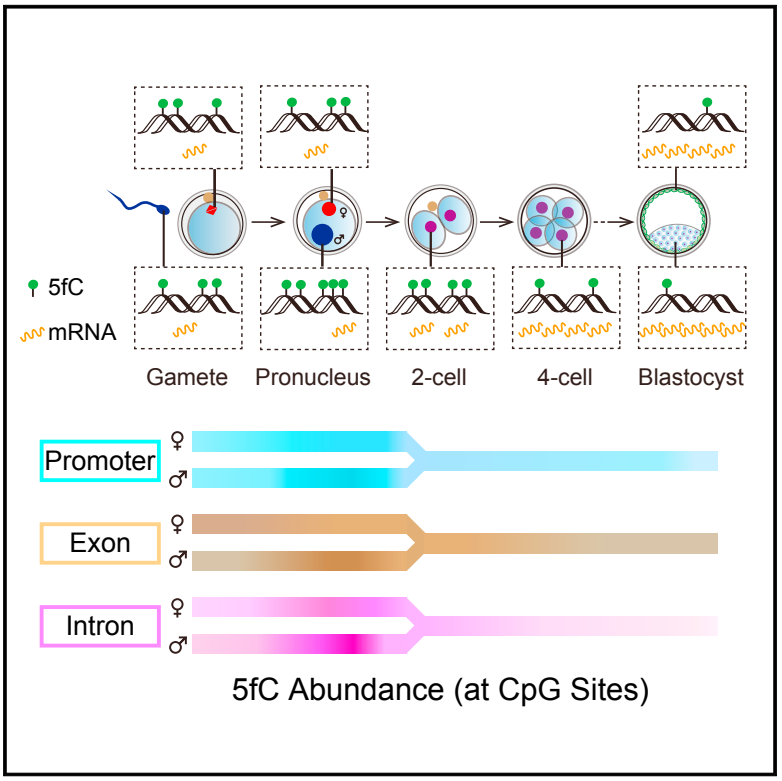
Resource

Cell Stem Cell

Single-Cell 5-Formylcytosine Landscapes of Mammalian Early Embryos and ESCs at Single-Base Resolution

Graphical Abstract



Highlights

- A bisulfite-free method detects 5fC at the single-cell and single-base resolution
- 5fC is highly dynamic and enriched in functionally important genomic regions
- Intrinsic 5fC heterogeneity is revealed for mouse early embryos, mEpiSC, and mESCs
- Emergence of promoter 5fC production precedes the upregulation of gene expression

Authors

Chenxu Zhu, Yun Gao, Hongshan Guo, ..., Kehkooi Kee, Fuchou Tang, Chengqi Yi

Correspondence

tangfuchou@pku.edu.cn (F.T.), chengqi.yi@pku.edu.cn (C.Y.)

In Brief

CLEVER-seq is a single-cell, single-base resolution whole-genome 5fC-sequencing technology. CLEVER-seq reveals intrinsic 5fC heterogeneity in mouse early embryos, EpiSCs, and ESCs and dissects its highly patterned genomic distributions. During mouse pre-implantation development, 5fC shows parental-specific signatures, and promoter 5fC production precedes the upregulation of corresponding gene expression.

Data Resources

GSE84833

Zhu et al., 2017, Cell Stem Cell *20*, 720–731
May 4, 2017
http://dx.doi.org/10.1016/j.stem.2017.02.013

CellPress

Cell Stem Cell
Resource

Single-Cell 5-Formylcytosine Landscapes of Mammalian Early Embryos and ESCs at Single-Base Resolution

Chenxu Zhu,[1,7] Yun Gao,[2,7] Hongshan Guo,[2,7] Bo Xia,[1,7,8] Jinghui Song,[1] Xinglong Wu,[2,3] Hu Zeng,[1] Kehkooi Kee,[4] Fuchou Tang,[2,3,5,*] and Chengqi Yi[1,3,6,9,*]

[1]State Key Laboratory of Protein and Plant Gene Research, School of Life Sciences, Peking University, Beijing 100871, PRC
[2]Biodynamic Optical Imaging Center and Beijing Advanced Innovation Center for Genomics, School of Life Sciences, Peking University, Beijing 100871, PRC
[3]Peking-Tsinghua Center for Life Sciences, Peking University, Beijing 100871, PRC
[4]Center for Stem Cell Biology and Regenerative Medicine, Department of Basic Medical Sciences, School of Medicine, Tsinghua University, Beijing 100084, China
[5]Ministry of Education Key Laboratory of Cell Proliferation and Differentiation, Peking University, Beijing 100871, PRC
[6]Department of Chemical Biology and Synthetic and Functional Biomolecules Center, College of Chemistry and Molecular Engineering, Peking University, Beijing 100871, PRC
[7]Co-first author
[8]Present address: Sackler Institute of Graduate Biomedical Sciences and Institute for Computational Medicine, New York University School of Medicine, New York, NY 10016, USA
[9]Lead Contact
*Correspondence: tangfuchou@pku.edu.cn (F.T.), chengqi.yi@pku.edu.cn (C.Y.)
http://dx.doi.org/10.1016/j.stem.2017.02.013

SUMMARY

Active DNA demethylation in mammals involves ten-eleven translocation (TET) family protein-mediated oxidation of 5-methylcytosine (5mC). However, base-resolution landscapes of 5-formylcytosine (5fC) (an oxidized derivative of 5mC) at the single-cell level remain unexplored. Here, we present "CLEVER-seq" (chemical-labeling-enabled C-to-T conversion sequencing), which is a single-cell, single-base resolution 5fC-sequencing technology, based on biocompatible, selective chemical labeling of 5fC and subsequent C-to-T conversion during amplification and sequencing. CLEVER-seq shows intrinsic 5fC heterogeneity in mouse early embryos, Epi stem cells (EpiSCs), and embryonic stem cells (ESCs). CLEVER-seq of mouse early embryos also reveals the highly patterned genomic distribution and parental-specific dynamics of 5fC during mouse early pre-implantation development. Integrated analysis demonstrates that promoter 5fC production precedes the expression up-regulation of a clear set of developmentally and metabolically critical genes. Collectively, our work reveals the dynamics of active DNA demethylation during mouse pre-implantation development and provides an important resource for further functional studies of epigenetic reprogramming in single cells.

INTRODUCTION

5-methylcytosine (5mC) can be sequentially oxidized by TET family proteins, producing 5-hydroxylmethylcytosine (5hmC), 5-formylcytosine (5fC), and 5-carboxylcytosine (5caC); the latter two are excised by thymine DNA glycosylase (TDG), collectively defining the paradigm of active DNA demethylation in mammals (He et al., 2011; Ito et al., 2010, 2011; Kriaucionis and Heintz, 2009; Maiti and Drohat, 2011; Pfaffeneder et al., 2011; Tahiliani et al., 2009). During mammalian pre-implantation development, active DNA demethylation is shown to play a role in the epigenetic reprogramming in mouse early embryos (Hackett and Surani, 2013; Saitou et al., 2012). For instance, 5mC-oxidized derivatives have been found to accumulate during the late zygotic stage and are asymmetrically distributed between the paternal and maternal genomes (Inoue et al., 2011; Inoue and Zhang, 2011). More recently, we and others have demonstrated that both the paternal and maternal genomes undergo genome-wide Tet3-dependent active DNA demethylation (Guo et al., 2014; Shen et al., 2014). Furthermore, base-resolution maps of 5hmC and 5fC are generated for mouse two-cell embryos, demonstrating the existence of 5mC-oxidized derivatives in the parental genomes (Mooijman et al., 2016; Wang et al., 2014).

In recent years, single-cell epigenome sequencing technologies have revolutionized the study of DNA (hydroxyl-)methylation (Farlik et al., 2015; Guo et al., 2013; Mooijman et al., 2016; Smallwood et al., 2014), chromatin accessibility (Buenrostro et al., 2015; Cusanovich et al., 2015; Jin et al., 2015), histone modification (Rotem et al., 2015), and chromosome conformation and structure (Kind et al., 2015; Nagano et al., 2013). However, single-cell sequencing technologies for 5fC remain unexplored. Although several methods have been developed for 5fC detection in bulk samples (Booth et al., 2014; Lu et al., 2015; Neri et al., 2015; Raiber et al., 2012; Shen et al., 2013; Song et al., 2013; Sun et al., 2015; Wu et al., 2014; Xia et al., 2015), sequencing 5fC at the single-cell level is very challenging. First, 5fC is present at extremely low abundance: although it is stable in adult brains and found in many cell types and all major organs, it is present at a level of 20–200 ppm of cytosines, which is

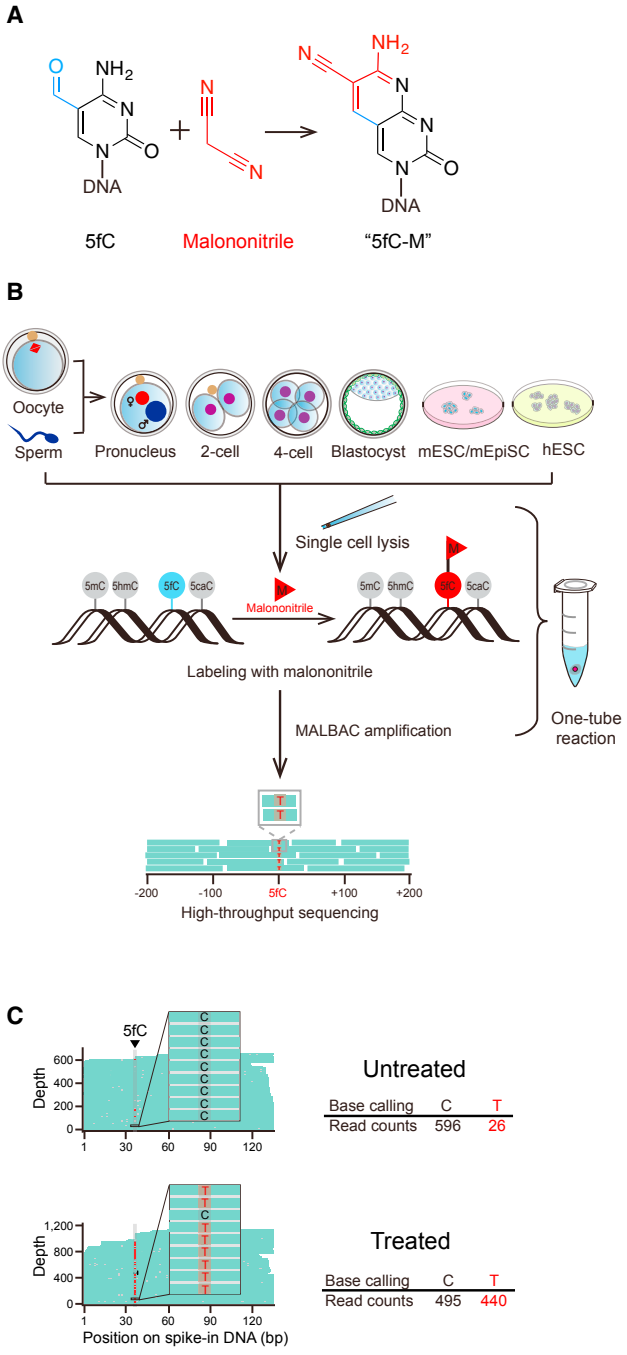


Figure 1. Concept and Validation of CLEVER-Seq for Single-Cell Studies

(A) Scheme for malononitrile-mediated selective labeling of 5fC, based on the principle of the Friedländer reaction. "5fC-M" stands for the adduct and is read as a T in sequencing.

