Published in final edited form as:

Curr Opin Chem Biol. 2012 December ; 16(5-6): 516–524. doi:10.1016/j.cbpa.2012.10.002.

Nucleic Acid Modifications with Epigenetic Significance

Ye Fu and Chuan He

Department of Chemistry and Institute for Biophysical Dynamics, The University of Chicago, 929 East 57th Street, Chicago, Illinois 60637, USA.

Abstract

Epigenetic modifications influence gene expression without alterations to the underlying nucleic acid sequence. In addition to the well-known 5-methylcytosine (5mC), 5-hydroxymethylcytosine (5hmC), 5-formylcytosine (5fC), and 5-carboxycytosine (5caC) have recently been discovered in genomic DNA, which all result from iterative oxidation of 5mC by the TET (Ten-Eleven-Translocase) family of enzymes. Recent studies have proposed the roles of these oxidized cytosines in mediating active demethylation of 5mC. Through affinity-based genome-wide sequencing and oxidation-assisted base-resolution sequencing methods, 5hmC is found to be dynamically regulated during development, and is enriched mainly in distal regulatory elements in human and mouse embryonic cells. Among RNA modifications, *N*⁶-methyladenosine (*m*⁶A) is a widespread yet poorly-studied base modification in mRNA and non-coding RNA. The recent discovery that *m*⁶A in RNA is the major substrate of the fat mass and obesity associated (FTO) protein draws attention to the potential regulatory functions of reversible RNA methylations, which can be dynamic, and could be important in many fundamental cellular functions.

Introduction

Epigenetics refers to the inheritable changes of gene expression and cellular identity caused by mechanisms other than alterations that occur to the underlying genetic sequence [1]. DNA modification and histone modifications are two well-known categories of chemical modifications that epigenetically control cell lineage and cell fate [2]. For instance, 5-methylcytosine (5mC) controls pluripotency status and cell fate mainly through its repressive roles [3]. Active demethylation of 5mC happens after fertilization to retrieve the pluripotential state of embryonic stem cells through as yet undefined mechanisms [David2] [4]. The recently discovered [David3] 5-hydroxymethylcytosine (5hmC), 5-formylcytosine (5fC), and 5-carboxycytosine (5caC) open a new paradigm of active DNA demethylation through 5mC oxidation (Figure 1[David4]) [5-9]. Other than serving as potential demethylation intermediates, the abundance of these versatile modifications is also regulated, and these modifications can potentially sculpt the epigenetic landscape of a cell by affecting gene expression. In contrast to well-recognized DNA and histone modifications, the epigenetic function of RNA modifications has just begun to emerge. Our effort in searching for reversible RNA modifications led to the discovery that *N*⁶-methyladenosine (*m*⁶A) can be demethylated by the fat mass and obesity associated (FTO) protein (Figure 2) [David5], which suggests that reversible RNA methylation may also function as an

© 2012 Elsevier Ltd. All rights reserved.

Corresponding author: He, Chuan (chuanhe@uchicago.edu).

Publisher's Disclaimer: This is a PDF file of an unedited manuscript that has been accepted for publication. As a service to our customers we are providing this early version of the manuscript. The manuscript will undergo copyediting, typesetting, and review of the resulting proof before it is published in its final citable form. Please note that during the production process errors may be discovered which could affect the content, and all legal disclaimers that apply to the journal pertain.

“epigenetic marker” by possibly regulating not only mRNA but also non-coding RNAs that play roles in gene expression regulation [David6]. Here we summarize the current understanding on the demethylation pathways of 5mC in DNA by TET, and m⁶A in RNA by FTO, respectively; we also briefly summarize the newly developed sequencing methods for 5hmC in DNA and m⁶A in RNA, and discuss new insights revealed.

Involvement of 5hmC, 5fC, and 5caC in the Active DNA Demethylation Pathway

5mC is an important epigenetic marker that controls transcription dynamics and maintains cell identity. 5mC marks a repression region by recruiting 5mCpG-selective binding proteins, and functions in gene transcription regulation, genomic imprinting, and retrotransposon silencing. While *de novo* methylation is catalyzed by DNA methyltransferase 3A (DNMT3A) and DNMT3B, replication-dependent methylation is maintained by DNMT1 (Figure 1a) [3]. In order to reprogram the epigenetic marker to regain totipotency or pluripotency, mammalian cells undergo at least two waves of global DNA demethylation after fertilization and during germ cell development [10]. 5mC can be lost during DNA replication at the newly synthesized DNA strand, termed as replication-dependent passive demethylation; or it could be removed in an active demethylation pathway independent of replication.

After fertilization, 5mC in the maternal genome is lost mainly through passive demethylation. 5mC in the paternal genome is lost rapidly in several hours, which may suggest the existence of a global active demethylation pathway [4]. In primordial germ cells, 5mC can be lost in both active and passive demethylation pathways [11]. Recently, it was shown that various pathways of active DNA demethylation may exist in different cell types or specific DNA regions.

