Author Manuscript

Author Manuscript

Author Manuscript

Author Manuscript

HHS Public Access

Author manuscript

Immunol Rev. Author manuscript; available in PMC 2019 August 13.

Published in final edited form as:

Immunol Rev. 2015 January ; 263(1): 36–49. doi:10.1111/imr.12242.

New themes in the biological functions of 5-methylcytosine and 5-hydroxymethylcytosine

Erika L. Moen[1,2], **Christopher J. Mariani**[1,3], **Hayley Zullow**[1], **Meselle Jeff-Eke**[1], **Edward Litwin**[1], **John N. Nikitas**[1], **Lucy A. Godley**[1,2,4]

[1]Section of Hematology/Oncology, Department of Medicine, The University of Chicago, Chicago, IL, USA.

[2]Committee on Cancer Biology, The University of Chicago, Chicago, IL, USA.

[3]Committee on Molecular Pathogenesis and Molecular Medicine, The University of Chicago, Chicago, IL, USA.

[4]Comprehensive Cancer Center, The University of Chicago, Chicago, IL, USA.

Summary

5-methylcytosine (5-mC) and 5-hydroxymethylcytosine (5-hmC) play a critical role in development and normal physiology. Alterations in 5-mC and 5-hmC patterns are common events in hematopoietic neoplasms. In this review, we begin by emphasizing the importance of 5-mC, 5-hmC, and their enzymatic modifiers in hematological malignancies. Then, we discuss the functions of 5-mC and 5-hmC at distinct genic contexts, including promoter regions, gene bodies, intron-exon boundaries, alternative promoters, and intragenic microRNAs. Recent advances in technology have allowed for the study of 5-mC and 5-hmC independently and specifically permitting distinction between the bases that show them to have transcriptional effects that vary by their location relative to gene structure. We extend these observations to their functions at enhancers and transcription factor binding sites. We discuss dietary influences on 5-mC and 5-hmC levels and summarize the literature on the effects of folate and vitamin C on 5-mC and 5-hmC, respectively. Finally, we discuss how these new themes in the functions of 5-mC and 5-hmC will likely influence the broader research field of epigenetics.

Keywords

5-methylcytosine; 5-hydroxymethylcytosine; epigenetics; transcriptional regulation

Introduction

Since the discovery of the presence of cytosine methylation in mammalian genomes, an abundant amount of research has demonstrated its importance in normal physiology and disease. 5-methylcytosine (5-mC) plays a critical role in genomic imprinting, X-

Correspondence to: *Lucy A. Godley*, Section of Hematology/Oncology, Department of Medicine, The University of Chicago, 5841 S. Maryland Ave., MC 2115, Chicago, IL 60637, USA, Tel.: +1 773 702 4140, Fax: +1 773 702 9268, lgodley@medicine.bsd.uchicago.edu.

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

chromosome inactivation, and tissue-specific gene expression patterns (1–3). DNA methylation is catalyzed through an enzymatic reaction that requires the DNA methyltransferases (DNMTs) and S-adenosyl methionine (SAM) as the methyl group donor. In somatic tissues, DNA methylation occurs almost exclusively at the C5 position of a cytosine in a 5′-CpG-3′ dinucleotide. The maintenance DNA methyltransferase DNMT1 specifically binds hemimethylated DNA during cell division to ensure the methylated cytosine is inherited on the daughter strand, which contains the palindromic CpG sequence. The two *de novo* DNMTs, DNMT3A and DNMT3B, are required for establishing new DNA methylation patterns and have been shown to contribute to maintaining DNA methylation marks along with DNMT1 (4). DNA methylation is associated with heterochromatin, which is facilitated in part by its interaction with proteins that specifically bind 5-mC, termed methyl-binding proteins (MBPs).

5-mC can be oxidized to 5-hydroxymethylcytosine (5-hmC) by the ten-eleven translocation (TET) family of proteins that use oxygen, Fe(II), and α-ketoglutarate as substrates in the enzymatic reaction (5, 6). The TETs can continue to oxidize 5-hmC to 5-formylcytosine (5-fC) and 5-carboxylcytosine (5-caC) (Fig. 1). These reactions are believed to play a role in active demethylation of cytosines and function to facilitate transcription (7, 8). For many years, 5-mC was thought to be a stable, irreversible epigenetic mark, so the discovery of this active demethylation pathway suggests that DNA methylation patterns may be more susceptible to variation than previously appreciated. Among these covalently modified bases, 5-hmC is more abundant and stable than 5-fC and 5-caC, and 5-hmC may have its own biological function(s) distinct from 5-mC in addition to its role in active demethylation. In this review, we emphasize the importance of 5-mC and 5-hmC and their respective enzymatic modifiers in hematological malignancies briefly. For more detail, please see these additional excellent recent reviews (9–12). Then, we focus on the emerging research related to the functions of cytosine modifications, primarily 5-mC and 5-hmC, in transcriptional regulation and discuss external factors that can influence the levels and distributions of these cytosine modifications in the genome. Finally, we highlight promising avenues for future research of the roles of 5-mC and 5-hmC in normal physiology and disease.

5-methylcytosine in hematological malignancies

The DNMTs play a critical role in cellular differentiation. DNMT3A is required for hematopoietic stem cell (HSC) differentiation, and loss of *DNMT3A* results in both increased and decreased methylation at individual loci (13). This aberrant methylation profile contributes to upregulation of HSC multipotency genes and downregulation of differentiation factors (13). Loss of differentiation markers is a common event in hematological malignancies, and therefore several groups have studied the roles of the DNMTs in initiation and progression of these diseases (14–19). Interestingly, exome sequencing and whole genome sequencing studies of leukemias have shown that several of the recurrent mutations identified are in genes that encode epigenetic modifiers (12). Somatic heterozygous mutations of *DNMT3A* have been found to predict a shorter overall survival in patients with acute myeloid leukemias (AML) (14, 15, 20). In fact, *DNMT3A* mutations are one of the most frequently mutated genes in AML and occur in up to 36% of cytogenetically normal AML patients (19). Another study investigating levels of the DNMTs

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

in normal hematopoietic cells and leukemias found that *DNMT3B* was highly expressed in $CD34^+$ bone marrow cells and leukemia cells and expressed to lower levels in differentiated cells, suggesting it may be promoting an immature DNA methylation phenotype in the leukemia cells (16). Dysregulation of the *de novo* DNMTs in leukemia can also contribute to disease pathogenesis by hypermethylating and silencing tumor suppressor genes (17, 18).

Drugs that target DNA methylation have been studied clinically for the treatment of hematological malignancies. Hypomethylating agents, such as 5-aza-2′-deoxycytidine (decitabine) and 5-azacytidine (azacitidine), are FDA-approved for the treatment of myelodysplastic syndromes (MDS), and are used clinically to treat MDS and AML (9, 21–23). The precise mechanisms of action of these drugs are not fully understood, but they are known to incorporate into DNA as a cytosine analog. Decitabine and azacitidine have an identical ring structure, but decitabine contains a deoxyribose sugar, whereas azacitidine contains a ribose sugar. Consequently, decitabine is incorporated exclusively into DNA, and azacitidine is incorporated mainly into RNA, with approximately 5–10% being converted to decitabine and incorporated into DNA. Although decitabine and azacitidine both show an improved response rate among patients with MDS (30–60% and 40–60%, respectively) compared to standard chemotherapy treatment, azacitidine showed a more promising improvement in overall survival (23–25). The clinical benefit seen in patients is thought to be in part due to reactivating tumor suppressor genes through regional DNA demethylation, but as mentioned previously, the exact mechanism of action is not fully defined (26). Collectively, researchers in the field have demonstrated that the normal distribution of methylated cytosines is disrupted in hematological diseases, partially due to the impaired function of the DNMTs (26).

5-hydroxymethylcytosine in hematological malignancies

The three TET proteins, TET1, TET2, and TET3, are all capable of oxidizing 5-mC, but they are believed to function in distinct biological contexts. For instance, murine Tet1 has been shown to regulate mouse embryonic stem cell (ESC) pluripotency, whereas both Tet1 and Tet2 regulate cell line-age specification (27). In human ESCs, levels of *TET*2 expression were found to be low, and *TET*2 levels increased during hematopoietic differentiation (28). TET2 was demonstrated to promote hematopoietic differentiation through regulation of 5-hmC at the *NANOG* promoter (28). Tet3 is believed to be critical in the process of fertilization, given that murine Tet3 mediates genome-wide oxidation of 5-methylcytosine of the zygotic paternal DNA, finally clarifying the mechanism of rapid demethylation of the paternal genome that occurs immediately following fertilization (29–31).

Similar to DNA methylation and DNMTs, alterations in 5-hmC and TET enzymatic activity have been observed in hematological malignancies (10). *TET*2 was found to be mutated in approximately 15% of patients with various myeloid malignancies (32), and AML patients with intermediate-risk cytogenetics whose cancers have *TET*2 mutations have an unfavorable prognosis (33). Functional studies have shown that loss of *Tet2* in mice resulted in increased HSC self-renewal, myeloid transformation, and pleiotropic hematopoietic abnormalities (34–40).

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

As described above, the TET enzymes use α-ketoglutarate (also known as 2-oxoglutarate) as a cofactor in the conversion of 5-mC to 5-hmC. The isocitrate dehydrogenase (IDH) enzymes catalyze the enzymatic conversion of isocitrate to α-ketoglutarate. Mutations in *IDH*1 and *IDH*2 antagonize TET-mediated oxidation of 5-mC through aberrant conversion of isocitrate to 2-hydroxyglutarate, which acts as an oncometabolite by interfering with α-ketoglutarate-dependent enzymes, including the TETs. *IDH1*/2 mutations can therefore decrease normal 5-hmC levels and are associated with poorer prognosis of certain subsets of cytogenetically normal AML (41). A novel inhibitor of mutant IDH2 developed by Agios, AGI-6780, was shown to induce differentiation of primary human AML cells *in vitro* (42). Preliminary data from an ongoing phase I clinical trial of AGI-6780 in patients with relapsed or refractory AML or MDS show clinical promise, with 7 of 10 patients achieving complete remission at the time this study was presented in April 2014 (43). A full appreciation of the distinct functions of the epigenetic modifiers and the effects they have on cytosine modifications and transcription will improve our understanding of how alterations in normal 5-mC and 5-hmC patterning can contribute to disease pathogenesis.

