# 3. Role of 5-Hydroxymethylation and TET enzymes in remodelling the epigenome

Ksenia Skortsova, Phillippa Taberlay, Susan J. Clark, Clare Stirzaker

DNA methylation plays an important role in the epigenetic regulation of the genome. The discovery that 5-methylcytosine (5mC) can be oxidized to 5-hydroxymethylcytosine (5hmC) by the ten-eleven-translocation (TET) proteins has prompted wide interest in the potential roles of 5hmC in remodeling the methylation landscape during pluripotency, development and disease. In particular, 5hmC is proposed to act as an intermediate in active demethylation of DNA. In this review we discuss the known and potential biological functions of 5hmC and the TET enzymes at different stages of normal development and differentiation.

## Introduction

DNA cytosine methylation is one of the key epigenetic regulators of gene expression programs in normal development and disease. DNA methylation of the mammalian genome is particularly dynamic during early embryonic development followed by 'dynamic homeostasis' of the methylation landscape in normal-functioning somatic cells. However, the mechanisms controlling DNA methylation dynamics and maintenance are not entirely understood. The ability of TET proteins to further oxidize 5hmC[1], with the potential to result in cytosine demethylation[2], provides a new mechanism that may contribute to DNA methylation pattern dynamics in the early embryo[3,4], normal cell biology[5,6] and disease processes[7]. Notably, 5hmC levels vary substantially in somatic mammalian tissues[8] and the abundance and genomic distribution of 5hmC are dramatically altered during development[6,9,10]. The importance of the dynamic interplay between 5mC and 5hmC for maintaining normal DNA methylation patterns and gene expression and the causes and conse-

[keywords]
DNA methylation, DNA hydroxymethylation, TET enzymes

Role of 5-hydroxymethylation and TET enzymes in remodelling the epigenome
Ksenia Skortsova[1]/Phillippa Taberlay[1,2]/Susan J. Clark[1,2]/Clare Stirzaker[1,2] : Genomics and Epigenetics Division, Garvan Institute of Medical Research[1]/St Vincent's Clinical School, University of New South Wales[2]



This PDF was provided by Yodosha.co.jp <FOR PERSONAL USE ONLY>

quences of an imbalance between the two are key questions yet to be answered. Here we discuss the known and potential biological functions of 5hmC and the TET enzymes at different stages during normal development and differentiation.

## 1 TET enzymes and DNA hydroxymethylation provide new epigenetic regulatory functions

### 1) DNA binding properties and catalytic function of the TET proteins

The mammalian TET family of Fe(II)- and 2-oxoglutarate-dependent dioxygenases consists of three known members TET1, TET2 and TET3, that arose from a common ancestor after the divergence of jawed vertebrates[11]. All three family members contain a C-terminal catalytic dioxygenase (CD) domain that catalyzes the oxidation of 5mC to 5hmC, 5-formylcytosine (5fC) and 5-carboxycytosine (5caC)[1) 12)] (**Fig. 1**). Additionally, TET1 and TET3 have a N-terminal zinc finger CXXC DNA binding domain. TET2 is devoid of this domain due to a chromosomal inversion that excludes the exon containing the CXXC domain but forms a new, separate gene called *IDAX*[11] (**Fig. 1**). There is evidence that the IDAX protein, which binds to DNA *via* its own CXXC domain, recruits TET2 to DNA by direct interaction with its catalytic domain[13]. Therefore, TET1, TET2 and TET3 all have the ability to bind to DNA, either directly or indirectly. However, the fact that the CXXC domain sequence differs between TET1, TET3 and IDAX[14] may potentiate the existence of a distinct sets of genomic targets for each of TET proteins. This, together with the evidence of distinct TET and IDAX expression patterns throughout different stages of development and different tissues, suggests that TET and 5hmC may have multiple roles and different mechanisms of action.

