell development [123]. Demethylation at this time has also been proposed to be responsible for reinstating the pluripotent state in PGCs. Clearly, those regions whose expression had been repressed are now turned on, allowing for the germ cell lineage acquisition of pluripotent characteristics, perhaps a pre-requisite to achievement of totipotency. Similar to the spared erasure of paternally imprinted genes in zygotic development, certain transposons such as IAPs remain moderately methylated at this time [34].

In contrast to zygotic development, the demethylation of imprinted regions is absolutely critical because germ cells possess imprints that are uniquely exclusive to the gender of the organism. As indicated in the bottom half of Figure 3, an analogous round of de novo methylation also takes place later in oogenesis and spermatogenesis. Re-methylation along with the re-installation of imprints starts at approximately E15 in males and, in females, after birth [139].

3.1. DNA demethylation pathways

As illustrated above, it is a well ingrained fact that epigenetic reprogramming does take place in zygotic development and PGC formation. More recently, TET protein activity has been observed in relation to the timing of particular developmental events. Consequently, levels of 5hmC have been monitored and correlated with the activity of TET proteins at each stage of interest. Truly novel is the fact that this reprogramming involves a mass oxidation event, that is, the generation of 5hmC. The function of 5hmC in these stages is a highly debated and controversial area of research. We have already addressed the first disputed issue regarding the stability of this oxidized base, but we will now draw our attention to the second debate: What are the possible demethylation pathways by which reprogramming takes place? We will define demethylation as the eventual replacement of 5mC with unmodified cytosine. Possible mechanisms for this can be broadly grouped into passive and active pathways. In analyzing the possibilities, it is important to note that absolutely no consensus exists as to which mechanism predominates and that there is experimental evidence to both confirm and rebuke nearly all pathways.

3.1.1. Passive demethylation

Passive demethylation is fairly well established as one of perhaps several operating processes. As mentioned above, the maintenance DNMT places methylation patterns on a progeny strand using the parental strand as a template, similar to the synthesis of a new strand of DNA whereby DNA polymerase utilizes the parent strand as a template for construction of a complementarily base-paired DNA strand. Like the DNA double helix, the maintenance of methylation patterns also occurs through a semi-conservative pathway. This particular setup yields a hemi-methylated (partly methylated) pattern on the DNA. UHRF1 binding and PCNA recruitment are the first steps in the faithful copying of methylation patterns from one strand to another, thereby ensuring the transmission of methylation patterns through mitotic division. From the perspective of demethylation pathways, if this faithful re-installation of methylation patterns is disturbed, every generation will subsequently exhibit 50% less methylation. Eventually, such patterns will undergo what is called passive replication-dependent dilution.
As mentioned, this appears to be the way in which the methylation patterns of the zygotic maternal genome are erased. Thus, the plausibility of this mechanism is not debated. Paternal genome erasure prior to replication is significant evidence that defies the notion of passive demethylation as the only demethylation pathway.

As far as active or replication-independent pathways go, the central tenets of a variety of these pathways will be discussed in the following. A common thread that runs beneath a number of these is the coupling of TET protein activity with a well-known DNA repair pathway, base excision repair (BER), whereby a nitrogenous base is excised by a glycosylase enzyme and replaced with a new one. Some sort of replacement mechanism is practically inevitable given that an unmodified cytosine must take the place of a previous base in order to complete demethylation.

3.1.2. Pathways in which 5fC and 5caC are involved

The first of these proposed pathways involves the other two oxidized bases, 5-formylcytosine (5fC) and 5-carboxylcytosine (5caC) that can be generated by further oxidation of 5hmC by TET enzymes [70, 85, 86]. Their presence at very low quantities in the genome (5fC: 0.03% of 5mC in mouse ES cells; 5caC: 0.01% of 5mC in mouse ES cells) [70, 85, 86], has been attributed to the preference of TET enzymes towards 5mC, not 5hmC, as a substrate, as well as an exceedingly fast removal of the oxidized bases [85, 140]. Thymine DNA glycosylases (TDGs) are capable of excising both bases once they are produced. Kinetic studies by Maiti & Drohat clearly demonstrated that TDG excises 5fC and 5caC, but not 5hmC [141]. Neither when left for a longer time nor performed at a lower temperature (for enzyme stability maintenance) was there any significant 5hmC excision [141]. Excision of the oxidized bases opens the door for base excision repair (BER) action that completes the guanine base on the opposite strand with a brand new cytosine, hence, completing the demethylation pathway. Several labs have confirmed that the proposed is a highly plausible mechanism [85, 141, 142].

Interestingly, a number of studies have shown that TDG exhibits strong specificity in its action: it excises bases which are complementarily base paired with G and/or those which are also followed by guanine [143-146]. Such an observation is entirely consistent with the proposed role of TDG and CpG sites [147] whereby the cytosine has undergone a further modification, yet still fulfills the remainder of the criteria. More research must be undertaken to confirm whether TDG action on modified cytosines, followed by BER, is a plausible demethylation mechanism in vivo. Whether such a pathway is of sufficient speed to play a role in demethylation is one concern which must be addressed. Another comes about when we consider the exceedingly minute quantities of both of the bases in the genome: How prevalent can this pathway realistically be? Is it enough to account for mass demethylation in both zygotic and PGC development?

There is limited evidence that supports a decarboxylation pathway of 5caC, mediated by proteins present in lysates of ES cells [148]. Although the factors that catalyze this reaction have yet to be concisely identified, Schiesser et al. utilized isotope tracing to detect the direct conversion of 5caC to cytosine without interference from BER [148]. The existence of a
deformylation pathway involved in the direct conversion of 5fC to cytosine has also been postulated.

3.1.3. Direct removal of 5-methylcytosine

What about a simple removal of the 5mC base followed by TDG action? Scientists have noted that, due to the strength of the carbon-carbon bond needed to be broken, an enzyme with unwieldy catalytic power would be needed to achieve such a feat [140]. Methyl-CpG binding domain protein 2 (MBD2) was first and foremost reported to catalyze just such a reaction. By its very nature, MBD2 is known to bind to methylated cytosine, so it was often questioned how it could be responsible for mediating such an active reaction if it first must attach itself [28, 75]. Numerous other downfalls of this theory emerged and no lab was apparently able to replicate the results, so the theory of MBD2-mediated direct removal of the methyl group of 5mC fell out of favour. Hence, the inevitability of some sort of TDG/BER activity is further substantiated [149].

A model plant used in genetics and epigenetics studies, Arabidopsis thaliana, is the only source that has thus far provided evidence for such a direct removal of 5mC (reviewed in [150]). The existence of ROS1, DME, DML2, and DML7-a family of four 5mC DNA glycosylases has been acknowledged in this plant [150]. Their preference for acting on 5mC in double stranded DNA, has led researchers to claim that these enzymes are indeed necessary for active demethylation of particular genes. The function of these glycosylases is similar to that of BER, in that a nick created leaves an abasic site which is rapidly repaired. Yet, the enzymes and mechanisms in mammals appear to be quite distinct from those that have been identified in the plant [151].

