

UCSF

UC San Francisco Electronic Theses and Dissertations

Title

The dynamic patterning and functional roles of DNA modifications in olfactory sensory neurons

Permalink

<https://escholarship.org/uc/item/91p0q8c4>

Author

Colquitt, Bradley Mark

Publication Date

2013

Peer reviewed|Thesis/dissertation

The dynamic patterning and functional roles of DNA modifications
in olfactory sensory neurons

by

Bradley Colquitt

DISSERTATION

Submitted in partial satisfaction of the requirements for the degree of

DOCTOR OF PHILOSOPHY

in

Neuroscience

in the

GRADUATE DIVISION

of the

UNIVERSITY OF CALIFORNIA, SAN FRANCISCO

To my parents Mark and Melony

Acknowledgements

Taken as is, this work presents a line of experiments, data, and interpretations that might seem straight and easily come by. In reality, of course, science is messy, characterized more by its necessary failures and cruel slights of hand than its outright successes. So a good portion of my time in graduate school was spent exploring what eventually turned out to be dead-ends or at least passages that were too steep or too odorous to attempt. To spare the sanities of both the author and the reader, these less productive forays have been largely omitted here. However these unsuccessful attempts, although frustrating and often painful, were absolutely essential to framing the other moments – those that will stick with me – that were so thrilling and deeply rewarding.

Research can be all-consuming and exhausting, and I would like to thank everyone that helped me along the way. First, I was extraordinarily fortunate to have stumbled into Stavros's lab. When I first met Stavros, he was unproven and unwashed, a freshly minted assistant professor with no lab members and a penchant for spending a lot of money. But his passion for science and experimental creativity were immediately apparent, and I knew that, if nothing else, working with him would be at least interesting. Stavros exudes an inspiring confidence that, depending on my mood and the status of my project, is either mystically grand or curiously deluded. Overall, this confidence is well placed; Stavros has an astounding ability to make sense of confusing masses of data and then find the experiments that will clarify them. He was a great mentor and example of how to be a good scientist.

I would also like to thank the members of the Lomvardas lab. Scientific research is a human process and has its own local culture and mores. This culture can be terrible. The Lomvardas lab's was not. Either out of luck or skill, Stavros has attracted a group of dedicated, intelligent,

and deeply curious people that love producing ideas and data (and thoroughly testing ideas and data). They made lab a fun and exciting place to be.

Extending out from the lab, my graduate work took place within the remarkable scientific community at UCSF. My committee members John Rubenstein, Nirao Shah, and Barbara Panning were excellent guides and often provided much needed perspective. The UCSF Neuroscience Graduate Program was my first choice school, and I am fortunate to have had the chance to train with such a great set of scientists.

My friends and family have been incredibly supportive during graduate school, which often absorbed some of the time I could have been spending with them. My lovely girlfriend Aimee was an invaluable counterpoint to the stresses of graduate school and has made my life so much fuller. My sister Brittany and my brother Greg remind me of our weird, silly past. They've kept that past going despite growing up to be real people themselves who are actively pursuing their own dreams. Finally, this was all ultimately made possible by my parents Mark and Melony who, in addition to conceiving me, have given me an unfaltering source of love and support. My interest in the natural world was clear when I was just a kid. My parents recognized this and encouraged its development by keeping the supply of books and chemistry sets free and flowing.

Enjoy.

Contributions

Portions of this research are reprinted material. Chapter 2 was published in *Proceedings of National Academy of Sciences*: Colquitt B. M., Allen W. E., Barnea G., Lomvardas S.. Alteration of genic 5-hydroxymethylcytosine patterning in olfactory neurons correlates with changes in gene expression and cellular identity. *PNAS*, 110(36), 14682-14687. doi:10.1073/pnas.1302759110. Chapter 3 is currently under review for publication.

William Allen contributed some of the initial code for manipulating sequencing data. Gilad Barnea supervised William Allen and guided the research design of Chapter 2. Eirene Markenscoff-Papadimitriou performed several of the RNA- and ChIP-seq library preparations presented in Chapter 3. Stavros Lomvardas oversaw the design of the experiments and the writing of the manuscripts associated with Chapters 2 and 3. Bradley Colquitt designed, performed, and analyzed all other experiments and wrote the bulk of the text for the Chapters 2 and 3 and all of Chapters 1 and 4.

ABSTRACT OF THE DISSERTATION

by Bradley Mark Colquitt

Doctor of Philosophy in Neuroscience

University of California, San Francisco, 2013

The dazzling complexity of multicellular organisms is generated from just a single genome, which must be interpreted differently in each cell type to produce a unique view of this common material. Epigenetic modifications, which influence the structure and functional properties of the genome but do not change its primary sequence, are essential to this act of interpretation during development. In particular, neurodevelopment is a time of dramatic reconfiguration of epigenetic patterning, including major changes to the DNA modification landscape. The nuclear base 5-hydroxymethylcytosine (5hmC), an oxidized derivative of the classic DNA modification 5-methylcytosine (5mC), is highly enriched within the genomes of neurons relative to other cell types and is predominantly localized to active and accessible genomic regions. However, it is unclear how 5hmC patterns are established during neurogenesis and how the modified base contributes to neuronal gene regulation.

Here, I provide insight into neuronal 5hmC biology by studying the patterning and functional roles of the modification during olfactory sensory neuron (OSN) development. Using a genome-wide approach to map 5mC and 5hmC distributions in defined developmental stages of the olfactory epithelium – from multipotent stem cells to neuronal progenitors to mature olfactory sensory neurons – I found that 5hmC levels increase during neuronal differentiation, particularly within the bodies of actively transcribed genes. Altering 5hmC patterning in mOSNs by overexpressing an enzyme that regulates its production demonstrated that 5hmC within the gene

body modulates transcriptional output. Moreover, this manipulation disrupted the expression of olfactory receptors and axon guidance molecules, suggesting that 5hmC patterning is critical for the formation of neuronal identity.

In the final section, I examine connections between *de novo* DNA methylation and 5hmC patterning in olfactory sensory neurons. I show that Dnmt3a, a *de novo* DNA methyltransferase implicated in neuronal development and function, is necessary to generate elevated levels of 5hmC in OSNs. Moreover, Dnmt3a-dependent 5mC and 5hmC patterning is found within accessible regions of the genome, and its loss globally disrupts gene expression. Lastly, I provide evidence that neuronal DNA modifications are critical for neuronal function by demonstrating that odorant-induced gene activation is impaired in the absence of Dnmt3a.

Table of Contents

Acknowledgements	iv
Contributions	vi
Abstract	vii
List of figures	xi
List of tables	xi
Chapter 1: Introduction.....	1
Diversity from a common source	2
Patterning and functions of DNA methylation	4
Phylogenetic and genomic distribution.....	6
Enzymatic regulation	9
Function	12
Dynamic DNA methylation landscapes.....	16
Extending the DNA modification repertoire.....	20
Genomic distribution of oxidized 5mC.....	21
Enzymatic regulation	23
Varieties of function.....	25
Neurogenesis of olfactory sensory neurons	29
Chapter 2: 5hmC patterning and functional role during olfactory sensory neuron development	36
Abstract	38
Introduction	39
Results.....	41

Discussion	48
Methods	51
Figures	57
Chapter 3: Dnmt3a generates the epigenetic substrate for 5-hydroxymethylcytosine in neurons	78
Abstract	80
Main Text	81
Methods	92
Chapter 4: General discussion	121
Oxidation targeting	122
Possible mechanisms	123
Construction of the neuronal epigenome	125
Chapter 5: References	129

List of figures

Figure 1. Schematic of the mouse main olfactory epithelium (MOE).	35
Figure 2. Developmental 5hmC patterning in the MOE.	58
Figure 3. Relationships between 5hmC levels and gene transcription.	60
Figure 4. Dynamic patterning of 5hmC at developmentally regulated genes.	62
Figure 5. Tet3 overexpression alters 5hmC patterning.	64
Figure 6. 5hmC alterations modulate gene expression.	66
Figure 7. Tet3 overexpression disrupts mOSN axonal targeting.	68
Figure 8. Validation of FACS and DIP.	70
Figure 9. Relationships between gene-body 5hmC and transcription in HBCs, GBCs, and during in vitro transcription.	72
Figure 10. Quantitative PCR confirmation of gene-body 5hmC association with developmental gene expression.	74
Figure 11. Supplemental analysis of Tet3-tg mOSNs.	76
Figure 12. Dnmt3a is necessary for neuronal 5hmC patterning within accessible chromatin.	100
Figure 13. Dnmt3a-dependent 5mC and 5hmC patterning segregates with differential histone modifications.	102
Figure 14. Dnmt3a is required for neuronal 5mC and 5hmC patterning at enhancers.	104
Figure 15. Loss of Dnmt3a disrupts global and activity-induced transcription in neurons.	106
Figure 16. Supplemental analysis of mOSN development in Dnmt3a KO MOEs.	108
Figure 17. General analysis of Dnmt3a-dependent 5mC and 5hmC patterning in mOSNs.	110
Figure 18. Representative gene loci in H3K27ac/H3K27me3 cluster 3.	112
Figure 19. Supplemental analysis of transcription profiles in Dnmt3a WT, Het, and KO mOSNs.	114
Figure 20. Supplemental analysis of odorant-induced gene expression.	116
Figure 21. Speculative model of the contributions of Dnmt3a and Tet3 to establishing a polarized neuronal genome.	128

List of tables

Table 1. Chapter 2 sequencing metadata.	77
Table 2. Chapter 3 sequencing metadata.	118
Table 3. Odorant-induced genes.	120

Chapter 1: Introduction

Diversity from a common source

With few exceptions, each cell within a multicellular organism contains the same genetic content. Yet complex organisms possess an astounding complement of different types of cells that vary in both shape and function. From elongated myocytes in muscle to highly branched neurons in the brain, and from insulin-secreting beta cells in the pancreas to light-detecting photoreceptors in the retina, each cell is defined by an identical set of genetic instructions. The fundamental challenge in the construction of multicellular organisms is how to use this common information set to produce these varied cell types. A cell must be able to selectively interpret the genome – using some genes, ignoring others – to create its characteristic properties.

This interpretation occurs through a process called ‘gene regulation’, an umbrella term for the ensemble of diverse molecular strategies that cells employ to control their genes. Gene regulation can be broadly divided into two categories: specification and competence. To be expressed, a gene must be specified for activation (the converse applies for gene repression). In addition, for a gene to be expressed, it must be competent for, or capable of, expression. These two classifications are not strictly delimited. They should be viewed as two intersecting regulatory modes that together orchestrate gene expression. The mechanisms that specify, or drive, gene expression can instruct competence state. Likewise, a gene’s ability to be transcribed, perhaps defined during an earlier developmental time point or by other specification mechanisms, can influence what specification mechanisms are able to influence gene activity.

The concept of genome interpretation hinges on the ability to parse the sequence content of genomic DNA. A class of proteins called transcription factors is operationally defined by the ability to bind to particular sequences of DNA and to influence transcriptional activity of associated genes. Working alongside this sequence-specific factors are modifications to the

chromatin environment, loosely termed ‘epigenetics’, that alter the probability that particular gene will be active or repressed.

The word ‘epigenetics’, much like its vague etymological root (‘upon’ + genetics), has been used by different researchers to refer to different biological processes over time. Conrad Waddington provided perhaps the first use of the word as a reference to the processes through which a genotype produces diverse phenotypes during development (Bird, 2007; Waddington, 1957). A more specific, yet still quite broad, definition states that an epigenetic phenomenon is one that 1) is heritable through a mitotic or meiotic division and 2) does not alter the primary genetic sequence (Russo et al., 1996). Theoretically, such an epigenetic process would allow a cell to convey its developmental and environmental experience to subsequent cellular generations. Taken more broadly, epigenetics could allow transfer of information about the parent’s history to its progeny.

However, this second definition – as something heritable that does not alter the genome – is a negative definition, describing largely what epigenetics is not. Such an open definition has allowed researchers to sketch out a positive definition in molecular terms, fitting a diverse range of regulatory processes under the banner of epigenetics. A more modern definition of epigenetics has come to refer to the set of chemical modifications to chromatin that alter its structure and/or function. Two sets of chromatin modifications have been studied in detail: direct modification of genomic DNA and modification of the histone that compose nucleosomes. These broad classes are implicated in wide array of functions, from gene repression and activation to the conditioning of distal regulatory elements, and are dynamically regulated across developmental time. The enzymatic machinery responsible for controlling these modifications is highly intertwined, producing a coordinated system of epigenetic regulation.

In this work, I focus on the first class, the direct modification of genomic DNA, and examine its described role in gene regulation during development, with particular emphasis on neurodevelopment. The field of DNA modifications has undergone tremendous change in the past four years with the discovery of additional modified DNA bases, which has brought interesting new complexity to our understanding of epigenetics. The second portion of this chapter will discuss new roles for this expanded set of DNA modifications in gene regulation. Finally, I will describe the neurogenesis of mouse olfactory sensory neurons, the model system used for my thesis work, and the properties of this system make it amenable to the dissection of epigenetic regulatory roles during neurodevelopment.

Patterning and functions of DNA methylation

A molecular understanding of the mechanisms underlying epigenetic processes began with the discovery of methylated cytosines in nucleic acid extracted from calf thymus (Hotchkiss, 1948). In particular, one methylated species, a cytosine with a methyl group at the 5-position of the nucleoside, 5-methylcytosine, was particularly abundant. This modification satisfies the two criteria of an epigenetic feature described above. 1) Cytosine methylation does not *per se* alter the genomic sequence: 5mC strongly prefers a complementary G and likewise will continue to read as a cytosine through cell division and is interpreted as a cytosine during transcription. 2) It is heritable across cell divisions. In mammals, 5mC occurs almost exclusively at CG dinucleotides (although important exceptions to this rule have been discovered and will be discussed later). Because the reverse complement of CG is CG, placement within this dinucleotide context provides a mechanism to propagate the modified cytosine during DNA synthesis to the newly synthesized, unmodified strand (Holliday and Pugh, 1975; Riggs, 1975). In this way methylation patterns can serve as a form of genomic memory and may reflect the life

history of a particular cell lineage. Evidence that this mode of methylation replication occurs first came through the observation that *in vitro* methylated DNA fragments will remain methylated after up to 100 generations within mouse cells lines (Stein et al., 1982; Wigler et al., 1981). Moreover, methylation of cytosines outside of the CG context was lost while those within the CG context were selectively retained.

Several methods exist to detect the presence of 5mC at given genomic regions, with resolutions ranging from single cytosines to broader domains of hundreds of basepairs. The earliest method developed takes advantage of a type of bacterial immune system, the type II restriction-modification system, in which a methyltransferase is paired with a restriction endonuclease (Wilson, 1988). Each enzyme targets a specific sequence in the bacterial genome; if the site is methylated, cleavage by the endonuclease will be inhibited. Therefore researchers can use these restriction enzymes to probe areas of methylated DNA in other species by assaying which regions are cut and which remain intact (Laird, 2010). A few of these enzymes, such as *HpaII* and *SmaI*, have a matching isoschizomer, *MspI* for *HpaII*, or neoschizomer, *XmaI* for *SmaI*, which can be used to control site frequency and digestion efficiency.

The gold standard for these methods is bisulfite sequencing which provides modification information at single basepair resolution. Exposure to sodium bisulfite deaminates unmethylated cytosines. Therefore, if the cytosine is methylated remain a cytosine; if it is unmethylated it will convert to a thymine. This sequence difference can then be interpreted using either primers sensitive to such a conversion or by sequencing a given region and comparing that sequence to an untreated reference sequence. Modern methods of sequencing have allowed researchers to assess genome-wide methylation patterns at single-base resolution (Laird, 2010) which has provided unprecedented detail of DNA methylation landscapes in number of systems (Lister et

al., 2013; 2008; 2009). However, acquisition of data with this breadth and detail can be expensive for vertebrate-sized genomes.

Methylation levels can also be probed using affinity enrichment methods that are easy to perform and considerably cheaper than bisulfite-based approaches when scaled for genome-wide analysis. Using either antibodies or recombinant protein domains that bind to methylated cytosines, one can enrich a pool of genomic DNA segments for methylated molecules, similar to the approach used in chromatin immunoprecipitation of histone modifications or transcription factors. However, this approach is both less sensitive (isolated methylated cytosines are not detectable) and has lower resolution (approximately 100 bp) than bisulfite sequencing. Moreover, reported methylation levels from affinity enrichment display a non-linear dependence on cytosine density, making comparisons of methylation levels between loci with different cytosine content challenging (Laird, 2010).

With modern massively parallel sequencing technology, the DNA sample that results from each of these techniques can be fully sequenced, yielding genome-wide maps of methylation abundance. This combined level of resolution and breadth has transformed the study of epigenetics and has allowed researchers to test the applicability of some of the ‘principles’ of DNA methylation – largely determined through the analysis of single loci with a limited number of cell types – to the full range of genomic features and cellular states.

Phylogenetic and genomic distribution

Although widespread across phylogeny, the genomic abundance of DNA methylation is not uniform across species. In fact, considerable range of total methylation levels exists within metazoans, from essentially undetectable levels in nematodes to extensive methylation in

vertebrates. These differences have led to the hypothesis that DNA methylation provides a distinct regulatory function in vertebrates that is not found in simpler mammals (Bird, 1980). However, although some model organisms, most notably *Caenorhabditis elegans* and *Drosophila melanogaster*, appear to have little methylation within their genomes, most phylogenetic lineages contain at least one member with a methylated genome (Bestor, 1990). Moreover, the genomes of plants and filamentous fungi (*Neurospora*) are highly methylated (Martienssen, 2001), suggesting that DNA methylation is ancestral to eukaryotes. Indeed eukaryotic DNA methyltransferases share strong homology with bacterial 5mC methyltransferases (Bestor et al., 1988; Pósfai et al., 1989). It has been hypothesized (Bestor, 1990) that the primary driver in the use of DNA methylation systems is the size of the genome and the relative abundance of parasitic DNA elements that have incorporated within it. Genomes with fewer than 10^8 nucleotides (such as *C. elegans*, *Drosophila*, and yeast) tend to not have detectable methylation, while those with larger genomes with an accompanying expansion of repeat content, contain DNA methylation.

In addition, the genomic distributions of DNA methylation widely vary across species (Suzuki and Bird, 2008). As mentioned, *C. elegans* has no detectable 5mC and its genome does not encode a typical eukaryotic DNA methyltransferase (Bird, 2002). The genome of *Drosophila* contains sparse methylation (Gowher et al., 2000; Lyko et al., 2000; Tweedie et al., 1999) that primarily occurs at CT and CA dinucleotides within genes and parasitic transposons (Salzberg et al., 2004). Other invertebrates, such as ascidians and echinoderms, have mosaic methylation patterns (Bird et al., 1979; Simmen et al., 1999; Tweedie et al., 1997), in which 5mC is highly concentrated in particular areas of the genome. In general, these methylated areas span regions of little transcriptional activity, such as repetitive satellite sequences (Bird, 1980).

In contrast, vertebrate genomes contain extensive ‘global’ DNA methylation, in which 5mC is broadly distributed and is not confined to a particular type of genomic element. Approximately 70% of CGs within human somatic cells are methylated (Ehrlich et al., 1982) or about 1% of the total number of DNA bases. This global pattern of DNA methylation has had profound effects on the nucleotide composition of the vertebrate genome. Cytosines are sensitive to spontaneous deamination. When unmethylated, deaminated cytosine becomes uracil, which can be identified as a mutation by the DNA repair machinery; when methylated, a deaminated cytosine is identical to a thymine (Bird, 1980). This mutation can then be interpreted incorrectly during DNA synthesis and thus be stably propagated. As a result, CGs occur at approximately one-fifth of the expected frequency in vertebrates (Swartz et al., 1962) but are found at expected frequencies in the poorly methylated genomes of *Drosophila*, *C. elegans*, and *S. cerevisiae* (Deaton and Bird, 2011).

Although CG dinucleotides are infrequent across the genome on average, the vertebrate genome contains isolated pockets containing greater than expected numbers of CGs called CpG islands (CGI) (Bird, 1986; 1987; Lister et al., 2013). There are estimated to be ~29,000 CGIs within the human genome (Colquitt et al., 2013; Lander et al., 2001) of which 50% are associated with the 5’ ends of genes (Bird, 1987; Illingworth et al., 2010; Wu and Zhang, 2011). The remainder, termed ‘orphan’ CGIs, are equally split between intergenic and intragenic sites (Illingworth et al., 2010; Magklara et al., 2011). These regions are largely unmethylated in all tissues examined regardless of the transcriptional activities of associated genes (Bird, 2002; Nicolay et al., 2006). This lack of methylation is reasoned to be the cause of the higher than expected CG frequencies.

Enzymatic regulation

The mammalian genome contains three enzymes with DNA methyltransferase activity: a so-called ‘maintenance’ methyltransferase, *DNA methyltransferase 1 (Dnmt1)* and two *de novo* methyltransferases, *Dnmt3a* and *Dnmt3b*. Each of these proteins contributes in distinct ways to the DNA methylation landscape, at both different genomic locations and developmental stages.

As mentioned above, the placement of methylated cytosines within the CG context provides a mechanism to propagate the methylated state to the newly synthesized, but unmodified, strand during DNA synthesis (Creighton et al., 2010; Holliday and Pugh, 1975; Riggs, 1975). *Dnmt1* strongly binds and methylates hemimethylated CG *in vitro* (Bestor, 1992; Pradhan et al., 1999; Visel et al., 2007). Moreover, the enzyme is localized to replication forks during S phase (Colquitt et al., 2013; Leonhardt et al., 1992; Magklara et al., 2011) and interacts with proliferating cell nuclear antigen (PCNA), a processivity factor for DNA polymerase (Chuang et al., 1997; Colquitt et al., 2013). In addition, its deletion in mice results in global demethylation and embryonic lethality (Li et al., 1992; Wu et al., 2010), strongly supporting its role in maintaining methylation patterns across cell division. Although this mechanism in theory could perfectly maintain DNA methylation patterns established in previous generations, maintenance methylation is not without error; considerable variation of methylation patterning on both *in vitro* (Mombaerts, 2004; Rubin and Katz, 1999; Stein et al., 1982) and *in vivo* (Lister et al., 2013; Stöger et al., 1997) sequences has been observed in clonal cell populations, suggesting that methylation patterns are continually in flux and are not simply copied by rote (Bird, 2002; Meissner et al., 2005; Ramsahoye et al., 2000).

While *Dnmt1* maintains, *Dnmt3a* and *Dnmt3b* mediate *de novo* methylation during development. The enzymes share 84% homology, differing at only 44 residues. In addition, both enzymes are

able to methylate unmethylated cytosines and show equal preference for both fully unmethylated and unmethylated CGs (Lister et al., 2013; Okano et al., 1998). Both can be broadly divided into N-terminal regulatory and C-terminal regulatory domains. The regulatory domain contains PHD, PWWP, and ADD domains, which are thought to mediate the targeting of the enzymes to specific chromatin sites, likely through interactions with other proteins. In particular, in mouse embryonic stem cells both Dnmt3a and Dnmt3b target pericentromeric heterochromatin, even under hypomethylated conditions (Bachman et al., 2001; Okano et al., 1999), and associate with the heterochromatic proteins HP1 and SUV39H1 through the PHD domain (Fuks et al., 2003; Shykind et al., 2004). Likewise, the PWWP domain is necessary to target both enzymes to chromatin (Ge et al., 2004; Gong et al., 2003) and mediates the interaction between Dnmt3a and H3K36me3, a histone modification associated with actively transcribed genes (Colquitt et al., 2013; Dhayalan et al., 2010; Magklara et al., 2011). Interestingly, the ADD domain binds to unmethylated lysine 4 of histone 3 (H3K4), a residue that is typically methylated at active transcription start sites, thus providing a mechanistic connection between gene activity and the hypomethylated state of TSS associated CGIs (Ooi et al., 2007; Otani et al., 2009).

However the two enzymes exhibit important differences in target specificity and enzymatic mode of methylation. Dnmt3a shows a strong preference for cytosines in both CG and CA contexts, while Dnmt3b methylates only CG (Gowher and Jeltsch, 2001; Heinz et al., 2010). In addition, Dnmt3b acts in ‘processive’ manner *in vitro*, i.e. the kinetics of Dnmt3b-mediated methylation on a long piece of DNA are consistent with a model in which the enzyme scans along DNA, methylating multiple cytosines without losing contact with the DNA (Anders and Huber, 2010; Gowher, 2002). In contrast, Dnmt3a operates in a ‘distributive’ manner, methylating single sites

without travelling along the DNA to additional cytosines (Gowher, 2002; Gowher and Jeltsch, 2001; McLean et al., 2010).

A third member of the Dnmt3 family, *Dnmt3l*, shares homology with *Dnmt3a* and *Dnmt3b* but lacks catalytic activity and evolved specifically in the therian lineage (Liao et al., 2013; Yokomine et al., 2006). *Dnmt3l* is highly expressed during gametogenesis is necessary for the establishment of germline methylation patterns (Anders and Huber, 2010; Bourchis et al., 2001). Although itself catalytically dead, *Dnmt3l* forms tetrameric complexes with *Dnmt3a*, a configuration that enhances the catalytic efficiency and processivity of *Dnmt3a* (Holz-Schietinger and Reich, 2010; Jia et al., 2007; Jurkowska et al., 2008).

Accompanying these different molecular functions, *Dnmt3a* and *Dnmt3b* have different expression patterns and roles during development. The genetic ablation of both genes results abnormal *de novo* methylation in embryonic stem cells and early embryogenesis (Okano et al., 1999). Loss of *Dnmt3b* results embryonic lethality and reduced methylation of centromeric minor satellite repeats. In contrast, mice lacking *Dnmt3a* survive through embryogenesis but die 3 to 4 weeks after birth. Both genes are highly expressed in early embryos but assume differential expression patterns during later development: *Dnmt3a* is highly expressed within ectoderm at E7.5 and is later ubiquitously expressed at E9.5. *Dnmt3b* is predominantly expressed in ectoderm, especially forebrain and eyes, but remains weakly expressed throughout the embryo (Okano et al., 1999).

Interestingly, the expression patterns of the *de novo* DNA methyltransferases exhibit considerable temporal dynamics in the embryonic nervous system (Feng et al., 2005). *Dnmt3a* is expressed in neural precursor cells in the subventricular zone (SVZ) between E10.5 and E17.5

during neurogenesis, but is restricted to post-mitotic neurons in postnatal cortex, olfactory bulb, hippocampus, striatum, and cerebellum. Moreover its expression in the CNS increases from birth until three weeks after birth then declines in the adult. *Dnmt3b* is expressed in the ventricular layer of the subventricular zone between E10.5 and E13.5, when astroglial neuronal progenitors first appear at the onset of neurogenesis (Götz and Huttner, 2005). These spatially and temporally defined expression patterns suggest that each enzyme is engaged in neurodevelopment and post-mitotic function in distinct ways.

Function

The contribution of DNA methylation to gene regulation has been intensely studied since its discovery. Although there is some debate about how the modification functions in particular areas of the genome, 5mC is strongly associated with transcriptionally silent regions. Evidence for this repressive role comes from the transcriptional properties of unmethylated CGIs. Whether present within traditional gene promoters or found in intergenic regions, these unmethylated domains are highly bound by RNA polymerase II (RNAPII) and are enriched for H3K4me3, a histone modification associated with transcriptional activity (Illingworth et al., 2010). Moreover, global run-on sequencing, as measure of transcript production, indicates that unmethylated CGIs are actively transcribed (Illingworth et al., 2010). Orphan CGIs, as defined above, are more sensitive to methylation than those found near TSSs (Illingworth et al., 2010; Maunakea et al., 2010), and their tissue-specific methylation correlates with reduced H3K4me3 enrichment and transcript production. Together, these results suggest that areas of enriched CG density are platforms for transcription initiation that are strongly sensitive to methylation status (Deaton and Bird, 2011).

Additional evidence for this repressive role comes from the silencing of repetitive regions of the genome that, while not classically considered genes, have transcriptional potential. The genomic landscapes of vertebrates are dominated by repetitive genetic elements; depending on detection method, 50-80% of the human genome is repetitive (de Koning et al., 2011; Kuhn et al., 2007), an indication of the tremendous effect these invasive DNA elements have had on genome evolution. A major class of repeat is the major and minor satellite repeats found adjacent to centromeres. These are highly methylated, the loss of which causes hyperacetylation (a chromatin state associated with transcriptional activity) and chromosomal abnormalities (Gopalakrishnan et al., 2009). Transposable elements, non-endogenous DNA elements, comprise the second major class of repetitive element and are also highly methylated. Injection of mice with the cytosine analog 5-azacytidine (which cannot be methylated and results in a global reduction of methylation) results in the activation of normally silent moloney murine leukemia provirus (Jaenisch et al., 1985). In addition the loss of Dnmt1 in mouse embryos results in the activation of the transposable element intracisternal A particle (Walsh et al., 1998).

These effects could be mediated by either decreasing or increasing the binding efficiency of specific trans factors to methylated sites, and both mechanisms appear to be at work in the interpretation of 5mC landscapes. The blocking of protein binding appears to be an ancestral mechanism as it forms the basis for self/non-self recognition in bacteria. The *E. coli* genome is methylated at specific sites with renders them unrecognizable by endogenous restriction enzymes. Similarly, the presence of 5mC within promoters can block the binding of transcription factors in human cell lines, in agreement with the absence of methylation at TSS associated CGIs (Watt and Molloy, 1988). Moreover, the binding of CCCTC-binding factor (CTCF), a

transcription factor believed to insulate or functionally segregate areas of the genome from each other, is strongly inhibited by 5mC (Bell et al., 1999; Hark et al., 2000).

Alongside this inhibitory function, several proteins in eukaryotes specifically bind to methylated cytosines. In fact, the number of methyl-cytosine binding proteins correlates with 5mC abundance and generality of distribution, with the largest numbers of these proteins found in vertebrates and plants (Bird, 2002). There are five proteins in mammals known to bind to methylated DNA. Four of these – methyl-CpG binding protein 2 (MeCP2)(Lewis et al., 1992), methyl-CpG binding domain protein 1 (MBD1)(Meehan et al., 1989), MBD2, and MBD4 (Hendrich and Bird, 1998)– contain an MBD domain (methyl binding domain), while the fifth Kaiso does not and is unrelated (Prokhortchouk et al., 2001). MBD3, while sharing homology to the other MBD family members, does not strongly bind to methylated DNA (Hendrich and Bird, 1998). In contrast, MeCP2 is able to bind single methylated CGs (Nan et al., 1993).

In keeping with the association between 5mC and transcriptional repression, MBDs can recruit transcriptional repressors to methylated loci. In fact, there is evidence that at some loci it is not 5mC *per se* that inhibits transcription but the factors that it recruits. For example, methylated transgenes become transcriptionally active in the absence of MBD1 despite maintaining unaltered levels of 5mC (Boyes and Bird, 1991). Moreover the MBDs associate with proteins with established functions in transcriptional repression, such as Sin3a (Jones et al., 1998) and histone deacetylase complexes (Feng and Zhang, 2001; Nan et al., 1998; Ng et al., 1999).

Although this inhibitory role is well established in a number of systems, there are important exceptions. In *Arabidopsis thaliana*, A substantial fraction of genes (33%) contain enriched levels of 5mC within their gene-bodies (Cokus et al., 2008; Zhang et al., 2006; Zilberman et al.,

2007). Interestingly, these genes are moderately transcribed and several are constitutively active, ‘housekeeping’ genes. However the ‘atypical’ localization of 5mC within gene-bodies is not strictly associated with plants, as the globally methylated genomes of vertebrates also contain 5mC within genes. This does not appear to be a passive association: during X chromosome inactivation, CGIs are hypermethylated on the inactive X and establish a silent transcriptional state (Mohandas et al., 1981; Venolia et al., 1982). In contrast, gene bodies on the active X are highly methylated resulting in higher overall methylation levels on the active versus inactive X chromosome (Hellman and Chess, 2007). 5mC levels in autosomal genes also positively correlate with transcription in ESCs (Lister et al., 2009). Notably, this correlation is largely generated by strand-specific 5mC within the CH context, suggesting that Dnmt3a, through interactions with gene-body H3K36me3, establishes this transcriptionally permissive methylation signature.

It is currently not clear how gene body methylation might influence the expression of transcribed genes. However, 5mC is particularly enriched at the boundaries between exons and introns (Laurent et al., 2010). Moreover, methylation has been shown to regulate CTCF binding to exon-associated regions which influences the probability of exon inclusion (Shukla et al., 2011). The rate of RNA polymerase elongation is a major determinant of exon usage (la Mata et al., 2003), suggesting that methylation state and CTCF binding may regulate splicing by altering the rate of polymerase passage through a given exon.

A complementary model suggests that 5mC can exclude the deposition of repressive histone modifications that spread during differentiation (Wu et al., 2010). Neural stem cells (NSCs) from the subventricular zones of *Dnmt3a*-null mice exhibit differentiation defects as evidenced by reduced numbers of interneurons migrating to the olfactory bulb. Interestingly, the absence of

Dnmt3a in NSCs both upregulates non-neuronal genes and downregulates active, pro-neural genes. These downregulated genes typically display elevated levels of 5mC flanking their TSSs, which is lost in Dnmt3a-nulls and is replaced by elevated levels of H3K27me3, a repressive histone modification that dynamically regulated during development. These data suggest the 5mC facilitates gene expression not through directly positive way (e.g. recruiting a histone acetylase) but by establishing a protected zone that is competent for transcription.

Dynamic DNA methylation landscapes

DNA modification patterning in mammals undergoes two waves of dramatic rearrangement, one during early embryonic development and the other during gametogenesis. On fertilization, methylation levels on the paternal genome sharply reduce before DNA synthesis begins (Reik et al., 2001), suggesting active removal of methyl groups. However, a number of regions retain elevated levels of methylation, most notably some imprinted loci, such as *H19* (Olek and Walter, 1997; Tremblay et al., 1997), and IAP and retrotransposon repeats (Kim et al., 2004). The maternal genome however is steadily demethylated at a rate consistent with the gradual diluting out of methylated cytosines by replication, likely due to the exclusion of Dnmt1 from the nucleus during this time (Ratnam et al., 2002). Around implantation, global DNA methylation levels are restored to somatic levels by Dnmt3a and Dnmt3b (Reik et al., 2001).

In addition, gametogenesis is host to a second wave of massive methylation loss and gain. By E13 to E14, the methylation levels in primordial germ cells, the precursors to mature sperm and oocytes, have dropped substantially (Kafri et al., 1992; Monk et al., 1987). This loss occurs globally, erasing the methylation found at differentially methylated imprinted regions as well as that found at single-copy genes. In males, remethylation occurs around E16 to E17 in

prospermatogonia (Brandeis et al., 1993; Kafri et al., 1992). Remethylation in the female germ line however occurs later, during the development of the oocytes (Reik et al., 2001).