### 2) Mechanisms of action of TET enzymes and 5hmC stability

The main proposed role of TET-mediated oxidation of 5mC to 5hmC is centered on the DNA demethylation process[11]. The two proposed mechanisms of DNA demethylation *via* 5hmC involve either passive demethylation through DNA replication or active demethylation through DNA repair (**Fig. 1**). During cell division patterns of cytosine methylation are maintained by the activities of DNA methyltransferase DNMT1 and Ubiquitin-Like with PHD and Ring Finger Domain 1 UHRF1[15]. However, oxidation of 5mC to 5hmC compromises DNMT1 binding and hence cytosine methylation maintenance, leading to passive DNA demethylation[15]. The concept of active DNA demethylation gained support following the recent discovery that DNA demethylation occurs not only in dividing cells but also in post-mitotic cells[16]. The predominantly accepted mechanism of active DNA demethylation in mammals involves TET-mediated sequential oxidation of 5hmC to 5fC and 5caC, followed by excision by thymine DNA glycosylase (TDG) in a base excision repair (BER) pathway[11] (**Fig. 1**). Thus, the formation of modified cytosine derivatives by the TET enzymes potentially leads to DNA demethylation, thereby playing a key role in the 'dynamic homeostasis' of the methylation landscape.

5hmC is also purported to be a 'stable' epigenetic modification in contrast to its 'transient intermediate' role in DNA demethylation[17]. The oxidation of 5mC to 5hmC impairs binding of several methyl-binding proteins[8] suggesting that 5hmC opposes 5mC function. Furthermore, the existence of proteins that bind 5hmC with higher affinity than 5mC, such as Ubiquitin-Like with PHD and Ring Finger Domains 2 (UHRF2) and Methyl-CpG binding Protein 2 (MeCP2) suggesting a substantive role for 5hmC. Potentially, the function of 5hmC as a transient intermediate in the



