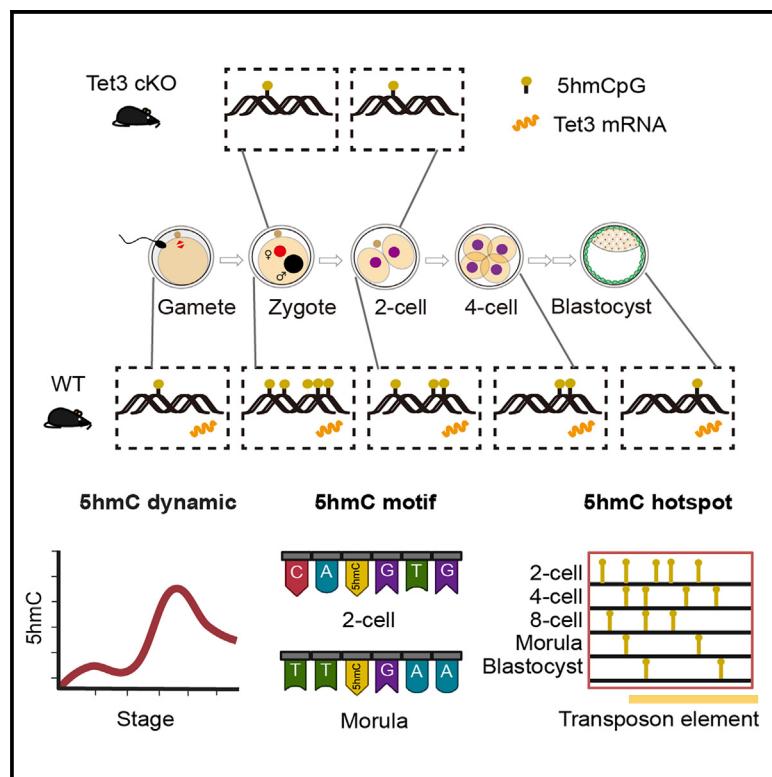


## Single-cell 5-hydroxymethylcytosine landscapes of mouse early embryos at single-base resolution

### Graphical abstract



### Authors

Dongsheng Bai, Jinmin Yang,  
Xiaohui Xue, ..., Chenxu Zhu,  
Fuchou Tang, Chengqi Yi

### Correspondence

czhu@nygenome.org (C.Z.),  
tangfuchou@pku.edu.cn (F.T.),  
chengqi.yi@pku.edu.cn (C.Y.)

### In brief

Bai et al. developed and applied schmC-CATCH for quantitative, base-resolution, genome-wide 5hmC profiling of mouse gametes and preimplantation embryos. They revealed that TET-mediated DNA hydroxymethylation initiates immediately after fertilization and identified embryonic 5hmC hotspots. This work elucidates the dynamics 5hmC landscapes during embryonic development and provides resource for epigenetic reprogramming studies.

### Highlights

- schmC-CATCH quantitatively sequences 5hmC at single-cell and single-base resolutions
- Two waves of 5hmC accumulation during zygote development
- Different developmental stages prefer varying 5hmC flanking sequences
- Embryonic 5hmC hotspots are associated with young TEs



## Resource

# Single-cell 5-hydroxymethylcytosine landscapes of mouse early embryos at single-base resolution

Dongsheng Bai,<sup>1,15</sup> Jinmin Yang,<sup>1,15</sup> Xiaohui Xue,<sup>2,3,15</sup> Yun Gao,<sup>2,15</sup> Yan Wang,<sup>2,3</sup> Mengge Cui,<sup>1</sup> Bo He,<sup>1,4</sup> Hu Zeng,<sup>1</sup> Huifen Xiang,<sup>5,6</sup> Zijian Guo,<sup>7</sup> Lan Zhu,<sup>8,9</sup> Juan Gao,<sup>10</sup> Chenxu Zhu,<sup>11,12,\*</sup> Fuchou Tang,<sup>2,4,\*</sup> and Chengqi Yi<sup>1,4,13,14,16,\*</sup>

<sup>1</sup>State Key Laboratory of Gene Function and Modulation Research, School of Life Sciences, Peking University, Beijing 100871, PRC

<sup>2</sup>Biomedical Pioneering Innovation Center, School of Life Sciences, Peking University, Beijing 100871, PRC

<sup>3</sup>Peking University-Tsinghua University-National Institute of Biological Sciences Joint Graduate Program, Academy for Advanced Interdisciplinary Studies, Peking University, Beijing 100871, PRC

<sup>4</sup>Peking-Tsinghua Center for Life Sciences, Peking University, Beijing 100871, PRC

<sup>5</sup>Department of Obstetrics and Gynecology, The First Affiliated Hospital of Anhui Medical University, Hefei 230022, PRC

<sup>6</sup>NHC Key Laboratory of Study on Abnormal Gametes and Reproductive Tract, Anhui Medical University, Hefei 230022, PRC

<sup>7</sup>State Key Laboratory of Coordination Chemistry, Coordination Chemistry Institute, School of Chemistry and Chemical Engineering, Nanjing University, Nanjing 210023, PRC

<sup>8</sup>Department of Obstetrics and Gynecology, National Clinical Research Center for Obstetric & Gynecologic Diseases, The State Key Laboratory for Complex, Severe, and Rare Diseases, Peking Union Medical College Hospital, Chinese Academy of Medical Sciences and Peking Union Medical College, Beijing 100730, PRC

<sup>9</sup>The State Key Laboratory of Common Mechanism Research for Major Diseases, Institute of Basic Medical Sciences Chinese Academy of Medical Sciences, School of Basic Medicine Peking Union Medical College, Beijing 100730, PRC

<sup>10</sup>Shanghai Institute of Biochemistry and Cell Biology, Center for Excellence in Molecular Cell Science, Chinese Academy of Sciences, Shanghai 200031, PRC

<sup>11</sup>New York Genome Center, New York, NY 10013, USA

<sup>12</sup>Department of Physiology and Biophysics, Institute for Computational Biomedicine, Weill Cornell Medicine, New York, NY 10065, USA

<sup>13</sup>Department of Chemical Biology and Synthetic and Functional Biomolecules Center, College of Chemistry and Molecular Engineering, Peking University, Beijing 100871, PRC

<sup>14</sup>Beijing Advanced Center of RNA Biology (BEACON), Peking University, Beijing 100871, PRC

<sup>15</sup>These authors contributed equally

<sup>16</sup>Lead contact

\*Correspondence: czhu@nygenome.org (C.Z.), tangfuchou@pku.edu.cn (F.T.), chengqi.yi@pku.edu.cn (C.Y.)

<https://doi.org/10.1016/j.celrep.2025.115520>

## SUMMARY

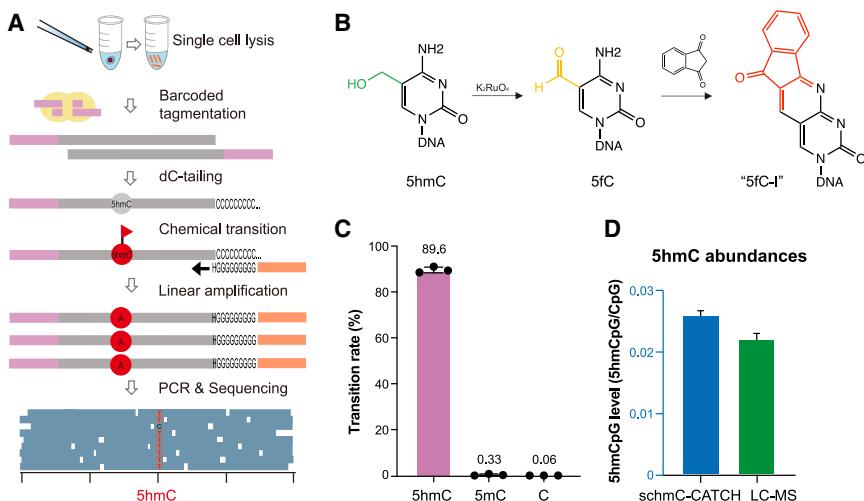
DNA methylation and hydroxymethylation are extensively reprogrammed during mammalian early embryogenesis, and studying their regulatory functions requires comprehensive DNA hydroxymethylation maps at base resolution. Here, we develop single-cell 5-hydroxymethylcytosine (5hmC) chemical-assisted C-to-T conversion-enabled sequencing (schmC-CATCH), a method leveraging selective 5hmC labeling for a quantitative, base-resolution, genome-wide landscape of the DNA hydroxymethylome in mouse gametes and pre-implantation embryos spanning from the zygote to blastocyst stage. We revealed that, in addition to late zygotic stages, onset of ten-eleven translocation (TET)-mediated DNA hydroxymethylation initiates immediately after fertilization and is characterized by the distinct 5hmC patterns on the parental genomes shaped by TET3 demethylase. We identified persistent clusters of 5hmC hotspots throughout early embryonic stages, which are highly associated with young retroelements. 5hmC is also associated with different regulatory elements, indicating a potential regulatory function during early embryogenesis. Collectively, our work elucidates the dynamics of active DNA demethylation during mouse preimplantation development and provides a valuable resource for functional studies of epigenetic reprogramming in early embryos.

## INTRODUCTION

5-methylcytosine (5mC) can be sequentially oxidized to 5-hydroxymethylcytosine (5hmC), 5-formylcytosine (5fC), and 5-carboxycytosine (5caC) by ten-eleven translocation (TET)-family proteins; the involvement of base excision repair promotes the latter two to be reversed back into unmodified cyto-

sines, constituting the TET-mediated active DNA demethylation paradigm. During mammalian embryogenesis, parental DNA methylation is extensively reprogrammed.<sup>1–3</sup> Both the paternal and maternal genomes undergo active demethylation, coupled with 5hmC accumulation and passive dilution through early embryonic cleavages,<sup>4–6</sup> while the maternal genome is protected by Stella and undergoes limited TET3-mediated demethylation.<sup>7,8</sup>





**Figure 1. Concept and validation of schmC-CATCH for single-cell, base-resolution 5hmC profiling**

(A) The schmC-CATCH workflow.  
(B) Scheme for the two-step chemical-assisted selective labeling of 5hmC. 5fC-I represents the final labeling adduct and is read as “T” in sequencing.

(C) Conversion rate on schmC-CATCH-treated synthetic spike-ins. The efficiency and specificity of chemical treatment on 5hmC, 5mC, and unmodified C are shown ( $n = 3$ ). Data are represented as mean  $\pm$  SD.  
(D) The DNA hydroxymethylation level in mESCs, quantified by whole-genome schmC-CATCH (single cell, left) and MS (bulk sample, right; Song et al., 2013).<sup>22</sup> Data are represented as mean  $\pm$  SD.

Early studies have characterized the methylation status of gametes and early embryos<sup>1–3</sup>; however, as the main product of active demethylation,<sup>9</sup> a whole-genome, quantitative picture of the DNA hydroxymethylation map across early embryonic stages is still needed to understand the role of DNA methylation dynamics during this process. Because bisulfite sequencing cannot distinguish between 5mC and 5hmC,<sup>10</sup> efforts have been made in recent years to specifically detect 5hmC at single-base resolution as well as at the single-cell level.<sup>11–20</sup> Yet, current methods for global profiling of 5hmC lack the sensitivity or quantitative ability to investigate preimplantation embryos across different stages at the single-cell level. Therefore, the global and high-resolution pattern of active DNA demethylation during early embryogenesis remains an open question.

To address this, we develop schmC-CATCH (single-cell 5hmC chemical-assisted C-to-T conversion-enabled sequencing), a single-cell 5hmC sequencing method based on selective chemical transition of 5hmC and subsequent 5hmC-to-T during amplification and sequencing, in combination with highly efficient genome amplification and library preparation. Using schmC-CATCH, we obtain a quantitative, genome-wide landscape of the DNA hydroxymethylome in wild-type mouse gametes and preimplantation embryos from zygote (pronuclear stage 1–5, PN1–5) to the blastocyst stage and Tet3-depleted zygote and 2-cell embryos. Our work provided an allele-specific DNA hydroxymethylation map with unprecedented resolution and revealed the dynamics of active DNA demethylation in mouse preimplantation embryos.

## RESULTS

### A single-cell, base-resolution method for quantitative and whole-genome 5hmC analysis

To address the challenges in dissecting single-cell 5hmC patterns, we sought to come up with a quantitative and genome-wide profiling method. We developed schmC-CATCH, an assay combining bisulfite-free chemical labeling with a tailored library construction process for highly sensitive and precise 5hmC detection (Figure 1A).<sup>18</sup> Previously, we have developed a bisul-

fite-free approach for 5hmC labeling and pull-down based on potassium ruthenate-mediated oxidation and subsequent Friedländer labeling.<sup>18</sup> Because the chemistry is very mild and efficient, we envisioned that it would be possible to further establish a reaction condition for single-cell 5hmC analysis. A major challenge in our previous 5hmC method is that the labeling chemical 1,3-indandione precipitates, leading to the loss of DNA. This could be particularly deleterious for single cell 5hmC analysis. Hence, we carefully optimized the reaction condition and managed to identify an unsaturated 1,3-indandione solution that does not cause DNA loss. In addition, our chemistry is bio-compatible with DNA amplification; no purification step is needed, preserving the single cell genomic DNA. Moreover, compared to malononitrile-mediated 5fC labeling, whose efficiency is suboptimal,<sup>21</sup> we achieved nearly complete labeling efficiency under this new reaction condition, allowing quantitative 5hmC analysis (Figures 1B and S1A). Last, the bisulfite-free strategy enables subtraction-free 5hmC readout, which avoids unacceptable noise signals and template degradation for single-cell analysis.

A second key step in the design of schmC-CATCH is the library preparation process. Tn5 tagmentation is a popular strategy for single-cell genomic analysis; however, only about one-half of genomic DNA is transformed into a library with the conventional strategy. To improve the genome coverage, we integrated homo-adaptor Tn5 tagmentation with efficient single-strand DNA ligation to achieve an efficient, one-pot library preparation step (Figures 1A and S1B). We also predesigned cell-specific DNA barcodes on the homo-adaptor; hence, the preindexed single-cell samples can be pooled together for multiplexed reactions. Finally, both strands of the tagmented DNAs were subjected to efficient single-stranded DNA library preparation, which has been well established previously (Figure 1A).