Beyond these two key developmental periods, when many of the initial regulatory programming of the organism takes place, DNA methylation patterns are relatively constant in developed somatic tissue. Trans-factors expression and binding and the patterning of other epigenetic regulators are notably dynamic during the specification and differentiation of mature cell types; *de novo* DNA methylation is relatively less active in most tissues. However, the DNA methylation landscape does appear to be progressively refined during the development of somatic lineages (Smith and Meissner, 2013). This refinement may channel or stabilize particular cell fates as the expression of certain regulatory factors or terminal genes is restricted by silencing 5mC. Alternatively, the presence of 5mC within the bodies of actively transcribed genes suggests that developmental 5mC patterning may also aid in the facilitation of expression programs.

In particular, both the maintenance and alteration of DNA methylation patterning is critical for neurodevelopment. Although the loss of *Dnmt1* in post-mitotic neurons does not affect neuronal survival, its deletion in neural precursor cells (using a *Nestin-Cre* driver) results in hypomethylation and reduced daughter cell survival (Fan et al., 2001). Disruption of *de novo* DNA methylation through the ablation of *Dnmt3a* in neural precursor cells produces mice that are born healthy but display neuromuscular defects, degenerate in adulthood, and die prematurely at 3-4 weeks, around the time of extensive synaptogenesis (De Felipe et al., 1997; Nguyen et al., 2007).

Recent studies characterizing DNA methylation patterns during neurogenesis demonstrated that *de novo* methylation is far more extensive than previously appreciated (Lister et al., 2013; Varley et al., 2013; Xie et al., 2012). Surprisingly, this new neuronal methylation occurs predominantly in the CH (mCH) context with a bias toward CA in whole brain (Xie et al., 2012) and frontal cortex (Lister et al., 2013). Moreover, this non-canonical methylation is not a minor component of the methylation landscape, occurring on substantial fraction of all methylated cytosines (25.6%) compared to the proportion of CG methylation (57.2%). Notably, the gain of this methylation signature occurs around 1 to 4 weeks post-birth, coincident with the expression of *Dnmt3a* (Lister et al., 2013). In contrast to ESCs (Lister et al., 2009), mCH in gene bodies is negatively correlated with transcription (Lister et al., 2013; Xie et al., 2012), suggesting that mark is differentially interpreted in the two systems. In support of this negative role, genes that are differentially expressed in glia versus neurons gain 5mC in both CG and CH contexts in neurons, while genes that are expressed in neurons exhibit a neuron-specific depletion of mCH.

Moreover, as noted above, this time coincides with increased synaptogenesis, suggesting that these *de novo* methylation patterns interact with newly established neuronal signaling. Indeed the loss of DNA methyltransferase activity in post-mitotic neurons impairs various aspects of neuronal function. Double knockout of both *Dnmt1* and *Dnmt3a* using *Camk2a-Cre*, broadly expressed in post-mitotic neurons, results in a smaller hippocampus and reduced LTP in CA1 (Feng et al., 2010). Moreover, these mice display defects in hippocampal-dependent learning and memory in the Morris water maze, with apparently greater impairment of memory consolidation over acquisition. However, the generality of the genetic ablation make it difficult to assign a causal relationship between DNA methylation activity and neuronal function; gene

expression programs may be globally disrupted in these cells, affecting their health and function in a non-specific way.

Perhaps more compelling evidence of crosstalk between DNA methylation patterns and neuronal activity comes from work that directly assayed 5mC levels at specific loci after some form of stimulation. Early work using primary mouse cortical cells demonstrated that the induction of *brain-derived neurotrophic factor (Bdnf)* in response to depolarization by potassium chloride correlates with the demethylation of one of the *Bdnf* promoters after 2 to 4 days treatment (Martinowich et al., 2003). A more recent study assessed methylation levels at ~200,000 CGs in dentate granule cells in the hippocampus after widespread neuronal activation using electroconvulsive stimulation (ECS) (Guo et al., 2011b). A slight increase in the number of methylated CGs (1.4%) was detected four hours after stimulation. Interestingly, about a third of these *de novo* methylated cytosines remained methylated after 24 hours, and a substantial fraction (67%) were also methylated after a three-day course of voluntary running. However, the authors report no correlation between this ECS-induced methylation and changes to the expression of associated genes.

Although these data point to a role for dynamic DNA methylation in regulating gene expression in response to neuronal signaling, no firm evidence exists of a direct connection within an *in vivo* system. This lack of evidence is likely due in part to more complex role for DNA methylation than is currently appreciated. Much of neuronal *de novo* methylation occurs either within gene bodies or intergenic sites, where the function of the modification has been less studied, and may have a more subtle influence on gene expression, e.g. altering the probability of gene activation. Moreover, the fact that gene expression correlates with gene body DNA methylation levels in opposite ways in ECSs and neurons indicates that the interpretation of these 5mC landscapes

may different regulatory logic in different systems. Nonetheless, it is clear that DNA methylation is not a simply employed epigenetic modification, necessary only for the silencing of repeats and establishing regulatory states during early development, but is actively regulated in somatic lineages through development and by environmental signaling.

Extending the DNA modification repertoire

For decades, 5mC was the only modified DNA base known to exist at appreciable levels in mammalian genomes. However in 2009 substantial levels of a derivative of 5mC, 5-hydroxymethylcytosine (5hmC), were reported in the genomes of mouse embryonic stem cells and cerebellar neurons (Kriaucionis and Heintz, 2009b; Tahiliani et al., 2009). The discovery immediately expanded the potential functional repertoire of direct modifications to genomic DNA and suggested that we must think about the broader range of DNA modifications, not just DNA methylation, in trying to decipher the role of epigenetics in genome function.

5hmC was originally studied for its role in the genome defense system of T-even bacteriophages (Revel and Lubia, 1970). The cytosines of the bacteriophage genome are entirely in the hydroxymethylated form and are additionally modified by a glucosyl group that is added to the reactive hydroxyl. This set of modifications protects the phage genome from both the endogenous bacterial restriction enzymes and its own host-targeting nucleases. A similar modification system exists in the trypanosome *Trypanosoma brucei*, the pathogen that cause sleeping sickness (Borst and Sabatini, 2008). Here, thymines undergo a similar hydroxylation of the 5-position methyl group followed by glucosylation forming what is called 'base J'. Although its function is unknown, the base is thought to be involved in the silencing of variant surface glycoprotein expression, which allows the pathogen to escape detection by the host immune system (Borst and Sabatini, 2008).

Genomic distribution of oxidized 5mC

The finding that 5hmC appears to be particularly enriched in ESCs and neurons suggested that levels of modified base were generally variable across tissues. Indeed, although 5mC levels vary little across tissues (approximately 4% of cytosines in all tissues assayed), the amount of global 5hmC depends on tissue (Globisch et al., 2010). 5hmC levels are indeed highest in the CNS (0.3% to 0.7% of cytosines), intermediate in nasal epithelium, kidney, bladder, heart, skeletal muscle and lung (0.15% to 0.17%), and low in pituitary, spleen, liver, and testes (0.03% to 0.06%) (Globisch et al., 2010; Kriaucionis and Heintz, 2009b; Münzel et al., 2010). These percentages are substantial relative to the 5mC abundances; 5hmC constitutes 0.6% and 0.2% of cytosines in isolated Purkinje and granule cells from the cerebellum, representing 40% the number of methylated cytosines (Kriaucionis and Heintz, 2009b).

The presence of 5hmC within mammalian genomes was discovered using thin layer chromatography and liquid chromatography-mass spectrometry methods that assessed the abundance of the modified base within total genomic DNA (Globisch et al., 2010; Kriaucionis and Heintz, 2009b; Tahiliani et al., 2009). To extend these techniques, antibodies that selectively bind 5hmC and chemical labeling methods that specifically react with the hydroxyl group were developed to provide genomic position information (Huang et al., 2012; Jin et al., 2011; Song et al., 2011; Stroud et al., 2011). As with 5mC sensitive affinity enrichment techniques, these methods are relatively inexpensive and simple to perform but at a cost to resolution and sensitivity. Bisulfite mediated detection is complicated by the fact that deaminated 5hmC is read as a thymine during sequencing, making it indistinguishable from 5mC (Huang et al., 2010). A way around this conflict was developed by incorporating the enzymatic protection used by T-even bacteriophages in a method termed Tet-assisted bisulfite sequencing (TAB-seq) (Yu et al.,

2012). Before treatment with bisulfite, genomic DNA is incubated with beta-glucosyltransferase, adding a protective glucosyl group to the reactive hydroxyl of 5hmC, followed by incubation with recombinant TET1 (an enzyme that catalyzes the oxidation of 5mC as will be discussed in more detail below), which fully oxidizes unprotected 5mC. Subsequent bisulfite treatment produces DNA in which 5hmC are read as cytosines and 5mC is read as thymine. Comparison of these sequences with those from standard bisulfite reactions allows one to discriminate between 5mC and 5hmC.

The genome-wide distribution of 5hmC is complementary to that of 5mC. This pattern is evident in immunofluorescent detection of the two modifications (and will be elaborated on in Chapter 2). While 5mC is enriched in densely compacted pericentromeric repeats, 5hmC is enriched in accessible, euchromatic regions (Mellén et al., 2012). Moreover, biochemical fractionation of chromatin from neural tissues by DNase I accessibility shows that 5hmC is largely enriched in open regions of the genome (Mellén et al., 2012).

In agreement with this association with accessible regions of the genome, 5hmC is enriched at genomic elements typically associated with transcription and active regulation. In ESCs, neural progenitor cells, and neurons, 5hmC is enriched at gene loci, extending 500-2000 bp upstream of the TSS and extending into the gene-body (Ficz et al., 2011; Hahn et al., 2013; Pastor et al., 2011; Song et al., 2011; Szulwach et al., 2011a; 2011b; Wu et al., 2011a). However, 5hmC is depleted from the TSSs of active genes, similar to 5mC. The depletion of both modifications from the TSS is consistent with a repressive role for both in transcription initiation. In contrast, gene-body and TSS flanking 5hmC levels positively correlate with transcriptional output in mouse ESCs, brain, and liver (Mellén et al., 2012; Song et al., 2011; Szulwach et al., 2011b; Wu

et al., 2011a). This bidirectional association with transcription suggests that 5hmC is interpreted differentially depending on its genomic context.

5hmC is also found enriched at active enhancers in human and mouse ESCs (Ficz et al., 2011; Pastor et al., 2011; Stroud et al., 2011; Szulwach et al., 2011b; Williams et al., 2011; Wu et al., 2011a; Xu et al., 2011; Yu et al., 2012) and CNS neurons (Hahn et al., 2013; Lister et al., 2013). However, the transcription factor binding site itself is fully depleted of both 5mC and 5hmC, similar to the pattern seen at TSSs. Interestingly, during differentiation of mouse neural P19 and adipocyte 3T3-L1 cell lines, 5hmC is enriched on activated enhancers that are associated with genes expressed in neurons and adipose tissue, respectively (Sérandour et al., 2012). Moreover, this gain of 5hmC precedes acquisition of H3K27 acetylation, a histone modification associated with enhancer activity, suggesting that dynamic 5mC oxidation is an early event in conditioning enhancers for activity.

Enzymatic regulation

Coincident with the discovery of 5hmC in mammals, a conserved family of enzymes, the Ten-eleven translocation (Tet) family, was shown to catalyze the oxidation of 5mC to 5hmC (Tahiliani et al., 2009). Interestingly, the founding member of the family, Tet1, was originally discovered as a fusion partner with the H3K4 methyltransferase MLL (mixed lineage leukemia) in acute myeloid leukemia (Lorsbach et al., 2003; Ono et al., 2002), although its function was unknown until its connection with 5mC oxidation. The other members of the family were identified in a computational search for mammalian enzymes with homology to the two *T. brucei* enzymes, JBP1/2, that catalyze the oxidation of thymine during base J synthesis. An extended search for homologous proteins indicated that Tet homologues are found across metazoans with the conserved catalytic domain also found in fungi and plants (Tahiliani et al., 2009).

There are three members of the Tet family (Tet1/2/3) in jawed vertebrates that appear to have been generated by a triplication event from a common metazoan ancestor (Iyer et al., 2009; Pastor et al., 2013). Although both Tet1 and Tet3 contain an N-terminal CXXC domain, Tet2 lost this domain through a chromosomal inversion event, producing an independent CXXC-containing gene IDAX (Pastor et al., 2013). The CXXC domain can be divided into three subfamilies based on sequence homology (Frauer et al., 2011). Subfamily 1 binds clustered unmethylated CGs (Birke et al., 2002; Lee et al., 2001) and is present in a range of chromatin-modifying enzymes, including the H3K4 methyltransferase MLL (mixed lineage leukemia), Dnmt1, and MBD1. The CXXC domains of Tet1, Tet3, and IDAX lie in the third subfamily and have more nuanced binding properties: Tet1 binds to CG regardless of modification state (Xu et al., 2011; Zhang et al., 2010), Tet3 binds to unmethylated cytosines with no preference for the CG context (Xu et al., 2011), and IDAX binds to unmethylated CGs similarly to those in subfamily 1 (Ko et al., 2013).

All three members are able to convert 5mC to 5hmC (Ito et al., 2010) through a 2-oxoglutarate (2OG) – and Fe(II)-dependent dioxygenase domain also found in the trypanosome JBPs (Tahiliani et al., 2009). The core of the domain contains a double-stranded β -helix (DSBH) fold, which is present in all members of the superfamily. Although the exact mechanism of oxidation is unknown, biochemical and structural information is available for a distantly related homolog, AlkB, which flips the targeted cytosine out of the double helix into proximity with the catalytic Fe^{2+} present in the DSBH fold (Loenarz and Schofield, 2011).

Another member of the 2OG-dependent oxygenase family, thymine 7-hydroxylase, can progressively oxidize thymine as part of the thymidine salvage pathway in fungi (Smiley et al., 2005). The mammalian Tet proteins can execute a similar progressive oxidation of 5mC

producing first 5-formylcytosine (5fC) then 5-carboxylcytosine (5caC) (He et al., 2011; Ito et al., 2011). These additionally oxidized forms are considerably less abundant than either 5mC or 5hmC. In ESCs, approximately 0.002% and 0.0003% cytosines are in the 5fC and 5caC forms, respectively, compared to 0.13% in 5hmC (Ito et al., 2011), suggesting that the additional oxidation of 5hmC is either less efficient than the oxidation of 5mC or limited by some regulatory factor.

Interestingly, recent work demonstrated that 5hmC is restricted to the CG context in both ESCs (Yu et al., 2012) and cortical forebrain neurons (Lister et al., 2013), in contrast to abundant 5mC in both CG and CH contexts. This difference suggests that Tet activity may be specific for 5mC within the CG context or may only fully oxidize 5mC to 5fC and 5caC forms with present in the CH context. Either way, the presence of apparently stable levels of 5hmC in CGs suggests that the modification may itself act as an epigenetic mark.

Varieties of function

The complementary distribution of 5hmC with 5mC, its enrichment in specific tissues and genomic loci, and the conserved system of enzymes that regulate its production all suggest that 5hmC is an integral component of gene regulation. One early model proposed that 5hmC inhibits remethylation after DNA synthesis, thus initiating passive demethylation. Support for this model comes from *in vitro* data indicating that both UHRF1, a protein that tethers Dnmt1 to the replication fork (Bostick et al., 2007; Sharif et al., 2007), and Dnmt1 bind less efficiently to oligonucleotides containing asymmetrical 5hmC (Hashimoto et al., 2012; Valinluck and Sowers, 2007). However, clear evidence that this mechanism operates *in vivo* has not been demonstrated (Pastor et al., 2013).

The ability of the Tet enzymes to progressively oxidize 5mC provides a long-sought mechanism of methylation removal. Although demethylation pathways have been worked out for several histone modifications, no comparably well-defined pathway had been demonstrated for DNA methylation – a difference that supported views about the lability of histone modifications and stability of DNA modifications. Several putative mechanisms have been described, from deamination by AID (activation-induced cytidine deaminase) followed by DNA repair (Rai et al., 2008) to reverse catalysis by the *de novo* DNA methyltransferases (Kangaspeska et al., 2008; Métivier et al., 2008). However, recent work has provided strong evidence for Tet-mediated removal of methylation in operation in both ESCs and neural tissues (He et al., 2011). Both of the downstream products of this oxidation, 5fC and 5caC, can be repaired by DNA repair machinery. Thymine DNA glycosylase (TDG) can excise 5fC and 5caC, but has no activity against 5hmC (He et al., 2011; Maiti and Drohat, 2011). In addition AID can deaminate 5hmC to produce 5-hydroxymethyluracil, which can then be excised and repaired by TDG and the base excision repair pathway (Cortellino et al., 2011).

In fact, the potential of 5hmC to act as the first step toward demethylation, and the excitement that this potential generates, has resulted the broad interpretation of 5hmC biology within this context. However 5hmC may also be stably patterned in the genome and act as an epigenetic modification itself via either the recruitment or inhibition of trans factor binding to DNA. High levels of the modification in post-mitotic neuronal genomes support this alternate model since if 5hmC were simply an intermediate within a fully progressive oxidation pathway it would be rapidly depleted in the absence of new 5mC substrate. A caveat to this reasoning is that some portions of the genome may undergo a continuous methylation-demethylation cycle through the combined action of *de novo* DNA methyltransferases and dioxygenases. In addition, several

studies have identified proteins with 5hmC-binding activity, including several proteins classically studied for the role in 5mC interpretation. Mbd3, a protein containing a methyl-CpG binding domain, has stronger affinity for 5hmC than 5mC, and its depletion in ESCs results in the mislocalization of Tet1 and widespread alteration of the expression of Tet1 target genes (Yildirim et al., 2011). MeCP2, in a surprising twist, has been shown to bind to 5hmC with similar affinity as with 5mC, in contrast with an earlier report (Valinluck et al., 2004), and has been argued to be the dominant 5hmC binder in neurons (Mellén et al., 2012). Moreover a MeCP2 mutation (R113C) associated with Rett syndrome was shown to preferentially inhibit the binding of the protein to 5hmC, establishing an attractive connection between the neuronal enrichment of 5hmC and neurological defects that define Rett syndrome.

It is likely that both the ‘intermediate’ and ‘epigenetic’ models are partly correct and operate in distinct ways. A recent proteomics study that characterized the factors that bind to 5hmC, 5fC, and 5caC in mouse ESCs, *in vitro*-differentiated neural precursor cells (NPCs), and whole identified a protein (among hundreds of others) Uhrf2 that binds to 5hmC specifically in NPCs (Spruijt et al., 2013). The incubation of recombinant Uhrf2 with Tet1 and methylated DNA templates appears to limit the formation of 5fC and 5caC and leads to the accumulation of 5hmC (Spruijt et al., 2013), thus suggesting that Tet enzymes can mediate the oxidative cycle or stable 5hmC production depending on the presence of regulatory factors.

Although the mode of action is unclear, the alteration of Tet protein levels and accompanying changes to 5hmC patterning have dramatic effects on transcription. Most of the early work exploring the role of 5hmC in gene regulation was done in ESCs, and although some caution must be taken in extending these effects to other systems, this work provides important guidance about possible functions. A common theme emerging from these studies is the dual role of 5hmC

in either facilitating or repressing transcription. The knockdown of Tet1 results in the upregulation of repressed, PRC2 bound lineage-specific genes (Wu et al., 2011b). Correspondingly, 5hmC and Tet1 are enriched at bivalent promoters associated with developmentally transcription factors (Pastor et al., 2011; Williams et al., 2011; Wu et al., 2011a; Xu et al., 2011). Tet1 interacts with the transcriptional repressor Sin3a, providing a plausible mechanism underlying these effects (Williams et al., 2011). However, In addition to these repressive effects, the depletion of Tet1 downregulates a subset of actively transcribed genes, including several pluripotency-related factors (e.g. *Nanog* and *Esrrb*)(Ficz et al., 2011; Wu and Zhang, 2011).

The reduction of Tet activity in ESCs has controversial effects on their pluripotency. One study reported that the knockdown of Tet1 in ESCs causes an increase of methylation at the *Nanog* promoter resulting in reduced expression of the gene (Ito et al., 2010). Correspondingly, Tet1, Tet2, and global 5hmC levels decrease upon differentiation of ESCs. Moreover, knockdown of Tet1 in pre-implantation embryos resulted in bias toward trophoectoderm, supporting a role for 5hmC in maintaining the pluripotent state. However other studies report no alteration of ESC potency with reduced Tet activity (Dawlaty et al., 2013; 2011).

In contrast to the decrease of 5hmC levels upon differentiation of ESCs, 5hmC levels increase during neurogenesis with enrichment in both neuronal progenitors and mature neurons in the CNS (Ruzov et al., 2011). This increase appears to mediated by Tet3 which is expressed in somatic tissues including the nervous system (Ito et al., 2010). In particular, levels of 5hmC increase approximately 5-fold during postnatal development of the mouse hippocampus and cerebellum (Szulwach et al., 2011b). Interestingly, these increases are not strictly accompanied

by a reduction of 5mC (Hahn et al., 2013; Szulwach et al., 2011b), suggesting that Tet activity might be coupled to *de novo* methylation during neurogenesis, a model I explore in Chapter 3.

Neurogenesis of olfactory sensory neurons

These epigenetic and transcriptional transitions that contribute to the construction of neuronal cell identity are complex and an understanding of them requires the analysis of defined cellular populations within a model tissue. The main olfactory epithelium (MOE), the primary sensory tissue of the olfactory system, has several features that make it an excellent system for such analysis. Compared with the central nervous system, it contains a relatively accessible neuronal population that is of mostly homogenous identity (excepting the diversity of olfactory receptor expression). Moreover, as described in more detail below, the neuronal population is continuously regenerated allowing one to study neurogenesis at any time in the life of the animal.

The MOE is located at the dorsal-caudal portion of the nasal cavity in terrestrial mammals and together with the lamina propria form the olfactory mucosa, a highly convoluted structure composed of several turbinates. The MOE itself has, like all epithelia, a characteristic basal-apical orientation and cells positioned within largely distinct and ordered layers. It is also a ‘pseudostratified’ epithelium meaning that cells in most layers send projections to both basal and apical surfaces. The peripheral olfactory system is first formed from the olfactory placode, visible as a thickening of non-neural ectoderm that appear around embryonic day 9 (E9) in the mouse (Cuschieri and Bannister, 1975). Subsequently, the MOE appears at embryonic day 10 (E10) when the olfactory placode invaginates to form the olfactory pit. At E11, olfactory sensory neurons (OSNs), the primary sensory neurons of the tissue, appear and begin to send axonal projections in fascicles through the cribiform plate to the primordial olfactory bulb (OB) at the

most anterior portion of the forebrain. Also during this time, the primary layers of the epithelium are established.

The MOE is composed of several defined cell types arranged along the apical-basal axis and characterized by different developmental features and cellular functions (Fig. 1)(Frisch, 1967). Moving from apical to basal, we find the non-neuronal sustentacular cells (Sus)(Carr et al., 1991), followed by a relatively large compartment of bipolar OSNs, a thin layer of globose basal cells (GBCs), and finally a single layer of horizontal basal cells (HBCs) positioned directly superior to the basal lamina. In addition, Bowman's glands, the primary producers of mucus in the tissue, extend across the basal-apical axis. Olfactory ensheathing cells (OEC), which share characteristics with Schwann cells, reside in the lamina propria subtending the MOE and ensheath mOSN axons projecting to the OB (Ramón-Cueto and Avila, 1998).

OSNs can be broadly divided into two classes, immature (iOSNs) and mature (mOSNs), using specific markers. Immature OSNs are generally defined by expression of *Growth associated protein 43 (Gap43)*, a cytoplasmic protein associated with axonal outgrowth and synaptic plasticity (Benowitz and Routtenberg, 1997). In this stage, OR expression is first detected and the initial steps of axonal targeting and dendrite formation take place. Mature OSNs are conventionally defined by the expression of *Olfactory marker protein (Omp)*(Margolis, 1972) which is implicated in the functional maturation of OSN signaling (Lee et al., 2011). Maturation is in part driven by olfactory receptor expression and mOSNs are host to the stable expression of single OR (Dalton et al., 2013; Lyons et al., 2013). Thus, these two stages correspond to distinct developmental and functional and qualities: iOSNs with the initial activation of neuronal gene expression programs and relatively labile cellular definition, and mOSNs with relatively with stable expression programs and consistent transduction of olfactory information.

A defining feature of the MOE, one that makes it an excellent model to study neurogenesis, is that OSNs are continually replaced during the life of the animal (Graziadei and Graziadei, 1979). The average lifespan of OSNs is approximately three months (Moulton, 1974). Although there is some disagreement as to the cause of continual neuronal death in the MOE, it is likely that cell death is caused by environmental insults suffered to due to the relatively exposed position of the sensory organ and not due to an internal ‘clock’ that results in apoptosis and a predetermined time (Farbman, 1990). Because of this turnover, the multiple stages of neurogenesis – from progenitor maintenance and differentiation to axon guidance and the establishment of neuronal expression programs – can be studied within the adult animal. In effect, at a given time point each of the progenitor cells that contribute to the OSN population is present, allowing one separate out these developmental stages and analyze their epigenetic and transcriptional states. Moreover, the stratified epithelium character of the tissue allows one to visually distinguish developmental stages on the basis of basal-apical positioning.

GBCs, located immediately basal to the OSN layer, are a mitotically active population of neuronal progenitors (Caggiano et al., 1994). The dominant model of their developmental contribution describes GBCs are committed neuronal progenitors, thus dividing to only produced OSNs (Caggiano et al., 1994; Cau et al., 1997). However, there is some evidence that GBCs, particularly those considered more basal, can also generate non-neuronal sustentacular cells (Gokoffski et al., 2011), however this appears to be rare event. Approximately 98% of the mitotic activity within the MOE can attributed to the globose basal cells, which divide about once a day (Huard and Schwob, 1995). In the normally functioning, non-injured OE, GBCs go through two to three symmetric cell divisions to replace mOSNs lost through cell death (Mackay-Sim and Kittel, 1991).

One of the hallmarks of this population is the expression of cascade of proneural transcription factors that are essential to specification of neuronal identity (Nicolay et al., 2006). This cascade is made up of the successive expression of basic helix-loop-helix (bHLH) transcription factors (TFs) in daughter cells produced from several rounds of division. This sequence of bHLH expression during OSN neurogenesis is deeply conserved with parallels in other neurogenic systems, e.g. in *Drosophila* proneural clusters (Campuzano and Modolell, 1992) and mammalian embryonic neural tube (Lee, 1997). *Mammalian achaete scute homolog 1 (Mash1)* initiates the cascade (Gordon et al., 1995; Guillemot et al., 1993) and is required for downstream production of neuronal progenitors and olfactory sensory neurons (Cau et al., 1997; Murray et al., 2003), followed by *Neurogenin 1 (Neurog1)* immediately before the final division into OSNs, and finally *Neurogenic differentiation (NeuroD)* in iOSNs (Cau et al., 1997; 2002; Manglapus et al., 2004). Evidence for this succession is found from both genetic ablation of individual factors (i.e. loss of *Mash1* reduces *Neurog1* and *NeuroD* expression while loss of *Neurog1* affects only *NeuroD* and not *Mash1* expression (Cau et al., 2002) and the timing of bHLH TF expression after lesion of the MOE using methyl bromide (Manglapus et al., 2004).

Despite some early controversy surrounding the identity of the multipotent stem cell in the MOE (Duggan and Ngai, 2007), HBCs have emerged as a clear favorite (Iwai et al., 2008; Leung et al., 2007). These cells are normally quiescent, asymmetrically dividing approximately every 60 days (Mackay-Sim and Kittel, 1991), but become mitotically active when the tissue is injured, e.g. exposure to the damaging agents methimazole or methyl bromide (MeBr) (Schwob, 2002). This regeneration is remarkably thorough and rapid. Inhalation of low doses of MeBr eliminates greater than 95% of the Sus and OSN population, but by eight to ten weeks after exposure, the MOE is fully reconstituted (Hurt et al., 1988; Schwob et al., 1995).

In Chapters 2 and 3, I describe work investigating the roles of DNA modifications in gene regulation along this developmental lineage. Not much is known about DNA modification patterning in this system, especially 5hmC patterning. However, studies examining the expression patterns of the *de novo* DNA methyltransferases identified a developmental switch between *Dnmt3a* and *Dnmt3b* use (MacDonald et al., 2005; Watanabe et al., 2006). During embryonic and early postnatal development, *Dnmt3a* is expressed throughout the neuronal layer but is restricted to a tight band of immature olfactory sensory neurons in adults. On the other hand, *Dnmt3b* is broadly expressed in mitotically active progenitors cells (possibly GBCs) in embryos and sparsely expressed in the most basal layer (HBCs and possibly GBCs) in adults. This expression pattern suggests the *Dnmt3b* is necessary to establish ‘early’ (relative to this lineage) 5mC patterns, while *Dnmt3a* mediates 5mC patterning that is specific to neurons. Indeed, a fundamental role for *Dnmt3a* in establishing neuronal DNA modification patterns is described in Chapter 3.

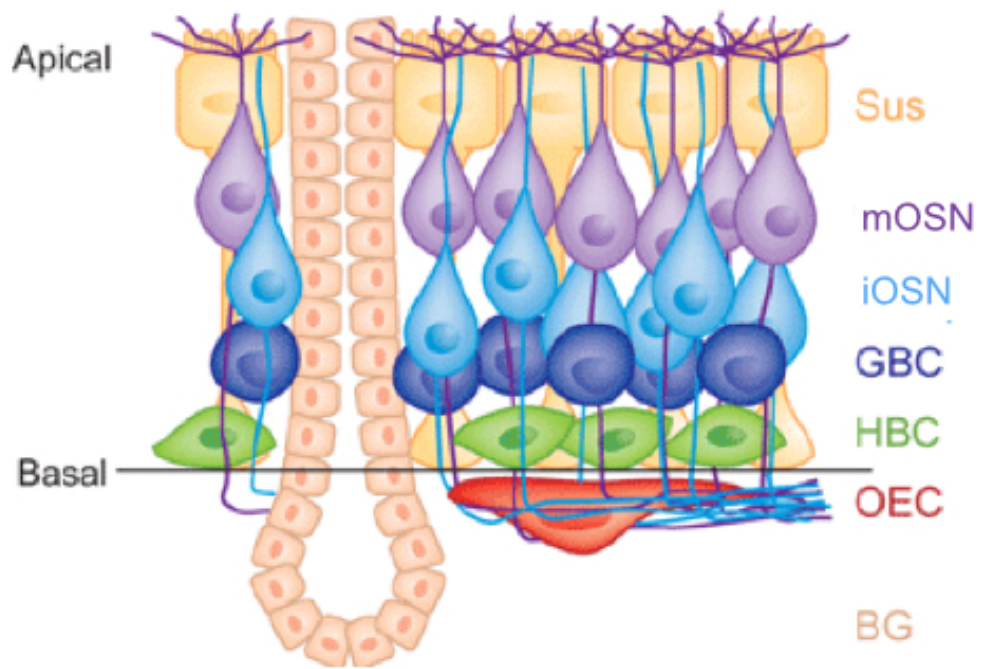


Figure 1. Schematic of the mouse main olfactory epithelium (MOE).

From apical to basal: non-neuronal sustentacular cells (Sus), mature olfactory sensory neurons (mOSN), immature olfactory sensory neurons (iOSN), globose basal cells (GBCs), horizontal basal cells (HBCs), olfactory ensheathing cells (OEC), and Bowman gland (BG). Figure adapted from {Duggan:2007ka}.

**Chapter 2: 5hmC patterning and functional role during olfactory
sensory neuron development**

Published title: Alteration of genic 5-hydroxymethylcytosine patterning in olfactory neurons correlates with changes in gene expression and cellular identity.

Bradley M. Colquitt¹, William E. Allen², Gilad Barnea², Stavros Lomvardas³

¹Neuroscience Graduate Program, University of California, San Francisco, CA, 94158, USA

²Department of Neuroscience, Brown University, Providence, RI 02912, USA

³Department of Anatomy, University of California, San Francisco, CA 94158, USA

Abstract

The modified DNA base 5-hydroxymethylcytosine (5hmC) is enriched in neurons where it may contribute to gene regulation and cellular identity. To determine how 5hmC influences gene expression in an *in vivo* neuronal population, we assessed the patterning and function of the base along the developmental lineage of the main olfactory epithelium – from multipotent stem cells through neuronal progenitors to mature olfactory sensory neurons (mOSNs). We find that 5hmC increases over gene bodies during mOSN development with substantial patterning between the progenitor and mOSNs stages. Although gene body 5hmC levels correlate with gene expression in all three developmental cell types, this association is particularly pronounced within mOSNs. Overexpression of Tet3 in mOSNs markedly alters gene body 5hmC levels and gene expression in a manner consistent with a positive role for 5hmC in transcription. Moreover, Tet3 overexpression disrupts olfactory receptor expression and the targeting of axons to the olfactory bulb, key molecular and anatomical features of the olfactory system. Our results suggest a physiologically significant role for gene body 5hmC in transcriptional facilitation and the maintenance of cellular identity independent of its function as an intermediate to demethylation.

Introduction

The functional repertoire of DNA modifications has expanded recently with the discovery of a derivative of 5mC, 5-hydroxymethylcytosine (5hmC), in mammalian genomes (Kriaucionis and Heintz, 2009b; Tahiliani et al., 2009). This modified base is enriched in embryonic stem cells (ESCs) and several adult tissues, with particular enrichment within the nervous system and bone marrow (Kriaucionis and Heintz, 2009b; Ruzov et al., 2011). The modification is generated by a conserved family of enzymes, the Tet family of methylcytosine dioxygenases (Tet1, Tet2, and Tet3), which show variable levels of expression across tissues (Ito et al., 2010; Tahiliani et al., 2009). Work in neural systems (Song et al., 2011; Szulwach et al., 2011b) and ESCs (Ficz et al., 2011; Pastor et al., 2011; Stroud et al., 2011; Szulwach et al., 2011a; Wu et al., 2011a) has found that 5hmC is concentrated on putative intergenic enhancers and gene bodies where levels of the base positively correlate with transcription.

Despite substantial work in ES cells describing the localization of 5hmC, its association with protein factors, and the effects of its manipulation on gene expression, it is still unclear how the modified base influences gene expression within a neuronal system. Several models describe 5hmC as an intermediate to a demethylated state (Branco et al., 2011; Cortellino et al., 2011; Guo et al., 2011a; Ito et al., 2010; Valinluck et al., 2004; Williams et al., 2011; Wu and Zhang, 2011). However, it remains possible that 5hmC influences transcription not only via the eventual lack of cytosine modification but directly through interactions with the chromatin environment or through the recruitment of binding proteins. Recent work suggests that 5hmC patterning persists at some loci during the development of the central nervous system (Szulwach et al., 2011b). In addition, the recent identification of protein factors that bind 5hmC suggests that the modified

base may be interpreted directly as an epigenetic mark (Mellén et al., 2012; Spruijt et al., 2013; Yildirim et al., 2011).