Ten-Eleven-Translocate proteins (TET1-3), which were identified as a fusion partner of the MLL gene in acute myeloid leukemia (AML), are mammalian homologs of trypanosomal thymidine hydroxylase JBP1 and JBP2 [5]. They contain a conserved α -ketoglutarate (α -KG) oxygenase double-stranded beta-helix (DSBH) core fold, a CXXC-type zinc-binding domain, and a cysteine-rich region located on the N-terminal of the DSBH domain. With the detection of substantial amounts of 5hmC in genomic DNA from mouse brain and embryonic stem cells [5,6], TET1 was first shown to catalyze the conversion of 5mC to 5hmC in cultured cells and *in vitro* in a α -KG/Fe(II)-dependent manner [5]. It was later [David7] reported that TET proteins can also enzymatically generate 5fC and 5caC from 5mC *in vitro* and *in vivo* [7-9]. Non-heme Fe(II)/ α -KG-dependent dioxygenases typically utilize an iron(II) center to activate dioxygen and hydroxylate an inert C-H bond of the substrate [12]. Since hydroxylation is a two-electron oxidation, α -KG serves as a co-factor to provide the other two electrons needed for the four-electron reduction of oxygen. The reaction mechanism involves two reaction phases: the first phase of reaction is the activation of dioxygen to form an iron(IV)-oxo intermediate (Figure 1b) [13]; the second phase is the substrate oxidation step, in which iron(IV)-oxo abstracts an H atom of the target C-H bond to eventually yield a hydroxylated product. 5mC is oxidized to 5hmC, 5fC, and 5caC stepwise by TET proteins using the similar mechanism (Figure 1b). It is proposed that the hydrate form of 5fC is generated upon oxidation of 5hmC, and this hydrate form is the real substrate for the oxidation of 5fC to 5caC (Figure 1b) [7]. These oxidation products are also found in DNA isolated from mouse embryonic stem (ES) cells and mouse organs [7-9]. While 5fC and 5caC are found to be specifically recognized and excised by thymine-DNA glycosylase (TDG) [9,14, 15], depletion of TDG in mouse ES cells leads to accumulation of 5caC. These data suggest that oxidation of 5mC by Tet proteins (mainly Tet1 and Tet2 in mouse ES cells), followed by TDG-mediated base excision of 5fC and 5caC and subsequent base excision repair (BER), constitutes a pathway for active DNA demethylation (Figure 1a) [9].

Oxidation of 5mC to 5hmC, 5fC, and 5caC by Tet3 is involved in the active demethylation pathway in zygotic paternal DNA [16, 17]. The conversion of 5mC to 5hmC, 5fC, and 5caC leads to the rapid loss of 5mC in the paternal genome. Recently, these oxidation products have been thought of as diluted in a passive replication-dependent manner during preimplantation development. The presence of 5hmC, 5fC, and 5caC indicates they could be functionally important at this stage [17, 18].

Other than oxidation to 5caC, deamination of 5hmC to 5hmU by Apobec1 or AID, followed by base excision repair (BER) repair has also been suggested [David8] as a potential active demethylation pathway in brain and primordial germ cells (Figure 1a) [19]. Tet1 and Apobec1 are shown to be involved in neuronal activity-induced, region-specific, active DNA demethylation in the dentate gyrus of the adult mouse brain *in vivo* [19]. In mouse primordial germ cells, AID deficiency affects the genome-wide erasure of DNA methylation, which suggests that the deamination of 5hmC to 5hmU could be involved in active demethylation in primordial germ cells [20]. The presence of 5hmU in the mammalian genome still needs confirmation, and AID/APOBEC deaminases are recently shown to disfavor modified cytosines [21], however.

Having revealed these various pathways of methylation and demethylation, it is still very possible that other pathways may exist, in particular, the global demethylation of paternal genome during imprinting is still an unsolved mystery.

Sequencing methods for genome-wide analysis of 5hmC

An understanding of the exact biological functions of 5hmC requires knowledge of the precise locations of this base modification. Several affinity-based methods have been developed with resolutions of several hundred to one thousand bases [22-30]. In these methods, after 5hmC in DNA is labeled and immunoprecipitated, DNA is sequenced by high-throughput parallel sequencing. Several groups [David9] including ours have used T4 bacteriophage β -glucosyltransferase (β -GT) to label 5hmC [23-25]. We showed that β -GT can transfer either normal glucose or an engineered glucose moiety containing an azide group onto the hydroxyl group of 5-hmC. The azide group is then chemically modified with biotin for detection, affinity enrichment, and sequencing [25]. When a regular glucose is transferred to 5hmC by β -GT, it can be selectively oxidized by periodate oxidation, and labeled by biotin [23], or be selectively pulled down with J-binding protein 1 coupled to magnetic beads [24]. Treatment of genomic DNA with sodium bisulphite can also convert 5hmC to cytosine 5-methylenesulphonate (CMS), in which immunoprecipitation can be performed with a specific antiserum against CMS [23]. Bio-orthogonal chemistry and nucleic acid chemistry have played major roles in these new approaches. Newly developed 5hmC antibody has also helped to develop hydroxymethylated DNA immunoprecipitation [26, 27].

