Genome **Biology**

RESEARCH HIGHLIGHT

Genome-wide mapping of the sixth base

Dinh Diep[1,2] and Kun Zhang*[1,2]

See Research article: http://genomebiology.com/2011/12/6/R54

Abstract

Mapping of 5-hydroxylmethylcytosine in mammalian genomes has unveiled its unique role in the epigenetic regulation of gene expression.

In over half a century since its initial discovery in bacteriophage DNA, 5-hydroxylmethylcytosine (5hmC) has been mostly a rare and uncharacterized phenomenon. Starting in 2009, a series of studies discovered the presence of 5hmCs in Purkinje neurons, adult brain, mouse cerebellum, and also in mouse and human embryonic stem cells [1,2]. The 5hmC base was also identified as an intermediate of active DNA demethylation by way of hydrolyzing 5-methylcytosine (5mC) by members of the ten-eleven translocation (Tet) enzyme family. These discoveries have motivated a multitude of recent studies focusing on resolving the mystery of how 5hmC is dynamically created and removed during stem cell development and differentiation, characterizing the genome-wide localization of 5hmC as a new epigenetic marker, and determining the roles played by 5hmC in epigenetic regulation.

Genome-wide maps of 5hmCs in embryonic stem cells

A report in this issue of *Genome Biology* by Stroud and colleagues [3] represents one of the recent efforts in mapping the localization of 5hmCs throughout the genome in embryonic stem cells and related cell types. In the work by Shroud *et al.* on the mapping of 5hmCs in human embryonic stem cells (hESCs), genomic fragments containing 5hmCs were enriched by immuno-precipitation with 5hmC-specific antibodies, followed by Illumina massively parallel sequencing and quantification of enrichment based on read depth. To eliminate artifacts

*Correspondence: kzhang@bioeng.ucsd.edu
[1]Department of Bioengineering, University of California at San Diego, 9500 Gilman Drive, MC0412, La Jolla, CA 92093-0412, USA
Full list of author information is available at the end of the article

BioMed Central

due to non-specific antibodies, two data sets were generated with different commercial antibodies, and the consistency seemed to be high.

With this map, Shroud and colleagues showed that 5hmCs tend to associate with genic regions, including both promoters and gene bodies (particularly exons). In intergenic regions, 5hmCs co-localize with enhancers marked by the activating histone modifications H3K4me1 and H3K27ac. Importantly, enhancers enriched for 5hmCs appear to associate strongly with hESC-specific genes, suggesting a role for 5hmCs in gene regulation through enhancers. In addition to enhancers, other DNA-protein interaction regions, in particular transcription factor (such as NANOG and OCT4) binding sites, have also been found to be enriched for 5hmCs. This suggests a potential secondary regulatory mechanism by 5hmCs through the blocking of DNMT1, a methyl-transferase that generates 5mC, and MeCP2, a transcriptional repressor that binds to methylated promoters. Such a mechanism would ensure that no 5mC is present to prevent the binding of enhancers or transcription factors. Finally, Shroud *et al.* reported an interesting observation of GC skewness in 5hmC-enriched regions, wherein G residues are enriched over C residues from the 5' ends of the regions, and C residues are enriched over G residues from the 3' ends, although the functional roles of such GC skewness remain elusive.

Four other recent studies on mouse embryonic stem cells (mESCs) have revealed a very similar distribution of 5hmCs [4-7]. Notably, Pastor *et al.* [4] developed two novel methods for the genome-wide mapping of 5hmCs. One of these methods, called GLIB (glucosylation, periodate oxidation, biotinylation), uses three enzymatic and chemical reactions to label 5hmCs with biotins, followed by pull-down with streptavidin-coated magnetic beads, and direct single-molecule sequencing on a HeliScope. Compared with affinity pull-down by anti-bodies, GLIB seems to have lower background noise and no bias towards CpG-dense regions. Single-molecule sequencing also eliminated any potential artifact due to bias in PCR amplification. The second novel method is similar to the methods used by the other three groups [3,5,6] in that 5hmC-containing DNA fragments are enriched by antibodies, coupled with massive parallel

sequencing. However, one unique aspect of the second method is that genomic DNA is first treated with bisulfite to convert 5hmC into cytosine 5-methylenesulfonate prior to immunoprecipitation. Since the sulfonate group is larger than the hydroxyl group, antibody binding could be more specific and less dependent on CpG density.