To further validate the designed approach at the single-cell level, we sought to evaluate single-cell libraries from cultured mouse embryonic stem cells. For trace amounts of genomic DNA from single cells, we found that carrier RNA substantially relieved template dropout during buffer exchange and that a linear amplification before exponential PCR is beneficial for

library uniformity. Because of the this design, schmC-CATCH achieved a high mean coverage of 2.3 million CpG sites per cell ( $\geq 1 \times 2.3$  million,  $\geq 3 \times 2.1$  million,  $\geq 5 \times 1.9$  million) (Table S1), which represents 9.40%–13.67% in the whole genome. We observed a significantly higher mapping rate (>95%;, Figure S1C) for schmC-CATCH compared with bisulfite sequencing.<sup>23–27</sup> Using a stringent criterion, we identified about 25,000 5hmCpG sites at single-base resolution from mouse embryonic stem cells (mESCs) with a low false discovery rate (FDR) (0.33% for 5mC and 0.06% for unmodified C) and average conversion rate of 89.6% for 5hmC (Figure 1C). In addition, the 5hmC level obtained using schmC-CATCH data recapitulates liquid chromatography-mass spectrometry (MS)-based results (Figure 1D), demonstrating the quantitative power of schmC-CATCH. Furthermore, we found that the aggregated single-cell profile correlates well with the quantitative bulk 5hmC dataset (Pearson  $r = 0.6$ )<sup>6</sup> (Figure S1D).

### Mapping 5hmC landscapes in mouse gametes and early embryos

Next, we sought to generate comprehensive 5hmC profiles of mouse gametes and early embryonic cells to dissect the role of active demethylation in preimplantation stages. We collected sperms, MII (Metaphase II) oocytes, and early embryos, including zygotes; 2-cell, 4-cell, and 8-cell embryos; morula; and blastocysts by crossing female C57BL/6J mice with male DBA/2N mice (Figure 2A). Previous studies have revealed TET3 is the major demethylase during early mouse embryonic development.<sup>28,29</sup> To further validate the function of this enzyme during this process, we additionally generated Tet3 conditional knockout (cKO) mice and obtained schmC-CATCH datasets for zygotes and 2-cell embryos. All samples were extensively washed to remove any somatic, gametic, or polar body contaminants. In total, we performed 5hmC analysis for 537 individual cells from gametes and preimplantation embryos. Each individual cell generated, on average, 4.04 Gb of sequencing data (clean reads, 29.34 million) and identified 702–61,764 5hmCpG sites (Table S2), which is in excellent agreement with the absolute levels of 5hmC measured by ultra-performance liquid chromatography-mass spectrometry (Figure S1E).<sup>4</sup>

The average hydroxymethylation profiles were then plotted across genes and flanking intergenic regions (Figure 2B). As expected, for both wild-type and Tet3-depleted embryos, we observed a drop in 5hmC signal at the transcription start site (TSS); interestingly, we found a strong 5hmC signal just downstream of the transcription end site (TES), which is different from the 5hmC profiles in bulk mESCs and human tissue. Significantly, unusual strong open chromatin at the TES region is also observed with chromatin accessibility analysis.<sup>30</sup> Zygotic gene activation (ZGA) is the critical event that governs the transition from maternal to embryonic control of development. In the mouse, ZGA occurs during the 2-cell stage with genome-wide gene expression. By dividing genes into three groups according to their RNA expression level and plotting the average 5hmC profile along gene bodies, we discovered an enrichment of 5hmC in and around gene bodies of more highly expressed genes but more depletion in TSS regions (Figure S2A). Next, we performed pseudotime analysis based on the 5hmC genome-wide distribu-

tion and constructed the developmental trajectory of mouse early embryos (Figure 2C). As a dynamic modification, we found that single-cell 5hmC profiles have the ability to mark continuous developmental stages (Figures S2B–S2E). Interestingly, the zygotes and 2-cell embryos collected from Tet3 cKO mice show higher similarity to oocytes from wild-type mice, confirming that this demethylase is responsible for shaping the 5hmC landscape during this process (Figures 2B, 2C, and S2E). In a representative view, continuous stages of early embryo displayed quite different 5hmC patterns at single-base resolution, and depletion of Tet3 diminished the prominent accumulation of 5hmC after fertilization (Figure 2D). Taken together, these data suggest that schmC-CATCH has the power to examine the global 5hmC patterns in early mouse development.

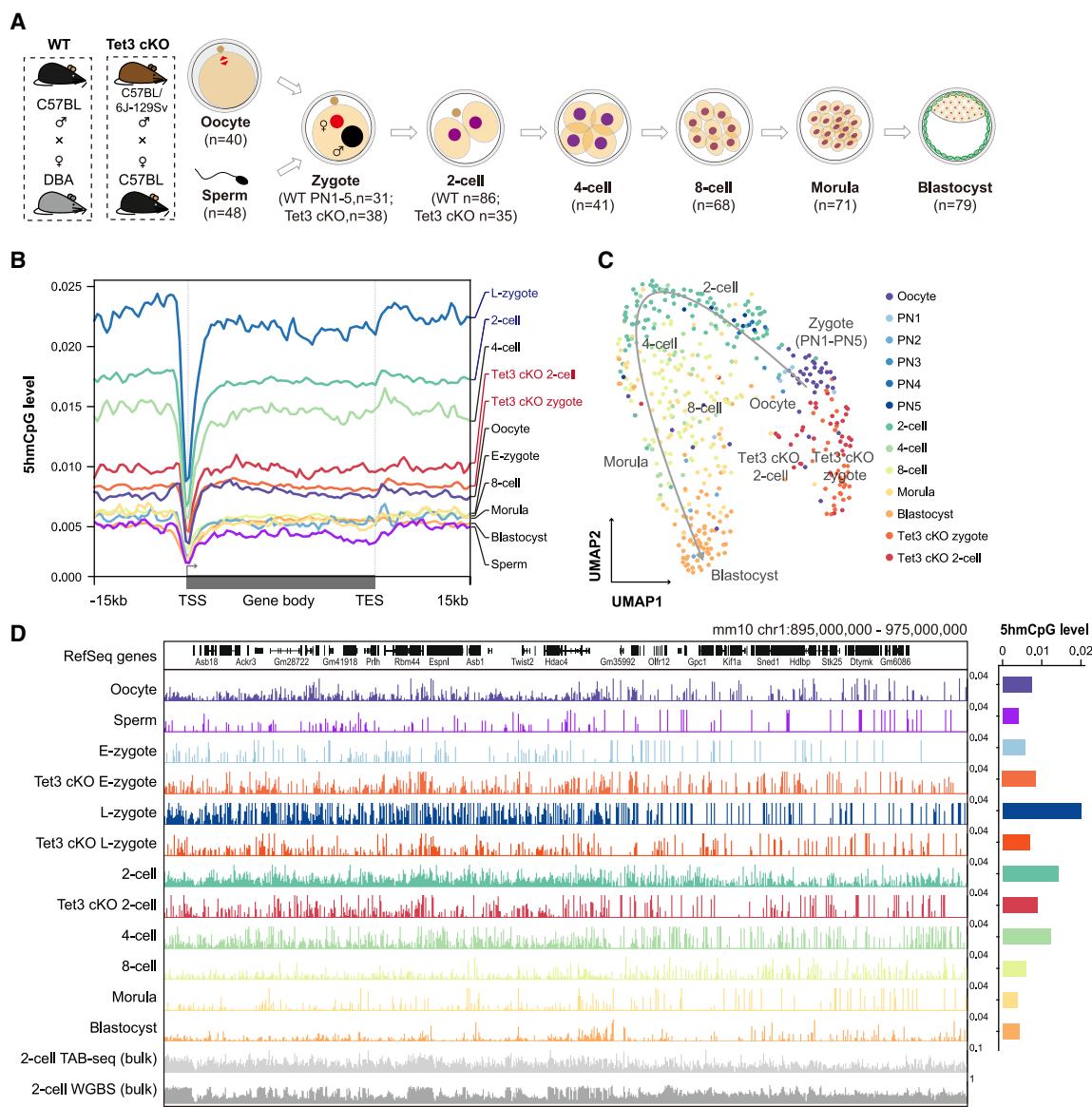
### Two waves of 5hmC accumulation during zygote development

As the major product of TET-driven 5mC oxidation, 5hmC is recognized as a hallmark of active DNA demethylation. We identified allelic 5hmC sites using SNPs belonging to the two parental strains.<sup>31</sup> During the zygotic stage, the DNA methylation levels of the pronuclei decrease as they moved closer to each other.<sup>32</sup> To elucidate the beginning of TET-driven active demethylation with higher temporal resolution, we collected and profiled PN1–PN5 zygotes throughout the pronucleus stages to track the refined pattern of 5hmC kinetics. Interestingly, we observed a first wave of 5hmC accumulation immediately at the PN1 zygotic stage, hinting that TET proteins quickly bind and oxidize inherited 5mC once fertilization occurs (Figure 3A). Such TET3-dependent active demethylation of the paternal genome could be functional, given that DNA demethylation-induced DNA lesions have been shown to signal the cell -cycle machinery and contribute to the checkpoint activation.<sup>33</sup>

A stronger wave of paternal 5hmC accumulation was detected at the late zygote stage (PN3–PN5). It supports the theory that generation of 5hmC is highly coupled to DNA methylation maintenance during S phase,<sup>34</sup> in which TET proteins and DNA methyltransferases (DNMTs) are paradoxically co-expressed and tend to achieve rapid DNA methylation turnover (Figure S3A). Thus, our results revealed two consecutive waves of 5hmC accumulation for the paternal genome, at fertilization and genome replication, which may play different roles. As for the maternal genome, there is no significant 5hmC accumulation during the zygotic stages (Figure 3A).

### Parental genomes confer distinct hydroxymethylation patterns

It is well known that, at the end of the zygote stage, the 5hmC level on the paternal genome is much higher than that of the maternal genome, as seen by immunofluorescence and MS analysis using separated pronuclei. Nevertheless, it is difficult to track parental hydroxymethylation after pronucleus fusion. By sequencing and separating reads into parental genomes, we were able to describe the parental DNA hydroxymethylation across consecutive developmental stages. First, 5hmC can be detected in sperm and MII oocytes, showing comparable levels (Figure 3B). Second, upon fertilization, we observed dramatic differences in hydroxymethylation abundance and dynamics



**Figure 2. Landscapes of DNA hydroxymethylome in early mouse embryos**

(A) Samples isolated for single-cell hydroxymethylation analysis with replicate numbers. PN1–PN5 zygotes were included in our analysis, and the number of cells collected from each stage is shown.

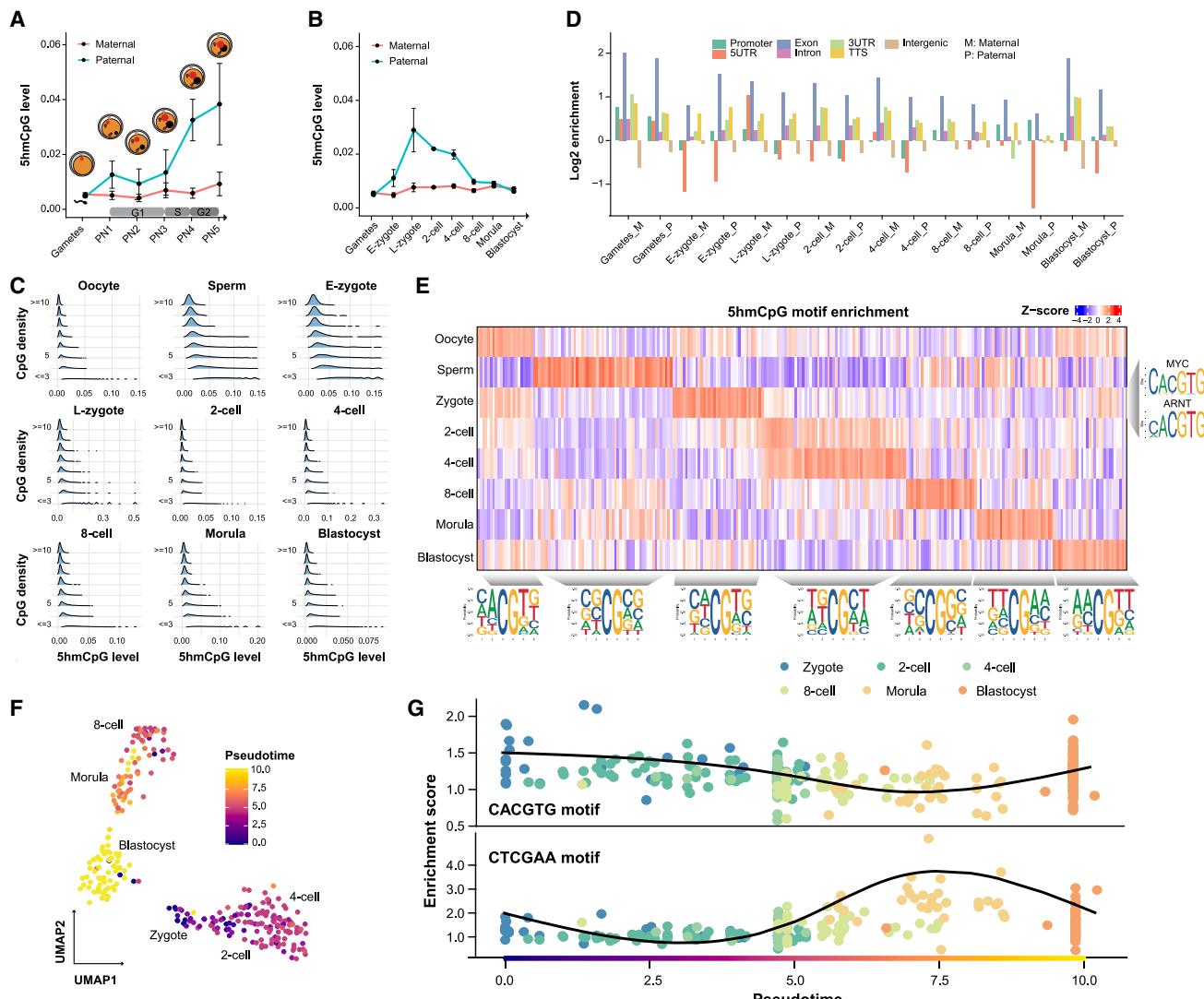
(B) Averaged DNA hydroxymethylation levels along the gene bodies and 15 kb upstream of the transcription start sites (TSSs) and 15 kb downstream of the transcription end sites (TESs) of all RefSeq genes.

(C) Uniform Manifold Approximation and Projection (UMAP) of the genome-wide 5hmCpG distribution pattern of mouse gametes and early embryos. Each dot represents one cell.