3.1.4. AID/APOBEC activity

The second well investigated, yet still highly debated active replication-independent demethylation pathway is one which also utilizes a DNA repair pathway. AID (activation-induced cytidine deaminase) and APOBEC (apolipoprotein B mRNA editing enzymes, catalytic polypeptide) family members are primarily known for their ability to deaminate cytosine to uracil [152]. In vitro studies have demonstrated the further ability of AID/APOBEC to deaminate 5mC in DNA, to thymine, resulting in a T:G mismatch [149]. The excision conducted by a TDG glycosylase would subsequently result in an abasic site (apurinic and apyrimidinic [AP] site) which would then be subject to repair by BER. The final product of such a chain reaction would be the regeneration of cytosine in place of the modified base [149].

There are a number of variations and caveats of this and related processes. For instance, AID and APOBEC have frequently been seen to function in a single-stranded DNA environment [149]. Deamination in single-stranded DNA has been experimentally observed in vitro and in an E. coli assay [149]. Despite this, AID and APOBEC are both present at appreciable levels in oocytes as well as PGCs, so their potential role in developmental demethylation has not been dismissed [149].

An alternate pathway through which AID/APOBEC are thought to function involves the deamination of 5hmC to 5-hydroxymethyluracil (5hmU) and its subsequent removal by a TDG
glycosylase or BER pathway [153]. This process would result in the eventual replacement by cytosine. The over-expression of TET1 and AID have been observed to lead to a global accumulation of 5hmU [153]. Two potential pathways may compete with one another because both TDG as well as selective monofunctional uracil-DNA glycosylase 1 (SMUG1) have strong affinity towards the mismatch that would be generated upon deamination: G:hmU [153, 154]. Hence, one of these mechanisms is thought to play a vital role in mediating the downstream commencement of BER following 5mC hydroxylation. Such observations substantiate what many studies propose: that TDG plays an absolutely essential role in active demethylation, regardless of what mechanism succeeds it.

Perhaps what could be considered the downfall of the AID/APOBEC pathway was elucidated in a series of experiments and simulations conducted by Nabel et al. [152]. Insisting that the role of this deamination-based mechanism is limited, the researchers examined the plausibility of this pathway by investigating the biochemistry of natural and non-natural modified cytosines. The series of substrates represented a 150 fold difference in reactivity, with the size of the substituent being an integral determinant of susceptibility to deamination. At one end of the scale was unmodified cytosine, the behavior of which as a deaminase substrate was unequivocally superior to all other larger natural and unnatural substrates. The bulky 5hmC not only sterically hindered the activity of AID/APOBEC, but its poor hydrophobic character was also implicated in substantially decreasing its reactivity, in comparison to all other bases. The researchers consolidated their results and firmly stated that AID/APOBEC family members preferentially deaminate unmodified cytosine, yet strongly discriminate against any and all 5-substituted cytosine substrates [152]. Clearly, this result contradicts the notion of 5hmC deamination to 5hmU, and more so favours the other possibility of 5mC deamination to thymine. The hope is that future research will work towards eliminating such discrepancies amongst multiple lines of work as we gain a better understanding of the plausibility of these pathways.

3.1.5. DNMT dehydroxymethylation

In addition to AID and APOBEC, the activity of DNMTs themselves has recently been proposed as a possible 5mC demethylation mechanism. In 2012, Chen et al. demonstrated that DNMT enzymes can remove the hydroxymethyl group of 5hmC in vitro [155]. Hence, 5hmC is directly converted to cytosine [156]. Although it is still unknown whether such a reaction takes place in cells [84], it has been established that reducing conditions favour the generally accepted methyltransferase activity of DNMT3A and 3B while oxidizing conditions are ideal for dehydroxymethylation activity [155]. Participation in both methylation and demethylation results in rapid cycling between the removal and replacement of methylation patterns – a phenomenon that has been reported in two papers to date [157, 158]. In spite of our analysis of a limited number of experiments, both papers refer to the interplay of TDG and BER in this transcriptional cycling. Particular repair proteins were, in fact, recruited to site-specific locations, likely assisting in the reconstruction of the site [140]. Physiological implications must also be taken into account. S-adenosylmethionine (SAM) is required, not only at low levels for the DNMT3A/B mediated reaction, but also at specific levels for other biochemical reactions
Scientists continue to ponder the other biological effects which would result from rapid SAM level fluctuations.

Quite clearly, numerous potential active demethylation pathways do exist and the validity of one over the other is constantly being reassessed. Further oxidation to 5fC or 5caC, direct removal of the methyl group, AID/APOBEC, TDG, and BER pathways have all been proposed and are now under scientific scrutiny. Numerous sources highlight the possibility of more than one of these pathways operating simultaneously with one or more of the others.

4. Case study I: vitamin C

4.1. Scurvy

Lethargy, extreme muscle and joint pain, gum disease, loosening of teeth, and emotional instability. Often mistaken for syphilis, leprosy, or dysentery, such were the symptoms that characterized “the plague of the sea” [159]. Now a rare disease, scurvy took the lives of two-thirds of Vasco de Gama’s men, eighty per cent of Magellan’s mates, and terrorized many of those who took to the sea during the Age of Exploration. Little did the famed 15th century explorers know that they were losing members of their crew to a vitamin deficiency. Sailors who boarded ships destined for long journeys relied mainly on cured and salted meats as well as dried grains, and suffered greatly for their deprivation of perishable fruits and vegetables. It was not until its identification and isolation in the 1930s that the association between vitamin C (ascorbate) and its role in scurvy was made. In fact, ascorbic acid, from which the L-enantiomer, ascorbate, is derived, actually means “anti-scurvy” [160].

Unlike many animals, primates, humans, guinea pigs, and a few varieties of fish, cannot synthesize vitamin C de novo in the liver [161]. In humans, for instance, the final step of vitamin C biosynthesis cannot be performed: the conversion of L-gulono-gamma-lactone into ascorbic acid. The gene that encodes gulonolactone, the enzyme which catalyzes this reaction, is present in the human genome, but is inactive as a result of mutation accumulation that turned it into a pseudogene [162]. Vitamin C is known to be a critical component of health, while lack of it is truly a detriment to the body because of its dual function in a number of enzymatic reactions as well as its role as an antioxidant.

4.2. Collagen biosynthesis

Without a doubt, vitamin C presence is absolutely vital in the regular synthesis of collagen. Being the most abundant protein in mammals and composed mainly of connective tissue [163], the triple helical construction of collagen requires the successful completion of a number of biochemical events. One of these is the essential enzymatic hydroxylation of proline (Pro) residues into hydroxyproline (Hyp) by prolyl-4-hydroxylase (P4H) [164]. A member of a dioxygenase superfamily, P4H decarboxylates the co-substrate, α-ketoglutarate, and utilizes Fe²⁺ as a co-factor, yielding hydroxyproline and an oxidized Fe³⁺ [165]. For full catalytic activity, ascorbate is required as another co-factor in the reaction [166, 167]. In the absence of ascorbate,
this reaction can perform at a maximal rate, but only to a certain extent [164]. A buildup of Fe$^{3+}$ requires the activity of ascorbate as a potent reducing agent to regenerate the reduced form of iron: Fe$^{2+}$[168, 169]. In this way, the catalytic potential of P4H is maintained at an efficient level.