Cytosine modifications at distinct genic contexts

Transcriptional start sites

Promoter regions of coding genes are usually unmethylated in somatic cells to support an open chromatin state and accessibility to transcription factors (Fig. 2A). Approximately 70% of gene promoters contain a CpG island (44), and CpG islands in the promoter region of active genes are often nucleosome-depleted, associated with histone 3 lysine 4 (H3K4) tri-methylation, and flanked by the histone variant H2A.Z (45). Hypermethylation of CpG islands in promoter regions is associated with repression of gene transcription, and this is known to play an important role in many cellular and physiological contexts, including development. For instance, imprinted genes and stably repressed genes on the inactive X chromosome have highly methylated promoter regions (1, 2). Methylation of CpGs adjacent to the CpG island, termed ‘CpG shores’, are also inversely correlated with gene expression and correspond to many tissue-specific differentially methylated regions (DMRs), demonstrating that variation of methylation in regions adjacent to CpG islands is evident and may in fact be functionally important (46). The causal link between DNA methylation and gene silencing is complex, and in some cases, such as the *Hprt* gene on the inactive X chromosome, chromosome inactivation precedes promoter methylation (47). It has been difficult to discern the exact timing of *de novo* methylation and gene silencing, but it has been shown that *de novo* methylation of promoter regions may require nucleosome recruitment and additional repressive histone modifications to silence the locus effectively, suggesting that DNA methylation is likely working with other chromatin modifiers to regulate transcription (48, 49).

Another potential confounding factor in understanding the function of 5-mC in regulating transcription at promoter regions is that, until recently, techniques used to study cytosine methylation were unable to distinguish between 5-mC and 5-hmC, and these two cytosine modifications may have different effects on transcription potential. A study investigating the role of 5-hmC at promoter regions found that, in contrast with 5-mC, 5-hmC was found to be

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

enriched in TSSs with intermediate to high CpG density (50). The balance between hydroxymethylation and methylation at gene promoters is likely critical for fine-tuning gene expression patterns (Fig. 2A and B). Indeed, a decline of the TET enzyme levels during differentiation was associated with decreased hydroxymethylation levels at promoter regions of genes specifically expressed in ESCs. This loss of hydroxymethylation corresponded with increased methylation and gene silencing (51). TET-mediated oxidation of 5-mC to unmodified cytosine could be an important contributing factor to demethylation of promoter regions, as TET1 has been shown to actively demethylate promoter regions in the human brain (52). However, another study showed that knockdown of TET1 resulted in upregulation and downregulation of genes, suggesting that TET1 and 5-hmC may promote or inhibit transcription, depending on the context. In fact, TET1 was found to bind a significant proportion of Polycomb group target genes and associated with the SIN3A co-repressor complex, further indicating a role in transcriptional repression (50). In support of this, it has been observed that the presence of 5-hmC in the promoter, but not the gene body, has a repressive effect on *in vitro* transcription, potentially due to its inhibition of essential transcription factors (53). More work that distinguishes between these two modifications will be paramount for our understanding of the effects they have on regulating transcription potential at gene promoters.

Gene bodies

Over a decade ago, high cytosine methylation levels were observed in exons, but the function was unknown (54). It was only more recently that we learned that a small portion of those 5-mC bases is likely 5-hmC, because both bases are protected from sodium bisulfite conversion (55, 56). In the following sections, we use the term ‘DNA methylation’ to discuss the findings of these studies, with the acknowledgement of the caveat that most techniques used sodium bisulfite conversion and therefore did not distinguish between 5-mC and 5-hmC. We also discuss the more recent work that looks specifically at 5-hmC to uncover its biological function at different intragenic contexts.

The implications of intragenic gene body methylation have not been studied as extensively as promoter methylation, but several studies have observed a positive correlation between gene body methylation and gene expression levels (57–59). There is evidence of a bi-directional relationship between gene body methylation and transcription, where each can promote the other, as one study demonstrated that intragenic nucleosomes and H3K36me3, a histone mark associated with transcriptional elongation, recruited DNMTs to facilitate methylation of intragenic DNA (60). It has also been observed that DNA methylation interacts with nucleosome positioning, and nucleosomal DNA is more highly methylated than flanking DNA (61). Furthermore, the H2A.B histone variant is associated with highly methylated intragenic DNA, and these regions are enriched in actively transcribed loci (62).

It is possible that intragenic DNA methylation is more critical for transcription of some genes compared with others. For instance, our group has shown that gene body DNA methylation levels of *MGMT* were predictive of expression in glioblastoma patients whose tumors had an unmethylated *MGMT* promoter, suggesting that DNA methylation may contribute to variation in expression levels of genes with an unmethylated promoter region

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

(63). In addition, we demonstrated that decitabine-induced hypomethylation across the gene body corresponded with a decrease in gene expression (63). Decitabine has been shown to preferentially hypomethylate CpGs that are not in CpG islands (64), and its effects on intragenic CpGs and corresponding expression changes warrants additional study.

The contribution of intragenic 5-hmC to regulating gene expression has been gaining attention, and similar to promoter regions, it is likely that there is a balance between 5-mC and 5-hmC along gene bodies that regulates transcription. 5-hmC has been found to be enriched at intragenic regions, particularly at the 3′ end of actively transcribed genes that are TET1 targets (65). The brain is a tissue with particularly high levels of 5-hmC, suggesting that this base may be particularly important in the brain (66). A study of neuronal cells has shown that the ratio of 5-hmC to 5-mC predicts expression better than either mark on its own (67). Interestingly, genes with the highest intragenic CpG density are enriched with 5-hmC, and levels of 5-hmC in gene bodies have been shown to increase during neuronal differentiation, without a significant change in 5-mC levels (68). Like 5-mC, there is evidence that 5-hmC along gene bodies is also positively correlated with gene expression (69–71). In mature olfactory sensory neurons, 5-hmC levels across the gene body correlated with gene expression (69). In addition, overexpression of *TET3* in these cells significantly altered gene body 5-hmC levels and gene expression in a manner consistent with a positive role for 5-hmC in transcription (69). Therefore, the authors of this study suggest a role for 5-hmC in maintenance of cellular identity independent of its function as an intermediate to demethylation (69).

Intron-exon boundaries

A research area that has been gaining traction over the past several years has been the role of intragenic cytosine modifications in regulating mRNA processing decisions. In 2010, it was observed that exons are more highly methylated than introns, and that exonic and intronic DNA may consist of distinct chromatin features (72). An example of how transcript splicing may partially depend on differentially methylated intragenic boundaries is demonstrated in the *CD45* mammalian model system for alternative splicing, which is regulated by the interaction between DNA methylation and CTCF (73). CTCF promotes inclusion of weak upstream exons by mediating local RNA polymerase II (Pol II) pausing, and DNA methylation inhibits CTCF binding and upstream exon inclusion at the *CD45* locus (73). Exonic DNA methylation may also cooperate with nucleosomes and H3K36me3 to ensure appropriate splicing of transcripts with different expression levels (74). This phenomenon appears to be conserved across species, as an investigation of DNA methylation profiles among honey bees found a strong correlation between methylation patterns and splicing sites, including alternative exons (75). Furthermore, the correlation of DMRs in honey bees with the frequency of alternatively spliced exons suggests that methylation likely plays a role in selecting which exons are included in mature transcripts (76). It has also been shown that greater than 14% of alternatively spliced genes were associated with a tissue-specific DMR in mouse retina and brain, and that these genes were enriched for developmental genes (77). Another study using chemical labeling of 5-hmC and 5-mC reported that there is substantial tissue specificity of these cytosine modifications at exon-intron boundaries in human and mouse, and constitutive exons contain higher 5-hmC levels relative to

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

alternatively spliced exons (78). In addition, 5-hmC was reported to be enriched at the boundaries of exons that are highly expressed in mouse ESCs, and Tet2 depletion resulted in significant loss of 5-hmC at these regions (79). Furthermore, single base resolution of 5-hmC and 5-mC using whole genome TET-assisted bisulfite sequencing (TAB-seq) demonstrated that 5-hmC peaks, but not 5-mC peaks, are enriched at the 5′ splicing sites at the exon-intron boundaries in the brain, suggesting a mechanistic link between 5-hmC and splicing (80). This study also reported a transcription-correlated 5-hmC bias toward the sense strand and a 5-mC bias toward the antisense strand of gene bodies (80). Fig. 2A depicts the distributions of 5-mC and 5-hmC at exon-intron boundaries described above.

Alternative promoters

A significant proportion of transcripts are transcribed from alternative promoters, and thus epigenetic regulation of these promoters is likely a common occurrence (81). Interestingly, methylation of alternative promoters within genes has been shown to regulate alternative transcripts at a tissue specific level (82). In the brain, approximately one-third of all intragenic CpG islands are methylated and show a large degree of tissue-specific methylation (82). The remaining intragenic CpG islands that are unmethylated likely function as alternative promoters, because they overlap with H3K4me3, a histone modification that marks active promoters (82). This evidence, along with the role of cytosine modifications in regulating splicing, suggests that 5-mC and 5-hmC are critical components in determining abundance of alternative transcripts.

DNA methylation and non-coding RNAs

Another function of gene body DNA methylation is the regulation of non-coding RNAs located in the intronic regions of genes. DNA methylation has in fact been shown to regulate several intragenic miRNAs (83, 84). Inhibition of DNA methylation via genetic deletion of DNMT1 or treatment with a hypomethylating agent can reactivate several miRNAs, suggesting that methylation of miRNA promoter regions is a mechanism of miRNA silencing (85–87).

Intriguingly, non-coding RNAs may also be regulating DNA methylation levels of nearby genes. A recent study investigated the involvement of a non-coding RNA originating within the *CEBPA* locus in regulating *CEBPA* methylation. The non-coding RNA was found to interact with DNMT1, resulting in prevention of *CEBPA* gene methylation and an increase in *CEBPA* mRNA production (88). In another example, a non-coding RNA species known as pRNA (promoter-associated RNA), which was complementary to the rDNA promoter, mediated *de novo* DNA methylation of rRNA genes (89). The authors of this study describe a phenomenon where pRNA formed a DNA:RNA triplex which acted as a binding platform for DNMT3B, resulting in DNMT3B recruitment to the rDNA genes (89).

DNA:RNA structures can also act to protect gene promoters from DNA methylation. For example, DNA:RNA structures known as ‘R loops’ can form at gene promoters that have significant strand asymmetry in the distribution of guanines and cytosines (GC skew) immediately downstream of the transcription start site (90). R loop formation during transcription was shown to protect the CpG island within the promoter region from

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

DNMT3B, suggesting that these R loop structures can affect the epigenetic status of genes in cis (90). These interactions between a non-coding RNAs and DNMTs are distinct from the studies that show microRNAs target the DNMTs for degradation (91, 92), and add a layer of complexity to the interplay between DNA methylation and non-coding RNAs.