It is difficult to assess the stability and function of the modified base in a mitotic population, such as ESCs where continual division may mask persistent 5hmC patterning, or in a fully developed tissue, such as the adult brain where comparisons with previous developmental stages are challenging. To better understand how 5hmC influences developmental gene regulation, we characterized its patterning and function in an *in vivo* developmental system, the mouse main olfactory epithelium (MOE). The MOE continuously regenerates over the life of the animal, such that the major neurogenic cell types are simultaneously present in adulthood. Using genetic and cell surface markers, we isolated distinct neurodevelopmental stages from the MOE and defined their 5hmC, 5mC, and transcriptional states through next-generation sequencing approaches. Our analysis focused on three cell types representing three stages of neurodevelopment. Horizontal basal cells (HBCs), located in a single layer along the basal lamina, are quiescent multipotent cells that can repopulate all the cell types of the MOE in response to injury (Leung et al., 2007). Globose basal cells (GBCs), located apical to the HBC layer, are mitotically active progenitors and can produce both neuronal and non-neuronal cell types (Huard et al., 1998). Finally, mature olfactory sensory neurons (mOSNs) are terminally differentiated primary sensory neurons and comprise the majority of cells within the adult MOE.

Results

To initially assess global changes of 5mC and 5hmC abundances across MOE neurogenesis, we stained sections of the MOE from P0 animals using antibodies that specifically recognize the two modified bases (Ito et al., 2010). 5hmC is substantially enriched in neurons and neuronal progenitors relative to basal stem cells and apical, non-neuronal sustentacular cells (Fig. 2A, top). In contrast, global levels of 5mC are relatively constant across the tissue. Within the nucleus (Fig. 2A, bottom), 5hmC and 5mC occupy largely complementary compartments: 5hmC is diffusely localized within the nucleus yet excluded from the central pericentromeric heterochromatin focus (Clowney et al., 2012), where 5mC is enriched. This apparent increase in 5hmC levels during differentiation coincides with the upregulation of Tet3, the most highly expressed Tet family member in the MOE (Fig. 2B).

To develop genomic maps of 5hmC across MOE neurogenesis, we sorted *Olfactory marker protein* (OMP)-expressing mOSNs, *Neurogenin1*-expressing GBCs, and *Inter-cellular Adhesion Molecule 1*-expressing HBCs from the MOEs of eight-week old mice using genetic and cell surface markers (Fig. 8A) (Magklara et al., 2011). We then purified genomic DNA from these populations and performed DNA immunoprecipitation coupled to next-generation sequencing using anti-5hmC antibodies (hmeDIP-seq, see Table 1 for metadata). Two biological replicates of each hmeDIP-seq library were generated. Unless otherwise noted, subsequent analyses were performed using averages of the replicates. We also performed meDIP-seq in these populations to follow the relative changes of 5hmC/5mC ratios during differentiation. The ability of these antibodies to discriminate among cytosine-, 5mC- and 5hmC-containing templates was verified by DIP-qPCR (Fig. 8B). To correlate epigenetic signatures with transcriptional state, we used expression analyses performed in these cell lines by RNA-seq (Magklara et al., 2011).

Genome-wide correlations of the DNA modification datasets give an initial impression of the dynamics and stability of the modifications across mOSN development. In agreement with the developmental relationships between the stages, 5hmC levels are more highly correlated between mOSN/GBC (Pearson $R = 0.78$) and GBC/HBC ($R = 0.75$) stages than those in mOSN/HBC ($R = 0.70$). In addition, differential peak analysis of 5hmC indicates that most *de novo* patterning occurs within mOSNs (mOSN/HBC peak number = 6,746, GBC/HBC = 2,145) and is strongly associated with exons and introns (Fig. 2C).

5hmC levels within the gene body (from transcription start site, TSS, and transcription end site, TES) positively correlate with transcription in mOSNs (Pearson $R = 0.21$, $p < 2.2E-16$) and to a lesser extent in GBCs and HBCs ($R = 0.16$, $p < 2.2E-16$; $R = 0.14$, $p < 2.2E-16$, respectively; Fig. 3A and B, and Fig. 9A). As in other systems (Szulwach et al., 2011b; Wu et al., 2011a), we observe a dichotomy between the correlation of 5hmC levels and transcription that depends on genic location: although gene body 5hmC levels linearly increase with transcription across the full range of the transcriptional output (Fig. 3C), the TSSs of active genes are depleted of the modified base (Fig. 3D). It is possible that the differential patterning of 5hmC on the TSS and gene body produces differential effects on transcription. To test this model, we performed a cell-free transcription assay using a linear template with a CMV promoter or transcribed region containing unmodified cytosines, 5mC, or 5hmC (Fig. S1D). We find that the presence of either 5mC or 5hmC in the promoter region has a deleterious effect on transcription yet have little consequence when they are restricted to the transcribed region, suggesting that 5hmC within the gene body may require interactions with the chromatin environment to exert an effect.

The significant increases of 5hmC seen over gene elements as OSN differentiation progresses suggest that 5hmC patterning may contribute to the definition of the neuronal progenitor and

neuronal stages. To further examine the developmental patterning of 5hmC, we characterized 5hmC levels associated with developmentally regulated genes. Inspection of the modification patterning across two representative genes, *Ncam1* and *Egfr*, which are differentially expressed between HBCs and mOSNs reveals a dynamic interplay between 5hmC and 5mC at different developmental stages (Fig. 4A). Substantial enrichment of 5hmC is seen over the *Ncam1* gene body only within the cell types in which it actively transcribed. Accordingly, 5mC is depleted from the 5' portion of the gene upon its transcription in mOSNs. A similar though weaker enrichment of 5hmC is seen across the *Egfr* locus only in HBCs, while 5mC gene body levels remain unchanged. 5hmC and 5mC DIPs using independent sorted populations and qPCR analysis of five differentially expressed genes (*Ncam1*, *Ebf1*, *Neurog1*, *Egfr*, and *Notch2*) supports this pattern (Fig. 10A and B).

The positive correlation between 5hmC gene body levels and transcription suggests that 5hmC patterning may simply reflect the transcriptional status of a gene within a given cell type. To determine if 5hmC levels faithfully track with differential expression across the three cell types, we generated five sets of differentially expressed genes – mOSN/GBC-common, GBC/HBC-common, mOSN-specific, GBC-specific, and HBC-specific (Table S2) – and computed the 5hmC profiles flanking the TSSs of each set (Fig. 4B-F). mOSN/GBC-common genes experience an elevation of gene body 5hmC in the GBC and mOSN stages (Fig. 4B) with higher levels in the mOSN stage. Similarly, mOSN- (Fig. 4D) and HBC-specific (Fig. 4F) genes exhibit elevated gene body 5hmC within the mOSN and HBC populations, respectively, again supporting the general association between 5hmC levels and gene expression. However within the bodies of GBC-specific (Fig. 4E) genes, the timing of 5hmC elevation is more complicated. These genes exhibit slight increases in 5hmC levels within the GBC stage that then persist into

the mOSN stage. Thus, transcriptionally downregulated genes may retain elevated levels of 5hmC within mOSNs, indicating that, although there is a close association between gene body 5hmC levels and transcription, the presence of the modified base is likely not sufficient for gene activation.

The dynamic patterning of gene body 5hmC across the mOSN developmental lineage suggests that the modified base plays a key role in defining the identity of differentiated cell types. To test this model we employed a genetic strategy to alter the 5hmC landscape of mOSNs. It is likely that the modification state of each locus reflects the steady state of methylation, hydroxymethylation, and demethylation reactions. Therefore, the increase of Tet3 activity in mOSNs may bias loci with low 5hmC levels toward increased hydroxymethylation and loci with high 5hmC levels toward demodification. With this model in mind, we generated a tetO-Tet3-ires-EGFP transgenic mouse, which we crossed into a background containing OMP-ires-tTA to drive expression in mOSNs. This genetic strategy allowed specific expression of transgenic Tet3 (Tet3-tg) and the bicistronic GFP reporter in approximately half of mOSNs ($49\pm 12\%$ as determined by FACS). By immunofluorescence, Tet3 protein level is increased in Tet3-tg expressing MOEs and colocalizes with increased levels of 5hmC (Fig. 5A).

To assess the consequences of Tet3 overexpression in these neurons, we FAC-sorted two biological replicates of control (OMP-ires-GFP) and Tet3-transgenic (Tet3-tg) mOSNs and analyzed the patterning of 5hmC by hmeDIP-seq and expression levels by rRNA-depleted RNA-seq. Differential 5hmC peak analysis indicates significant alterations of 5hmC levels within regions 1 to 3 kb upstream of TSSs, CDSs, 3' UTRs, and introns (Fig. 5B). Interestingly, these regions exhibit both increases and decreases of 5hmC depending on control levels of 5hmC (Table S3). Both exons (Fig. 5C) and introns (Fig. 5D) with moderate levels of 5hmC in control

mOSNs (middle 10%) exhibit a slight gain of the mark in Tet3-tg (Tet3-tg / control 5hmC ratio \pm s.d. exons = 1.38 \pm 0.57, Wilcoxon $p < 2.2E-16$; introns = 1.23 \pm 0.42, Wilcoxon $p < 2.2E-16$), while those with high levels (top 10%) exhibit a significant reduction (exons = 0.49 \pm 0.23, Wilcoxon $p < 2.2E-16$; introns = 0.56 \pm 0.25, Wilcoxon $p < 2.2E-16$). These opposing effects are consistent with the ability of Tet3 to oxidize 5hmC as well as 5mC; genes with high levels of 5hmC in control mOSNs may recruit sufficient levels of Tet3 to induce further oxidation, while those with moderate levels achieve only partial oxidation from 5mC to 5hmC. Moreover, gene body 5hmC levels are shifted toward 3' ends (Fig. 5D), perhaps reflecting differential recruitment of Tet3 by 5'-enriched transcriptional elongation complexes. Regions that lose 5hmC do not exhibit a complementary increase of apparent fC (Fig. 11A), suggesting that the loss of 5hmC does not result in the enrichment of additionally oxidized forms but a return to an unmodified state. In contrast to the bidirectional changes to 5hmC levels in gene bodies, Tet3-tg 5hmC levels at promoters and TSSs are generally reduced (Fig. 11B) and do not exhibit a complex dependency on control 5hmC levels (Fig. 11C).

Two example gene loci, *Stxbp2* and *Plxna1*, demonstrate both the increase and decrease, respectively, of gene body 5hmC levels in Tet3-tg mOSNs (Fig. 6A). Interestingly, the expression levels of these genes track with Tet3-induced 5hmC changes. An analysis of differentially expressed genes (Bayes factor ≥ 20 , see Methods and Table S4) indicates that transcriptionally downregulated genes (Fig. 6B, N=1621) exhibit a significant depletion of gene body 5hmC (Wilcoxon $p = 5.7E-18$), while upregulated genes (Fig. 6C, N=938) exhibit a modest, but insignificant, increase of gene body 5hmC. We examined the general association between gene body 5hmC and expression by ordering genes according to their gene body 5hmC levels in either control or Tet3-tg mOSNs using principal component analysis (Fig. 6D and E). Strikingly,

the fold change of control to Tet3-tg mOSN gene expression reflects alterations to gene body 5hmC: genes that lose gene body 5hmC are substantially downregulated while those that gain 5hmC are slightly upregulated. Notably, these transcriptional effects positively correlate with 5hmC levels at the gene body (Pearson $R = 0.12$, $p < 2.2E16$) and not the TSS (± 500 bp of TSS, $R = -0.02$, $p = 0.01$).

The most highly expressed mOSN-specific genes, such as olfactory receptors (ORs) and the guidance molecules involved in the wiring of olfactory neurons, are transcriptionally downregulated in Tet3-tg mOSNs (1071 of 1082 expressed ORs and 21 of 25 expressed guidance molecules, Fig. 7A and Table S4). By immunofluorescence, we find no overlap between OR-expressing mOSNs and Tet3-tg mOSNs (Fig. 7B and C). Consistent with the general effects described above, 5hmC is significantly depleted within gene bodies encoding guidance molecules in Tet3-tg mOSNs (Fig. 7D). The same relationship between OR downregulation and 5hmC levels cannot be established for ORs likely due to technical reasons; each OR is expressed in only 0.1% of the cells in the MOE and activating epigenetic marks cannot be detected on these genes in mixed mOSN populations (Magklara et al., 2011). Moreover, OR genes have a high AT content (Clowney et al., 2011) which reduces the number of modifiable cytosines. As a result the levels of 5hmC on OR genes are extremely low in this mixed mOSN population, possibly below the sensitivity of hmeDIP.

The proper expression of these genes governs the correct targeting of OSN axons to discrete glomeruli in the olfactory bulb. To examine if the transcriptional misregulation caused by Tet3 overexpression disrupts this targeting, we analyzed *Olf151* glomerulus formation in the bulbs of *Olf151-Cre/floxed-Tomato* mice in the presence or absence of Tet3 transgene expression (Fig. 7E). In every case examined (N=3), Tet3 overexpression produced mistargeted axons and the

formation of additional glomeruli. Consistent with the RNA-seq and IF data, there is no detectable overlap between GFP-positive and Tomato-positive axons (Fig. 7E), indicating a lack of *Olf151* expression in Tet3-tg mOSNs. It is therefore likely that the targeting deficit seen in Tomato-positive axons is due to non-cell autonomous disruption of the glomerular map.

Discussion

Our experiments describe the alterations to the 5mC and 5hmC genomic landscapes across the neurogenic lineage of the main olfactory epithelium from multipotent stem cell to mature olfactory sensory neuron. Within the MOE, 5hmC is globally enriched in the genomes of neuronal progenitors and neurons, consistent with previous reports assaying 5hmC levels in neural tissues (Ruzov et al., 2011; Szulwach et al., 2011b). In concordance with the general positive correlation between transcription and gene body 5hmC levels, gene body 5hmC is associated with cell type-specific gene expression. However, genes that are active in progenitors but silenced in neurons retain elevated levels of gene body 5hmC in the neuronal stage, suggesting that gene body 5hmC patterning in neurons does not always simply reflect transcriptional status.

We note different relationships between 5hmC levels and transcription depending on the location of the modification. The depletion of 5hmC at the TSSs of transcribed genes is consistent with a role for this modification in the demethylation of these sites. Although the genomic occupancy of the Tet family members has not been determined in a neuronal system, several studies have characterized the binding of Tet1 in ESCs where the protein appears to be predominantly localized at TSSs (Williams et al., 2011; Wu and Zhang, 2011). In addition to their ability to convert 5mC to 5hmC, the Tet family members are able to catalyse the conversion to subsequent oxidative states, 5-formylcytosine (5-fC) and 5-carboxylcytosine (5-caC), which may be acted on by DNA repair machinery to produce a fully demodified state (He et al., 2011; Ito et al., 2011). Therefore, the high occupancy of Tet at TSSs may drive full oxidation and demodification within these regions. In contrast, Tet protein occupancy within gene bodies and enhancers may be relatively transitory, preventing subsequent oxidation from occurring, or the interaction of the

Tet proteins with regulatory factors may alter its activity. Therefore, 5hmC is likely a multifunctional modification that is employed in either the demethylation pathway or as an epigenetic modification itself depending on the genomic location of the mark. Consistent with differential role of 5hmC in TSSs versus within the promoter and TSS abolishes transcription, whereas gene body hydroxymethylation has no measurable effects. Although these experiments cannot exclude that gene body hydroxymethylation is just a byproduct of transcriptional elongation, an attractive hypothesis proposes that 5hmC facilitates transcriptional elongation in the context of chromatin.

In this vein, the effects of Tet3 overexpression in mOSNs on 5hmC patterning and gene transcription are consistent with a direct role for 5hmC in gene regulation. The loss of gene body 5hmC resulted in a reduction of expression, while the gain of 5hmC associated with increased expression. This relationship extended on average across all genes and does not appear to be confined to defined gene classes, such as housekeeping or developmentally regulated genes. Although we are unable to exclude indirect effects of Tet3 overexpression, such a general association suggests that gene body 5hmC patterning is a common modulator of transcriptional output, perhaps directly altering the local chromatin environment or indirectly recruiting recognition factors. Recent studies indicate that both Mbd3 (Yildirim et al., 2011) and MeCP2 {Mellen, 2012 #304} (Mellén et al., 2012) can bind 5hmC. Interestingly, as in other neuronal cell types, MeCP2 expression is highly upregulated in mOSNs (MacDonald et al., 2010), and its loss leads to mOSN axons targeting deficits (Degano et al., 2009). The modulatory character of gene body 5hmC levels suggests that the precise tuning of these levels may be necessary to maintain constant levels of expression within mature, post-mitotic populations such as olfactory sensory neurons. Indeed, initial OR expression and axonal targeting occurs just before the mOSN stage

and the onset of Tet3 overexpression, suggesting that the effects seen here are due to the disruption of mechanisms that maintain gene expression and not those that initiate it. Through this control, gene body 5hmC patterning may aid in establishing stable cellular identities.

Methods

Animal care and use. Mice were treated in compliance with the rules and regulations of IACUC under protocol approval number AN084169-01.

Immunofluorescence. Sections of postnatal C57BL/6 MOE were fixed with 3:1 methanol:acetic acid, fragmented with 1N HCl, and neutralized with 100 mM Tris pH 8. Sections were blocked then incubated with anti-5mC (#39649) and anti-5hmC (#39791) antibodies (Active Motif) overnight at 4°C. Tet3 immunostaining was performed on coronal MOE sections fixed with 4% PFA and using anti-Tet3 antibody (Santa Cruz, sc-139186).

Fluorescence activated cell sorting. Mature OSNs were isolated from the MOEs of OMP-ires-GFP knockin animals (Shykind et al., 2004). GBCs were isolated from the MOEs of Ngn-GFP BAC transgenics (Gong et al., 2003). HBCs were isolated with anti-mouse CD54 (*Icam1*) conjugated to PE (Biolegend). Isolation and preparation were performed as described in (Magklara et al., 2011). Briefly, single cell suspensions from 2-10 MOEs were made using the Papain dissociation system (Worthington). MOE were dissected into Earl's Buffered Saline Solution (EBSS), minced in papain dissociation solution (Worthington), and incubated at 37°C for 45 minutes. Cells were washed once in 1:10 inhibitor solution and once in PBS. For a given assay (DIP-seq/qPCR or RNA-seq), 2E5 to 1E6 cells were used. Population purity was assayed through a limited resorting of the sorted cells. Purities of ~90-95% were routinely achieved.

meDIP and hmeDIP-seq. Genomic DNA was isolated from the cell populations described in the main text, prepared for single- or paired-end Illumina sequencing using standard protocols, and immunoprecipitated as in (Weber et al., 2005) with minor modifications. Specifically, sorted cell gDNA was sonicated to 200-1000 bp with a mean size of ~400 bp and end repaired (NEBNext

End Repair). Adenosine was added to 3' ends (Klenow, 3' to 5' exo-, NEB), and 10:1 Illumina adaptor:samples were ligated onto the samples. 20-50 ng of adapted genomic DNA was diluted in MB (10 mM phosphate buffer pH 7.0, 140 mM NaCl, 0.05% Triton X-100), denatured at 95°C for 10 minutes, brought to 300 μ L MB, and precipitated with 0.5 μ g anti-5mC or 5hmC antibodies (as used in immunofluorescence) overnight at 4°C. Separately, 10 μ L per IP of Protein A and G Dynabeads (Invitrogen) were blocked with 2 mg BSA and 2 mg yeast tRNA in 1 mL MB overnight at 4°C. Beads and IPs were combined the next day and rotated for 3 hours at 4°C. Samples were washed 4 times in MB, eluted twice in 100 μ L elution buffer (0.1 M NaHCO₃, 1% SDS, 1 mM EDTA) by shaking at 37°C for 15 minutes, phenol:chloroform:isoamyl extracted, ethanol precipitated, and resuspended in 50 μ L TE. Samples were then amplified with Illumina single-read or TruSeq PCR primers for 15 cycles using Phusion DNA polymerase (NEB).

Quantitative PCR. DIP libraries were prepared as above and analyzed using Maxima SYBR Green (Fermentas).

RNA-seq. Total RNA was isolated from a given cell population or tissue using Trizol (Invitrogen) followed by TURBO DNase treatment (Ambion). mRNA sequencing libraries were prepared using either Superscript III reverse transcription using a poly-dT primer (Invitrogen) followed by addition of adaptors using standard techniques or FastTrack MAG Micro mRNA isolation (Invitrogen) followed by ScriptSeq (Epicentre). Control (OMP-GFP) and Tet3-tg (OMP-tTA / tetO-Tet3-EGFP) RNA-seq samples were prepared from rRNA-depleted total RNA (Invitrogen, RiboMinus) using ScriptSeq v2. Samples were amplified for 15 cycles using Phusion DNA polymerase (NEB) and Illumina single-read or ScriptSeq primers.

TetO-Tet3-ires-EGFP transgenic line. Full-length Tet3 was PCR amplified from MOE cDNA and cloned into the pTRE2 vector (Clontech) containing ires-EGFP. Transgenic lines were generated by pronuclear microinjection of the linearized construct into FVB founders. Transgene expression was driven by crossing into the OMP-tTA line (Yu et al., 2004).

In vitro transcription. Modified promoter and transcribed regions were amplified from the HeLa Scribe positive control (Promega) using 5mC or 5hmC dNTP mixes (Zymo). Mixed modification templates were obtained by digesting both fragments with BglII, immobilizing the biotinylated promoter with streptavidin conjugated to magnetic beads (Invitrogen, M-280), ligating promoters and transcribed regions (T4 DNA Ligase, NEB), and magnetically pulling down the ligation products. *In vitro* transcription was performed on these products using the HeLa Scribe Kit (Promega) with α -³²P-UTP. Reactions were purified then run on a 6% acrylamide TBE-urea gel.

Sequencing alignment and initial processing. Read mapping was performed using Bowtie2 (Ben Langmead and Salzberg, 2012) (default parameters, mm9). PCR duplicates and reads with MAPQ<30 were removed. Tracks were generated consisting of read counts binned into 25 bp normalized by the number of millions of reads and scaled to reflect the average number of reads per 1kb to obtain reads per million (RPM, i.e. $1E6 * 1E3 / (\text{window size} * \text{total number of reads})$). For visualization, biological replicates were averaged and tracks were smoothed using a 125 bp moving average window. RNA-seq mapping was performed using Tophat2 (Kim et al., 2013)(no GTF). Reads with MAPQ<30 were removed.

Genome-wide correlations. Pairwise Pearson correlation values were computed as in (Wang et al., 2008). Read data binned into 1000 bp windows were divided into groups of 100, correlation

values were computed within these groups, and these values were averaged to obtain a genome-wide correlation value.

Differential peak analysis. Regions with differential enrichment of 5hmC and 5mC were identified using HOMER (Heinz et al., 2010). Tag directories (GC-normalized) for each sample and replicated were created independently from processed BAM files. Differential regions were called within either GBC or mOSN using HBC as input, 'histone' presets, peak size of 1000 bp, and at least 3 fold enrichment over input (`findPeaks -style histone -size 1000 -F 3`). Control and Tet3-tg 5hmC differential peaks were similarly identified. Differential peak sets from each replicate were intersected (requiring an overlap of at least 50% of peak size) to create common peak sets for feature intersection analysis. Genomic positions for gene elements (*1 to 3 kb upstream, 1 kb upstream, CGI, CDS, 3' UTR, 5' UTR*) were obtained from the mm9 UCSC refGene table. Intergenic sites (*intergenic*) were determined determining the complement of refGene regions extended by 5kb. Repetitive elements were subtracted from these regions using the mm9 rmsk table. Conserved intergenic regions (*intergenic conserved*) were obtained by intersecting this intergenic set with the vertebrate conserved elements (`phastConsElements30way`).

Aligned feature profiles. For aligned gene profiles, refGene position information was obtained from the UCSC genome browser, and 50 200 bp windows were generated upstream, downstream, and spaced evenly throughout gene bodies. TSS profiles were generated using 25 bp windows extending from the 5' most annotated TSS for each gene. For each profile, the mean of a given aligned position was computed (excluding values at extreme 1%) and 95% confidence intervals were calculated through bootstrap resampling for 1000 iterations using the R package `boot`. Significant differences at a given aligned position between two samples were computed by

permutation as follows. Positions were binned by 8 average profile values were computed in these bins. For a given position and set of two samples, the differences between the mean values from 1000 label randomizations were computed. The number of differences from the randomized data exceeding the true difference was used as p-values. FDRs were computed from these p-values using the Benjamini-Hochberg method.

Analysis of 5hmC/5mC and transcription. Fragments per kilobase per millions of reads (FPKM) values were computed for each sample using Cufflinks (Trapnell et al., 2010) with assembly to the iGenomes Ensembl GTF file and multi-read correction. Aligned gene profiles were grouped into quartiles by associated FPKM values. Transcriptional indices were generated by grouping FPKM-ranked genes into 100 groups (143 genes each). Guidance molecules here are defined as genes with these prefixes: *Efn*, *Eph*, *Dscam*, *Nphs*, *Kirrel*, *Nrp*, *Plxn*, *Sema*, *Pcdh*, *Ctnn*, *Robo*, and *Slit*.

Differential expression analysis. To define sets of genes with developmentally differential expression, the expression of status of each gene was determined using normal mixture model-based clustering via the R package *mclust*. Cufflinks-generated $\log_2(\text{FPKM}+1)$ values were fit independently for each sample to a two Gaussian mixture model with equal variance (R command: `Mclust(values, G=2, modelName="E")`). The resulting classification was used to generate ‘specific’ and ‘common’ gene expression sets. To identify genes with differential expression, we followed the approach used in (Katz et al., 2010), which applied a two-sided point null hypothesis test. For each gene g in our data D , we computed $d_g = A_g - B_g$, where A_g and B_g are the FPKM of g in the two samples A and B . The null hypothesis H_0 states $d_g = 0$ and the alternative hypothesis H_g states $d_g \neq 0$. To select between the two hypotheses, we computed the Bayes Factor $BF = \frac{p(D|H_1)}{p(H_1)}$ /

$p(D|H0)p(H0)$. Specifically, we applied a Kernel Density Estimator to compute the distribution $p(d | D, H1)$, and estimated each Bayes Factor $BFg \cong 1/p(dg=0 | H1, D)$ (the Savage-Dickey density ratio with a prior density $p(dg = 0 | H1) = 1$).

Principal component analysis. Gene body 5hmC levels were sampled for each gene using 50 200bp windows spaced from TSS to TES (genes less than 10 kb long were excluded). Principal component analysis was performed on the resulting matrices using the R function *prcomp*. The original data matrix was then ordered by principal component 1 or 2. Ordered matrices were then averaged by every 50 genes and plotted as heatmaps.

Acknowledgments: This work was supported by the NIH (DP2 OD006667, R01MH091661), NSF (Graduate Research Fellowship), and the Rett Syndrome Trust.

Figures

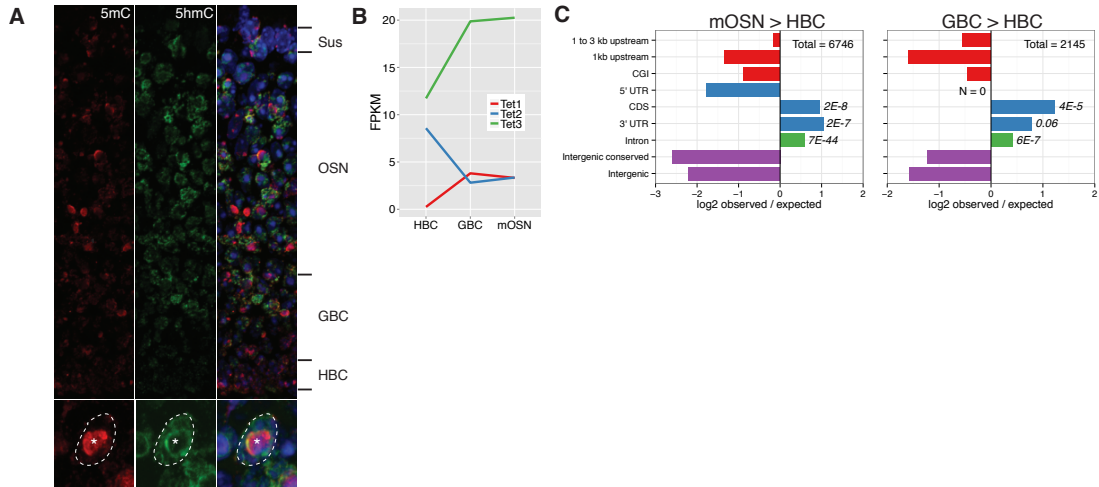


Figure 2. Developmental 5hmC patterning in the MOE.

(A) top, MOE coronal section from a P0 mouse stained for 5hmC (green), 5mC (red) and DNA (DAPI, blue). Major developmental layers are indicated. bottom, outlined OSN nucleus, * pericentromeric condensation. (B) Tet family member expression levels within HBCs, GBCs, and mOSNs. (C) Intersections of differential 5hmC peaks in mOSN vs. HBC (left) and GBC vs. HBC (right) with genomic features. Differential peaks were identified in each biological replicate individually then merged for a consensus set. Values are log₂ ratios of observed versus expected number of peaks within a given region. Significance values are one-sided Fisher-test false discovery rates.

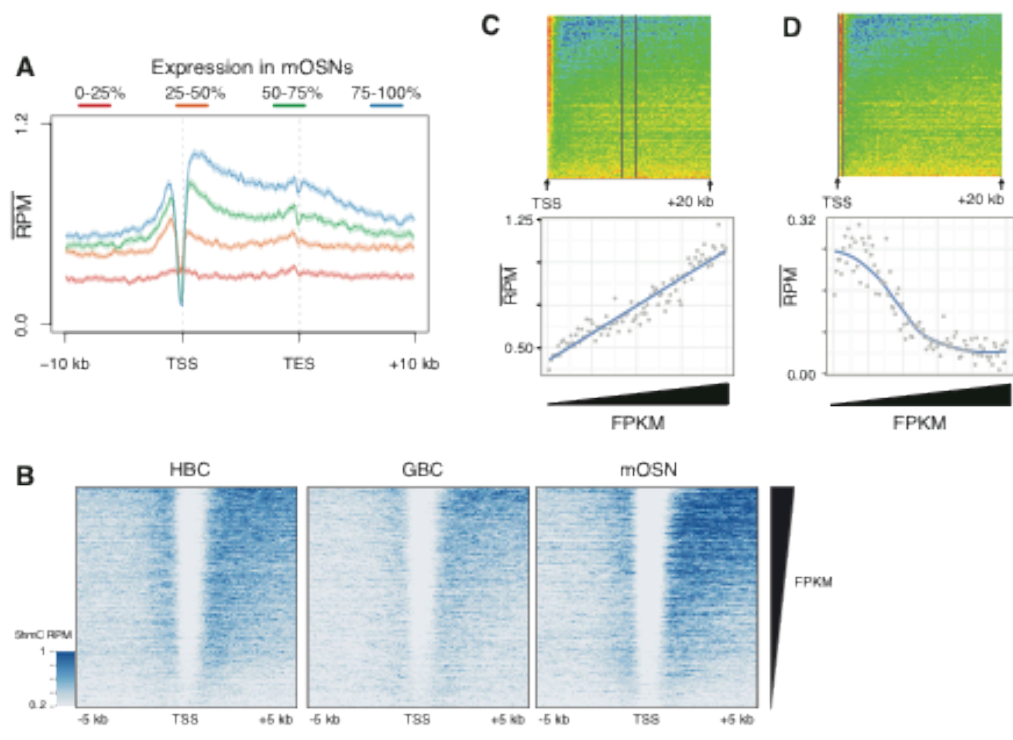


Figure 3. Relationships between 5hmC levels and gene transcription.

(A) Aligned gene profiles of mOSN 5hmC levels split by mRNA FPKM quartiles in mOSNs. Error is 95% bootstrap confidence interval. (B) HBC, GBC, and mOSN 5hmC flanking TSSs sorted by FPKM in each cell type. (C, D) Mean mOSN 5hmC values within regions extending (C) 9 kb to 11 kb within the gene body and (D) from the TSS to 500 bp downstream. X-axis is index of increasing FPKMs. In (C), linear regression adjusted R-squared = 0.86.

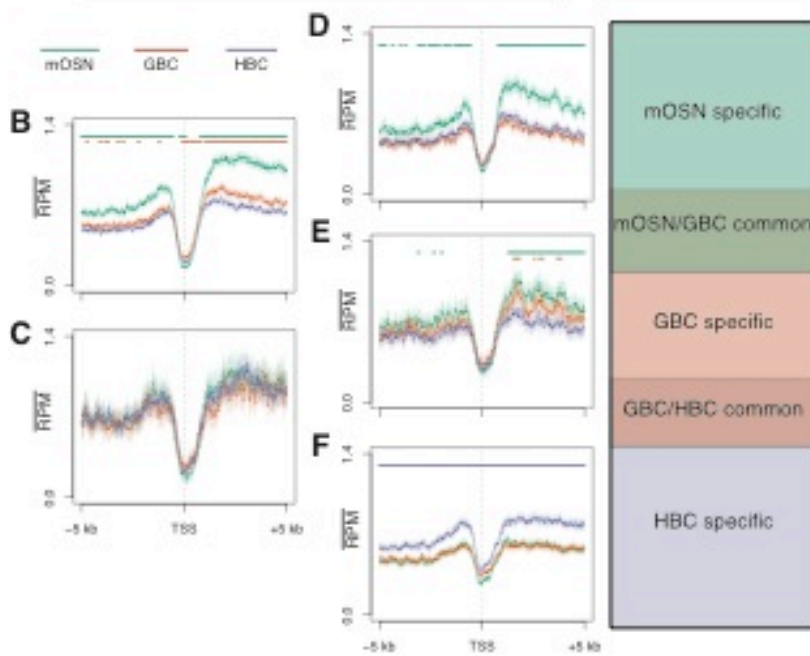
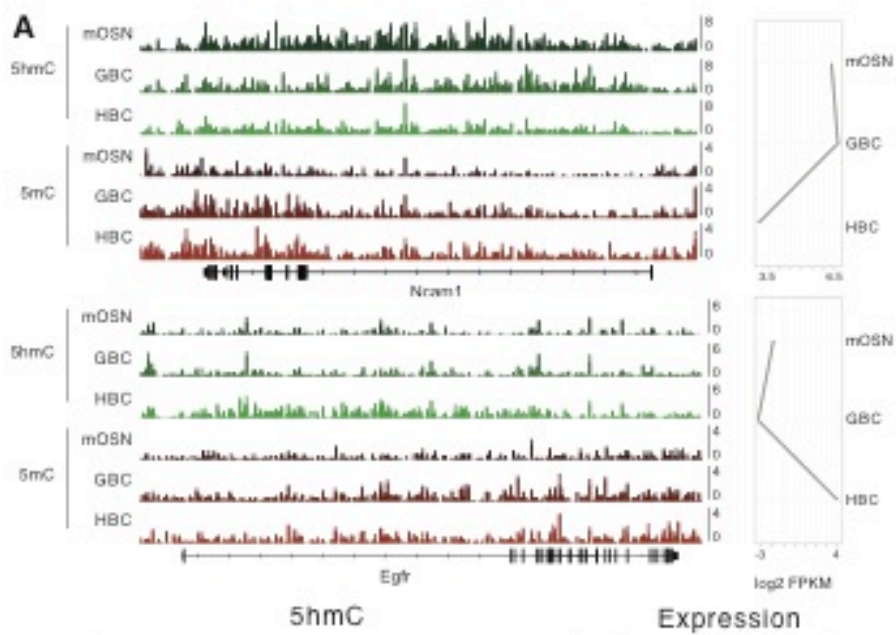


Figure 4. Dynamic patterning of 5hmC at developmentally regulated genes.

