MOLECULAR BIOLOGY

Nascent DNA methylome mapping reveals inheritance of hemimethylation at CTCF/cohesin sites

Chenhuan Xu and Victor G. Corces*

The faithful inheritance of the epigenome is critical for cells to maintain gene expression programs and cellular identity across cell divisions. We mapped strand-specific DNA methylation after replication forks and show maintenance of the vast majority of the DNA methylome within 20 minutes of replication and inheritance of some hemimethylated CpG dinucleotides (hemiCpGs). Mapping the nascent DNA methylome targeted by each of the three DNA methyltransferases (DNMTs) reveals interactions between DNMTs and substrate daughter cytosines en route to maintenance methylation or hemimethylation. Finally, we show the inheritance of hemiCpGs at short regions flanking CCCTC-binding factor (CTCF)/cohesin binding sites in pluripotent cells. Elimination of hemimethylation causes reduced frequency of chromatin interactions emanating from these sites, suggesting a role for hemimethylation as a stable epigenetic mark regulating CTCF-mediated chromatin interactions.

Cytosine DNA methylation in mammals is maintained mainly by the canonical DNA maintenance methyltransferase DNMT1 during each cell cycle (1, 2). By interacting with proliferating cell nuclear antigen (PCNA) and ubiquitin-like–containing PHD and RING finger domains 1 (UHRF1) during DNA replication, DNMT1 is recruited to replication foci and loaded onto hemiCpGs to methylate the nascent cytosines (Cs) (3–5). Although the onset of this process is closely coupled with the entry into S phase, the kinetics of maintenance methylation and the content of the nascent DNA methylome have never been studied quantitatively on a genome-wide scale (6). Furthermore, although various biochemical (7, 8) and genetic perturbation experiments (6, 9–12) have strongly suggested the involvement of the de novo methyltransferase DNMT3A/3B in maintenance methylation, direct evidence of in vivo interaction between DNMTs and hemiCpGs is missing.

To gain insights into these key aspects of maintenance methylation, we used nascent DNA bisulfite sequencing (nasBS-seq) to measure cytosine methylation frequency strand-specifically on nascent chromatin across the genome (fig. S1, A and B, and supplementary materials). We first labeled H9 human embryonic stem cells (H9-hESCs) with the nucleotide analog ethynyl-deoxyuridine (EdU) for 20 min as a pulse condition. Libraries for a chase condition were also made by labeling the cells with EdU for 20 min and growing them for another 8 hours in the absence of EdU to monitor the maintenance methylation at a later time point within the cell cycle. We obtained 357 to 544 million uniquely mapped and deduplicated alignments from libraries for each strand, covering 5.6 to 9.8 billion Cs (fig. S2, A and B), converging to 24 million parent-daughter CpG dyads (pdCpGs) from either pulse or chase (fig. S2C). The methylation frequency was highly reproducible between replicates for each library (fig. S2D). Methylation frequency between parental Cs (pC) and daughter Cs (dC) in the same pdCpGs (Fig. 1A and fig. S3A), and between the same Cs in pulse and chase (fig. S3B), were highly correlated. Although Cs in the context of CH are not symmetrically methylated in mammals (13), their methylation was also maintained on the dCs in the other nascent DNA duplex (fig. S3C). The methylation frequency of dCs was globally maintained in both pulse and chase irrespective of the nature of the genomic features investigated (fig. S3D). These results suggest that the vast majority of the DNA methylome is maintained within 20 min after passage of replication forks and thereafter.