(D) Representative genome browser view showing the modification fraction of 5hmC sites in each developmental stage. The right side of the bar plot shows the average 5hmCpG level for each. Bulk 2-cell Tet-assisted bisulfite sequencing dataset: GSM1386028; bulk 2-cell whole-genome bisulfite sequencing dataset: GSM1386021.

for the paternal and maternal genomes, and this differential hydroxymethylation level exists until the 8-cell stage. Consistent with previous evidence,<sup>5,35</sup> the paternal genome is preferentially demethylated with extensive 5hmC accumulation across stages, while limited activity of TET enzymes is observed on the maternal genome, with only slight fluctuation of 5hmC level from gamete to blastocyst stage (Figure S3B).

From the PN5 zygote (4N) to the 2-cell embryo (2N) stage, there is no DNA replication and, hence, no passive dilution of DNA modifications. Yet, there is a significant 5hmC reduction after the first embryonic cleavage, suggesting that 5hmC is subjected to further oxidation and removal, which is supported by the observation that no significant increase in global 5hmC levels is present upon Tet3 depletion in late zygotes and 2-cell embryos (Figure S3C).



**Figure 3. Paternal and maternal genomes confer distinct hydroxymethylation features**

- (A) Averaged DNA hydroxymethylation levels of individual male and female pronuclei of zygotes at different time points. PN1–PN2, G1 phase before DNA replication; PN3–PN4, S phase during DNA replication; G2 phase, G2 phase after DNA replication. Data are mean  $\pm$  95% confidence interval.
- (B) Hydroxymethylation landscape across each stage of early mouse embryos. Early zygotes represent PN1 and PN2 zygotes, and late zygotes represent PN3–PN5 zygotes. Data are mean  $\pm$  95% confidence interval (CI).
- (C) Distribution of hydroxymethylation values across local CpG densities for early embryonic stages.
- (D) Relative enrichment analysis of 5hmCpG sites located in different genomic elements on parental genomes. M, maternal genome; P, paternal genome.
- (E) Heatmap showing 5hmCpG motif enrichment across developmental stages. Each column represents a motif, and each row corresponds to an embryonic stage.
- (F) UMAP showing the single-cell trajectory of early embryo development based on genome-wide 5hmCpG motif enrichment.
- (G) Scatterplots showing the 5hmCpG enrichment at CACGTG (top) and CTCGAA (bottom) motifs for each cell across pseudotime. Each dot represents a single cell and is colored according to the pseudotime score.

Notably, compared with the first cleavage, there is less reduction of hydroxymethylation from the 2-cell to the 4-cell embryo stage. Considering the presence of passive dilution during this process, there ought to exist newly generated 5hmC via active DNA demethylation. It's reasonable to attribute the DNA demethylation activity during the first two cleavages to the high level of TET3 (Figure S3A). From the third cleavage, paternal 5hmC is gradually reduced to the same level as the maternal genome.

Active demethylation occurs at almost all types of genomic elements, yet there are limited 5hmC sites on CpG islands (CGIs) with high CpG densities (Figure S3D). Gametes and early embryos show a clear inverse correlation between CpG densities and hydroxymethylation levels (Figure 3C), which is stronger when compared to that of DNA methylation<sup>2</sup>; this pattern is independent of local 5mCpG modification levels (Figure S3E). We next separately analyzed the genomic distribution of 5hmC on

the two parental genomes. While the overall abundance of 5hmC is much higher for the paternal genome, the genomic distribution pattern of 5hmC appears to be less distinct compared to the maternal genome, except at the blastocyst stage (Figures 3D and S3F). However, during the zygote stage, hydroxymethylation is more accumulated for intragenic regions including CGIs ( $p < 0.0001$ ) and 5' UTR (5' untranslated regions;  $p < 0.05$ ) on the maternal genome despite limited overall changes between these two stages, hinting at a potential regulatory role. In addition, we observed similar hydroxymethylation dynamics on both of the maternal and paternal CGIs and 5' UTR despite their distinct overall modification levels (Figure S3D).

### Temporal precedence of CpG flanking bases in active DNA demethylation

The distinct 5hmC distribution pattern prompted us to explore whether TET exhibits differential sequence preferences during embryonic development. We calculated the average 5hmCpG modification levels for all 256 sequence compositions (NNCGNN) and observed the local base sequences enriched for each developmental stage (Figure 3E). For example, oocyte 5hmC is enriched for NACGTN, while NCCGGN is highly hydroxymethylated in 8-cell embryos. Interestingly, we found that, in zygotes, the most 5hmC-modified 6-mer sequence is CACGTG, coinciding with the core binding motif of MYC and ARNT (Figure 3E). A previous study found that c-MYC is expressed in mouse oocytes and early preimplantation embryos,<sup>36</sup> while the core binding sequences of basic-helix-loop-helix-family transcription factors (TFs; such as c-MYC) are generally hypomethylated<sup>37</sup>; thus, TET-mediated active DNA demethylation may prioritize the removal of 5mC from c-MYC binding sequences in zygotes to activate developmental genes.

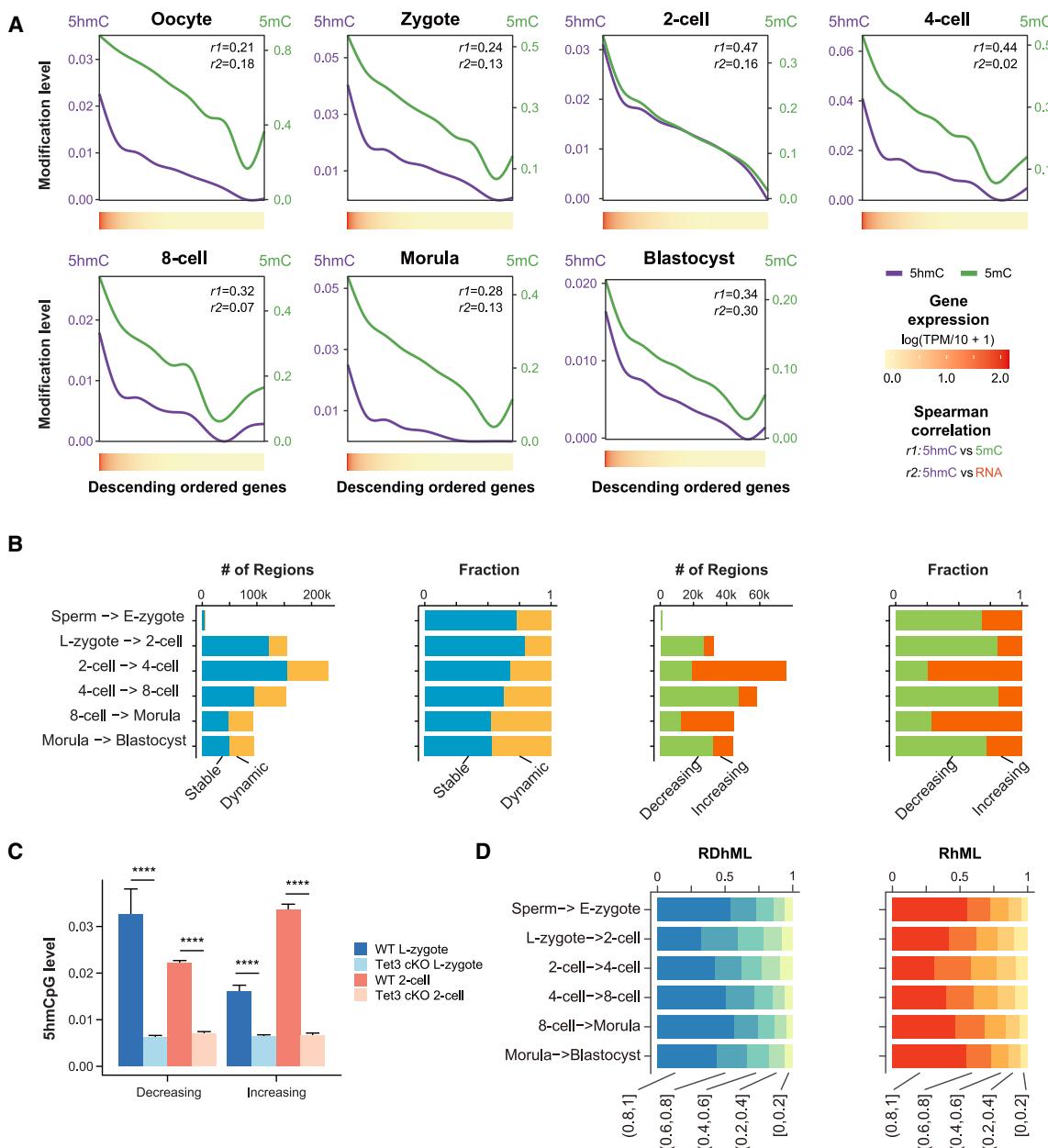
Next, we further explored the 5hmC modification levels of different CG flanking sequences in individual cells. Cell clustering based on the hydroxymethylation levels of NNCGNN motif could clearly separate embryos from different stages (Figure S3I). We then performed pseudotime analysis based on 5hmCpG 6-mer enrichment and identified the CG motifs displaying dynamic modification levels (Figures 3F and 3G). For example, we again observed that the CACGTG motif is highly modified in zygotes and 2-cell embryos and decreased in later stages. Another example is the CTCGAA motif, which displays high 5hmCpG levels in the morula and blastocyst but not in earlier stages. This core sequence motif can be recognized by KLF4, a known DNA methylation reader.<sup>38</sup> Klf4 is highly expressed in mESCs and quickly downregulated upon differentiation.<sup>39</sup> The active demethylation of CT5mCGAA could promote embryonic development by rapidly deactivating KLF4 targets before reducing the abundance of this TF protein. Collectively, the dynamic hydroxymethylation on CpGs with different flanking sequences, shaped by the generation and further oxidation of 5hmC by TET, could enhance the developmental gene regulation programs.

### Dynamics of active DNA demethylation in early embryonic development

Early works provided evidence of high DNMT1 maintenance activity and limited *de novo* DNMT3A/B methylation activity in early

embryos.<sup>34</sup> Published whole-genome bisulfite sequencing data in mouse preimplantation embryos identified the capacity to maintain methylation during the global reprogramming.<sup>40</sup> Interestingly, we observed a significant positive correlation of 5hmC signals with 5mC in gene bodies during the mouse early developmental stages (Figure 4A). This could be explained by the hypothesis that 5hmC may be mainly derived from newly deposited 5mC by DNMT1/DNMT3.<sup>4</sup> Consistent with the results from mESCs and tissues, we observed a positive correlation of gene body 5hmC signals with gene activities.<sup>41–43</sup> To further explore the relationships between 5mC and 5hmC in gene regulation, we identified 616 and 196 genes showing positive and negative relationships, respectively (Figure S4A). The negatively correlated genes tend to have higher numbers of genic CpGs and show the highest 5hmCpG modification levels in 2-cell embryos and blastocyst, coinciding with the two stages with the highest relative 5hmCpG levels when normalized by local 5mCpG levels (Figures S2B and S4B–S4D). Interestingly, for 2-cell embryos, the promoters and genic regions of negatively correlated 5mCpG–5hmCpG genes show higher enrichment of RNA polymerase II (RNA Pol II), H3K4me3, and H3K36me3 chromatin immunoprecipitation sequencing (ChIP-seq) signals, corresponding to active transcription, while they are depleted from the repressive H3K27me3 histone modification (Figures S4E and S4F). Thus, TET-mediated active DNA demethylation, by leaving 5hmC modification, may prevent the corresponding sites from being considered as targets of both *de novo* and maintenance methylation. This is synergistic with the dramatic global DNA demethylation, which is required for transcription activation during this period.<sup>4</sup>

Current *in vivo* evidence from preimplantation embryos suggests that 5hmC is passively removed through replication.<sup>5</sup> We next searched for several possibilities for the fate of DNA hydroxymethylation. Pre-existing 5hmC can be further oxidized to be eventually removed by base excision repair (BER) or diluted by DNA replication.<sup>44,45</sup> To distinguish between these two possibilities, we compared the hydroxymethylation levels between each pair of consecutive embryonic stages (Figure 4B). The results showed that, for the first and third cleavage (late zygote > 2 cells and 4 cells > 8 cells), the latter stage had more decreasing differentially hydroxymethylated regions (DhMRs) than the former stage, which is consistent with the global dynamics of hydroxymethylation (Figure 3B). In order to study the magnitude of 5hmC variations between stages in more detail, we adopted two concepts: relative de-hydroxymethylation level (RDhML) and relative hydroxymethylation level (RhML). Taking the pair of 2-cell > 4-cell stage as an example, RDhML represents  $(5\text{hmC}_{2\text{-cell}} - 5\text{hmC}_{4\text{-cell}}) / 5\text{hmC}_{2\text{-cell}}$ , and RhML stands for  $(5\text{hmC}_{4\text{-cell}} - 5\text{hmC}_{2\text{-cell}}) / 5\text{hmC}_{4\text{-cell}}$ . It's obvious that a large fraction of the 5hmC-increasing regions had RhMLs in the range of (0.6, 1] across stage pairs, which means those 5hmC-increasing regions may be newly oxidized by TETs in the later embryonic stage (Figures 4B–4D). 5hmC is known not to be maintained by DNMT1,<sup>46</sup> so the difference in 5hmC levels between pairs of adjacent stages is affected by genome replication; i.e., if there is no new generation or further oxidation, then the level of 5hmC in the latter embryonic stage should be half of that of the former stage. We observed a varying fraction of 5hmC-decreasing regions showing more than 50% of



**Figure 4. Dynamics of active DNA demethylation across early embryonic stages**

(A) The variation trend and correlation of 5hmC and 5mC modification levels in the gene body, ranked by the decrease of RNA expression across the early development process.

(B) Stacked bar plots showing the pairwise compared 5 kb tiles between consecutive developmental stages. Tiles with no significant changes are defined as stable tiles and otherwise as dynamic tiles (fold change  $>1$ ,  $p < 0.05$ ). The dynamic tiles are further divided into increasing and decreasing tiles according to whether the 5hmCpG level of the previous stage is lower or higher than that of the next stage. Shown from left to right are the absolute number of stable and dynamic 5 kb tiles and their relative fractions and the absolute number of increasingly or decreasingly hydroxymethylated 5 kb tiles and their relative fractions.

(C) The 5hmCpG level of WT and Tet3 cKO embryos in dynamic DhMRs. Wilcoxon rank-sum tests were performed between WT and Tet3 cKO embryos at the corresponding stage (ns,  $p > 0.05$ ; \* $p < 0.05$ ; \*\* $p < 0.01$ ; \*\*\* $p < 0.001$ ; \*\*\*\* $p < 0.0001$ ). Data are mean  $\pm$  SE.

(D) The fold change of 5hmCpG and 5mCpG in 5-kb tiles between adjacent stages. The tiles in red and blue indicate the increasing and decreasing DhMR in (B), respectively. Higher RDhML values indicate a greater extent of decrease. Similarly, higher RhML values represent a greater extent of increase.