Despite the technical differences between these mapping methods, the five studies reached very similar conclusions about the distribution of 5hmCs at the genome level. They reported that 5hmC, similar to 5mC, is enriched in promoters and gene bodies in hESCs and mESCs. In addition, 5hmCs are preferentially present in promoters that have been found to be repressive toward transcription of the associated genes [3,7]. Interestingly, while 5mC is distributed mainly at the 3' end of transcriptional start sites, 5hmC is distributed symmetrically at the 5' and 3' ends of transcriptional start sites [4]. The majority of high 5hmC promoters are marked by the H3K4me3 activation mark alone, and smaller percentages of such promoters co-localize with the repressive bivalent H3K4me3 and H3K27me3 marks. However, if normalized by the total number of genes carrying these histone marks, 5hmCs are actually enriched in H3K4me3 and H3K27me3 bivalent regions [3,4,6,8]. This led to the hypothesis that 5hmC binds to genes that may be poised for transcription upon differentiation. Interestingly, genes associated with 5hmC in the gene bodies are actively transcribed in mESCs and mouse cerebellum [4,5,9]. Xu *et al.* [6] and Wu *et al.* [5] noted that the correlation of 5hmC levels with gene expression is stronger at the 3' ends of genes and also for genes associated with Tet1.

Concerted and opposing functions for 5hmC and Tet1

In the current working model for DNA demethylation, which still has many gaps to be filled, 5hmC is the first intermediate product of 5mC demethylation catalyzed by Tet proteins, including Tet1 [9]. Intensive efforts have been invested in the characterization of Tet1-binding sites, as well as the relationships between Tet1 activities and localization of 5hmCs [6-8,10]. Tet1-binding sites are enriched at hypomethylated regions in mESCs [6,8]. Additionally, while Tet1 may bind to unmodified cytosine, 5hmC or 5mC, it preferentially binds to unmodified cytosines [6]. There is substantial overlap of Tet1-binding sites and 5hmC-enriched regions, which are distributed at promoters, exons, introns and intergenic regions [2]. Tet1-binding sites are enriched at regions with the repressive bivalent H3K4me3 and H3K27me3 marks as well as univalent H3K4me3 and H3K27me3 marks [6]. Tet1 knockdown causes a decrease in 5hmC levels and an increase in 5mC levels. Genes with increased 5mC also show a reduction in expression after Tet1 depletion by short hairpin RNA (shRNA) or after embryonic stem cell differentiation [8].

Tet1 appears to have a dual role in gene regulation: the Tet1-associated genes that have low 5hmC and high 5mC are generally repressed by Tet1 binding, while the genes with high 5hmC and low 5mC are positively regulated by Tet1 [6]. It has been hypothesized that Tet1 might have additional functions in addition to hydrolyzing 5mCs. In a search of Tet1-interacting protein partners that might mediate such functions, Williams *et al.* [8] identified the SIN3A co-repressor complex using double-epitope tagging and mass spectrometry. ChIP-seq experiments revealed a significant overlap between Tet1 and SIN3A binding. Tet1 helps recruit SIN3A, but the interaction is not symmetric. Furthermore, shRNA knockdown of Sin3A seems to increase the expression of a subset of Tet1 repressed genes, indicating a new role in transcriptional repression by Tet1 [8]. Taking the results together, an emerging notion is that Tet1 regulates gene expression through at least three mechanisms, including demethylation of 5mCs, binding to 5hmCs to insulate DNA methyltransferase or methyl-CpG binding domain activities, and transcriptional repression by direct interaction and recruitment of the SIN2A co-repressor complex.