The current lack of single-base resolution sequencing methods capable of detecting the relative abundance of 5hmC at specific cytosine sites hampers a more detailed understanding of 5hmC's function. The bisulfite-sequencing method has been used as the gold standard in determining the methylation status with single-base resolution in genomic DNA [2]. Treatment of DNA with sodium bisulfite followed by PCR leads to the conversion of cytosine to thymine through hydrolytic deamination. However, 5mC remains as cytosine in this reaction. It was shown that 5hmC resists deamination similar to 5mC under traditional bisulfite conditions, while 5fC and 5caC exhibit deamination similar to unmodified cytosine after bisulfite treatment. Taking advantage of these different properties, two groups including us have developed base-resolution sequencing methods for 5hmC, namely Tet-assisted bisulfite sequencing (TAB-Seq) (Figure 3a) [31] and oxidative bisulfite sequencing (oxBS-Seq) [32], respectively (Figure 3b). In TAB-Seq, after protection of

5hmC with glucose by β -GT, 5mC is converted to 5caC by a Tet enzyme. Both C and 5caC are deaminated to U/5caU upon bisulfite treatment, while glucose-protected 5hmC remains as C in subsequent sequencing. This method directly reads 5hmC with single-base resolution and relative abundance at the modification site. In reference to traditional bisulfite data, which provides a sum of 5mC+5hmC, the position and relative abundance of 5mC at each modification site can be obtained. In oxBS-Seq, selective chemical oxidation of 5hmC to 5-formylcytosine (5fC) by KRuO_4 enables subsequent conversion of 5fC to U upon bisulfite treatment. Then, the absolute position and abundance of 5mC can be obtained through sequencing. The comparison of the sequence with traditional bisulfite sequencing result, in which 5hmC remains unchanged, will yield the position and amount of 5hmC.

Distribution and function of 5hmC

Brain tissue contains the highest absolute amount of 5hmC in the genome and accumulates during brain development [25, 33]. The level of 5hmC in ES[David10] cells is also high and 5hmC has been proposed to play roles in differentiation [22, 23]. With affinity-based approaches, 5hmC is shown to be enriched at promoters, enhancers, CCCTC-binding factor (CTCF)[David11]-binding sites, gene bodies, and exons, suggesting a role for this modification in gene regulation [22-30]. The balance between pluripotency and lineage commitment is linked with the balance between 5hmC and 5mC in the genome [22]. 5hmC also contributes to the 'poised' chromatin signature with dual histone 3 lysine 27 trimethylation (H3K27me3) and histone 3 lysine 4 trimethylation (H3K4me3) marks found at developmentally-regulated genes in ES cells [23].

Given the increased [David12]sensitivity of TAB-Seq and oxBS-Seq, we can create a more detailed map of all the 5hmC sites inside cells [31, 32]. Although 5mC occurs in three nucleotide sequence contexts -- CG, CHG, and CHH -- nearly all 5hmCs exist in the CG context. Distal-regulatory elements including p300-binding sites, predicted enhancers, CTCF-binding sites, and DNase I hypersensitive sites are more enriched with 5hmC than at other genic regions (Figure 3c), where 5mC are correspondingly lower than the genome average. In transcription starting sites (TSSs), low steady-state levels of 5hmC are observed, which can be explained by the high turnover of those 5mC in TSSs due of the enrichment of TET1 at these regions. And high levels of 5hmC and reciprocally low levels of 5mC are observed near but not on transcription factor-binding sites. Although 5hmC and 5mC often occur together at the same position, the relative abundance of 5hmC varies significantly among distinct functional sequence elements.

More sequencing depth is required for base-resolution methods to detect 5hmC at relatively low abundance. Genic 5hmC likely exists at low abundance (around or less than 5%); currently used approaches would therefore not have detected it. The presence of many of these low abundance sites highlights the biases in affinity-based 5hmC mapping, which can amplify frequent but weak signals found in gene bodies to overshadow rare but stronger ones at distal-regulatory elements.

Current studies raise novel topics including the mechanism of regulating 5hmC dynamics and sequence-specificity by TETs. Regulation of TET enzyme, together with any sequence-specific factors directing the TET machinery to a specific genomic region, will direct 5hmC deposition at specific sites. From a kinetic point of view, the conversion of 5mC to 5hmC catalyzed by TET enzymes is faster than the conversion of 5hmC to 5fC and 5caC [8]. The steady-state distribution of modifications at a single site would be largely dependent on its accessibility to TET enzymes. A site with a high steady-state percentage of 5hmC could be a result of different TET-catalyzed rates of conversion of 5mC to 5hmC as compared to 5hmC to 5fC. Or, a subsequent conformation change of the genome structure may block further oxidation by the TET enzyme after 5hmC is deposited. The heterogeneity of epigenetic

markers at a single site among the same cell population raises the question of how these cells orchestrate population behavior during different biological processes.