Ascorbate deficiency, the situation which arose when sailors went to sea for several months at a time, leads to the incomplete hydroxylation of proline residues in collagen, which results in its improper folding [170]. Non-functional collagen in blood vessels and bones has been attributed to the severe bone and blood vessel related symptoms that came to characterize scurvy. As a point of interest, the symptoms that extended beyond the changes in bones and blood vessels, have also been linked to vitamin C deficiency. Prolyl hydroxylases are only one of a number of enzymes whose mechanism of action relies on this vitamin. The synthesis of norepinephrine from dopamine, for instance, is catalyzed by dopamine beta hydroxylase, another enzyme which utilizes vitamin C as a co-factor [171]. Required for the generation of metabolic energy via mitochondrial breakdown of fatty acids, carnitine synthesis [172], also necessitates the utilization of ascorbate. Thus, the neurological dysfunctioning of the sailors and their symptoms of lassitude can both be attributed to a lack of this indispensable vitamin.

Figure 4. Structure of ascorbate. The vital nature of vitamin C (ascorbate) is exemplified by its dual function in the human body, acting as both a cofactor as well as an antioxidant. Collagen biosynthesis necessitates the hydroxylation of proline residues by prolyl-4-hydroxylase – one of a number of biochemical reactions that require the presence of vitamin C. Belonging to the same family as P4H, TET proteins were proposed as possible enzymes whose catalytic action might also have been augmented by the presence of vitamin C. Recent studies have found that ascorbate is responsible for widespread, yet specific demethylation, by way of inducing the generation of 5hmC by TET enzymes.

Thus far, we have explored an interesting historical event that revealed the enormous impact of vitamin C on human health. The indisputably important role of vitamin C is further highlighted by its role in collagen biosynthesis. What’s more intriguing is the fact that prolyl hydroxylases belong to the same family as the TET enzymes and share numerous structural features. The possibility of vitamin C interacting with the TET proteins and potentially contributing to the epigenetic landscape of modifications has become an area of scientific interest.

A recent report from Minor et al. conclusively stated that ascorbate induces the generation of 5hmC by TET enzymes, likely acting as a co-factor in the reaction [173]. Utilizing mouse embryonic fibroblasts (MEFs), the addition of ascorbate in a time-and dose-dependent manner...
was unmistakably seen to correlate with increased 5hmC levels. A novel role for vitamin C in epigenetic modifications has been brought to the forefront. In cells with knocked down TET expression, ascorbate still enhanced the generation of 5hmC by at least 3 fold above basal levels [173]. Rather than protein synthesis, it was shown that the mere accumulation of ascorbate inside the cells was all that was necessary in order to observe increased 5hmC levels. The researchers noted that a detailed mechanism for ascorbate’s interaction with the TET enzymes has yet to be elucidated, but the similarity between TET and P4H catalytic action, points towards ascorbate performing a comparable role in TETs as it does for P4H [173].

Due to its enhancement of cell proliferation, vitamin C is a commonly used medium supplement in ES cell culture. A Stem Cells report from Chung et al. showed that vitamin C, at the concentration used in commercial media (50 µg/mL), promotes widespread DNA demethylation of human ES cells, yet, with remarkable specificity for the location of demethylation [174]. This study showed that the ascorbate-mediated hypomethylation of CpG island shores was highly correlated with gene expression at these sites. A precise set of 1,847 genes experienced altered gene expression as a result of ascorbate addition. From this, researchers speculated that DNA topology may have been changed such that histone demethylase and acetylase activity subsequently altered the accessibility of DNA to other factors [174].

Clearly, numerous connections are being made between vitamin C and the epigenetic landscape. Not only does the vitamin act as a co-factor in enzymatic reactions, but its biochemistry allows it to participate in an even wider array of interactions (Figure 4). As its demethylation potential is remarkably specific in locale, it was first speculated that this vitamin must be interacting with other enzymes to execute its action. Indeed, this interaction has been confirmed in the way that TET enzymes are thought to modulate the actual demethylation with ascorbate acting as a co-factor. Although the exact antioxidant behavior of vitamin C is not thoroughly illuminated here, the potential role of the general class of antioxidants is revealed in the context of the pathogenesis of atherosclerosis. Hence, vitamin C is intricately linked with another underlying theme of this chapter: cardiovascular disease.

4.3. Epigenetic mechanisms and the vascular endothelium

Endothelial cells, those that line the lumen of all blood vessels in mammals, mark the interface between blood flow and the entrance of substances into tissues and organs. The endothelium, itself, is composed of a thin layer of endothelial cells and an even thinner layer of basal lamina beneath the cells. Vascular smooth muscle cells neighbour the endothelial cells, but, as will be discussed later, exhibit remarkably distinct gene expression. Vasorelaxation was argued, by Furchgott and Zawadzki, to rely on an intact endothelial lining, from which an endothelium-derived relaxing factor (EDRF) would be released [175]. Soon after, the EDRF was recognized to be nitric oxide (NO) [176, 177]. Nitric oxide is now considered to be an incredibly vital component for maintaining vasculature tone due to its antiatherogenic and antithrombotic properties [177]. Nitric oxide generation is highly implicated in vasodilation, as well as its inhibition of platelet aggregation, leukocyte adhesion, and vascular smooth muscle cell proliferation [178-180].
A number of biochemical pathways in aerobic cells produce oxygen-derived free radicals as by-products. Free radicals are molecules with a single unpaired electron, exhibiting great chemical instability and high reactivity. Donating or accepting an electron normally results in the generation of another free radical. Normal cellular respiration produces $O_2^-$, the superoxide anion, while the hydroxyl radical OH is the by-product of the reaction intended to eliminate $O_2^-$. Humans have evolved a number of strategies to combat the production of these reactive oxygen species (ROS). For this purpose, humans are able to utilize diet-derived or de novo antioxidants in an effort to convert the damaging species into less toxic substances. Despite this effort, when the levels of reactive oxygen species are appreciably greater than antioxidant levels, a condition known as oxidative stress results. Free radicals, which are further generated from oxidative stress, induce tissue damage and are responsible for detrimental modification of lipids and proteins in the vasculature.

The damaging impacts of oxidative stress are seen nowhere as vividly as they are in the endothelium. Endothelial function is impaired by the conditions and NO production is diminished. The effects are disastrous: reduced vasodilation, enhanced platelet activation, and increased vascular smooth muscle migration and proliferation [181]. In the way that each of these outcomes lead to dysfunctioning of the endothelium, all have severe consequences for the pathogenesis of atherosclerosis. A condition primarily of the intima and a pre-cursor for a number of severe organismal dysfunctions, atherosclerosis can affect arteries as large as the aorta or as small as the tertiary branches of coronary arteries [182].