(B) General procedure of the single-cell CLEVER-seq technology. Briefly, individual single cells were picked into the lysis buffer using mouth pipette, and malononitrile was added into the same test tube to selectively label 5fC sites. After 20 hr of incubation, MALBACs were directly performed as one-tube reaction to amplify the DNA for high-throughput sequencing. C-to-T conversions were used to identify 5fC sites (see STAR Methods for details).

(C) Chemical-labeling-induced C-to-T conversion of 5fC in a double-stranded 5fC-containing spike-in sequence. In the untreated samples, a low rate of C-to-T transition (primarily due to the mildly mutagenic nature of 5fC) was observed. In contrast, in other malononitrile-treated samples, high C-to-T conversion was repeatedly observed. It is worth mentioning that the remaining C read counts in the treated samples, which are comparable to the T read counts, come from the complementary strand of the spike-in control sequence. The G nucleotide complementary to 5fC is not labeled and does not undergo conversion, hence giving rise to the C signals during sequencing. For fair calculation of the observed C-to-T rate, a factor of 2 was multiplied to the ratio of T/(T+C) reads. See also Figure S1 and Table S1.

approximately 100-fold to 1,000-fold lower than that of 5mC (Bachman et al., 2015; Ito et al., 2011; Pfaffeneder et al., 2011). In addition, 5fC is indistinguishable from 5caC and regular Cs during bisulfite treatment (Song et al., 2012); hence, the existing bisulfite-based, single-cell sequencing technologies may not be directly applicable to 5fC detection. Moreover, the multiple conversion steps or purification procedures of the existing 5fC-detection methods may lead to the degradation and severe loss of genomic DNA from single cells, further complicating single-cell-sequencing technologies. To date, genome-wide base-resolution 5fC profiling is generally performed with tens of thousands or even millions of cells as starting materials (Song et al., 2012).

Here, we developed a novel single-cell 5fC-sequencing technique, termed "CLEVER-seq" (chemical-labeling-enabled C-to-T conversion sequencing), to map the genome-wide base-resolution 5fC landscapes in pluripotent stem cells and mouse early embryos at single-cell level. CLEVER-seq is based on a new biocompatible chemical-labeling reaction of 5fC and subsequent C-to-T conversion during amplification and sequencing. Our work reveals the dynamics of active DNA demethylation during early mouse pre-implantation development and paves the way for future functional dissection of epigenetic reprogramming at single-cell level.

RESULTS

Biocompatible and Selective Chemical Labeling of 5fC Enables C-to-T Conversion during Sequencing

We have previously developed a bisulfite-free method for 5fC detection in bulk samples (fC-CET), based on 1,3-indandione-mediated Friedländer labeling of 5fC and subsequent C-to-T conversion during sequencing (Figure S1A; Shen et al., 2011; Xia et al., 2015). However, 1,3-indandione (and its azido derivative) has only limited solubility in water; hence, although fC-CET is highly selective and robust for 5fC profiling in bulk samples, the resulting aqueous suspension makes it difficult to reproducibly label 5fC in a solution of very small amount (for instance, 2–5 μL) typical for single-cell analysis. In addition, the excess 1,3-indandione chemical needs to be removed before the labeled genomic DNA is subjected to library preparation and DNA amplification, in which several enzymes are involved. These multiple purification steps of 1,3-indandione-mediated labeling may also be unsuitable for 5fC profiling at the single-cell level. Hence, a different chemical that has not only better solubility in aqueous solution but also better biocompatibility is needed.

To identify one chemical that can meet these criteria, we expanded our scope of small-molecule screening for selective chemical labeling of 5fC. One stable and commercially available chemical, malononitrile, appeared very promising: it reacted selectively with 5fC and is very soluble in water (Figures 1A

Table 1. Summary of the Number of Uniquely Covered 5fCpG Sites in Each Single Cell and Merged 5fCpG Sites for Each Developmental Stage, See also Table S2

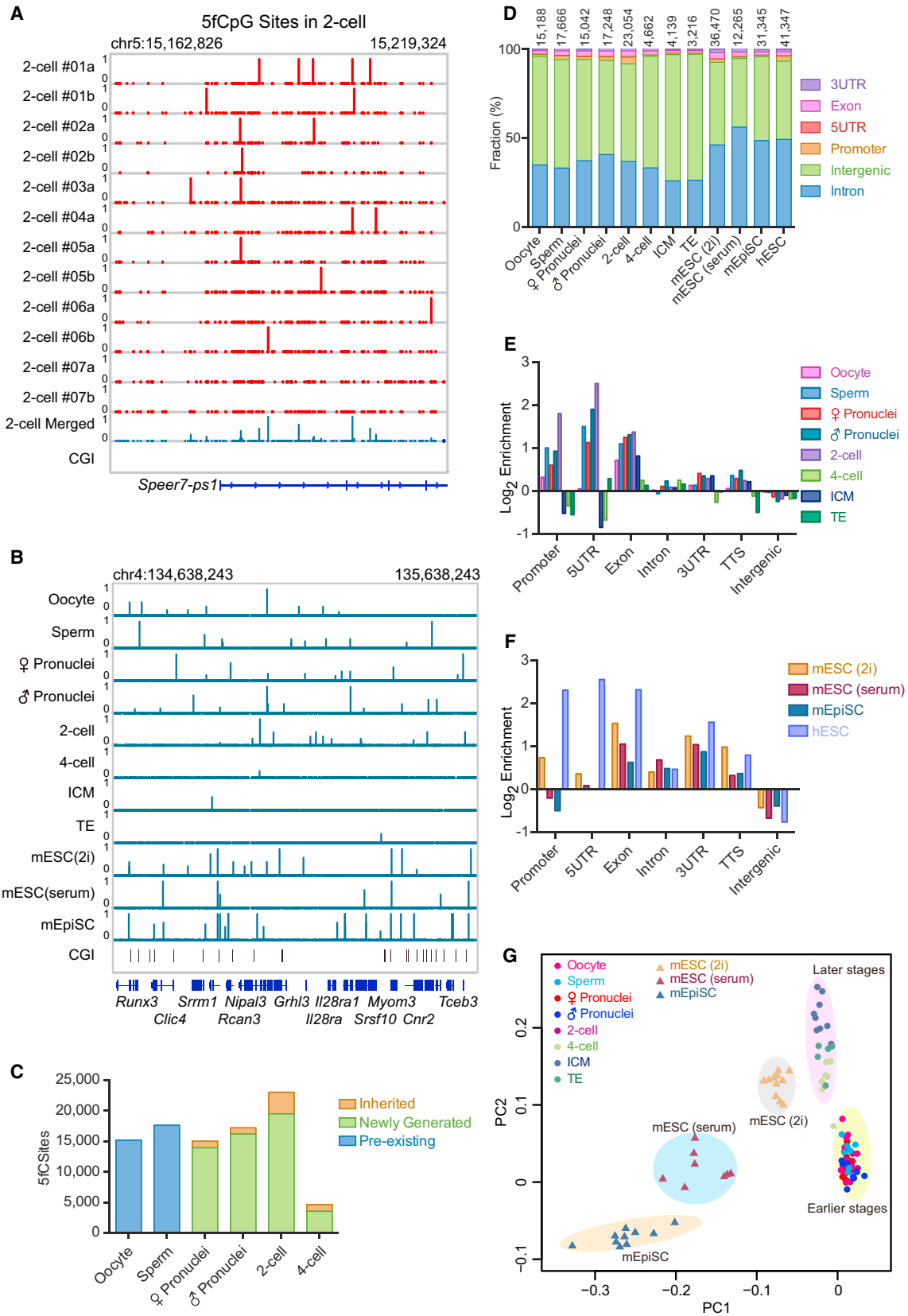
Stage	Cell No. 1	Cell No. 2	Cell No. 3	Cell No. 4	Cell No. 5	Cell No. 6	Cell No. 7	Cell No. 8	Cell No. 9	Cell No. 10	Cell No. 11	Cell No. 12	Merged
Oocyte	3,100	1,402	1,542	1,042	1,364	773	1,866	826	904	1,676	1,474	-	15,131
Sperm	2,036	2,494	3,525	2,319	2,627	2,354	3,198	-	-	-	-	-	17,665
Female pronuclei	1,067	2,447	1,706	372	1,676	1,617	2,152	2,152	1,783	1,241	-	-	15,041
Male pronuclei	1,088	2,262	1,505	2,256	2,837	1,672	4,799	1,673	-	-	-	-	17,496
Two-cell	2,038	1,373	1,808	1,157	1,983	920	1,938	3,000	2,457	2,457	1,929	1,755	21,813
Four-cell	353	1,585	863	720	337	862	471	-	-	-	-	-	4,662
ICM	544	245	932	783	1,197	414	712	419	108	-	-	-	4,139
TE	942	882	739	693	647	-	-	-	-	-	-	-	3,216
mESC (2i)	7,082	5,895	3,527	5,031	5,839	5,018	6,037	5,030	5,699	5,523	3,814	4,038	36,477
mESC (serum)	3,330	3,963	2,487	4,426	2,429	3,921	3,921	5,876	-	-	-	-	12,265
mEpiSC	11,973	8,869	10,465	14,774	11,520	7,118	6,843	8,702	6,676	7,628	-	-	31,345
hESC	8,796	10,123	10,575	12,948	12,848	11,146	7,779	7,714	9,689	9,094	-	-	41,347

and S1B–S1D). Under a wide range of pH (pH 5–8), salt and buffer conditions, malononitrile can be readily dissolved to close to molar concentration, hence making it convenient and reliable to perform labeling reactions in a solution of tiny volume (Table S1). In addition, the labeling reaction causes no observable DNA degradation, which is desirable for preserving the single-cell genomic DNA during chemical treatment (Figures S1E and S1F). More importantly, this new chemical labeling reaction is highly biocompatible: even at several hundred millimolar concentration, malononitrile does not significantly inhibit the polymerase activity of several commonly used DNA polymerases (Figure S1G; Table S1). Hence, we were able to perform the selective labeling of 5fC and subsequent DNA amplification in one single test tube, with no need of any purification procedures to remove the chemical (Figure 1B). Furthermore, because the 4' amino group of 5fC is no longer a competent proton donor in the adduct (or "5fC-M," where M denotes malononitrile) and thus fails to base pair with dG, the labeled 5fC is read as a dT during PCR (Figure S1H), similar to our previous 5fC-detection method (Xia et al., 2015). We further demonstrated that such C-to-T conversion is independent of the 5fC density and its flanking sequences (Figure S1I). We proposed that this one-tube labeling reaction, or CLEVER-seq, could be used to detect single-cell 5fC landscape at single-base resolution.