This PDF was provided by Yodosha.co.jp <FOR PERSONAL USE ONLY>

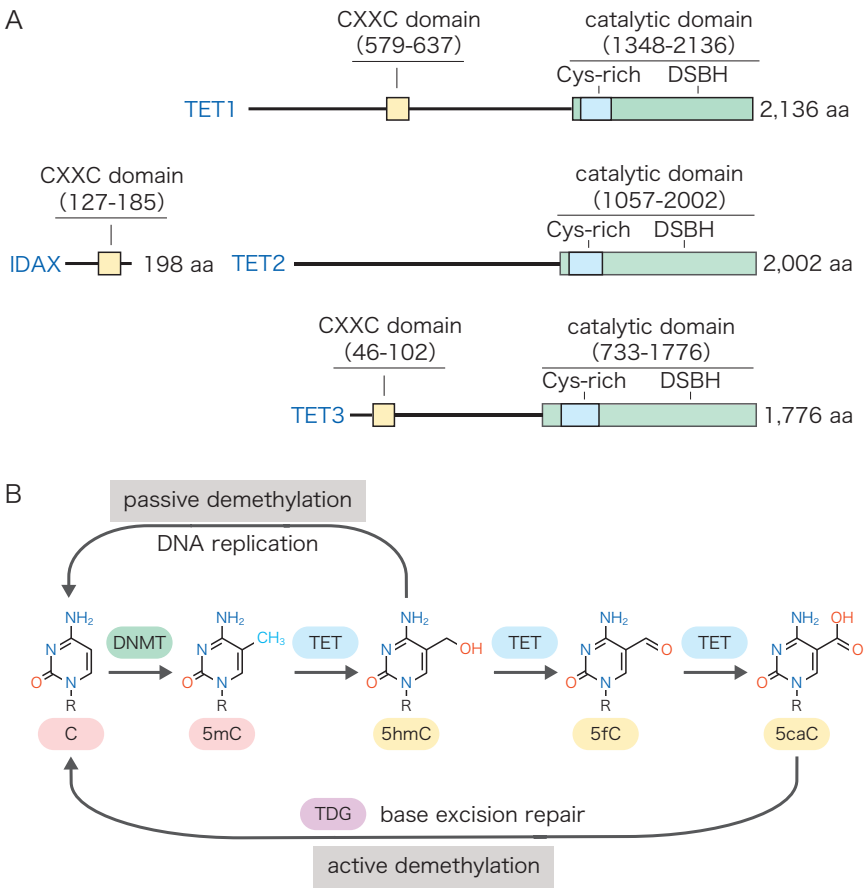


**Fig. 1** Mechanisms of DNA demethylation mediated by TET enzymes and their known domain structure

**A)** Ten-eleven-translocation (TET) proteins contain a carboxy-terminal catalytical domain (CD) harboring a cysteine-rich insert and a double stranded $\beta$-helix (DSBH) domain. DSBH domain contains binding sites for co-factor Fe(II) and co-substrate 2-oxoglutarate, essential for TET catalytic activity. TET1 and TET3 proteins also contain an amino-terminal DNA-binding CXXC domain, which was detached from TET2 protein due to chromosomal inversion resulting in a formation of separate gene IDAX. The number of amino acids shown depicts human TET proteins. **B)** 5-methylcytosine (5mC) formation is implemented by DNA methyltransferase proteins (DNMT), and can be sequentially oxidized by Ten-eleven translocation (TET) proteins to 5-hydroxymethylcytosine (5hmC), 5-formylcytosine (5fC) and 5-carboxymethylcytosine (5caC). 5caC can be excised by Thymine DNA glycosylase (TDG) and replaced with an unmodified cytosine through base excision repair (BER). BER-mediated DNA demethylation pathway is referred to as active demethylation. Alternatively, the oxidation of 5mC to 5hmC can compromise DNMT1-mediated maintenance during replication, resulting in a passive dilution of DNA methylation.

process of DNA demethylation or as a stable epigenetic modification is likely to depend on the developmental or differentiation stage, tissue type and genomic region in which TET-mediated 5mC oxidation takes place[17] and whether the process is occurring in normal or diseased cells.