Demethylation of m⁶A in RNA

Despite extensive knowledge gains on the epigenetic role of DNA modifications, only recently have we turned our attention to the functional role of internal modifications in RNA in epigenetics and epitranscriptomics. Extensive modifications and editing events have been documented in various coding and non-coding RNA species, although most of their functions remain elusive. In contrast to the conventional static view of these RNA modifications, more recent evidence reveals their dynamic nature. Epigenetic RNA regulation can ensure rapid responses to input signals by avoiding the slow kinetics of transcription regulation. Following the idea that RNA modifications are potentially reversible, we made the recent discovery that the FTO protein is an m⁶A RNA demethylase. The development of a sequencing method for m⁶A in RNA came next. Indeed, our work on reversible modifications may have already initiated a new era of focus on reversible RNA modifications in gene expression regulation [35].

N⁶-methyladenosine (m⁶A) is a widespread modification in mRNA, snRNA, and rRNA[36]. Although widespread in the transcriptome, representing 0.1%-0.4% of total adenosine residues, the distribution of this modification, which has been found in mRNA in a consensus sequence [G/A/U][G>A]m⁶AC[U>A>C] [37-39], is poorly understood. Partial purification of the methyltransferase reveals a large complex of about 800kD that installs m⁶A in mRNA, with a catalytic unit binding to a SAM co-factor [40]. Inducer of meiosis 4 (IME4), which is part of the methyltransferase in yeast and *Drosophila*, is crucial in the induction of yeast sporulation, and controls development and gametogenesis in *Drosophila* [41-43]. It was shown that m⁶A-modified RNA may be involved in the Notch signaling pathway during soma-germ line interactions in *Drosophila* [44].

FTO has been found to affect human obesity and energy homeostasis in both genomic and mouse model studies [45-48], and more recently its genotype has been associated with phenotypic variability of body mass index [49]. FTO belongs to the non-heme Fe(II)- and α -KG-dependent dioxygenase AlkB family proteins, which also includes the human AlkB homologs ALKBH1-8 [50]; some of these enzymes catalyze hydroxylation of a methyl group attached to a nitrogen atom to afford a hemiaminal intermediate which decomposes to eliminate the methyl group (Figure 2b) [51]. FTO has previously been shown to oxidatively demethylate m³T and m³U in single-stranded DNA (ssDNA) and single-stranded RNA (ssRNA) *in vitro*, although yielding exceedingly low activities. The crystal structure of FTO reveals an extra loop, which serves as the basis for its preference for single-stranded nucleic acids as substrate. We discovered that FTO efficiently demethylates RNA oligonucleotides containing m⁶A *in vitro* (Figure 2a). Knockdown of FTO in HeLa and 293FT cell lines leads to an increase of m⁶A level in mRNA, and *vice versa*. FTO is partially localized in nuclear speckles, in which methyltransferase likely resides. FTO and the methyltransferase may therefore coordinately regulate the dynamic methylation status of mRNA.

Although[David13] m⁶A in mRNA has been known for more than forty years, understanding of the function of m⁶A has been hampered by a lack of an effective sequencing method. Recently, two groups have developed an antibody-based, affinity-purification method in order to reveal the topology of the human and mouse transcriptome-wide m⁶A modification landscape [38, 39]. A comprehensive study showed that 12,000 m⁶A sites in the transcripts of more than 7,000 human genes are identified [David14][38]. This modification shows a preference around stop codons, in 3'-UTRs and within long internal exons, which are highly conserved in both human and mouse. Some of them are stimulus-dependent and dynamically modulated by interferon- γ treatment. [38]. An

intriguing finding indicates that the most highly expressed miRNAs have a greater chance of targeting transcripts that contain m⁶A in their 3'-UTR, which suggests that miRNA levels may control methylation of their target transcripts [39]. In addition, several putative m⁶A-binding proteins have been identified [35].

As facilitated by the m⁶A-binding protein, functions of m⁶A in internal mRNA may include tuning pre-mRNA processing, mRNA trafficking, translation, and stability. On the other hand, the methylation at a specific region may disrupt the binding of the RNA-binding protein on the unmethylated sequences. This layer of regulation may partially explain the lack of correlation between mRNA copy numbers and protein abundance.

In addition to mRNA, a large portion of poly(A) RNA is long non-coding RNA (lncRNA) [52]. Recent studies have revealed that regulation of genome activity by non-coding RNA makes RNA another constituent of the eukaryotic genome structure, in interplay with histones [53, 54]. The finding that m⁶A is also widespread in lncRNA raises the possibility that RNA modification may contribute to genome structure. Potential dynamic regulation of m⁶A affecting the function of non-coding RNA is an intriguing question to explore. Demethylation of m⁶A could fine-tune lncRNA function [David15] in a rapid manner in response to environmental signals because it bypasses transcription, which can be beneficial in various signaling events.

Because the methylation in RNA is not stoichiometric, bulk analysis may not reveal the dynamics of a single RNA transcript. A high-throughput method for quantitative and single-base resolution sequencing is highly demanded for further understanding of its dynamics and functional role. To understand how a single m⁶A site in a single RNA may affect its life cycle and metabolism will be facilitated by single-molecule RNA imaging and high-throughput sequencing. RNA has been shown to be potentially inheritable in mouse zygotes [55] and epi-transcriptome modifications could also be potentially inheritable, particularly with its effects on non-coding RNAs and even pre-miRNAs, which may in turn affect the genetics of the offspring. This topic could be further explored as RNA epigenetics [56].