Single-Cell 5fC Sequencing at Single-Base Resolution with CLEVER-Seq

To identify whole-genome and base-resolution 5fC landscape at the single-cell level, we combined CLEVER-seq with multiple annealing- and looping-based amplification cycles (MALBACs) (which have been proven to be a robust single-cell genome amplification method; Zong et al., 2012). To test whether our approach is suitable for single-cell analysis, we first performed chemical labeling of 5fC on genomic DNA samples from individual mouse embryonic stem cells at naive state (mESCs) (grown in 2i medium). The chemical-labeling-induced C-to-T conversion of 5fC in the sequencing reads was used to obtain the genome-wide 5fC sites at single-base resolution. To directly assess the conversion rate in the context of genomic DNAs, we spiked in a synthesized DNA oligo containing one fully modified 5fC base; we also spiked in unmodified lambda DNA to evaluate the false positive rate of 5fC detection caused by undesired base transition. Under our optimized labeling conditions, a C-to-T conversion rate of ~86.4%, averaged from 12 CLEVER-seq experiments in mESCs (2i), was observed for the 5fC site, thus suggesting that more than four-fifths of the 5fC sites were reproducibly detected (Figures 1C and S2A). Importantly, although 5fC is known to be mildly mutagenic (Kamiya et al., 2002; Münzel et al., 2011), the C-to-T conversion rate was ~18-fold higher after malononitrile treatment and the labeling reaction did not cause undesired base mutation for regular cytosines in the spiked-in lambda DNA control (Figure S2A). Thus, our method shows both high sensitivity and high specificity.

With ~3× mean sequencing depth of a single-cell sample, we consistently obtained ~11.3 and ~6.4 million unique CpG sites at ≥1× and ≥3× coverage, respectively, which represents 27% and 15% CpG sites in the entire genome (Table S2). Control single-cell samples without chemical labeling recovered similar numbers of CpG sites, demonstrating that our one-tube chemical labeling strategy had a minimal effect on the MALBACs reactions. To identify genome-wide, single-5fC sites for individual single cells, we required the coverage of a given CpG site to be ≥3× and a C-to-T conversion ratio to be ≥65%; we also defined several metrics to remove SNPs and potential C-to-T mutations caused by PCR errors (Figure S2B). Using such stringent bioinformatics criteria, we achieved an overall signal-to-noise ratio of ~12 between the treated and untreated samples and identified ~3,500–7,000 5fCpG sites for each mESC (2i) (Tables 1 and S2). At the single-base level, 5fC exhibited substantial heterogeneity among individual mESCs (2i) (Figure S2C); nevertheless, in terms of their genomic distribution, 5fC sites in all 12 single mESCs (2i) were found to be enriched at similar genomic regions (Figure S2D). Moreover, the genomic pattern of the merged 5fC sites from 12 single mESCs (2i) resembled that of 5fC sites from bulk mESCs (2i), which we obtained previously using a different 5fC-labeling chemical and a distinct amplification and library preparation procedure (Xia et al., 2015). Furthermore, our single-cell 5fC data strongly correlate with the existing 5fC



(legend on next page)

data from bulk samples (Figure S2E; Wang et al., 2014; Wu et al., 2014). These results demonstrate the reliability of CLEVER-seq in single-cell 5fC profiling. Finally, our approach was able to detect copy number variations (CNVs) in an individual mESC (2i) at megabase resolution, in addition to its capability of identifying genome-wide 5fC sites at single-base resolution for a single cell (Figure S2F).

CLEVER-Seq Identifies Dynamic 5fC Landscapes at Base Resolution in Mouse Gametes and Early Embryos

We next applied CLEVER-seq to mouse sperm, metaphase II oocytes, zygotes, two-cell embryos, four-cell embryos, inner cell mass (ICM), and trophectoderm (TE) isolated from the blastocysts. The male and female pronuclei within each individual PN3-PN4 stage zygote were physically isolated by micromanipulation; the individual blastomeres of two-cell and four-cell embryos were also separated. In addition to mESCs (2i), we also performed CLEVER-seq to mESCs (serum), mouse Epi stem cells (mEpiSCs), and human embryonic stem cells (hESCs). In total, we sequenced more than 1,500 GB data for 139 single cells and identified 3,216–41,347 5fCpG sites from these different stages (Figures 2A, 2B, and S2G–S2I; see Tables 1 and S2 for details). In these different stages, we observed both inherited and newly generated 5fC sites during early pre-implantation development; yet only 7.1%, 6.8%, 17.5%, and 23.2% of the 5fC sites were inherited from sperm, oocytes, zygotes, and two-cell embryos, respectively, suggesting the highly dynamic formylation landscape after fertilization (Figure 2C). We also estimated the modification fraction for each 5fCpG based on the digitized readout of 5fC from an average of ten single cells from each developmental stage. For all developmental stages analyzed, 5fC sites tended to have a modification fraction of ~20%–40% (Figure S2G), consistent with those measured in bulk two-cell embryos and mESCs using different methods (Booth et al., 2014; Wang et al., 2014).

We next characterized the genomic distribution pattern of 5fC (Figures 2D–2F). Although approximately half of 5fC sites were located in intergenic regions, the intragenic regions had higher 5fC levels than the intergenic regions (Figures 2D–2F). Of the different genomic elements, 5fC exhibited clear enrichment in exon, intron, 3′ UTRs, and transcriptional termination sites (TTS) for all the stem cells analyzed in this study (Figure 2F). 5fC in hESCs correlated with poised enhancer (Figure S2J), similar to previous findings in mESCs (Shen et al., 2013; Song et al., 2013), and 5fC in two-cell embryos correlated with both proximal and distal H3K27ac (Figure S2K). For gametes and mouse early embryos, we also observed a large fraction of 5fC sites in repeat elements: 5fC is weakly enriched in long interspersed nuclear element (LINE) and short interspersed nuclear element (SINE) but is depleted from long terminal repeat (LTR) region (Figure S2L). Within both LINE and SINE, 5fC tends to be relatively more depleted in the evolutionarily older subfamilies, such as L2 and MIR (Figure S2M).

Heterogeneity of 5fC at the Single-Cell Level

With the 5fC landscapes in these single-cell samples, we next analyzed the potential heterogeneity of 5fC. We first performed principal-component analysis (PCA) to compare the genome-wide 5fC distribution. Sperm, oocytes, pronuclei, and two-cell embryos cluster together, whereas four-cell embryos, ICM, and TE cells cluster together and away from the earlier-stage embryos (Figure 2G). Because mESCs (2i) form a distinct cluster, we also applied this analysis to mESCs (serum) and mEpiSCs. Notably, these two types of cells form different clusters and show stronger heterogeneity than the mESCs (2i) (Figure 2G). To further analyze the heterogeneity of these different samples, we calculated their variance of 5fC abundance. For the pluripotent stem cell samples, we found that 5fC landscapes of mESCs (serum) and mEpiSCs are more heterogeneous than mESCs (2i); among gametes and early embryos, we observed the smallest 5fC heterogeneity in sperm cells (Figure S2N). Lastly, we further compared the 5fC variance (Smallwood et al., 2014) in different genomic regions and found that the intergenic regions are always more heterogeneous than the intragenic regions for all the single-cell samples we analyzed (Figure S2O). It is worth mentioning that 5fC distribution in promoters and exons appears to be the least heterogeneous, hinting regulated active DNA demethylation to these key functional genomic regions (Figure S2O).

Single-Cell 5fC Profiles Provide Direct Evidence of Active DNA Demethylation for Both Paternal and Maternal Genomes

We next sought to quantify 5fC levels of mouse early embryos using the 5fC profiles by CLEVER-seq (see STAR Methods). We first showed that 5fC levels calculated using CLEVER-seq data recapitulate the relative 5fC ratio between sperm and mESCs (2i) obtained by quantitative mass spectrometry (Figure S3A); additionally, 5fC levels by CLEVER-seq are in excellent agreement with that by sequencing bulk samples (Figure S3B). Thus, the 5fC levels measured by CLEVER-seq could be used to estimate the actual 5fC levels of the mouse early embryos and stem cells (Figures 3A and S3C).

Figure 2. The Landscape of DNA Formylation in Mouse Pluripotent Stem Cells and Early Embryos

(A) Representative snapshots of base-resolution maps of 5fC sites along the single-copy gene loci in two-cell embryos.

(B) Representative snapshots showing the modification fraction of 5fC sites in each developmental stage.

(C) Stacked barplot showing the number of the 5fC sites in gametes and early embryos. In parental pronuclei and two-cell embryos, newly generated and inherited 5fC sites are indicated with different colors. A 5fC site was defined as being newly generated if it was identified as 5fC in one stage and as a regular C in the previous stage; a 5fC site was defined as inherited if it was identified as 5fC in two consecutive developmental stages.

(D) Stacked barplot showing the distribution of 5fC in different non-overlapping genomic elements.

(E) Relative enrichment of 5fC in different genomic elements in each stage of mouse early embryos.

(F) Relative enrichment of 5fC in different genomic elements in human and mouse stem cells.

(G) PCA of the genome-wide 5fC distribution pattern (averaged in 10-kb bins) of mouse gametes, early embryos, EpiSCs, as well as naive (2i) and primed (serum) ESCs. The yellow circle indicates the cluster of single cells from earlier stages, including gametes, zygotes, and two-cell embryos; the pink circle indicates the cluster of single cells from later stages, including four-cell embryo cells isolated from ICM and TE.

See also Figure S2.