## 2 TET proteins and DNA hydroxymethylation in normal development

### 1) 5hmC and 5mC dynamics in embryonic development

The best-studied example of 5hmC playing a



This PDF was provided by Yodosha.co.jp <FOR PERSONAL USE ONLY>

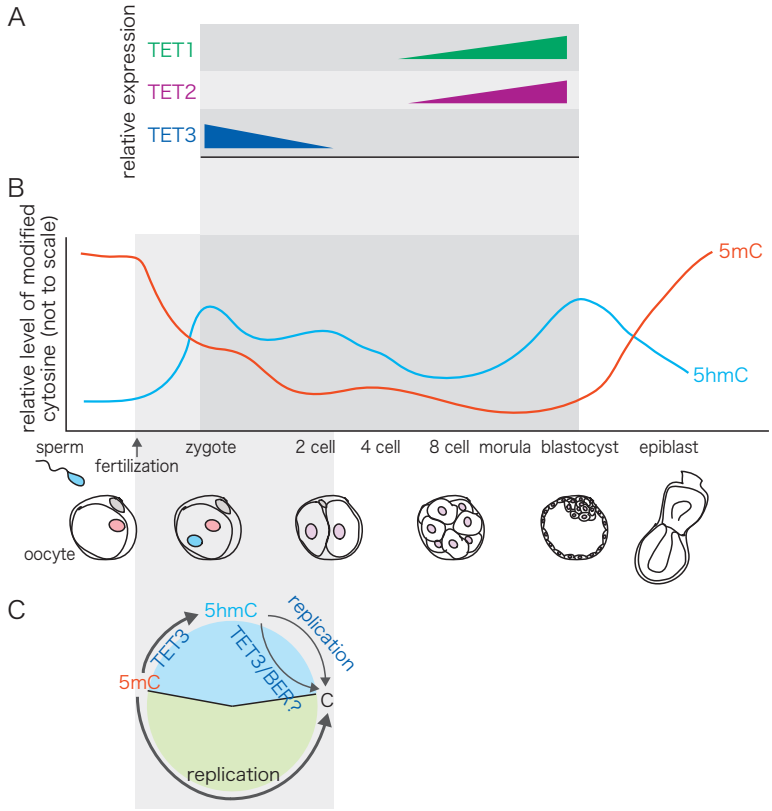


**Fig. 2** DNA methylation, hydroxymethylation and TET proteins dynamics during embryonic development

**A**) Relative levels of TET1, 2 and 3 protein expression during early embryonic development. TET3 is the only TET enzyme expressed in the zygote with reduced expression by the blastocyst stage when TET1 and TET2 become increasingly expressed. **B**) Relative changes of 5mC and 5hmC inherited from gametes, occurring in both maternal and paternal genomes in the forming zygote to establish totipotent developmental potential. DNA methylation erasure coincides with 5hmC accumulation in the zygote. **C**) Both maternal and paternal genomes undergo DNA demethylation via passive replication-dependent removal of 5mC (depicted in green) as well as TET3-mediated 5mC oxidation to 5hmC, followed by either (i) replication-dependent removal of 5hmC or (ii) replication-independent mechanism (depicted in blue).

role as a transient intermediate in DNA demethylation is in early embryonic development where two waves of DNA methylation reprogramming take place. The first wave occurs immediately after fertilization, when both maternal and paternal genomes in the forming zygote undergo substantial DNA demethylation to establish totipotent developmental potential (**Fig. 2**). During and following implantation of the embryo, a wave of *de novo* DNA re-methylation takes place. This newly established methylation pattern is retained in somatic tissues, with one more round of DNA methylation erasure occurring in primordial germ cells[11]. The mechanisms underlying removal of DNA methylation in the zygote have been hotly debated; with 5hmC serving as a potential intermediate in DNA demethylation, its involvement in active demethylation at the early developmental stages has been of special interest. Additionally, TET proteins possess distinct expression profiles during development and differentiation; TET3 is the only TET enzyme expressed in the zygote, fol-



This PDF was provided by Yodosha.co.jp <FOR PERSONAL USE ONLY>

lowed by reduced expression by the blastocyst stage[9) 18) 19)], whereas TET1 and TET2 are increasingly expressed (**Fig. 2**). Until recently it was proposed, based on immunostaining experiments, that the paternal genome in the fertilized zygote undergoes TET3–mediated demethylation[9) 18) 19)], whereas the maternal genome is protected from TET3–mediated 5mC oxidation by the maternal factor PGC7 (Stella)[20)] and undergoes exclusively passive replication–dependent dilution of DNA methylation[19)]. However, more quantitative and sensitive approaches to immunostaining, including reduced representation bisulfite sequencing (RRBS)[21)], have revealed replication–dependent *as well as* TET3–mediated 5mC oxidation–dependent demethylation mechanisms in both paternal and maternal genomes[3) 4)] (**Fig. 2**). Moreover, TET3 knockout in mice caused abnormal development and perinatal lethality, highlighting the extreme importance of TET3 role in DNA methylation reprogramming in the developing zygote[18)]. Importantly, since RRBS covers only a portion of total number of CpG sites in the genome (approx 4%[22)]), whole genome methylation analysis at single–nucleotide resolution would be necessary to accurately assess the contribution of the active and passive demethylation pathways.