4.4. Pathogenesis of atherosclerosis

The lipid deposition associated with atherosclerosis is derived from an excess of low-density lipoprotein (LDL), the so-called “bad cholesterol,” in the plasma. Endothelial damage, such that would occur during oxidative stress conditions, allows for the leakage of LDL into the sub-endothelial space via LDL receptors in endothelial cells [183, 184]. There, the free radical and other oxidized agents attack incoming LDL molecules and generate oxidized LDL (ox-LDL) [184]. Endothelial cells and smooth muscle cells have the capacity to oxidize LDL utilizing free radicals, the presence of which initiated the oxidative stress conditions [185-187]. Expression of adhesive cell surface glycoproteins such as VCAM-1 and ICAM-1 is induced as a result of mildly oxidized LDL [188-190]. Moreover, a variety of molecules within the intravascular space are responsible for attracting and modifying monocytes. An environment such as this, is now rich in modified lipoproteins, chemoattractants, and growth factors, all facilitating the differentiation of incoming monocytes into macrophages. These macrophages are quite powerful in that they are capable of a more relentless oxidation of LDL, acting as phagocytes and promoting smooth muscle cell migration and proliferation [188-191]. These lipid filled macrophages accumulate rapidly and begin to form the foam cells that characterize an atherosclerotic lesion on its way to becoming a clinically significant and potentially severe condition. Because LDL leakage is un-regulated, greater accumulation will lead to the bursting of the cell and an explosion of free radicals and ox-LDL – an event which upholds the cycling of cytotoxicity, worsening the damage to the endothelium.
Clearly, the oxidation of LDL exhibits an enormously harmful effect on the endothelium and drastically alters the dynamics of its cellular function. Hence, endothelial injury and dysfunction have been proposed as the initiating events of atherosclerosis. In the presence of cardiovascular risk factors including hyperlipidemia, hypercholesterolemia, smoking, diabetes, and hypertension, the likelihood of cardiovascular disease is highly augmented.

For quite some time, increased antioxidant intake has been investigated for its potential biochemical role in reducing the levels of ROS. Antioxidants are known to protect cells from the effects of free radical damage either by disturbing their mode of production or by attenuating their proliferation once they have formed. They accomplish the former by converting free radical reactants into stable species and the latter takes place via neutralizing the effects of free radical molecules. More precisely, these antioxidants often act by donating an unpaired electron to the unstable free radical, themselves now in the position of becoming a pro-oxidative species. Yet, the structure of antioxidants is usually such that their unpaired electron can be stabilized by adapting one of several possible resonance structures [192-194]. Hence, antioxidants appear to be a very viable preventative measure towards the onset of atherosclerosis and potentially subsequent cardiovascular disease. Long-term studies must be conducted to judge the dose-and time-dependency of the plethora of potential antioxidants. Many of these substances play several integral roles in the body and any beneficial outcomes may not be a direct result of a substance’s anti-atherosclerotic properties, per se. Furthermore, atherosclerosis pathogenesis is outstandingly complicated and it is now known that the condition’s progression is linked to a number of redox-sensitive events that extend beyond the oxidation of LDL [195]. A great deal more investigation will have to be undertaken in an attempt to determine which components of the oxidative stress state are critical for atherosclerosis and the precise mechanism by which antioxidants function.

4.5. Epigenetic regulation of eNOS

Until this point, nitric oxide production and inhibition have been discussed in the context of atherosclerotic plaque formation and their potential role in cardiovascular disease. Moreover, from an epigenetic perspective, the endothelium and its biochemical properties also present an incredibly fascinating model. In mammals, three isoforms of nitric oxide synthase (NOS), encoded by genes on distinct chromosomes: neuronal NOS (nNOS), inducible NOS (iNOS), and endothelial NOS (eNOS), are responsible for the production of nitric oxide [196-198]. The latter of these, eNOS, is known to be the dominant producer of NO and has been found to be constitutively expressed in the endothelium [199]. The eNOS gene, itself, contains 26 exons over the length of 21 kb of genomic DNA, maps to chromosome 7q35-36, and is present as a single copy in the haploid genome [196-198, 200]. Additionally, the gene’s promoter has no proximal CpG island and a TATA box is also absent [201]. eNOS deficiency, in animals, has been linked to pulmonary hypertension, atypical vascular remodeling, and flawed angiogenesis [181, 202]. eNOS deficiency in humans has been observed in pulmonary hypertensive individuals and in the coating of advanced plaques [203]. Much attention has been given to discovering and documenting the regulation of eNOS in vitro as well as in vivo.
Transcriptional, posttranscriptional, and posttranslational pathways have been linked to the intriguing regulation of eNOS [199]. In addition to von Willebrand factor (vWF), vascular-endothelial cadherin (VE-cadherin), intercellular adhesion molecule-2 (ICAM-2), and the vascular endothelial growth factor receptors (VEGFR1 and VEGFR2), eNOS is one of many proteins whose expression is largely endothelial-cell specific [181]. This remarkable nuance has allowed for gene regulation study, from a predominantly epigenetic perspective.

The classic cis/trans mentality would be the simplest way to account for the endothelial-cell specificity that is observed in eNOS and the other abovementioned proteins. By this way of thinking, the cis-acting elements would bind the 5′ regulatory regions of target genes in near proximity. Trans-acting factors would be recruited, but they themselves, would have to be cell-restricted in expression. This arrangement would describe gene expression regulated by a master transcription factor. Several well documented instances of master regulation, in mammalian biology, include NeuroD in neurons [204], smooth muscle cell myocardin [205], peroxisome proliferator-activated receptor-γ in adipocytes [206], with the most profound example being MyoD in skeletal muscle cells [207]. In spite of all of these examples, a master regulator in endothelial cells has yet to be discovered [181]. There are some transcription factors, such as Hey2, Vezf1, FoxA9 [181], and KLF2 [208] which are preferentially expressed in the vascular endothelium, but can also be found in other locales. Much research is dedicated to establishing the prominence of each of these in the vascular endothelium, especially KLF2 (kruppel-like factor 2). Cis-DNA binding elements in the 5′-regulatory regions have been found in a number of the endothelial-cell restricted genes, and the current train of thought tends towards the cooperative activity of a unique combination of transcription factors present in endothelial cells and not in any other cell type [181].

4.6. Role of DNA methylation & eNOS expression

A differentially methylated region (DMR) that had been detected in the proximal promoter of eNOS in expressing cell types has been observed to be unmethylated or lightly methylated whereas that same DMR in nonexpressing cell types was heavily methylated [209]. Combination of chromatin immunoprecipitation (ChIP) and quantitative real-time PCR revealed the preferential attraction of transcription factors Sp1, Sp3, and Ets1 to the eNOS proximal promoter of endothelial cells, even though they were also present in the tested nonexpressing cell type: vascular smooth muscle cells (VSMC) [209]. Not only did RNA polymerase II preferentially bind to the proximal promoter of eNOS in endothelial cells, but the transcriptionally repressive methyl-CpG-binding domain protein, MeCP2, was attracted to the promoter of VSMCs [210]. Upon treatment with the DNMT inhibitor, 5-azacytidine, those heavily methylated CpG sites were found to be unmethylated, while expression of the eNOS mRNA was seen to increase [209]. Hence, human eNOS is the first identified constitutively expressed vascular endothelium gene whose specificity in expression, is at least partially dictated by DNA methylation [209].
4.7. A histone code for the endothelium

Despite the prominence of DNA methylation pathways in these intricate biochemical sequences, other chromatin based mechanisms, especially histone posttranslational modifications, have been studied in the context of eNOS. Transcriptionally activating marks such as the acetylation of H3 and H4 as well as the di-and tri-methylation of H3K4 have been experimentally detected in nucleosomes of the eNOS proximal promoter in expressing, but not nonexpressing human cell types [211]. As would be predicted, histone deacetylases (HDAC1 but not HDAC 2 or 3) were found at the promoter region of nonendothelial cell types [211]. We used ChIP analysis which recognized, at the eNOS promoter in endothelial cells, the discernible enrichment of acetylated H3K9 and H4K12. The above observations support the existence of a histone code unique to the endothelium, whose precise posttranslational modifications are, to a degree, responsible for gene expression regulation in the endothelium [209, 211].