Non-CG methylation

Although the majority of cytosine methylation occurs on CpG dinucleotides, the presence of non-CG methylation has been discovered in human and mouse ESCs and human spleen cells with the implementation of whole genome bisulfite sequencing (72, 93, 94). Laurent *et al.* (72) demonstrated that human embryonic stem cells (hESCs) had the highest level of non-CG methylation compared with fibroblast-like cells differentiated from hESCs, primary neonatal foreskin fibroblasts, and peripheral monocytes. Of the non-CG methylation events, methylation of CA dinucleotides was more common than CT or CC, and both CG and non-CG methylation decreased during differentiation (72, 95). Non-CG methylation has been observed in mouse germinal vesicle oocytes at maternally methylated DMRs in addition to non-DMRs (96). However, the function of these epigenetic marks is not well understood. A more recent study demonstrated that non-CG methylation in mouse germinal vesicle oocytes is dependent on Dnmt3a and Dnmt3L, which form a tetramer complex to act as a *de novo* DNA methyltransferase (95, 97). The pattern of non-CG methylation is concordant with CG methylation, with low levels in the promoter regions of genes and higher levels across gene bodies (95). As mentioned above, somatic mutations in *DNMT3A* are known to be a poor prognostic marker in AML, but the effects of mutant DNMT3A on DNA methylation patterns are inconclusive (98). One group reported similar global levels of 5-methylcytosine in *DNMT3A* wildtype and mutant AML samples (99). Given the implication that DNMT3A is required for non-CG methylation, it may be interesting to study the effects of the *DNMT3A* mutations on non-CG methylation and whether those epigenetic modifications are deregulated in cancers.

Cytosine modifications at gene regulatory elements

Enhancers

Although the core promoter region around transcribed genes is sufficient to assemble the Pol II machinery, transcriptional activity is often regulated further by distant sites known as enhancers (100) (Fig. 3). Chromosome conformation capture techniques have allowed researchers to map interactions between genomic elements that are in close spatial proximity but are tens to hundreds of kilobases apart (101). These physical interactions are critical for distal regulatory elements such as enhancers. Several studies have identified a role for CTCF and cohesin in establishing DNA loops and enhancer function (102–105). Cohesin does not bind DNA directly, but it can regulate tissue-specific expression by stabilizing the interaction between transcription factors and enhancer elements (106). Interestingly, somatic mutations in components of the cohesin complex, including *STAG1, STAG2, SMC1A, SMC3*, and *RAD21* have been identified in patients with myeloid neoplasms, resulting in a loss of cohesin binding sites on chromatin (107–110). It is thought that these mutations may impair the cohesin complex and its regulation of gene transcription, which may contribute to development of the myeloid malignancies (109). Enhancers are frequently nucleosome

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

depleted and are therefore DNase hypersensitive. They can bind both activating and repressive regulators and are associated with characteristic histone marks (e.g. H3K4me1 and H3K27ac) (100). The importance of methylation in regulating enhancer function is demonstrated by the fact that enhancers with methylation-dependent activity display differential methylation levels even between closely related cell types (111). Other evidence that epigenetic modifications regulate enhancer function is found in experiments examining the response to glucocorticoid receptor (GR) stimulation, which is heterogeneous among different cell types. GR binding to enhancers is restricted to regions of open chromatin that either pre-exist before GR stimulation or are induced upon GR stimulation in a manner that involves cytosine demethylation. Epigenetic regulation of GR binding to enhancers accounts for cell-type specific responses (112).

In mouse ESCs and neuronal progenitor cells, genome-wide profiling of methylation identified a subpopulation of genomic regions as low methylation regions (LMRs) displaying an average methylation level of 30%. These regions are CpG poor, DNase hypersensitive, colocalize with distinct chromatin marks, and augment transcriptional activity in reporter assays, which led to classification of these LMRs as enhancer regions (93). Enrichment of Tet1 at these positions (93), combined with numerous reports that 5-hmC is enriched at enhancers (51, 65, 113–121) supports the notion that Tet-mediated hydroxymethylation is important in maintaining low methylation levels at these positions (Fig. 3). Given that the DNA-binding proteins (DBP) REST and CTCF are necessary and sufficient to establish LMRs (93), this observation suggests that REST and CTCF direct TET dioxygenases to LMRs to protect them from hypermethylation. This hypothesis was tested by Feldmann *et al.* (116), who focused specifically on the function of REST, since CTCF deletion is cytotoxic to ESCs. REST deletion resulted in 5-hmC loss and 5-mC accumulation at LMRs, supporting a model where DBPs such as REST and CTCF recruit Tet enzymes to protect these regions from methylation. This is consistent with findings that DBPs, such as PRDM14, direct Tet enzymes to specific loci to regulate transcription through cytosine demethylation (122).

Transcription factor binding sites

Association of 5-hmC with tissue specific transcription factor binding sites has been demonstrated in both static and dynamic systems. In mouse ESCs, 5-hmC is enriched at enhancers with binding sites for pluripotency transcription factors, such as Nanog, Sox2, and Oct4 (65, 114, 123). In differentiating neural cells and adipocytes, there are dynamic changes of 5-hmC at enhancers regions. In these differentiation studies, 5-hmC is lost at enhancers with binding sites for pluripotency-related transcription factors, whereas enhancers with binding sites for lineage specific transcription factors, such as Meis1 and PPARγ, gain 5-hmC in differentiation. Moreover, gain of 5-hmC is associated with other chromatin changes favoring a more accessible state, such as gain of H3K27ac (117).

Our group also found that DNA loci that gain 5-hmC density during erythropoiesis were associated with activating histone marks and numerous transcription factors known to be important for erythropoiesis, such as GATA1, GATA2, KLF1, STAT5A, and STAT1 (70). Combined with multiple observations that 5-hmC is enriched at binding sites of

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

pluripotency-related transcription factors in ESCs (described above), these observations suggest a model where multiple transcription factors may be able to recruit TET dioxygenases to their target loci. A physical interaction between NANOG and TET1/TET2 and NANOG and TET1 co-occupancy of specific binding sites supports this model (123). TET activity at transcription factor binding sites may serve to release MBPs, recruit specific 5-hmC binding proteins, and/or demethylate nearby CpGs (11). Although numerous transcription factors have methylation-inhibited binding properties, none of these transcription factors has yet been shown to bind 5-hmC, but this remains a possibility requiring further work. Interestingly, in the differentiating and replicative systems used by Serandour *et al.* (117) and Madzo *et al.* (70), 5-hmC marks persist over several days, suggesting that 5-hmC may confer its own epigenetic function beyond demethylation.

The studies discussed above have largely used affinity based 5-hmC detection techniques to map 5-hmC within the genome. Although these techniques have provided useful maps of 5-hmC with resolution on the order of hundreds of base pairs, few studies to date have examined 5-hmC at single base pair resolution in the context of transcription factor binding sites (124). Genome-wide sequencing of 5-hmC at single base pair resolution in mouse ESCs revealed that 5-hmC is depleted at the core binding sequences of proteins such as CTCF and Nanog with a bimodal distribution of 5-hmC adjacent to these sites (125). Affinity-based techniques would not be able to distinguish this small difference in 5-hmC position.

Similar to the results from this single base-resolution study, our laboratory has mapped 5-hmC changes induced by hypoxia in neuroblastoma cells with affinity-based techniques and at single-base resolution (71). Affinity-based sequencing of 5-hmC gains in hypoxia overlap with hypoxia inducible factor-1 (HIF-1) binding sites. One of these HIF-1 binding sites is at the *CA9* transcripiton start site. Single-base resolution mapping of 5-hmC at this site, showed that 5-hmC accumulated 74 bp upstream of the HIF-1 binding site, but not at the core HIF-1 binding motif. Nevertheless, the CpG within the core HIF-1 binding site is demethylated in hypoxia. This observation leaves open the possibility that TET-mediated oxidation of CpGs within transcription factor binding sites could drive an active demethylation process that is too fast for 5-hmC accumulation at these CpGs to be observed (71).

TET enzymes can further oxidize 5-hmC to 5-fC and 5-caC (7, 8). Although these bases are present at levels orders of magnitude lower than 5-hmC in the genome (7), mapping of 5-fC in mESCs reveals that 5-fC is enriched at LMRs and enhancer regions (126). Among enhancers, 5-fC is especially enriched at poised and promoter-linked enhancers. At these, sites 5-fC likely plays a role in active demethylation processes that maintain low methylation levels. Consistent with this observation, 5-fC was enriched at CTCF, p300, DHS, and H3K4me1 marks (126). Unlike 5-hmC, however, 5-fC was not enriched at transcription factor binding sites including Nanog and Oct4 (126). Collectively, these studies have demonstrated cytosine modifications can affect accessibility of enhancer regions and transcription factor binding sites to regulate gene transcription.

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

Dietary influences on modified cytosine levels

Folate

Altered genomic patterns of cytosine modification, as well as mutations of the *TET* and *DNMT* epigenetic modifiers, are well-established phenomena in hematological malignancies (127–129). As a result, clarifying the roles of dietary coenzymes in establishing epigenetic marks and in modulating the function of cytosine modifying enzymes is of particular interest for understanding epigenetic regulation of hematological function.

Folate is an essential, water-soluble B-vitamin that is obtained from a diet of fruits and vegetables (130, 131). Derivatives of folate and folic acid, which is the synthetic form of folate that is found in supplements and fortified foods, are utilized in one-carbon metabolism and are the main effectors of trans-methylation reactions, such as cytosine methylation (131). Specifically, folate-derived 5,10-methylenetetrahydrofolate (5mTHF) is crucial for the establishment of 5-mC, as 5mTHF is required for the maintenance of appropriate methionine levels in the cell. The enzyme S-adenosylmethionine transferase (MAT) links methionine and a molecule of ATP to generate the principal cytosine methyl donor in the cell, SAM (Fig. 4). Due to its function in one-carbon metabolism, folate is considered a major dietary methyl donor.

The influence of dietary methyl donors on the establishment of 5-mC first began to be fully recognized through studies using *agouti* mice in the mid-1990s. In these mice, intracisternal A-particle (IAP) sequences are interspersed throughout the *agouti* gene, which determines coat color directly and influences body size, predisposition to metabolic disorders, and susceptibility to tumorigenesis through pleiotropic effects (132). In 1994, a critical study showed that the level of expression of the *agouti* gene depends upon the methylation status of the inserted IAP 5′ long terminal repeat (LTR); specifically, increased expression of the *agouti* gene was observed with decreased levels of IAP LTR methylation (132). In 1998, another group discovered that feeding pregnant *agouti* mice methyl-supplemented diets (e.g. excess folic acid, vitamin B_{12}, and betaine) determines the phenotypic outcome and IAP LTR methylation status of offspring (133). *Agouti* offspring whose mothers had been fed the methyl-supplemented diet were darker in color, leaner, and exhibited a lower incidence of metabolic disorders and tumor formation than the large, yellow, highly *agouti*-expressing offspring whose mothers had been fed non-supplemented diets (133). This research established a role for dietary folate in modulating 5-mC levels.