(A) 5hmC and 5mC tracks for two representative genes, *Ncam1* and *Egfr*, with different expression paths across mOSN, GBC, and HBC cell types. Scale is reads per million. (B-F) 5hmC TSS profiles in mOSN, GBC, and HBC stages grouped into 5 developmental expression sets: (B) mOSN/GBC common (N = 2577), (C) GBC/HBC common (N = 281), (D) mOSN-specific (N = 893), (E) GBC-specific (N = 557), (F) HBC-specific (N = 1566). Bars indicate a significant difference (see Methods) between the profile of the correspondingly colored track and the HBC profile (B-D), the GBC profile (E), or the mOSN profile (F).

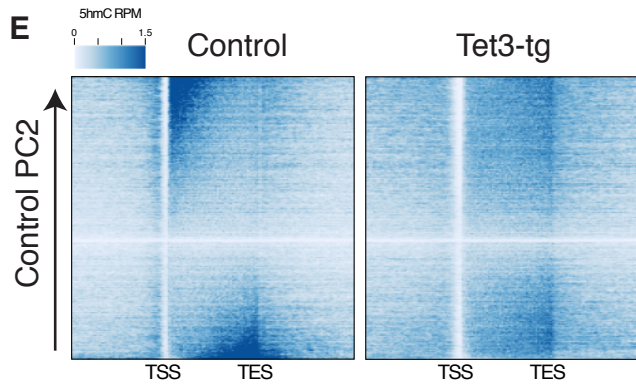
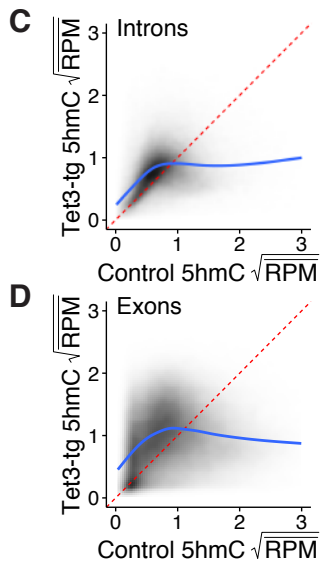
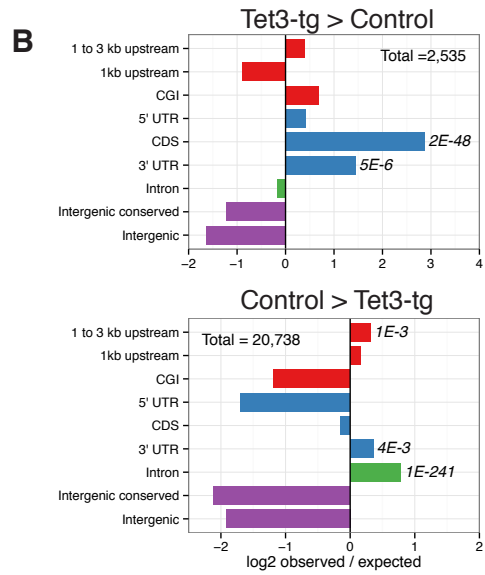
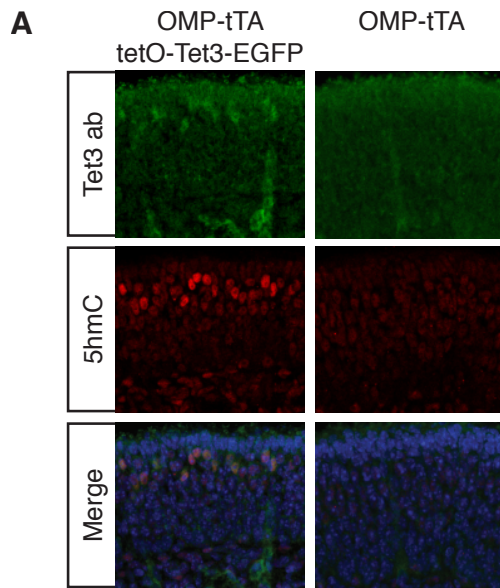


Figure 5. Tet3 overexpression alters 5hmC patterning.

(A) Immunofluorescence of Tet3 and 5hmC on coronal MOE sections of P0 mice either expressing tetO-Tet3-EGFP in mOSNs, *left*, or not, *right*. (B) Differential 5hmC peak analysis and genomic feature intersections. Values are log₂ ratios of observed versus expected number of peaks within a given region. *Top*, regions enriched in Tet3-tg versus control mOSNs. *Bottom*, regions enriched in control versus Tet3-tg mOSNs. Significance values are one-sided Fisher-test false discovery rates. (C, D) Comparison of square-root mean 5hmC RPM in control and Tet3 overexpressing mOSNs within introns (C) and exons (D). Fitted lines are predicted values from a generalized additive model – R command: `gam(y~s(x, bs='cs')`. (E) Heatmaps of genic 5hmC in control, *left*, and Tet3-tg, *right*, mOSNs. Genes are ordered by control 5hmC gene body levels according to principal component 2 (See Methods).

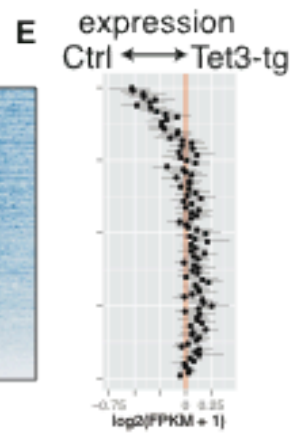
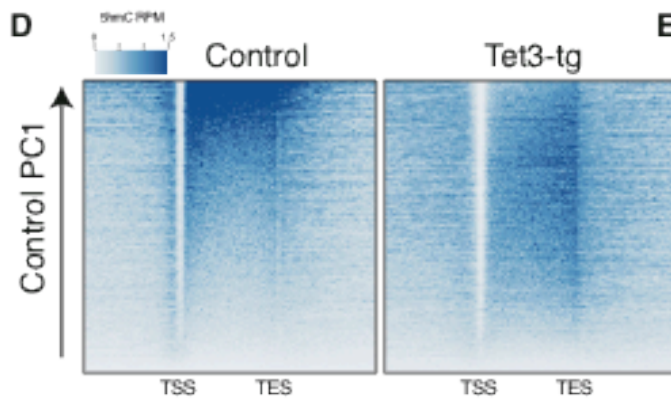
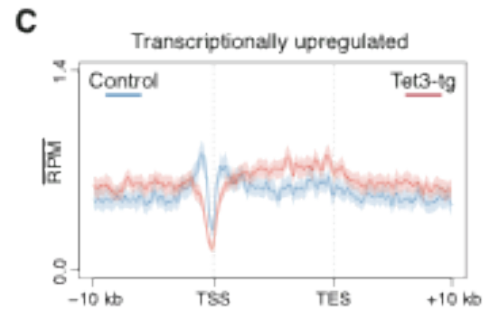
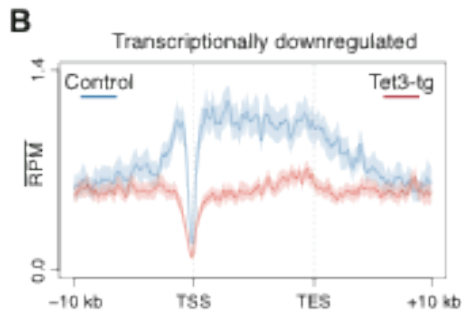
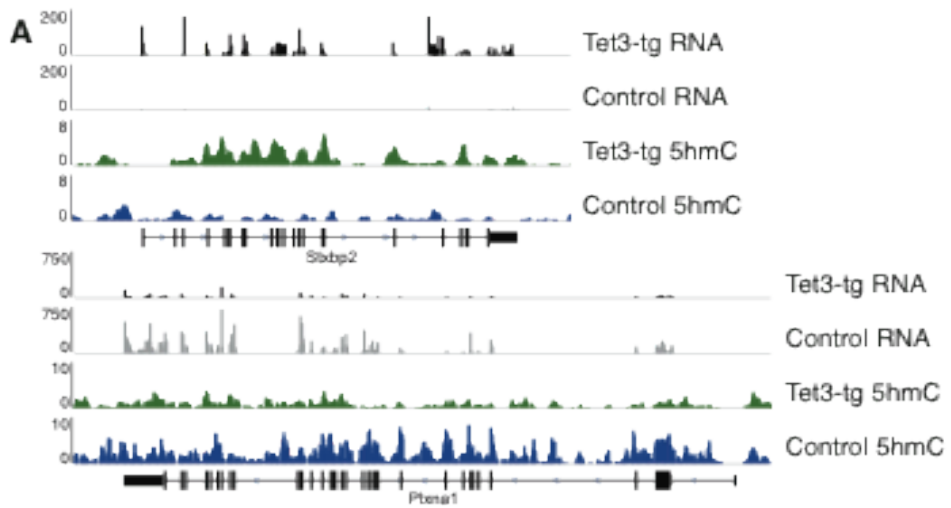


Figure 6. 5hmC alterations modulate gene expression.

(A) rRNA-depleted RNA (intersected with exons) and 5hmC levels in Tet3-tg and control mOSNs for two representative genes. Y-axis is RPM. (B, C) 5hmC profiles over genes transcriptionally downregulated (B) or upregulated (C) in Tet3-tg relative to control mOSNs. (D) Heatmaps of gene body 5hmC in control and Tet3-tg mOSNs ordered by control 5hmC gene body levels according to principal component 1 (PC1, see Methods). (E) Difference between Tet3-tg and control $\log_2(\text{FPKM} + 1)$ values of genes along the ordering used in (D). Values shown are the average of these genes grouped into 100 bins. Error bars are 95% bootstrap confidence intervals.

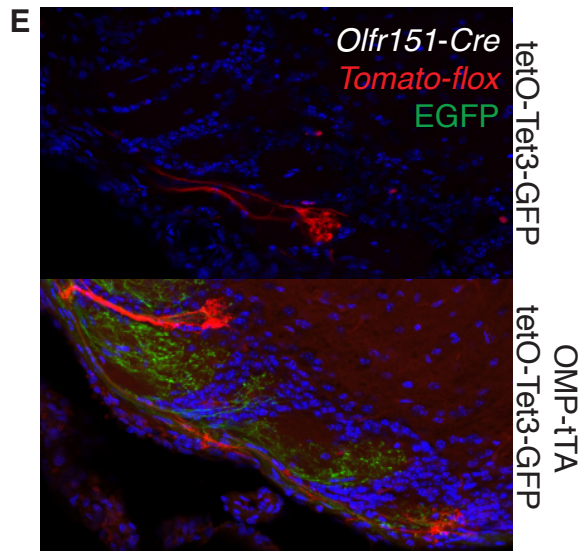
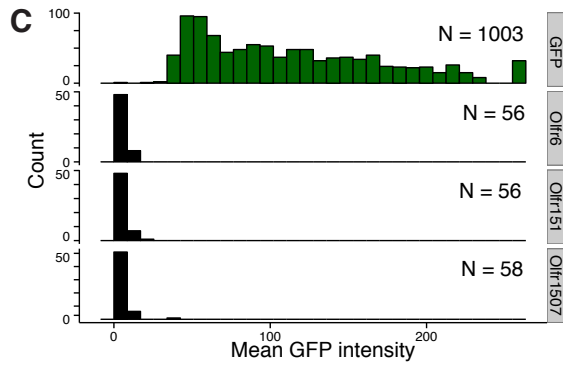
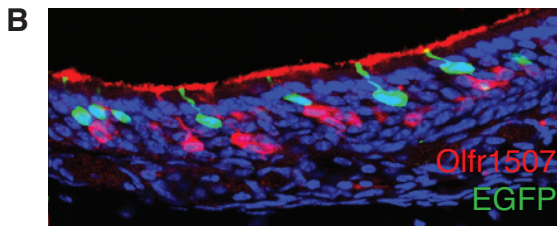
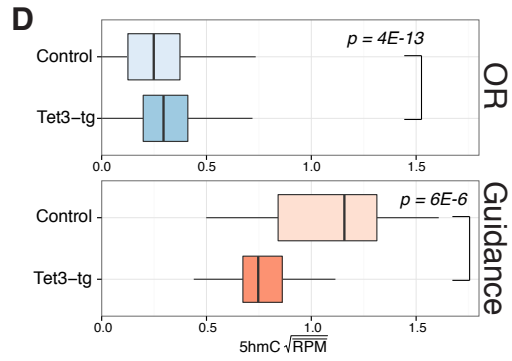
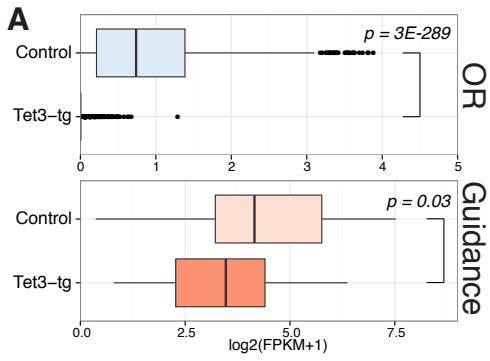
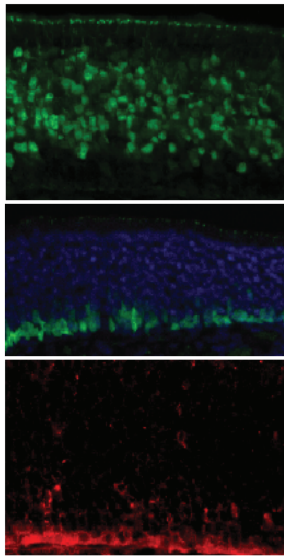


Figure 7. Tet3 overexpression disrupts mOSN axonal targeting.

(A) Boxplots of rRNA-depleted RNA $\log_2(\text{FPKM} + 1)$ values of olfactory receptors (OR) and axon guidance molecules (defined in Methods) in control and Tet3-tg mOSNs. P-values from Wilcoxon two-sided rank-sum test. (B) Immunofluorescent staining of Olfr1507 demonstrating mutually exclusive expression with Tet3-tg. (C) Quantification of GFP intensities within GFP-, Olfr6-, Olfr151-, and Olfr1507-positive OSNs in OMP-tTA / tetO-Tet3-GFP MOEs. (D) Square-root mean 5hmC RPM of ORs and axon guidance molecules with a $\log_2(\text{FPKM} + 1) \geq 2$ in either genotype. P-values from Wilcoxon two-sided rank-sum test. (E) Analysis of OSN axon targeting to the olfactory bulb in *Olfr151-Cre/ Tomato-floxed* animals under control Tet3 expression, *top*, or Tet3 overexpression in mOSNs, *bottom*.

A



B

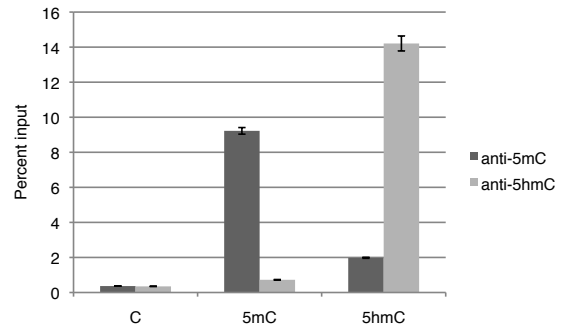


Figure 8. Validation of FACS and DIP.

(A) Coronal sections of adult MOEs from *OMP-ires-GFP*, *Neurog1-GFP*, and C57BL/6 mice stained with anti-ICAM1-PE. (B) DIP-qPCR was performed using a 200 bp linear template containing unmodified cytosines, 5mC, or 5hmC. Error bars are SEM of 3 qPCR replicates.

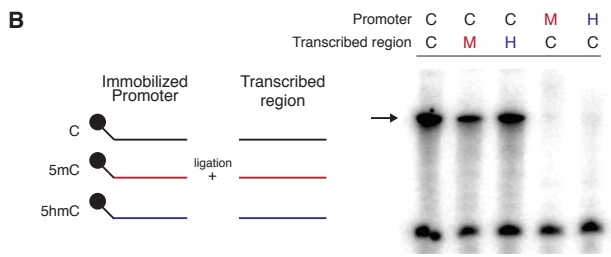
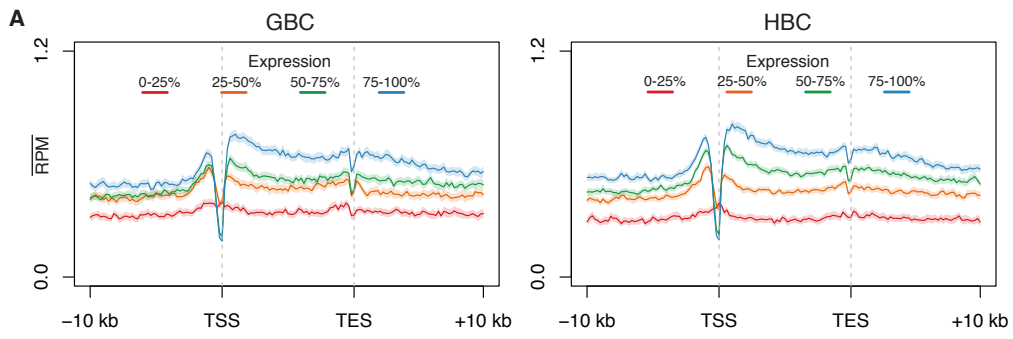


Figure 9. Relationships between gene-body 5hmC and transcription in HBCs, GBCs, and during in vitro transcription.

(A) Aligned gene profiles of GBC and HBC 5hmC levels split by gene expression within each cell type. Error is 95% bootstrap confidence interval. (B) *In vitro* transcription assay testing the effects of 5mC and 5hmC within either the CMV promoter or a 200 bp transcribed fragment. *Left*, schematic of unmodified and modified template construction. *Right*, Digital autoradiograph of transcribed product (arrow) with modified templates. ‘C’ – unmodified; ‘M’ – 5mC; ‘H’ – 5hmC.

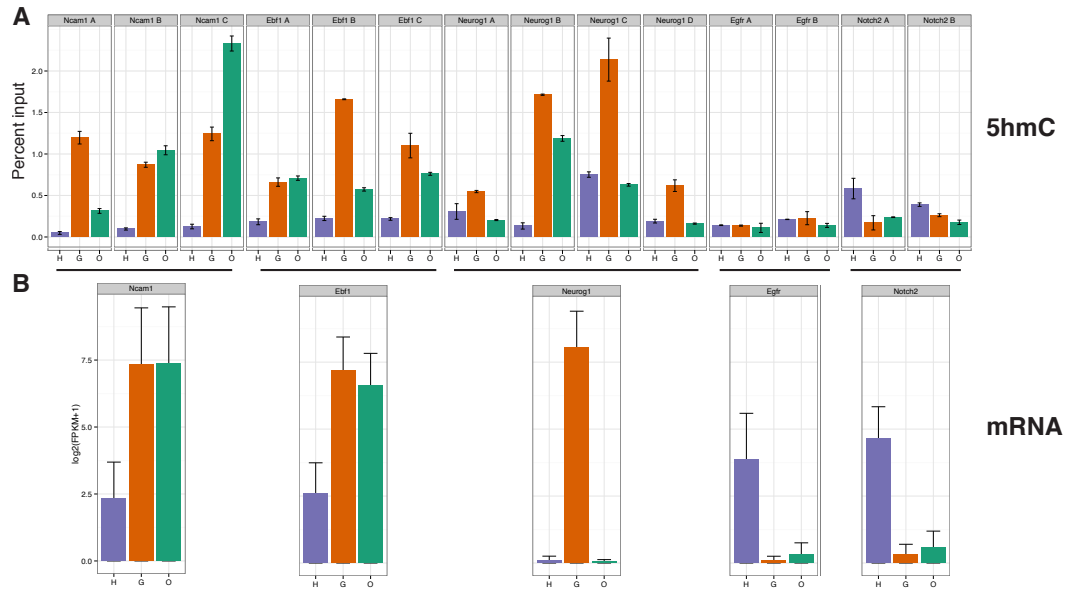


Figure 10. Quantitative PCR confirmation of gene-body 5hmC association with developmental gene expression.

(A) 5hmC and 5mC DIP-qPCR analysis of five differentially expressed genes. Primer sets amplify regions located within the gene body and, in the case of Neurog1 A-C sets, flanking the gene body. Cell types are HBC (H), GBC (G), mOSN (O). Error bars are the range of qPCR duplicates. (B) $\log_2(\text{FPKM} + 1)$ values from sorted cell mRNA-seq datasets. Error bars are Cufflinks generated confidence values. For each, the lower confidence bar is equal to 0 and has been omitted for clarity.

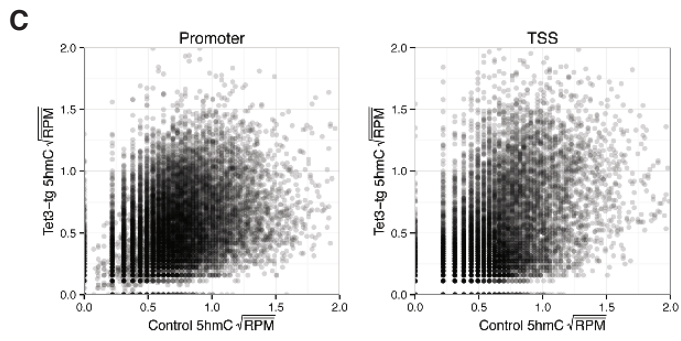
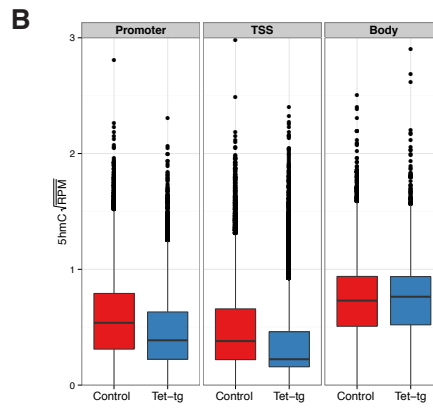
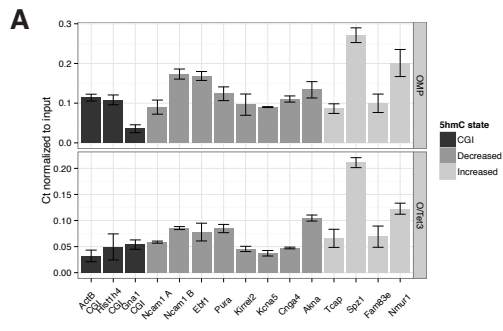


Figure 11. Supplemental analysis of Tet3-tg mOSNs.

(A) DIP-qPCR analysis of 5fC levels in Tet3-tg and control mOSNs. CGI, CpG island. Error bars are the range of qPCR replicates. (B) Distributions of 5hmC levels in control and Tet3-tg mOSNs in promoters (-1kb to TSS), TSSs (-500 bp to +500 bp flanking TSS), and gene bodies (TSS to TES). (C) Control versus Tet3-tg square-root mean RPM values in promoters and TSSs.

Sample	Sequencing type	Number of cycles	Mappable read number
mOSN 5hmC	Single-end	42	21,948,297
GBC 5hmC	Single-end	42	20,460,620
HBC 5hmC	Single-end	42	20,548,870
mOSN 5mC	Single-end	42	17,437,633
GBC 5mC	Single-end	42	22,660,035
HBC 5mC	Single-end	42	20,621,144
mOSN mRNA	Single-end	35	26,211,869
GBC mRNA	Single-end	35	30,373,962
HBC mRNA	Paired-end	75	2,706,246
O/Tet3 5hmC	Paired-end	50	40,063,153
O/Tet3 5mC	Paired-end	50	30,880,849
OMP ribominus RNA	Paired-end	50	50,565,427
O/Tet3 ribominus RNA	Paired-end	50	55,653,440

Table 1. Chapter 2 sequencing metadata.

Chapter 3: Dnmt3a generates the epigenetic substrate for 5-hydroxymethylcytosine in neurons

Authors: Bradley M. Colquitt¹, Eirene Markenscoff-Papadimitriou¹, Stavros Lomvardas^{2*}

Affiliations:

¹Neuroscience Graduate Program, University of California, San Francisco, CA, 94158, USA

²Department of Anatomy, University of California, San Francisco, CA 94158, USA

*Correspondence to: stavros.lomvardas@ucsf.edu

Abstract

During differentiation, neurons experience a reorganization of DNA modification patterns within their genomes. However, the mechanisms governing this developmental patterning are currently unclear. We find that the *de novo* DNA methyltransferase Dnmt3a is necessary for elevated levels of 5-hydroxymethylcytosine (5hmC), a derivative of 5-methylcytosine (5mC), in mature olfactory sensory neurons (mOSNs). A genome-wide analysis of Dnmt3a-dependent 5mC and 5hmC patterning in isolated mOSNs indicates that Dnmt3a is necessary to modify highly accessible regions, including neural enhancers and the transcription start sites of transcribed genes. The loss of Dnmt3a results in the global disruption of gene expression via the activation of silent genes, the reduction of mOSN-expressed transcripts, and the suppression of odorant-induced transcriptional responses. Together, these results demonstrate that Dnmt3a is necessary to establish not only repressive 5mC but also elevated 5hmC associated with active regions of the neuronal genome.

Main Text

The sculpting of DNA modification patterns during cellular differentiation is essential to the construction of cellular identity. Once thought to be largely static in somatic tissues after embryogenesis, the patterning of cytosine modifications is now appreciated to be dynamic in certain cellular and genomic contexts. In particular, 5-hydroxymethylcytosine (5hmC), an oxidized derivative of 5-methylcytosine (5mC), is highly enriched in neurons (Globisch et al., 2010; Kriaucionis and Heintz, 2009a; Ruzov et al., 2011), increases in abundance during neurogenesis (Hahn et al., 2013; Szulwach et al., 2011b), and is localized to gene bodies, regions upstream of transcription start sites (TSSs), and enhancer elements (Colquitt et al., 2013; Hahn et al., 2013; Mellén et al., 2012; Song et al., 2011; Szulwach et al., 2011b). To date, it is unclear how neuron-specific 5hmC patterning is established during the transition from neuronal progenitor to differentiated neuron. Recent studies (Hahn et al., 2013; Szulwach et al., 2011b) noted that the increase of 5hmC during neurodevelopment is not strictly accompanied by a reduction of 5mC, suggesting that *de novo* DNA methylation and cytosine oxidation, mediated by the Tet family of enzymes (Wu and Zhang, 2011), are tightly coupled during this developmental period. Moreover, substantial *de novo* 5mC patterning was recently found in cortical forebrain neurons relative to non-neuronal cell types (Lister et al., 2013).

Dnmt3a, one of two *de novo* DNA methyltransferases in mammals, is expressed in neural precursor cells and neurons during late embryogenesis and in post-mitotic neurons in the postnatal central nervous system (CNS) (Feng et al., 2005). Its ablation specifically within the CNS (Nguyen et al., 2007) recapitulates the early death (3-4 weeks) seen in mice with constitutive loss of the enzyme (Okano et al., 1999). In addition, its loss within post-mitotic neurons is associated with deficits in long-term potentiation, learning and memory (Feng et al.,

2010), and sensitization to cocaine (LaPlant et al., 2010). Recent work identified a role for Dnmt3a in both the repression and facilitation of gene expression in neural stem cells (Wu et al., 2010), suggesting that the enzyme's effects extend beyond establishing repressive 5mC. Here, we tested the hypothesis that Dnmt3a-mediated *de novo* DNA methylation contributes to 5hmC patterning within neurons and is necessary to define neuronal regulatory and transcriptional states. To explore this, we analyzed the relationships between Dnmt3a-dependent 5mC and 5hmC patterning, chromatin accessibility, histone modifications, and transcription within the MOE and isolated mOSNs.

We find that *Dnmt3a* transcription increases along the developmental lineage of olfactory sensory neurons, from horizontal basal cells (HBCs) – the multipotent progenitors of the tissue (Leung et al., 2007) – through globose basal cells (GBCs) – the neuronal progenitors of mOSNs (Caggiano et al., 1994) – to mOSNs (Fig. 12A), in agreement with previous work (MacDonald et al., 2005; Watanabe et al., 2006). Notably, this developmental increase is similar to the enrichment of 5hmC along the mOSN developmental path we described in earlier work (Colquitt et al., 2013). In contrast, *Dnmt3b*, the other mammalian *de novo* DNA methyltransferase, is weakly expressed in HBCs but is not expressed within GBCs or mOSNs. Thus, Dnmt3a is the primary source of *de novo* 5mC within mOSNs. In three-week old Dnmt3a (WT) wildtype mice Dnmt3a protein is most abundant in the immature neuronal stage, between GBCs and mOSNs, but is absent from basal stem and apical non-neuronal layers (Fig. 12B and Fig. 16A). Dnmt3a heterozygous-null (Het) MOEs display intermediate Dnmt3a protein levels, while homozygous-null (KO) MOEs show complete loss of Dnmt3a. Importantly, in Dnmt3a KO MOEs we do not find significant alterations of the expression of the other DNA methyltransferases (except for a modest increase of *Dnmt3b* expression) or of the three Tet members, which convert 5mC to

5hmC (Fig. 16B). In addition, mOSN differentiation is largely unaffected in Dnmt3a KO MOEs, as determined by unaltered percentages of mOSNs and GBCs within the tissue (Fig. 16C), unaffected mitosis rates (Fig. 16D), and mild increases in apoptosis rates (0.3% of total cells in WT to 0.9% in KOs. Fig. 16D and E).

To determine the effects of Dnmt3a loss on global 5mC and 5hmC levels, we immunostained Dnmt3a WT and KO MOEs with antibodies specific for the two modified bases (Fig. 12C). Interestingly, we find no significant difference in 5mC levels along the basal-apical axis ($p = 0.39$, Student's t-test, $N=3$). However, we detect a significant reduction of 5hmC in Dnmt3a KO MOEs within the mOSN layer ($p = 0.03$, Student's t-test, $N=3$), despite initially reaching levels similar to those in WT within more basal layers (Fig. 12C).

These results indicate that 5hmC levels are reduced in the neuronal stage in the absence of Dnmt3a expression. To determine what genomic regions exhibit Dnmt3a-dependent 5mC and 5hmC patterning, we isolated mOSNs via fluorescence-activated cell sorting (FACS) from the MOEs of three-week old Dnmt3a WT, Het, and KO mice and performed DNA immunoprecipitation against 5mC and 5hmC coupled to massively parallel sequencing (DIP-seq, Table 2). Genome-wide, 5mC and 5hmC levels are highly correlated among Dnmt3a genotypes; however, regions with high levels of 5hmC in WT mOSNs exhibit reduced levels in KO mOSNs (Fig. 17A and B).

A recent study described a positive relationship between chromatin accessibility and neuronal DNA methylation in the CH dinucleotide context, hypothesized to be established by Dnmt3a (Lister et al., 2013). To directly test this association, we compared 5mC and 5hmC levels in Dnmt3a WT and KO mOSNs with DNase I accessibility in Dnmt3a Het MOEs. In agreement

with Dnmt3a activity occurring in accessible regions, we find positive correlations between chromatin accessibility and Dnmt3a-dependent 5mC and 5hmC patterning ($R=0.10$ and $R=0.17$, Fig. 12D and S2C). Moreover, in the absence of Dnmt3a, regions flanking local DNase I hypersensitive sites in the MOE exhibit depletions of 5hmC, and to a lesser extent 5mC (Fig. 12E). Notably, 5hmC levels in Dnmt3a Het mOSNs are present at intermediate levels flanking these sites. In a previous study (Colquitt et al., 2013), we overexpressed Tet3 in mOSNs (Tet3-tg) and found that regions with high levels of 5hmC in WT mOSNs exhibited depletions of the mark in Tet3-tg mOSNs, consistent with the role of Tet3 in progressive oxidation of modified cytosines (Wu and Zhang, 2011). From this data, we identified a positive correlation between Tet3-tg/wildtype differential 5hmC and Dnmt3a-dependent 5hmC (Pearson $R = 0.18$, Fig. 12D and S2D), suggesting that Dnmt3a-mediated *de novo* 5mC is closely coupled to neuronal Tet-mediated oxidation.

We used two methods to identify genomic regions that contain Dnmt3a-dependent 5mC and 5hmC – peak analysis using a small peak size (500 basepairs, bp) and differential enrichment analysis in broader windows (5-50 kilobases, kb). Dnmt3a-dependent 5mC ($N=955$) and 5hmC ($N=5767$) 500 bp regions are significantly associated with different genomic regions: although both are found within introns, Dnmt3a-mediated 5mC peaks are localized to CpG islands (CGIs), CDSs, and 3' UTRs while Dnmt3a-mediated 5hmC peaks are associated with regions upstream of TSSs (Fig. 17E). Despite these differences, each peak set tends to be found near genes implicated in development, e.g. somatic stem cell gene regulation and sensory organ morphogenesis (Table S2). Likewise, broad Dnmt3a-dependent 5mC ($N=1162$) and 5hmC ($N=1805$) regions identified using the second method are strongly associated with genes (2025 of 2549, $p = 4E-58$, Fisher's exact-test).

Extensive remodeling of histone modifications occurs along the mOSN developmental lineage (Magklara et al., 2011) and may interact with *de novo* cytosine modification. To integrate Dnmt3a-dependent DNA modification patterning into the broader epigenetic landscape of the MOE, we performed CHIP-seq against H3K4me1 (a mark associated with genes and enhancer elements), H3K27ac (TSSs and active enhancer elements), and H3K27me3 (repressed loci) using native chromatin preparations isolated from the MOEs of three-week old Dnmt3a WT mice. We hierarchically clustered the 5-50kb set of regions with Dnmt3a-dependent 5mC or 5hmC by their average H3K27ac and H3K27me3 levels, producing three clusters (enriched for H3K27ac only, H3K27me3 only, or both modifications, Fig. 13A). This grouping by histone modifications resulted in a partial segregation of 5mC and 5hmC levels (Fig. 13B): Regions enriched for H3K27ac (cluster 1, example in Fig. 13C, N=1219) are enriched for 5hmC while those enriched for H3K27me3 (cluster 2, example in Fig. 13C, N=317) are correspondingly enriched for 5mC. Interestingly, a substantial fraction of these regions (N=913) are enriched for both modifications and possess high levels of both DNA modifications (cluster 3, examples in Fig. 13D and Figs. S3A-D). Each cluster is associated with genes expressed in neural tissues but with different functions, e.g. cluster 1 with signal transduction and both 2 and 3 with transcriptional regulation (Table S3). Notably, the set of genes within the doubly modified H3K27ac/H3K27me3 cluster is enriched for genes implicated in neural development including several transcription factors -- *Neurog1*, *Dlx5*, *Lhx2*, and *Emx2* -- known to function during OSN neurogenesis (Nicolay et al., 2006). These regions also contain elevated DNase I accessibility that precisely overlaps with the area of H3K27ac, H3K27me3, and Dnmt3a-dependent 5mC and 5hmC enrichment (Fig. 13E and Figs. S3A-D). This dual presence of histone modifications associated with activation and

repression within single loci may be due to developmental heterogeneity within the tissue or may reflect differential monoallelic epigenetic states.