### 2) 5hmC and TET enzymes regulate pluripotent state and extraembryonic lineage commitment

Embryonic stem cells (ESCs) can be maintained in a pluripotent state or triggered towards differentiation *in vitro*. ESCs, therefore, provide an excellent model for studying the mechanisms underlying the retention of DNA methylation patterns in the non–differentiated state, and for investigating triggers of DNA methylation dynamics during differentiation. While expression of TET1 and TET2 is undetectable in the zygote, both are highly expressed in ESCs[1) 23)], which, together with the abundance of 5hmC[12) 24)] suggests an essential role for TET activity and hydroxymethylation. However, current literature encompasses contradictory studies regarding the role of TET1 and TET2 proteins in regulation of ESC pluripotency[25) 26)]. TET1 was first reported as an important regulator of ESC pluripotency when knockdown of TET1, but not TET2 or TET3, impaired ESC self–renewal via hypermethylation and inactivation of the key stem cell factor *Nanog*[23)]. Similarly, other studies revealed TET1 knockdown in ES cells[25)] decreased expression of other pluripotency–related genes. Downregulation of these genes was associated with increased 5mC and decreased 5hmC at their promoters and increased propensity to extraembryonic lineage differentiation[25)]. Consistent with this, during ESC differentiation to embryoid bodies (EB), both TET1 and TET2 are rapidly downregulated[25) 27)] coinciding with decreased expression of key stemness genes associated with significant increased promoter 5mC and decreased 5hmC levels[25)]. Together these studies support that the balance between promoter methylation and hydroxymethylation is linked to the relationship between pluripotent state and extraembryonic lineage commitment.

Conversely, other studies have revealed that despite TET1 or TET2 depletion in ESCs, pluripotency was retained, along with the ability to form all three germ layers, though with skewed developmental potential towards trophectoderm or mesendoderm, in the case of TET1 or TET2 KD, respectively[26) 28)]. ESCs with double knockout of both TET1 and TET2, which eliminated any potential compensatory functions, also remained pluripotent, although they show a skewed differentiation towards extraembryonic lineage[29)]. Potentially such differences could be explained by mESC background differences used in these studies or different degrees of siRNA–mediated TET knockdown.



This PDF was provided by Yodosha.co.jp <FOR PERSONAL USE ONLY>

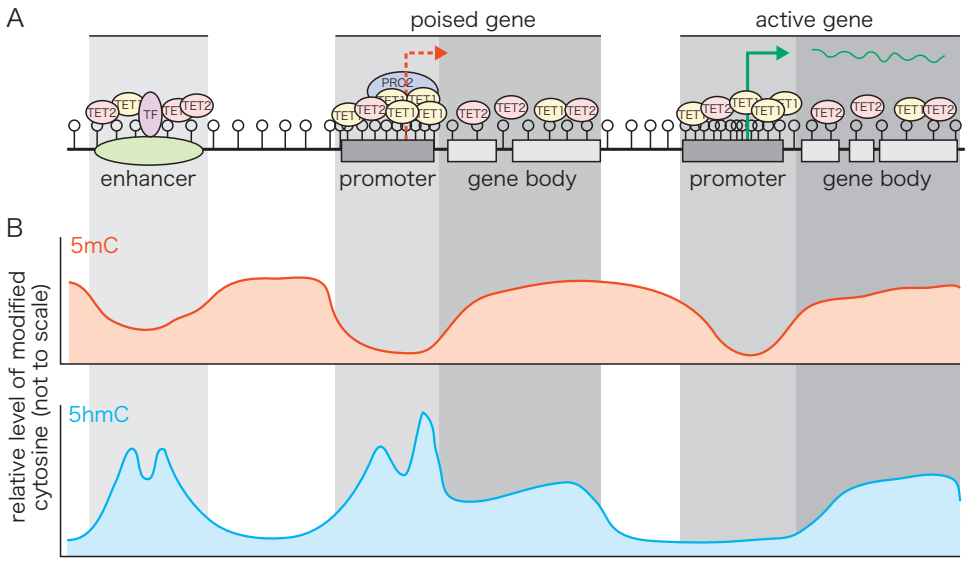


**Fig. 3** Genomic distribution of TET proteins, DNA methylation and hydroxymethylation

**A)** Schematic representation of TET proteins enrichment at enhancers and promoters of poised genes together with polycomb repressive complex 2 (PRC2) as well as promoters of active genes. Whereas TET1 is predominantly enriched at promoters, TET2 mainly occupies gene bodies and enhancers. Lollipops depict CpG dinucleotides highlighting the CpG density at those regions. **B)** Methylation and hydroxymethylation profiles at enhancers and promoters of active and poised genes. 5hmC is enriched at enhancers, surrounding the transcription factor (TF) binding site, coinciding with the local depletion of 5mC. Unlike promoters of active genes, 5hmC is enriched at the promoters of poised genes with a local depletion at the TSS. 5hmC, together with 5mC, is enriched at gene bodies of both active and poised genes.