These studies instigated an outpouring of research on modifying levels of 5-mC through dietary intake. Research examining the effects of folate deficiency and dietary supplementation on genomic 5-mC has been abundant, though largely inconsistent (134). Observed correlations between folate supplementation and 5-mC levels are often context-dependent, depending on the organism and organ site studied, the dose and timing of folate administration, the number and resolution of CpGs examined, e.g. locus-specific versus global levels of 5-mC, and other confounding factors, such as age, diet, and lifestyle, especially in human studies (135). Regardless, across the majority of studies, folate deficiency is associated with global hypomethylation, while increased folate intake is correlated with elevated levels of global 5-mC (130). However, due to the complex nature of

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

epigenetic regulation, examining site-specific variations in 5-mC due to dietary folate are likely to be more physiologically relevant than are global 5-mC levels (135).

A recent study examining the relationship between maternal folic acid intake and 5-mC distribution in the murine cerebral hemisphere has shown that folate levels are correlated with site-specific hyper- and hypomethylation (136). These data show that offspring whose mothers were fed a folic acid-supplemented diet exhibited a differential 5-mC pattern to offspring whose mothers were fed a low folic acid diet at CpG sites within promoters and gene bodies, as well as at non-CpG sites. Interestingly, high maternal folic acid intake was associated with both hyper- and hypomethylation of cytosine residues across all the genomic elements studied, suggesting that folate status may augment programs to establish 5-mC in a site-specific manner during gestation (136). Furthermore, the genes that were affected by maternal folate intake and the distribution of 5-mC across the genome differed across male and female pups in the study, suggesting that the site-specific establishment of modified cytosine residues may be highly regulated during development and, at least in part, influenced by folate status.

The influence of folate status on site-specific 5-mC establishment has also recently been addressed in the human genome. Metastable alleles, which stochastically acquire or resist methylation during gestation and whose modification status is stably inherited through mitotic divisions, are recognized in humans (137). Similar to the *agouti* gene in mice, the methylation status of human metastable alleles is determined by maternal dietary intake. In a recent study from Dominguez-Salas *et al.* (138), cytosine modification of meta-stable alleles were monitored in the offspring of rural Gambian women whose diets vary according to the rainy and dry seasons. During the rainy season, Gambian women exhibited higher levels of folate, methionine, vitamin B12, and SAM. The offspring of women who conceived during the rainy season had increased levels of CpG methylation at six of the seven metastable alleles examined (138), demonstrating that maternal folate intake impacts site-specific 5-mC establishment.

Therefore, recent studies that focus on specific sites in the genome whose modification status is influenced by folate status are making promising strides toward clarifying the effects of dietary folate on the epigenome. However, many questions remain concerning the role of folate in the establishment of cytosine modifications. Apart from the *agouti* gene, the functional and phenotypic consequences of differentially modified metastable alleles in both mice and humans have not yet been established. Furthermore, the method used to determine methylation status in most, if not all, of the discussed studies is bisulfite sequencing; this method does not differentiate between modified cytosines, e.g. 5-mC versus 5-hmC, which may affect conclusions concerning methylation status in these studies.

Vitamin C

Differentiating between 5-mC and 5-hmC has become even more important in studies examining the influence of dietary cofactors on the epigenome, as new evidence has shown that ascorbic acid (vitamin C) influences 5-hmC levels and TET enzyme function (Fig. 4). Vitamin C is an essential vitamin that acts as an antioxidant in the cell under physiological conditions. Vitamin C is a crucial cofactor in reactions catalyzed by α-ketoglutarate-

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

dependent dioxygenases, such as the TET enzymes, as it may reduce Fe(III) to the dioxygenase cofactor Fe(II) (139). However, a potential role for vitamin C in promoting the establishment of 5-hmC has not been appreciated until very recently.

Several studies have recently shown that vitamin C induces increased 5-hmC levels *in vitro* in λ DNA, in cultured ES cells, and *in vivo* in mice using a variety of techniques (140–143). Furthermore, it has been shown that increased levels of other highly oxidized modifications (e.g. 5-fC and 5-caC) occurs upon vitamin C administration, suggesting that the coenzyme stimulates active DNA demethylation (142). A concomitant decrease in levels of 5-mC with vita-min C treatment detected using ultra high performance liquid chromatography supports the hypothesis that vitamin C promotes active DNA demethylation (142). Furthermore, another publication showed that vitamin C may also promote the establishment of 5-hmC for regulatory functions, as the authors show that 5-hmC is gained in a gene-specific manner upon vitamin C treatment to promote a blastocyst-like state in ES cells (143). As a result, these studies reveal site-specific establishment of 5-hmC and demethylation due to vitamin C administration.

Additionally, two studies have demonstrated that vitamin C controls 5-hmC levels by modulating the activity of TET enzymes. Using low doses, Yin *et al.* (142) showed that vitamin C enhanced Tet1- and Tet2-mediated production of 5-hmC in ES cells. By measuring the levels of intrinsic fluorescence of a tryptophan-rich catalytic Tet domain, Yin *et al.* (142) revealed the enhancement of 5-hmC production occurs through a direct interaction between vitamin C and the Tet catalytic domain. This interaction may not just affect activity but also the functionality of the TET enzymes. Chen *et al.* (144) showed that TET1 may act as a positive or negative regulator of somatic cell reprogramming depending on the concentration of vitamin C in the reaction medium; at low doses, vitamin C increases reprogramming efficiency of somatic cells to stem cells, whereas at high doses, vitamin C delays reprogramming and inhibits induction of genes associated with reprogramming in their experimental system. Although the exact mechanism of vitamin C induction of 5-hmC and the nature of its interaction with TET dioxygenases remain unclear, vitamin C clearly influences the composition of the epigenome.

The influence of dietary coenzymes on cytosine modification has been, and continues to be, an area of active study. Appreciating the impact of folate, vitamin C, and other cofactors on cytosine modifications will be valuable to understanding the contribution of diet to health, disease, and genomic function.

Conclusions and areas for future research

In this review, we have described several new themes on the biological functions of 5-mC and 5-hmC. Beginning at the transcriptional start site and extending along the body of genes, it is clear that these cytosine modifications have different epigenetic functions dependent upon their genic context. The idea that cytosine modifications, particularly 5-mC, regulate transcription potential at the promoter of genes has been well-established. However, it is becoming more apparent that 5-mC and 5-hmC are also playing a role in elongation and mRNA processing, which suggests their influences on gene expression levels are more

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

multifaceted than previously thought. As genome-wide DNA methylation and hydroxymethylation techniques become more widely used and affordable, we expect to see further distinction of their various roles in regulation transcription. It is also likely that there are some tissue- and cell type-specific effects of 5-mC and 5-hmC in transcriptional regulation; for instance, 5-hmC is significantly more prevalent in the brain compared with other somatic tissues (66, 67). Future work investigating genome-wide distributions of 5-mC and 5-hmC and their associations with gene expression levels in various tissues will shed more light on those differences.

Ongoing research on the interactions between non-coding RNAs and epigenetic modifiers is an area that will likely lead to the discovery of novel mechanisms of epigenetic regulation. We highlighted several papers that discuss how non-coding RNAs can promote and protect DNA from *de novo* methylation. The mechanisms of DNMT3A and DNMT3B recruitment to target genes are not well-understood, but these studies suggest that non-coding RNAs may be playing an important role in targeting DNMTs to specific genes for methylation. Furthermore, although interactions between non-coding RNAs and TETs have yet to be described, this could be a potential mechanism for TET recruitment to target genes for hydroxymethylation. The DNMTs, TETs, and non-coding RNAs are all disrupted in hematological malignancies, and offsetting the balance of just one of these epi-genetic factors could lead to dysregulation of the entire epigenetic profile of the cell. It is exciting to consider the extent of opportunities for research in this area, especially within the realm of hematological malignancies, given the importance of epigenetic modifiers in disease pathogenesis. For instance, discovering cooperative mechanisms of action between epigenetic modifiers could lead to development of novel therapeutic strategies to improve patient care.

In this review, we also emphasized the role of dietary cofactors, folate and vitamin C, in regulating levels of 5-mC and 5-hmC respectively. Folate deficiency is associated with global hypomethylation due to its key role in one-carbon metabolism. Vitamin C has been demonstrated to promote TET activity and influence 5-hmC levels at specific loci (see above). Inter-individual variation in cytosine modification patterns is influenced by numerous factors, including environmental exposures such as diet. The diet of patients with hematological malignancies could consequently affect the epigenetic profile of their cancer or response to epigenetic therapies and warrants additional study.

The establishment of 5-mC patterns by the DNMTs and 5-hmC patterns by the TETs is a critical part of normal physiology that is disrupted in hematological malignancies. Beyond understanding the basic biology behind transcriptional regulation, future investigation within this field could have important clinical implications. The balance of the activity of these enzymes is essential for proper epigenetic programming, and this balance is upset in hematological malignancies, especially in those cancers with mutations in *DNMT3A*, *TET2*, or *IDH1/2*. Targeting epigenetic modifiers for therapeutic interventions has yielded modest success so far and represents a promising treatment avenue for patients with hematological malignancies.

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

Acknowledgements

The authors have no conflicts of interest to declare.