5hmCpG loss across the development stages (varying from 69.64% during 2-cell  $>$  4-cell to 80.87% during 8-cell  $>$  morula), which means that, in the majority of the 5hmC-decreasing regions,

5hmC is unstable and can be further oxidized to 5fC and 5caC, followed by BER (Figures 4D and S4G).<sup>45</sup> The regions showing evident further oxidation of 5hmC (RDhML: [0.5, 1]) displayed

higher enrichments in CGIs, while is less enriched in promoters, exons, and 3' UTR regions (Figure S4H). These regions are also enriched for bivalent promoters and poised enhancers, similar to the previous observations in mESCs.<sup>11</sup> Collectively, these results show that TET-driven active DNA demethylation safeguards specific genomic regions from hypermethylation during the first several cleavages and that, as a transient intermediate of TET oxidation, 5hmC can be actively removed by further oxidation with a short lifetime or passively diluted by DNA replication.

### 5hmC hotspots associated with young TEs in early embryos

Repeat elements have been shown to play a role in gene regulation of early embryonic development.<sup>47</sup> Enrichment analysis showed that 5hmCpG sites were enriched in specific repeat elements (Figures S3G and S3H), and we further investigated DNA hydroxymethylation on transposable elements (TEs). We found that the evolutionary age of TEs seems to determine the DNA hydroxymethylation status; evolutionary younger TEs, including L1, Alu, and ERVK, show higher 5hmC levels upon fertilization and throughout early development for the maternal genome, while old TEs, including L2, MIR, and ERVL, exhibit relatively lower 5hmC levels (all  $p < 0.003$ ). For the paternal genome, young TEs showed consistently higher 5hmC levels after the 8-cell stage (all  $p < 0.003$ ). Especially the sharp increase of L1 5hmC on both parental genomes at the early zygote stage indicates that there is high TET activity on L1 shortly after fertilization. Previous studies demonstrate that L1 expression is critical for early embryogenesis<sup>47,48</sup>; we observed that DNA hydroxymethylation level coordinates with L1 expression from 2-cell embryo to blastocyst stage (Figure 5A), supporting a regulatory role of 5hmC in L1 expression. We also observed a similar pattern for long terminal repeats (LTRs) and SINEs (Figure 5A). Besides, both paternal and maternal 5hmC share more similar abundance levels and more consistent dynamic patterns on younger TEs across stages, which is quite different from the old TEs (Figure 5A).

When we used Integrated Genome Viewer to scan 5hmC signals at various embryonic stages, we found that there seem to be conserved 5hmC signals in early development. After a series of attempts, we identified nearly 2,500 5hmCpG-rich genomic loci of 1-kb length, which we termed “hotspots of DNA hydroxymethylation” (Figures 5B, 5D, 5E, and S4I). The 5hmC hotspots are enriched in H3K9me3-associated heterochromatin (Hc-H), and we assumed that these regions can endure consistent DNA hydroxymethylation due to non-repressive H3K9me3 and consistent DNA methylation acquisition in early developmental stages (Figure 5C).<sup>49,50</sup> Notably, these 5hmC hotspots reside primarily in promoters of the TEs LINE and LTRs, especially the younger ones, such as L1 and ERVK (Figure 5E). In addition, we found that these 5hmC hotspots significantly colocalize with stable H3K9me3 peaks (chi-square test with CpG density-controlled 1-kb tiles as a control, odds ratio = 2.0,  $p = 1.5e-11$ ), which is non-repressive in early developmental stages<sup>49</sup> (Figure 5D); the correlation is much weaker or even depleted in ESCs and somatic cells, in which H3K9me3 acts as a repressive mark.<sup>43,51</sup> Previous studies have found that the sperms are hypermethylated in some specific families of LINE

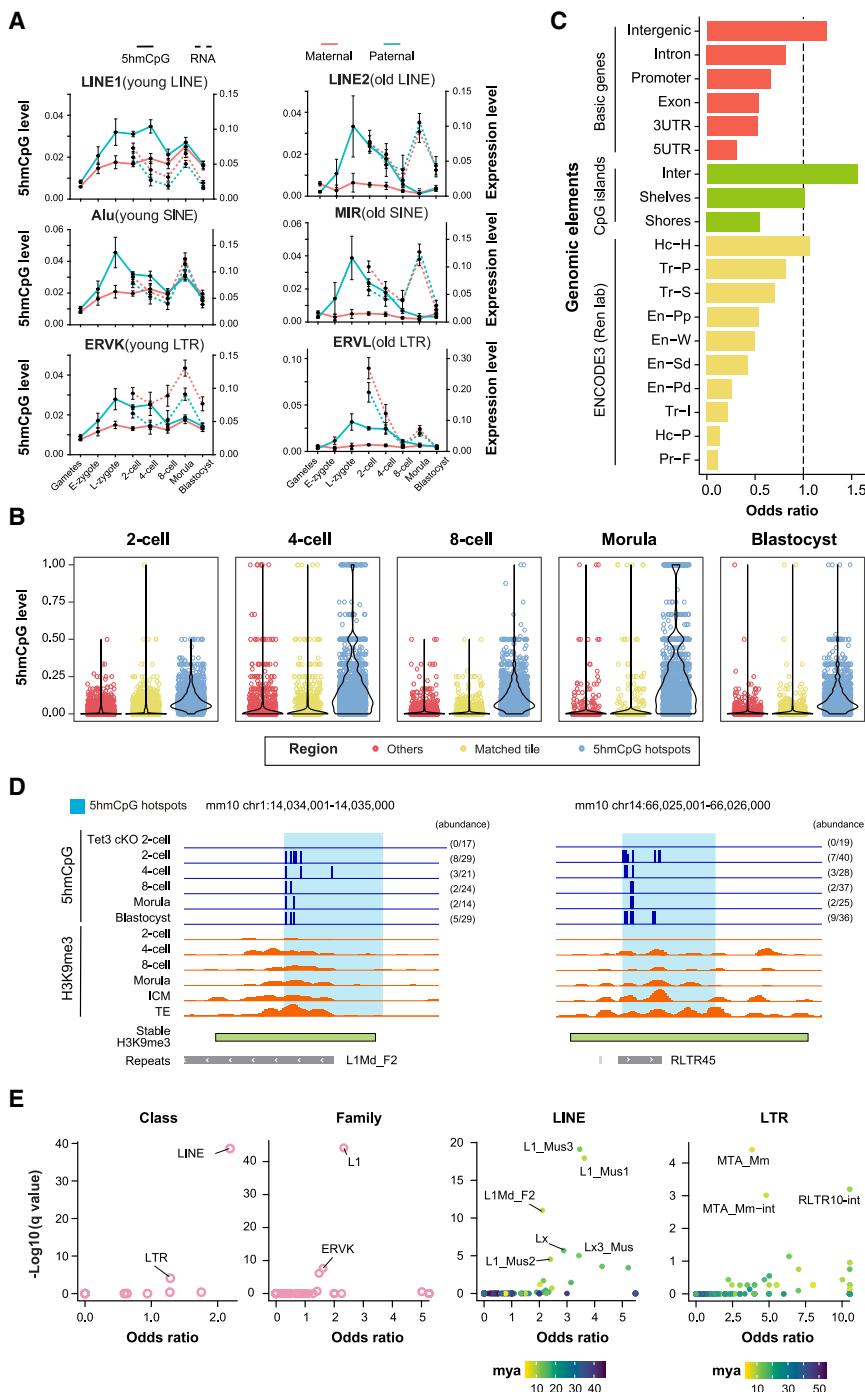
and LTR TEs; these sperm-specific hypermethylated elements undergo dramatic DNA demethylation after fertilization.<sup>2</sup> Our result shows that such an active DNA demethylation event occurs continuously during the preimplantation embryo development period.

### Involvement of TET3-mediated DNA hydroxymethylation in early development

Because DNA hydroxymethylation tends to exist in gene bodies, promoters, and enhancers, we then investigated a potential regulatory role of 5hmC in early development. Previous studies have defined different sets of genes whose promoters are marked with differential histone modifications<sup>52</sup>; thus, we analyzed the hydroxymethylation levels for these genes. We found that, at the 2-cell stage, H3K4me3-marked promoters and bivalent promoters exhibit strong depletion of 5hmC around the TSS regions, which is not the case for H3K27me3-marked promoters or unmarked promoters (Figure 6A). However, there is higher 5mC-5hmC turnover ratio at active and bivalent promoters (Figure S5A). Variation in 5hmC modification levels of wild-type (WT) 2-cell at the TSS region among different classes of promoter were significantly higher than those of Tet3-cKO 2-cell and WT MII oocytes (Figures S5B and S5C; WT 2-cell vs. Tet3-cKO 2-cell:  $p < 0.001$ ; WT 2-cell vs. WT oocyte:  $p < 0.001$ ), suggesting involvement of TET3-mediated active DNA demethylation during early embryonic stages. Next, we asked whether the association of 5hmC with active histone marks observed in promoter regions also presents in distal regulatory elements. We found that both promoter-proximal and -distal regions enriched for 5hmC are modified with the active histone mark H3K27ac, indicating that 5hmC may be an epigenetic mark for active regulatory elements (Figure 6B). We also observed that the enrichment of the H3K27ac signal is positively correlated with the average local 5hmCpG modification levels (Figure S5D). These results agreed with the active regulatory role of 5hmC in early embryos.<sup>53</sup>

Besides histone marks, CpG density also affects promoter hydroxymethylation and gene expression. To investigate this issue, we stratified genes into three groups based on the CpG densities in their promoters: high-CpG promoter genes, intermediate-CpG promoter genes, and low-CpG promoter genes. There is newly generated TET3-dependent 5hmC, especially in the zygote paternal genome, on promoters with intermediate to low CpG density, regardless of the local 5mC abundance (Figures 6C and S5E). To predict the potential biological functions that involve 5hmC modification on promoters, we identified the Gene Ontology (GO) terms that are enriched for the genes with 5hmC sites (Figures S5F–S5H), especially for morula, which is associated with cell fate decision,<sup>54</sup> including histone modification, regulation of cell development, cell-cell signaling, and cell fate commitment (Figure 6D).

As regulatory regions are hotspots of binding sites of TFs, which are known to be recruited or inhibited by the presence of DNA modifications,<sup>55,56</sup> we next sought to examine whether the 5hmC-modified distal regions may harbor binding sites for early development-related TFs. HOMER,<sup>57</sup> which is a motif enrichment analysis tool, found a set of TFs whose binding sites



are enriched in 5hmC-modified distal elements throughout early embryonic development (Figure 6E). These include HIF-1b(Arnt), HIF-1a, OCT4, ESRRB, GSC, NPAS4, OTX2, KLF4, KLF5, NR5A2, GATA4, GATA3, GATA1, and AP1, the major TFs for early development. For example, OTX2 and ESRRB have been identified as mediators of enhancer priming in embryonic cells.<sup>58</sup> We also found OCT4 motif enrichment in the 5hmC-marked region, which coincides with the increased chromatin accessibility

**Figure 5. 5hmC hotspots associated with young TEs**

(A) Average DNA hydroxymethylation levels of LINEs (L1 and L2), SINEs (Alu and MIR), and LTRs (ERVK and ERVL) and relative expression levels (dataset: PRJNA616184) across early embryonic stages. Hydroxymethylation and expression levels are shown with solid and dotted lines, respectively. Data are mean  $\pm$  95% CI.

(B) Violin plots showing the 5hmCpG level in hotspots (blue) compared to CpG density-controlled 1-kb tiles (yellow) and the other 1-kb tiles (red) across stages.

(C) The colocalization of hotspots to the genomic elements provided by ENCODE3 (Hc, heterochromatin; Hc-H, H3K9me3-associated Hc; Hc-P, Polycomb-associated Hc; Tr, transcription; Tr-P, permissive Tr; Tr-S, strong Tr; Tr-I, initiation Tr; En, enhancer; En-Pp, poised TSS-proximal En; En-W, weak TSS-distal En; En-Sd, strong TSS-distal En; En-Pd, poised TSS-distal En; Pr, promoter; Pr-F, flanking Pr). The black dashed line indicates the threshold of odds ratio = 1. Only genomic elements significantly enriched by Fisher's exact test and adjusted by FDR ( $q < 0.05$ ) compared to CpG density-controlled regions are shown.

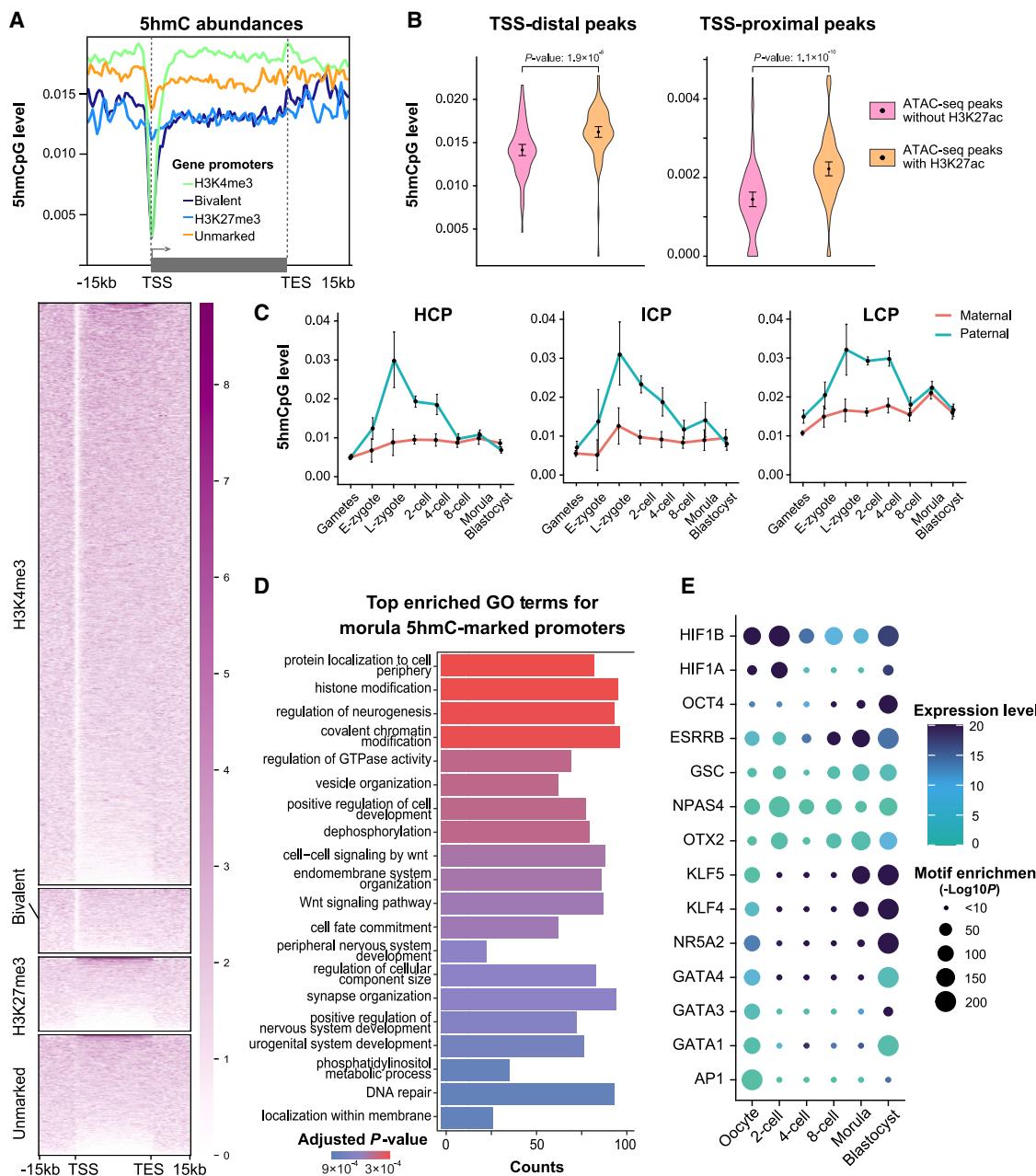
(D) The representative views of hydroxymethylation hotspots during early development. They are located in the promoter of the repeat elements LINE (left) and LTR (right) and overlap with the stable H3K9me3 peaks.

