Genome **Biology**

RESEARCH HIGHLIGHT

Deep C diving: mapping the low-abundance modifications of the DNA demethylation pathway

John P Thomson, Jennifer M Hunter and Richard R Meehan*****

Abstract

Two new studies imply that the reprogramming of 5-methylcytosine via TET- and TDG-family enzymes is both widespread throughout the genome and functionally significant.

In the mammalian genome, the dinucleotide CpG acts as a unique signaling module that can regulate the local chromatin environment through the recruitment of specific chromatin modifying proteins [1]. Although it is thought to be context specific, the general enzymatic acquisition of methylation at CpG dinucleotides by DNA methyltransferase enzymes (DNMTs) over promoter regions tends to be associated with gene silencing events and heterochromatin formation. The maintenance of 5-methylcytosine (5mC) modification patterns has since been implicated in many important roles in normal cell function during mammalian development and disease progression [1]. Although it is widely understood how DNA can become enzymatically methylated, less is known regarding the active removal of 5mC at specific loci, aside from the potential for passive loss during cell division in the absence of DNMT activity. In 2009, a second form of DNA modification, that of 5-hydroxymethylcytosine (5hmC), was rediscovered, and enzymatic oxidation reactions (involving the ten-eleven translocation (TET) proteins) responsible for generating 5hmC from 5mC were identified [2]. Subsequent work has since identified the downstream, TET-dependent, oxidative derivatives of 5hmC, those of 5-formylcytosine (5fC) and 5-carboxylcytosine (5caC) [2]. This has led to the proposal of an active DNA demethylation cycle relying on the initial oxidation of 5mC into 5hmC, through the TET family of enzymes, before further oxidation to the 5fC and 5caC derivatives (Figure 1a). In contrast to the more abundant

5hmC modification, these lower-abundance downstream intermediates are proposed to be removed by base excision repair mechanisms that are highly reliant on the thymine DNA glycosylase (TDG) protein, ultimately resulting in the replacement of modified cytosine with non-modified cytosine.

Mapping the patterns of 5fC and 5caC in mouse embryonic stem cells

Genome-wide patterns of 5mC and now 5hmC are becoming well characterized in a host of cell and tissue types with ever increasing complexity, ultimately driven by a host of recent technological advances [3,4]. In contrast, there is a distinct lack of understanding regarding the distributions of the 5fC and 5caC modified sequences, largely due to a lack of accurate methods to detect these low-abundance modifications; mass spectroscopy indicates that 5fC is at 2% and 5caC at 0.5% of the levels of 5hmC in mouse embryonic stem cells (mESCs), which in turn is only 4% as abundant as 5mC [5]. Two recent studies report on novel techniques for mapping both 5fC and 5caC modifications, as well as addressing the functionality of TET/TDG 5mC oxidation events that occur throughout the genome [6,7]. Through the use of highly specific antibodies raised against both 5fC and 5caC, researchers led by Yi Zhang at Harvard University are able to map the genome-wide distributions of both derivatives of 5hmC [6]. In an analogous set of experiments, Song and colleagues from the laboratory of Chuan He at the University of Chicago expand upon their already successful chemical capture techniques to enrich for 5hmC-marked DNA [7]. In short, by first modifying all endogenous 5hmC by glucosylation, they can then specifically reduce 5fC-marked cytosines to 5hmC through the addition of sodium borohydride ($NaBH_4$) and then glycosylate these sites with a modified glucose group (6-azide-glucose) to which a disulfite biotin linker is attached for subsequent enrichment. In addition, the group also adapt techniques to visualize the 5fC modification at single-base resolution (fCAB-seq), overcoming the issues of discrimination between the modified forms of cytosine that arise in traditional bisulfite-based mapping. By employing these novel techniques, both studies

*Correspondence: richard.meehan@igmm.ed.ac.uk
MRC Human Genetics Unit, Institute of Genetics and Molecular Medicine, University of Edinburgh, Crewe Road, Edinburgh EH4 2XU, UK