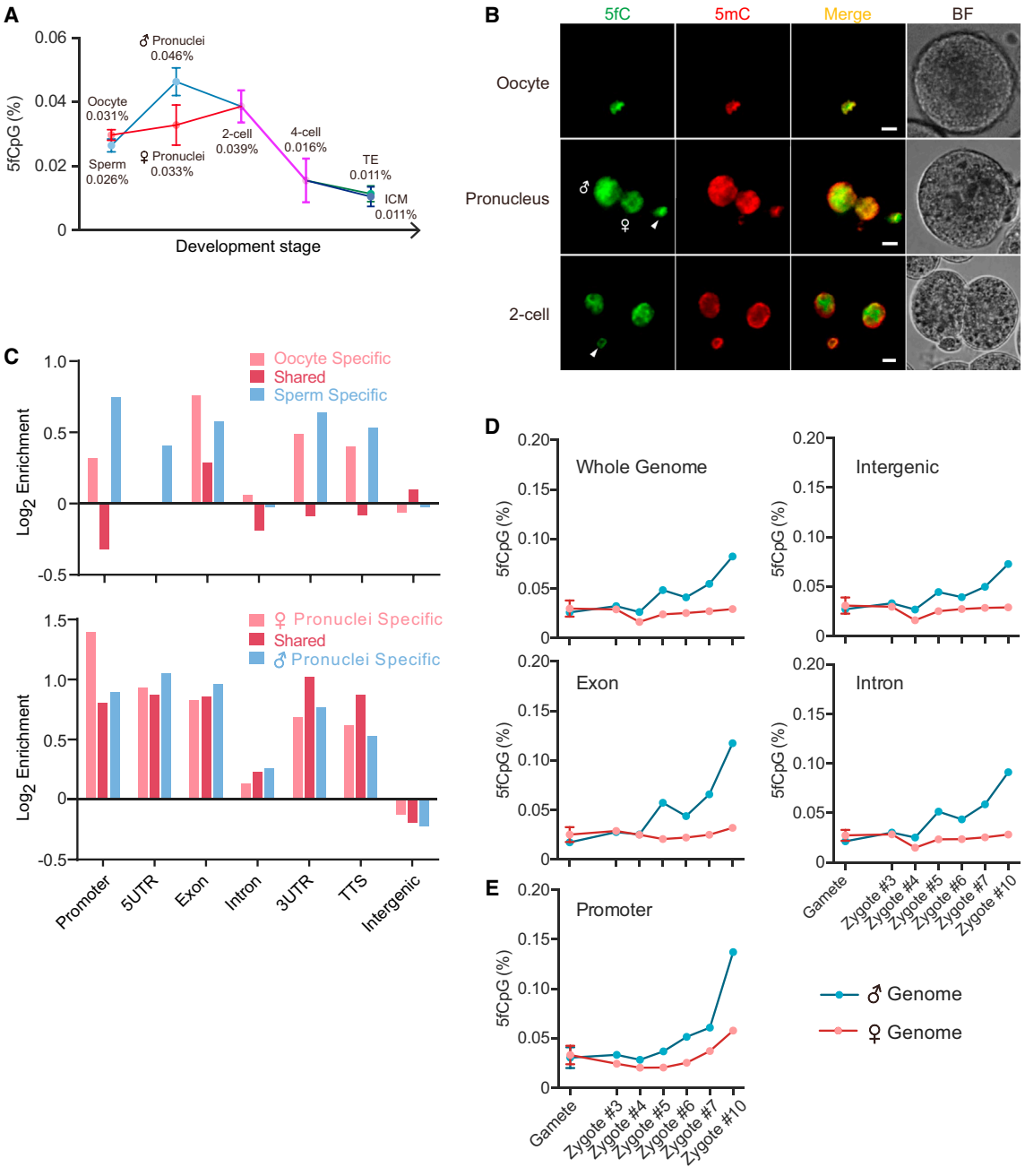


Figure 3. Dynamics of DNA Formylation in Mouse Early Embryos

(A) 5fC levels quantified using CLEVER-seq data for different mouse developmental stages. The values represent the percentage of 5fCpG in total CpG sites. Data are mean ± SEM.

(B) Immunostaining of 5mC and 5fC in mouse oocytes, zygotes, and two-cell embryos. Triangles indicate the polar bodies of the zygote and two-cell embryo, whereas male and female symbols indicate the male and female pronucleus, respectively. The bar represents 10 μm. BF, bright field.

(C) Relative 5fC enrichment in different annotated regions. Maternal- and paternal-genome-specific 5fC-marked regions are shown in pink and blue, respectively; the shared 5fC-marked regions are shown in red. The upper and bottom panels are results for gametes and pronuclei of zygotes, respectively.

(D and E) The DNA formylation dynamics in the whole genome, intergenic, exon and intron regions (D), and promoter regions (E) of the paternal (blue) and maternal (pink) genomes of each individual zygote taken at different zygotic timing within the PN3–PN4 stage. The 5fC levels for the gametes are also shown as the first points of each curve. The values represent the percentage of 5fCpG in total CpG sites.

See also Figure S3.

We found that 5fC level is nearly doubled in the paternal genome after fertilization, demonstrating dramatic active DNA demethylation of the paternal genome (Figure 3A). We also observed a mild increase of 5fC level for the maternal genome after fertilization (Figure 3A); notably, the genomic DNAs in the PN3–PN4 female pronuclei were at least partially replicated, indicating that active DNA demethylation occurs to the maternal genome as well. Additionally, the 5fC level in the two-cell embryos was approximately equal to the average abundance of the male and female pronuclei (Figure 3A), although only 17.5% of the 5fC sites observed in two-cell embryos were inherited from PN3–PN4 zygotes (Figure 2C). Finally, we performed immunostaining experiments using 5mC and 5fC antibodies and verified the presence of 5mC and 5fC in the embryos after fertilization (Figure 3B).

Synergistic Production of 5fC in Paternal and Maternal Genome after Fertilization

To compare the 5fC profiles between the paternal and maternal genomes, we binned the genome into consecutive 1-kb tiles and classified the tiles into either 5fC-marked or unmarked regions. 5fC-marked regions can be both specific and shared between the paternal and maternal genome in the gametes and zygotes (Figures 3C and S3D). In the gametes, whereas the sperm- and oocyte-specific 5fC-marked regions were both enriched in promoter, 3′ UTR, and TTS, the shared 5fC-marked regions were depleted in these elements, thus suggesting that 5fC modification may be present within similar genomic elements but at different genes (Figure 3C). Interestingly, after fertilization, the shared 5fC-marked regions were enriched in very similar genomic elements to those in both the paternal- and maternal-specific 5fC-marked regions in zygotes. This observation demonstrated that production of 5fC, and hence active DNA demethylation, are synergistic in the paternal and maternal genome after fertilization. Moreover, we performed gene ontology (GO) analysis to characterize genes with synergistic production of promoter 5fC in both parental genomes and found that these genes are enriched in the GO terms of trophectodermal cell differentiation ($p = 1.6 \times 10^{-4}$), blastocyst formation ($p = 5.3 \times 10^{-4}$), and others, among which are key transcription factors, such as *Tfap2c* and *Foxd2*. Such synergistic production of 5fC implies coordinated regulation of active DNA demethylation in both the maternal and paternal genomes after fertilization.

5fC Levels of Paired Pronuclei and Two-Cell Embryos

Out of the 18 single pronuclei used in this study, 12 (six pairs) were successively obtained from six zygotes at the PN3–PN4 stage; out of the 12 single blastomeres of the two-cell embryos, ten were paired blastomeres from five embryos. We next compared the 5fC levels among these paired single-cell samples. For the zygotes, the 5fC levels of the male pronuclei were always higher than those in the corresponding female pronuclei (Figure 3D). Moreover, a substantial increase of the whole-genome 5fC level was observed for the male pronuclei successively obtained at the PN3–PN4 stage, but no significant increase in 5fC level was observed for the female pronuclei. Notably, when we divided the whole genome into separate functional elements, a similar trend was observed for these elements, except for the promoter of the female pronuclei, which also showed a substantial increase of 5fC level (Figures 3D, 3E, and S3E). Therefore, active DNA demethylation tends to be genome wide for the paternal genome and is more restricted to particular functional elements, such as promoters, for the PN3–PN4 maternal genome. Different from that of pronuclei, no significant difference of 5fC levels was observed for the paired two-cell blastomeres (Figure S3F). Yet, a comparison of 5fC data in the two-cell embryos with bulk two-cell 5hmC profile (Wang et al., 2014) indicates that active DNA demethylation preferentially occurs to functional elements in the two-cell embryos as well (Figure S3G).

5fC Production Correlates with 5mC Demethylation

We next analyzed the loss of 5mC (by integrating published methylome data; Smith et al., 2012; see STAR Methods) for the 5fC-marked regions. 5fC-marked regions in both parental pronuclei exhibit high level of 5mC loss, suggesting positive correlation between 5fC production and 5mC demethylation during fertilization (Figure 4A). This is consistent with the knowledge that the zygotic-stage embryos undergo the most dramatic active DNA demethylation. Starting from the two-cell stage, the correlation between 5fC production and 5mC demethylation becomes less prominent (Figure 4A). We next compared 5fC production and 5mC demethylation in different genomic regions. For both intragenic and intergenic regions, the evident production of 5fC in zygotes is accompanied by dramatic loss of 5mC. From the two-cell to the four-cell stage, generation of 5fC is limited, coinciding with few sizeable 5mC changes. Interestingly, although the genome-wide loss of 5mC is less prominent in the intragenic regions, the production of 5fC is comparable between intragenic and intergenic regions in zygotes (Figure 4B). The enrichment of 5fC is most pronounced in promoters, which also have the lowest level of DNA methylation, suggesting that active DNA demethylation preferentially occurs at key regulatory elements during epigenetic reprogramming in the mouse early embryos (Figures 4B and S4).