(E) From left to right: enrichment analysis of hotspots with the repeat elements in the class level, family level and subfamily level of LINE and LTR. Center right and right: color coding based on million years of age using the Jukes-Cantor model.

at the blastocyst stage.<sup>30</sup> HIF-1b is constitutively expressed but could only bind unmethylated CpG<sup>59,60</sup>; its enrichment in the 5hmC-marked region suggests that active DNA demethylation could be an *a priori* requirement for its binding. We also analyzed the 5hmC-modified proximal regions. There are similar sets of TFs between distal and proximal regulator regions, including HIF-1b(Arnt), ESRRB, GSC, OTX2, KLF4, and KLF5 (Figure S5I). Again, the generation of 5hmC on zygotic TF binding motifs is dependent on TET3, as evidenced by the similar 5hmC TF motif enrichments between oocytes and Tet3 cKO zygotes (Figure S5J). Collectively, these results suggest that Tet3 and 5hmC may have a regulatory function in early development, in addition to their roles as demethylation intermediates.

#### The 5hmC profile defines the cell and strand lineage of early embryos

In early embryo cleavages, we observed dramatic strand bias of 5hmC, which means that TET-driven active DNA demethylation



**Figure 6. Prevalent presence of DNA hydroxymethylation in regulatory genomic elements**

(A) The distribution of 5hmC among different chromatin states of TSS in 2-cell embryos. Colors represent different chromatin states of the TSS (top). Each row indicates a RefSeq gene (bottom).

(B) 5hmC level on distal ATAC-seq (assay for transposase-accessible chromatin using sequencing) peaks without H3K27ac (peaks), distal ATAC-seq peaks with H3K27ac (peaks), promoter ATAC-seq peaks without H3K27ac (peaks), and promoter ATAC-seq peaks with H3K27ac (peaks) in 2-cell embryos.  $p$  values are based on a paired U test. Data are mean  $\pm$  95% CI.

(C) Dynamics of DNA hydroxymethylation on promoters with different CpG densities. LCP, low CpG density promoter; ICP, intermediate CpG density promoter; HCP, high CpG density promoter. Data are mean  $\pm$  95% CI.

(D) GO term enrichment in genes with enhanced promoter hydroxymethylation in the morula stage.

(E) Motif analysis of 5hmC-marked regions (5hmC site  $\pm$  100 bp). The size of dots indicates  $-\log_{10}(p$  value), and the colors of dots indicate gene expression level (bulk RNA sequencing dataset: GSE98150).

occurs predominantly on one strand. The phenomenon is supported by previous observations.<sup>4,61</sup> Since we are able to separate the two parental genomes, we examined whether both display such strand bias. Interestingly, despite the overall dramatic difference of 5hmC of the paternal and maternal genomes, both display biased 5hmC distribution, supporting a similar preference of TET activity in strand selection for oxidation (Figure 7A). The biased distribution can be detected across early embryo stages, especially for 2-cell and 4-cell embryos, and Tet3 is responsible for this asymmetry (Figure S6A). Yet, starting from 8-cell embryos, we found a decline of the bimodal distribution and gradual establishment of a unimodal distribution (Figure 7A).

Based on these observations, we sought to investigate the potential inheritance of strand-asymmetric DNA hydroxymethylation. Inspired by nonrandom chromosome segregation during asymmetrical division of stem cells,<sup>62,63</sup> we analyzed the chromosome segregation from 2-cell to 4-cell embryos. For paired 4-cell embryo cells, we found that there are always two highly modified strands for each chromosome (Figure 7B). At the same time, we took advantage of the existence of strand bias and SCE (sister chromatid exchange) events that occur at the same location in the genome to identify sister cells within 2-cell or 4-cell embryos (Figures 7C, S6B, and S6C). Interestingly, two highly modified strands for each chromosome were always distributed into two non-sister cells in the same 4-cell embryo. Therefore, we proposed a model of DNA hydroxymethylation dynamics based on strand-specific 5hmC (Figure 7D). In this model, active demethylation occurs on newly deposited 5mC sites in the newly synthesized DNA strand, while 5mC sites on the inherited strand are still preserved after two rounds of cleavage. Thus, the 5hmC modification generated during this process can be used as an index or indicator of strand age for the reconstruction of strand-inherited lineages. Previous studies found that early embryos as early as the 4-cell stage initiate cell fate decisions by modulating the balance of pluripotency and differentiation.<sup>64</sup> Some specific genes related to cell fate decision, such as *LincGET* coded on the Watson/Crick strand of chromosome 7, have been reported<sup>65</sup> (Figure S6D).

## DISCUSSION

Active DNA demethylation plays an important role in early embryonic development. Here, we develop a quantitative 5hmC profiling method at single-cell and single-base resolution, and it should be readily compatible with combinatorial indexing platforms, offering the opportunity for ultra-high-throughput single-cell 5hmC profiling. In addition, schmC-CATCH could also be modified to enable simultaneous profiling of 5hmC and RNA expression in single cells. Hence, we anticipate schmC-CATCH and its future derivatives to have wide applications in different biological and clinical scenarios.

5hmC profiles help us understand the active DNA demethylation process. In this study, we found a first wave of 5hmC accumulation at fertilization before DNA replication, suggesting that TETs participate in the active DNA demethylation of 5mC inherited from gametes. The distinct 5hmC dynamics in paternal and maternal genomes suggested differentiated roles of TETs on the two alleles. The maternal active demethylation is more

focused on shaping the epigenetic status of regulatory elements, whereas paternal active demethylation serves as a safeguard from hypermethylation at whole-genome scale.

Another surprising finding is the presence of stable hydroxymethylation that co-exists with the deposition of H3K9me3 and is enriched in retrotransposon elements, especially in LINEs and LTRs. TEs are highly active during early development.<sup>48</sup> H3K9me3 has been reported to be initially non-repressive for gene expression and simultaneously plays a role in bookmarking the subsequent transmission into constitutive heterochromatin.<sup>49</sup> The 5hmC hotspots could cooperate with the non-repressive characteristics of H3K9me3 modification.

As the first epigenetic modification showing strand- and cell-biased patterns in mammalian early development, it is tempting to speculate regarding a potential role of asymmetric chromosomal 5-hydroxymethylation. Whether the different 5hmC status of specific genomic regions or chromosomes in individual cells of the 4-cell embryo could produce different outputs on gene expression,<sup>64</sup> especially cell fate-related genes, remains to be determined. In addition, the cell and strand lineage reconstructed by 5hmC may act as an indicator to investigate other epigenetic mark distributions or to examine the “immortal strand” hypothesis in the initial period of the life cycle to figure out the determinants of cell fate decisions during embryo cleavages.<sup>66</sup>

## Limitations of the study

In addition to active DNA demethylation, the early embryos undergo fast cell division and the replication-dependent passive dilution of 5mC, which may function independent from or in coordination with active DNA demethylation. Here, we focused on analyzing 5hmC and TET3-mediated active DNA demethylation and compared them with independently generated 5mC datasets, which may not reflect the precise relationships of these two modifications within individual cells. Similarly, comparisons with public gene expression profiles can only reveal the shared 5hmC-gene expression relationships across multiple embryos, but whether 5hmC may potentially contribute to asymmetric cell division remains unclear.

## RESOURCE AVAILABILITY

### Lead contact

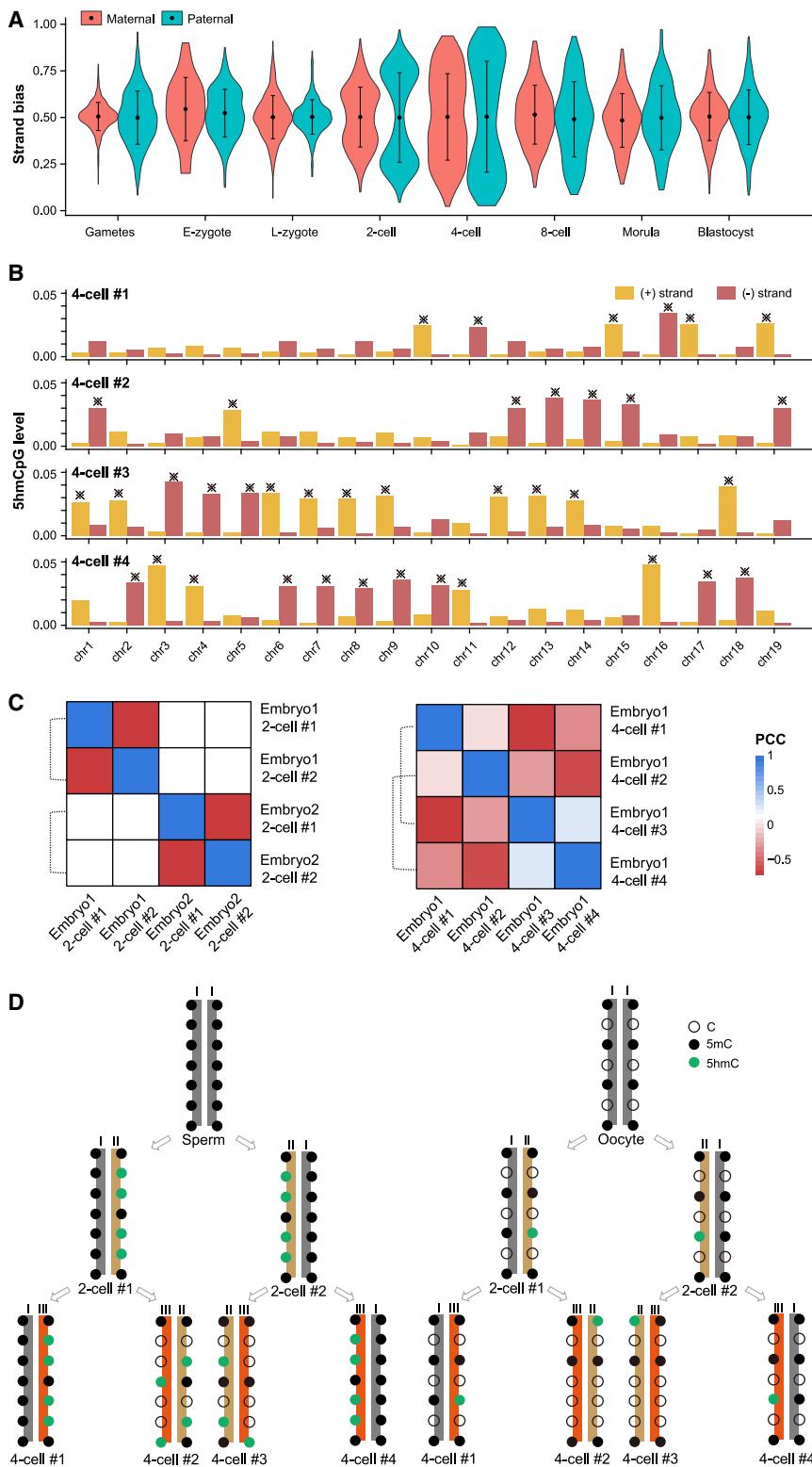
Further information and requests for resources and reagents should be directed to and will be fulfilled by the lead contact, Chengqi Yi ([chengqi.yi@pku.edu.cn](mailto:chengqi.yi@pku.edu.cn)).

### Materials availability

This study did not generate new unique reagents.

### Data and code availability

- Raw and processed single-cell 5hmC sequencing data have been deposited at NCBI and are publicly available as of the date of publication. Data of the methylation sequencing, RNA sequencing, ChIP sequencing of histone modifications and RNA Pol II binding, and ATAC-seq are public resources. Accession numbers are listed in the [key resources table](#). The chromatin state of the TSS in MII oocytes and preimplantation was taken from Liu et al.<sup>52</sup>
- All code are publicly available on GitHub (<https://github.com/yangjim221/schmC-CATCH>) as of the date of publication.



**Figure 7. The 5hmC profiles define the cell and strand lineage of early embryos**

(A) Violin plots of 5hmCpG strand bias for both the maternal and paternal genome. Data are mean  $\pm$  SD.

(B) The difference of 5hmCpG levels between Watson/Crick stands at the chromosome level. All 4 individual cells are from one single 4-cell embryo. Two strands with highest 5hmC levels are marked with an asterisk.

(C) Heatmaps showing the correlation coefficients in mouse 2-cell embryo and 4-cell embryo lineages. The Pearson correlation coefficients between strand bias in any two cells were calculated.

(D) Parental strand lineages constructed based on 5hmC bias to track each strand generated during the first two embryo cleavages.

- Any additional information required to re-analyze the data reported in this paper is available from the [lead contact](#) upon request.

### ACKNOWLEDGMENTS

We thank Prof. Guoliang Xu for providing the Tet3 cKO mice. We thank Yuan Zhuang and Prof. Wei Tao for technical assistance. We thank the State Key Laboratory of Gene Function and Modulation Research at Peking University in Beijing, China, for technical help. We carried out data analysis on the High-Performance Computing Platform at the School of Life Sciences, Peking University in Beijing, China. This study is supported by the Beijing Natural Science Foundation (Z220013 to C.Y.), the National Key R&D Program of China (2023YFC3402200), National Natural Science Foundation of China (91953201, 92153303, and 22425071 to C.Y. and 22207003 to B.H.), the China Postdoctoral Science Foundation (2021M700234 to B.H.), and the Central Guidance for local Science and Technology Development Fund Project (2024ZYD0106 to B.H.). This work was supported by the New Cornerstone Science Foundation through the XPLORER Prize.