BioMed Central
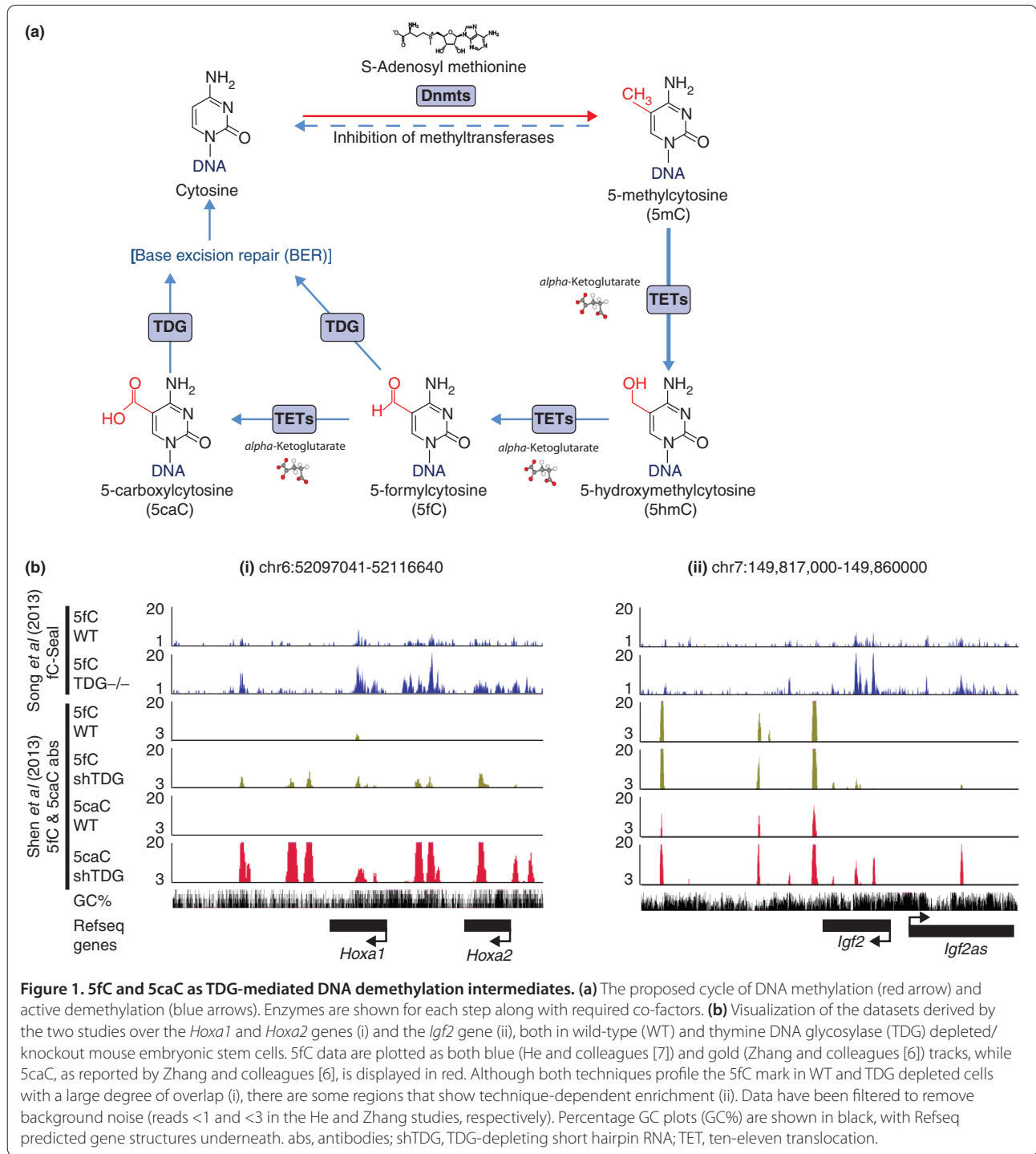
Figure 1. 5fC and 5caC as TDG-mediated DNA demethylation intermediates. (a) The proposed cycle of DNA methylation (red arrow) and active demethylation (blue arrows). Enzymes are shown for each step along with required co-factors. **(b)** Visualization of the datasets derived by the two studies over the *Hoxa1* and *Hoxa2* genes (i) and the *Igf2* gene (ii), both in wild-type (WT) and thymine DNA glycosylase (TDG) depleted/knockout mouse embryonic stem cells. 5fC data are plotted as both blue (He and colleagues [7]) and gold (Zhang and colleagues [6]) tracks, while 5caC, as reported by Zhang and colleagues [6], is displayed in red. Although both techniques profile the 5fC mark in WT and TDG depleted cells with a large degree of overlap (i), there are some regions that show technique-dependent enrichment (ii). Data have been filtered to remove background noise (reads <1 and <3 in the He and Zhang studies, respectively). Percentage GC plots (GC%) are shown in black, with Refseq predicted gene structures underneath. abs, antibodies; shTDG, TDG-depleting short hairpin RNA; TET, ten-eleven translocation.