Production of Promoter 5fC Precedes the Upregulation of Gene Expression

To investigate the temporal relationship between 5fC production and gene expression, we integrated our single-cell 5fC profiles with existing single-cell RNA sequencing (RNA-seq) data of mouse early embryos (Deng et al., 2014). We selected genes with promoter 5fC sites that were pre-existed in gametes or newly generated in the parental pronuclei and two-cell and four-cell embryos, respectively. We then calculated the normalized fold changes in the expression levels of these genes across consecutive developmental stages (Figure 5A). We found that the expression level of genes with promoter 5fC sites in oocyte and sperm were upregulated from both the mid-two-cell stage and to the eight-cell stage (oocyte) or to the four-cell stage (sperm), respectively. For genes with promoter 5fC sites newly generated in the male pronuclei, female pronuclei, and two-cell embryos, upregulation of gene expression only emerges from the mid-two-cell stage (for both parental pronuclei) or from the four-cell stage (two-cell embryos), and all lasts to the eight-cell stage. Therefore, both the promoter 5fC sites pre-existed in gametes and gained after fertilization precedes the upregulation

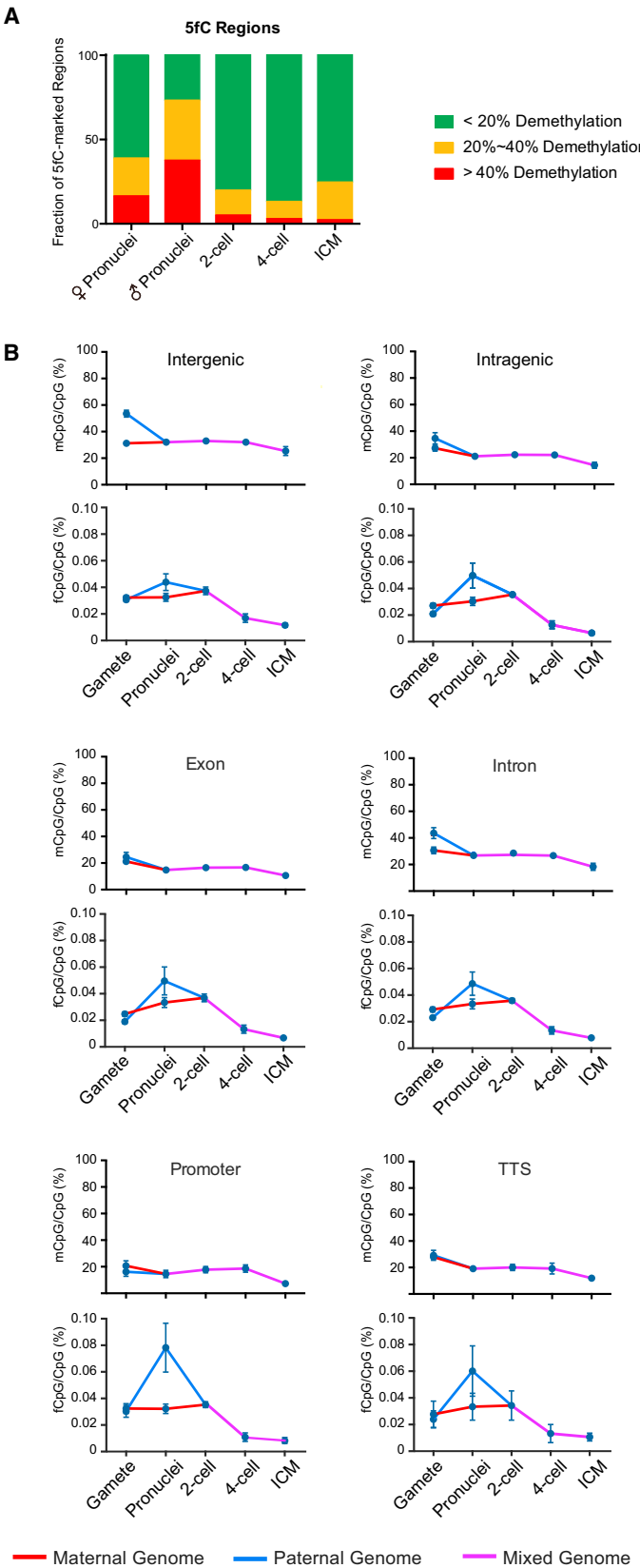


Figure 4. DNA Formylation Is Coupled with DNA Demethylation in Mouse Early Embryos

(A) Stacked barplot showing the fraction of 5fC-marked regions with different degrees of DNA demethylation.

(B) 5fC generation in different genomic regions is coupled with 5mC demethylation in mouse early embryos. The dynamics of active demethylation in male and female genomes is also distinct. Red and blue lines indicate the DNA formylation and methylation levels in maternal (red) and paternal (blue) genomes, and the purple line represents that of the mixed diploid genome. Regions include the intergenic region, intragenic region, exon, intron, promoter, and transcription termination site (TTS).

Data are mean ± SEM. See also Figure S4.

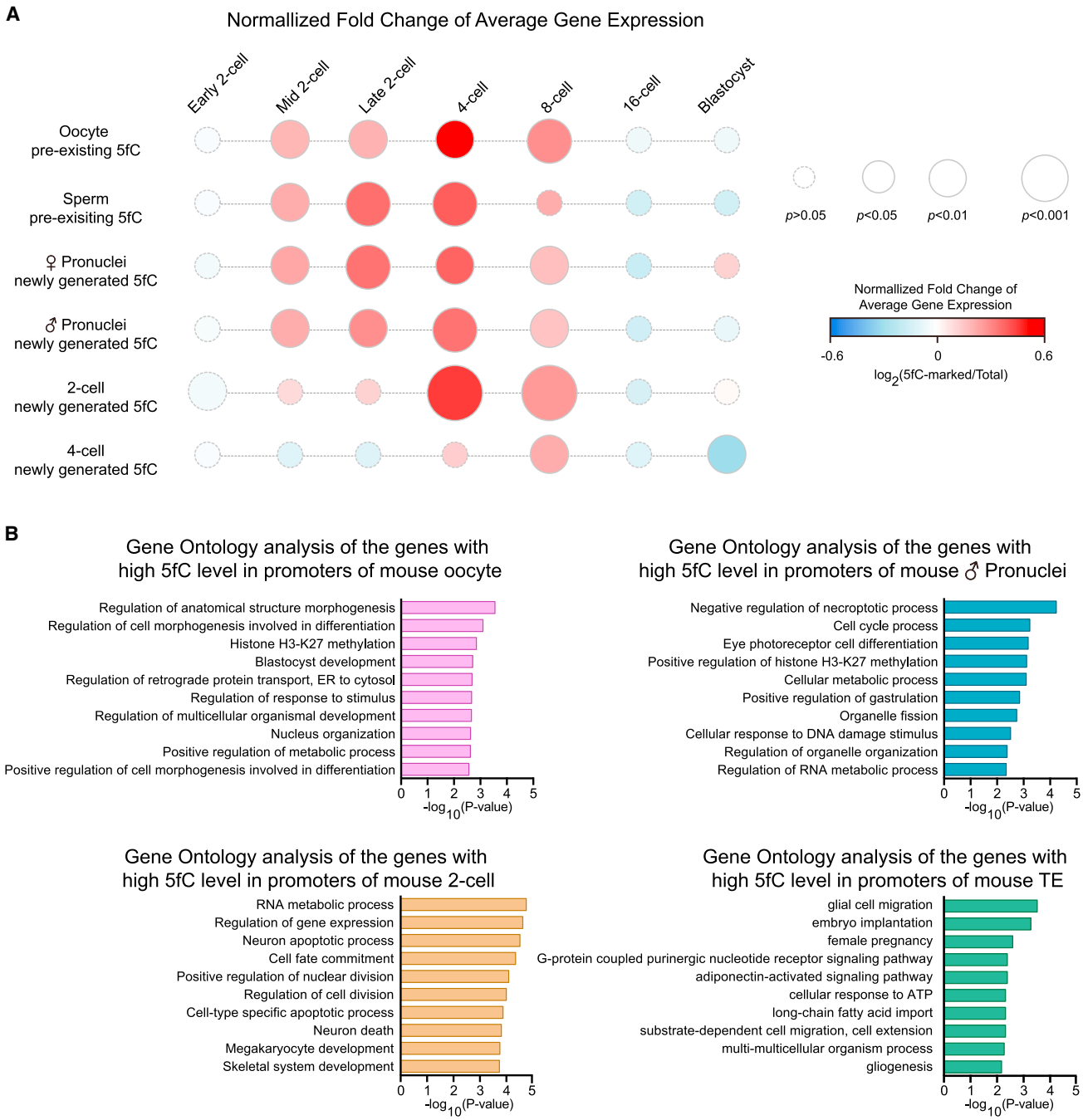


Figure 5. Generation of Promoter 5fC Precedes the Upregulation of Gene Expression in Mouse Early Embryos

(A) The normalized fold changes of gene expression levels between consecutive stages for genes with promoter 5fC in mouse early embryos. The color key, ranging from blue to red, indicates the normalized fold changes from low to high, respectively. The sizes of the circles represent the p values of normalized fold changes (between genes with promoter 5fC and total genes) of gene expression between consecutive stages (*Wilcoxon test*); to allow visualization, statistically significant and insignificant changes are represented with solid and dashed lines, respectively. The RNA-seq dataset is referred from Deng et al. (2014).

(B) GO analysis of genes whose promoters are enriched with 5fC sites in oocytes (top left panel), male pronuclei (top right panel), two-cell embryos (bottom left panel), and TE (bottom right panel), respectively.

See also Figure S5.

of gene expression. Importantly, this temporal relationship was observed only for the promoter 5fC; no temporal regulation pattern was found when intragenic 5fC sites were used for gene expression analyses (Figure S5A).

Promoter 5fC Occurs Preferentially to Genes Important for Pre-implantation Development

To predict the potential biological functions that involve 5fC modification, we identified the GO terms that are enriched for

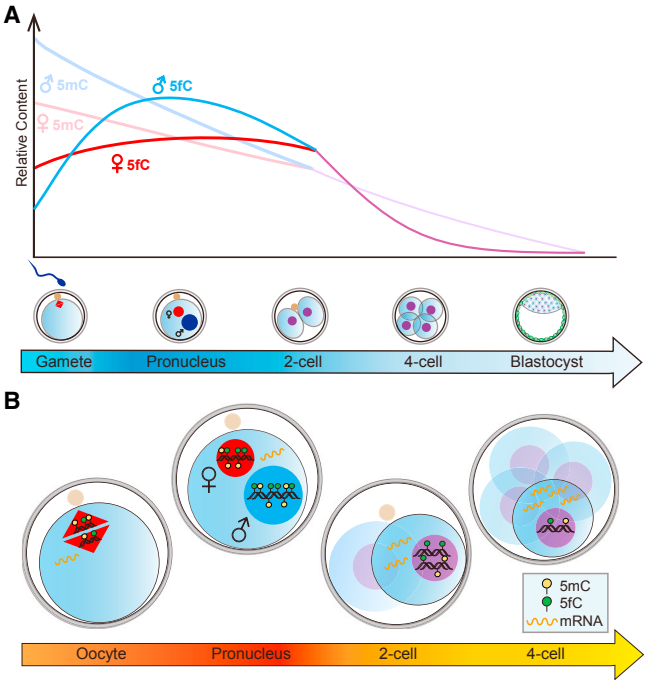


Figure 6. Models of the DNA Formylation Landscapes in Mouse Early Embryos

(A) Schematics of relative level of DNA formylation in parental genomes in mouse early embryos. The DNA methylation level is also plotted, showing the drastic DNA demethylation process in mouse preimplantation development.