### AUTHOR CONTRIBUTIONS

Conceptualization, D.B., F.T., and C.Y.; methodology, D.B. and J.Y.; investigation, D.B., J.Y., and X.X.; writing—original draft, D.B., J.Y., X.X., C.Z., and C.Y.; writing—review & editing, all authors; funding acquisition, C.Y. and B.H.; resources, J.Y., Y.G., Y.W., and M.C.; supervision, F.T. and C.Y.

### DECLARATION OF INTERESTS

The authors declare no competing interests.

### STAR METHODS

Detailed methods are provided in the online version of this paper and include the following:

- KEY RESOURCES TABLE
- EXPERIMENTAL MODEL AND STUDY PARTICIPANT DETAILS
  - Animals
  - mESC culture
- METHOD DETAILS
  - Oligonucleotide and model DNA synthesis
  - schmC-CATCH library preparation and sequencing
  - Sanger sequencing of the labeled model DNA
- QUANTIFICATION AND STATISTICAL ANALYSIS
  - Quality control of reads
  - Alignment
  - Identification of 5hmC sites
  - Embedding using UMAP
  - 5hmC level normalized by 5mC level
  - Expression level of transposable element (TE)
  - Ridgeplot of CpG density and 5hmC signal
  - Identification of DhMRs in the single-cell
  - Identification of regions enriched in 5hmC signaling at most early embryonic stages (hotspot)

### SUPPLEMENTAL INFORMATION

Supplemental information can be found online at <https://doi.org/10.1016/j.celrep.2025.115520>.

Received: July 29, 2024

Revised: December 30, 2024

Accepted: March 14, 2025

Published: April 3, 2025

### REFERENCES

- Guo, H., Zhu, P., Yan, L., Li, R., Hu, B., Lian, Y., Yan, J., Ren, X., Lin, S., Li, J., et al. (2014). The DNA methylation landscape of human early embryos. *Nature* 511, 606–610. <https://doi.org/10.1038/nature13544>.
- Smith, Z.D., Chan, M.M., Mikkelsen, T.S., Gu, H., Gnrke, A., Regev, A., and Meissner, A. (2012). A unique regulatory phase of DNA methylation in the early mammalian embryo. *Nature* 484, 339–344. <https://doi.org/10.1038/nature10960>.
- Zhu, P., Guo, H., Ren, Y., Hou, Y., Dong, J., Li, R., Lian, Y., Fan, X., Hu, B., Gao, Y., et al. (2018). Single-cell DNA methylome sequencing of human preimplantation embryos. *Nat. Genet.* 50, 12–19. <https://doi.org/10.1038/s41588-017-0007-6>.
- Amouroux, R., Nashun, B., Shirane, K., Nakagawa, S., Hill, P.W., D'Souza, Z., Nakayama, M., Matsuda, M., Turp, A., Ndjetehe, E., et al. (2016). De novo DNA methylation drives 5hmC accumulation in mouse zygotes. *Nat. Cell Biol.* 18, 225–233. <https://doi.org/10.1038/ncb3296>.
- Inoue, A., and Zhang, Y. (2011). Replication-dependent loss of 5-hydroxymethylcytosine in mouse preimplantation embryos. *Science* 334, 194. <https://doi.org/10.1126/science.1212483>.
- Wang, L., Zhang, J., Duan, J., Gao, X., Zhu, W., Lu, X., Yang, L., Zhang, J., Li, G., Ci, W., et al. (2014). Programming and inheritance of parental DNA methylomes in mammals. *Cell* 157, 979–991. <https://doi.org/10.1016/j.cell.2014.04.017>.
- Li, Y., Zhang, Z., Chen, J., Liu, W., Lai, W., Liu, B., Li, X., Liu, L., Xu, S., Dong, Q., et al. (2018). Stella safeguards the oocyte methylome by preventing de novo methylation mediated by DNMT1. *Nature* 564, 136–140. <https://doi.org/10.1038/s41586-018-0751-5>.
- Nakamura, T., Arai, Y., Umehara, H., Masuhara, M., Kimura, T., Taniguchi, H., Sekimoto, T., Ikawa, M., Yoneda, Y., Okabe, M., et al. (2007). PGC7/Stella protects against DNA demethylation in early embryogenesis. *Nat. Cell Biol.* 9, 64–71. <https://doi.org/10.1038/ncb1519>.
- Kohli, R.M., and Zhang, Y. (2013). TET enzymes, TDG and the dynamics of DNA demethylation. *Nature* 502, 472–479. <https://doi.org/10.1038/nature12750>.
- Plongthongkum, N., Diep, D.H., and Zhang, K. (2014). Advances in the profiling of DNA modifications: cytosine methylation and beyond. *Nat. Rev. Genet.* 15, 647–661. <https://doi.org/10.1038/nrg3772>.
- 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. <https://doi.org/10.1016/j.cell.2012.04.027>.
- Booth, M.J., Branco, M.R., Ficz, G., Oxley, D., Krueger, F., Reik, W., and Balasubramanian, S. (2012). Quantitative sequencing of 5-methylcytosine and 5-hydroxymethylcytosine at single-base resolution. *Science* 336, 934–937. <https://doi.org/10.1126/science.1220671>.
- Schutsky, E.K., DeNizio, J.E., Hu, P., Liu, M.Y., Nabel, C.S., Fabyanic, E.B., Hwang, Y., Bushman, F.D., Wu, H., and Kohli, R.M. (2018). Nondestructive, base-resolution sequencing of 5-hydroxymethylcytosine using a DNA deaminase. *Nat. Biotechnol.* 36, 1083–1090. <https://doi.org/10.1038/nbt.4204>.
- Liu, Y., Hu, Z., Cheng, J., Siejka-Zielinska, P., Chen, J., Inoue, M., Ahmed, A.A., and Song, C.X. (2021). Subtraction-free and bisulfite-free specific sequencing of 5-methylcytosine and its oxidized derivatives at base resolution. *Nat. Commun.* 12, 618. <https://doi.org/10.1038/s41467-021-20920-2>.
- Han, D., Lu, X., Shih, A.H., Nie, J., You, Q., Xu, M.M., Melnick, A.M., Levine, R.L., and He, C. (2016). A Highly Sensitive and Robust Method for Genome-wide 5hmC Profiling of Rare Cell Populations. *Mol. Cell* 63, 711–719. <https://doi.org/10.1016/j.molcel.2016.06.028>.
- 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. <https://doi.org/10.1038/nbt.1732>.
17. 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. <https://doi.org/10.1038/nature10102>.
  18. Zeng, H., He, B., Xia, B., Bai, D., Lu, X., Cai, J., Chen, L., Zhou, A., Zhu, C., Meng, H., et al. (2018). Bisulfite-Free, Nanoscale Analysis of 5-Hydroxymethylcytosine at Single Base Resolution. *J. Am. Chem. Soc.* 140, 13190–13194. <https://doi.org/10.1021/jacs.8b08297>.
  19. Yan, R., Cheng, X., Gu, C., Xu, Y., Long, X., Zhai, J., Sun, F., Qian, J., Du, Y., Wang, H., and Guo, F. (2023). Dynamics of DNA hydroxymethylation and methylation during mouse embryonic and germline development. *Nat. Genet.* 55, 130–143. <https://doi.org/10.1038/s41588-022-01258-x>.
  20. Cao, Y., Bai, Y., Yuan, T., Song, L., Fan, Y., Ren, L., Song, W., Peng, J., An, R., Gu, Q., et al. (2023). Single-cell bisulfite-free 5mC and 5hmC sequencing with high sensitivity and scalability. *Proc. Natl. Acad. Sci. USA* 120, e2310367120. <https://doi.org/10.1073/pnas.2310367120>.
  21. Zhu, C., Gao, Y., Guo, H., Xia, B., Song, J., Wu, X., Zeng, H., Kee, K., Tang, F., and Yi, C. (2017). Single-Cell 5-Formylcytosine Landscapes of Mammalian Early Embryos and ESCs at Single-Base Resolution. *Cell Stem Cell* 20, 720–731.e5. <https://doi.org/10.1016/j.stem.2017.02.013>.
  22. Song, C.X., Szulwach, K.E., Dai, Q., Fu, Y., Mao, S.Q., Lin, L., Street, C., Li, Y., Poidevin, M., Wu, H., et al. (2013). Genome-wide profiling of 5-formylcytosine reveals its roles in epigenetic priming. *Cell* 153, 678–691. <https://doi.org/10.1016/j.cell.2013.04.001>.
  23. Guo, H., Zhu, P., Guo, F., Li, X., Wu, X., Fan, X., Wen, L., and Tang, F. (2015). Profiling DNA methylome landscapes of mammalian cells with single-cell reduced-representation bisulfite sequencing. *Nat. Protoc.* 10, 645–659. <https://doi.org/10.1038/nprot.2015.039>.
  24. Smallwood, S.A., Lee, H.J., Angermueller, C., Krueger, F., Saadeh, H., Peat, J., Andrews, S.R., Stegle, O., Reik, W., and Kelsey, G. (2014). Single-cell genome-wide bisulfite sequencing for assessing epigenetic heterogeneity. *Nat. Methods* 11, 817–820. <https://doi.org/10.1038/nmeth.3035>.
  25. Luo, C., Keown, C.L., Kurihara, L., Zhou, J., He, Y., Li, J., Castanon, R., Lucero, J., Nery, J.R., Sandoval, J.P., et al. (2017). Single-cell methylomes identify neuronal subtypes and regulatory elements in mammalian cortex. *Science* 357, 600–604. <https://doi.org/10.1126/science.aan3351>.
  26. Luo, C., Rivkin, A., Zhou, J., Sandoval, J.P., Kurihara, L., Lucero, J., Castanon, R., Nery, J.R., Pinto-Duarte, A., Bui, B., et al. (2018). Robust single-cell DNA methylome profiling with snmC-seq2. *Nat. Commun.* 9, 3824. <https://doi.org/10.1038/s41467-018-06355-2>.
  27. Mulqueen, R.M., Pokholok, D., Norberg, S.J., Torkenczy, K.A., Fields, A.J., Sun, D., Sinnamon, J.R., Shendure, J., Trapnell, C., O’Roak, B.J., et al. (2018). Highly scalable generation of DNA methylation profiles in single cells. *Nat. Biotechnol.* 36, 428–431. <https://doi.org/10.1038/nbt.4112>.
  28. Shen, L., Inoue, A., He, J., Liu, Y., Lu, F., and Zhang, Y. (2014). Tet3 and DNA replication mediate demethylation of both the maternal and paternal genomes in mouse zygotes. *Cell Stem Cell* 15, 459–471. <https://doi.org/10.1016/j.stem.2014.09.002>.
  29. 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. <https://doi.org/10.1038/nature10443>.
  30. Wu, J., Huang, B., Chen, H., Yin, Q., Liu, Y., Xiang, Y., Zhang, B., Liu, B., Wang, Q., Xia, W., et al. (2016). The landscape of accessible chromatin in mammalian preimplantation embryos. *Nature* 534, 652–657. <https://doi.org/10.1038/nature18606>.
  31. Fan, X., Tang, D., Liao, Y., Li, P., Zhang, Y., Wang, M., Liang, F., Wang, X., Gao, Y., Wen, L., et al. (2020). Single-cell RNA-seq analysis of mouse pre-implantation embryos by third-generation sequencing. *PLoS Biol.* 18, e3001017. <https://doi.org/10.1371/journal.pbio.3001017>.
  32. Guo, H., Zhu, P., Wu, X., Li, X., Wen, L., and Tang, F. (2013). Single-cell methylome landscapes of mouse embryonic stem cells and early embryos analyzed using reduced representation bisulfite sequencing. *Genome Res.* 23, 2126–2135. <https://doi.org/10.1101/gr.161679.113>.
  33. Ladstatter, S., and Tachibana-Konwalski, K. (2016). A Surveillance Mechanism Ensures Repair of DNA Lesions during Zygotic Reprogramming. *Cell* 167, 1774–1787.e13. <https://doi.org/10.1016/j.cell.2016.11.009>.
  34. Sen, M., Mooijman, D., Chialastri, A., Boisset, J.C., Popovic, M., Heindryckx, B., Chuva de Sousa Lopes, S.M., Dey, S.S., and van Oudenaren, A. (2021). Strand-specific single-cell methylomics reveals distinct modes of DNA demethylation dynamics during early mammalian development. *Nat. Commun.* 12, 1286. <https://doi.org/10.1038/s41467-021-21532-6>.
  35. Wossidlo, M., Nakamura, T., Lepikhov, K., Marques, C.J., Zakhartchenko, V., Boiani, M., Arand, J., Nakano, T., Reik, W., and Walter, J. (2011). 5-Hydroxymethylcytosine in the mammalian zygote is linked with epigenetic reprogramming. *Nat. Commun.* 2, 241. <https://doi.org/10.1038/ncomms1240>.
  36. Suzuki, T., Abe, K.I., Inoue, A., and Aoki, F. (2009). Expression of c-MYC in nuclear speckles during mouse oocyte growth and preimplantation development. *J. Reprod. Dev.* 55, 491–495. <https://doi.org/10.1262/jrd.09-069a>.
  37. Prendergast, G.C., Lawe, D., and Ziff, E.B. (1991). Association of Myn, the Murine Homolog of Max, with C-Myc Stimulates Methylation-Sensitive DNA-Binding and Ras Cotransformation. *Cell* 65, 395–407. [https://doi.org/10.1016/0092-8674\(91\)90457-A](https://doi.org/10.1016/0092-8674(91)90457-A).
  38. Aksoy, I., Giudice, V., Delahaye, E., Wianny, F., Aubry, M., Mure, M., Chen, J., Jauch, R., Bogu, G.K., Nolden, T., et al. (2014). Klf4 and Klf5 differentially inhibit mesoderm and endoderm differentiation in embryonic stem cells. *Nat. Commun.* 5, 3719. <https://doi.org/10.1038/ncomms4719>.
  39. Jiang, J., Chan, Y.S., Loh, Y.H., Cai, J., Tong, G.Q., Lim, C.A., Robson, P., Zhong, S., and Ng, H.H. (2008). A core Klf circuitry regulates self-renewal of embryonic stem cells. *Nat. Cell Biol.* 10, 353–360. <https://doi.org/10.1038/ncb1698>.
  40. Yang, H., Bai, D., Li, Y., Yu, Z., Wang, C., Sheng, Y., Liu, W., Gao, S., and Zhang, Y. (2022). Allele-specific H3K9me3 and DNA methylation co-marked CpG-rich regions serve as potential imprinting control regions in pre-implantation embryo. *Nat. Cell Biol.* 24, 783–792. <https://doi.org/10.1038/s41556-022-00900-4>.
  41. He, B., Zhang, C., Zhang, X., Fan, Y., Zeng, H., Liu, J., Meng, H., Bai, D., Peng, J., Zhang, Q., et al. (2021). Tissue-specific 5-hydroxymethylcytosine landscape of the human genome. *Nat. Commun.* 12, 4249. <https://doi.org/10.1038/s41467-021-24425-w>.
  42. 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. <https://doi.org/10.1016/j.molcel.2011.04.005>.
  43. Cui, X.L., Nie, J., Ku, J., Dougherty, U., West-Szymanski, D.C., Collin, F., Ellison, C.K., Sieh, L., Ning, Y., Deng, Z., et al. (2020). A human tissue map of 5-hydroxymethylcytosines exhibits tissue specificity through gene and enhancer modulation. *Nat. Commun.* 11, 6161. <https://doi.org/10.1038/s41467-020-20001-w>.
  44. Shen, L., Wu, H., Diep, D., Yamaguchi, S., D’Alessio, A.C., Fung, H.L., Zhang, K., and Zhang, Y. (2013). Genome-wide analysis reveals TET- and TDG-dependent 5-methylcytosine oxidation dynamics. *Cell* 153, 692–706. <https://doi.org/10.1016/j.cell.2013.04.002>.
  45. Wu, X., and Zhang, Y. (2017). TET-mediated active DNA demethylation: mechanism, function and beyond. *Nat. Rev. Genet.* 18, 517–534. <https://doi.org/10.1038/nrg.2017.33>.
  46. Branco, M.R., Ficz, G., and Reik, W. (2011). Uncovering the role of 5-hydroxymethylcytosine in the epigenome. *Nat. Rev. Genet.* 13, 7–13. <https://doi.org/10.1038/nrg3080>.