Conclusions and outlook

In summary, recent insights into the regulation of DNA and RNA epigenetics have shown great potential to promote further understanding of biological regulatory systems. Oxidation of 5mC to 5hmC, 5fC, and 5caC by TET proteins has been shown to be important for active demethylation of genomes. Oxidative demethylation of m⁶A in RNA by FTO suggested a dynamic and regulatory role for m⁶A in RNA. All these recent discoveries add multiple layers of epigenetic regulation into the known epigenetic blueprint. Participation of Fe(II)/ α -KG-dependent dioxygenases in oxidative manipulation of these various DNA and RNA base modifications makes them important factors for tuning epigenetic regulation. Further exploration of the mechanisms underlying these enzymes will enable our understanding of their selectivity and explore small-molecule tuning of their activity. Application of the new sequencing methods to map these modifications in cells at different temporal and spatial points will provide additional mechanistic insights into the underlying dynamics that regulate epigenetic status. Other methods, such as live cell single molecular imaging, will help characterize the functions of these modifications. Chemical biology will play major roles in these future developments.

Acknowledgments

The authors thank the US National Institutes of Health (NIH) for support (GM071440, GM088599) and a Catalyst Award from the Chicago Biomedical Consortium with support from the Searle Funds at The Chicago Community

Trust. We thank Chun-Xiao Song and Miao Yu for helpful discussions. We thank Sarah F. Reichard, MA for editing the manuscript.

References and recommended reading

Papers of particular interest, published within the period of review, have been highlighted as:

■ of special interest

■■ of outstanding interest

1. Bird A. Perceptions of epigenetics. *Nature*. 2007; 447:396–398. [PubMed: 17522671]
2. Fuks F. DNA methylation and histone modifications: teaming up to silence genes. *Curr Opin Genet Dev*. 2005; 15:490–495. [PubMed: 16098738]
3. Sasaki H, Matsui Y. Epigenetic events in mammalian germ-cell development: reprogramming and beyond. *Nat Rev Genet*. 2008; 9:129–140. [PubMed: 18197165]
4. Reik W, Dean W, Walter J. Epigenetic reprogramming in mammalian development. *Science*. 2001; 293:1089–1093. [PubMed: 11498579]
5. Tahiliani M, Koh KP, Shen Y, Pastor WA, Bandukwala H, Brudno Y, Agarwal S, Iyer LM, Liu DR, Aravind L, et al. Conversion of 5-methylcytosine to 5-hydroxymethylcytosine in mammalian DNA by MLL partner TET1. *Science*. 2009; 324:930–935. [PubMed: 19372391]
6. 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]
- 7■ Pfaffeneder T, Hackner B, Truss M, Munzel M, Muller M, Deiml CA, Hagemeyer C, Carell T. The discovery of 5-formylcytosine in embryonic stem cell DNA. *Angew Chem Int Ed Engl*. 2011; 50:7008–7012. [PubMed: 21721093] [This paper discovered 5-formylcytosine in embryonic stem cell genomic DNA.]
- 8■■ Ito S, Shen L, Dai Q, Wu SC, Collins LB, Swenberg JA, He C, Zhang Y. Tet proteins can convert 5-methylcytosine to 5-formylcytosine and 5-carboxylcytosine. *Science*. 2011; 333:1300–1303. [PubMed: 21778364] [This paper discovered the formation of 5fC and 5caC catalyzed by Tet, it should be read alongside Reference [11■■].]
- 9■■ He YF, Li BZ, Li Z, Liu P, Wang Y, Tang Q, Ding J, Jia Y, Chen Z, Li L, et al. Tet-mediated formation of 5-carboxylcytosine and its excision by TDG in mammalian DNA. *Science*. 2011; 333:1303–1307. [PubMed: 21817016] [Published at the same time of Reference [10■■], this paper describes the formation and excision of 5caC in DNA.]
10. Wu SC, Zhang Y. Active DNA demethylation: many roads lead to Rome. *Nat Rev Mol Cell Biol*. 2010; 11:607–620. [PubMed: 20683471]
11. Hackett JA, Zyllicz JJ, Surani MA. Parallel mechanisms of epigenetic reprogramming in the germline. *Trends Genet*. 2012; 28:164–174. [PubMed: 22386917]
12. Solomon EI, Decker A, Lehnert N. Non-heme iron enzymes: contrasts to heme catalysis. *Proc Natl Acad Sci U S A*. 2003; 100:3589–3594. [PubMed: 12598659]
13. Krebs C, Galonic Fujimori D, Walsh CT, Bollinger JM Jr. Non-heme Fe(IV)-oxo intermediates. *Acc Chem Res*. 2007; 40:484–492. [PubMed: 17542550]
14. Maiti A, Drohat AC. Thymine DNA glycosylase can rapidly excise 5-formylcytosine and 5-carboxylcytosine: potential implications for active demethylation of CpG sites. *J Biol Chem*. 2011; 286:35334–35338. [PubMed: 21862836]
- 15■ Zhang L, Lu X, Lu J, Liang H, Dai Q, Xu GL, Luo C, Jiang H, He C. Thymine DNA glycosylase specifically recognizes 5-carboxylcytosine-modified DNA. *Nat Chem Biol*. 2012; 8:328–330. [PubMed: 22327402] [This structure paper describes specific 5caC recognition by mammalian TDG.]
- 16■ Gu TP, Guo F, Yang H, Wu HP, Xu GF, Liu W, Xie ZG, Shi L, He X, Jin SG, et al. The role of Tet3 DNA dioxygenase in epigenetic reprogramming by oocytes. *Nature*. 2011; 477:606–610. [PubMed: 21892189] [This paper reported the function of Tet3 in DNA demethylation following fertilization.]