References

1. Bartolomei MS, Tilghman SM. Genomic imprinting in mammals. Annu Rev Genet 1997;31:493–525. [PubMed: 9442905]
2. Panning B, Jaenisch R. RNA and the epigenetic regulation of X chromosome inactivation. Cell 1998;93:305–308. [PubMed: 9590161]
3. Nagase H, Ghosh S. Epigenetics: differential DNA methylation in mammalian somatic tissues. FEBS J 2008;275:1617–1623. [PubMed: 18331347]
4. Liang G, et al. Cooperativity between DNA methyltransferases in the maintenance methylation of repetitive elements. Mol Cell Biol 2002;22:480–491. [PubMed: 11756544]
5. Ito S, D'Alessio AC, Taranova OV, Hong K, Sowers LC, Zhang Y. Role of Tet proteins in 5mC to 5hmC conversion, ES-cell self-renewal and inner cell mass specification. Nature 2010;466:1129–1133. [PubMed: 20639862]
6. Tahiliani M, et al. Conversion of 5-methylcytosine to 5-hydroxymethylcytosine in mammalian DNA by MLL partner TET1. Science 2009;324:930–935. [PubMed: 19372391]
7. Ito S, et al. Tet proteins can convert 5-methylcytosine to 5-formylcytosine and 5-carboxylcytosine. Science 2011;333:1300–1303. [PubMed: 21778364]
8. He YF, et al. Tet-mediated formation of 5-carboxylcytosine and its excision by TDG in mammalian DNA. Science 2011;333:1303–1307. [PubMed: 21817016]
9. Jasielec J, Saloura V, Godley LA. The mechanistic role of DNA methylation in myeloid leukemogenesis. Leukemia 2014;28:1765–1773. [PubMed: 24913729]
10. Madzo J, Vasanthakumar A, Godley LA. Perturbations of 5-hydroxymethylcytosine patterning in hematologic malignancies. Semin Hematol 2013;50:61–69. [PubMed: 23507484]
11. Mariani CJ, Madzo J, Moen EL, Yesilkanal A, Godley LA. Alterations of 5-hydroxymethylcytosine in human cancers. Cancers (Basel) 2013;5:786–814. [PubMed: 24202321]
12. Abdel-Wahab O, Levine RL. Mutations in epigenetic modifiers in the pathogenesis and therapy of acute myeloid leukemia. Blood 2013;121:3563–3572. [PubMed: 23640996]
13. Challen GA, et al. Dnmt3a is essential for hematopoietic stem cell differentiation. Nat Genet 2012;44:23–31.
14. Thol F, et al. Incidence and prognostic influence of DNMT3A mutations in acute myeloid leukemia. J Clin Oncol 2011;29:2889–2896. [PubMed: 21670448]
15. Yan XJ, et al. Exome sequencing identifies somatic mutations of DNA methyltransferase gene DNMT3A in acute monocytic leukemia. Nat Genet 2011;43:309–315. [PubMed: 21399634]
16. Mizuno S, et al. Expression of DNA methyltransferases DNMT1, 3A, and 3B in normal hematopoiesis and in acute and chronic myelogenous leukemia. Blood 2001;97:1172–1179. [PubMed: 11222358]
17. Melki JR, Vincent PC, Clark SJ. Concurrent DNA hypermethylation of multiple genes in acute myeloid leukemia. Cancer Res 1999;59:3730–3740. [PubMed: 10446989]
18. Jiang Y, et al. Aberrant DNA methylation is a dominant mechanism in MDS progression to AML. Blood 2009;113:1315–1325. [PubMed: 18832655]
19. Marcucci G, et al. Age-related prognostic impact of different types of DNMT3A mutations in adults with primary cytogenetically normal acute myeloid leukemia. J Clin Oncol 2012;30:742–750. [PubMed: 22291079]
20. Yamashita Y, et al. Array-based genomic resequencing of human leukemia. Oncogene 2010;29:3723–3731. [PubMed: 20400977]
21. Zeidan AM, Linhares Y, Gore SD. Current therapy of myelodysplastic syndromes. Blood Rev 2013;27:243–259. [PubMed: 23954262]
22. Kantarjian HM, et al. Multicenter, randomized, open-label, phase III trial of decitabine versus patient choice, with physician advice, of either supportive care or low-dose cytarabine for the

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

treatment of older patients with newly diagnosed acute myeloid leukemia. J Clin Oncol 2012;30:2670–2677. [PubMed: 22689805]

23. Fenaux P, et al. Efficacy of azacitidine compared with that of conventional care regimens in the treatment of higher-risk myelodysplastic syndromes: a randomised, open-label, phase III study. Lancet Oncol 2009;10:223–232. [PubMed: 19230772]
24. Silverman LR, et al. Randomized controlled trial of azacitidine in patients with the myelodysplastic syndrome: a study of the cancer and leukemia group B. J Clin Oncol 2002;20:2429–2440. [PubMed: 12011120]
25. Kantarjian H, et al. Decitabine improves patient outcomes in myelodysplastic syndromes: results of a phase III randomized study. Cancer 2006;106:1794–1803. [PubMed: 16532500]
26. Galm O, Herman JG, Baylin SB. The fundamental role of epigenetics in hematopoietic malignancies. Blood Rev 2006;20:1–13. [PubMed: 16426940]
27. Koh KP, et al. Tet1 and Tet2 regulate 5-hydroxymethylcytosine production and cell lineage specification in mouse embryonic stem cells. Cell Stem Cell 2011;8:200–213. [PubMed: 21295276]
28. Langlois T, et al. TET2 deficiency inhibits mesoderm and hematopoietic differentiation in human embryonic stem cells. Stem Cells 2014;32:2084–2097. [PubMed: 24723429]
29. Gu TP, et al. The role of Tet3 DNA dioxygenase in epigenetic reprogramming by oocytes. Nature 2011;477:606–610. [PubMed: 21892189]
30. Iqbal K, Jin SG, Pfeifer GP, Szab o PE. Reprogramming of the paternal genome upon fertilization involves genome-wide oxidation of 5-methylcytosine. Proc Natl Acad Sci USA 2011;108:3642–3647. [PubMed: 21321204]
31. Wossidlo M, et al. 5-Hydroxymethylcytosine in the mammalian zygote is linked with epigenetic reprogramming. Nat Commun 2011;2:241. [PubMed: 21407207]
32. Delhommeau F, et al. Mutation in TET2 in myeloid cancers. N Engl J Med 2009;360:2289–2301. [PubMed: 19474426]
33. Chou WC, et al. TET2 mutation is an unfavorable prognostic factor in acute myeloid leukemia patients with intermediate-risk cytogenetics. Blood 2011;118:3803–3810. [PubMed: 21828143]
34. Moran-Crusio K, et al. Tet2 loss leads to increased hematopoietic stem cell self-renewal and myeloid transformation. Cancer Cell 2011;20:11–24. [PubMed: 21723200]
35. Quivoron C, et al. TET2 inactivation results in pleiotropic hematopoietic abnormalities in mouse and is a recurrent event during human lymphomagenesis. Cancer Cell 2011;20:25–38. [PubMed: 21723201]
36. Ko M, et al. Ten-Eleven-Translocation 2 (TET2) negatively regulates homeostasis and differentiation of hematopoietic stem cells in mice. Proc Natl Acad Sci USA 2011;108:14566–14571. [PubMed: 21873190]
37. Li Z, et al. Deletion of Tet2 in mice leads to dysregulated hematopoietic stem cells and subsequent development of myeloid malignancies. Blood 2011;118:4509–4518. [PubMed: 21803851]
38. Shide K, et al. TET2 is essential for survival and hematopoietic stem cell homeostasis. Leukemia 2012;26:2216–2223. [PubMed: 22469782]
39. Dawlaty MM, et al. Combined deficiency of Tet1 and Tet2 causes epigenetic abnormalities but is compatible with postnatal development. Dev Cell 2013;24:310–323. [PubMed: 23352810]
40. Muto T, et al. Concurrent loss of Ezh2 and Tet2 cooperates in the pathogenesis of myelodysplastic disorders. J Exp Med 2013;210:2627–2639. [PubMed: 24218139]
41. Marcucci G, et al. IDH1 and IDH2 gene mutations identify novel molecular subsets within de novo cytogenetically normal acute myeloid leukemia: a Cancer and Leukemia Group B study. J Clin Oncol 2010;28:2348–2355. [PubMed: 20368543]
42. Wang F, et al. Targeted inhibition of mutant IDH2 in leukemia cells induces cellular differentiation. Science 2013;340:622–626. [PubMed: 23558173]
43. Stein E, et al. Clinical safety and activity in a phase I trial of AG-221, a first in class, potent inhibitor of the IDH2-mutant protein, in patients with IDH2 mutant positive advanced hematologic malignancies [abstract]. Proc Ann Meeting AACR. San Diego, CA, 2014.

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

44. Saxonov S, Berg P, Brutlag DL. A genome-wide analysis of CpG dinucleotides in the human genome distinguishes two distinct classes of promoters. Proc Natl Acad Sci USA 2006;103:1412–1417. [PubMed: 16432200]
45. Kelly TK, et al. H2A.Z maintenance during mitosis reveals nucleosome shifting on mitotically silenced genes. Mol Cell 2010;39:901–911. [PubMed: 20864037]
46. Doi A, et al. Differential methylation of tissue-and cancer-specific CpG island shores distinguishes human induced pluripotent stem cells, embryonic stem cells and fibroblasts. Nat Genet 2009;41:1350–1353. [PubMed: 19881528]
47. Lock LF, Takagi N, Martin GR. Methylation of the Hprt gene on the inactive X occurs after chromosome inactivation. Cell 1987;48:39–46. [PubMed: 3791414]
48. Ooi SK, et al. DNMT3L connects unmethylated lysine 4 of histone H3 to de novo methylation of DNA. Nature 2007;448:714–717. [PubMed: 17687327]
49. You JS, Kelly TK, De Carvalho DD, Taberlay PC, Liang G, Jones PA. OCT4 establishes and maintains nucleosome-depleted regions that provide additional layers of epigenetic regulation of its target genes. Proc Natl Acad Sci USA 2011;108:14497–14502. [PubMed: 21844352]
50. Williams K, et al. TET1 and hydroxymethylcytosine in transcription and DNA methylation fidelity. Nature 2011;473:343–348. [PubMed: 21490601]
51. Ficz G, et al. Dynamic regulation of 5-hydroxymethylcytosine in mouse ES cells and during differentiation. Nature 2011;473:398–402. [PubMed: 21460836]
52. Guo JU, Su Y, Zhong C, Ming GL, Song H. Hydroxylation of 5-methylcytosine by TET1 promotes active DNA demethylation in the adult brain. Cell 2011;145:423–434. [PubMed: 21496894]
53. Robertson J, Robertson A, Klungland A. The presence of 5-hydroxymethylcytosine at the gene promoter and not in the gene body negatively regulates gene expression. Biochem Biophy Res Commun 2011;411:40–43.
54. Nguyen C, et al. Susceptibility of nonpromoter CpG islands to de novo methylation in normal and neoplastic cells. J Natl Cancer Inst 2001;93:1465–1472. [PubMed: 11584062]
55. Huang Y, Pastor WA, Shen Y, Tahiliani M, Liu DR, Rao A. The behaviour of 5-hydroxymethylcytosine in bisulfite sequencing. PLoS ONE 2010;5:e8888. [PubMed: 20126651]
56. Jin SG, Kadam S, Pfeifer GP. Examination of the specificity of DNA methylation profiling techniques towards 5-methylcytosine and 5-hydroxymethylcytosine. Nucleic Acids Res 2010;38:e125. [PubMed: 20371518]
57. Jones PA. The DNA methylation paradox. Trends Genet 1999;15:34–37. [PubMed: 10087932]
58. Ball MP, et al. Targeted and genome-scale strategies reveal gene-body methylation signatures in human cells. Nat Biotechnol 2009;27:361–368. [PubMed: 19329998]
59. Moen EL, et al. Genome-wide variation of Cytosine modifications between European and african populations and the implications for complex traits. Genetics 2013;194:987–996. [PubMed: 23792949]
60. Hahn MA, Wu X, Li AX, Hahn T, Pfeifer GP. Relationship between gene body DNA methylation and intragenic H3K9me3 and H3K36me3 chromatin marks. PLoS ONE 2011;6: e18844. [PubMed: 21526191]
61. Chodavarapu RK, et al. Relationship between nucleosome positioning and DNA methylation. Nature 2010;466:388–392. [PubMed: 20512117]
62. Chen Y, Chen Q, McEachin RC, Cavalcoli JD, Yu X. H2A.B facilitates transcription elongation at methylated CpG loci. Genome Res 2014;24:570–579. [PubMed: 24402521]
63. Moen EL, Stark AL, Zhang W, Dolan ME, Godley LA. The role of gene body cytosine modifications in MGMT expression and sensitivity to temozolomide. Mol Cancer Ther 2014;13:1334–1344. [PubMed: 24568970]
64. Hagemann S, Heil O, Lyko F, Brueckner B. Azacytidine and decitabine induce gene-specific and non-random DNA demethylation in human cancer cell lines. PLoS ONE 2011;6:e17388. [PubMed: 21408221]
65. Wu H, et al. Genome-wide analysis of 5-hydroxymethylcytosine distribution reveals its dual function in transcriptional regulation in mouse embryonic stem cells. Genes Dev 2011;25:679–684. [PubMed: 21460036]