In addition to gene elements, Dnmt3a-dependent 5hmC is significantly associated with potential MOE enhancers (intergenic H3K4me1 peaks, Fig. 17E). We find positive correlations at these regions between flanking 5hmC levels and both H3K27ac levels ($R=0.33$) and DNase hypersensitivity ($R=0.22$, Fig. 14A) while only weak correlations between these features and flanking 5mC levels ($R=-0.04$ and 0.07 , respectively). To further investigate the relationship between enhancer activity and 5hmC levels, we compared enhancer 5hmC levels and the expression of nearby genes along the mOSN developmental lineage. We clustered the 5000 potential enhancers with the highest 5hmC variance across HBCs, GBCs, and mOSNs by their 5hmC levels, producing 8 clusters each representing a developmental 5hmC path, e.g. mOSN-specific or GBC-specific (Fig. 14B) and computed the median expression levels of nearby genes. In support of an active role for 5hmC in enhancer function, we find a striking correlation between developmental enhancer 5hmC patterning and nearby gene expression.

Enhancers have been previously functionally classified as ‘poised’ or ‘active’ using H3K4me1 and H3K27ac levels (Creyghton et al., 2010); poised enhancers are enriched for only H3K4me1, while active enhancers display enrichment of both modifications. To generate a similar classification for MOE enhancers, H3K27ac peaks were intersected with intergenic H3K4me1 peaks, yielding H3K27ac-negative H3K4me1 peaks (‘poised’) and H3K27ac-positive peaks (‘active’). We find that 5mC levels in Dnmt3a KO mOSNs are reduced relative to WT flanking enhancers regardless of classification (Fig. 14C). Similarly, 5hmC is depleted from the flanking regions of both poised and active enhancers in Dnmt3a KOs, although the depletion is strongest

at regions flanking active enhancers. As with DNase hypersensitive sites in general, 5hmC in Dnmt3a Het mOSNs is present at intermediate levels flanking both poised and active enhancers.

The reduction of 5hmC levels at active enhancers in the absence of Dnmt3a may indicate that the enzyme modifies all enhancers regardless of activity or only those that are active within neural tissues. To resolve this, we computed Dnmt3a KO versus WT 5hmC ratios at VISTA enhancers (Visel et al., 2007) and found that those with Dnmt3a-dependent 5hmC patterning are strongly associated with activity in the nose and other neural or face tissues (forebrain, midbrain, hindbrain, neural tube, and facial mesenchyme, Fig. 14D). Correspondingly, VISTA enhancers classified as active in these tissues have depleted 5hmC in Dnmt3a KO mOSNs, while those with activities in non-associated tissues retain 5hmC patterning in the absence of Dnmt3a (Fig. 14E). Moreover, VISTA enhancers with at least two-fold reduction of 5hmC are significantly associated with homeodomain transcription factors implicated in neurodevelopment (Table S4). Thus, Dnmt3a appears to provide the 5mC substrate necessary to produce elevated levels of 5hmC flanking enhancers and, when assaying a neuronal cell type such as mOSNs, does so at enhancers engaged in the regulation of neuronal development.

The gain of Dnmt3a-mediated 5hmC within neurons suggests that this patterning plays a role in regulating neuronal gene expression programs. To test this, we prepared RNA-seq libraries from FACS-isolated Dnmt3a WT, Het, and KO mOSNs (Tables S1 and S5). Overall, reads per kilobase per million reads (RPKM) values from each genotype are highly correlated (Fig. 19A). To assess global gene expression, we averaged RPKM values across genotypes and replicates to obtain a transcriptional index. By plotting expression fold change against this index, we find a striking bidirectional change of expression between WT and KO mOSNs that is dependent on transcriptional output (Fig. 15A). On average, lowly expressed or silent genes (as determined by

Gaussian-based clustering of RPKMs) are transcriptionally upregulated (mean $\log_2(\text{KO}/\text{WT RPKM}) \pm \text{SD} = 0.9 \pm 1.1$) while highly expressed genes are weakly downregulated (-0.3 ± 0.4). Moreover, genes with an absolute KO-WT fold change of at least 2-fold exhibit an intermediate disruption of expression levels in Het mOSNs (Fig. 19B). To assess how Dnmt3a-sensitive genes are regulated during mOSN development, we clustered genes with an absolute $\log_2(\text{KO} / \text{WT RPKM}) > 1$ by their expression in Dnmt3a wildtype mOSNs, GBCs, and HBCs, using previously published datasets (Fig. 19C-D)(Colquitt et al., 2013; Magklara et al., 2011). We find that upregulated genes (N=1845) are typically expressed in HBCs, GBCs, or not at all in the MOE (Fig. 19C) while downregulated genes (N=387) are expressed in mOSNs (Fig. 19D).

Transcription start sites are highly accessible (Fig. 19E) and may, like enhancers, be subject to Dnmt3a-mediated patterning. Across all genes, 5mC levels are modestly reduced flanking TSSs in Dnmt3a KOs while 5hmC levels flanking TSSs in both Hets and KOs are reduced (Fig. 15B). Moreover, 5hmC in Dnmt3a KO mOSNs is present at levels similar to those found in Dnmt3a WT GBCs (data from (Colquitt et al., 2013)). Ordering these TSS-flanking levels by the fold change of KO versus WT expression revealed complementary depletions of 5mC and 5hmC depending on expression fold change: upregulated genes exhibit TSS-flanking depletions of 5mC in Dnmt3a KOs, while unaffected and downregulated genes exhibit TSS-flanking depletions of 5hmC in both Dnmt3a Hets and KOs (Fig. 15D). This dual response suggests that Dnmt3a is necessary to establish both repressive, 5mC-enriched and activating, 5hmC-enriched regulatory states. Interestingly, whether Dnmt3a acts as a repressor or facilitator of transcription correlates with the strength (observed/expected number of CpGs) of associated CpG islands (CGIs): genes that are upregulated in Dnmt3a KO mOSNs tend to have weaker CGIs while those that are downregulated tend to have stronger CGIs (Fig. 19F).

A previous study observed increased levels of the repressive histone modification H3K27me3 at the TSSs of genes downregulated in neural stem cells in the absence of Dnmt3a (Wu et al., 2010). To determine if a similar increase occurs within the olfactory epithelium, we performed H3K27me3 ChIP-seq using Dnmt3a KO MOEs. In contrast to the previous report, we do not observe an increase of H3K27me3 levels at genes with reduced expression in KO mOSNs and instead detect a modest depletion of H3K27me3 at the TSSs of genes with increased expression (Figs. 4C and D). This distinction may reflect key differences between the developmental stages of the two populations: in progenitor cells, Dnmt3a-mediated methylation may be necessary to antagonize the spread of Polycomb activity, while in differentiated neurons such as mOSNs, Dnmt3a-mediated methylation may positively influence gene expression largely as a substrate for neuronal-specific oxidation to 5hmC.

These transcriptional defects led us to investigate transcriptional responses to odorant-induced neuronal activation in Dnmt3a-deficient animals. We exposed mice to a concentrated mixture of previously characterized odorants (1:1:1 amyl acetate:acetophenone:octanal, am:ac:oc) (Mombaerts, 2004; Rubin and Katz, 1999) and assayed transcriptional levels in their MOEs relative to those of mice exposed to water. We detected maximal *Fos*, *Egr1*, and *Nr4a1* induction after 1 hr exposure (Fig. 19A); after this time point, expression returned to unexposed levels. To globally assay odorant-induced transcription, we prepared RNA-seq libraries from the MOEs of Dnmt3a WT, Het, and KO mice that had been exposed to 1 hour of am:ac:oc or corresponding unexposed controls (Tables S1 and S6). Across genotypes, we identified 25 genes that were significantly upregulated and one gene that was downregulated (FDR<0.05) in response to odorant exposure (Table 3). Among these genes was the transcription factor *Nr4a2*, which strikingly exhibits both significant depletions of 5mC and 5hmC levels in its gene body and the

reduction of odorant-induced expression in Dnmt3a KOs (Fig. 15E). In contrast, the 5mC and 5hmC levels and odorant-induced expression of another gene *Pcdh10* are unaffected by the loss of Dnmt3a. This association between reduced 5mC, 5hmC, and reduced gene activation generalizes across odorant-induced genes (Fig. 15F). On average, the odorant-induced transcriptional responses of these 25 genes are significantly reduced in Dnmt3a Hets ($p = 4E-4$, paired two-sided Wilcoxon) and KOs ($p = 5E-5$, Fig. 15G and S5B), while basal levels remain unchanged (Fig. 20B). Similarly, the gene bodies of these odorant-induced genes are significantly depleted of both 5mC ($p=5E-4$) and 5hmC ($p=3E-4$) in Dnmt3a KO relative to WT mOSNs (Fig. 20C).

Our results indicate that Dnmt3a is necessary to establish the full 5mC and 5hmC landscapes of olfactory sensory neurons and, in turn, to fully execute neuronal gene expression. In neurons, 5hmC occurs nearly exclusively on cytosines in CG dinucleotides while *de novo* 5mC occurs broadly in multiple contexts (Lister et al., 2013). This suggests that Dnmt3a, which has been shown to provide both CG and CH methylation in non-neural systems (Meissner et al., 2005; Ramsahoye et al., 2000), generates a 5mC platform on which neuronal Tet activity can provide elevated levels of 5hmC. Although the exact mechanisms remain to be determined, the titration of this balance between 5mC and 5hmC may modulate transcriptional output. Moreover, our finding that odorant-induced transcriptional responses require Dnmt3a, combined with the observation that substantial *de novo* 5mC occurs in neurons around the time of synaptogenesis (Lister et al., 2013), suggests that neuronal DNA modification patterns are essential for activity-sensitive transcriptional programs and may play a crucial role in regulating experience-dependent plasticity.

Acknowledgments:

Sequencing data is deposited in GEO accession GSE52464. This work was supported by the NIH (DP2 OD006667, R01MH091661), NSF (Graduate Research Fellowship), and the Rett Syndrome Trust. BC designed the research project, performed most experiments, analyzed the data, and wrote the manuscript. EMP performed several ChIP-seq assays. SL designed the research project and wrote the manuscript.

Methods

Animal care and use. Mice were treated in compliance with the rules and regulations of IACUC under protocol approval number AN099395-01.

Dnmt3a-null mice. Dnmt3a constitutive null mice (Okano et al., 1999) were obtained from MGI (2182412).

Immunofluorescence. Sections of postnatal C57BL/6 MOE were fixed with 3:1 methanol:acetic acid, fragmented with 1N HCl, and neutralized with 100 mM Tris pH 8. Sections were blocked then incubated with anti-5mC (#39649) and anti-5hmC (#39791) antibodies (Active Motif) overnight at 4°C. Immunostaining for proteins was performed on coronal MOE sections from 3-week old animals fixed with 4% PFA. Antibodies: anti-Dnmt3a (Santa Cruz, sc-10231).

Immunofluorescence quantification. Basal-apical profiles were computed using a custom ImageJ macro. Selection boxes were sized around images of straight sections of OE with the basal-apical axis oriented left-right. Width of the selection box was sized to span the full OE as determined by DAPI staining and height was set to 20 μm . For each OE section, the bounding box was stepped down the OE to compute non-overlapping immunofluorescence profiles spanning at least 160 μm . This process was repeated with three different OE sections per biological replicate. These profiles were then average within each replicate.

Fluorescence activated cell sorting. Mature OSNs were isolated from the MOEs of OMP-ires-GFP knockin animals (Shykind et al., 2004). GBCs were isolated from the MOEs of Ngn-GFP BAC transgenics (Gong et al., 2003). HBCs were isolated with anti-mouse CD54 (*Icam1*) conjugated to PE (Biolegend). Isolation and preparation were performed as described in (Colquitt et al., 2013; Magklara et al., 2011). Briefly, single cell suspensions from 1-10 MOEs

were made using the Papain dissociation system (Worthington). MOE were dissected into Earl's Buffered Saline Solution (EBSS), minced in papain dissociation solution (Worthington), and incubated at 37°C for 45 minutes. Cells were washed once in 1:10 inhibitor solution and once in PBS. For a given assay (DIP-seq or RNA-seq), 2E5 to 1E6 cells were used. Population purity was assayed by resorting a small fraction of the sorted cells. Purities of ~90-95% were routinely achieved.

DNA immunoprecipitation (DIP)-seq. Genomic DNA was isolated from the cell populations described in the main text, prepared for paired-end Illumina sequencing using standard protocols, and immunoprecipitated as in Weber et al. (Weber et al., 2005) with minor modifications. Specifically, sorted cell gDNA was sonicated to 200-1000 bp with a mean size of ~400 bp and end repaired (NEBNext End Repair). Adenosine was added to 3' ends (Klenow, 3' to 5' exo-, NEB), and 10:1 Illumina adaptor:samples were ligated onto the samples. 20-100 ng of adapted genomic DNA was diluted in MB (10 mM phosphate buffer pH 7.0, 140 mM NaCl, 0.05% Triton X-100), denatured at 95°C for 10 minutes, brought to 300 µL MB, and precipitated with 0.5 µg anti-5mC or 5hmC antibodies (as used in immunofluorescence) overnight at 4°C. Separately, 10 µL per IP of Protein A and G Dynabeads each (Invitrogen) were blocked with 2 mg BSA and 2 mg yeast tRNA in 1 mL MB overnight at 4°C. Beads and IPs were combined the next day and rotated for 3 hours at 4°C. Samples were washed 4 times in MB, eluted twice in 100 µL elution buffer (0.1 M NaHCO₃, 1% SDS, 1 mM EDTA) by shaking at 37°C for 15 minutes, phenol:chloroform:isoamyl extracted, ethanol precipitated, and resuspended in 50 µL TE. Samples were then amplified with Illumina TruSeq PCR primers for 15 cycles using Phusion DNA polymerase (NEB).

RT-qPCR. MOE total RNA was isolated by TRIzol extraction (Invitrogen) and treated with TURBO DNase (Invitrogen). 1 µg of this RNA was used for first strand cDNA synthesis using Superscript III First-Strand Synthesis System (Invitrogen). qPCR was performed using Maxima SYBR Green (Fermentas).

RNA-seq. Total RNA was isolated from a given cell population or tissue using TRIzol (Invitrogen), the aqueous phase of which was passed through a Zymo RNA Clean and Concentrator 5 column, DNase-treated on column using 1 unit of TURBO DNase for 15 minutes at 37°C, and eluted twice using 10 µL water. rRNA-depleted sequencing libraries were prepared by either depleting 200 ng of total RNA via two rounds of RiboMinus (Invitrogen) followed by library construction using ScriptSeq v2 or direct application of 100 ng of total RNA to the Nugen Encore Complete RNA-seq system. Samples were amplified for 15 cycles using Phusion DNA polymerase (NEB) and provided ScriptSeq or Nugen primers.

DNase accessibility-seq. Nuclei from the olfactory epithelium were isolated as described in Magklara et al. and resuspended in nuclease digestion buffer (0.32 M sucrose, 50 mM Tris-HCl pH 7.5, 4 mM MgCl₂, 1 mM CaCl₂, 0.1 mM PMSF). 3 million nuclei were brought to 250 µL nuclease digestion buffer, pre-warmed for one minute at 37°C, and incubated with 20 U DNase I (Ambion) for 2 or 5 minutes. Reactions were stopped with the addition of 28 µL 0.5 M EDTA and placement on ice. Reactions were then incubated with 200 µg proteinase K (Ambion) 55°C overnight, extracted using phenol:chloroform:isoamyl (Invitrogen), and ethanol precipitated. Samples were resuspended in TE, incubated with 10 µg RNase A (Roche) at 37°C for 30 minutes, extracted using phenol:chloroform:isoamyl (Invitrogen), and ethanol precipitated. Samples were resuspended in TE, and one half of each reaction was run on a 1% agarose/1x TAE gel. Regions 100-500 bp were excised and purified using Qiagen Gel Extraction Kit.

Illumina sequencing libraries were then prepared as described in meDIP and hmeDIP-seq section and amplified for 9 cycles using Phusion DNA polymerase (NEB) and Illumina TruSeq primers. Sequencing data from both time points was combined due to high correlation.

Chromatin immunoprecipitation (ChIP)-seq. Nuclei were isolated from the MOEs of 3-week old Dnmt3a wildtype and knockout mice and native ChIP was performed as described (Magklara et al., 2011) using anti-H3K27ac (Millipore, cma309), anti-H3K4me1 (Abcam, ab8895), and anti-H3K27me3 (Millipore, 07-449) antibodies. For library preparation, ChIP DNA was sonicated for 120-180 seconds on a Covaris S220 and prepared for sequencing using the Ovation Ultralow Library Kit (Nugen).

Odorant exposure. Mice were divided into two clean cages containing no bedding and placed into separate dark cabinets. After 10 minutes in these cages, a 4 cm x 4 cm piece of Whatman paper with either 80 μ L of water or a 1:1:1 mixture of pure amyl acetate:acetophenone:octanal (all purchased from Sigma-Aldrich) was placed at one end of the cage.

For odorant exposure time course experiments, Dnmt3a wildtype mice were kept in these conditions for 15, 30, 60, 120 minutes or exposed for 60 minutes to either water or odorant then transferred to a new clean cage containing no odorant for an additional 60 minutes. At the appropriate time, mice were euthanized using CO₂. Their MOEs were dissected into RNAlater (Ambion) on ice, transferred to RLT plus buffer (Qiagen), and homogenized. RNA was isolated using the RNeasy Plus Mini Kit (Qiagen). cDNA was generated by Superscript III First-Strand Synthesis System (Invitrogen).

For RNA-seq experiments, Dnmt3a wildtype, heterozygous-null, homozygous-null mice were treated as above for 60 minutes and euthanized using CO₂. Their MOEs were dissected into

TRIzol (Invitrogen) and homogenized. RNA was isolated using the standard TRIzol protocol and treated with TURBO DNase (Invitrogen). RNA-seq libraries were then prepared using 100 ng of total RNA and the Encore Complete RNA-seq Library Kit (Nugen).

Sequencing alignment and initial processing. DIP-seq, DNase-seq, and ChIP-seq read mapping was performed using Bowtie2 (Ben Langmead and Salzberg, 2012) (default parameters, mm9). RNA-seq mapping was performed using STAR (Dobin et al., 2013), mapping to the Illumina iGenomes Ensembl NCBIM37 GTF and mm9 using default parameters. Reads with MAPQ < 30 (all samples) and PCR duplicates (DIP- and ChIP-seq only) were removed. Tracks were generated consisting of read counts binned into 25 bp normalized by the number of millions of reads and scaled to reflect the average number of reads per 1kb to obtain reads per kilobase per million reads (RPKM, i.e. $1E6 * 1E3 / (\text{window size} * \text{total number of reads})$). Biological replicates were averaged for visualization and subsequent analysis. For visualization, tracks were smoothed using a 125 bp moving average window.

Genome-wide correlations. Read data was binned into specified window sizes across each chromosome, bins with zero reads were removed, and Pearson correlation values were computed across each chromosome. These chromosomal correlations values were averaged to obtain a genome-wide correlation value.

Differential peak analysis. Regions with differential enrichment of 5mC and 5hmC were identified using HOMER (Heinz et al., 2010). Tag directories (GC-normalized) for each sample and replicate were created independently from processed BAM files. Only read 1 was used for paired-end samples. Differential regions were called in Dnmt3a WT samples using Dnmt3a KO as input, 'histone' presets, peak size of 500 bp, and at least 2 fold enrichment over input

(findPeaks -style histone -size 500 -F 2). Differential peak sets from each replicate were intersected (requiring an overlap of at least 50% of peak size) to create common peak sets for feature intersection analysis. Genomic positions for gene elements (*1 to 3 kb upstream, 1 kb upstream, CGI, CDS, 3' UTR, 5' UTR*) were obtained from the mm9 UCSC refGene table. Intergenic sites (*intergenic*) are defined as the complement of refGene regions extended by 5kb. Conserved intergenic regions (*intergenic conserved*) were obtained by intersecting this intergenic set with the vertebrate conserved elements (phastConsElements30way). For differential enrichment analysis within broader regions, the genome was first divided into non-overlapping 5000 bp window and the number of reads within each window for each sample was computed. P-values of differential enrichment were computed from this count matrix using the R package *DESeq2* (Anders and Huber, 2010). Regions with FDR-adjusted p-value of less than 0.05 were classified as significant. If these regions were within 20 kilobases of one another, they were joined into one region.

Gene ontology analysis. Ontologies statistically associated with regions of interest were determined using GREAT (McLean et al., 2010) with the ‘basal plus extension’ setting.

Aligned feature profiles. TSS profiles were generated using 25 bp windows extending from the 5' most annotated TSS for each gene. For each profile, the mean of a given aligned position was computed (excluding values at extreme 1%) and 95% confidence intervals were calculated through bootstrap resampling for 1000 iterations using the R package *boot*. Profiles flanking intergenic H3K4me1 peaks were similarly generated.

Genomic feature heatmaps. 200 25 bp windows we generated starting from the midpoint of a given genomic region (for example H3K4me1 peaks). For a given set of regions, these windows

were arranged into a matrix with regions for rows and positions flanking region as columns. In the case of H3K4me1 peaks, ordering of these rows was achieved by performing principal component analysis (R command *prcomp*) on H3K27ac values and ordering by PC1. Ordered matrices were then averaged by every 100 rows and plotted as heatmaps.

RNA-seq analysis. Reads per kilobase per millions of reads (RPKM) values for each gene were computed by first counting the number of reads per gene using *featureCounts* (Liao et al., 2013) and the Illumina iGenomes Ensembl NCBI37 GTF. These counts were then normalized across samples using the *estimateSizeFactors* function in the R package *DESeq* (Anders and Huber, 2010) and by gene length in kilobases. For subsequent analysis, including global fold-change analysis and clustering, sample RPKMs were obtained by averaging across biological replicates. Differential expression of odorant-induced genes was performed using the *edgeR* package (Robinson et al., 2010). Model-based determination of expressed versus unexpressed genes was performed using averaged RPKM values and the *Mclust* function (G=2) from the R package *mclust* (Fraley and Raftery, 2002). Genes with an uncertainty greater than 0.1 were excluded from subsequent analysis.

Figures

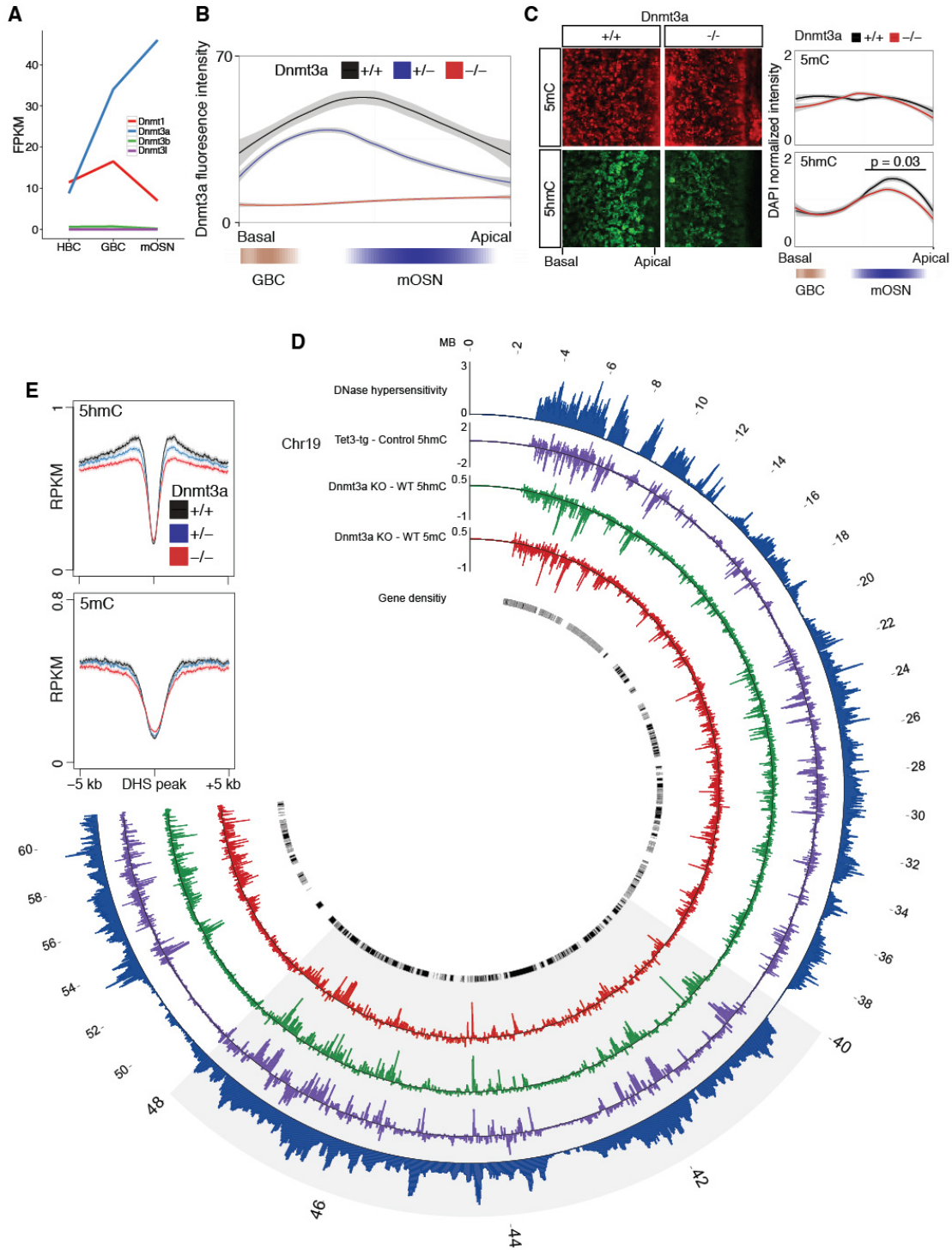


Figure 12. Dnmt3a is necessary for neuronal 5hmC patterning within accessible chromatin.

(A) RNA-seq fragments per kilobase per million (FPKM) of *Dnmt* family members in horizontal basal cells (HBCs, multipotent stem cells), globose basal cells (GBCs, neuronal progenitors), and mature olfactory sensory neurons (mOSNs, primary sensory neurons). Data from (Colquitt et al., 2013). (B) Basal-apical quantification of Dnmt3a immunoreactivity in *Dnmt3a* *+/+*, *+/-*, and *-/-* MOEs. Lines are average intensity across three biological replicates, error is 95% confidence interval. Positions of GBCs and mOSNs determined using Neurog1-EGFP and OMP-EGFP MOEs, respectively. (C) *Left*, Immunofluorescence of 5mC and 5hmC in *Dnmt3a* *+/+* and *-/-* MOE. *Right*, Quantification of 5mC and 5hmC intensity along basal-apical axis. Lines are average intensity across three biological replicates, error is 95% confidence interval. (D) Chromatin accessibility and Dnmt3a-dependent DNA modification patterning across chromosome 19. *Outside row*, MOE Dnmt3a Het DNase accessibility. *Second row from outside*, Difference between 5hmC levels in Tet3-tg mOSNs (OMP-tTA x tetO-Tet3-GFP) and control mOSNs (OMP-GFP). Data from (Colquitt et al., 2013). Dnmt3a KO minus WT 5hmC (*third row from outside*) and 5mC (*fourth row from outside*) levels in mOSNs. 5mC and 5hmC values are averages across biological replicates. Y-axes in RPKM. (E) Average 5hmC and 5mC levels spanning localized MOE Dnmt3a Het DNase hypersensitive sites (N=12,419). Error is bootstrapped 95% confidence intervals.

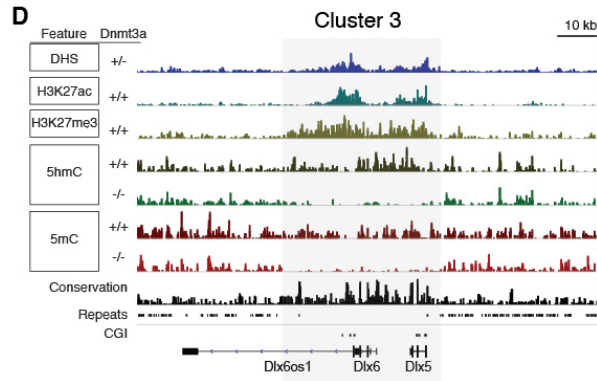
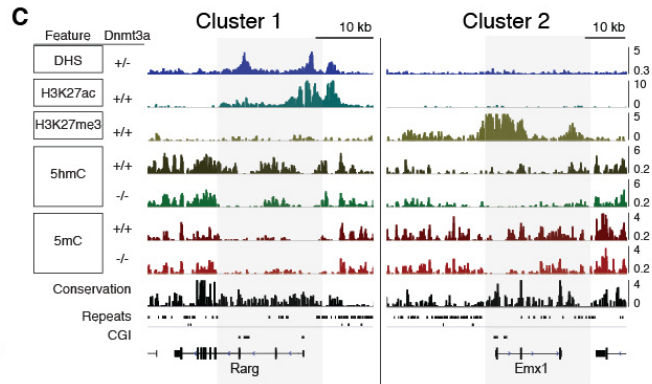
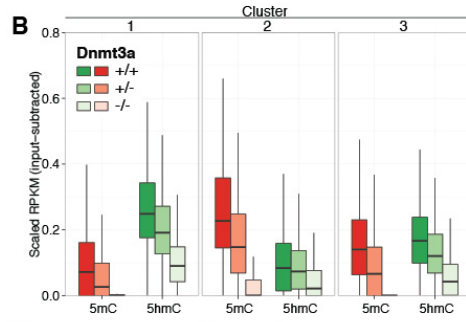
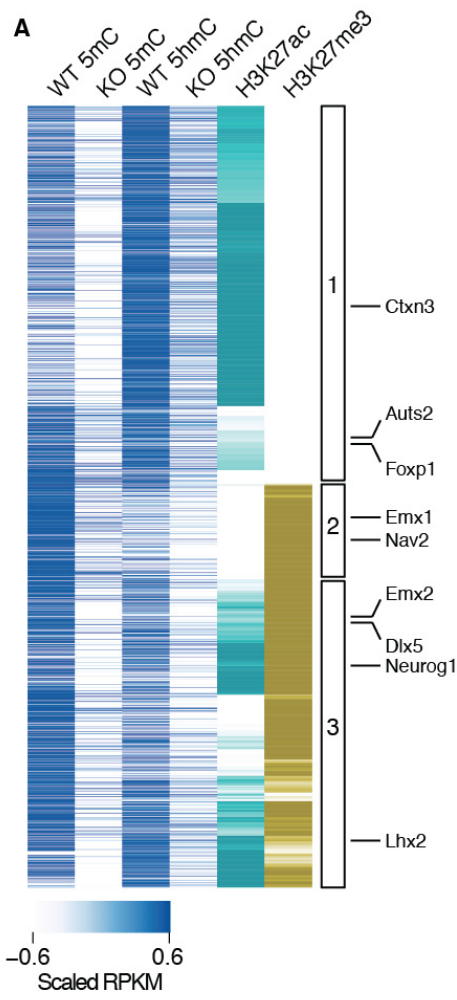


Figure 13. Dnmt3a-dependent 5mC and 5hmC patterning segregates with differential histone modifications.

(A) Hierarchical clustering of the union of Dnmt3a-dependent 5mC or 5hmC regions by average H3K27ac and H3K27me3 levels. Displayed are average 5mC and 5hmC RPKM in Dnmt3a WT and KO mOSNs and average H3K27ac and H3K27me3 RPKM in Dnmt3a WT MOE. Each track is input-subtracted and scaled to a common range. Numbered boxes indicate positions of three broad clusters from hierarchical clustering. Representative genes associated with each cluster are indicated at far right. (B) Average levels of 5mC and 5hmC in WT, Het, and KO mOSNs within each of the three broad clusters indicated in (A). (C) Tracks of representative genes found in cluster 1 (Rarg) and cluster 2 (Emx1). CGIs, CpG islands. Shaded boxes denote regions containing Dnmt3a-dependent 5mC and 5hmC. Y-axis is RPKM. H3K27ac and H3K27me3 are subtracted from Dnmt3a WT native chromatin input tracks. (D) As (C) with representative genes found in cluster 3 (Dlx5/6).

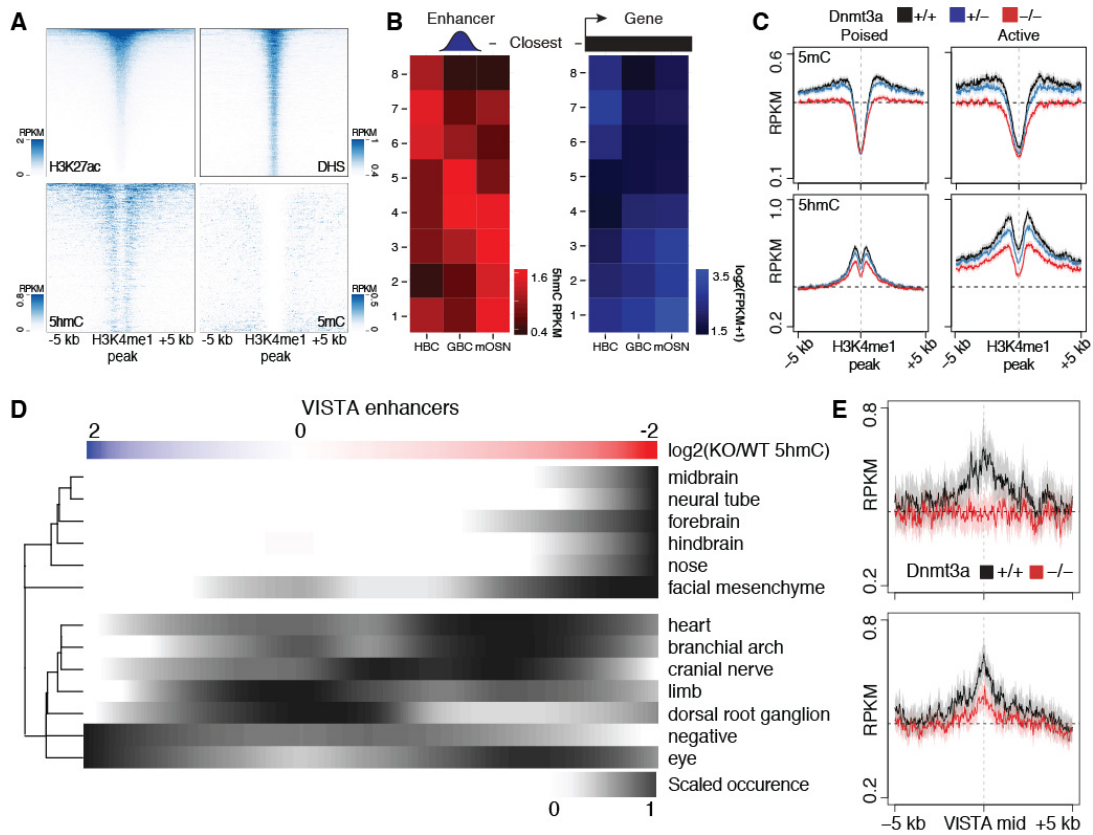


Figure 14. Dnmt3a is required for neuronal 5mC and 5hmC patterning at enhancers.