Despite the high correlation of methylation frequency between pC and dC, the two Cs in many CpGs showed differential methylation frequency (Fig. 1B and fig. S4A), suggesting the existence of hemiCpGs with a spectrum of frequencies (fig. S4B). The strand-specific nascent DNA methylome enables the resolution of different types of hemiCpGs with respect to the parent-daughter axis. Using a highly stringent cutoff for differences in 5-methylcytosine (mC) [$\Delta mC \leq -75\%$, or $\geq 75\%$, $\Delta mC = m(pC - dC)$], we obtained a list of 23,305 CpGs with at least one dyad showing hemimethylation either in pulse or chase (Fig. 1C). The vast majority (96%) of them were hemimethylated in only one dyad and failed to reproduce the methylation pattern in the other condition (fig. S4, C to F), suggesting that they may represent the rapid DNA methylation turnover events abundant in pluripotent cells (14). In contrast, the methylation pattern of the remaining 4% CpGs hemimethylated in both dyads in a concordant way (Cs on either two Watson or two Crick strands are methylated) was highly consistent between pulse and chase (Fig. 1D), suggesting that concordant hemiCpGs were stably inherited through S phase. By concatenating the data from pulse and chase (see supplementary materials), we expanded the category of concordant hemiCpGs to include 2467
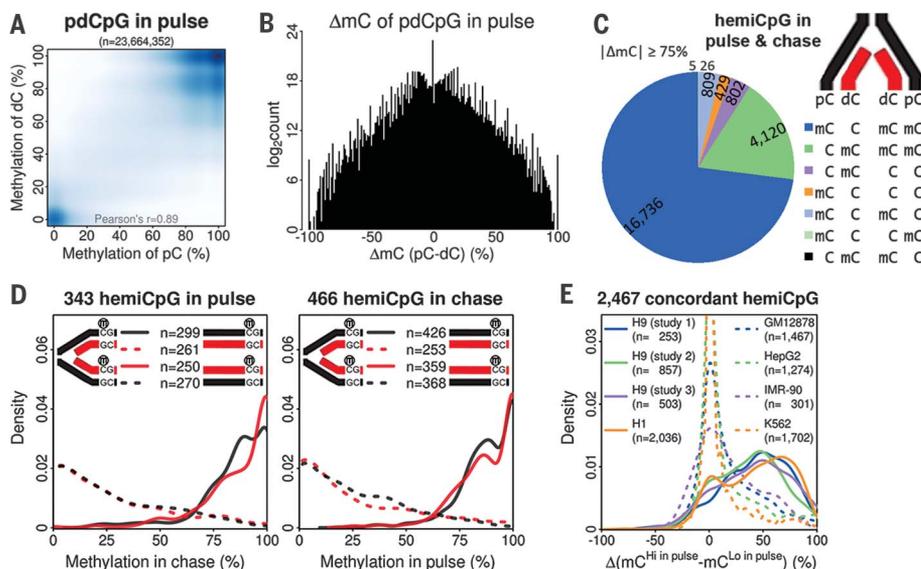

Fig. 1. The vast majority of the DNA methylome is maintained 20 min after replication. (**A**) Correlation of methylation frequency between pCs and dCs within the same pdCpGs in pulse. (**B**) Count of pdCpGs in pulse with differential ΔmC values. (**C**) Different types of hemiCpGs with all four Cs mapped at least four times. (**D**) For concordant hemiCpGs in pulse, the distribution of methylation frequency of four Cs in chase is shown (left), and vice versa (right). (**E**) All concordant hemiCpGs were intersected with WGBS data sets from other human cells. The distribution of ΔmC values is shown for each data set.

Department of Biology, Emory University, 1510 Clifton Road NE, Atlanta, GA 30322, USA.
*Corresponding author. Email: vcorces@emory.edu

CpGs and confirmed their stable inheritance across up to six passages (>12 cell divisions) (fig. S4, G and H). The ΔmC values of these CpGs were compared with several whole-genome BS-seq (WGBS) data sets in various human cells. Unexpectedly, the majority of them are conserved in other pluripotent cells but are absent in non-pluripotent cells (Fig. 1E), suggesting that hemimethylation could be cell type–specific and well conserved across related cell lineages.

WGBS only reports the independent methylation frequency of two Cs in the same CpGs. To obtain methylation status of CpGs per se, we developed a computational method called in silico strand annealing (iSA) to resolve the nasBS-seq data and identify pairs of alignments sharing exactly the same two ends between strands of parent^{Watson} and daughter^{Crick} and between daughter^{Watson} and parent^{Crick} (Fig. 2A and supplementary materials). We employed a "moving-ends" statistical test to justify that most of these pairs (26- to 111-fold enrichment over random pairing) represent distinct nascent double-stranded DNA (dsDNA) fragments (fig. S5A). iSA enabled us to call intramolecule CpGs (intraCpGs) from single dsDNA fragments and to determine their methylation state to be one of four types: methylation (intraCpG^{me}), unmethylation (intraCpG^{unme}) or pC- or dC- hemimethylation (intraCpG$^{hemi-pC}$ or intraCpG$^{hemi-dC}$). About 4.5 and 2.1 million intraCpGs were called from all replicates in pulse and chase, respectively (Fig. 2B). The two conditions showed nearly identical fractions for all four types, including a surprisingly high and consistent 14% combined fraction of intraCpG^{hemi}. Next, we used iSA to resolve published WGBS data sets in mouse early embryonic stages (15) and showed that hemiCpGs account for 4 to 18% of the DNA methylome (Fig. 2C) and is relatively