(B) A model demonstrates that the production of promoter 5fC precedes the upregulation of gene expression in mouse early embryos.

the genes with 5fC sites in their promoter regions (Figures 5B and S5B). Genes with promoter 5fC were enriched in GO terms, including regulation of anatomical structure morphogenesis ($p = 2.7 \times 10^{-4}$) and blastocyst development ($p = 1.9 \times 10^{-3}$) for oocytes and negative regulation of necroptotic process ($p = 5.9 \times 10^{-5}$), cell cycle process ($p = 5.8 \times 10^{-4}$), and cellular responses to DNA damage stimulus ($p = 3.1 \times 10^{-3}$) for paternal pronuclei. For the two-cell embryos, genes with promoter 5fC were enriched in GO terms, including RNA metabolic process ($p = 1.7 \times 10^{-5}$), regulation of gene expression ($p = 2.4 \times 10^{-5}$), and cell fate commitment ($p = 4.5 \times 10^{-5}$). In agreement with our observations, previous single-cell RNA-seq of mouse pre-implantation embryos has identified stage-specific gene expression modules, which include the regulation of transcription for the four-cell and eight-cell embryos (Xue et al., 2013). Collectively, these observations suggested active DNA demethylation as a regulatory mechanism of gene expression during mouse early pre-implantation development.

DISCUSSION

In this work, we developed the CLEVER-seq method and identified the base-resolution 5fC landscapes for mouse gametes, pre-implantation embryos, and pluripotent stem cells at single-cell level. A key feature that enabled CLEVER-seq to be used for single-cell analyses is its excellent biocompatibility. Comparing to our previous 5fC detection method for bulk samples, CLEVER-seq is based on a new chemical-labeling reaction that is compatible with enzymes needed for DNA amplification and sequencing. Such biocompatibility of CLEVER-seq allowed us to perform 5fC labeling and DNA amplification in a one-tube reaction fashion and is valuable for single-cell analysis. For instance, in this work, we combined CLEVER-seq with MALBACs and were consequently able to not only sequence the single-cell 5fC profiles at single-base resolution but also obtain information about the copy number variation in the same cells (Figure S2F). Moreover, if this biocompatible labeling approach is combined with available single-cell techniques for genome, transcriptome, and/or epigenome sequencing (Dey et al., 2015; Hou et al., 2016; Macaulay et al., 2015), CLEVER-seq could have wider application in multi-dimensional omics profiling of the same individual cell.

CLEVER-seq revealed dynamic DNA formylation during mouse early pre-implantation development (Figure 6A). Because the majority of 5fC sites appeared to be stage specific, it is tempting to speculate the regulatory mechanism for the dynamic 5fC profiles. Previous evidences, including ours, have shown that the production of 5fC depends on Tet3, but the removal of 5fC is not dependent on the Tdg DNA glycosylase (Guo et al., 2014; Shen et al., 2011, 2014). One possibility is that 5fC may be further oxidized by Tet3 to 5caC, which cannot be detected by CLEVER-seq. Moreover, 5fC and 5caC have been suggested to undergo deformylation and decarboxylation reaction (Schiesser et al., 2013), although to what extent such reactions can occur in a biological system remains unclear. Of course, it is possible that the passive dilution of 5fC may also contribute to the rapid removal of 5fC after fertilization (Inoue et al., 2011).

CLEVER-seq revealed that active DNA demethylation is synergistic after fertilization. The observation that 5fC is present in similar genomic regions of different genes between sperm and oocytes suggested that the 5fC profiles of gametes are gender-specifically established during gametogenesis. Remarkably, synergistic active DNA demethylation in the zygote produced 5fC at key functional genomic elements for the same set of genes in the parental genomes, suggesting a regulated process of active DNA demethylation after fertilization. Thus, a potential regulatory mechanism that enables the synergistic active demethylation may be of interests for future studies.

Our integrated analysis shows that the production of promoter 5fC, but not 5fC sites from other genomic regions, precedes the upregulation of gene expression (Figures 5 and 6B). It has been shown that 5fC recruits transcription regulators as well as DNA repair proteins in mESCs (Iurlaro et al., 2013; Spruijt et al., 2013). Although it remains to be determined whether 5fC recruits transcription activators in mouse early embryos, the removal of promoter 5fC via DNA repair machinery might require considerable time. Alternatively, the passive removal of 5fC via DNA replication would dilute the 5fC level to half after one round of cell division. In fact, on the basis of the temporal relationship shown in Figure 5, the duration of each round of cell division appears to be the minimal interval between the production of promoter 5fC and the upregulation of gene expression. Additionally, it is also possible that other time-consuming regulatory mechanism might be involved in the delayed gene expression.

CLEVER-seq has provided, to the best of our knowledge, the first single-cell, single-base resolution 5fC landscape during

mouse pre-implantation development. Our work demonstrates the highly dynamic and synergistic nature of active DNA demethylation after fertilization, reveals the temporal and potentially causal relationship between epigenetic formylation and upregulation of gene expression, and provides an important resource for further functional studies of epigenetic reprogramming in single cells.

STAR★METHODS

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

- KEY RESOURCES TABLE
- CONTACT FOR REAGENT AND RESOURCE SHARING
- EXPERIMENTAL MODEL AND SUBJECT DETAILS
  - Animals
  - mESC culture
  - hESC culture
- METHOD DETAILS
  - Oligonucleotide and model synthesis
  - Malononitrile-based selective labeling of 5fC
  - Sanger sequencing of the labeled model DNA
  - LC-MS/MS quantification of 5fC in genomic DNA
  - Single-cell CLEVER-seq
  - CLEVER-seq (reduced representative)
- QUANTIFICATION AND STATISTICAL ANALYSIS
  - Immunofluorescence of 5mC and 5fC
  - Data processing
- DATA AND SOFTWARE AVAILABILITY
  - Software
  - Data Resources

SUPPLEMENTAL INFORMATION

Supplemental Information includes five figures and three tables and can be found with this article online at http://dx.doi.org/10.1016/j.stem.2017.02.013.

AUTHOR CONTRIBUTIONS

C.Y., F.T., C.Z., Y.G., H.G., and B.X. designed the experiments; B.X. and C.Y. conceived the labeling strategy; B.X. characterized the labeling reaction; C.Z., Y.G., and H.G. performed the experiments with the help of J.S., X.W., H.Z., and K.K.; C.Z. performed bioinformatics analysis with the help of H.G.; and C.Y. and F.T. supervised the project. C.Y., F.T., C.Z., Y.G., and H.G. wrote the manuscript with contributions from all authors.

ACKNOWLEDGMENTS

Part of the analysis was performed on the Computing Platform of the Center for Life Science. C.Y. and F.T. were supported by grants from the National Basic Research Program of China and the National Natural Science Foundation of China (91519325, MOST2016YFC0900300, 21522201, and 2014CB964900 to C.Y. and 31271543 and 81561138005 to F.T.). F.T. was also supported by the Foundation for Innovative Research Groups of the National Natural Science Foundation of China (no. 81521002). B.X., C.Z., and C.Y. are co-inventors on filed patents (WO2015043493 and 201710111600.9) for the labeling strategies and sequencing methods reported herein.

Received: August 22, 2016
Revised: December 14, 2016
Accepted: February 23, 2017
Published: March 23, 2017

REFERENCES

Bachman, M., Uribe-Lewis, S., Yang, X., Burgess, H.E., Iurlaro, M., Reik, W., Murrell, A., and Balasubramanian, S. (2015). 5-formylcytosine can be a stable DNA modification in mammals. Nat. Chem. Biol. *11*, 555–557.

Bao, S., Tang, F., Li, X., Hayashi, K., Gillich, A., Lao, K., and Surani, M.A. (2009). Epigenetic reversion of post-implantation epiblast to pluripotent embryonic stem cells. Nature *461*, 1292–1295.

Booth, M.J., Marsico, G., Bachman, M., Beraldi, D., and Balasubramanian, S. (2014). Quantitative sequencing of 5-formylcytosine in DNA at single-base resolution. Nat. Chem. *6*, 435–440.

Buenrostro, J.D., Wu, B., Litzenburger, U.M., Ruff, D., Gonzales, M.L., Snyder, M.P., Chang, H.Y., and Greenleaf, W.J. (2015). Single-cell chromatin accessibility reveals principles of regulatory variation. Nature *523*, 486–490.

Cusanovich, D.A., Daza, R., Adey, A., Pliner, H.A., Christiansen, L., Gunderson, K.L., Steemers, F.J., Trapnell, C., and Shendure, J. (2015). Multiplex single cell profiling of chromatin accessibility by combinatorial cellular indexing. Science *348*, 910–914.

Deng, Q., Ramsköld, D., Reinius, B., and Sandberg, R. (2014). Single-cell RNA-seq reveals dynamic, random monoallelic gene expression in mammalian cells. Science *343*, 193–196.

Dey, S.S., Kester, L., Spanjaard, B., Bienko, M., and van Oudenaarden, A. (2015). Integrated genome and transcriptome sequencing of the same cell. Nat. Biotechnol. *33*, 285–289.

Farlik, M., Sheffield, N.C., Nuzzo, A., Datlinger, P., Schönegger, A., Klughammer, J., and Bock, C. (2015). Single-cell DNA methylome sequencing and bioinformatic inference of epigenomic cell-state dynamics. Cell Rep. *10*, 1386–1397.

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.

Guo, F., Li, X., Liang, D., Li, T., Zhu, P., Guo, H., Wu, X., Wen, L., Gu, T.P., Hu, B., et al. (2014). Active and passive demethylation of male and female pronuclear DNA in the mammalian zygote. Cell Stem Cell *15*, 447–458.

Hackett, J.A., and Surani, M.A. (2013). Beyond DNA: programming and inheritance of parental methylomes. Cell *153*, 737–739.

He, Y.F., Li, B.Z., Li, Z., Liu, P., Wang, Y., Tang, Q., Ding, J., Jia, Y., Chen, Z., Li, L., et al. (2011). Tet-mediated formation of 5-carboxylcytosine and its excision by TDG in mammalian DNA. Science *333*, 1303–1307.

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.

Hou, Y., Guo, H., Cao, C., Li, X., Hu, B., Zhu, P., Wu, X., Wen, L., Tang, F., Huang, Y., and Peng, J. (2016). Single-cell triple omics sequencing reveals genetic, epigenetic, and transcriptomic heterogeneity in hepatocellular carcinomas. Cell Res. *26*, 304–319.

Inoue, A., and Zhang, Y. (2011). Replication-dependent loss of 5-hydroxymethylcytosine in mouse preimplantation embryos. Science *334*, 194.

Inoue, A., Shen, L., Dai, Q., He, C., and Zhang, Y. (2011). Generation and replication-dependent dilution of 5fC and 5caC during mouse preimplantation development. Cell Res. *21*, 1670–1676.

Ito, S., D'Alessio, A.C., Taranova, O.V., Hong, K., Sowers, L.C., and Zhang, Y. (2010). Role of Tet proteins in 5mC to 5hmC conversion, ES-cell self-renewal and inner cell mass specification. Nature *466*, 1129–1133.