17. Inoue A, Shen L, Dai Q, He C, Zhang Y. Generation and replication-dependent dilution of 5fC and 5caC during mouse preimplantation development. *Cell Res.* 2011; 21:1670–1676. [PubMed: 22124233]
18. Inoue A, Zhang Y. Replication-dependent loss of 5-hydroxymethylcytosine in mouse preimplantation embryos. *Science.* 2011; 334:194. [PubMed: 21940858]
19. 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]
20. Popp C, Dean W, Feng S, Cokus SJ, Andrews S, Pellegrini M, Jacobsen SE, Reik W. Genome-wide erasure of DNA methylation in mouse primordial germ cells is affected by AID deficiency. *Nature.* 2010; 463:1101–1105. [PubMed: 20098412]
21. Nabel CS, Jia H, Ye Y, Shen L, Goldschmidt HL, Stivers JT, Zhang Y, Kohli RM. AID/APOBEC deaminases disfavor modified cytosines implicated in DNA demethylation. *Nat Chem Biol.* 2012; 8:751–758. [PubMed: 22772155]
22. Ficiz G, Branco MR, Seisenberger S, Santos F, Krueger F, Hore TA, Marques CJ, Andrews S, Reik W. Dynamic regulation of 5-hydroxymethylcytosine in mouse ES cells and during differentiation. *Nature.* 2011; 473:398–402. [PubMed: 21460836]
23. Pastor WA, Pape UJ, Huang Y, Henderson HR, Lister R, Ko M, McLoughlin EM, Brudno Y, Mahapatra S, Kapranov P, et al. Genome-wide mapping of 5-hydroxymethylcytosine in embryonic stem cells. *Nature.* 2011; 473:394–397. [PubMed: 21552279]
24. Robertson AB, Dahl JA, Ougland R, Klungland A. Pull-down of 5-hydroxymethylcytosine DNA using JBP1-coated magnetic beads. *Nat Protoc.* 2012; 7:340–350. [PubMed: 22281869]
25. Song CX, Szulwach KE, Fu Y, Dai Q, Yi C, Li X, Li Y, Chen CH, Zhang W, Jian X, et al. Selective chemical labeling reveals the genome-wide distribution of 5-hydroxymethylcytosine. *Nat Biotechnol.* 2011; 29:68–72. [PubMed: 21151123]
26. Jin SG, Wu X, Li AX, Pfeifer GP. Genomic mapping of 5-hydroxymethylcytosine in the human brain. *Nucleic Acids Res.* 2011; 39:5015–5024. [PubMed: 21378125]
27. 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]
28. Williams K, Christensen J, Pedersen MT, Johansen JV, Cloos PA, Rappaport J, Helin K. TET1 and hydroxymethylcytosine in transcription and DNA methylation fidelity. *Nature.* 2011; 473:343–348. [PubMed: 21490601]
29. Wu H, D'Alessio AC, Ito S, Wang Z, Cui K, Zhao K, Sun YE, Zhang Y. 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]
30. Xu Y, Wu F, Tan L, Kong L, Xiong L, Deng J, Barbera AJ, Zheng L, Zhang H, Huang S, et al. Genome-wide regulation of 5hmC, 5mC, and gene expression by Tet1 hydroxylase in mouse embryonic stem cells. *Mol Cell.* 2011; 42:451–464. [PubMed: 21514197]
31. Yu M, Hon GC, Szulwach KE, Song CX, Zhang L, Kim A, Li X, Dai Q, Shen Y, Park B, et al. Base-Resolution Analysis of 5-Hydroxymethylcytosine in the Mammalian Genome. *Cell.* 2012; 149:1368–1380. [PubMed: 22608086] [This paper invented TAB-Seq as a base-resolution sequencing method for 5hmC, and mapped the 5hmC sites in mouse and human ES cells. It shows nearly all 5hmC is in CpG context, and 5hmC is mostly enriched in distal regulatory elements. It also shows 5hmC is strand asymmetric and prefers G-rich strand. It can be read together with Reference [30].]
32. Booth MJ, Branco MR, Ficiz G, Oxley D, Krueger F, Reik W, Balasubramanian S. Quantitative sequencing of 5-methylcytosine and 5-hydroxymethylcytosine at single-base resolution. *Science.* 2012; 336:934–937. [PubMed: 22539555] [This paper developed oxBS-seq method to map and quantify 5hmC (from quantifying 5mC first) at CpG islands in mouse ES cells. It directly reads out 5hmC at single-base resolution with absolute abundance.]
33. Munzel M, Globisch D, Bruckl T, Wagner M, Welzmler V, Michalakakis S, Muller M, Biel M, Carell T. Quantification of the sixth DNA base hydroxymethylcytosine in the brain. *Angew Chem Int Ed Engl.* 2010; 49:5375–5377. [PubMed: 20583021]