depleted at transcription start sites (TSSs) (Fig. 2D). Murine intracisternal A-particle (IAP) retrotransposons are resistant to demethylation during early embryogenesis (15). Indeed, in inner cell mass (ICM) cells, intraCpG^{me} accounts for 47% of the DNA methylome in IAPs versus 14% genome-wide, whereas intraCpG^{hemi} accounts for 17 versus 16% genome-wide. Notably, ICM cells have the highest frequency of hemiCpGs on gene bodies, where it correlates slightly with transcription level, although it anticorrelates with transcription at promoters (fig. S5B), suggesting a pleiotropic role of hemiCpGs on gene expression. iSA was also used to resolve WGBS and Tet-assisted bisulfite sequencing (TAB-seq) data sets in H1-hESCs (fig. S5C) (16) and showed that although 5-hydroxymethylcytosine (5hmC) preferentially exists in hemimethylated form (fig. S5D), the vast majority of hemiCpGs discovered by nasBS-seq/WGBS is contributed by mC (fig. S5E). The ΔmC values from WGBS highly correlate with the frequency of hemiCpGs resolved by iSA (fig. S5F), suggesting that ΔmC values from WGBS can serve as a proxy for the frequency of hemimethylation.

The use of 20-min EdU labeling achieved a synchronization of genomic fragments by their replicative "age" of 10 min on average (0 to 20 min after passage of the local replication fork). The same frequency of hemiCpGs in pulse and chase suggests that the maintenance methylation reaction happens in a subminute scale (a 1-min-long methylation reaction would result in the pulse sample having 5% more hemiCpGs than chase), preventing nasBS-seq from revealing the rich plethora of pC-methylated hemiCpGs en route to maintenance methylation. We thus postulated that an enrichment of binding events between DNMTs and nascent chromatin would achieve both spatial and temporal enrichment of such

transient interactions and would help identify cognate substrate CpGs maintained by a certain DNMT. Hence, we used chromatin immuno-precipitation on nascent chromatin followed by bisulfite sequencing (nasChIP-BS-seq) to specifically map the nascent DNA methylome targeted by DNMT1, DNMT3A, or DNMT3B in both pulse and chase (fig. S6, A and B). Unexpectedly, in pulse but not in chase, all three DNMT-targeted methylomes of the two daughter strands showed incomplete methylation, which was most apparent at centers of alignments, indicative of the precise location of DNMTs (fig. S6C). Analysis of the data using iSA revealed that ~42, 46, and 44% of all DNMT1-, DNMT3A- and DNMT3B-targeted CpGs in pulse were intraCpG$^{hemi-pC}$, respectively, whereas the same category only contributed 7, 6, and 5% in chase, respectively (Fig. 3A and fig. S6D).