Ito, S., Shen, L., Dai, Q., Wu, S.C., Collins, L.B., Swenberg, J.A., He, C., and Zhang, Y. (2011). Tet proteins can convert 5-methylcytosine to 5-formylcytosine and 5-carboxylcytosine. Science *333*, 1300–1303.

Iurlaro, M., Ficz, G., Oxley, D., Raiber, E.A., Bachman, M., Booth, M.J., Andrews, S., Balasubramanian, S., and Reik, W. (2013). A screen for hydroxymethylcytosine and formylcytosine binding proteins suggests functions in transcription and chromatin regulation. Genome Biol. *14*, R119.

Jin, W., Tang, Q., Wan, M., Cui, K., Zhang, Y., Ren, G., Ni, B., Sklar, J., Przytycka, T.M., Childs, R., et al. (2015). Genome-wide detection of DNase

I hypersensitive sites in single cells and FFPE tissue samples. Nature *528*, 142–146.

Kamiya, H., Tsuchiya, H., Karino, N., Ueno, Y., Matsuda, A., and Harashima, H. (2002). Mutagenicity of 5-formylcytosine, an oxidation product of 5-methylcytosine, in DNA in mammalian cells. J. Biochem. *132*, 551–555.

Kent, W.J., Baertsch, R., Hinrichs, A., Miller, W., and Haussler, D. (2003). Evolution's cauldron: duplication, deletion, and rearrangement in the mouse and human genomes. Proc. Natl. Acad. Sci. USA *100*, 11484–11489.

Kind, J., Pagie, L., de Vries, S.S., Nahidiazar, L., Dey, S.S., Bienko, M., Zhan, Y., Lajoie, B., de Graaf, C.A., Amendola, M., et al. (2015). Genome-wide maps of nuclear lamina interactions in single human cells. Cell *163*, 134–147.

Kriaucionis, S., and Heintz, N. (2009). The nuclear DNA base 5-hydroxymethylcytosine is present in Purkinje neurons and the brain. Science *324*, 929–930.

Krueger, F., and Andrews, S.R. (2011). Bismark: a flexible aligner and methylation caller for Bisulfite-Seq applications. Bioinformatics *27*, 1571–1572.

Li, H., Handsaker, B., Wysoker, A., Fennell, T., Ruan, J., Homer, N., Marth, G., Abecasis, G., and Durbin, R.; 1000 Genome Project Data Processing Subgroup (2009). The sequence alignment/map format and SAMtools. Bioinformatics *25*, 2078–2079.

Lu, X., Han, D., Zhao, B.S., Song, C.X., Zhang, L.S., Doré, L.C., and He, C. (2015). Base-resolution maps of 5-formylcytosine and 5-carboxylcytosine reveal genome-wide DNA demethylation dynamics. Cell Res. *25*, 386–389.

Ludwig, T.E., Bergendahl, V., Levenstein, M.E., Yu, J., Probasco, M.D., and Thomson, J.A. (2006). Feeder-independent culture of human embryonic stem cells. Nat. Methods *3*, 637–646.

Macaulay, I.C., Haerty, W., Kumar, P., Li, Y.I., Hu, T.X., Teng, M.J., Goolam, M., Saurat, N., Coupland, P., Shirley, L.M., et al. (2015). G&T-seq: parallel sequencing of single-cell genomes and transcriptomes. Nat. Methods *12*, 519–522.

Maiti, A., and Drohat, A.C. (2011). Thymine DNA glycosylase can rapidly excise 5-formylcytosine and 5-carboxylcytosine: potential implications for active demethylation of CpG sites. J. Biol. Chem. *286*, 35334–35338.

Mikkelsen, T.S., Ku, M., Jaffe, D.B., Issac, B., Lieberman, E., Giannoukos, G., Alvarez, P., Brockman, W., Kim, T.K., Koche, R.P., et al. (2007). Genome-wide maps of chromatin state in pluripotent and lineage-committed cells. Nature *448*, 553–560.

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.

Münzel, M., Lischke, U., Stathis, D., Pfaffeneder, T., Gnerlich, F.A., Deiml, C.A., Koch, S.C., Karaghiosoff, K., and Carell, T. (2011). Improved synthesis and mutagenicity of oligonucleotides containing 5-hydroxymethylcytosine, 5-formylcytosine and 5-carboxylcytosine. Chemistry *17*, 13782–13788.

Nagano, T., Lubling, Y., Stevens, T.J., Schoenfelder, S., Yaffe, E., Dean, W., Laue, E.D., Tanay, A., and Fraser, P. (2013). Single-cell Hi-C reveals cell-to-cell variability in chromosome structure. Nature *502*, 59–64.

Neri, F., Incarnato, D., Krepelova, A., Rapelli, S., Anselmi, F., Parlato, C., Medana, C., Dal Bello, F., and Oliviero, S. (2015). Single-base resolution analysis of 5-formyl and 5-carboxyl cytosine reveals promoter DNA methylation dynamics. Cell Rep. *10*, 674–683.

Pfaffeneder, T., Hackner, B., Truss, M., Münzel, M., Müller, M., Deiml, C.A., Hagemeier, C., and Carell, T. (2011). The discovery of 5-formylcytosine in embryonic stem cell DNA. Angew. Chem. Int. Ed. Engl. *50*, 7008–7012.

Raiber, E.A., Beraldi, D., Ficz, G., Burgess, H.E., Branco, M.R., Murat, P., Oxley, D., Booth, M.J., Reik, W., and Balasubramanian, S. (2012). Genome-wide distribution of 5-formylcytosine in embryonic stem cells is associated with transcription and depends on thymine DNA glycosylase. Genome Biol. *13*, R69.

Rotem, A., Ram, O., Shoresh, N., Sperling, R.A., Goren, A., Weitz, D.A., and Bernstein, B.E. (2015). Single-cell ChIP-seq reveals cell subpopulations defined by chromatin state. Nat. Biotechnol. *33*, 1165–1172.

Saitou, M., Kagiwada, S., and Kurimoto, K. (2012). Epigenetic reprogramming in mouse pre-implantation development and primordial germ cells. Development *139*, 15–31.

Schiesser, S., Pfaffeneder, T., Sadeghian, K., Hackner, B., Steigenberger, B., Schröder, A.S., Steinbacher, J., Kashiwazaki, G., Höfner, G., Wanner, K.T., et al. (2013). Deamination, oxidation, and C-C bond cleavage reactivity of 5-hydroxymethylcytosine, 5-formylcytosine, and 5-carboxycytosine. J. Am. Chem. Soc. *135*, 14593–14599.

Shen, Q., Wang, L., Yu, J., Liu, M., Qiu, J., Fang, L., Guo, F., and Tang, J. (2011). Synthesis of quinolines via Friedländer reaction in water and under catalyst-free conditions. Synthesis *44*, 389–392.

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.

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*, 459–470.

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.

Smith, Z.D., Chan, M.M., Mikkelsen, T.S., Gu, H., Gnirke, A., Regev, A., and Meissner, A. (2012). A unique regulatory phase of DNA methylation in the early mammalian embryo. Nature *484*, 339–344.

Song, C.X., Yi, C., and He, C. (2012). Mapping recently identified nucleotide variants in the genome and transcriptome. Nat. Biotechnol. *30*, 1107–1116.

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.

Spruijt, C.G., Gnerlich, F., Smits, A.H., Pfaffeneder, T., Jansen, P.W., Bauer, C., Münzel, M., Wagner, M., Müller, M., Khan, F., et al. (2013). Dynamic readers for 5-(hydroxy)methylcytosine and its oxidized derivatives. Cell *152*, 1146–1159.

Sun, Z., Dai, N., Borgaro, J.G., Quimby, A., Sun, D., Corrêa, I.R., Jr., Zheng, Y., Zhu, Z., and Guan, S. (2015). A sensitive approach to map genome-wide 5-hydroxymethylcytosine and 5-formylcytosine at single-base resolution. Mol. Cell *57*, 750–761.

Tahiliani, M., Koh, K.P., Shen, Y., Pastor, W.A., Bandukwala, H., Brudno, Y., Agarwal, S., Iyer, L.M., Liu, D.R., Aravind, L., and Rao, A. (2009). Conversion of 5-methylcytosine to 5-hydroxymethylcytosine in mammalian DNA by MLL partner TET1. Science *324*, 930–935.

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.

Wu, H., Wu, X., Shen, L., and Zhang, Y. (2014). Single-base resolution analysis of active DNA demethylation using methylase-assisted bisulfite sequencing. Nat. Biotechnol. *32*, 1231–1240.

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.

Xia, B., Han, D., Lu, X., Sun, Z., Zhou, A., Yin, Q., Zeng, H., Liu, M., Jiang, X., Xie, W., et al. (2015). Bisulfite-free, base-resolution analysis of 5-formylcytosine at the genome scale. Nat. Methods *12*, 1047–1050.

Xue, Z., Huang, K., Cai, C., Cai, L., Jiang, C.Y., Feng, Y., Liu, Z., Zeng, Q., Cheng, L., Sun, Y.E., et al. (2013). Genetic programs in human and mouse early embryos revealed by single-cell RNA sequencing. Nature *500*, 593–597.

Ying, Q.L., Wray, J., Nichols, J., Batlle-Morera, L., Doble, B., Woodgett, J., Cohen, P., and Smith, A. (2008). The ground state of embryonic stem cell self-renewal. Nature *453*, 519–523.

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.

Zong, C., Lu, S., Chapman, A.R., and Xie, X.S. (2012). Genome-wide detection of single-nucleotide and copy-number variations of a single human cell. Science *338*, 1622–1626.