In summary, current studies are attempting to decipher the relative importance of both DNA methylation and histone posttranslational modifications in gene expression regulation in the vascular endothelium. The ways in which these patterns are initially laid down in development are likely complex and understanding them is still well into the future. It has been speculated that disturbances of these epigenetic pathways may have implications for human health and disease. As we have seen, decreased eNOS expression, and hence, reduced NO production, can have profound consequences for conditions such as atherosclerosis.

4.8. LncRNA and the cardiovascular system

With the traditional four base static genetic code failing to account for the missing heritability of complex genetic disease phenotypes, the dynamic epigenetic code is becoming widely accepted as an important determinant of cardiovascular pathologies [212]. Epigenetics provides a heritable link between our environment and diseases classically linked to genetics. It should come as no surprise, then, that if lncRNAs can act as epigenetic regulators they might have a crucial role to play in the development and homeostasis of the cardiovascular system (Figure 5).

Advances in genome sequencing have allowed for new genetic screening techniques that are more accurate than ever. Within the last decade, genome wide association studies (GWAS) have paved the way for understanding genetic variation in common complex diseases. While traditional family based and linkage analysis studies are effective in identifying rare disease causing mutations, they fail to uncover genetic variation amongst common diseases such as cardiovascular disease [213]. Advances in genomic sequencing have lead to the establishment of databases like the Human Genome Project and the Human Haplotype Map Project that have allowed GWAS to become a useful and accurate tool for correlating single nucleotide polymorphisms (SNPs) to disease phenotype when implemented correctly. Using microarray technologies, up to a million SNPs can be assessed.

Interest in cardiovascular epidemiology has led to a number of GWA studies examining coronary artery disease (CAD) [214]. These studies found an unexpected risk association at the 9p21 locus, a locus with few protein coding genes. This locus showed the strongest genetic
correlation to CAD, with ~20-25% of the Caucasian population homozygous for the risk haplotype corresponding to a 30-40% increase in CAD development [215]. Interestingly, however, the sequence of a recently described IncRNA, ANRIL (antisense non-coding RNA in the INK4 locus), was found to overlap with this genomic region [216]. These discoveries put the spotlight on ANRIL as a new potential regulator of cardiovascular integrity. As such, many recent studies have shed light on the functions of ANRIL. ANRIL is expressed in endothelial cells, vascular smooth muscle cells, and leukocytes, all of which are involved in the pathogenesis of atherosclerosis [217]. One of the key findings from studies with ANRIL was its function as an epigenetic regulator of the INK4a/ARF/INK4b locus. This locus encodes three proteins, p15, p14ARF, and p16, all with important implications as cell cycle regulators, which together slow cell proliferation. Initiation and maintenance of INK4a/ARF/INK4b gene repression is mediated by PRC2 and PRC1, respectively. As demonstrated by Xist and HOTAIR, PRC complexes seem to have a common relation to IncRNAs. ANRIL is no exception, and interacts with the SUZ12 subunit of PRC2 and the chromobox 7 (CBX7) subunit of PRC1, recruiting them in cis to the INK4a/ARF/INK4b locus. Indeed, knockdowns of ANRIL show a decrease in CBX7 and a subsequent increase in INK4a/ARF/INK4b gene expression. Although the exact mechanism of ANRIL’s association to cardiovascular diseases has yet to be elucidated, the preliminary correlation has been made and further work may yet confirm ANRIL as the important regulatory molecule it is hypothesized to be.

Figure 5. ANRIL is expressed in tissues associated with atherosclerosis. Expression of ANRIL in leukocytes, VSMCs, and endothelial cells acts in cis to repress the neighboring INK4a/ARF/INK4b locus harboring important cell cycle regulators. Silencing occurs via the recruitment of PRC1 and PRC2 complexes and subsequent increase in the repressive H3K27me3 histone modification.

More recently, new studies have identified a number of IncRNAs associated with the cardiovascular system. Specifically, two IncRNA have emerged as potential regulators of murine cardiogenesis. One group of researchers has identified a novel IncRNA, deemed Braveheart
(Bvht), which plays a role in cardiac development [218]. Bvht is an ~590 nucleotide mouse-specific transcript that has been shown to be an essential regulator of cardiac commitment in embryonic stem cells. Knockdown of Bvht in mouse ES cells, which were allowed to differentiate into cardiomyocytes, showed a substantial decrease in the percent of contractile cells compared to the control. Although the mechanism is unclear, evidence suggests that Bvht functions upstream of the MesP1 pathway, a known regulator of cardiac mesoderm development [219]. Notably, Bvht, in agreement with an emerging theme among lncRNAs, was also shown to interact with PRC complexes [218]. Bvht interacts with the SUZ12 component of the PRC2 complex. This provides some insight into its function as a possible epigenetic regulator of cardiac gene expression, although it is not known whether this interaction acts to keep PRC2 from repressing genes needed for cardiogenesis, or if it functions to recruit the repressive complex to genes that inhibit this process.

Shortly after the discovery of Bvht, another group reported an lncRNA in mice, named Fendrr, which was similarly involved in the development of the heart [220]. Fendrr is expressed in the lateral mesoderm of the developing embryo, where it is required for differentiation into lateral mesoderm derivatives, most notably the cardiac mesoderm and ventral body wall. Accordingly, Fendrr null mutant embryos are embryonic lethal and display distinct phenotypic changes in the heart, such as cardiac hypoplasia and impaired function. Once again, epigenetic regulation by an lncRNA is the likely culprit for these outcomes. Fendrr null mutants show an increase in the activating H3K4me3 mark, laid down by the TrxG/MLL complex, at the promoters of the cardiac development regulators GATA6 and Nkx2-5. Co-immunoprecipitation assays also show an interaction between Fendrr and TrxG/MLL. Additionally, co-immunoprecipitation assays also show an interaction of Fendrr with TrxG/MLL and PRC2. Fendrr interacts with the PRC2 complex in both cis and trans to regulate transcription factors in the lateral and cardiac mesoderm. Fendrr is able to bind to a region near the neighbouring Foxf1 gene, where it is believed to anchor and guide PRC2 to the promoter region and increase H3K27me3.