(A) MOE H3K27ac, MOE DNase hypersensitivity (DHS), mOSN 5hmC, and mOSN 5mC levels over MOE intergenic H3K4me1 peaks ordered by H3K27ac levels. (B) Analysis of the relationship between enhancer 5hmC levels (left) and the expression of closest genes (within 50 kb, right) within HBCs, GBCs, and mOSNs. (C) Average 5mC and 5hmC levels flanking ‘poised’ and ‘active’ enhancers in mature olfactory sensory neurons (mOSNs) from Dnmt3a +/+, +/-, and -/- MOEs. Error bars are bootstrapped 95% confidence intervals. (D) Association of neural enhancer activity with Dnmt3a-dependent 5hmC. A matrix of VISTA enhancer activities (rows) was ordered by $\log_2(\text{Dnmt3a KO/WT 5hmC RPKM})$ (columns). Each row was then loess smoothed and normalized, and the matrix was hierarchically clustered. (E) Average 5hmC levels at VISTA enhancers in Dnmt3a +/+ and -/- mOSNs. Top, VISTA enhancers with activities in the top cluster in (D). Bottom, VISTA enhancers with activities in the bottom cluster in (D).

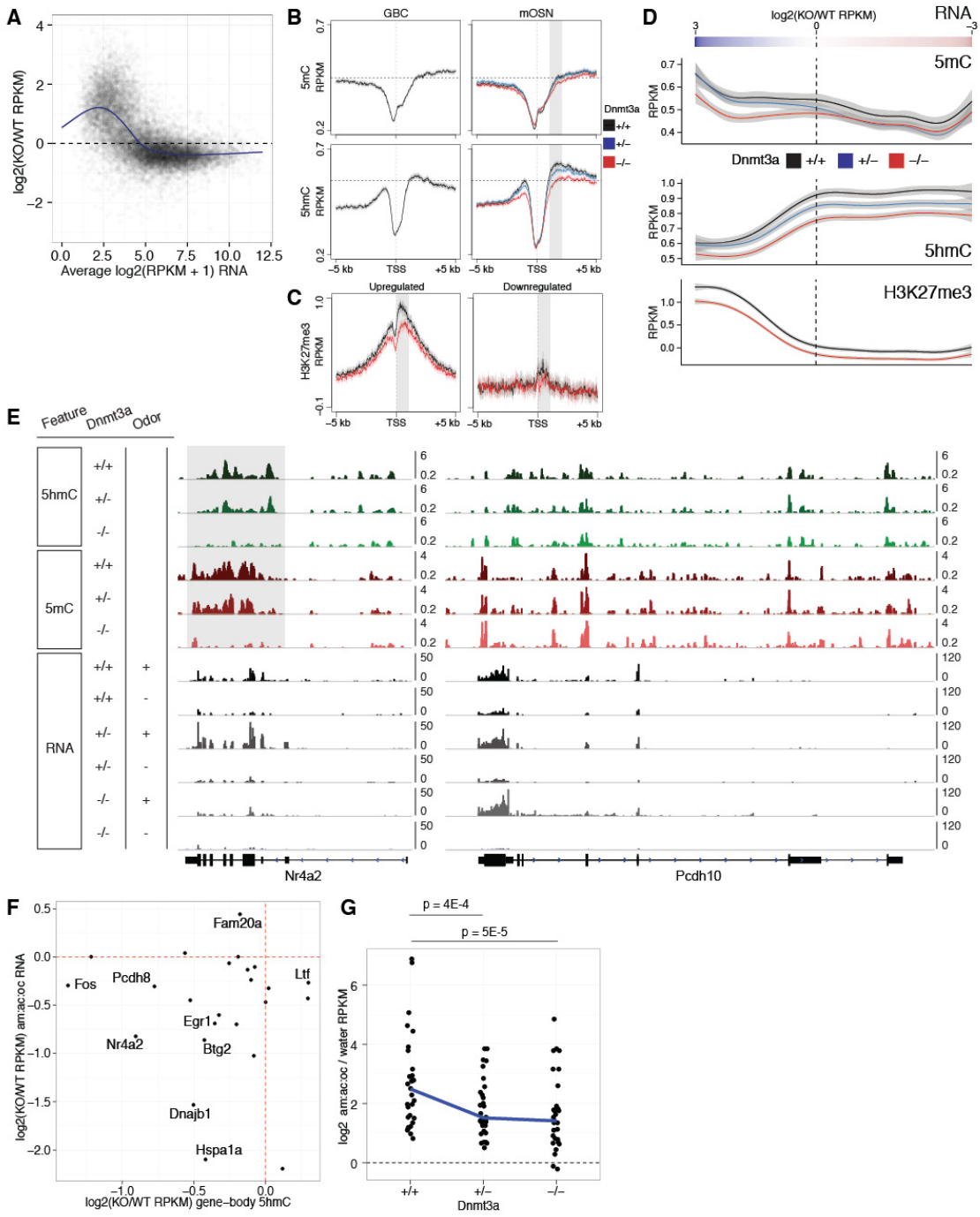
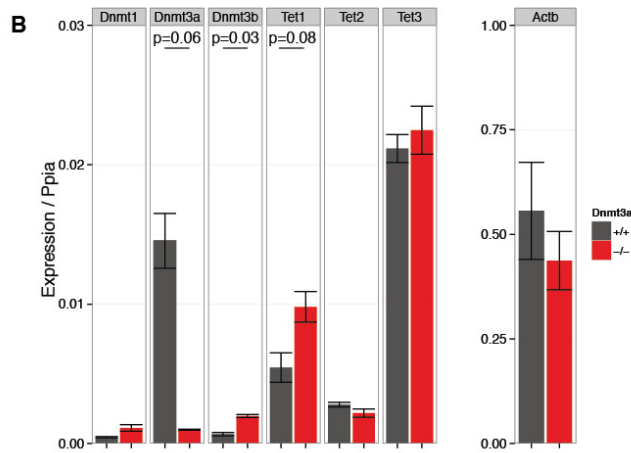
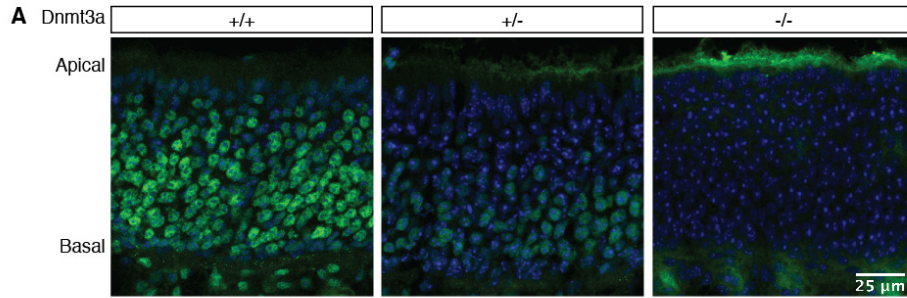


Figure 15. Loss of Dnmt3a disrupts global and activity-induced transcription in neurons.

(A) Gene expression in Dnmt3a WT and KO mOSNs. *X-axis*, average $\log_2(\text{RPKM}+1)$ across Dnmt3a WT and KO for each gene. *Y-axis*, $\log_2(\text{KO/WT RPKM}+1)$ for each gene. (B) Average 5mC and 5hmC levels flanking TSSs in GBCs and Dnmt3a *+/+*, *+/-*, and *-/-* mOSNs. Shaded areas indicate regions used for analysis shown in (D). (C) Average H3K27me3 levels flanking TSSs in Dnmt3a *+/+* and *-/-* MOEs. ‘Upregulated’ is $\log_2(\text{KO/WT RNA RPKM}) > 1$. ‘Downregulated’ is $\log_2(\text{KO/WT RNA RPKM}) < -1$. Shaded areas indicate regions used for analysis shown in (D). (D) Average mOSN 5mC and 5hmC levels 1-2kb downstream and MOE H3K27me3 levels 0-1kb downstream of TSSs ordered by KO/WT expression fold change. (E) 5hmC, 5mC, and RNA tracks of two representative odorant-induced genes, *Nr4a2* and *Pcdh10*. Odor ‘+’, 1 hour exposure to 1:1:1 amyl acetate:acetophenone:octanal (am:ac:oc); odor ‘-’, 1 hour exposure to a water. (F) Gene-body 5hmC loss is correlated with reduced odorant-induced expression. *X-axis*, average $\log_2(\text{KO/WT RPKM})$ 5hmC across gene-bodies; *y-axis*, $\log_2(\text{KO/WT})$ RPKM of induced RNA expression of 25 odorant-induced genes. (G) Odorant-induced gene expression (\log_2 am:ac:oc/water RPKM) of 25 odorant-induced genes within Dnmt3a *+/+*, *+/-*, and *-/-* MOEs. Blue line intersects at medians. P-values are two-sided Wilcoxon.



C

mOSN				
Genotype	Number of animals	Percentage single cells	SEM	Relative to wildtype
+/+	4	45	3	1.00
+/-	5	49	2	1.09
-/-	9	45	1	1.00

GBC				
Genotype	Number of animals	Percentage single cells	SEM	Relative to wildtype
+/+	2	28	2	1.00
+/-	2	32	1	1.14
-/-	3	32	2	1.14

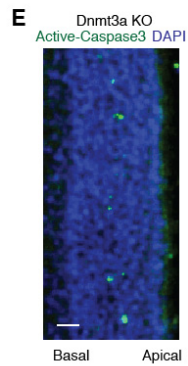
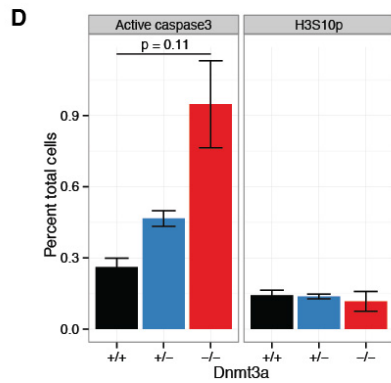


Figure 16. Supplemental analysis of mOSN development in Dnmt3a KO MOEs.

(A) Dnmt3a immunofluorescence of coronal MOE sections from postnatal day 21 Dnmt3a wildtype (+/+), heterozygous-null (+/-), and homozygous-null (-/-) mice. (B) RT-qPCR of whole MOE RNA for DNA methyltransferases (*Dnmt1*, *Dnmt3a*, *Dnmt3b*), Tet family members (*Tet1/2/3*), and beta-actin (*Actb*). Values are normalized to the expression levels of the control genes *Ppia*. P-values from two-sided Student's t-test, N=3. (C) FACS mOSN and GBC abundance quantification. Quantification of GFP positive cells from 3-week old Dnmt3a +/+, +/-, and -/- mice carrying either the OMP-ires-EGFP (mOSN) or Neurog1-EGFP (GBC) all. Single cells were determined as the number of cells after gating for forward and side scatters. (D) Quantification of the number of apoptotic (anti-active caspase-3 positive) and mitotic (anti-H3S10p) within the MOE represented as a percentage of the total number of cells (determined from DAPI staining) within the quantified area. P-value is from two-sided paired Student's t-test, N=4. (E) Representative image of anti-active caspase-3 staining within Dnmt3a KO MOEs. Positive cells are positioned at approximately the immature olfactory sensory neurons layer. Scale bar, 25 μ m.

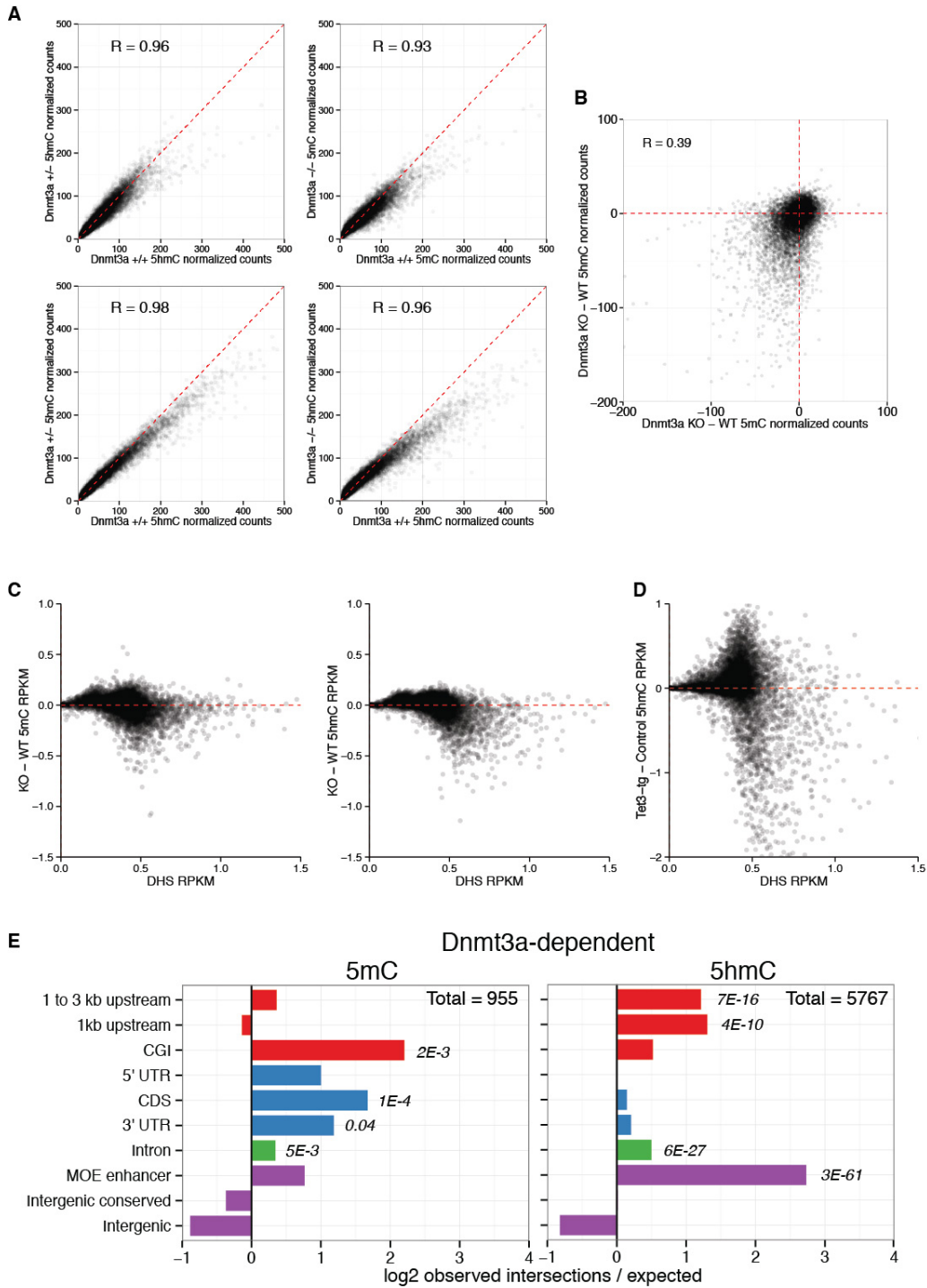


Figure 17. General analysis of Dnmt3a-dependent 5mC and 5hmC patterning in mOSNs.

(A) Scatter plots of average 5mC and 5hmC normalized (by DESeq median) read counts within genome-wide 5 kilobase windows (here 10,000 random (~2% of total number) windows are shown). The x-axis for each plot is Dnmt3a WT normalized counts. Y-axes are either from Dnmt3a Het (*left*) or KO (*right*). Correlation values are Pearson R. (B) Scatter plots of KO – WT 5mC (x-axis) and 5hmC (y-axis) normalized read counts within genome-wide 5 kilobase windows (shown 50,000 windows). Correlation values are Pearson R. (C) Dnmt3a KO – WT 5mC (*left*) and 5hmC (*right*) RPKM versus Dnmt3a Het DNase I hypersensitivity (DHS). (D) Tet3-tg – Control mOSN 5hmC RPKM (Colquitt, 2013) versus Dnmt3a Het DNase I hypersensitivity (DHS). (E) Genomic feature-peak intersections represented as the ratio of the observed number of peaks that intersect with a given feature set over the expected number of peaks (determined the relative size of the feature set). Peaks were generated using HOMER comparing WT versus KO 5mC or 5hmC (size 500 bp, minimum peak stitching distance 100 bp fold difference > 2, FDR 0.1%). Benjamini-Hochberg corrected P-values from Fisher's exact test.

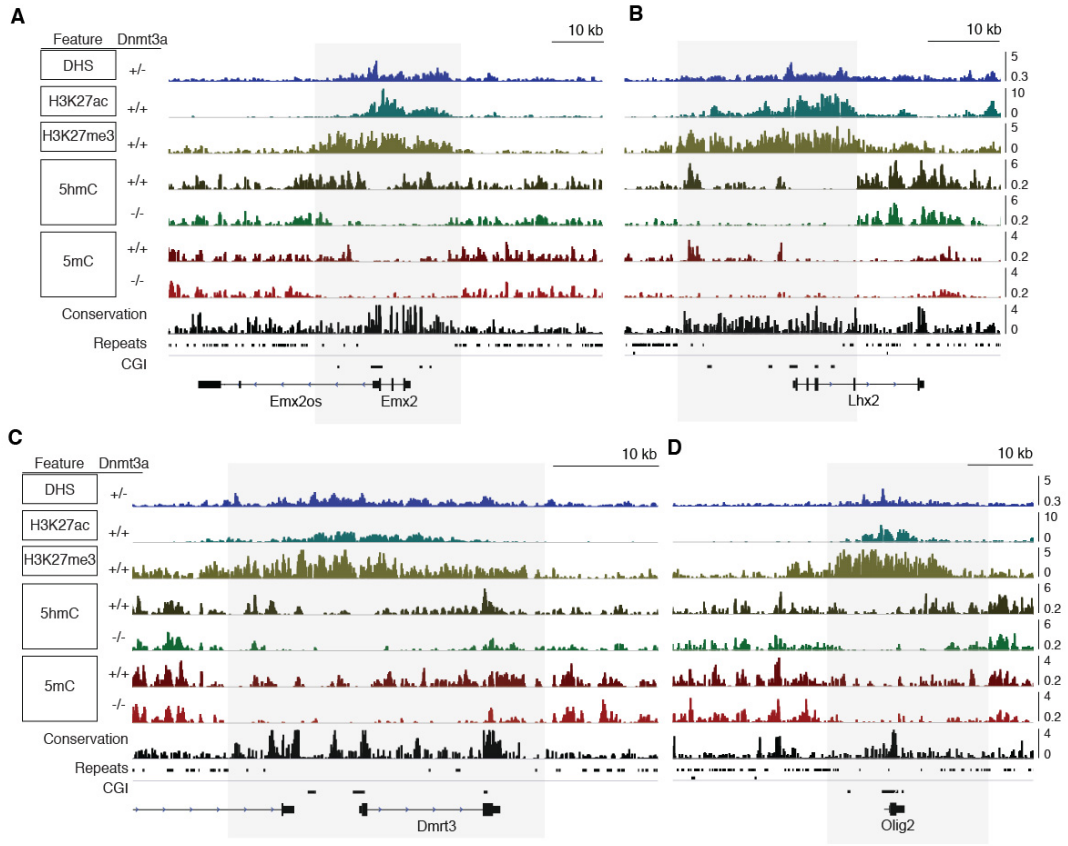


Figure 18. Representative gene loci in H3K27ac/H3K27me3 cluster 3.

(A) *Emx2*, (B) *Lhx2*, (C) *Dmrt3*, (D) *Olig2*. DHS -- DNase I hypersensitivity; Conservation – phyloP30Way; Repeats - Repeatmasker SINE, LINE, LTR, DNA, and RNA repeats; CGI - CpG islands, cgpIslandExt table (UCSC Genome Browser). Shaded boxes indicate are of Dnmt3a-dependent 5mC and 5hmC. Y-axis is RPKM. H3K27ac and H3K27me3 tracks are subtracted by Dnmt3a WT native chromatin input tracks.

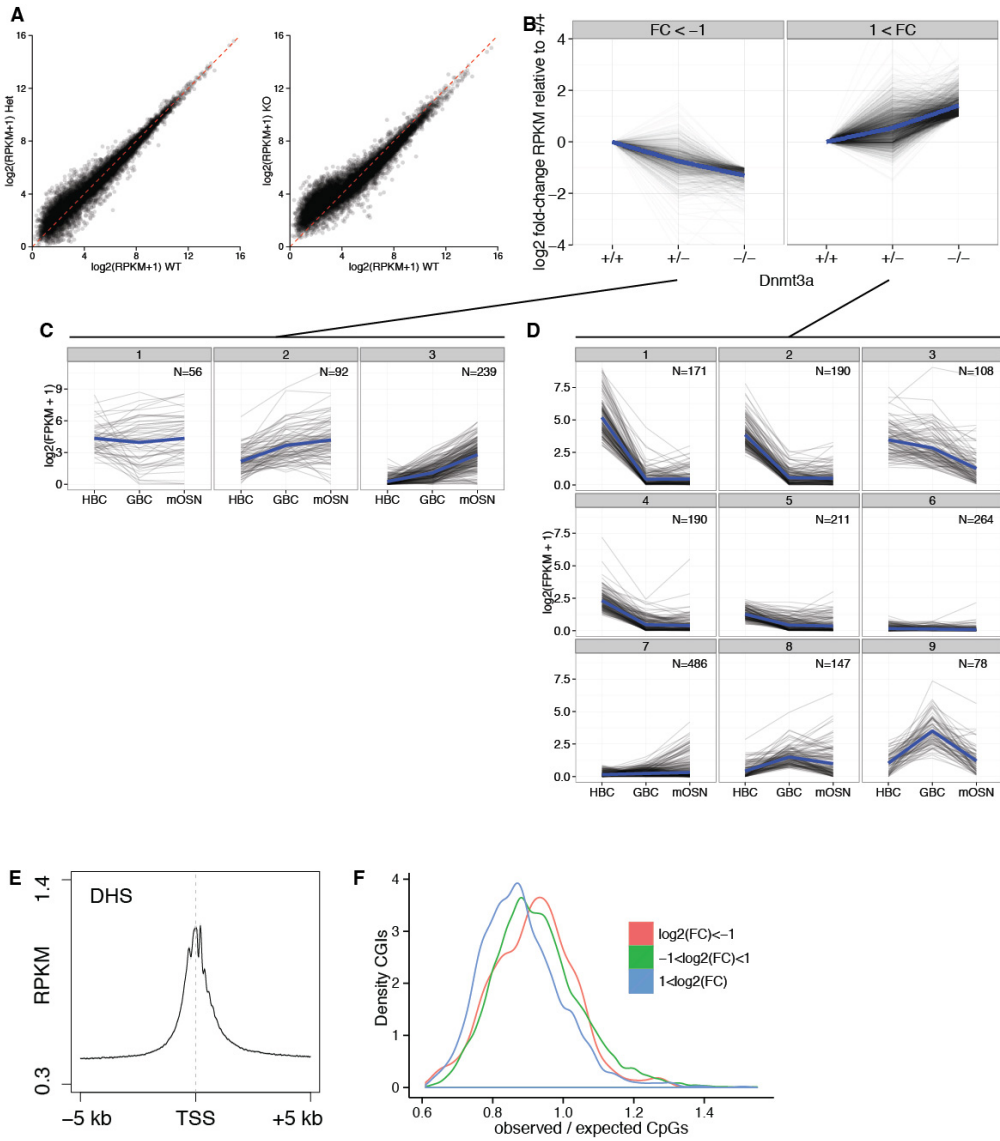


Figure 19. Supplemental analysis of transcription profiles in Dnmt3a WT, Het, and KO mOSNs.

(A) Comparison of Het (*left*) and KO (*right*) mOSN rRNA-depleted RNA-seq $\log_2(\text{RPKM}+1)$ to WT. (B) RNA RPKM values in Dnmt3a WT, Het, and KO mOSNs normalized to wildtype (\log_2). Displayed are values from genes that have an absolute \log_2 fold change greater than 1. (C) Developmental of mRNA expression in horizontal basal cells (HBC), globose basal cells (GBC), and mature olfactory sensory neurons (mOSNs, Colquitt, 2013) of genes with Dnmt3a $\log_2(\text{KO}/\text{WT RPKM}) < 1$. Clustering was performed using HOPACH (“cosangle” distance metric) and trimming to the highest level to produce 3 broad clusters. (D) As (B) but with genes with Dnmt3a $\log_2(\text{KO}/\text{WT RPKM}) > 1$. HOPACH clustering generated 9 broad clusters. (E) Average DNase hypersensitivity in Dnmt3a Het MOEs across all transcription start sites. (F) Density plots of the observed/expected number CpGs of CGIs within 1 kb of a gene. CGIs are separated by the \log_2 KO/WT RNA-seq RPKM fold-change of associated genes.

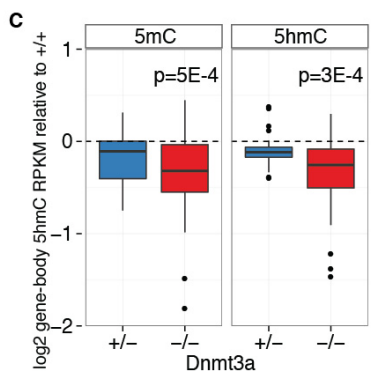
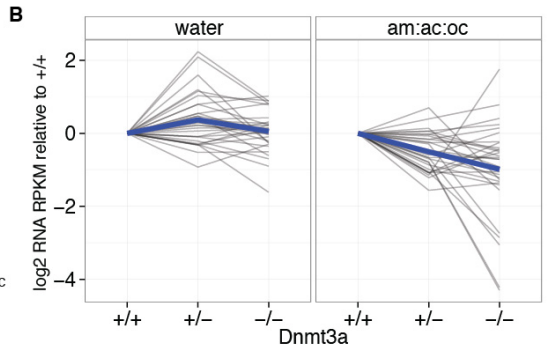
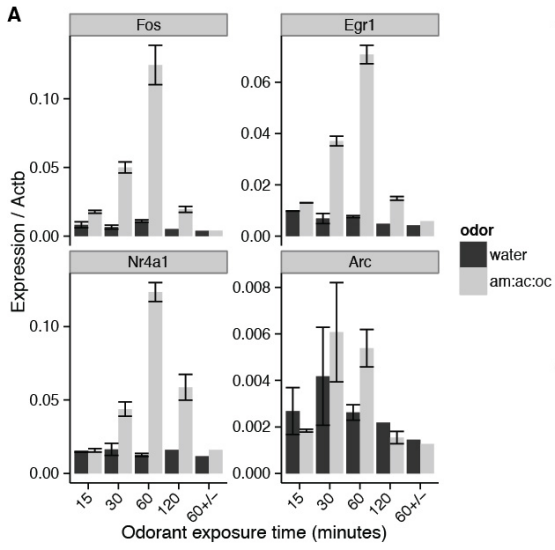


Figure 20. Supplemental analysis of odorant-induced gene expression.

(A) RT-qPCR of *Fos*, *Egr1*, *Nr4a1*, and *Arc* in MOEs from 3-week old wildtype mice that had been exposed to either water (odor-) or 1:1:1 amyl acetate:acetophenone:octanal (am:ac:oc, odor+) for 15, 30, 60, or 120 minutes or exposed for 60 minutes then allowed to recover in a new cage for 60 minutes (60+/-). Expression levels are expressed as the fraction of *Actb* expression. Error bars are the range of two biological replicates. (B) Wildtype-normalized RPKM values of 25 odorant-induced genes in either the unexposed (water) or 1 hour odor (1:1:1 am:ac:oc) conditions. Blue bars intersect with the median of each genotype. (C) Average 5hmC and 5mC levels over gene-bodies of odorant-induced genes, normalized by wildtype levels. P-values are two-sided Wilcoxon.

Sample	Mouse age	Sequencing type	Number of cycles	Mapped and filtered read number
Dnmt3a +/- mOSN 5hmC	3 weeks	Paired-end	2x50	88,940,807
Dnmt3a +/- mOSN 5hmC rep1	3 weeks	Paired-end	2x50	44,996,941
Dnmt3a +/- mOSN 5hmC rep2	3 weeks	Paired-end	2x50	120,559,770
Dnmt3a -/- mOSN 5hmC rep1	3 weeks	Paired-end	2x50	45,795,005
Dnmt3a -/- mOSN 5hmC rep2	3 weeks	Paired-end	2x50	63,718,121
Dnmt3a +/- mOSN 5mC	3 weeks	Paired-end	2x50	82,177,179
Dnmt3a +/- mOSN 5mC rep1	3 weeks	Paired-end	2x50	37,777,674
Dnmt3a +/- mOSN 5mC rep2	3 weeks	Paired-end	2x50	60,368,514
Dnmt3a -/- mOSN 5mC rep1	3 weeks	Paired-end	2x50	47,981,097
Dnmt3a -/- mOSN 5mC rep2	3 weeks	Paired-end	2x50	51,473,796
Dnmt3a +/- mOSN rRNA-depleted RNA rep1	3 weeks	Paired-end	2x50	155,407,945
Dnmt3a +/- mOSN rRNA-depleted RNA rep2	3 weeks	Paired-end	2x50	167,389,197
Dnmt3a +/- mOSN rRNA-depleted RNA rep1	3 weeks	Paired-end	2x50	65,308,180
Dnmt3a +/- mOSN rRNA-depleted RNA rep2	3 weeks	Paired-end	2x50	77,991,518
Dnmt3a -/- mOSN rRNA-depleted RNA rep1	3 weeks	Paired-end	2x50	135,615,380
Dnmt3a -/- mOSN rRNA-depleted RNA rep2	3 weeks	Paired-end	2x50	138,510,404
Dnmt3a +/- MOE H3K4me1 rep1	3 weeks	Paired-end	2x50	150,876,233
Dnmt3a +/- MOE H3K4me1 rep2	3 weeks	Paired-end	2x50	191,665,025
Dnmt3a +/- MOE H3K27ac rep1	3 weeks	Paired-end	2x50	75,308,226
Dnmt3a +/- MOE H3K27ac rep2	3 weeks	Paired-end	2x50	80,653,918
Dnmt3a +/- MOE H3K27me3 rep1	3 weeks	Paired-end	2x50	52,632,947
Dnmt3a +/- MOE H3K27me3 rep2	3 weeks	Paired-end	2x50	40,300,679
Dnmt3a -/- MOE H3K27me3 rep1	3 weeks	Paired-end	2x50	90,780,313
Dnmt3a -/- MOE H3K27me3 rep2	3 weeks	Paired-end	2x50	31,358,084
Dnmt3a +/- MOE DNase	3 weeks	Paired-end	2x50	114,301,834
Dnmt3a +/- MOE rRNA-depleted RNA H2O	3 weeks	Paired-end	2x50	82,225,026
Dnmt3a +/- MOE rRNA-depleted RNA H2O	3 weeks	Paired-end	2x50	83,624,526
Dnmt3a -/- MOE rRNA-depleted RNA H2O	3 weeks	Paired-end	2x50	57,808,710
Dnmt3a +/- MOE rRNA-depleted RNA am:ac:oc	3 weeks	Paired-end	2x50	73,259,282
Dnmt3a +/- MOE rRNA-depleted RNA am:ac:oc	3 weeks	Paired-end	2x50	70,823,704
Dnmt3a -/- MOE rRNA-depleted RNA am:ac:oc	3 weeks	Paired-end	2x50	71,941,860
Dnmt3a +/- MOE native input	3 weeks	Paired-end	2x50	156,825,857
Dnmt3a -/- MOE native input	3 weeks	Paired-end	2x50	85,109,054

Table 2. Chapter 3 sequencing metadata.

Mapping and filtering performed as in methods. Filtered read number for ChIP and DIP samples is the number of reads after filtering for both quality and read duplication.

Ensembl_id	gene	logFC	logCPM	PValue	FDR	
ENSMUSG00000021250	Fos		3.99	5.24	2.4507E-42	3.1012E-38
ENSMUSG00000038418	Egr1		3.56	6.52	5.0314E-30	3.1834E-26
ENSMUSG00000023034	Nr4a1		3.43	5.77	2.407E-28	1.0153E-24
ENSMUSG00000028341	Nr4a3		3.15	4.22	3.4412E-14	1.0886E-10
ENSMUSG00000091971	Hspa1a		4.67	8.12	2.8133E-11	7.1199E-08
ENSMUSG00000090877	Hspa1b		4.66	8.14	3.597E-11	7.5862E-08
ENSMUSG00000020423	Btg2		1.91	4.47	8.125E-11	1.3298E-07
ENSMUSG00000024042	Sik1		1.72	4.70	8.4068E-11	1.3298E-07
ENSMUSG00000044786	Zfp36		1.75	3.89	1.6825E-08	2.3656E-05
ENSMUSG00000049100	Pcdh10		1.71	7.36	6.0511E-08	7.6571E-05
ENSMUSG00000026826	Nr4a2		1.99	4.43	8.6609E-08	9.1904E-05
ENSMUSG00000074063	Osgin1		1.80	4.30	8.7154E-08	9.1904E-05
ENSMUSG00000005483	Dnajb1		2.32	5.22	1.0459E-07	0.0001018
ENSMUSG00000038894	Irs2		1.23	5.95	1.6135E-07	0.00014584
ENSMUSG00000051022	Hs3st1		1.42	3.32	4.0042E-07	0.0003378
ENSMUSG00000020893	Per1		1.22	5.79	6.7447E-07	0.00053342
ENSMUSG00000036422	Pcdh8		1.30	4.31	2.2074E-06	0.0016431
ENSMUSG00000075014	Gm10800		1.13	5.28	7.6291E-06	0.00514421
ENSMUSG00000032496	Ltf		1.49	6.54	7.724E-06	0.00514421
ENSMUSG00000004951	Hspb1		2.09	4.02	9.9482E-06	0.0062942
ENSMUSG00000024190	Dusp1		1.15	4.98	1.7568E-05	0.01058611
ENSMUSG00000020614	Fam20a		1.14	3.71	4.2112E-05	0.02422192
ENSMUSG00000032526	Deb1		-1.24	2.12	5.1998E-05	0.02860819
ENSMUSG00000027376	Prom2		1.28	2.53	5.4355E-05	0.02865843
ENSMUSG00000026819	Slc25a25		1.20	4.53	7.3118E-05	0.03700964
ENSMUSG00000006204	5430419D17Rik		0.99	6.47	8.8522E-05	0.04308289

Table 3. Odorant-induced genes.