report the genome-wide patterns of 5fC, in addition to 5caC in the Zhang study, in wild-type (WT) mESCs [6,7]. Typically sequence reads for both modifications are small in number in WT mESCs, consistent with a low abundance, but there is a suggestion of moderate levels of 5fC at repeat regions. The overall genomic distribution of 5fC and 5caC appears to be distinct from 5hmC in WT cells

[6], but this view should be interpreted with caution due to the relatively fewer number of reads for 5fC and 5caC compared with 5hmC. Both studies recognized that enhancement of 5fC and 5caC levels in cells would improve data interpretation, so they derived similar biological strategies to improve the signal-to-noise ratio for their respective assay systems.

Visualizing sites of active demethylation by blocking base excision repair

As the 5fC and 5caC derivatives are believed to be committed for rapid removal by base excision repair-mediated mechanisms involving the protein TDG, the patterns of these two marks at steady state may not accurately reflect where demethylation is dynamically occurring in WT cells. To solve this problem the TDG protein was reduced to low levels either by short hairpin RNA interference [6] or through genetic manipulation in mESCs [7], to allow for the accumulation of both demethylation intermediates following TET-mediated oxidation of 5mC and 5hmC (Figure 1b). This increased the absolute levels of each modification and enhanced data quality and interpretation. Upon loss of TDG activity, many ectopic regions of 5fC and 5caC become apparent over genic and promoter-proximal regions; this contrasts with an earlier study that found 5fC enrichment in CpG islands (CGIs) of promoters and exons using a different assay technique [8]. The earlier study suggested CGI promoters, in which 5fC was relatively more enriched compared with 5mC or 5hmC, corresponded to transcriptionally active genes. In the present studies, upon relating the TDG-mediated changes of 5fC and 5caC to the transcriptional activities of associated genes, both groups suggest that TDG-mediated 5fC/5caC excision occurs preferentially at transcriptionally inactive promoters, implying a potential inhibitory role for the oxidative products at promoter proximal regions. No doubt these differing views will be amicably resolved in the future.

Many of the ectopic 5fC and 5caC peaks were found to correspond to regions bound by transcription factors such as Oct4 and Nanog, which themselves play key roles in the maintenance of pluripotency, as well as at sites of Polycomb-group protein binding. These results imply that TET/TDG-mediated 5mC oxidation may be a key event in the targeting of chromatin modifying proteins and transcription factors to specific loci. Interestingly, both of the studies report that upon TDG reduction/removal, the majority of ectopic 5fC and 5caC is found at non-repetitive regions of the genome outside of promoters and exons, particularly over enhancer elements. After inhibition of TDG activity, the genomic distribution patterns of 5caC and 5fC are comparable with that of 5hmC, which was not so obvious for the WT cells [6]. Closer analysis reveals a strong enrichment for both 5fC and 5caC at poised (H3K4me1 but not H3K27ac marked) enhancer elements, implying that 5mC oxidation may be crucial for the priming of such regulatory regions. Comparison to transcription factor binding site data indicated that TDG-dependent regulation of 5fC occurs preferentially at Tet1-, Tet2-, p300- and CTCF-binding regions in mESCs [7].

Interpreting DNA demethylation

As TET/TDG-mediated changes to cytosine modification states have now been shown to occur over a large number of genes and regulatory elements, this work reveals the potential for active DNA demethylation throughout the genome. Functionally, it is difficult to interpret how such modifications affect the overall epigenomic and transcriptomic landscape of the cells. The relationship between transcriptional state and DNA demethylation appears to be a complex affair. Upon depletion of TDG, only a small proportion of genes actually change in their expression state (99 genes with P-values <0.01 and a fold change >1.5-fold; or 1,192 genes with P-values <0.01 alone). In contrast, relative global changes in the levels of both 5fC and 5caC are extensive. Mass spectrometry analysis indicates that global levels of 5fC and 5caC increased by 5.6-fold and 8.4-fold, respectively, in response to TDG knockdown; 5mC and 5hmC levels were not altered [6]. Furthermore, ectopic peaks of 5fC and 5caC accumulate outside of promoters and enhancers, such as those at the 3' ends of genes, at sites that do not align to annotated regions of TDG binding [9]. As such, other proteins may be able to facilitate the base excision repair of the oxidative products of 5mC/5hmC in the absence of TDG.