Furthermore, binding sites for all three DNMTs showed an enrichment of CpGs over flanking sequences in pulse (fig. S6E). In chase, the enrichment diminished for DNMT1 and DNMT3B, whereas DNMT3A showed an enrichment of methylated CpGs (fig. S6E), suggesting that DNMTs may have differential occupancy preferences on nascent and mature chromatin (17, 18). When viewed through the pulse nasBS-seq, the vast majority of DNMT-targeted intraCpG$^{hemi-pC}$ were fully methylated (Fig. 3B), suggesting that they were methylated shortly after the binding events. In DNMT1 knockout (KO) cells (12), both DNMT1-targeted intraCpG$^{hemi-pC}$ and intraCpG^{me} showed higher than average reduction of methylation, suggesting that their methylation state is maintained by DNMT1 (Fig. 3C). Under DNMT3A/3B double KO (12), both DNMT3A- and 3B-targeted intraCpG$^{hemi-pC}$ showed significantly higher reduction of methylation than the targeted intraCpG^{me}
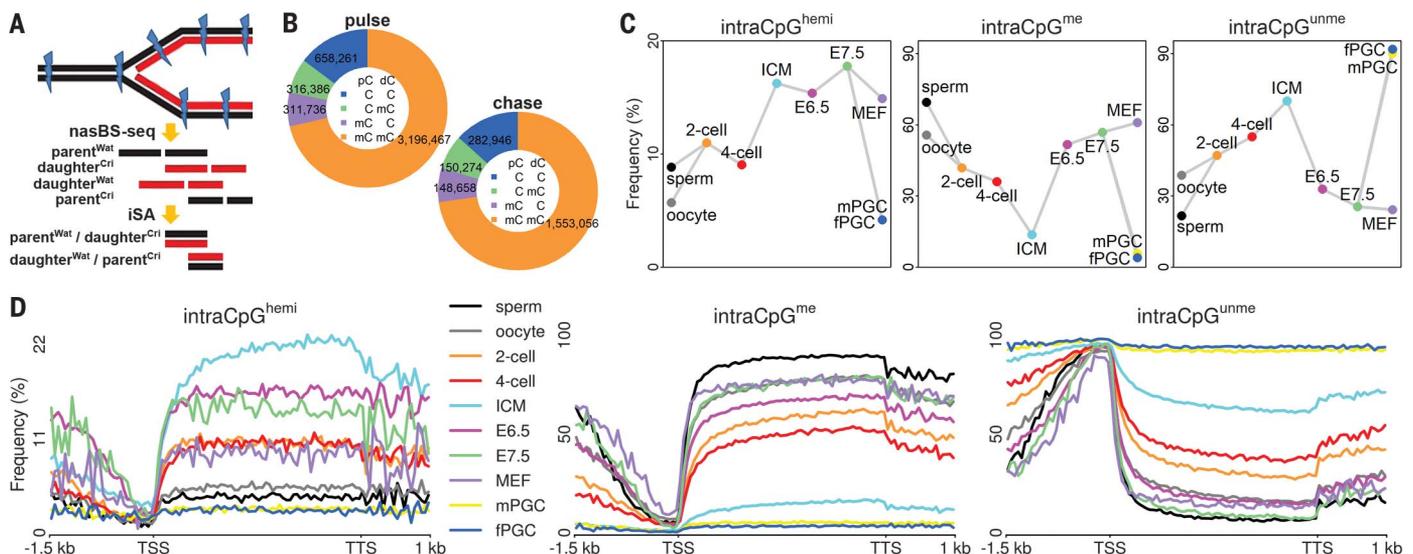
Fig. 2. HemiCpG is an important component of the DNA methylome. (**A**) Schematic representation of the principles underlying the iSA method. (**B**) The fraction of all four types of intraCpGs in pulse and chase. (**C**) The frequency of three types of intraCpGs (with two types of intraCpGs^{hemi} combined) at different mouse embryonic stages. ICM, inner cell mass; MEF, mouse embryonic fibroblast; mPGC/fPGC, male/female primordial germ cell. (**D**) The frequency of three types of intraCpGs at genic regions at different mouse embryonic stages.