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

66. Kriaucionis S, Heintz N. The nuclear DNA base 5-hydroxymethylcytosine is present in Purkinje neurons and the brain. Science 2009;324:929–930. [PubMed: 19372393]
67. Mell en M, Ayata P, Dewell S, Kriaucionis S, Heintz N. MeCP2 binds to 5hmC enriched within active genes and accessible chromatin in the nervous system. Cell 2012;151:1417–1430. [PubMed: 23260135]
68. Hahn MA, et al. Dynamics of 5-hydroxymethylcytosine and chromatin marks in Mammalian neurogenesis. Cell Rep 2013;3:291–300. [PubMed: 23403289]
69. Colquitt BM, Allen WE, Barnea G, Lomvardas S. Alteration of genic 5-hydroxymethylcytosine patterning in olfactory neurons correlates with changes in gene expression and cell identity. Proc Natl Acad Sci USA 2013;110:14682–14687. [PubMed: 23969834]
70. Madzo J, et al. Hydroxymethylation at gene regulatory regions directs stem/early progenitor cell commitment during erythropoiesis. Cell Rep 2014;6:231–244. [PubMed: 24373966]
71. Mariani CJ, et al. TET1-mediated hydroxymethylation facilitates hypoxic gene induction in neuroblastoma. Cell Rep 2014;7:1343–1352. [PubMed: 24835990]
72. Laurent L, et al. Dynamic changes in the human methylome during differentiation. Genome Res 2010;20:320–331. [PubMed: 20133333]
73. Shukla S, et al. CTCF-promoted RNA polymerase II pausing links DNA methylation to splicing. Nature 2011;479:74–79. [PubMed: 21964334]
74. Choi JK. Contrasting chromatin organization of CpG islands and exons in the human genome. Genome Biol 2010;11:R70. [PubMed: 20602769]
75. Lyko F, Foret S, Kucharski R, Wolf S, Falckenhayn C, Maleszka R. The honey bee epigenomes: differential methylation of brain DNA in queens and workers. PLoS Biol 2010;8: e1000506. [PubMed: 21072239]
76. Foret S, et al. DNA methylation dynamics, metabolic fluxes, gene splicing, and alternative phenotypes in honey bees. Proc Natl Acad Sci USA 2012;109:4968–4973. [PubMed: 22416128]
77. Wan J, Oliver VF, Zhu H, Zack DJ, Qian J, Merbs SL. Integrative analysis of tissue-specific methylation and alternative splicing identifies conserved transcription factor binding motifs. Nucleic Acids Res 2013;41:8503–8514. [PubMed: 23887936]
78. Khare T, et al. 5-hmC in the brain is abundant in synaptic genes and shows differences at the exon-intron boundary. Nat Struct Mol Biol 2012;19:1037–1043. [PubMed: 22961382]
79. Huang Y, et al. Distinct roles of the methylcytosine oxidases Tet1 and Tet2 in mouse embryonic stem cells. Proc Natl Acad Sci USA 2014;111:1361–1366. [PubMed: 24474761]
80. Wen L, et al. Whole-genome analysis of 5-hydroxymethylcytosine and 5-methylcytosine at base resolution in the human brain. Genome Biol 2014;15:R49. [PubMed: 24594098]
81. Kimura K, et al. Diversification of transcriptional modulation: large-scale identification and characterization of putative alternative promoters of human genes. Genome Res 2006;16:55–65. [PubMed: 16344560]
82. Maunakea AK, et al. Conserved role of intragenic DNA methylation in regulating alternative promoters. Nature 2010;466:253–257. [PubMed: 20613842]
83. Iorio MV, Piovan C, Croce CM. Interplay between microRNAs and the epigenetic machinery: an intricate network. Biochim Biophys Acta 2010;1799:694–701. [PubMed: 20493980]
84. Kulis M, Queiros AC, Beekman R, Martín-Subero JI. Intragenic DNA methylation in transcriptional regulation, normal differentiation and cancer. Biochim Biophys Acta 2013;1829:1161–1174. [PubMed: 23938249]
85. Lujambio A, et al. A microRNA DNA methylation signature for human cancer metastasis. Proc Natl Acad Sci USA 2008;105:13556–13561. [PubMed: 18768788]
86. Lujambio A, et al. Genetic unmasking of an epigenetically silenced microRNA in human cancer cells. Cancer Res 2007;67:1424–1429. [PubMed: 17308079]
87. Toyota M, et al. Epigenetic silencing of microRNA-34b/c and B-cell translocation gene 4 is associated with CpG island methylation in colorectal cancer. Cancer Res 2008;68:4123–4132. [PubMed: 18519671]
88. Di Ruscio A, et al. DNMT1-interacting RNAs block gene-specific DNA methylation. Nature 2013;503:371–376. [PubMed: 24107992]

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

89. Schmitz KM, Mayer C, Postepska A, Grummt I. Interaction of noncoding RNA with the rDNA promoter mediates recruitment of DNMT3b and silencing of rRNA genes. Genes Dev 2010;24:2264–2269. [PubMed: 20952535]

90. Ginno PA, Lott PL, Christensen HC, Korf I, Ch edin F. R-loop formation is a distinctive characteristic of unmethylated human CpG island promoters. Mol Cell 2012;45:814–825. [PubMed: 22387027]

91. Zhao S, et al. MicroRNA-126 regulates DNA methylation in CD4 + T cells and contributes to systemic lupus erythematosus by targeting DNA methyltransferase 1. Arthritis Rheum 2011;63:1376–1386. [PubMed: 21538319]

92. Fabbri M, et al. MicroRNA-29 family reverts aberrant methylation in lung cancer by targeting DNA methyltransferases 3A and 3B. Proc Natl Acad Sci USA 2007;104:15805–15810. [PubMed: 17890317]

93. Stadler MB, et al. DNA-binding factors shape the mouse methylome at distal regulatory regions. Nature 2011;480:490–495. [PubMed: 22170606]

94. Woodcock DM, Simmons DL, Crowther PJ, Cooper IA, Trainor KJ, Morley AA. Delayed DNA methylation is an integral feature of DNA replication in mammalian cells. Exp Cell Res 1986;166:103–112. [PubMed: 2427346]

95. Shirane K, et al. Mouse oocyte methylomes at base resolution reveal genome-wide accumulation of non-CpG methylation and role of DNA methyltransferases. PLoS Genet 2013;9: e1003439. [PubMed: 23637617]

96. Tomizawa S, et al. Dynamic stage-specific changes in imprinted differentially methylated regions during early mammalian development and prevalence of non-CpG methylation in oocytes. Development 2011;138:811–820. [PubMed: 21247965]

97. Jia D, Jurkowska RZ, Zhang X, Jeltsch A, Cheng X. Structure of Dnmt3a bound to Dnmt3L suggests a model for de novo DNA methylation. Nature 2007;449:248–251. [PubMed: 17713477]

98. Shah MY, Licht JD. DNMT3A mutations in acute myeloid leukemia. Nat Genet 2011;43:289–290. [PubMed: 21445072]

99. Ley TJ, et al. DNMT3A mutations in acute myeloid leukemia. N Engl J Med 2010;363:2424–2433. [PubMed: 21067377]

100. Shlyueva D, Stampfel G, Stark A. Transcriptional enhancers: from properties to genome-wide predictions. Nat Rev Genet 2014;15: 272–286. [PubMed: 24614317]

101. Dekker J, Rippe K, Dekker M, Kleckner N. Capturing chromosome conformation. Science 2002;295:1306–1311. [PubMed: 11847345]

102. Schaaf CA, et al. Genome-wide control of RNA polymerase II activity by cohesin. PLoS Genet 2013;9:e1003382. [PubMed: 23555293]

103. Seitan VC, et al. Cohesin-based chromatin interactions enable regulated gene expression within preexisting architectural compartments. Genome Res 2013;23:2066–2077. [PubMed: 24002784]

104. Merkenschlager M, Odom DT. CTCF and cohesin: linking gene regulatory elements with their targets. Cell 2013;152:1285–1297. [PubMed: 23498937]

105. DeMare LE, et al. The genomic landscape of cohesin-associated chromatin interactions. Genome Res 2013;23:1224–1234. [PubMed: 23704192]

106. Faure AJ, et al. Cohesin regulates tissue-specific expression by stabilizing highly occupied cis-regulatory modules. Genome Res 2012;22:2163–2175. [PubMed: 22780989]

107. Kon A, et al. Recurrent mutations in multiple components of the cohesin complex in myeloid neoplasms. Nat Genet 2013;45:1232–1237. [PubMed: 23955599]