47. Perchardé, M., Lin, C.J., Yin, Y., Guan, J., Peixoto, G.A., Bulut-Karslioglu, A., Biechele, S., Huang, B., Shen, X., and Ramalho-Santos, M. (2018). A LINE1-Nucleolin Partnership Regulates Early Development and ESC Identity. *Cell* 174, 391–405.e19. <https://doi.org/10.1016/j.cell.2018.05.043>.
48. Jachowicz, J.W., Bing, X., Pontabry, J., Bošković, A., Rando, O.J., and Torres-Padilla, M.E. (2017). LINE-1 activation after fertilization regulates global chromatin accessibility in the early mouse embryo. *Nat. Genet.* 49, 1502–1510. <https://doi.org/10.1038/ng.3945>.
49. Burton, A., Brochard, V., Galan, C., Ruiz-Morales, E.R., Rovira, Q., Rodriguez-Terrones, D., Kruse, K., Le Gras, S., Udayakumar, V.S., Chin, H.G., et al. (2020). Heterochromatin establishment during early mammalian development is regulated by pericentromeric RNA and characterized by non-repressive H3K9me3. *Nat. Cell Biol.* 22, 767–778. <https://doi.org/10.1038/s41556-020-0536-6>.
50. Richard Albert, J., Au Yeung, W.K., Toriyama, K., Kobayashi, H., Hirasawa, R., Brind'Amour, J., Bogutz, A., Sasaki, H., and Lorincz, M. (2020). Maternal DNMT3A-dependent de novo methylation of the paternal genome inhibits gene expression in the early embryo. *Nat. Commun.* 11, 5417. <https://doi.org/10.1038/s41467-020-19279-7>.
51. Kubura, M., Okano, M., Kimura, H., Kawamura, F., and Tada, M. (2012). Chromosome-wide regulation of euchromatin-specific 5mC to 5hmC conversion in mouse ES cells and female human somatic cells. *Chromosome Res.* 20, 837–848. <https://doi.org/10.1007/s10577-012-9317-9>.
52. Liu, X., Wang, C., Liu, W., Li, J., Li, C., Kou, X., Chen, J., Zhao, Y., Gao, H., Wang, H., et al. (2016). Distinct features of H3K4me3 and H3K27me3 chromatin domains in pre-implantation embryos. *Nature* 537, 558–562. <https://doi.org/10.1038/nature19362>.
53. 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. <https://doi.org/10.1186/gb-2011-12-6-r54>.
54. Guo, G., Huss, M., Tong, G.Q., Wang, C., Li Sun, L., Clarke, N.D., and Robson, P. (2010). Resolution of cell fate decisions revealed by single-cell gene expression analysis from zygote to blastocyst. *Dev. Cell* 18, 675–685. <https://doi.org/10.1016/j.devcel.2010.02.012>.
55. Zhu, H., Wang, G., and Qian, J. (2016). Transcription factors as readers and effectors of DNA methylation. *Nat. Rev. Genet.* 17, 551–565. <https://doi.org/10.1038/nrg.2016.83>.
56. Yin, Y., Morgunova, E., Jolma, A., Kaasinen, E., Sahu, B., Khund-Sayeed, S., Das, P.K., Kivioja, T., Dave, K., Zhong, F., et al. (2017). Impact of cytosine methylation on DNA binding specificities of human transcription factors. *Science* 356, eaaj2239. <https://doi.org/10.1126/science.aaaj2239>.
57. Heinz, S., Benner, C., Spann, N., Bertolino, E., Lin, Y.C., Laslo, P., Cheng, J.X., Murre, C., Singh, H., and Glass, C.K. (2010). Simple combinations of lineage-determining transcription factors prime cis-regulatory elements required for macrophage and B cell identities. *Mol. Cell* 38, 576–589. <https://doi.org/10.1016/j.molcel.2010.05.004>.
58. Buecker, C., Srinivasan, R., Wu, Z., Calo, E., Acampora, D., Faial, T., Simeone, A., Tan, M., Swigut, T., and Wysocka, J. (2014). Reorganization of enhancer patterns in transition from naive to primed pluripotency. *Cell Stem Cell* 14, 838–853. <https://doi.org/10.1016/j.stem.2014.04.003>.
59. D'Anna, F., Van Dyck, L., Xiong, J., Zhao, H., Berrens, R.V., Qian, J., Bieniasz-Krzywiec, P., Chandra, V., Schoonjans, L., Matthews, J., et al. (2020). DNA methylation repels binding of hypoxia-inducible transcription factors to maintain tumor immunotolerance. *Genome Biol.* 21, 182. <https://doi.org/10.1186/s13059-020-02087-z>.
60. Dunwoodie, S.L. (2009). The role of hypoxia in development of the Mammalian embryo. *Dev. Cell* 17, 755–773. <https://doi.org/10.1016/j.devcel.2009.11.008>.
61. Mooijman, D., Dey, S.S., Boisset, J.C., Crosetto, N., and van Oudenaarden, A. (2016). Single-cell 5hmC sequencing reveals chromosome-wide cell-to-cell variability and enables lineage reconstruction. *Nat. Biotechnol.* 34, 852–856. <https://doi.org/10.1038/nbt.3598>.
62. Huh, Y.H., Cohen, J., and Sherley, J.L. (2013). Higher 5-hydroxymethylcytosine identifies immortal DNA strand chromosomes in asymmetrically self-renewing distributed stem cells. *Proc. Natl. Acad. Sci. USA* 110, 16862–16867. <https://doi.org/10.1073/pnas.1310323110>.
63. Ranjan, R., and Chen, X. (2022). Mitotic drive in asymmetric epigenetic inheritance. *Biochem. Soc. Trans.* 50, 675–688. <https://doi.org/10.1042/BST20200267>.
64. Goolam, M., Scialdone, A., Graham, S.J.L., Macaulay, I.C., Jedrusik, A., Hupalowska, A., Voet, T., Marioni, J.C., and Zernicka-Goetz, M. (2016). Heterogeneity in Oct4 and Sox2 Targets Biases Cell Fate in 4-Cell Mouse Embryos. *Cell* 165, 61–74. <https://doi.org/10.1016/j.cell.2016.01.047>.
65. Wang, J., Wang, L., Feng, G., Wang, Y., Li, Y., Li, X., Liu, C., Jiao, G., Huang, C., Shi, J., et al. (2018). Asymmetric Expression of LincGET Biases Cell Fate in Two-Cell Mouse Embryos. *Cell* 175, 1887–1901.e18. <https://doi.org/10.1016/j.cell.2018.11.039>.
66. Rando, T.A. (2007). The immortal strand hypothesis: segregation and reconstruction. *Cell* 129, 1239–1243. <https://doi.org/10.1016/j.cell.2007.06.019>.
67. Krueger, F., and Andrews, S.R. (2016). SNPsplits: Allele-specific splitting of alignments between genomes with known SNP genotypes. *F1000Res.* 5, 1479. <https://doi.org/10.12688/f1000research.9037.2>.
68. De Coster, W., D'Hert, S., Schultz, D.T., Cruts, M., and Van Broeckhoven, C. (2018). NanoPack: visualizing and processing long-read sequencing data. *Bioinformatics* 34, 2666–2669. <https://doi.org/10.1093/bioinformatics/bty149>.
69. Li, H. (2018). Minimap2: pairwise alignment for nucleotide sequences. *Bioinformatics* 34, 3094–3100. <https://doi.org/10.1093/bioinformatics/bty191>.
70. He, J., Babarinde, I.A., Sun, L., Xu, S., Chen, R., Shi, J., Wei, Y., Li, Y., Ma, G., Zhuang, Q., et al. (2021). Identifying transposable element expression dynamics and heterogeneity during development at the single-cell level with a processing pipeline scTE. *Nat. Commun.* 12, 1456. <https://doi.org/10.1038/s41467-021-21808-x>.
71. Juhling, F., Kretzmer, H., Bernhart, S.H., Otto, C., Stadler, P.F., and Hoffmann, S. (2016). metilene: fast and sensitive calling of differentially methylated regions from bisulfite sequencing data. *Genome Res.* 26, 256–262. <https://doi.org/10.1101/gr.196394.115>.
72. Sheffield, N.C., and Bock, C. (2016). LOLA: enrichment analysis for genomic region sets and regulatory elements in R and Bioconductor. *Bioinformatics* 32, 587–589. <https://doi.org/10.1093/bioinformatics/btv612>.
73. Wang, C., Liu, X., Gao, Y., Yang, L., Li, C., Liu, W., Chen, C., Kou, X., Zhao, Y., Chen, J., et al. (2018). Reprogramming of H3K9me3-dependent heterochromatin during mammalian embryo development. *Nat. Cell Biol.* 20, 620–631. <https://doi.org/10.1038/s41556-018-0093-4>.
74. Yu, G., Wang, L.G., Han, Y., and He, Q.Y. (2012). clusterProfiler: an R package for comparing biological themes among gene clusters. *OMICS* 16, 284–287. <https://doi.org/10.1089/omi.2011.0118>.

## STAR★METHODS

### KEY RESOURCES TABLE

REAGENT or RESOURCE	SOURCE	IDENTIFIER
<b>Chemicals, peptides, and recombinant proteins</b>		
HCG	Sigma	Cat#1297001
PMSG	San-Sheng	Cat# 21958956
Tyrode's solution	Sigma	Cat# T1788
Trypsin-EDTA (0.05%)	GIBCO	Cat#25300-062
PD0325901	Selleckchem	Cat# S1036
CHIR99021	Selleckchem	Cat#S2924
fetal bovine serum (FBS)	GIBCO	Cat#16141079
5-Hydroxymethyl-dC II-CE Phosphoramidite	Glen Research	Cat#10-1510-02
5-Formyl dC III CE Phosphoramidite	Glen Research	Cat# 10-1564-02
5-Me-dC-CE Phosphoramidite	Glen Research	Cat# 10-1060-02
Tris-Ac	Sigma	Cat#93337
EDTA	Invitrogen	Cat#AM9260G
DTT	Sigma	Cat#D9779
Protease	Qiagen	Cat#19157
Tn5 transposome	Vazyme	Cat#S601-01
T4 DNA ligase buffer	NEB	Cta#B0202S
Exonuclease I	NEB	Cat#M0293S
Carrier RNA	Ambion	Cat#4382878
10× EX Taq buffer	Takara	Cat#RR001A
dCTP	NEB	Cat#N0446S
Terminal transferase	NEB	Cat#M0315S
Potassium perruthenate	Sigma	Cat#334537
NaOH	Alfa Aesar	Cat#A18395
Malononitrile	J&K	Cat#261700
KAPA2G Robust HS DNA polymerase	Roche	Cat#KK5023
Kapa HF Hotstart DNA polymerase	Roche	Cat#KR0369
<b>Critical commercial assays</b>		
Micro Bio-Spin P-6 SSC column	Bio-rad	Cat#732-6200
AMPure XP beads	Beckman	Cat#B23319
<b>Deposited data</b>		
schmC-CATCH sequencing data	This Paper; NCBI GEO	GEO: GSE217627
The RNA sequencing data of DNA methylation-related genes	NCBI GEO	GEO:PRJNA616184
The ChIP-seq data of H3K9me3	NCBI GEO	GEO:GSE98149
The ChIP-seq data of H3K4me3	NCBI GEO	GEO:GSE73952
The ChIP-seq data of H3K27me3	NCBI GEO	GEO:GSE73952
The ChIP-seq data of H3K36me3	NCBI GEO	GEO:GSE112834
The ChIP-seq data of H3K27ac	NCBI GEO	GEO:GSE79935
The data of RNA Polymerase II binding	NCBI GEO	GEO:GSE135457
ATAC-seq data	NCBI GEO	GEO:GSE66581
<b>Experimental models: Cell lines</b>		
Mouse: mESC	Fuchou Tang lab	N/A

(Continued on next page)