(Fig. 3C), suggesting that these intraCpG$^{hemi-pC}$ are better candidates for DNMT3-maintained CpG than the targeted intraCpG^{me}. Indeed, these two types of CpGs showed mutually exclusive distribution, suggesting that they are subject to different regulation (fig. S6F). We next asked if nasChIP-BS-seq can also capture the substrate state of dCs en route to de novo methylation by examining the methylation state of targeted dCs in either inherited hemiCpGs or maintained CH methylation. In both cases, the yet-to-be-methylated dCs showed extensive hypomethylation in DNMT3A/3B nasChIP-BS-seq but not in nasBS-seq (Fig. 3D and fig. S6G). These results suggest that nasChIP-BS-seq can visualize the transient interactions
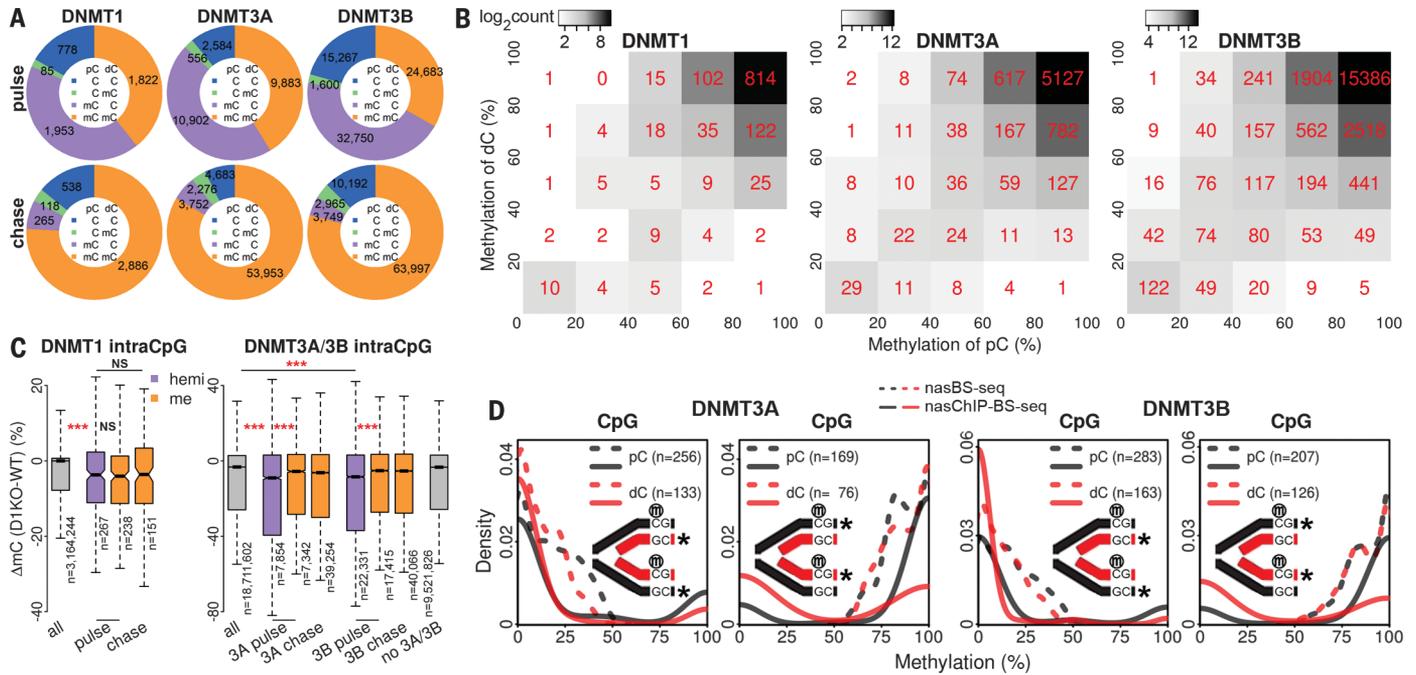

Fig. 3. Transient interactions between DNMTs and substrate dCs in both maintenance and de novo methylation. (**A**) The fraction of all four types of DNMT-targeted intraCpGs in pulse and chase. (**B**) Counts of all DNMT-targeted intraCpGs$^{hemi-pC}$ in pulse allocated to the appropriate cells according to their methylation frequency in pulse nasBS-seq. (**C**) Reduction of methylation under DNMT1 KO (24 hours) or DNMT3A/3B double KO (late) is shown for all CpGs, DNMT-targeted intraCpGs$^{hemi-pC}$, and intraCpGs^{me} in pulse and chase and unmapped CpGs. ***P < 0.001. NS, not significant. (**D**) Distribution of methylation frequency of the four Cs in concordant hemiCpGs viewed through nasBS-seq and DNMT3A/3B nasChIP-BS-seq. The asterisks mark the two Cs inspected in each panel.
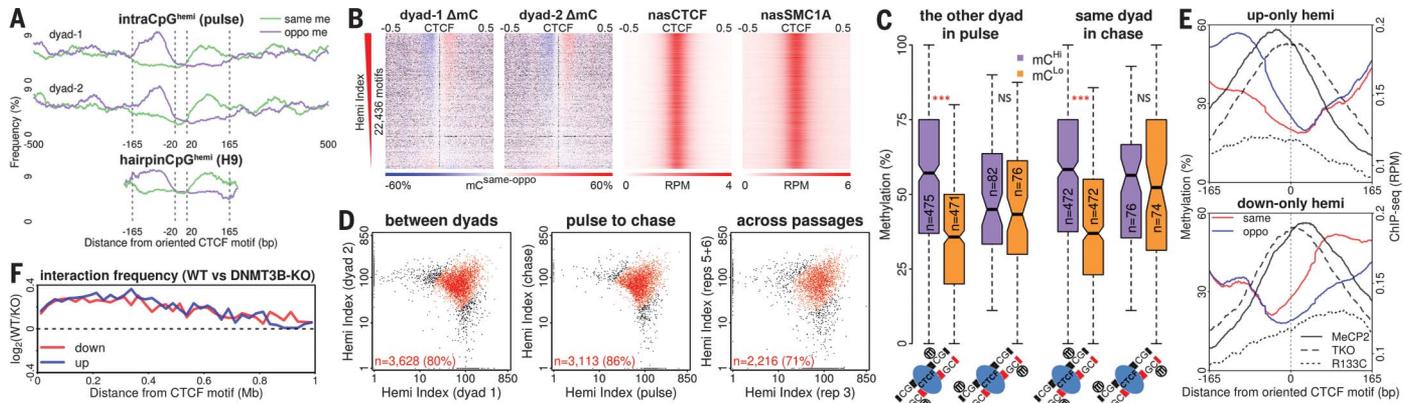