108. Thol F, et al. Mutations in the cohesin complex in acute myeloid leukemia: clinical and prognostic implications. Blood 2014;123:914–920. [PubMed: 24335498]

109. Rocquain J, et al. Alteration of cohesin genes in myeloid diseases. Am J Hematol 2010;85:717–719. [PubMed: 20687102]

110. Network CGAR. Genomic and epigenomic landscapes of adult de novo acute myeloid leukemia. N Engl J Med 2013;368:2059–2074. [PubMed: 23634996]

111. Schmidt D, et al. A CTCF-independent role for cohesin in tissue-specific transcription. Genome Res 2010;20:578–588. [PubMed: 20219941]

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

112. Wiench M, et al. DNA methylation status predicts cell type-specific enhancer activity. EMBO J 2011;30:3028–3039. [PubMed: 21701563]
113. Wu H, Zhang Y. Tet1 and 5-hydroxymethylation: a genome-wide view in mouse embryonic stem cells. Cell Cycle 2011;10:2428–2436. [PubMed: 21750410]
114. Stroud H, Feng S, Morey Kinney S, Pradhan S, Jacobsen SE. 5-Hydroxymethylcytosine is associated with enhancers and gene bodies in human embryonic stem cells. Genome Biol 2011;12:R54. [PubMed: 21689397]
115. Gan H, et al. Dynamics of 5-hydroxymethylcytosine during mouse spermatogenesis. Nat Commun 2013;4:1995. [PubMed: 23759713]
116. Feldmann A, Ivanek R, Murr R, Gaidatzis D, Burger L, Schübeler D. Transcription factor occupancy can mediate active turnover of DNA methylation at regulatory regions. PLoS Genet 2013;9:e1003994. [PubMed: 24367273]
117. Sérandour AA, et al. Dynamic hydroxymethylation of deoxyribonucleic acid marks differentiation-associated enhancers. Nucleic Acids Res 2012;40:8255–8265. [PubMed: 22730288]
118. Szulwach KE, et al. 5-hmC-mediated epigenetic dynamics during postnatal neurodevelopment and aging. Nat Neurosci 2011;14:1607–1616. [PubMed: 22037496]
119. Sun Z, et al. High-resolution enzymatic mapping of genomic 5-hydroxymethylcytosine in mouse embryonic stem cells. Cell Rep 2013;3:567–576. [PubMed: 23352666]
120. Pastor WA, et al. Genome-wide mapping of 5-hydroxymethylcytosine in embryonic stem cells. Nature 2011;473:394–397. [PubMed: 21552279]
121. Szulwach KE, et al. Integrating 5-hydroxymethylcytosine into the epigenomic landscape of human embryonic stem cells. PLoS Genet 2011;7:e1002154. [PubMed: 21731508]
122. Okashita N, et al. PRDM14 promotes active DNA demethylation through the ten-eleven translocation (TET)-mediated base excision repair pathway in embryonic stem cells. Development 2014;141:269–280. [PubMed: 24335252]
123. Costa Y, et al. NANOG-dependent function of TET1 and TET2 in establishment of pluripotency. Nature 2013;495:370–374. [PubMed: 23395962]
124. Song CX, Yi C, He C. Mapping recently identified nucleotide variants in the genome and transcriptome. Nat Biotechnol 2012;30:1107–1116. [PubMed: 23138310]
125. Yu M, et al. Base-resolution analysis of 5-hydroxymethylcytosine in the Mammalian genome. Cell 2012;149:1368–1380. [PubMed: 22608086]
126. Song CX, et al. Genome-wide profiling of 5-formylcytosine reveals its roles in epigenetic priming. Cell 2013;153:678–691. [PubMed: 23602153]
127. Fathi AT, Abdel-Wahab O. Mutations in epigenetic modifiers in myeloid malignancies and the prospect of novel epigenetic-targeted therapy. Adv Hematol 2012;2012:469592. [PubMed: 21811504]
128. Solary E, Bernard OA, Tefferi A, Fuks F, Vainchenker W. The Ten-Eleven Translocation-2 (TET2) gene in hematopoiesis and hematopoietic diseases. Leukemia 2014;28:485–496. [PubMed: 24220273]
129. Stumpel DJ, et al. Specific promoter methylation identifies different subgroups of MLL-rearranged infant acute lymphoblastic leukemia, influences clinical outcome, and provides therapeutic options. Blood 2009;114:5490–5498. [PubMed: 19855078]
130. Crider KS, Yang TP, Berry RJ, Bailey LB. Folate and DNA methylation: a review of molecular mechanisms and the evidence for folate's role. Adv Nutr 2012;3:21–38. [PubMed: 22332098]
131. Blom HJ, Smulders Y. Overview of homocysteine and folate metabolism. With special references to cardiovascular disease and neural tube defects. J Inherit Metab Dis 2011;34:75–81. [PubMed: 20814827]
132. Michaud EJ, van Vugt MJ, Bultman SJ, Sweet HO, Davisson MT, Woychik RP. Differential expression of a new dominant agouti allele (Aiapy) is correlated with methylation state and is influenced by parental lineage. Genes Dev 1994;8:1463–1472. [PubMed: 7926745]
133. Wolff GL, Kodell RL, Moore SR, Cooney CA. Maternal epigenetics and methyl supplements affect agouti gene expression in Avy/a mice. FASEB J 1998;12:949–957. [PubMed: 9707167]

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

134. Anderson OS, Sant KE, Dolinoy DC. Nutrition and epigenetics: an interplay of dietary methyl donors, one-carbon metabolism and DNA methylation. J Nutr Biochem 2012;23:853–859. [PubMed: 22749138]

135. Ly A, Hoyt L, Crowell J, Kim YI. Folate and DNA methylation. Antioxid Redox Signal 2012;17:302–326. [PubMed: 22332737]

136. Barua S, Kuizon S, Chadman KK, Flory MJ, Brown WT, Junaid MA. Single-base resolution of mouse offspring brain methylome reveals epigenome modifications caused by gestational folic acid. Epigenetics Chromatin 2014;7:3. [PubMed: 24484737]

137. Rakyan VK, Blewitt ME, Druker R, Preis JI, Whitelaw E. Metastable epialleles in mammals. Trends Genet 2002;18:348–351. [PubMed: 12127774]

138. Dominguez-Salas P, et al. Maternal nutrition at conception modulates DNA methylation of human metastable epialleles. Nat Commun 2014;5:3746. [PubMed: 24781383]

139. McDonough MA, Loenarz C, Chowdhury R, Clifton IJ, Schofield CJ. Structural studies on human 2-oxoglutarate dependent oxygenases. Curr Opin Struct Biol 2010;20:659–672. [PubMed: 20888218]

140. Minor EA, Court BL, Young JI, Wang G. Ascorbate induces ten-eleven translocation (Tet) methylcytosine dioxygenase-mediated generation of 5-hydroxymethylcytosine. J Biol Chem 2013;288:13669–13674. [PubMed: 23548903]

141. Dickson KM, Gustafson CB, Young JI, Züchner S, Wang G. Ascorbate-induced generation of 5-hydroxymethylcytosine is unaffected by varying levels of iron and 2-oxoglutarate. Biochem Biophys Res Commun 2013;439:522–527. [PubMed: 24021282]

142. Yin R, et al. Ascorbic acid enhances Tet-mediated 5-methylcytosine oxidation and promotes DNA demethylation in mammals. J Am Chem Soc 2013;135:10396–10403. [PubMed: 23768208]

143. Blaschke K, et al. Vitamin C induces Tet-dependent DNA demethylation and a blastocyst-like state in ES cells. Nature 2013;500:222–226. [PubMed: 23812591]

144. Chen J, et al. Vitamin C modulates TET1 function during somatic cell reprogramming. Nat Genet 2013;45:1504–1509. [PubMed: 24162740]

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

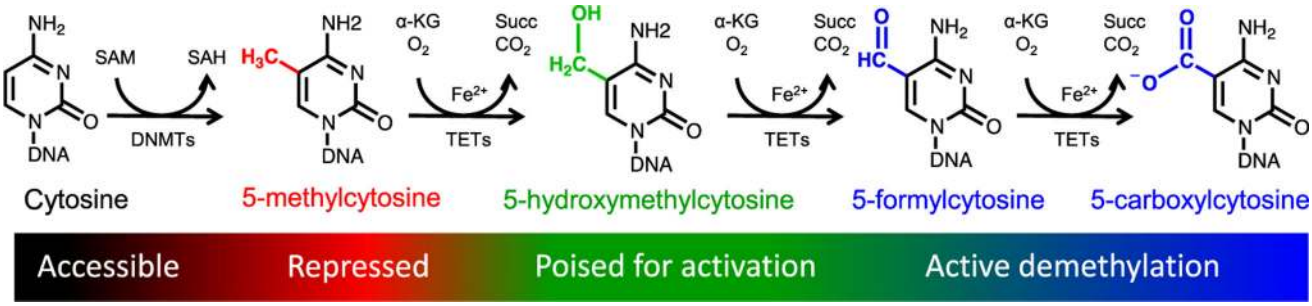

Fig. 1. Cytosine modification pathway.

The cytosine modification pathway from unmodified cytosine to carboxylcytosine is depicted. The most common and understood function of each species of cytosine is described below. Additional functions of 5-mC and 5-hmC are described in the text.