The lncRNAs mentioned are likely only scratching the surface of important regulators of cardiovascular function. In fact, one recent study using high-throughput sequencing techniques to study VSMC in rats has identified 24 novel lncRNA differentially expressed in response to AngII [221]. Similarly, another study has shown, using high throughput RNA-seq techniques, that 135 non-coding genes show differential expression in failing murine hearts [222]. Together, these studies illustrate the complexity of the cardiovascular transcriptome. With thousands of putative lncRNAs, more studies will no doubt uncover similar associations to the cardiovascular system. Further insight into the function of these lncRNAs is crucial to realizing their contribution to epigenetics, and is a probable direction of many future studies.

4.9. Epigenetic inheritance

A vital component of the definition of epigenetics and one which has briefly been mentioned in this chapter is that of the heritability of epigenetic patterns. These patterns carry great significance for the determination and preservation of defined and consistent gene expression programs that underlie cell fate decisions during and beyond early development. Despite
flexibility of these patterns in totipotent and pluripotent stages, this plasticity wanes as development continues. The inheritance of such patterns is, therefore, of the utmost importance. The following section will explore the transmission of such information during the replication stages that take place in mitosis and meiosis, specifically analyzing DNA methylation and histone modification patterns within the context of these two processes.

DNA methylation is a significant epigenetic process that nicely fulfills the criterion for heritability. It has been proposed that DNA methylation may in fact act as a source of epigenetic memory, via the mitotic pathway. Previously, we offered a widely accepted and elegant sequence of events to describe the generational inheritance of DNA methylation patterns. De novo methyltransferases initially place methylation marks on DNA and, following replication, the maintenance DNMT marks the nascent strand with the same patterns exhibited by the original parent strand. Quite simply, DNA methylation absence demarcates active promoter regions, namely CpG islands, such that the promoters in which they are found remain potentially transcriptionally active throughout development and maturity. Conversely, segments of the genome that lack promoter activity become methylated and carry the flags of repression throughout the life of the organism. Our understanding of methylation pattern inheritance through mitosis is becoming relatively well elucidated, while more studies must be dedicated to comprehending the ways in which these patterns are initially laid down.

The semi-conservative model of inheritance that applies to both DNA replication and the transmission of DNA methylation patterns, had been speculated [223], and is now seldom conceived to be the main mode of transmission of histone modifications. Although compelling, this model relies on inherent symmetry within mono-nucleosomes as well as the semi-conservative separation of H3-H4 tetramers [224] (which, together with H2A and H2B dimers, make up the nucleosome core). Although two copies of each of the core histones are present within the nucleosome, it is unknown, yet highly doubted, that histone modifications exist in a symmetric manner [224]. Given the observation that the vast majority of H3-H4 tetramers actually segregate in a conservative manner [225-228], the more likely candidates to act as templates for new patterns would actually be adjacent, pre-existing nucleosomes on the same strand [224]. A possible mechanism for this may be envisioned whereby a neighbouring nucleosome that will be used as a template, is marked by a chromatin binding protein or reader protein [229] which would recruit a writer protein. Such a system may act in repetitive sequences where a number of nucleosomes all require the same replication. A self-cycling process has been proposed for up keeping repressive H3K27me3 marks, whereby polycomb repressive complex 2 (PRC2) binds to its own methylation site [230].

Based on these observations, the literal notion of copying a code during the S phase of the cell cycle does not seem to be applicable to the intricate system of histone modifications. In the next section, we mention that DNA methylation patterns likely influence the propagation of histone modifications. It may be speculated that the semi-conservative model applies well to DNA methylation pattern transmission, and the flags installed on the newly synthesized strand then act as markers for prompting the appropriate histone modifications.

Histone variant inheritance outside of S phase has suggested that histone modification inheritance can be replication-independent, unlike the events described above that would
predominate at the replication fork. Histone variants are altered versions of one of the four core histone proteins. For example, H3.3 has been shown to be associated with transcriptionally active segments and is found abundantly in active histone marks [231-233]. H3.3 nucleosomes are known to be more dynamic and susceptible to displacement during transcription [234]. Consequently, the density of H3.3 nucleosomes is reduced during transcription. The replication-independent re-incorporation of H3.3 appears to be mediated by chaperone proteins, such as HIRA (Hir-related protein A) following the completion of transcription [235, 236]. Hence, this particular epigenetic modification can be maintained, despite dilution by replication.

4.10. Histone code dispute

Occasionally, the “histone code” is discussed as an entity whose existence is well established, while, at other times, it proves to be a highly debated topic. In the following, we will briefly touch upon these disputable characteristics and why a consensus has yet to be reached in this matter. Numerous tables have been compiled in an attempt to categorize the diverse range of possible histone modifications and the assortment of amino acid residues that are subject to these alterations. In contrast to the genetic code, the universality and predictability of the potential histone code seem to be much less clear-cut. In stark contrast to the genetic code, the consistency of histone patterns seems to vary dramatically from lower to higher eukaryotes [237]. Evidence has accumulated which argues against the notion of a simple binary relationship between modification and effect.

Turner thoroughly reviewed the concept of a histone code, in light of the extensive possibilities that result from combinations of modification and amino acid, from the perspective of semiotics: the study of signs and their use or meaning [237]. The genetic code is highlighted as the quintessential semiotic code in the way that following the central dogma of molecular biology (DNA→RNA→protein) requires the utilization of symbols, their meaning, and a code to make known their meaning. Even though the symbols of an epigenetic code would likely be combinations of histone modifications with DNA methylation marks, the number of these is far from being known. Such arrangements would increase in quantity with increasing number of cell types and would have to be pinpointed at different moments in differentiation. At this point, we can only speculate as to the number of different outcomes based on the indefinite capacity for combination. Thus, it can be seen that, in order to accurately describe the transmission of histone modifications as a “code,” a much more comprehensive synthesis of the variable arrangements must be attained.

4.11. Relationships between epigenetic pathways

Several attempts have been made at identifying the relationships between the discussed epigenetic pathways. DNA methylation and histone posttranslational modifications have been implicated in a wide variety of instances, many of which have been discussed in this chapter. Biochemical and genetic evidence have tended towards a bidirectional relationship between the two systems: histone methylation has been seen to guide DNA methylation and a DNA methylation template likely influences the reversible placement of histone modifications [238].
In the following, both of these options, in addition to the joint roles they play in repression, will be discussed.

We can now appreciate the dynamic nature of methylation patterns during early development. At approximately the implantation stage, a wave of de novo methylation inputs the methylation patterns that will be transferred to somatic cells in subsequent generations. At the same time, there is a mechanism, speculated to be the combinatorial binding of cis-acting elements and active demethylation, which serves to protect CpG islands from indiscriminate de novo methylation (40, 239-241). Interestingly, there is support for the notion that it is the presence of H3K4 mono-, di-, or tri-methylation that appears to protect CpG islands before the round of de novo DNA methylation begins (63, 242). Recent studies suggest that H3K4 methylation protects DNA from de novo methylation [243, 244]. RNA polymerase II is known to be bound primarily to CpG islands in the early embryo and is responsible for recruiting H3K4 methyltransferases (Table 1) (63, 243-246). Therefore, it is these regions that remain unmethylated as a result of early histone modifications.