Fig. 4. Inherited hemiCpGs flanking CTCF/cohesin sites may regulate chromatin interactions. (**A**) Frequency of motif or opposite strand-methylated (same me or oppo me) intraCpGs^{hemi} around oriented CTCF motifs co-occupied by CTCF/SMC1A from the two nascent DNA duplexes. Frequency of hairpinCpG^{hemi} from CTCF ChIP-hairpinBS-seq is also shown. (**B**) All CTCF motifs co-occupied by CTCF/SMC1A in pulse were ranked by their hemi-index. ΔmC of CpGs from the two nascent DNA duplexes and reads per million (RPM) for CTCF and SMC1A nasChIP-seq within a 1-kb window surrounding the motifs are shown. Black in the ΔmC heat maps represents missing data points. (**C**) All hemiCpGs (ΔmC ≥ 67% or ≤ −67%) from two flanking regions in (B) were retrieved. Methylation frequency of the two Cs in the other dyad in pulse (left) or the same dyad in chase (right) are shown. ***P < 0.001. NS, not significant. (**D**) The hemi-index of CTCF motifs showing HI > 50 in the pooled data were compared between two dyads, from pulse to chase, and across five passages. (**E**) Occupancy of WT and R133C mutant MeCP2 in WT mESC, and MeCP2 in DNMT1/3A/3B triple KO (TKO) mESCs was profiled around CTCF motifs showing upstream- or downstream-only hemimethylation in mESCs. (**F**) The ratio between interaction contacts from Hi-ChIP in WT and DNMT3B-KO HUES64 hESCs emanating from occupied CTCF motifs and extending up to ±1-Mb window is shown.

between a certain DNMT and substrate dCs in both maintenance and de novo methylation (fig. S6H).

To identify chromatin features associated with hemiCpGs, we examined the frequency of different types of intraCpG at different genomic features in H1-hESC. CTCF binding sites showed a very high ratio of intraCpG^{hemi} over intraCpG^{me} (fig. S7A). CTCF/cohesin binding sites orchestrate three-dimensional chromatin interactions across the mammalian genome (*19*). We then developed nasChIP-seq to map the binding landscape of CTCF and SMC1A (a cohesin subunit) on nascent chromatin in H9-hESC (fig. S7, B to E). By examining the average methylation profiles of the strands harboring the CTCF motif and the opposite strands around oriented CTCF motifs, we found two short regions flanking occupied CTCF motifs exhibiting an apparent spectrum of ΔmC with opposing orientation on the two sides (fig. S8A). The same pattern exists in H1-hESC, naïve H9-hESC, mouse ESC (mESC) (fig. S8A), and mouse embryos as early as the eight-cell stage (fig. S8B). Two independent methods, iSA and ChIP-hairpinBS-seq, confirmed at the single-molecule level that this spectrum of ΔmC indeed reflects an enrichment of hemiCpGs (Fig. 4A and fig. S8, C and D). The flanking hemiCpGs adopt a conformation of rotational symmetry with respect to CTCF motifs (fig. S8E), enabling us to search for the same pattern by screening the published ChIP-seq data sets of 60 chromatin-binding proteins in H1-hESC (*20*). This pattern is only exhibited by sites co-occupied by CTCF and RAD21 (a cohesin subunit) (fig. S8F). We also determined that this pattern is contributed by 5mC more than by 5hmC (fig. S8G) (*16*). Next, we built a hemi-index (HI) to quantitatively rank all CTCF motifs by the degree to which they associate with this pattern (see supplementary materials). The CTCF motifs from the two nascent DNA duplexes are highly concordant in HI (Fig. 4B), suggesting that these hemiCpGs were inherited during DNA replication. To confirm this, the methylation frequency was compared between the two dyads in pulse and between the same dyads in pulse and chase. The inheritance of ΔmC was observed in both cases, only from the enriched type of hemiCpGs (Fig. 4C). We also confirmed the inheritance at the level of CTCF motifs by comparing their HI between (i) two dyads, (ii) pulse and chase, and (iii) two cell populations with >10 cell divisions apart (Fig. 4D).