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

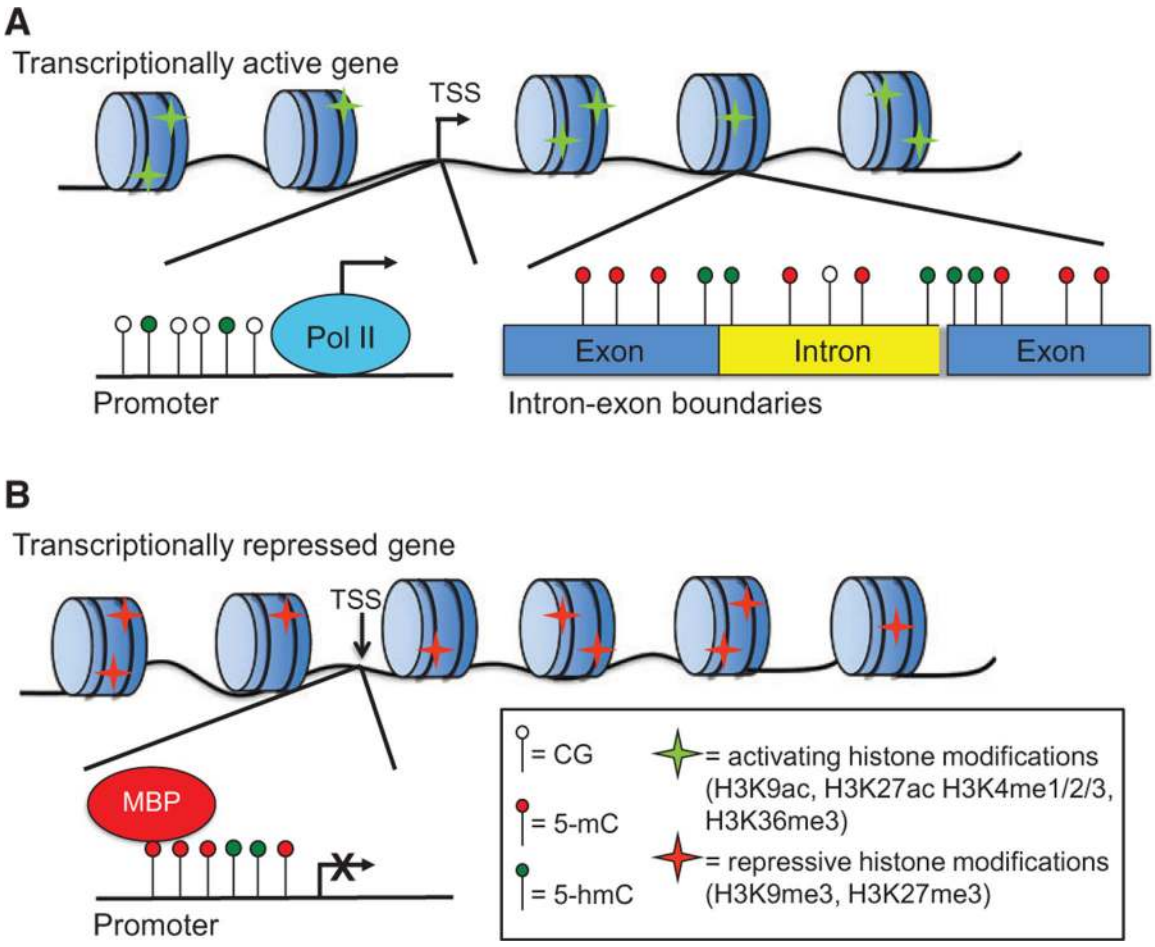

Fig. 2. Cytosine modification patterns for active versus repressed genes.

(A) Actively transcribed genes generally have a promoter region that is nucleosome-depleted and contains unmodified or hydroxymethylated CpGs. 5-mC and 5-hmC patterns at exon-intron boundaries are shown. (B) Repressed genes have a highly methylated promoter with more densely packed nucleosomes to prevent transcription.

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

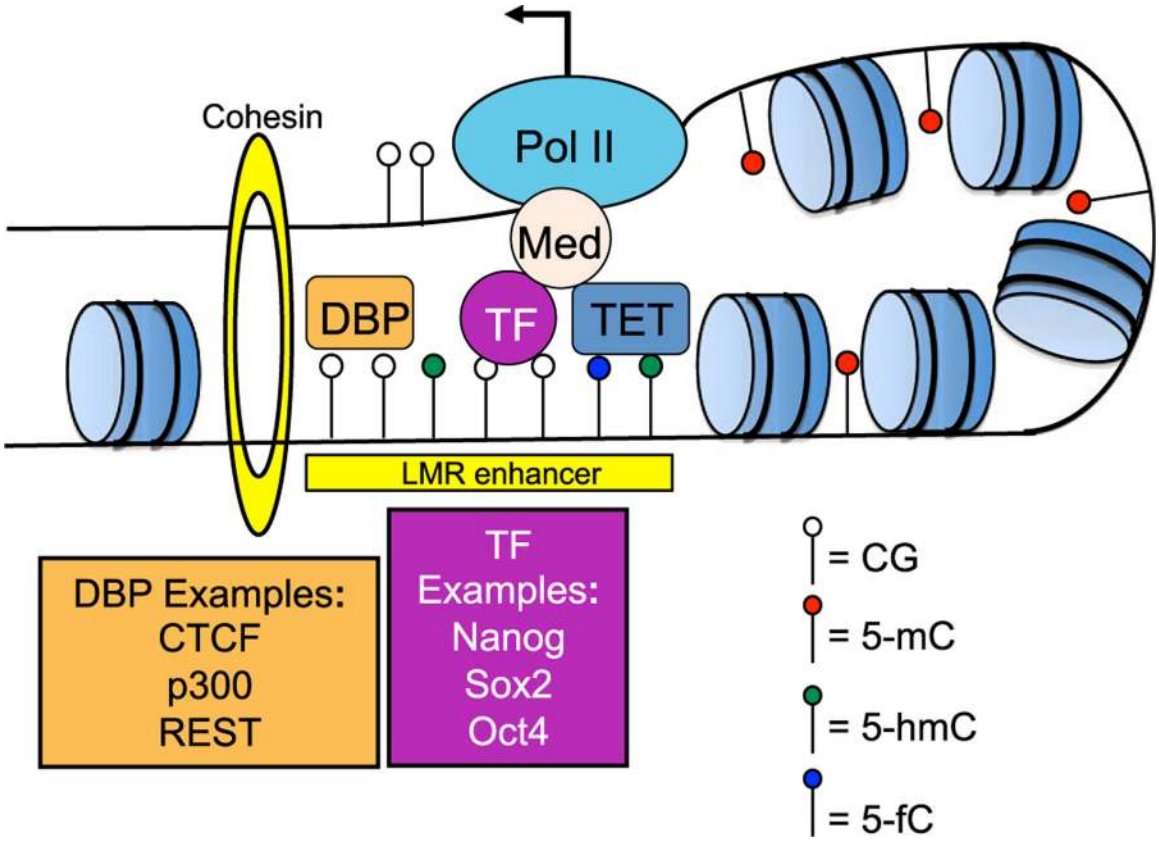

Fig. 3. Cytosine modifications at enhancer elements.

5-hmC at enhancer elements maintains an open chromatin state, facilitating access to DNA binding proteins (DBP), transcription factors (TF), and mediator proteins (Med) that are essential for transcription of target genes.

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

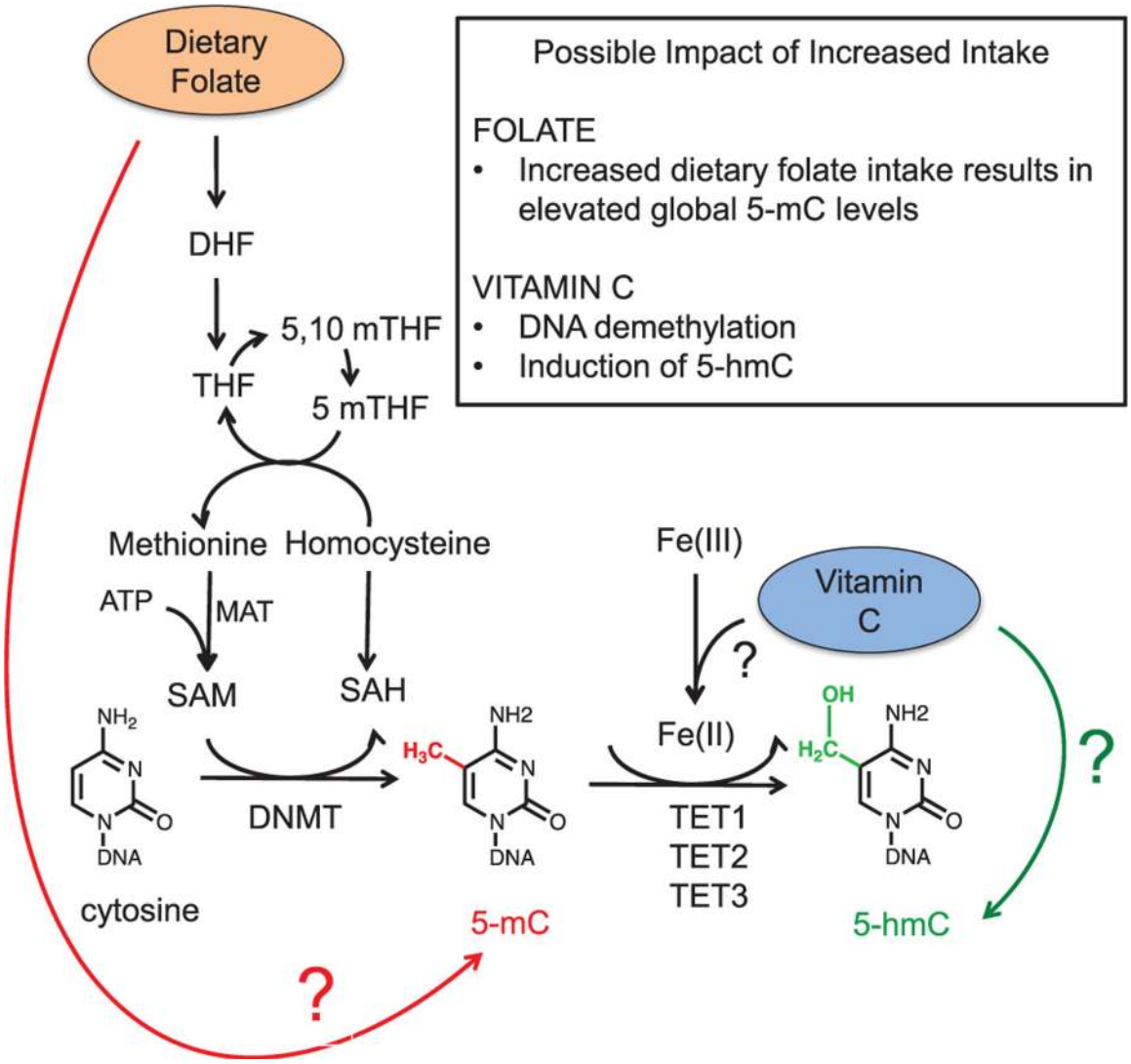

Fig. 4. Impact of folate and vitamin C on 5-mC and 5-hmC.

Dietary folate plays a role in one-carbon metabolism to affect levels of 5-mC. Vitamin C levels positively correlate with 5-hmC levels through mechanisms that are currently not completely understood. DHF, dihydrofolate; THF, tetrahydrofolate; 5,10mTHF, 5,10-methylenetetrahydrofolate; 5mTHF, 5-methylenetetrahydrofolate; ATP, adenosine triphosphate; MAT, S-adenosylmethionine transferase; SAM, S-adenosylmethionine; SAH, S-adenosylhomocysteine.