Statistical analysis performed using the R package *edgeR* treating WT, Het, KO RNA-seq data as biological replicates and using water or am:ac:oc-exposed at differential factor. Columns are Ensemble gene ID, official gene symbol, log am:ac:oc/water fold change (logFC), log counts per million averaged across samples (logCPM), p-value is from exact test for the negative binomial distribution as described in edgeR manual, and FDR are Benjamini-Hochberg corrected p-values.

Chapter 4: General discussion

In just the past few years, there has been sea change in our understanding of the complex dynamics and functional contributions of epigenetic modifications during the construction of multicellular organisms. The number of histone modifications alone is now thought to number approximately 130, comprising a diverse range of chemical moieties found throughout histones (Tan et al., 2011). The discovery of a regulated system of cytosine oxidation greatly expanded the set of DNA modifications, long believed to consist only 5-methylcytosine in vertebrates. With this expansion came the exciting potential to uncover new gene regulatory mechanisms that might underlie some of the diversity of cell types.

Oxidation targeting

In this work, I presented experiments that examined the role of 5-hydroxymethylcytosine, the initial oxidized derivative of 5mC, in olfactory sensory neuron development and function. As in other neural systems (Ruzov et al., 2011; Szulwach et al., 2011b), 5hmC abundance increases along the development lineage of mOSNs with enriched levels of the modified base in globose basal cells, the neuronal progenitors of the tissue, and mOSNs. This gain of 5hmC tends to occur in gene regions where levels of the modified base positively correlate with transcription. Correspondingly, gene-body 5hmC levels in developmentally regulated genes track with cell type-specific expression.

Gene-associated 5hmC patterning is tightly confined the transcriptional unit, suggesting that transcriptional machinery and Tet enzymes are closely coupled. Intriguingly, there is a genetic requirement for elongator complex proteins (Elp1, 3, and 4) during demethylation of the paternal genome in early embryogenesis (Okada et al., 2010), a process that is mediated by widespread Tet3-dependent 5mC oxidation (Gu et al., 2011). Biochemical interaction between Elp and Tet family members has not been demonstrated, but this result could help explain the precise

enrichment of 5hmC within gene-bodies and immediately upstream of the TSS. Such an association with the transcription complex could also provide an explanation for the enrichment of 5hmC flanking enhancers, which produce low levels of RNA (Kim et al., 2010). In ESCs, Tet1 is strongly localized to TSSs (Williams et al., 2011; Xu et al., 2011), however the genomic occupancy of Tet3, the major Tet family member in mOSNs, has not been determined in either ESCs or neurons. The CXXC domains of the three Tet enzymes have different sequence preferences *in vitro* (see Chapter 1) suggesting the enzymes may target different regions and generate oxidized 5mC in distinct contexts.

Possible mechanisms

The positive correlation between gene-body 5hmC levels and transcription has been observed in other systems, yet there has been no direct evidence that the modification actually contributes to gene regulation or if it is simply a byproduct of transcription. By altering levels of 5hmC through the overexpression of Tet3 in mOSNs, I showed that gene-body 5hmC positively influences transcriptional output. Gain of the mark results a tendency for higher expression; depletion of the mark strongly inhibits transcription. Because this trend was seen across genes and was not restricted to particular classes, I proposed that gene-body 5hmC acts as a general modulator of transcription. Such a mechanism could allow the fine-tuning of transcriptional output from actively engaged genes. Likewise, enriched 5hmC may establish a transcriptionally competent state, such that poised or lowly transcribed genes could quickly achieve high levels of expression upon induction.

It is still unclear how these regulatory effects are transduced. One could imagine that 5hmC recruits a set of proteins that facilitate transcriptional elongation. However, two of the proteins that bind 5hmC, Mbd3 in ESCs and MeCP2 in neurons, have traditionally been associated with

transcriptional repression. This repression is thought to occur via the secondary recruitment of histone deacetylases and methyltransferases by the methyl-binding proteins (Fuks, 2005; Jones et al., 1998). However the loss of MeCP2 in the nervous system has mild effects on transcription (Tudor et al., 2002), which is surprising considering its abundance in neurons (Skene et al., 2010) and the widespread distribution of its targets (5mC and now 5hmC). Moreover, experiments in which MeCP2 levels were altered in the hypothalamus indicate that the protein can both repress and activate transcription (Chahrour et al., 2008), consistent with its ability to bind both 5mC and 5hmC. Moreover, MeCP2 is phosphorylated in response to neuronal activity which alters its ability to regulate dendritic patterning and spine morphogenesis (Zhou et al., 2006). This suggests that the protein may be able to differentially influence gene expression depending on what it is binding and its post-translational modification state.

An alternate model suggests that 5hmC influences transcription not by altering transcription factor binding but by adjusting interactions between genomic DNA and nucleosomes. During transcriptional elongation, RNA polymerase must pass across or evict the nucleosomes that package genomic DNA, a process that requires the disruption of nucleosome-DNA interactions and translocation of nucleosomes from preferred positions (Henikoff, 2008; Izban and Luse, 1991; Studitsky et al., 1995; 1997). CG DNA methylation contributes to the stable positioning of nucleosomes both *in vitro* and in human T cells (Choi et al., 2009; Choy et al., 2010), and recent work in *Arabidopsis* found a strong enrichment of CG methylation within nucleosome-protected DNA (Chodavarapu et al., 2010). Moreover, DNA methylation and CG dinucleotides are distributed across the mammalian genome with an ~8-10 bp periodicity, roughly equal to one turn of the DNA helix (Glass et al., 2009; Jia et al., 2007; Lister et al., 2008), suggesting that methylated cytosines have a preferred orientation with respect to the nucleosome-DNA interface.

Oxidation of 5mC could interfere with this stabilizing effect, allowing nucleosomes to slide along the DNA strand more easily. Thus, elevated 5hmC may allow the ‘escape’ of nucleosomes to methylated or unmodified flanking regions with higher nucleosome affinity. Although attractive, this model does not yet have any experimental support.

Construction of the neuronal epigenome

The study of 5hmC in vertebrate neurons is a relatively new field replete with exciting possibilities, deep speculation, and mechanistic models that are only just beginning to take shape. It still remains to be determined how 5hmC patterning intersects with neuronal identity and function. Is there some property of neuron physiology or morphology that requires or is facilitated by the use of this modification? I believe that an examination of odorant-induced gene regulation in the context of altered DNA modification patterns will be a fruitful way to begin answering these questions. The data from the initial odorant exposure experiments suggest that OSNs have moderate immediate early gene activation that is compromised in the absence of Dnmt3a. However, due to the current lack of biological replicate data, I was not able to confidently assess if the loss Dnmt3a also results in the inappropriate activation of certain genes in response to odorant exposure. Such an analysis would provide insight into the functional restriction of the neuronal genome by its epigenome. That is, if the loss of Dnmt3a allows particular genes to be activated by neuronal activity, it may indicate that DNA modifications are necessary to control which genes are sensitive to the signaling environment of the neuron and those that are not.

This developmental restriction of competence has an appealing analogue in models that hypothesize that DNA methylation expanded in vertebrates to compensate for the increase in parasitic genome elements (Bird, 2002). DNA methylation may have allowed vertebrates to

reduce the effective size of their genomes, thereby simplifying its interpretation, by sequestering repetitive content while keeping host genomic material ‘exposed’ (Bestor, 1990). Similarly, neurons may employ complementary 5mC and 5hmC patterning to specify the repressed and engaged regions of their genomes during neurogenesis (Figure 21). Through this process, stable identity and consistent transcriptional responses could be established within a dynamic signaling environment.

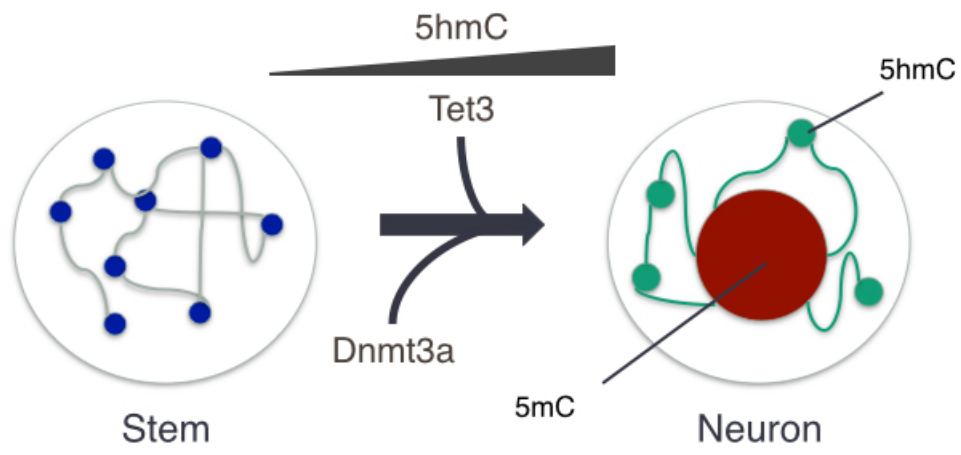


Figure 21. Speculative model of the contributions of Dnmt3a and Tet3 to establishing a polarized neuronal genome.

Cellular differentiation can be viewed as process of developmental restriction. On the left, a representative stem cell has both open developmental potential and chromatin structure. During neurogenesis, chromatin becomes more structured and polarized, here indicated by complementary distributions of 5mC and 5hmC. Dnmt3a and Tet3 cooperate to establish this neuronal epigenetic state.

Chapter 5: References

- Anders, S., and Huber, W. (2010). Differential expression analysis for sequence count data. *Genome Biol* *11*, R106.
- Bachman, K.E., Rountree, M.R., and Baylin, S.B. (2001). Dnmt3a and Dnmt3b are transcriptional repressors that exhibit unique localization properties to heterochromatin. *J Biol Chem* *276*, 32282–32287.
- Bell, A.C., West, A.G., and Felsenfeld, G. (1999). The protein CTCF is required for the enhancer blocking activity of vertebrate insulators. *Cell* *98*, 387–396.
- Ben Langmead, and Salzberg, S.L. (2012). Fast gapped-read alignment with Bowtie 2. *Nat Meth* *9*, 357–359.
- Benowitz, L.I., and Routtenberg, A. (1997). GAP-43: an intrinsic determinant of neuronal development and plasticity. *Trends Neurosci* *20*, 84–91.
- Bestor, T.H. (1990). DNA methylation: evolution of a bacterial immune function into a regulator of gene expression and genome structure in higher eukaryotes. *Philos. Trans. R. Soc. Lond., B, Biol. Sci.* *326*, 179–187.
- Bestor, T.H. (1992). Activation of mammalian DNA methyltransferase by cleavage of a Zn binding regulatory domain. *Embo J* *11*, 2611–2617.
- Bestor, T., Laudano, A., Mattaliano, R., and Ingram, V. (1988). Cloning and sequencing of a cDNA encoding DNA methyltransferase of mouse cells. The carboxyl-terminal domain of the mammalian enzymes is related to bacterial restriction methyltransferases. *J Mol Biol* *203*, 971–983.
- Bird, A.P. (1980). DNA methylation and the frequency of CpG in animal DNA. *Nucleic Acids Research* *8*, 1499–1504.
- Bird, A.P. (1986). CpG-rich islands and the function of DNA methylation. *Nature* *321*, 209–213.
- Bird, A.P., Taggart, M.H., and Smith, B.A. (1979). Methylated and unmethylated DNA compartments in the sea urchin genome. *Cell*.
- Bird, A. (2002). DNA methylation patterns and epigenetic memory. *Genes Dev* *16*, 6–21.
- Bird, A. (2007). Perceptions of epigenetics. *Nature* *447*, 396–398.
- Bird, A.P. (1987). CpG islands as gene markers in the vertebrate nucleus. *Trends Genet* *3*, 342–347.
- Birke, M., Schreiner, S., García-Cuéllar, M.-P., Mahr, K., Titgemeyer, F., and Slany, R.K. (2002). The MT domain of the proto-oncoprotein MLL binds to CpG-containing DNA and discriminates against methylation. *Nucleic Acids Research* *30*, 958–965.
- Borst, P., and Sabatini, R. (2008). Base J: Discovery, Biosynthesis, and Possible Functions. *Annu. Rev. Microbiol.* *62*, 235–251.

- Bostick, M., Kim, J.K., Estève, P.-O., Clark, A., Pradhan, S., and Jacobsen, S.E. (2007). UHRF1 plays a role in maintaining DNA methylation in mammalian cells. *Science* 317, 1760–1764.
- Bourchis, D., Xu, G.L., Lin, C.S., Bollman, B., and Bestor, T.H. (2001). Dnmt3L and the establishment of maternal genomic imprints. *Science* 294, 2536–2539.
- Boyes, J., and Bird, A. (1991). DNA methylation inhibits transcription indirectly via a methyl-CpG binding protein. *Cell* 64, 1123–1134.
- Branco, M.R., Ficz, G., and Reik, W. (2011). Uncovering the role of 5-hydroxymethylcytosine in the epigenome. *Nat Rev Genet* 13, 7–13.
- Brandeis, M., Kafri, T., Ariel, M., and Chaillet, J.R. (1993). The ontogeny of allele-specific methylation associated with imprinted genes in the mouse. *The EMBO ...*
- Caggiano, M., Kauer, J.S., and Hunter, D.D. (1994). Globose basal cells are neuronal progenitors in the olfactory epithelium: a lineage analysis using a replication-incompetent retrovirus. *Neuron* 13, 339–352.
- Campuzano, S., and Modolell, J. (1992). Patterning of the Drosophila nervous system: the achaete-scute gene complex. *Trends Genet* 8, 202–208.
- Carr, V.M., Farbman, A.I., Colletti, L.M., and Morgan, J.I. (1991). Identification of a new non-neuronal cell type in rat olfactory epithelium. *Neuroscience* 45, 433–449.
- Cau, E., Gradwohl, G., Fode, C., and Guillemot, F. (1997). Mash1 activates a cascade of bHLH regulators in olfactory neuron progenitors. *Development* 124, 1611–1621.
- Cau, E., Casarosa, S., and Guillemot, F. (2002). Mash1 and Ngn1 control distinct steps of determination and differentiation in the olfactory sensory neuron lineage. *Development* 129, 1871–1880.
- Chahrour, M., Jung, S.Y., Shaw, C., Zhou, X., Wong, S.T.C., Qin, J., and Zoghbi, H.Y. (2008). MeCP2, a key contributor to neurological disease, activates and represses transcription. *Science* 320, 1224–1229.
- Chodavarapu, R.K., Feng, S., Bernatavichute, Y.V., Chen, P.-Y., Stroud, H., Yu, Y., Hetzel, J.A., Kuo, F., Kim, J., Cokus, S.J., et al. (2010). Relationship between nucleosome positioning and DNA methylation. *Nature* 466, 388–392.
- Choi, J.K., Bae, J.-B., Lyu, J., Kim, T.-Y., and Kim, Y.-J. (2009). Nucleosome deposition and DNA methylation at coding region boundaries. *Genome Biol* 10, R89.
- Choy, J.S., Wei, S., Lee, J.Y., Tan, S., Chu, S., and Lee, T.-H. (2010). DNA methylation increases nucleosome compaction and rigidity. *J. Am. Chem. Soc.* 132, 1782–1783.
- Chuang, L.S., Ian, H.I., Koh, T.W., Ng, H.H., Xu, G., and Li, B.F. (1997). Human DNA-(cytosine-5) methyltransferase-PCNA complex as a target for p21WAF1. *Science* 277, 1996–2000.
- Clowney, E.J., LeGros, M.A., Mosley, C.P., Clowney, F.G., Markenskoff-Papadimitriou, E.C., Myllys, M., Barnea, G., Larabell, C.A., and Lomvardas, S. (2012). Nuclear aggregation of olfactory receptor genes governs their monogenic expression. *Cell* 151, 724–737.
- Clowney, E.J., Magklara, A., Colquitt, B.M., Pathak, N., Lane, R.P., and Lomvardas, S. (2011). High-throughput mapping of the promoters of the mouse olfactory receptor genes reveals a new

type of mammalian promoter and provides insight into olfactory receptor gene regulation. *Genome Res* 21, 1249–1259.

Cokus, S.J., Feng, S., Zhang, X., Chen, Z., Merriman, B., Haudenschild, C.D., Pradhan, S., Nelson, S.F., Pellegrini, M., and Jacobsen, S.E. (2008). Shotgun bisulphite sequencing of the *Arabidopsis* genome reveals DNA methylation patterning. *Nature* 452, 215–219.

Colquitt, B.M., Allen, W.E., Barnea, G., and Lomvardas, S. (2013). Alteration of genic 5-hydroxymethylcytosine patterning in olfactory neurons correlates with changes in gene expression and cell identity. *Proc Natl Acad Sci USA* 110, 14682–14687.

Cortellino, S., Xu, J., Sannai, M., Moore, R., Caretti, E., Cigliano, A., Le Coz, M., Devarajan, K., Wessels, A., Soprano, D., et al. (2011). Thymine DNA Glycosylase Is Essential for Active DNA Demethylation by Linked Deamination-Base Excision Repair. *Cell* 146, 67–79.

Creyghton, M.P., Cheng, A.W., Welstead, G.G., Kooistra, T., Carey, B.W., Steine, E.J., Hanna, J., Lodato, M.A., Frampton, G.M., Sharp, P.A., et al. (2010). Histone H3K27ac separates active from poised enhancers and predicts developmental state. *Proc Natl Acad Sci USA* 107, 21931–21936.

Cuschieri, A., and Bannister, L.H. (1975). The development of the olfactory mucosa in the mouse: light microscopy. *J. Anat.* 119, 277–286.

Dalton, R.P., Lyons, D.B., and Lomvardas, S. (2013). Co-Opting the Unfolded Protein Response to Elicit Olfactory Receptor Feedback. *Cell* 155, 321–332.

Dawlaty, M.M., Breiling, A., Le, T., Raddatz, G., Barrasa, M.I., Cheng, A.W., Gao, Q., Powell, B.E., Li, Z., Xu, M., et al. (2013). Combined Deficiency of Tet1 and Tet2 Causes Epigenetic Abnormalities but Is Compatible with Postnatal Development. *Dev Cell* 24, 310–323.

Dawlaty, M.M., Ganz, K., Powell, B.E., Hu, Y.-C., Markoulaki, S., Cheng, A.W., Gao, Q., Kim, J., Choi, S.-W., Page, D.C., et al. (2011). Tet1 Is Dispensable for Maintaining Pluripotency and Its Loss Is Compatible with Embryonic and Postnatal Development. *Cell Stem Cell* 9, 166–175.

De Felipe, J., Marco, P., Fairén, A., and Jones, E.G. (1997). Inhibitory synaptogenesis in mouse somatosensory cortex. *Cerebral Cortex*.

de Koning, A.P.J., Gu, W., Castoe, T.A., Batzer, M.A., and Pollock, D.D. (2011). Repetitive Elements May Comprise Over Two-Thirds of the Human Genome. *PLoS Genet* 7, e1002384.

Deaton, A.M., and Bird, A. (2011). CpG islands and the regulation of transcription. *Genes Dev* 25, 1010–1022.

Degano, A.L., Pasterkamp, R.J., and Ronnett, G.V. (2009). MeCP2 deficiency disrupts axonal guidance, fasciculation, and targeting by altering Semaphorin 3F function. *Mol Cell Neurosci* 42, 243–254.

Dhayalan, A., Rajavelu, A., Rathert, P., Tamas, R., Jurkowska, R.Z., Ragozin, S., and Jeltsch, A. (2010). The Dnmt3a PWWP Domain Reads Histone 3 Lysine 36 Trimethylation and Guides DNA Methylation. *J Biol Chem* 285, 26114–26120.

Dobin, A., Davis, C.A., Schlesinger, F., Drenkow, J., Zaleski, C., Jha, S., Batut, P., Chaisson, M., and Gingeras, T.R. (2013). STAR: ultrafast universal RNA-seq aligner. *Bioinformatics* 29, 15–21.

- Duggan, C.D., and Ngai, J. (2007). Scent of a stem cell. *Nature Neuroscience* *10*, 673–674.
- Ehrlich, M., Gama-Sosa, M.A., Huang, L.-H., Midgett, R.M., Kuo, K.C., McCune, R.A., and Gehrke, C. (1982). Amount and distribution of 5-methylcytosine in human DNA from different types of tissues or cells. *Nucleic Acids*
- Fan, G., Beard, C., Chen, R.Z., Csankovszki, G., Sun, Y., Siniaia, M., Biniszkiwicz, D., Bates, B., Lee, P.P., Kühn, R., et al. (2001). DNA Hypomethylation Perturbs the Function and Survival of CNS Neurons in Postnatal Animals. *The Journal of*
- Farbman, A.I. (1990). Olfactory neurogenesis: genetic or environmental controls? *Trends Neurosci* *13*, 362–365.
- Feng, J., Chang, H., Li, E., and Fan, G. (2005). Dynamic expression of de novo DNA methyltransferases Dnmt3a and Dnmt3b in the central nervous system. *J Neurosci Res* *79*, 734–746.
- Feng, J., Zhou, Y., Campbell, S.L., Le, T., Li, E., Sweatt, J.D., Silva, A.J., and Fan, G. (2010). Dnmt1 and Dnmt3a maintain DNA methylation and regulate synaptic function in adult forebrain neurons. *Nat Neurosci* *13*, 423–430.
- Feng, Q., and Zhang, Y. (2001). The MeCP1 complex represses transcription through preferential binding, remodeling, and deacetylating methylated nucleosomes. *Genes Dev* *15*, 827–832.
- Ficz, G., Branco, M.R., Seisenberger, S., Santos, F., Krueger, F., Hore, T.A., Marques, C.J., Andrews, S., and Reik, W. (2011). Dynamic regulation of 5-hydroxymethylcytosine in mouse ES cells and during differentiation. *Nature* *473*, 398–402.
- Fraley, C., and Raftery, A.E. (2002). Model-Based Clustering, Discriminant Analysis, and Density Estimation. *Journal of the American Statistical Association* *97*, 611–631.
- Frauer, C., Rottach, A., Meilinger, D., Bultmann, S., Fellinger, K., Hasenöder, S., Wang, M., Qin, W., Söding, J., Spada, F., et al. (2011). Different binding properties and function of CXXC zinc finger domains in Dnmt1 and Tet1. *PLoS ONE* *6*, e16627.
- Frisch, D. (1967). Ultrastructure of mouse olfactory mucosa. *Am. J. Anat.* *121*, 87–120.
- Fuks, F. (2005). DNA methylation and histone modifications: teaming up to silence genes. *Curr Opin Genet Dev* *15*, 490–495.
- Fuks, F., Hurd, P.J., Deplus, R., and Kouzarides, T. (2003). The DNA methyltransferases associate with HP1 and the SUV39H1 histone methyltransferase. *Nucleic Acids Research* *31*, 2305–2312.
- Ge, Y.-Z., Pu, M.-T., Gowher, H., Wu, H.-P., Ding, J.-P., Jeltsch, A., and Xu, G.-L. (2004). Chromatin targeting of de novo DNA methyltransferases by the PWWP domain. *J Biol Chem* *279*, 25447–25454.
- Glass, J.L., Fazzari, M.J., Ferguson-Smith, A.C., and Grealley, J.M. (2009). CG dinucleotide periodicities recognized by the Dnmt3a-Dnmt3L complex are distinctive at retroelements and imprinted domains. *Mamm. Genome* *20*, 633–643.
- Globisch, D.D., Münzel, M.M., Müller, M.M., Michalakakis, S.S., Wagner, M.M., Koch, S.S., Brückl, T.T., Biel, M.M., and Carell, T.T. (2010). Tissue distribution of 5-

- hydroxymethylcytosine and search for active demethylation intermediates. *PLoS ONE* 5, e15367–e15367.
- Gokoffski, K.K., Wu, H.-H., Beites, C.L., Kim, J., Kim, E.J., Matzuk, M.M., Johnson, J.E., Lander, A.D., and Calof, A.L. (2011). Activin and GDF11 collaborate in feedback control of neuroepithelial stem cell proliferation and fate. *Development* 138, 4131–4142.
- Gong, S.S., Zheng, C.C., Doughty, M.L.M., Losos, K.K., Didkovsky, N.N., Schambra, U.B.U., Nowak, N.J.N., Joyner, A.A., Leblanc, G.G., Hatten, M.E.M., et al. (2003). A gene expression atlas of the central nervous system based on bacterial artificial chromosomes. *Nature* 425, 917–925.
- Gopalakrishnan, S., Sullivan, B.A., Trazzi, S., Valle, Della, G., and Robertson, K.D. (2009). DNMT3B interacts with constitutive centromere protein CENP-C to modulate DNA methylation and the histone code at centromeric regions. *Human Molecular Genetics* 18, 3178–3193.
- Gordon, M.K., Mumm, J.S., Davis, R.A., Holcomb, J.D., and Calof, A.L. (1995). Dynamics of MASH1 expression in vitro and in vivo suggest a non-stem cell site of MASH1 action in the olfactory receptor neuron lineage. *Mol Cell Neurosci* 6, 363–379.
- Gowher, H. (2002). Molecular Enzymology of the Catalytic Domains of the Dnmt3a and Dnmt3b DNA Methyltransferases. *Journal of Biological Chemistry* 277, 20409–20414.
- Gowher, H., and Jeltsch, A. (2001). Enzymatic properties of recombinant Dnmt3a DNA methyltransferase from mouse: the enzyme modifies DNA in a non-processive manner and also methylates non-CpG [correction of non-CpA] sites. *J Mol Biol* 309, 1201–1208.
- Gowher, H., Leismann, O., and Jeltsch, A. (2000). DNA of *Drosophila melanogaster* contains 5-methylcytosine. *Embo J* 19, 6918–6923.
- Götz, M., and Huttner, W.B. (2005). The cell biology of neurogenesis. *Nat Rev Mol Cell Biol* 6, 777–788.
- Graziadei, P.P., and Graziadei, G.A. (1979). Neurogenesis and neuron regeneration in the olfactory system of mammals. I. Morphological aspects of differentiation and structural organization of the olfactory sensory neurons. *J Neurocytol* 8, 1–18.
- Gu, T.-P., Guo, F., Yang, H., Wu, H.-P., Xu, G.-F., Liu, W., Xie, Z.-G., Shi, L., He, X., Jin, S.-G., et al. (2011). The role of Tet3 DNA dioxygenase in epigenetic reprogramming by oocytes. *Nature* 477, 606–610.
- Guillemot, F., Lo, L.C., Johnson, J.E., Auerbach, A., Anderson, D.J., and Joyner, A.L. (1993). Mammalian achaete-scute homolog 1 is required for the early development of olfactory and autonomic neurons. *Cell* 75, 463–476.
- Guo, J.U.J., Su, Y.Y., Zhong, C.C., Ming, G.-L.G., and Song, H.H. (2011a). Hydroxylation of 5-methylcytosine by TET1 promotes active DNA demethylation in the adult brain. *Cell* 145, 423–434.
- Guo, J.U., Ma, D.K., Mo, H., Ball, M.P., Jang, M.-H., Bonaguidi, M.A., Balazer, J.A., Eaves, H.L., Xie, B., Ford, E., et al. (2011b). Neuronal activity modifies the DNA methylation landscape in the adult brain. *Nat Neurosci* 14, 1345–1351.

- Hahn, M.A., Qiu, R., Wu, X., Li, A.X., Zhang, H., Wang, J., Jui, J., Jin, S.-G., Jiang, Y., Pfeifer, G.P., et al. (2013). Dynamics of 5-hydroxymethylcytosine and chromatin marks in Mammalian neurogenesis. *Cell Reports* 3, 291–300.
- Hark, A.T., Schoenherr, C.J., Katz, D.J., Ingram, R.S., Levorse, J.M., and Tilghman, S.M. (2000). CTCF mediates methylation-sensitive enhancer-blocking activity at the H19/Igf2 locus. *Nature* 405, 486–489.
- Hashimoto, H., Liu, Y., Upadhyay, A.K., Chang, Y., Howerton, S.B., Vertino, P.M., Zhang, X., and Cheng, X. (2012). Recognition and potential mechanisms for replication and erasure of cytosine hydroxymethylation. *Nucleic Acids Research* 40, 4841–4849.
- He, Y.-F., Li, B.-Z., Li, Z., Liu, P., Wang, Y., Tang, Q., Ding, J., Jia, Y., Chen, Z., Li, L., et al. (2011). Tet-mediated formation of 5-carboxylcytosine and its excision by TDG in mammalian DNA. *Science* 333, 1303–1307.
- Heinz, S.S., Benner, C.C., Spann, N.N., Bertolino, E.E., Lin, Y.C.Y., Laslo, P.P., Cheng, J.X.J., Murre, C.C., Singh, H.H., and Glass, C.K.C. (2010). Simple combinations of lineage-determining transcription factors prime cis-regulatory elements required for macrophage and B cell identities. *Mol Cell* 38, 576–589.
- Hellman, A., and Chess, A. (2007). Gene body-specific methylation on the active X chromosome. *Science* 315, 1141–1143.
- Hendrich, B., and Bird, A. (1998). Identification and characterization of a family of mammalian methyl-CpG binding proteins. *Mol Cell Biol* 18, 6538–6547.
- Henikoff, S. (2008). Nucleosome destabilization in the epigenetic regulation of gene expression. *Nat Rev Genet* 9, 15–26.
- Holliday, R., and Pugh, J.E. (1975). DNA modification mechanisms and gene activity during development. *Science* 187, 226–232.
- Holz-Schietinger, C., and Reich, N.O. (2010). The Inherent Processivity of the Human de Novo Methyltransferase 3A (DNMT3A) Is Enhanced by DNMT3L. *J Biol Chem* 285, 29091–29100.
- Hotchkiss, R.D. (1948). The quantitative separation of purines, pyrimidines, and nucleosides by paper chromatography. *J Biol Chem* 175, 315–332.
- Huang, Y., Pastor, W.A., Shen, Y., Tahiliani, M., Liu, D.R., and Rao, A. (2010). The Behaviour of 5-Hydroxymethylcytosine in Bisulfite Sequencing. *PLoS ONE* 5, e8888.
- Huang, Y., Pastor, W.A., Zepeda-Martínez, J.A., and Rao, A. (2012). The anti-CMS technique for genome-wide mapping of 5-hydroxymethylcytosine. *Nat Protoc* 7, 1897–1908.
- Huard, J.M.J., and Schwob, J.E.J. (1995). Cell cycle of globose basal cells in rat olfactory epithelium. *Dev Dyn* 203, 17–26.
- Huard, J.M., Youngentob, S.L., Goldstein, B.J., Luskin, M.B., and Schwob, J.E. (1998). Adult olfactory epithelium contains multipotent progenitors that give rise to neurons and non-neural cells. *J Comp Neurol* 400, 469–486.
- Hurt, M.E., Thomas, D.A., Working, P.K., Monticello, T.M., and Morgan, K.T. (1988). Degeneration and regeneration of the olfactory epithelium following inhalation exposure to

methyl bromide: pathology, cell kinetics, and olfactory function. *Toxicol. Appl. Pharmacol.* *94*, 311–328.

Illingworth, R.S., Gruenewald-Schneider, U., Webb, S., Kerr, A.R.W., James, K.D., Turner, D.J., Smith, C., Harrison, D.J., Andrews, R., and Bird, A.P. (2010). Orphan CpG islands identify numerous conserved promoters in the mammalian genome. *PLoS Genet* *6*, e1001134.

Ito, S., Shen, L., Dai, Q., Wu, S.C., Collins, L.B., Swenberg, J.A., He, C., and Zhang, Y. (2011). Tet Proteins Can Convert 5-Methylcytosine to 5-Formylcytosine and 5-Carboxylcytosine. *Science* *333*, 1300–1303.

Ito, S., D'Alessio, A.C., Taranova, O.V., Hong, K., Sowers, L.C., and Zhang, Y. (2010). Role of Tet proteins in 5mC to 5hmC conversion, ES-cell self-renewal and inner cell mass specification. *Nature* *466*, 1129–1133.

Iwai, N.N., Zhou, Z.Z., Roop, D.R.D., and Behringer, R.R.R. (2008). Horizontal basal cells are multipotent progenitors in normal and injured adult olfactory epithelium. *Stem Cells* *26*, 1298–1306.

Iyer, L.M., Tahiliani, M., Rao, A., and Aravind, L. (2009). Prediction of novel families of enzymes involved in oxidative and other complex modifications of bases in nucleic acids. *Cell Cycle* *8*, 1698–1710.

Izban, M.G., and Luse, D.S. (1991). Transcription on nucleosomal templates by RNA polymerase II in vitro: inhibition of elongation with enhancement of sequence-specific pausing. *Genes Dev* *5*, 683–696.

Jaenisch, R., Schnieke, A., and Harbers, K. (1985). Treatment of mice with 5-azacytidine efficiently activates silent retroviral genomes in different tissues. *Proceedings of the ...*

Jia, D., Jurkowska, R.Z., Zhang, X., Jeltsch, A., and Cheng, X. (2007). Structure of Dnmt3a bound to Dnmt3L suggests a model for de novo DNA methylation. *Nature* *449*, 248–251.

Jin, S.-G., Wu, X., Li, A.X., and Pfeifer, G.P. (2011). Genomic mapping of 5-hydroxymethylcytosine in the human brain. *Nucleic Acids Research*.

Jones, P.L., Veenstra, G.J., Wade, P.A., Vermaak, D., Kass, S.U., Landsberger, N., Strouboulis, J., and Wolffe, A.P. (1998). Methylated DNA and MeCP2 recruit histone deacetylase to repress transcription. *Nat Genet* *19*, 187–191.

Jurkowska, R.Z., Anspach, N., Urbanke, C., Jia, D., Reinhardt, R., Nellen, W., Cheng, X., and Jeltsch, A. (2008). Formation of nucleoprotein filaments by mammalian DNA methyltransferase Dnmt3a in complex with regulator Dnmt3L. *Nucleic Acids Research* *36*, 6656–6663.

Kafri, T., Ariel, M., Brandeis, M., Shemer, R., Urven, L., McCarrey, J., Cedar, H., and Razin, A. (1992). Developmental pattern of gene-specific DNA methylation in the mouse embryo and germ line. *Genes Dev* *6*, 705–714.