## 3 Genomic distribution of TET proteins and hydroxymethylation in ESC

Genome-wide mammalian mapping studies have revealed 5hmC and TET protein enrichment at promoters, gene bodies and distal regulatory elements[30) 31)]. In ESCs, 5hmC enrichment at promoters is dependent on gene activity, promoter CpG density and chromatin state[30)~32)]. 5hmC is depleted from promoters of highly expressed genes but enriched at the bivalent promoters of poised genes[30) 33)] (**Fig. 3**). Interestingly, the distribution of 5hmC does not appear to be tightly correlated to the abundance of the TET proteins. For example, TET1 is enriched at the promoters of both poised genes, where it associates with the chromatin repressive complex PRC2, and actively transcribed genes[31) 34)] (**Fig. 3**). Intriguingly, TET1 depletion results in directional gene expression changes[34)] despite a minimal increase in DNA methylation[30) 34) 35)], suggesting that TET1 functions at gene promoters primarily through interactions with distinct histone marks and associated chromatin remodelers rather than through modulation of the methylation landscape. Whereas TET1 primarily binds gene promoters, TET2 is predominantly involved in the regulation of 5hmC hydroxymethylation profiles at gene bodies at the exons at exon-intron boundaries[32)] and distal regulatory elements[35)~38)].

5hmC is also enriched at enhancers and CTCF binding sites in a bimodal fashion surrounding the transcription factor (TF) binding motif[38)] (**Fig. 3**). 5hmC enrichment coincides with a local depletion of 5mC[38)], which increases in response to TET depletion[35) 37)], supporting that TET-mediated oxi-



This PDF was provided by Yodosha.co.jp <FOR PERSONAL USE ONLY>

dation of 5mC to 5hmC is essential for the retention of reduced 5mC levels at distal regulatory elements, in contrast to its function at promoters[28]. In fact TET depletion–mediated hypermethylation predominantly happens at 'weak' enhancers with lower enrichment of H3K27 acetylation and lower TF occupancy but higher baseline levels of 5mC methylation[37]. These data suggest that weak enhancers might be subject to a more dynamic interplay between DNMT–mediated methylation and TET–mediated demethylation, potentially governing enhancer activity through the mediation of TF binding.

Furthermore, during differentiation an increase of 5hmC enrichment, concomitant with cell–specific TF binding and H3K27 acetylation, has been observed at activated enhancers[39] as well as at newly forming CTCF binding sites[40]. Importantly, the relative stability of 5hmC observed in differentiated cells suggests its role as an independent mark involved in the regulation of enhancer activity by preventing the binding of DNA methylation machinery[39]. However, constant levels of 5hmC do not rule out the possibility of continuous turnover of methylated and unmethylated cytosines via 5hmC stage. Overall, TET enzymes and hydroxymethylation play critical roles in the functioning of regulatory elements, coinciding with the local chromatin environment.

## 4 Hydroxymethylation in brain

Among studied metabolically active non–proliferating adult tissues, brain has the highest 5hmC levels, reaching 40% of 5mC in Purkinje neurons, suggesting a specific role of 5hmC in brain function[41]. Different approaches have revealed that 5hmC accumulates in the brain during differentiation[10] and postnatal neurodevelopment[5,6]. TET2 and TET3 expression also increases with the onset of differentiation, with TET2/TET3 depletion compromising the process[10].