Another significant event of early development occurs at the time of gastrulation when targeted repression and de novo methylation of pluripotency preserving genes such as Oct3/4, takes place [247]. A sequential process ensues: binding of repressor molecules with the Oct3/4 promoter turns off transcription [248-250] and a complex containing the histone methyltransferase G9a as well as histone deacetylases are recruited. Deacetylation, one of the histone modifications associated with transcriptional repression, is the next step, resetting the lysine residues so that G9a is able to catalyze H3K9 methylation [251]. From here, the successive binding of heterochromatin protein 1 (HP1) facilitates the formation of heterochromatin [251]. Finally, G9a binding also recruits DNMT3A and 3B [252], meaning that the final stages of silencing are completed. Gastrulation is the time at which the embryo begins to separate into germ layers, losing the ability to maintain pluripotency. Accordingly, it is histone modifications, working in conjunction with DNA methylation pathways that direct the repression of the expression of pluripotency-associated genes.

In the above two examples, HP1 was discussed in the context of deacetylation, whereas H3K4 methylation was said to protect CpG islands from de novo DNA methylation in early development. There are other instances, however, where H3K4me acts as a binding site for HP1 [253, 254]. A major component of heterochromatin, HP1, is well known to contribute to the establishment and maintenance of transcriptionally silent heterochromatin [255]. In addition to these instances of histone modifications guiding DNA methylation patterns, there is evidence to suggest that, through cell division, DNA methylation is important for maintaining many histone modifications [251]. As discussed previously, DNA methylation patterns are faithfully transmitted to the next generation by means of the maintenance methyltransferase DNMT1 and prior UHRF1 and PCNA DNA binding. However, despite the importance of chromatin structure and composition in initiating transcriptional patterns, these same structures are liable to disruption as the replication fork and associated complexes make their way along the DNA during replication [251]. It has been inferred that DNA methylation patterns may act as an important indicator for the reconstruction of epigenetic patterns. That
Several studies have found that regions rich in unmethylated DNA are reformed in a rather open conformation, while those regions with more methylated DNA tend to be restructured in a sealed configuration [256, 257]. Acetylated (activating) histones are the primary component of those unmethylated, open regions. Conversely, it has also been shown that the nucleosomes of the compact, closed chromatin contain histones which are primarily non-acetylated [258, 259]. Also worth noting is the fact that methyl-CpG binding domain proteins such as MeCP2 and MBD2 also demonstrate the ability to attract histone deacetylases to methylated DNA regions [260, 261]. There is evidence that, in plants, DNA methylation may inhibit H3K4 methylation, a mark of activation, while perhaps also directing H3K9 dimethylation, a repressive indicator [259, 262, 263]. Thus, even though the mechanism of action may still be relatively unknown for a number of these hypotheses, evidence does suggest the potential of a DNA methylation template guiding the placement of histone modifications.

Earlier in this chapter, upon describing the structure of nucleosomes, we mentioned the fact that ATP-dependent chromatin remodeling complexes often execute an action of nucleosome sliding. We revisit this topic here, in light of our exploration of the relationships amongst epigenetic pathways. The regulation of gene expression has also been found to be affected by these chromatin remodelers, namely the SNF2 family. Containing helical domains but executing no helicase activity, these proteins disturb the histone/DNA contacts and are believed to function in both transcriptional activation and repression, depending on the context [238]. Briefly, we will highlight a few key avenues of research. Jeddeloh et al. were the first to link DNA methylation with these SNF2 helicases [264]. In A. thaliana, mutation in a protein called DDM1 (decreased in DNA methylation) resulted in 70% diminishment in global DNA methylation levels [264]. The mammalian lymphoid specific helicase (Lsh) has been found to be the most closely related SNF2 member and has been linked to cell proliferation [238, 265]. Murine knockout experiments have demonstrated normal development until birth, but death occurs soon after because of renal lesions, as well as defects in lymphoid development and proliferation [238, 266].

Based on the above examples, there certainly seems to be some sort of connectivity between the epigenetic pathways. Crosstalk is clearly happening between histone posttranslational modifications and DNA methylation patterns, despite the fact that they are often studied independently. Future research would certainly focus on deciphering both the direct and indirect ways in which these processes act in tandem with one another.

5. Case study II: Dutch winter famine

It is a rare and tragic instance when a serious famine strikes an urban and industrialized setting, but exactly this happened in the winter of 1944-45 in West Netherlands. Excellent documentation of the event and the people most affected by it has allowed for the study of nutritional deprivation, in the form of famine, on human reproduction. Under German occupation, the
Western Netherlands suffered an unusually harsh winter, the effects of which were exacerbated by fuel shortages. Homes were not heated, grass and tulip bulbs were consumed, and every scrap of furniture was burned in a desperate effort to stay alive [267]. Food rationing has become an inevitably direct consequence of wartime events, and the Dutch Winter Famine was no exception to this phenomenon. A 1,800 calorie limit had been upheld for four years before mid-1944 when it was reduced to 1,600. A stark decline placed the limit at 400 calories in April of 1945 [267]. Although children as well as expectant and nursing mothers did receive slightly greater quantities, essentially every person suffered from severe hunger [268].

Case studies dedicated to investigating the impact of malnourished mothers on their offspring came about as a result of the initial Dutch Famine Birth Cohort Study. There are 2414 singletons born between November of 1943 and February of 1947 in Amsterdam for whom detailed birth records were collected and who, at the ages of 50 to 58, were surveyed as to their physical and mental health [268]. In general, low birth weight has been correlated with a number of conditions later in life: hypertension, insulin resistance, and obesity [268-271]. A decline in birth weight of up to 300g was seen among those exposed to maternal malnutrition in the third trimester [272]. The results of the Cohort study clearly demonstrated that exposure to famine at any point of gestation was associated with glucose intolerance [273]. Interestingly, the first trimester proved to be the most decisive stage of gestation at which maternal malnutrition impacts offspring health. Increased rates of coronary artery disease, atherogenic lipid accumulation, disturbed blood coagulation, type II diabetes, and cardiovascular disease were exhibited by those exposed to famine in early gestation. This Cohort Study clearly demonstrated that, not only is maternal famine exposure during gestation associated with chronic disease in offspring later life, but that effects varied based on the timing in gestation of famine exposure.