In view of the low levels of these marks, it is impressive how comparable many of the conclusions are between the two studies, particularly as antibody-based methods of enrichment on low-abundance proteins and DNA modifications are challenging when compared with chemical capture based techniques (Figure 1b). Although semiquantitative, the relative enrichments of the modifications (particularly in TDG-depleted/knock-out cell lines) suggest that the marks may either be snapshots of active demethylation at key regulatory regions or 'memories' of recent transcription events. The impression is of a poised environment that is permissive for rapid transcription upon the binding of relevant factors, a feature that would be highly relevant to pluripotent cells undergoing developmentally induced reprogramming changes in response to signaling cascades. It will be interesting to determine the genome-wide patterns of both 5fC and 5caC in somatic samples containing globally higher levels of 5hmC modifications [10]. However, the data suggest that it will be a challenge to detect these low-abundance modifications in WT cells without first blocking endogenous base excision repair, but perhaps there are more surprises to come.

Abbreviations

5caC, 5-carboxylcytosine; 5fC, 5-formylcytosine; 5hmC, 5-hydroxymethylcytosine; 5mC, 5-methylcytosine; CGI, CpG island; DNMT, DNA methlytransferase; fCAB-seq, formyl-chemically assisted bisulfite sequencing; mESC, mouse embryonic stem cell; TDG, thymine DNA glycosylase; TET, ten-eleven translocation.

Competing interests

The authors declare that they have no competing interests.

References

1. Reddington JP, Pennings S, Meehan RR: **Non-canonical functions of the DNA methylome in gene regulation.** *Biochem J* 2013, **451:**13-23.
2. Koh KP, Rao A: **DNA methylation and methylcytosine oxidation in cell fate decisions.** *Curr Opin Cell Biol* 2013, **25:**152-161.
3. Shen L, Zhang Y: **5-Hydroxymethylcytosine: generation, fate, and genomic distribution.** *Curr Opin Cell Biol* 2013. doi: 10.1016/j.ceb.2013.02.017
4. Nestor CE, Ottaviano R, Reddington J, Sproul D, Reinhardt D, Dunican D, Katz E, Dixon JM, Harrison DJ, Meehan RR: **Tissue type is a major modifier of the 5-hydroxymethylcytosine content of human genes.** *Genome Res* 2012, **22:**467-477.
5. Ito S, Shen L, Dai Q, Wu SC, Collins LB, Swenberg JA, He C, Zhang Y: **Tet proteins can convert 5-methylcytosine to 5-formylcytosine and 5-CArboxylcytosine.** *Science* 2011, **333:**1300-1303.
6. Shen L, Wu H, Diep D, Yamaguchi S, D'Alessio AC, Fung HL, Zhang K, Zhang Y: **Genome-wide analysis reveals TET- and TDG-dependent 5-methylcytosine oxidation dynamics.** *Cell* 2013, **153:**692-706.
7. Song CX, Szulwach KE, Dai Q, Fu Y, Mao SQ, Lin L, Street C, Li Y, Poidevin M, Wu H, Gao J, Liu P, Li L, Xu GL, Jin P, He C: **Genome-wide profiling of 5-formylcytosine reveals its roles in epigenetic priming.** *Cell* 2013, **153:**678-691.
8. Raiber EA, Beraldi D, Ficz G, Burgess HE, Branco MR, Murat P, Oxley D, Booth MJ, Reik W, Balasubramanian S: **Genome-wide distribution of 5-formylcytosine in embryonic stem cells is associated with transcription and depends on thymine DNA glycosylase.** *Genome Biol* 2012, **13:**R69.
9. Dalton SR, Bellacosa A: **DNA demethylation by TDG.** *Epigenomics* 2012, **4:**459-467.
10. Thomson JP, Hunter JM, Lempiainen H, Muller A, Terranova R, Moggs JG, Meehan RR: **Dynamic changes in 5-hydroxymethylation signatures underpin early and late events in drug exposed liver.** *Nucleic Acids Res* 2013. doi: 10.1093/nar/gkt232