34. Yildirim O, Li R, Hung JH, Chen PB, Dong X, Ee LS, Weng Z, Rando OJ, Fazzio TG. Mbd3/NURD complex regulates expression of 5-hydroxymethylcytosine marked genes in embryonic stem cells. *Cell*. 2011; 147:1498–1510. [PubMed: 22196727]
35. Jia G, Fu Y, Zhao X, Dai Q, Zheng G, Yang Y, Yi C, Lindahl T, Pan T, Yang YG, et al. N⁶-methyladenosine in nuclear RNA is a major substrate of the obesity-associated FTO. *Nat Chem Biol*. 2011; 7:885–887. [PubMed: 22002720] [This is the first paper reporting the m⁶A demethylase activity of FTO enzyme. It revealed that m⁶A in RNA could have regulatory functions.]
36. Horowitz S, Horowitz A, Nilsen TW, Munns TW, Rottman FM. Mapping of N6-methyladenosine residues in bovine prolactin mRNA. *Proc Natl Acad Sci U S A*. 1984; 81:5667–5671. [PubMed: 6592581]
37. Narayan P, Ludwiczak RL, Goodwin EC, Rottman FM. Context effects on N6-adenosine methylation sites in prolactin mRNA. *Nucleic Acids Res*. 1994; 22:419–426. [PubMed: 8127679]
38. Dominissini D, Moshitch-Moshkovitz S, Schwartz S, Salmon-Divon M, Ungar L, Osenberg S, Cesarkas K, Jacob-Hirsch J, Amariglio N, Kupiec M, et al. Topology of the human and mouse m⁶A RNA methylomes revealed by m⁶A-seq. *Nature*. 2012; 485:201–206. [PubMed: 22575960] [This is the first paper describing the antibody-based genome-wide sequencing of m⁶A in human and mouse RNA. It shows m⁶A is mostly in 3' UTRs and long exons.]
39. Meyer KD, Saletore Y, Zumbo P, Elemento O, Mason CE, Jaffrey SR. Comprehensive Analysis of mRNA Methylation Reveals Enrichment in 3' UTRs and near Stop Codons. *Cell*. 2012; 149:1635–1646. [PubMed: 22608085] [Published almost at the same time of Reference [35], applying the similar approach, it shows distribution of m⁶A in 3' UTRs, and the position of m⁶A is related to miRNA binding site.]
40. Bokar JA, Shambaugh ME, Polayes D, Matera AG, Rottman FM. Purification and cDNA cloning of the AdoMet-binding subunit of the human mRNA (N6-adenosine)-methyltransferase. *RNA*. 1997; 3:1233–1247. [PubMed: 9409616]
41. Clancy MJ, Shambaugh ME, Timpte CS, Bokar JA. Induction of sporulation in *Saccharomyces cerevisiae* leads to the formation of N6-methyladenosine in mRNA: a potential mechanism for the activity of the IME4 gene. *Nucleic Acids Res*. 2002; 30:4509–4518. [PubMed: 12384598]
42. Bodi Z, Button JD, Grierson D, Fray RG. Yeast targets for mRNA methylation. *Nucleic Acids Res*. 2010; 38:5327–5335. [PubMed: 20421205]
43. Agarwala SD, Blitzblau HG, Hochwagen A, Fink GR. RNA methylation by the MIS complex regulates a cell fate decision in yeast. *PLoS Genet*. 2012; 8:e1002732. [PubMed: 22685417]
44. Hongay CF, Orr-Weaver TL. *Drosophila* Inducer of MEiosis 4 (IME4) is required for Notch signaling during oogenesis. *Proc Natl Acad Sci U S A*. 2011; 108:14855–14860. [PubMed: 21873203]
45. Dina C, Meyre D, Gallina S, Durand E, Korner A, Jacobson P, Carlsson LM, Kiess W, Vatin V, Lecoecur C, et al. Variation in FTO contributes to childhood obesity and severe adult obesity. *Nat Genet*. 2007; 39:724–726. [PubMed: 17496892]
46. Frayling TM, Timpson NJ, Weedon MN, Zeggini E, Freathy RM, Lindgren CM, Perry JR, Elliott KS, Lango H, Rayner NW, et al. A common variant in the FTO gene is associated with body mass index and predisposes to childhood and adult obesity. *Science*. 2007; 316:889–894. [PubMed: 17434869]
47. Fischer J, Koch L, Emmerling C, Vierkotten J, Peters T, Bruning JC, Ruther U. Inactivation of the Fto gene protects from obesity. *Nature*. 2009; 458:894–898. [PubMed: 19234441]
48. Church C, Moir L, McMurray F, Girard C, Banks GT, Teboul L, Wells S, Bruning JC, Nolan PM, Ashcroft FM, et al. Overexpression of Fto leads to increased food intake and results in obesity. *Nat Genet*. 2010; 42:1086–1092. [PubMed: 21076408]
49. Yang J, Loos RJF, Powell JE, Medland SE, Speliotes EK, Chasman DI, Rose LM, Thorleifsson G, Steinthorsdottir V, Mägi R, et al. FTO genotype is associated with phenotypic variability of body mass index. *Nature*. 2012 doi:10.1038/nature11401.
50. Gerken T, Girard CA, Tung YC, Webby CJ, Saudek V, Hewitson KS, Yeo GS, McDonough MA, Cunliffe S, McNeill LA, et al. The obesity-associated FTO gene encodes a 2-oxoglutarate-dependent nucleic acid demethylase. *Science*. 2007; 318:1469–1472. [PubMed: 17991826]