In less than 3 years since the first reports of 5hmC in mammalian cells [1,2], our understanding of the genome-wide localization of this new epigenetic marker, the mechanism of its formation and its functional roles has progressed at an incredible pace. However, these recent studies have just marked the beginning of an exciting area in the investigation of DNA demethylation. A new stage is set for further studies to answer important questions, such as the context-dependent functions of 5hmC and Tet1 binding, the roles of additional Tet proteins in 5hmC formation and transcriptional regulation, and the detailed mechanisms involved in the conversion of 5hmC to cytosine.

Author details
[1]Department of Bioengineering, University of California at San Diego, 9500 Gilman Drive, MC0412, La Jolla, CA 92093-0412, USA. [2]Bioinformatics and System Biology Graduate Program, University of California at San Diego, 9500 Gilman Drive, La Jolla, CA 92093-0412, USA.

References
1. Tahiliani M, Koh KP, Shen Y, Pastor WA, Bandukwala H, Brudno Y, Agarwal S, Iyer LM, Liu DR, Aravind L, Rao A: **Conversion of 5-methylcytosine to 5-hydroxymethylcytosine in mammalian DNA by MLL partner TET1.** *Science* 2009, **324**:930-935.
2. Kriaucionis S, Heintz N: **The nuclear DNA base 5-hydroxymethylcytosine is present in Purkinje neurons and the brain.** *Science* 2009, **324**:929-930.

3. Stroud H, Feng S, Kinney SM, Pradhan S, Jacobsen SE:
    **5-hydroxymethylcytosine is associated with enhancers and gene bodies
    in human embryonic stem cells.** *Genome Biol* 2011, 12:R54.
4. Pastor WA, Pape UJ, Huang Y, Henderson HR, Lister R, Ko M, McLoughlin EM,
    Brudno Y, Mahapatra S, Kapranov P, Tahiliani M, Daley GQ, Liu XS, Ecker JR,
    Milos PM, Agarwal S, Rao A: **Genome-wide mapping of
    5-hydroxymethylcytosine in embryonic stem cells.** *Nature* 2011,
    473:394-397.
5. Wu H, D'Alessio AC, Ito S, Wang Z, Cui K, Zhao K, Sun YE, Zhang Y: **Genome-
    wide analysis of 5-hydroxymethylcytosine distribution reveals its dual
    function in transcriptional regulation in mouse embryonic stem cells.**
    Genes Dev 2011, **25:**679-684.
6. Xu Y, Wu F, Tan L, Kong L, Xiong L, Deng J, Barbera AJ, Zheng L, Zhang H,
    Huang S, Min J, Nicholson T, Chen T, Xu G, Shi Y, Zhang K, Shi YG: **Genome-
    wide regulation of 5hmC, 5mC, and gene expression by Tet1 hydroxylase
    in mouse embryonic stem cells.** *Mol Cell* 2011, **42:**451-464.
7. Ficz G, Branco MR, Seisenberger S, Santos F, Krueger F, Hore TA, Marques CJ,
    Andrews S, Reik W: **Dynamic regulation of 5-hydroxymethylcytosine in
    mouse ES cells and during differentiation.** *Nature* 2011, **473:**398-402.
8. Williams K, Christensen J, Pedersen MT, Johansen JV, Cloos PA, Rappsilber J,
    Helin K: **TET1 and hydroxymethylcytosine in transcription and DNA
    methylation fidelity.** *Nature* 2011, **473:**343-348.
9. Guo JU, Su Y, Zhong C, Ming GL, Song H: **Hydroxylation of 5-methylcytosine
    by TET1 promotes active DNA demethylation in the adult brain.** *Cell* 2011,
    145:423-434.
10. Wu H, D'Alessio AC, Ito S, Xia K, Wang Z, Cui K, Zhao K, Eve Sun Y, Zhang Y:
    **Dual functions of Tet1 in transcriptional regulation in mouse embryonic
    stem cells.** *Nature* 2011, **473:**389-393.