5hmC is enriched at multiple genomic regions in brain tissue and therefore potentially serves different regulatory roles. Notably, 5hmC is depleted from gene promoters regardless of the gene activity[37] but is enriched at gene bodies, where levels correlate with the activity of related genes[6,24,42]. In addition, 5hmC enrichment has also been observed at exon–intron boundaries, suggesting a role in the splicing regulation[43,44]. The highest reported abundance of 5hmC in brain cells was found at both poised and active enhancers, pointing to its specific role at distal regulatory elements[39]. Indeed, functional studies have revealed that 5hmC and TET are involved in hypomethylation of enhancers concomitant with their activation during neurodevelopment[6]. Interestingly, 5hmC accumulation together with the loss of repressive histone mark H3K27me3 at gene bodies of activated genes during neuronal differentiation did not coincide with hypomethylation[10]. This finding suggests that oxidation of 5mC to 5hmC *without* subsequent demethylation can play a unique role as a regulator of gene expression. Importantly, TET depletion in brain compromises spatial learning, short–term memory[45] and behavioral adaptation[46,47], highlighting a critical role for TET proteins and hydroxymethylation in brain function.

## Concluding remarks

The discovery of TET proteins and their role in oxidizing 5mC to 5hmC has prompted wide interest and provided novel insights into the pathways underlying DNA demethylation through a 5hmC 'transient intermediate'. Levels of 5hmC vary between tissues, and are likely dependent on cell proliferation rates, with highest levels reported in the brain. To date, the majority of studies on 5hmC have focused on the embryo and



This PDF was provided by Yodosha.co.jp <FOR PERSONAL USE ONLY>

brain. One of the major challenges in the field has been how to accurately detect the presence of 5hmC and to distinguish it from 5mC. With the recent development of more sensitive single-base resolution methods that clearly identify 5hmC and its oxidized products, the roles of these modifications are being increasingly elucidated. Further studies using single-base resolution methods to interrogate the genomic distribution of 5hmC and its oxidized derivatives, the interplay between the TET proteins and their binding partners and the consequences of TET protein loss-of-function at different developmental stages and different tissues is critical to increase our understanding of the role of TET proteins and 5hmC in modulating the epigenomic landscape and gene expression patterns.

**Acknowledgments**
The authors thank Dr Brigid O'Gorman for reviewing the manuscript and artwork.

## References

1) Ito S, et al : Science, 333 : 1300-1303, 2011
2) He YF, et al : Science, 333 : 1303-1307, 2011
3) Shen L, et al : Cell Stem Cell, 15 : 459-470, 2014
4) Guo F, et al : Cell Stem Cell, 15 : 447-458, 2014
5) Wang T, et al : Hum Mol Genet, 21 : 5500-5510, 2012
6) Lister R, et al : Science, 341 : 1237905, 2013
7) Ficz G & Gribben JG : Genomics, 104 : 352-357, 2014
8) Branco MR, et al : Nat Rev Genet, 13 : 7-13, 2011
9) Wossidlo M, et al : Nat Commun, 2 : 241, 2011
10) Hahn MA, et al : Cell Rep, 3 : 291-300, 2013
11) Pastor WA, et al : Nat Rev Mol Cell Biol, 14 : 341-356, 2013
12) Tahiliani M, et al : Science, 324 : 930-935, 2009
13) Ko M, et al : Nature, 497 : 122-126, 2013
14) Long HK, et al : Biochem Soc Trans, 41 : 727-740, 2013
15) Hashimoto H, et al : Nucleic Acids Res, 40 : 4841-4849, 2012
16) Gavin DP, et al : Neuropharmacology, 75 : 233-245, 2013
17) Hahn MA, et al : Genomics, 104 : 314-323, 2014
18) Gu TP, et al : Nature, 477 : 606-610, 2011
19) Iqbal K, et al : Proc Natl Acad Sci U S A, 108 : 3642-3647, 2011
20) Nakamura T, et al : Nature, 486 : 415-419, 2012
21) Meissner A, et al : Nucleic Acids Res, 33 : 5868-5877, 2005
22) Stirzaker C, et al : Trends Genet, 30 : 75-84, 2014
23) Ito S, et al : Nature, 466 : 1129-1133, 2010
24) Song CX, et al : Nat Biotechnol, 29 : 68-72, 2011
25) Ficz G, et al : Nature, 473 : 398-402, 2011
26) Koh KP, et al : Cell Stem Cell, 8 : 200-213, 2011
27) Li T, et al : Mol Neurobiol, 51 : 142-154, 2015
28) Dawlaty MM, et al : Cell Stem Cell, 9 : 166-175, 2011
29) Dawlaty MM, et al : Dev Cell, 24 : 310-323, 2013
30) Wu H, et al : Genes Dev, 25 : 679-684, 2011
31) Xu Y, et al : Mol Cell, 42 : 451-464, 2011
32) Huang Y, et al : Proc Natl Acad Sci U S A, 111 : 1361-1366, 2014
33) Neri F, et al : Genome Biol, 14 : R91, 2013
34) Wu H, et al : Nature, 473 : 389-393, 2011
35) Lu F, et al : Genes Dev, 28 : 2103-2119, 2014
36) Stroud H, et al : Genome Biol, 12 : R54, 2011
37) Hon GC, et al : Mol Cell, 56 : 286-297, 2014
38) Yu M, et al : Cell, 149 : 1368-1380, 2012
39) Dubois-Chevalier J, et al : Nucleic Acids Res, 42 : 10943-10959, 2014
40) Sérandour AA, et al : Nucleic Acids Res, 40 : 8255-8265, 2012
41) Kriaucionis S & Heintz N : Science, 324 : 929-930, 2009
42) Mellén M, et al : Cell, 151 : 1417-1430, 2012
43) Wen L, et al : Genome Biol, 15 : R49, 2014
44) Khare T, et al : Nat Struct Mol Biol, 19 : 1037-1043, 2012
45) Zhang RR, et al : Cell Stem Cell, 13 : 237-245, 2013
46) Li X, et al : Proc Natl Acad Sci U S A, 111 : 7120-7125, 2014
47) Feng J, et al : Nat Neurosci, 18 : 536-544, 2015