51. Yi C, Jia G, Hou G, Dai Q, Zhang W, Zheng G, Jian X, Yang CG, Cui Q, He C. Iron-catalysed oxidation intermediates captured in a DNA repair dioxygenase. *Nature*. 2010; 468:330–333. [PubMed: 21068844]
52. Muers M. RNA: Genome-wide views of long non-coding RNAs. *Nat Rev Genet*. 2011; 12:742. [PubMed: 21989130]
53. Mitra SA, Mitra AP, Triche TJ. A central role for long non-coding RNA in cancer. *Front Genet*. 2012; 3:17. [PubMed: 22363342]
54. Gupta RA, Shah N, Wang KC, Kim J, Horlings HM, Wong DJ, Tsai MC, Hung T, Argani P, Rinn JL, et al. Long non-coding RNA HOTAIR reprograms chromatin state to promote cancer metastasis. *Nature*. 2010; 464:1071–1076. [PubMed: 20393566]
55. Rassoulzadegan M, Grandjean V, Gounon P, Vincent S, Gillot I, Cuzin F. RNA-mediated non-mendelian inheritance of an epigenetic change in the mouse. *Nature*. 2006; 441:469–474. [PubMed: 16724059]
56. He C. Grand challenge commentary: RNA epigenetics? *Nat Chem Biol*. 2010; 6:863–865. [PubMed: 21079590]

Highlights (for review)

Reversible modifications in DNA and RNA regulate gene expression. Demethylation pathway of 5mC through 5hmC, 5fC, and 5caC in DNA was discussed. Sequencing methods for 5hmC reveal its distribution and regulatory function. Demethylation of m⁶A in RNA was examined.

(a)

(b)

Figure 1.

(a) Dynamic regulation of 5mC, 5hmC, 5fC, and 5caC in DNA. DNMT3A and DNMT3B are responsible for *de novo* methylation to establish new methylation pattern, while 5mC is maintained during replication by DNMT1-catalyzed methylation of the newly synthesized DNA strand. TET enzymes can catalyze the oxidation of 5mC to 5hmC, 5fC, and 5caC in Fe(II) and α-ketoglutarate(α-KG)-dependent manner. 5fC and 5caC are recognized by DNA glycosylase TDG and converted to cytosine through base excision repair (BER) in an active demethylation. 5hmC may be demethylated by APOBEC3 or AID to form 5hmU, which could undergo base excision by TDG and BER in order to restore the unmodified cytosine. (b) Proposed mechanism of oxidation of 5mC to 5hmC, 5fC, and 5caC by iron(II)/α-

ketoglutarate (α -KG)-dependent Tet proteins. Dioxygen is activated by the iron(II) center to generate an peroxide intermediate that also covalently activates the α -KG cofactor. This intermediate further fragments to form an iron(IV)-oxo species which oxidizes the C-H bond of the substrate.

Figure 2.

(a) Methylation and demethylation of m^6A in RNA. Methylation of internal adenosines in RNA is catalyzed by a yet to be fully identified methyltransferase complex. FTO catalyzes the oxidative demethylation of m^6A in poly(A)-RNA with hydroxymethyladenosine (hm^6A) as a proposed intermediate. (b) FTO is an iron(II)/ α -KG-dependent dioxygenase, which activates dioxygen using the iron(II) center and the α -KG cofactor. Subsequent oxidation of the N-methyl group yields hm^6A as a potential intermediate, which further releases one molecule of formaldehyde to afford the unmethylated adenosine.

Figure 3.

(a) Single-base resolution sequencing of 5hmC. In TAB-Seq, 5hmC is first protected with glucose. After oxidation of 5mC to 5caC, only glucose-protected 5hmC is resistant to bisulfite deamination, and will be sequenced as C, while original C and 5mC will be sequenced as T. **(b)** In oxBS-Seq, DNA is either treated with bisulfite or first oxidized with NaRuO_4 , and then treated with bisulfite. Without oxidation, 5hmC is resistant to deamination. Once 5hmC is oxidized to 5fC it is prone to deamination. The site and abundance can be obtained by comparing the sequencing difference between these two treatments. **(c)** Schematic distribution of 5hmC. 5hmC is highest in distal-regulatory

elements including p300-binding sites, predicted enhancers, CTCF-binding sites, and DNase I hypersensitive sites compared to other regions such as gene bodies. Results suggest active demethylation at gene regulatory elements.

\$watermark-text

\$watermark-text

\$watermark-text