How can the solitary existence of the static DNA code account for such remarkable differences witnessed over the course of a single generation? Changes outside of the genome are needed to explain this trans-generational variation. To address this issue in more detail, Heijmans et al. focused on insulin-like growth factor 2 (IGF2), a maternally imprinted gene and a crucial factor in human growth and development, especially during gestation [274]. To investigate whether maternal famine exposure was associated with differences in methylation of IGF2, 60 periconceptionally exposed people were compared with their same-sex sibling. Overall, exposure was found to be associated with an average reduction of 5.2% methylation in the DMR of IGF2. All CpG sites, except for one, were seen to be less methylated in those who had been exposed [274]. Statistically significant data showed periconceptional exposure, but not later exposure, was associated with this decrease. It is hypothesized that the deficiency of methyl donors such as the amino acid methionine, from reduced caloric intake, may account for the hypomethylation of IGF2 [274]. This study was the first to record persistent epigenetic changes in response to transient environmental conditions.
6. Concluding remarks

In this chapter, we have aimed to excite readers about the emerging field of epigenetics and the potential role it will play in future research initiatives. Without underscoring the importance of the genetic code, we have attempted to describe epigenetics in such a way that its significance is comparatively equivalent to that of genetics. The birth of the current molecular biological paradigm occurred relatively recently in the history of science, especially the history of scientific revolutions. Whether or not epigenetics is paving the way for a new paradigm in biology, is a debate that is best left to the philosophers of science. Yet, the relevancy of epigenetics cannot be understated as we embark on future investigations into human health and disease.

Known histone posttranslational modifications were explored first, and the repressive, activating, or joint consequences of a majority of alterations were delineated. Due to its duration of study, great emphasis was placed on DNA methylation, its consequences, mediators, and mechanisms of action. Then, we introduced the re-discovery of the promising 6th base pair: 5hmC. Although its role remains highly elusive, a great deal of research has been devoted to understanding the demethylation pathways in which this base is likely involved. Many of these options and their respective caveats were explored in the text. The observations that have been made about TET enzyme, DNA methylation, and DNA hydroxymethylation levels in zygotic development and PGC formation were also elucidated in this chapter and concisely summarized in Figure 3. Long non coding RNAs were discussed in their broad context as well as the role they play in the cardiovascular system.

Connections across the entire scope of the chapter were highlighted in our exploration of vitamin C. Its fascinating role in the human body cannot be understated. In scurvy, this vitamin’s deficiency results in devastating physiological and psychological outcomes. Acting as both a co-factor for enzymatic reactions and as an antioxidant, the biochemical activity of ascorbate is also relevant in the context of cell culture and demethylation. Discovering the connections between P4H and TET enzyme structure has allowed the multi-faceted role of vitamin C in the epigenome, to also be elucidated. The pathogenesis of atherosclerosis was mechanized in such a way that the potential role of increased antioxidant intake was highlighted. Epigenetic regulation in the endothelium proves to be dynamic and highly changeable in the case of cardiovascular disease. eNOS gene regulation was thoroughly investigated as a representative case of highly endothelial-cell specific expression. Epigenetic inheritance at the replication fork was addressed in a DNA methylation and histone modification context. Crosstalk between these two pathways was discussed in the context of both DNA methylation patterns impacting those of histone modifications and vice versa. Finally, the Dutch Winter Famine was our second case study, this time further evaluating the role of transgenerational epigenetic inheritance in the context of severe maternal nutritional deprivation during pregnancy.
## List of non-standard Abbreviations and Acronyms

<table>
<thead>
<tr>
<th>Abbreviation</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>5caC</td>
<td>5-carboxylcytosine</td>
</tr>
<tr>
<td>5hmC</td>
<td>5-hydroxymethylcytosine</td>
</tr>
<tr>
<td>5fC</td>
<td>5-formylcytosine</td>
</tr>
<tr>
<td>5mC</td>
<td>5-methylcytosine</td>
</tr>
<tr>
<td>ATP</td>
<td>adenosine triphosphate</td>
</tr>
<tr>
<td>AID</td>
<td>activation-induced cytidine deaminase</td>
</tr>
<tr>
<td>APOBEC</td>
<td>apolipoprotein B mRNA editing enzymes, catalytic polypeptide</td>
</tr>
<tr>
<td>BER</td>
<td>base excision repair</td>
</tr>
<tr>
<td>ChiP</td>
<td>chromatin immunoprecipitation</td>
</tr>
<tr>
<td>CD domain</td>
<td>Cys-rich and double stranded-helix DSBH region</td>
</tr>
<tr>
<td>CpG</td>
<td>cytosine - phosphate - guanine</td>
</tr>
<tr>
<td>CGIs</td>
<td>CpG islands</td>
</tr>
<tr>
<td>DPPA3</td>
<td>developmental pluripotency-associated 3 protein</td>
</tr>
<tr>
<td>DNMTs</td>
<td>DNA methyltransferases</td>
</tr>
<tr>
<td>DMR</td>
<td>differentially methylated region</td>
</tr>
<tr>
<td>ES cell</td>
<td>embryonic stem cell</td>
</tr>
<tr>
<td>EC</td>
<td>endothelial cell</td>
</tr>
<tr>
<td>eNOS</td>
<td>endothelial nitric oxide synthase</td>
</tr>
<tr>
<td>EDRF</td>
<td>endothelium-derived relaxing factor</td>
</tr>
<tr>
<td>ES cell</td>
<td>embryonic stem cell</td>
</tr>
<tr>
<td>GWAS</td>
<td>genome-wide association studies</td>
</tr>
<tr>
<td>HATs</td>
<td>histone acetyltransferases</td>
</tr>
<tr>
<td>HDACs</td>
<td>histone deacetylases</td>
</tr>
<tr>
<td>HDMTs</td>
<td>histone demethylases</td>
</tr>
<tr>
<td>HOXC</td>
<td>homeobox X</td>
</tr>
<tr>
<td>IGF2</td>
<td>insulin-like growth factor</td>
</tr>
<tr>
<td>ICAM-2</td>
<td>intercellular cell adhesion molecule-2</td>
</tr>
<tr>
<td>IAPs</td>
<td>intracisternal A particles</td>
</tr>
<tr>
<td>Inc-RNA</td>
<td>long non-coding RNA</td>
</tr>
<tr>
<td>MBDs</td>
<td>methyl CpG binding domain proteins</td>
</tr>
<tr>
<td>miRNA</td>
<td>micro RNA</td>
</tr>
<tr>
<td>MEFs</td>
<td>mouse embryonic fibroblasts</td>
</tr>
<tr>
<td>NO</td>
<td>nitric oxide</td>
</tr>
<tr>
<td>PRC2</td>
<td>polycomb repressive complex 2</td>
</tr>
<tr>
<td>PGCs</td>
<td>primordial germ cells</td>
</tr>
</tbody>
</table>
Acknowledgements

We gratefully thank members of the Marsden lab for the critical review of this manuscript. P.A. Marsden holds the Keenan Chair in Medical Research and is a recipient of a Career Investigator Award from the Heart and Stroke Foundation of Canada and is supported by a grant from the Canadian Institutes of Health Research (CIHR MOP 79475) and a grant from the National Institutes of Health (NIH/HHLBI P01 HL076540-06A1).

Author details

Alexandra A. Majerski1, Anthony C. Quinton1 and Philip A. Marsden1,2*

*Address all correspondence to: p.marsden@utoronto.ca

1 The Keenan Research Centre and Li Ka Shing Knowledge Institute, St. Michael’s Hospital, Toronto, Canada

2 Department of Laboratory Medicine and Pathobiology, University of Toronto, Toronto, Canada

The authors have no conflicts of interest to disclose.
References