Methyl-CpG-binding domain (MBD) proteins can bind to both mCpG and mCA (*21, 22*), suggesting that their binding to mC is not selective for the methylation state of the other strand. To investigate their putative association with hemiCpGs, we analyzed published WGBS and MBD ChIP-seq in mESC (*23*). To overcome the insufficient resolution of ChIP-seq, we profiled the

occupancy of MBD proteins at CTCF motifs showing inherited hemimethylation either upstream-only or downstream-only. Indeed, MeCP2, Mbd1a, Mbd1b, Mbd2a, and Mbd2t all showed orientation-specific colocalization with hemimethylation (Fig. 4E and fig. S9). An MeCP2 mutant in the MBD domain (R133C) prominent in Rett syndrome showed significantly reduced colocalization with hemimethylation (Fig. 4E). Interestingly, in the absence of DNA methylation, MeCP2 loses the orientation-specific occupancy with no changes in occupancy level (Fig. 4E), whereas all other MBD proteins show reduced occupancy (fig. S9), suggesting that binding of MeCP2 shifts from a hemiCpG-dependent mode to a methylation-independent mode in the absence of DNA methylation. MeCP2 physically interacts with cohesin and regulates chromatin looping (*24–26*), compelling us to investigate the relationship between hemimethylation and CTCF-mediated chromatin interactions. We first determined that hemimethylation is not significantly altered under an acute and near-complete loss of CTCF protein (fig. S10, A and B) (*27*), suggesting that the inheritance of hemimethylation is CTCF-independent. In hESC, DNMT3B-KO alone is sufficient to eliminate most of the inherited hemimethylation at CTCF motifs with minimal impact on surrounding DNA methylation (fig. S10C) (*12*). ChIP-seq revealed that DNMT3B-KO led to no changes in RAD21 occupancy and a mild increase (~1.3-fold) in CTCF occupancy at CTCF motifs (fig. S10, D and E). We then performed RAD21 HiChIP (protein-centric chromatin conformation capture) in both wild-type (WT) and DNMT3B-KO hESC and found that loss of DNMT3B causes reduced interactions emanating from these CTCF motifs, extending up to 1 Mb apart (Fig. 4F), with no changes in interaction directionality bias (fig. S10D). This suggests that loss of hemimethylation renders the CTCF/cohesin complex to a less productive state, possibly through an altered mechanism of physical interaction with MeCP2.

Our results provide temporal and strand resolution of the nascent DNA methylome, identifying hemiCpGs with distinct methylation kinetics during DNA replication. Several studies have observed hemiCpGs in cells under heterogeneous cell cycle conditions using a hairpin adaptor-based strategy (*11, 28, 29*). Our study adds the resolution of the parent-daughter axis and the dimension of replication timing, and integrates a single-molecule perspective to the understanding of hemimethylation. The efficient reoccupancy by CTCF/cohesin and inheritance of flanking hemimethylation during DNA replication, and the colocalization with MBD proteins, support a model suggesting that CTCF sites actively engaged in chromatin interactions are marked by hemiCpGs shortly after passage of the local replication forks,

which may facilitate timely assembly of the interaction complex, possibly with the involvement of MBD proteins, to ensure the proper inheritance of chromatin interactome and gene expression programs.