**Continued**

REAGENT or RESOURCE	SOURCE	IDENTIFIER
Experimental models: Organisms/strains		
Mouse: DBA/2N	Vital River Laboratory	N/A
Mouse: C57BL/6J-129Sv	Guoliang Xu lab	N/A
Mouse: C57BL	Vital River Laboratory	N/A
Oligonucleotides		
Spike-in DNA oligos see <a href="#">Table S1</a>	This paper	N/A
Software and algorithms		
trim galore (v0.6.6)	N/A	<a href="http://www.bioinformatics.babraham.ac.uk/projects/trim_galore/">http://www.bioinformatics.babraham.ac.uk/projects/trim_galore/</a>
cutadapt(v2.10).	N/A	<a href="https://github.com/marcelm/cutadapt/">https://github.com/marcelm/cutadapt/</a>
Bowtie2(v2.4.1)	N/A	<a href="https://bowtie-bio.sourceforge.net/bowtie2">https://bowtie-bio.sourceforge.net/bowtie2</a>
SNPsplit(v0.5.0)	N/A	<a href="https://www.bioinformatics.babraham.ac.uk/projects/SNPsplit/">https://www.bioinformatics.babraham.ac.uk/projects/SNPsplit/</a>
Seurat (v4.2.0)	N/A	<a href="https://satijalab.org/seurat/">https://satijalab.org/seurat/</a>
NanoFilt (v2.8.0)	N/A	<a href="https://github.com/wdecoster/nanofilt">https://github.com/wdecoster/nanofilt</a>
Pychopper(v2.5.0)	N/A	<a href="https://github.com/nanoporetech/pychopper">https://github.com/nanoporetech/pychopper</a>
minimap2(v2.18-r1015)	N/A	<a href="https://github.com/lh3/minimap2">https://github.com/lh3/minimap2</a>
scTE(v1.0)	N/A	<a href="https://github.com/JiekaiLab/scTE">https://github.com/JiekaiLab/scTE</a>
HOMER	N/A	<a href="http://homer.salk.edu/homer/">http://homer.salk.edu/homer/</a>
Code for analysis	This paper	<a href="https://github.com/yangjim221/schmC-CATCH">https://github.com/yangjim221/schmC-CATCH</a>

## EXPERIMENTAL MODEL AND STUDY PARTICIPANT DETAILS

### Animals

All analyzed and wild type embryos were collected from 6- to 8-week-old C57BL/6N female mice mated with 12-week-old DBA/2N males (Vital River). Tet3 cKO embryos are collected from 6- to 8-week-old C57BL/6J-129Sv female mouse mated with 12-week-old C57BL male mouse. All mice were kept in 12h-light/12h-dark SPF-class Laboratory Animal Room and were offered food and water *ad libitum*. To induce ovulation, females were administered 7.5 IU of hCG intraperitoneally, 48 h after injection of 7.5 IU of PMSG (San-Sheng pharmaceutical Co. Ltd). Each set of embryos at a particular stage was flushed from the reproductive tract at defined time periods after hCG administration: 16.5–30 h(PN1-PN5), 43–45 h (2-cell), 50–52 h (4-cell), 65–67 h (8-cell), 74.5–76.5 h(morula) and 93–95 h (blastocysts) in M2 medium (Sigma M7167). Embryos were selected by cell number or morphology, with their zona pellucida gently removed by treatment with acidic Tyrode's solution (Sigma T1788) for several minutes. Naked embryos were further dissociated into single cells by using Accutase solution. The resulting cells were manually picked and treated with the lysis buffer for schmC-CATCH. All experimental procedures were carried out according to guidelines of Institutional Animal Care and Use Committee (IACUC) of Peking University, Beijing, China.

### mESC culture

mESCs were cultured on gelatin in DMEM containing 15% FBS, leukemia inhibiting factor, penicillin/streptomycin, L-glutamine, β-mercaptoethanol, and non-essential amino acids.

## METHOD DETAILS

### Oligonucleotide and model DNA synthesis

Oligonucleotides containing site-specific modified cytosines ([Table S3](#)) were synthesized on the Expedite 8909 nucleic acid synthesizer using commercially available phosphoramidites (Glen Research, 10-1510-02, 10-1564-02, 10-1060-02). Regular oligonucleotides were purchased from Sangon Biotech (Shanghai). Duplex 5hmC DNAs (hmC-spikein, [Table S3](#)) for NGS sequencing were prepared through ligation of short duplexes fragments (20-40bp) with sticky overhangs. Tri-spikein ([Table S3](#)) for Sanger sequencing was single strand DNA.

### schmC-CATCH library preparation and sequencing

The libraries of mESCs and early mouse embryos were prepared as below. Briefly, samples were lysed in lysis buffer (60 mM Tris-Ac pH 8.3 (Sigma, 93337), 2 mM EDTA pH 8.0 (Invitrogen, AM9260G), 15 mM DTT (Sigma, D9779), 0.5 μM carrier ssDNA

(5'-TCAGGTTTCCTGAA-3') for 30 min at 75°C, after that, add 0.5 μM 10 μg/μL Qiagen protease, and incubate for 4 h at 55°C. Immediately after lysis, nuclei were then incubated with the indexed Tn5 transposome and tagmentation buffer at 55°C for 15 min (Vazyme Biotech, S601-01). After the tagmentation, the stop buffer was added directly into the reaction to end the tagmentation. Pool 8 cell lysate into one tube, denature transposons at 72°C for 5 min and snap on ice 1 min, after that, add T4 DNA ligase buffer and 1 μL Exo I (NEB, M0293S) and incubate at 37°C for 60 min. After the reaction, DNA were purified with the 2×AMPure (Beckman) beads with 500 ng carrier RNA (Ambion, 4382878). Trace amount of hmC-spikein DNA (0.5 pg, [Table S3](#)) were spiked into the genomic DNA. dC-tailing was performed, 1 μL 10× EX Taq buffer (Takara, RR001A) and 1 μL 1mM dCTP (NEB, N0446S) into 28 μL eluted DNA, denature at 95°C for 1 min and snap on ice 1 min, after that, add 1 μL Terminal transferase (NEB, M0315S) and incubate at 37°C for 35 min, 75°C for 20 min. After the reaction, DNA were purified with the 2× AMPure XP beads (Beckman Coulter, B23319).

Potassium perruthenate (Sigma-Aldrich, 334537) solution was prepared according to published protocol, briefly: 0.15 mmol Potassium perruthenate (Aldrich, 334537) was added to 0.5 M NaOH solution (1 mL, Alfa Aesar, A18395) and vortexed to make sure all solid was dissolved; incubated the solution at 25°C for 2 days to produce potassium ruthenate solution ( $K_2RuO_4$ ). Before 5hmC oxidation, genomic DNA was purified with 1× AMPure XP beads (Beckman Coulter, B23319) followed by additional purification with Micro Bio-Spin P-6 SSC column (Bio-rad, 732-6200). The DNA was then denatured in 0.05 M NaOH for 30 min at 37°C followed by chilling on ice-water bath for 5 min. Next, 1.5 μL 1× oxidant was added to the denatured DNA sample (28.5 μL) and briefly mixed with tapping followed by incubation on ice for 1 h. The oxidized DNA was purified with a Micro Bio-Spin P-6 SSC column. To label the newly generated 5fC, the reaction was carried out in 10 mM pH 7.0 Tris buffer and 150 mM malononitrile (J&K, 261700) in a total volume of 35 μL at 37°C for 20 h, 850 r.p.m. in a thermomixer. After full speed centrifuge to remove any precipitant, perform purification with Micro Bio-Spin P-6 SSC column (Bio-rad, 732-6200).

To obtain the 5hmC profiles from the single cells, the 5hmC labeled DNA was subjected to multiple rounds of primer extension to record the 5hmC-to-T transition signal. Primer extension mix (8 μL of 5× KAPA 2G buffer, 0.8 μL of 10 mM dNTPs, 0.8 μL of 10 μM P5G9H primer, 0.6 μL KAPA2G Robust HS DNA polymerase (Roche, KK5023)) was added and primer extension was performed with the following program: Step 1: 95°C × 5 min; Step 2: 95°C × 15 S, 47°C × 60 s, 68°C × 8 min, and repeat Step 2 an additional 19 times; Step 3: 72°C × 10 min and hold at 4°C. 2 μL of linear amplification products were linear amplified with spike-in primer ([Table S3](#)) as described above to add PCR handle onto hmC-spike-in DNA, and amplified with P5 universal primer and P7 index primer ([Table S3](#)). Add 8 μL of KAPA HiFi buffer, 2 μL of 500 nM P5, 2 μL of 500nM P7, 1.6 μL Kapa HF Hotstart DNA polymerase directly to the linear amplification mixture. PCR was performed to amplify the library for 11-15 cycles using the following PCR conditions: 72°C for 3 min; 98°C for 30 s; and thermocycling at 98°C for 15 s, 65°C for 15 s and 72°C for 1 min; following by 72°C 5 min.

After the PCR reaction, libraries were purified with the 0.8× AMPure (Beckman) beads. The purified libraries were sequenced on an MGISEQ-2000 sequencer (MG) or NovaSeq sequencer (Illumina) with 150 bp paired-end mode.

### Sanger sequencing of the labeled model DNA

Single-strand tri-spike-in that contains both 5mC, 5hmC and 5fC ([Table S3](#)) was treated with chemical reactions described above, and after labeling, PCR amplification was performed using KAPA HF Hotstart DNA polymerase. PCR products were purified with DNA Clean & Concentrator kit (Vistech) for Sanger sequencing.

### QUANTIFICATION AND STATISTICAL ANALYSIS

#### Quality control of reads

Firstly, de-multiplexed fastq reads were trimmed to discard low quality bases and fixed sequences, such as adaptor sequences by trim galore (version: 0.6.6) and cutadapt(version: 2.10).

#### Alignment

The clean reads were aligned to mm10 reference using Bowtie2(version: 2.4.1) with default parameters. Reads with MAPQ <10 was excluded for next analysis. Allele assignment of sequencing reads was performed by SNPsplitsplit(version: 0.5.0) according to previous study.<sup>[67](#)</sup>

#### Identification of 5hmC sites

Because 5hmC sites were mostly located in CpG sites, we focused on this dinucleotide. Single cells with 1× CpG coverage >100,000 were kept for following analysis. For each site, the numbers of converted “T” bases as  $N_T$  and “C” bases as  $N_C$  were calculated. All sites with less than two total bases ( $N_T + N_C$ ) were removed for calling. To eliminate the noise, the sites with > 50% C-to-T conversion rate was remained. We also subtract SNPs from candidate sites to decrease the number of false positive sites. As required, the 5hmC abundance was calculated in different bin size. Parental 5hmC calling was the same with above described. Genomic annotation and motif analysis were computed by HOMER software.

#### Embedding using UMAP

Firstly, we binned the genome into non-overlapping 500-kb for each single cell. For each bin, the 5hmC level was calculated by dividing the sum of 5hmC sites by the sum of covered CpG sites. Because the matrix of single cell 5hmC level is sparse, we took

a imputation which is reported by previous study.<sup>25</sup> Bins with 70% or more of the single cell were conserved. And missing value was replaced by the average 5hmC level across all cells in each bin. Then we used Seurat (version: 4.2.0) to perform PCA and project single cells to two-dimensional space by UMAP.

### 5hmC level normalized by 5mC level

WGBS dataset was accessed from GSE97778. The raw data were processed by the methods in the paper, and the reads were assigned into paternal and maternal allele. The methylation level of each CpG site was calculated by the number of “T” and the number of “C”. The methylation level of genomic elements was determined by the average level of CpG sites in these elements. Then the 5hmC level is normalized by the 5mC level through the 5hmC/5mC on the corresponding element.

### Expression level of transposable element (TE)

Raw data of SCAN-seq was downloaded from SRA with the accession number: PRJNA616184. It was processed as mentioned in the publish with minor modification.<sup>31</sup> In short, the reads in low quality (Quality < 7) or having read length less than 100 bp was filtered by NanoFilt (v2.8.0)<sup>68</sup> and the remaining reads were processed with Pychopper(v2.5.0) (<https://github.com/epi2me-labs/pychopper>) to identify the full-length reads and be trimmed off poly-A sequence. Those reads were ready to be mapped to reference genome mm10 by minimap2(v2.18-r1015).<sup>69</sup> Allelic assignment was achieved with the C57BL/6N and DBA/2C specific SNPs, which were obtained from the website Mouse Genomes Project and whole genome sequencing (WGS), respectively. Only those reads with SNPs exclusively belonging to a single strain were assigned. And read counts of TEs were counted by scTE(v1.0)<sup>70</sup> with the annotation files of RefSeq genes and repeat elements download GENCODE and UCSC RepeatMasker track, respectively.

### Ridgeplot of CpG density and 5hmC signal

The CpG density was measured based on previous publish method.<sup>1</sup> For each single CpG site, the surrounding (upstream 50 bp and downstream 50 bp) number of CpG sites was calculated as its CpG density. For 1 kb tiles, their CpG density were the average of the CpG sites located within those tiles and converted to nearest integers. The 5hmC level for 1 kb tiles were the ratios of the number of 5hmC signal supporting reads to the total number of reads covering corresponding tiles. The function geom\_density\_ridges() from the R package ggridges(v0.5.3) (<https://github.com/cran/ggridges>) helped to plot the ridge plot of 5hmC level versus CpG density.

### Identification of DhMRs in the single-cell

In the single-cell level, the DhMRs were mainly identified by metilene (v0.2-8).<sup>71</sup> Files containing 5hmC levels in 5 kb tiles for single cells were first merged into a matrix and then filtered to keep those tiles covered by at least 3 cells for each stage. KS-test and Mann-Whitney-U test were used to calculate the p-values, which were further adjusted by Benjamini-Hochberg. Those tiles having adjust p-value less than 0.05 and fold change of 5hmC level between the stage pairs (compared to the later stage) greater than 2 or less than 0.5 were referred to as DhMRs. The remaining tiles were referred to be stable between the stage pairs. And among those DhMRs, increasing DhMRs indicates a higher 5hmC signal for the latter stage, while similarly the decreasing DhMRs reveals a higher 5hmC signal for the former stage.

### Identification of regions enriched in 5hmC signaling at most early embryonic stages (hotspot)

The hotspots were identified based on the tile strategy in pseudo-bulk level. A 1 kb tiles is considered a hotspot if it has an absolute level greater than 0.05 in at least 4 of the 5 stages from 2-cell to blastocyst. The enrichment analysis of repeat elements was performed by R package LOLA(v1.22.0),<sup>72</sup> which adopt the fisher's exact result between query regions and reference dataset. The reference dataset for repeat elements were prepared with the gtf file download from UCSC RepeatMasker. Only those classes or sub-families turned out to have FDR-adjusted q-values less than 0.05 were considered to be significantly overlapping the hotspots. The bulk ChIP-seq data of H3K9me3 were downloaded from GSE98149 and processed according to the published paper.<sup>73</sup> Peaks consistently exist in at least 4 of the 5 stages from 2-cell to blastocyst. The degree of overlapping hotspots and stable H3K9me3 peaks were compared to those shuffled regions in 1 kb length by their intersection with stable H3K9me3 peaks. The hotspots were annotated to nearest genes by rGREAT(v1.24.0) (<https://github.com/jokergoo/rGREAT>) with the rule “basalPlusExt” and further performed with gene ontology analysis by clusterProfiler(v4.0.0).<sup>74</sup>