＜Curriculum vitae＞
Ksenia Skortsova graduated with a Master's Degree in Molecular Biology in 2011 from Novosibirsk State University, Russia. She is currently a postgraduate student in the Epigenetics Laboratory in the Genomics and Epigenetics Division at the Garvan Institute of Medical Research, Australia. She has developed and refined the approaches to map the genome-wide distribution of 5-hydroxymethylation, enabling her to interrogate the role of DNA hydroxymethylation in the maintenance and remodelling of the DNA methylation landscape.

Phillippa C. Taberlay was awarded her PhD from the University of Tasmania in 2008 before accepting a post-doctoral research fellowship at the University of Southern California, USA. She then returned to join Prof. Susan Clark at the Garvan Institute of Medical Research in 2011 where she is Leader of the Chromatin Dynamics Group. Her interests are in chromatin remodelling complexes and three-dimensional chro-



This PDF was provided by Yodosha.co.jp <FOR PERSONAL USE ONLY>

matin structure in the context of epigenetic deregulation in cancer.

Susan Clark is Head of the Genomics and Epigenetics Division and Epigenetics Research Laboratory at the Garvan Institute of Medical Research, Sydney. Prof Clark graduated with a PhD at the University of Adelaide in 1982. Prof Clark and her group developed and implemented bisulphite sequencing for DNA methylation analysis in the early 1990s. Her DNA methylation studies have made extensive ground-breaking discoveries relating to DNA methylation patterns in normal and cancer genomes. Prof Clark is the founder of the Australian Epigenetic Alliance, a founding member of IHEC and a NHMRC Senior Principal Research Fellow.

Clare Stirzaker obtained her PhD at Macquarie University, Sydney in 1990. She has worked in the field of Epigenetics for the last 15 years and has played a major role in developing epigenome technologies that have enabled key questions to be answered in the field. She is currently Group leader of Epigenetic Deregulation in the Epigenetics Research Laboratory in the Genomics and Epigenetics Division, Garvan Institute of Medical Research, Sydney. Her research interest is in DNA methylation and understanding the mechanisms underpinning aberrant DNA methylation patterns in cancer.



This PDF was provided by Yodosha.co.jp <FOR PERSONAL USE ONLY>