REFERENCES AND NOTES

1. R. Holliday, J. E. Pugh, *Science* **187**, 226–232 (1975).
2. J. A. Law, S. E. Jacobsen, *Nat. Rev. Genet.* **11**, 204–220 (2010).
3. L. S. Chuang *et al.*, *Science* **277**, 1996–2000 (1997).
4. M. Bostick *et al.*, *Science* **317**, 1760–1764 (2007).
5. J. Sharif *et al.*, *Nature* **450**, 908–912 (2007).
6. G. Liang *et al.*, *Mol. Cell. Biol.* **22**, 480–491 (2002).
7. H. Gowher, A. Jeltsch, *J. Mol. Biol.* **309**, 1201–1208 (2001).
8. A. Aoki *et al.*, *Nucleic Acids Res.* **29**, 3506–3512 (2001).
9. T. Chen, Y. Ueda, J. E. Dodge, Z. Wang, E. Li, *Mol. Cell. Biol.* **23**, 5594–5605 (2003).
10. A. Tsumura *et al.*, *Genes Cells* **11**, 805–814 (2006).
11. J. Arand *et al.*, *PLOS Genet.* **8**, e1002750 (2012).
12. J. Liao *et al.*, *Nat. Genet.* **47**, 469–478 (2015).
13. Y. He, J. R. Ecker, *Annu. Rev. Genomics Hum. Genet.* **16**, 55–77 (2015).
14. Z. Shipony *et al.*, *Nature* **513**, 115–119 (2014).
15. L. Wang *et al.*, *Cell* **157**, 979–991 (2014).
16. M. Yu *et al.*, *Cell* **149**, 1368–1380 (2012).
17. S. Jeong *et al.*, *Mol. Cell. Biol.* **29**, 5366–5376 (2009).
18. S. Sharma, D. D. De Carvalho, S. Jeong, P. A. Jones, G. Liang, *PLOS Genet.* **7**, e1001286 (2011).
19. C. T. Ong, V. G. Corces, *Nat. Rev. Genet.* **15**, 234–246 (2014).
20. ENCODE Project Consortium, *Nature* **489**, 57–74 (2012).
21. R. R. Meehan, J. D. Lewis, S. McKay, E. L. Kleiner, A. P. Bird, *Cell* **58**, 499–507 (1989).
22. H. W. Gabel *et al.*, *Nature* **522**, 89–93 (2015).
23. T. Baubec, R. Ivánek, F. Lienert, D. Schübeler, *Cell* **153**, 480–492 (2013).
24. K. D. Kernohan *et al.*, *Dev. Cell* **18**, 191–202 (2010).
25. K. D. Kernohan, D. Vernimmen, G. B. Gloor, N. G. Bérubé, *Nucleic Acids Res.* **42**, 8356–8368 (2014).
26. S. Horike, S. Cai, M. Miyano, J. F. Cheng, T. Kohwi-Shigematsu, *Nat. Genet.* **37**, 31–40 (2005).
27. E. P. Nora *et al.*, *Cell* **169**, 930–944.e22 (2017).
28. C. D. Laird *et al.*, *Proc. Natl. Acad. Sci. U.S.A.* **101**, 204–209 (2004).
29. L. Zhao *et al.*, *Genome Res.* **24**, 1296–1307 (2014).

Science

Nascent DNA methylome mapping reveals inheritance of hemimethylation at CTCF/cohesin sites

Chenhuan Xu and Victor G. Corces

Hemimethylation drives chromatin assembly

Cytosine DNA methylation is a heritable and essential epigenetic mark. During DNA replication, cytosines on mother strands remain methylated, but those on daughter strands are initially unmethylated. These hemimethylated sites are rapidly methylated to maintain faithful methylation patterns. Xu and Corces mapped genome-wide strand-specific DNA methylation sites on nascent chromatin, confirming such maintenance in the vast majority of the DNA methylome (see the Perspective by Sharif and Koseki). However, they also identified a small fraction of sites that were stably hemimethylated and showed their inheritance at CTCF (CCCTC-binding factor)/cohesin binding sites. These inherited hemimethylation sites were required for CTCF and cohesin to establish proper chromatin interactions.
Science, this issue p. 1166; see also p. 1102

ARTICLE TOOLS	http://science.sciencemag.org/content/359/6380/1166
SUPPLEMENTARY MATERIALS	http://science.sciencemag.org/content/suppl/2018/03/07/359.6380.1166.DC1
RELATED CONTENT	http://science.sciencemag.org/content/sci/359/6380/1102.full
REFERENCES	This article cites 42 articles, 8 of which you can access for free http://science.sciencemag.org/content/359/6380/1166#BIBL
PERMISSIONS	http://www.sciencemag.org/help/reprints-and-permissions

Use of this article is subject to the Terms of Service