Kangaspeska, S., Stride, B., Métivier, R., Polycarpou-Schwarz, M., Ibberson, D., Carmouche, R.P., Benes, V., Gannon, F., and Reid, G. (2008). Transient cyclical methylation of promoter DNA. *Nature* *452*, 112–115.

Katz, Y., Wang, E.T., Airolidi, E.M., and Burge, C.B. (2010). Analysis and design of RNA sequencing experiments for identifying isoform regulation. *Nat Meth* *7*, 1009–1015.

- Kim, D., Pertea, G., Trapnell, C., Pimentel, H., Kelley, R., and Salzberg, S.L. (2013). TopHat2: accurate alignment of transcriptomes in the presence of insertions, deletions and gene fusions. *Genome Biol* 14, R36–R36.
- Kim, S.-H., Kang, Y.-K., Koo, D.-B., Kang, M.-J., Moon, S.-J., Lee, K.-K., and Han, Y.-M. (2004). Differential DNA methylation reprogramming of various repetitive sequences in mouse preimplantation embryos. *Biochem. Biophys. Res. Commun.* 324, 58–63.
- Kim, T.-K., Hemberg, M., Gray, J.M., Costa, A.M., Bear, D.M., Wu, J., Harmin, D.A., Laptewicz, M., Barbara-Haley, K., Kuersten, S., et al. (2010). Widespread transcription at neuronal activity-regulated enhancers. *Nature* 465, 182–187.
- Ko, M., An, J., Bandukwala, H.S., Chavez, L., Aijö, T., Pastor, W.A., Segal, M.F., Li, H., Koh, K.P., Lähdesmäki, H., et al. (2013). Modulation of TET2 expression and 5-methylcytosine oxidation by the CXXC domain protein IDAX. *Nature* 497, 122–126.
- Kriaucionis, S., and Heintz, N. (2009a). The Nuclear DNA Base 5-Hydroxymethylcytosine Is Present in Purkinje Neurons and the Brain. *Science*.
- Kriaucionis, S., and Heintz, N. (2009b). The nuclear DNA base 5-hydroxymethylcytosine is present in Purkinje neurons and the brain. *Science* 324, 929–930.
- Kuhn, R.M., Karolchik, D., Zweig, A.S., Trumbower, H., Thomas, D.J., Thakkapallayil, A., Sugnet, C.W., Stanke, M., Smith, K.E., Siepel, A., et al. (2007). The UCSC genome browser database: update 2007. *Nucleic Acids Research* 35, D668–D673.
- la Mata, de, M., Alonso, C.R., Kadener, S., Fededa, J.P., Blaustein, M., Pelisch, F., Cramer, P., Bentley, D., and Kornblihtt, A.R. (2003). A slow RNA polymerase II affects alternative splicing in vivo. *Mol Cell* 12, 525–532.
- Laird, P.W. (2010). Principles and challenges of genomewide DNA methylation analysis. *Nat Rev Genet* 11, 191–203.
- Lander, E.S., Linton, L.M., Birren, B., Nusbaum, C., Zody, M.C., Baldwin, J., Devon, K., Dewar, K., Doyle, M., FitzHugh, W., et al. (2001). Initial sequencing and analysis of the human genome. *Nature* 409, 860–921.
- LaPlant, Q., Vialou, V., Covington, H.E., Dumitriu, D., Feng, J., Warren, B.L., Maze, I., Dietz, D.M., Watts, E.L., Iñiguez, S.D., et al. (2010). Dnmt3a regulates emotional behavior and spine plasticity in the nucleus accumbens. *Nat Neurosci* 13, 1137–1143.
- Laurent, L., Wong, E., Li, G., Huynh, T., Tsirigos, A., Ong, C.T., Low, H.M., Kin Sung, K.W., Rigoutsos, I., Loring, J., et al. (2010). Dynamic changes in the human methylome during differentiation. *Genome Res* 20, 320–331.
- Lee, A.C., He, J., and Ma, M. (2011). Olfactory Marker Protein Is Critical for Functional Maturation of Olfactory Sensory Neurons and Development of Mother Preference. *J Neurosci* 31, 2974–2982.
- Lee, J.E. (1997). Basic helix-loop-helix genes in neural development. *Curr Opin Neurobiol* 7, 13–20.
- Lee, J.H., Voo, K.S., and Skalnik, D.G. (2001). Identification and characterization of the DNA binding domain of CpG-binding protein. *J Biol Chem* 276, 44669–44676.

- Leonhardt, H., Page, A.W., Weier, H.U., and Bestor, T.H. (1992). A targeting sequence directs DNA methyltransferase to sites of DNA replication in mammalian nuclei. *Cell* *71*, 865–873.
- Leung, C.T., Coulombe, P.A., and Reed, R.R. (2007). Contribution of olfactory neural stem cells to tissue maintenance and regeneration. *Nat Neurosci* *10*, 720–726.
- Lewis, J.D., Meehan, R.R., Henzel, W.J., Maurer-Fogy, I., Jeppesen, P., Klein, F., and Bird, A. (1992). Purification, sequence, and cellular localization of a novel chromosomal protein that binds to methylated DNA. *Cell* *69*, 905–914.
- Li, E., Bestor, T.H., and Jaenisch, R. (1992). Targeted mutation of the DNA methyltransferase gene results in embryonic lethality. *Cell* *69*, 915–926.
- Liao, Y., Smyth, G.K., and Shi, W. (2013). featureCounts: an efficient general-purpose read summarization program. *arXiv* *1305*, 3347.
- Lister, R., Mukamel, E.A., Nery, J.R., Urich, M., Puddifoot, C.A., Johnson, N.D., Lucero, J., Huang, Y., Dwork, A.J., Schultz, M.D., et al. (2013). Global Epigenomic Reconfiguration During Mammalian Brain Development. *Science* *341*, 1237905–1237905.
- Lister, R., O'Malley, R.C., Tonti-Filippini, J., Gregory, B.D., Berry, C.C., Millar, A.H., and Ecker, J.R. (2008). Highly integrated single-base resolution maps of the epigenome in Arabidopsis. *Cell* *133*, 523–536.
- Lister, R., Pelizzola, M., Dowen, R.H., Hawkins, R.D., Hon, G., Tonti-Filippini, J., Nery, J.R., Lee, L., Ye, Z., Ngo, Q.-M., et al. (2009). Human DNA methylomes at base resolution show widespread epigenomic differences. *Nature* *462*, 315–322.
- Loenarz, C., and Schofield, C.J. (2011). Physiological and biochemical aspects of hydroxylations and demethylations catalyzed by human 2-oxoglutarate oxygenases. *Trends Biochem Sci* *36*, 7–18.
- Lorsbach, R.B., Moore, J., Mathew, S., Raimondi, S.C., Mukatira, S.T., and Downing, J.R. (2003). TET1, a member of a novel protein family, is fused to MLL in acute myeloid leukemia containing the t(10;11)(q22;q23). *Leukemia* *17*, 637–641.
- Lyko, F., Ramsahoye, B.H., and Jaenisch, R. (2000). DNA methylation in *Drosophila melanogaster*. *Nature* *408*, 538–540.
- Lyons, D.B., Allen, W.E., Goh, T., Tsai, L., Barnea, G., and Lomvardas, S. (2013). An Epigenetic Trap Stabilizes Singular Olfactory Receptor Expression. *Cell* *154*, 325–336.
- MacDonald, J.L., Gin, C.S.Y., and Roskams, A.J. (2005). Stage-specific induction of DNA methyltransferases in olfactory receptor neuron development. *Developmental Biology* *288*, 461–473.
- MacDonald, J.L., Verster, A., Berndt, A., and Roskams, A.J. (2010). MBD2 and MeCP2 regulate distinct transitions in the stage-specific differentiation of olfactory receptor neurons. *Mol Cell Neurosci* *44*, 55–67.
- Mackay-Sim, A., and Kittel, P. (1991). Cell dynamics in the adult mouse olfactory epithelium: a quantitative autoradiographic study. *J Neurosci* *11*, 979–984.

- Magklara, A., Yen, A., Colquitt, B.M., Clowney, E.J., Allen, W., Markenscoff-Papadimitriou, E., Evans, Z.A., Kheradpour, P., Mountoufaris, G., Carey, C., et al. (2011). An epigenetic signature for monoallelic olfactory receptor expression. *Cell* *145*, 555–570.
- Maiti, A., and Drohat, A.C. (2011). Thymine DNA Glycosylase Can Rapidly Excise 5-Formylcytosine and 5-Carboxylcytosine: POTENTIAL IMPLICATIONS FOR ACTIVE DEMETHYLATION OF CpG SITES. *J Biol Chem* *286*, 35334–35338.
- Manglapus, G.L., Youngentob, S.L., and Schwob, J.E. (2004). Expression patterns of basic helix-loop-helix transcription factors define subsets of olfactory progenitor cells. *J Comp Neurol* *479*, 216–233.
- Margolis, F.L. (1972). A brain protein unique to the olfactory bulb. *Proc Natl Acad Sci USA* *69*, 1221–1224.
- Martienssen, R.A. (2001). DNA Methylation and Epigenetic Inheritance in Plants and Filamentous Fungi. *Science* *293*, 1070–1074.
- Martinowich, K., Hattori, D., Wu, H., Fouse, S., He, F., Hu, Y., Fan, G., and Sun, Y.E. (2003). DNA methylation-related chromatin remodeling in activity-dependent BDNF gene regulation. *Science* *302*, 890–893.
- Maunakea, A.K., Nagarajan, R.P., Bilenky, M., Ballinger, T.J., D'Souza, C., Fouse, S.D., Johnson, B.E., Hong, C., Nielsen, C., Zhao, Y., et al. (2010). Conserved role of intragenic DNA methylation in regulating alternative promoters. *Nature* *466*, 253–257.
- McLean, C.Y., Bristor, D., Hiller, M., Clarke, S.L., Schaar, B.T., Lowe, C.B., Wenger, A.M., and Bejerano, G. (2010). GREAT improves functional interpretation of cis-regulatory regions. *Nat Biotechnol* *28*, 495–501.
- Meehan, R.R., Lewis, J.D., McKay, S., Kleiner, E.L., and Bird, A.P. (1989). Identification of a mammalian protein that binds specifically to DNA containing methylated CpGs. *Cell* *58*, 499–507.
- Meissner, A., Gnirke, A., Bell, G.W., Ramsahoye, B., Lander, E.S., and Jaenisch, R. (2005). Reduced representation bisulfite sequencing for comparative high-resolution DNA methylation analysis. *Nucleic Acids Research* *33*, 5868–5877.
- Mellén, M., Ayata, P., Dewell, S., Kriaucionis, S., and Heintz, N. (2012). MeCP2 Binds to 5hmC Enriched within Active Genes and Accessible Chromatin in the Nervous System. *Neuron* *151*, 1417–1430.
- Métivier, R., Gallais, R., Tiffocche, C., Le Péron, C., Jurkowska, R.Z., Carmouche, R.P., Ibberson, D., Barath, P., Demay, F., Reid, G., et al. (2008). Cyclical DNA methylation of a transcriptionally active promoter. *Nature* *452*, 45–50.
- Mohandas, T., Sparkes, R.S., and Shapiro, L.J. (1981). Reactivation of an inactive human X chromosome: evidence for X inactivation by DNA methylation. *Science* *211*, 393–396.
- Mombaerts, P. (2004). Genes and ligands for odorant, vomeronasal and taste receptors. *Nat Neurosci* *5*, 263–278.

- Monk, M., Boubelik, M., and Lehnert, S. (1987). Temporal and regional changes in DNA methylation in the embryonic, extraembryonic and germ cell lineages during mouse embryo development. *Development* *99*, 371–382.
- Moulton, D.G. (1974). DYNAMICS OF CELL POPULATIONS IN THE OLFACTORY EPITHELIUM. *Ann NY Acad Sci* *237*, 52–61.
- Murray, R.C., Navi, D., Fesenko, J., Lander, A.D., and Calof, A.L. (2003). Widespread defects in the primary olfactory pathway caused by loss of *Mash1* function. *J Neurosci* *23*, 1769–1780.
- Münzel, M., Globisch, D., Brückl, T., Wagner, M., Welzmler, V., Michalakis, S., Müller, M., Biel, M., and Carell, T. (2010). Quantification of the Sixth DNA Base Hydroxymethylcytosine in the Brain. *Angewandte Chemie International Edition* *49*, 5375–5377.
- Nan, X., Meehan, R.R., and Bird, A. (1993). Dissection of the methyl-CpG binding domain from the chromosomal protein MeCP2. *Nucleic Acids Research* *21*, 4886–4892.
- Nan, X., Ng, H.-H., Johnson, C.A., Laherty, C.D., Turner, B.M., Eisenman, R.N., and Bird, A. (1998). Transcriptional repression by the methyl-CpG-binding protein MeCP2 involves a histone deacetylase complex. *Nature* *393*, 386–389.
- Ng, H.H., Zhang, Y., Hendrich, B., Johnson, C.A., Turner, B.M., Erdjument-Bromage, H., Tempst, P., Reinberg, D., and Bird, A.P. (1999). MBD2 is a transcriptional repressor belonging to the MeCP1 histone deacetylase complex. *Nat Genet* *23*, 58–61.
- Nguyen, S., Meletis, K., Fu, D., Jhaveri, S., and Jaenisch, R. (2007). Ablation of de novo DNA methyltransferase *Dnmt3a* in the nervous system leads to neuromuscular defects and shortened lifespan. *Dev Dyn* *236*, 1663–1676.
- Nicolay, D.J., Doucette, J.R., and Nazarali, A.J. (2006). Transcriptional regulation of neurogenesis in the olfactory epithelium. *Cell Mol Neurobiol* *26*, 803–821.
- Okada, Y., Yamagata, K., Hong, K., Wakayama, T., and Zhang, Y. (2010). A role for the elongator complex in zygotic paternal genome demethylation. *Nature* *463*, 554–558.
- Okano, M., Bell, D.W., Haber, D.A., and Li, E. (1999). DNA methyltransferases *Dnmt3a* and *Dnmt3b* are essential for de novo methylation and mammalian development. *Cell* *99*, 247–257.
- Okano, M., Xie, S., and Li, E. (1998). Cloning and characterization of a family of novel mammalian DNA (cytosine-5) methyltransferases. *Nat Genet* *19*, 219–220.
- Olek, A., and Walter, J. (1997). The pre-implantation ontogeny of the H19 methylation imprint. *Nat Genet* *17*, 275–276.
- Ono, R., Taki, T., Taketani, T., Taniwaki, M., Kobayashi, H., and Hayashi, Y. (2002). LCX, leukemia-associated protein with a CXXC domain, is fused to MLL in acute myeloid leukemia with trilineage dysplasia having t(10;11)(q22;q23). *Cancer Res.* *62*, 4075–4080.
- Ooi, S.K.T., Qiu, C., Bernstein, E., Li, K., Jia, D., Yang, Z., Erdjument-Bromage, H., Tempst, P., Lin, S.-P., Allis, C.D., et al. (2007). DNMT3L connects unmethylated lysine 4 of histone H3 to de novo methylation of DNA. *Nature* *448*, 714–717.
- Otani, J., Nankumo, T., Arita, K., Inamoto, S., Ariyoshi, M., and Shirakawa, M. (2009). Structural basis for recognition of H3K4 methylation status by the DNA methyltransferase 3A ATRX-DNMT3-DNMT3L domain. *EMBO Rep* *10*, 1235–1241.

- Pastor, W.A., Aravind, L., and Rao, A. (2013). TETonic shift: biological roles of TET proteins in DNA demethylation and transcription. *Nat Rev Mol Cell Biol* 14, 341–356.
- Pastor, W.A., Pape, U.J., Huang, Y., Henderson, H.R., Lister, R., Ko, M., McLoughlin, E.M., Brudno, Y., Mahapatra, S., Kapranov, P., et al. (2011). Genome-wide mapping of 5-hydroxymethylcytosine in embryonic stem cells. *Nature* 473, 394–397.
- Pósfai, J.J., Bhagwat, A.S.A., Pósfai, G.G., and Roberts, R.J.R. (1989). Predictive motifs derived from cytosine methyltransferases. *Nucleic Acids Research* 17, 2421–2435.
- Pradhan, S., Bacolla, A., Wells, R.D., and Roberts, R.J. (1999). Recombinant human DNA (cytosine-5) methyltransferase. I. Expression, purification, and comparison of de novo and maintenance methylation. *J Biol Chem* 274, 33002–33010.
- Prokhortchouk, A., Hendrich, B., Jørgensen, H., Ruzov, A., Wilm, M., Georgiev, G., Bird, A., and Prokhortchouk, E. (2001). The p120 catenin partner Kaiso is a DNA methylation-dependent transcriptional repressor. *Genes Dev* 15, 1613–1618.
- Rai, K., Huggins, I.J., James, S.R., Karpf, A.R., Jones, D.A., and Cairns, B.R. (2008). DNA demethylation in zebrafish involves the coupling of a deaminase, a glycosylase, and gadd45. *Cell* 135, 1201–1212.
- Ramón-Cueto, A., and Avila, J. (1998). Olfactory ensheathing glia: properties and function. *Brain Res. Bull.* 46, 175–187.
- Ramsahoye, B.H., Biniszkiwicz, D., Lyko, F., Clark, V., Bird, A.P., and Jaenisch, R. (2000). Non-CpG methylation is prevalent in embryonic stem cells and may be mediated by DNA methyltransferase 3a. *Proc Natl Acad Sci USA* 97, 5237–5242.
- Ratnam, S., Mertineit, C., Ding, F., Howell, C.Y., Clarke, H.J., Bestor, T.H., Chaillet, J.R., and Trasler, J.M. (2002). Dynamics of Dnmt1 Methyltransferase Expression and Intracellular Localization during Oogenesis and Preimplantation Development. *Developmental Biology* 245, 304–314.
- Reik, W., Dean, W., and Walter, J. (2001). Epigenetic reprogramming in mammalian development. *Science* 293, 1089–1093.
- Revel, H.R., and Lubia, S.E. (1970). DNA-glucosylation in T-even phage: genetic determination and role in phage-host interaction. *Annual Review of Genetics* 4, 177–192.
- Riggs, A.D. (1975). X inactivation, differentiation, and DNA methylation. *Cytogenet Cell Genet* 14, 9–25.
- Robinson, M.D., McCarthy, D.J., and Smyth, G.K. (2010). edgeR: a Bioconductor package for differential expression analysis of digital gene expression data. *Bioinformatics* 26, 139–140.
- Rubin, B.D., and Katz, L.C. (1999). Optical Imaging of Odorant Representations in the Mammalian Olfactory Bulb. *Neuron* 23, 499–511.
- Russo, V.E.A., Martienssen, R.A., and Riggs, A.D. (1996). Epigenetic mechanisms of gene regulation. (Cold Spring Harbor Laboratory Press).
- Ruzov, A., Tsenkina, Y., Serio, A., Dudnakova, T., Fletcher, J., Bai, Y., Chebotareva, T., Pells, S., Hannoun, Z., Sullivan, G., et al. (2011). Lineage-specific distribution of high levels of genomic 5-hydroxymethylcytosine in mammalian development. *Cell Research* 21, 1332–1342.

- Salzberg, A., Fisher, O., Siman-Tov, R., and Ankri, S. (2004). Identification of methylated sequences in genomic DNA of adult *Drosophila melanogaster*. *Biochem. Biophys. Res. Commun.* 322, 465–469.
- Schwob, J.E. (2002). Neural regeneration and the peripheral olfactory system. *Anat. Rec.* 269, 33–49.
- Schwob, J.E., Youngentob, S.L., and Mezza, R.C. (1995). Reconstitution of the rat olfactory epithelium after methyl bromide-induced lesion. *J Comp Neurol* 359, 15–37.
- Sérandour, A.A., Avner, S., Oger, F., Bizot, M., Percevault, F., Lucchetti-Miganeh, C., Paliarne, G., Gheeraert, C., Barloy-Hubler, F., Le Péron, C., et al. (2012). Dynamic hydroxymethylation of deoxyribonucleic acid marks differentiation-associated enhancers. *Nucleic Acids*
- Sharif, J., Muto, M., Takebayashi, S.-I., Suetake, I., Iwamatsu, A., Endo, T.A., Shinga, J., Mizutani-Koseki, Y., Toyoda, T., Okamura, K., et al. (2007). The SRA protein Np95 mediates epigenetic inheritance by recruiting Dnmt1 to methylated DNA. *Nature* 450, 908–912.
- Shukla, S., Kavak, E., Gregory, M., Imashimizu, M., Shutinoski, B., Kashlev, M., Oberdoerffer, P., Sandberg, R., and Oberdoerffer, S. (2011). CTCF-promoted RNA polymerase II pausing links DNA methylation to splicing. *Nature* 479, 74–79.
- Shykind, B.M., Rohani, S.C., O'Donnell, S., Nemes, A., Mendelsohn, M., Sun, Y., Axel, R., and Barnea, G. (2004). Gene switching and the stability of odorant receptor gene choice. *Cell* 117, 801–815.
- Simmen, M.W., Leitgeb, S., Charlton, J., Jones, S.J., Harris, B.R., Clark, V.H., and Bird, A. (1999). Nonmethylated transposable elements and methylated genes in a chordate genome. *Science* 283, 1164–1167.
- Skene, P.J., Illingworth, R.S., Webb, S., Kerr, A.R.W., James, K.D., Turner, D.J., Andrews, R., and Bird, A.P. (2010). Neuronal MeCP2 Is Expressed at Near Histone-Octamer Levels and Globally Alters the Chromatin State. *Mol Cell* 37, 457–468.
- Smiley, J.A., Kundracik, M., Landfried, D.A., Barnes, V.R., and Axhemi, A.A. (2005). Genes of the thymidine salvage pathway: thymine-7-hydroxylase from a *Rhodotorula glutinis* cDNA library and iso-orotate decarboxylase from *Neurospora crassa*. *Biochim Biophys Acta* 1723, 256–264.
- Smith, Z.D., and Meissner, A. (2013). DNA methylation: roles in mammalian development. *Nat Rev Genet* 14, 204–220.
- Song, C.-X., Szulwach, K.E., Fu, Y., Dai, Q., Yi, C., Li, X., Li, Y., Chen, C.-H., Zhang, W., Jian, X., et al. (2011). Selective chemical labeling reveals the genome-wide distribution of 5-hydroxymethylcytosine. *Nat Biotechnol* 29, 68–72.
- Spruijt, C.G., Gnerlich, F., Smits, A.H., Pfaffeneder, T., Jansen, P.W.T.C., Bauer, C., Münzel, M., Wagner, M., Müller, M., Khan, F., et al. (2013). Dynamic Readers for 5-(Hydroxy)Methylcytosine and Its Oxidized Derivatives. *Cell* 152, 1146–1159.
- Stein, R., Gruenbaum, Y., Pollack, Y., Razin, A., and Cedar, H. (1982). Clonal inheritance of the pattern of DNA methylation in mouse cells. *Proc Natl Acad Sci USA* 79, 61–65.

- Stöger, R., Kajimura, T.M., Brown, W.T., and Laird, C.D. (1997). Epigenetic variation illustrated by DNA methylation patterns of the fragile-X gene FMR1. *Human Molecular Genetics* 6, 1791–1801.
- Stroud, H., Feng, S., Morey Kinney, S., Pradhan, S., and Jacobsen, S.E. (2011). 5-Hydroxymethylcytosine is associated with enhancers and gene bodies in human embryonic stem cells. *Genome Biol* 12, R54.
- Studitsky, V.M., Clark, D.J., and Felsenfeld, G. (1995). Overcoming a nucleosomal barrier to transcription. *Cell* 83, 19–27.
- Studitsky, V.M., Kassavetis, G.A., Geiduschek, E.P., and Felsenfeld, G. (1997). Mechanism of transcription through the nucleosome by eukaryotic RNA polymerase. *Science* 278, 1960–1963.
- Suzuki, M.M., and Bird, A.P. (2008). DNA methylation landscapes: provocative insights from epigenomics. *Nat Rev Genet* 9, 465–476.
- Swartz, M.N., Trautner, T.A., and Kornberg, A. (1962). Enzymatic synthesis of deoxyribonucleic acid. XI. Further studies on nearest neighbor base sequences in deoxyribonucleic acids. *J Biol Chem* 237, 1961–1967.
- Szulwach, K.E., Li, X., Li, Y., Song, C.-X., Han, J.W., Kim, S., Namburi, S., Hermetz, K., Kim, J.J., Rudd, M.K., et al. (2011a). Integrating 5-hydroxymethylcytosine into the epigenomic landscape of human embryonic stem cells. *PLoS Genet* 7, e1002154.
- Szulwach, K.E., Li, X., Li, Y., Song, C.-X., Wu, H., Dai, Q., Irier, H., Upadhyay, A.K., Gearing, M., Levey, A.I., et al. (2011b). 5-hmC-mediated epigenetic dynamics during postnatal neurodevelopment and aging. *Nat Neurosci* 14, 1607–1616.
- Tahiliani, M., Koh, K.P., Shen, Y., Pastor, W.A., Bandukwala, H., Brudno, Y., Agarwal, S., Iyer, L.M., Liu, D.R., Aravind, L., et al. (2009). Conversion of 5-Methylcytosine to 5-Hydroxymethylcytosine in Mammalian DNA by MLL Partner TET1. *Science* 324, 930–935.
- Tan, M., Luo, H., Lee, S., Jin, F., Yang, J.S., Montellier, E., Buchou, T., Cheng, Z., Rousseaux, S., Rajagopal, N., et al. (2011). Identification of 67 Histone Marks and Histone Lysine Crotonylation as a New Type of Histone Modification. *Cell* 146, 1016–1028.
- Trapnell, C., Williams, B.A., Pertea, G., Mortazavi, A., Kwan, G., van Baren, M.J., Salzberg, S.L., Wold, B.J., and Pachter, L. (2010). Transcript assembly and quantification by RNA-Seq reveals unannotated transcripts and isoform switching during cell differentiation. *Nat Biotechnol* 28, 511–515.
- Tremblay, K.D., Duran, K.L., and Bartolomei, M.S. (1997). A 5' 2-kilobase-pair region of the imprinted mouse H19 gene exhibits exclusive paternal methylation throughout development. *Molecular and Cellular ...*
- Tudor, M., Akbarian, S., Chen, R.Z., and Jaenisch, R. (2002). Transcriptional profiling of a mouse model for Rett syndrome reveals subtle transcriptional changes in the brain. *Proc Natl Acad Sci USA* 99, 15536–15541.
- Tweedie, S., Charlton, J., Clark, V., and Bird, A. (1997). Methylation of genomes and genes at the invertebrate-vertebrate boundary. *Mol Cell Biol* 17, 1469–1475.

- Tweedie, S., Ng, H.H., Barlow, A.L., Turner, B.M., Hendrich, B., and Bird, A. (1999). Vestiges of a DNA methylation system in *Drosophila melanogaster*? *Nat Genet* 23, 389–390.
- Valinluck, V., and Sowers, L.C. (2007). Endogenous cytosine damage products alter the site selectivity of human DNA maintenance methyltransferase DNMT1. *Cancer Res.* 67, 946–950.
- Valinluck, V., Tsai, H.-H., Rogstad, D.K., Burdzy, A., Bird, A., and Sowers, L.C. (2004). Oxidative damage to methyl-CpG sequences inhibits the binding of the methyl-CpG binding domain (MBD) of methyl-CpG binding protein 2 (MeCP2). *Nucleic Acids Research* 32, 4100–4108.
- Varley, K.E., Gertz, J., Bowling, K.M., Parker, S.L., Reddy, T.E., Pauli-Behn, F., Cross, M.K., Williams, B.A., Stamatoyannopoulos, J.A., Crawford, G.E., et al. (2013). Dynamic DNA methylation across diverse human cell lines and tissues. *Genome Res* 23, 555–567.
- Venolia, L., Gartler, S.M., Wassman, E.R., Yen, P., Mohandas, T., and Shapiro, L.J. (1982). Transformation with DNA from 5-azacytidine-reactivated X chromosomes. *Proc Natl Acad Sci USA* 79, 2352–2354.
- Visel, A., Minovitsky, S., Dubchak, I., and Pennacchio, L.A. (2007). VISTA Enhancer Browser—a database of tissue-specific human enhancers. *Nucleic Acids Research* 35, D88–D92.
- Waddington, C.H. (1957). *The strategy of the genes. A discussion of some aspects of theoretical biology. With an appendix by H. Kacser. The Strategy of the Genes a Discussion of Some ...* ix+–262.
- Walsh, C.P., Chaillet, J.R., and Bestor, T.H. (1998). Transcription of IAP endogenous retroviruses is constrained by cytosine methylation. *Nat Genet* 20, 116–117.
- Wang, Z., Zang, C., Rosenfeld, J.A., Schones, D.E., Barski, A., Cuddapah, S., Cui, K., Roh, T.-Y., Peng, W., Zhang, M.Q., et al. (2008). Combinatorial patterns of histone acetylations and methylations in the human genome. *Nat Genet* 40, 897–903.
- Watanabe, D.D., Uchiyama, K.K., and Hanaoka, K.K. (2006). Transition of mouse de novo methyltransferases expression from Dnmt3b to Dnmt3a during neural progenitor cell development. *Neuroscience* 142, 11–11.
- Watt, F., and Molloy, P.L. (1988). Cytosine methylation prevents binding to DNA of a HeLa cell transcription factor required for optimal expression of the adenovirus major late promoter. *Genes Dev* 2, 1136–1143.
- Weber, M., Davies, J.J., Wittig, D., Oakeley, E.J., Haase, M., Lam, W.L., and Schübeler, D. (2005). Chromosome-wide and promoter-specific analyses identify sites of differential DNA methylation in normal and transformed human cells. *Nat Genet* 37, 853–862.
- Wigler, M., Levy, D., and Perucho, M. (1981). The somatic replication of DNA methylation. *Cell* 24, 33–40.
- Williams, K., Christensen, J., Pedersen, M.T., Johansen, J.V., Cloos, P.A.C., Rappaport, J., and Helin, K. (2011). TET1 and hydroxymethylcytosine in transcription and DNA methylation fidelity. *Nature* 473, 343–348.
- Wilson, G. (1988). Type II restriction—modification systems. *Trends in Genetics* 4, 314–318.

- Wu, H., D'Alessio, A.C., Ito, S., Wang, Z., Cui, K., Zhao, K., Sun, Y.E., and Zhang, Y. (2011a). Genome-wide analysis of 5-hydroxymethylcytosine distribution reveals its dual function in transcriptional regulation in mouse embryonic stem cells. *Genes Dev* 25, 679–684.
- Wu, H.H., and Zhang, Y.Y. (2011). Mechanisms and functions of Tet protein-mediated 5-methylcytosine oxidation. *Genes Dev* 25, 2436–2452.
- Wu, H., Coskun, V., Tao, J., Xie, W., Ge, W., Yoshikawa, K., Li, E., Zhang, Y., and Sun, Y.E. (2010). Dnmt3a-dependent nonpromoter DNA methylation facilitates transcription of neurogenic genes. *Science* 329, 444–448.
- Wu, H., D'Alessio, A.C., Ito, S., Xia, K., Wang, Z., Cui, K., Zhao, K., Sun, Y.E., and Zhang, Y. (2011b). Dual functions of Tet1 in transcriptional regulation in mouse embryonic stem cells. *Nature* 473, 389–393.
- Xie, W., Barr, C.L., Kim, A., Yue, F., Lee, A.Y., Eubanks, J., Dempster, E.L., and Ren, B. (2012). Base-Resolution Analyses of Sequence and Parent-of-Origin Dependent DNA Methylation in the Mouse Genome. *Cell* 148, 816–831.
- Xu, Y., Wu, F., Tan, L., Kong, L., Xiong, L., Deng, J., Barbera, A.J., Zheng, L., Zhang, H., Huang, S., et al. (2011). Genome-wide Regulation of 5hmC, 5mC, and Gene Expression by Tet1 Hydroxylase in Mouse Embryonic Stem Cells. *Mol Cell* 42, 451–464.
- Yildirim, O., Li, R., Hung, J.-H., Chen, P.B., Dong, X., Ee, L.-S., Weng, Z., Rando, O.J., and Fazio, T.G. (2011). Mbd3/NURD Complex Regulates Expression of 5-Hydroxymethylcytosine Marked Genes in Embryonic Stem Cells. *Cell* 147, 1498–1510.
- Yokomine, T., Hata, K., Tsudzuki, M., and Sasaki, H. (2006). Evolution of the vertebrate DNMT3 gene family: a possible link between existence of DNMT3L and genomic imprinting. *Cytogenet Cell Genet* 113, 75–80.
- Yu, C.R., Power, J., Barnea, G., O'Donnell, S., Brown, H.E.V., Osborne, J., Axel, R., and Gogos, J.A. (2004). Spontaneous neural activity is required for the establishment and maintenance of the olfactory sensory map. *Neuron* 42, 553–566.
- Yu, M., Hon, G.C., Szulwach, K.E., Song, C.-X., Zhang, L., Kim, A., Li, X., Dai, Q., Shen, Y., Park, B., et al. (2012). Base-Resolution Analysis of 5-Hydroxymethylcytosine in the Mammalian Genome. *Cell* 149, 1368–1380.
- Zhang, H., Zhang, X., Clark, E., Mulcahey, M., Huang, S., and Shi, Y.G. (2010). TET1 is a DNA-binding protein that modulates DNA methylation and gene transcription via hydroxylation of 5-methylcytosine. *Cell Research* 20, 1390–1393.
- Zhang, X., Yazaki, J., Sundaresan, A., Cokus, S., Chan, S.W.-L., Chen, H., Henderson, I.R., Shinn, P., Pellegrini, M., Jacobsen, S.E., et al. (2006). Genome-wide high-resolution mapping and functional analysis of DNA methylation in arabidopsis. *Cell* 126, 1189–1201.
- Zhou, Z., Hong, E.J., Cohen, S., Zhao, W.-N., Ho, H.-Y.H., Schmidt, L., Chen, W.G., Lin, Y., Savner, E., Griffith, E.C., et al. (2006). Brain-specific phosphorylation of MeCP2 regulates activity-dependent Bdnf transcription, dendritic growth, and spine maturation. *Neuron* 52, 255–269.

Zilberman, D., Gehring, M., Tran, R.K., Ballinger, T., and Henikoff, S. (2007). Genome-wide analysis of *Arabidopsis thaliana* DNA methylation uncovers an interdependence between methylation and transcription. *Nat Genet* 39, 61–69.

Publishing Agreement

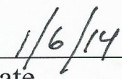
It is the policy of the University to encourage the distribution of all theses, dissertations, and manuscripts. Copies of all UCSF theses, dissertations, and manuscripts will be routed to the library via the Graduate Division. The library will make all theses, dissertations, and manuscripts accessible to the public and will preserve these to the best of their abilities, in perpetuity.

Please sign the following statement:

I hereby grant permission to the Graduate Division of the University of California, San Francisco to release copies of my thesis, dissertation, or manuscript to the Campus Library to provide access and preservation, in whole or in part, in perpetuity.



Author Signature



Date