STAR★METHODS

KEY RESOURCES TABLE

REAGENT or RESOURCE	SOURCE	IDENTIFIER
Antibodies		
Anti-5-Methylcytosine Antibody	Eurogentec	Cat#BI-MECY-1000; RRID: AB_11232874
5-Formylcytosine (5-fC) antibody (pAb)	Active Motif	Cat#61223
Goat anti-Rabbit IgG (H+L) Secondary Antibody, Alexa Fluor 594	Invitrogen	Cat#A-11012; RRID: AB_141359
Goat anti-Mouse IgG (H+L) Secondary Antibody, Cy5	Invitrogen	Cat#A10524; RRID: AB_2534033
Chemicals, Peptides, and Recombinant Proteins		
Malononitrile	J&K	Cat#261700
Hoechst 33342	Invitrogen	Cat#H3570
Trypsin-EDTA (0.05%)	GIBCO	Cat#25300-062
HTF medium	Millipore	Cat#MR-070-D
PD0325901	Selleckchem	Cat# S1036
CHIR99021	Selleckchem	Cat#S2924
LIF	Millipore	Cat#ESG1107
PMSG	Sigma	Cat#G4877
HCG	Sigma	Cat#1297001
fetal bovine serum (FBS)	GIBCO	Cat#16141079
Tyrode's solution, Acidic	Sigma-Aldrich	Cat#T1788
KSOM medium	Millipore	Cat#MR-121-D
M2 medium	Sigma	Cat#m7167
Acidic Tyrode's solution	Sigma	Cat#T1788
hESC-Qualified Matrix	Corning	Cat#354277
TeSR-E8 culture medium	Stemcell technologies	Cat#05940
DeepVent (exo-) DNA polymerase	NEB	Cat#M0259L
Proteinase	QIAGEN	Cat#19157
Glycogen	Roche	Cat#10901393001
Nuclease P1	Sigma-Aldrich	Cat#N8630
Alkaline Phosphatase	Sigma-Aldrich	Cat#P5931
Klenow Fragment (3′→5′ exo-)	NEB	Cat#M0212M
T4 DNA Ligase, HC (30 U/μL)	Thermo Scientific	Cat#EL0013
MightyAmp DNA Polymerase	Takara	Cat#R071A
Cobuddy DNA Polymerase	CWBIO	Cat#CW2396S
Taq DNA Polymerase	Transgen	Cat#AP111-01
Q5 DNA Polymerase	NEB	Cat#M0493L
DNA Degradase Plus	Zymo	Cat#E2020
DpnII	NEB	Cat#R0543L
Critical Commercial Assays		
MALBAC	Yikon Genomics	Cat#YK001B
NEBNext Ultra DNA Library Prep Kit for Illumina	NEB	Cat#E7370L
QIAquick PCR Purification kit	QIAGEN	Cat#28106
Deposited Data		
CLEVER-seq sequencing data	NCBI GEO	GEO: GSE84833
Experimental Models: Cell Lines		
Mouse: naive mESC	(Guo et al., 2013)	N/A
Mouse: prime mESC	(Ying et al., 2008)	N/A

(*Continued on next page*)

Continued		
REAGENT or RESOURCE	SOURCE	IDENTIFIER
Mouse: epiblast stem cell	(Bao et al., 2009)	N/A
Human: hESC	(Ludwig et al., 2006)	H9
Experimental Models: Organisms/Strains		
Mouse: C57BL/6	Beijing Vital River Laboratory	N/A
Sequence-Based Reagents		
Spike-in DNA oligos see Table S3	This paper	N/A
Software and Algorithms		
trim galore (version: 0.3.3)	N/A	http://www.bioinformatics.babraham.ac.uk/projects/trim_galore/
Bismark tool (version: 0.13)	(Krueger and Andrews, 2011)	http://www.bioinformatics.bbsrc.ac.uk/projects/bismark/
samtools (version: 0.1.18)	(Li et al., 2009)	http://samtools.sourceforge.net/
HOMER	(Heinz et al., 2010).	http://homer.salk.edu/homer/
Liftover	(Kent et al., 2003)	http://genome.ucsc.edu/cgi-bin/hgLiftOver
Other		
SNPs (database, version: VCFv4.1)		

CONTACT FOR REAGENT AND RESOURCE SHARING

Further information and requests for reagents should be directed to and will be fulfilled by the Lead Contact, Chenqi Yi (chengqi.yi@pku.edu.cn).

EXPERIMENTAL MODEL AND SUBJECT DETAILS

Animals

All the animal experiments were carried out in compliance with the ethical guidelines of the Peking University Institutional Animal Care and Use Committee (IACUC). The experimental animals in this study were kept at SPF-class Laboratory Animal Room. The C57BL/6-strain mice purchased from Beijing Vital River Laborartory were SPF-class wild-type. The female mice were 6-8 weeks and the male ones were 12 weeks, which were both in good healthy condition. The mice were raised in constant temperature (~22°C) and the relative humidity of 50% to 60%. The sunlight condition is 12 hr in the light followed by 12 hr in the dark. They were feed in sterile conditions, with full-nutrient forage and adequate water. The animal experiment manipulators got the specialized training and had the certification of experimental animal practice qualification. Before mating, the female mice were superovulated by injecting 5 IU PMSG (Sigma) followed by 5 IU HCG (Sigma) after 45 hr, while the male ones were without previous procedures. Superovulated female mice were either sacrificed to collect the metaphase II (MII) oocytes or mated with the males to isolate embryos. After daily vaginal plug check, E0.5 embryos (zygotes) and E1.5 embryos (2-cell embryos) were collected from the oviducts. The MII oocytes, zygotes and 2-cell embryos were further treated with acidic Tyrode's solution (Sigma) to remove the zona pellucida. Polar bodies were also removed carefully by repeated pipetting. Pronuclei were biopsied using a piezo assisted micromanipulator after stained with Hoechst 33342 (Invitrogen), and male and female pronuclei were further discriminated on the basis of their sizes and the relative distances to the polar bodies. 2-cell embryos were further cultured in KSOM medium (Millipore) to the next progressive cleavage stage to obtain the 4-cell embryos. While E3.5 blastocysts (cavitated) were collected from the uteri of the mated mice into the M2 medium (sigma), and followed by several rounds of washing. ICM and TE cells were further physically separated using the microneedles, and washed with PBS to remove the contaminants. Naked 2-cell, 4-cell embryos, ICM and TE cells were further dissociated into single cells by using Trypsin solution (GIBCO) and gentle pipetting.The spermatozoa were obtained from the caudal epididymides of adult male mice. After being incubated with HTF medium (Millipore), sperm cells with vigorous activity were chosen for further library preparation. All the oocytes, sperm cells, pronuclei of zygotes and blastomeres of 2-cell and 4-cell embryos, ICM and TE cells were washed extensively with PBS and confirmed to be free of any somatic contamination by viewing under a microscope.

mESC culture

Naive mouse Embryonic Stem Cells (mESCs) were regularly maintained on the pre-gelatinized dishes without feeders, supplemented with two inhibitors (2i): 1 μM PD0325901 (Selleckchem), 3 μM CHIR99021 (Selleckchem) in the presence of 1,000 U/ml leukemia inhibitory factor (LIF, Millipore) and 20% fetal bovine serum (FBS, GIBCO) in Dulbecco's modified eagle's medium (DMEM/F-12, GIBCO) (Ying et al., 2008). Primed mouse Embryonic Stem Cells (mESCs) were cultured with DMEM F12 (GIBCO) with 20% FBS (GIBCO) and also 1,000 U/ml LIF (Millopore). Mouse Epiblast Stem Cells (mEpiSCs) were cultured with DMEM F12 (GIBCO) supplemented with N2B27 and Activin A and bFGF (Bao et al., 2009).

hESC culture

Feeder-free human Embryonic Stem Cells (hESCs) (H9) were cultured on the hESC-Qualified Matrix (Corning) in complete TeSR-E8 culture medium (Stemcell technologies) with regular passaging (Ludwig et al., 2006).

METHOD DETAILS

Oligonucleotide and model synthesis

Oligonucleotides containing site-specific 5fC, 5mC, 5hmC, 5caC and 5fU were synthesized using the Expedite 8909 nucleic acid synthesizer using commercially available phosphoramidites (Glen Research). Deprotection and purification of oligonucleotides were carried out with Glen-Pak cartridges (Glen Reasearch) according to the manufacturer's instructions. Purified oligonucleotides were characterized by matrix-assisted laser desorption/ionization-time of flight (MALDI-TOF, ABI 7500). Regular oligonucleotides were purchased from Sangon Biotech (Shanghai). 139bp Model1_spike_in DNA was generated by primer extension using 1:1 ratio of single-stranded Model1 and Model1_Reverse sequences and subsequently purified by QIAquick PCR purification kit (QIAGEN).

Malononitrile-based selective labeling of 5fC

For malononitrile-mediated selective labeling of 5fC, 9-mer chemically synthesized model sequences containing 5fC, C, 5mC, 5hmC and 5fU respectively were used (Table S3, MS_5fC, MS_C, MS_5mC, MS_5hmC and MS_5fU). 4 μg oligonucleotide was incubated with 150 mM of malononitrile in 100 μL at 37°C for 20 hr in an Eppendorf tube in a thermomixer (Eppendorf, 850 rpm). After the reaction, we used ethanol precipitation to purify the DNAs with the help of glycogen (Invitrogen). Reaction were monitored by MALDI-TOF (ABI 7500).

For HPLC analysis of 9-mer model DNAs, 1 μg MS_5fC oligonucleotide was digested by nuclease P1 (Sigma, #N8630) and alkaline phosphatase (Sigma, #P5931) in 100 μL reaction solution at 37°C overnight. 10 μL of filtered reaction solution was injected into Agilent 1260 HPLC and separated by C-18 reverse phase column in 5%–50% acetonitrile (50 mM NH4Ac, pH 7.0); digested nucleosides were detected by UV A260 and A230.

Sanger sequencing of the labeled model DNA

For Sanger sequencing of model DNA, 20 pg of 76-mer model DNA (Table S3, Model1) was incubated with 150 mM of malononitrile with varying pH and buffers (Table S1) at 5 μL at 37°C for 20 hr in a thermomixer (Eppendorf, 850 rpm). After the reaction, we amplified the model DNA used MightyAmp (Takara, #R071A) with and/or without purification. The PCR reactions were carried out in 20 μL reactions containing primers Model1_Forward and Model1_Reverse. Model1_sequencing is used for Sanger sequencing of Model1. C-to-T conversion rate was measured by comparing the peak height of C and T at the corresponding sites of Sanger sequencing chromatography. For high-throughput sequencing of model DNA, 100 ng of model DNA (Table S3, NNfCNN, fCNfCN and fCNX_9fCN, "N" represents A, C, G or T and "X_9" is a defined sequence that is of 9 nucleotide) was incubated with 150 mM of malononitrile with varying pH and buffers (Table S1) at 50 μL at 37°C for 20 hr in a thermomixer (Eppendorf, 850 rpm). After reaction, model was purified by ethanol precipitation and mixed with unlabeled probes in in varying ratios (0, 20%, 40%, 60%, 80%, 100%). The mixed probes were then amplified using MightyAmp (Takara, #R071A), DeepVent (exo-) DNA polymerase (NEB, #M0259L), Q5 DNA Polymerase (NEB, #M0493L), Cobuddy DNA Polymerase (CWBIO, #CW2396S) and Taq DNA Polymerase (Transgen, #AP111-01).

LC-MS/MS quantification of 5fC in genomic DNA

4 μg of genomic DNA was digested by 20 U of DNA Degradase Plus (Zymo, #E2020) at 37°C overnight. After a brief desalting step with Bio-Spin P-6 Gel Columns (Bio-Rad, #7326227) and filtration, 10 μL (out of 40 μl) recovered solution was injected into LC-MS/MS (Agilent UPLC 1290 - MS/MS 6495). The nucleosides were separated by ultra-performance liquid chromatography on a C18 reverse phase column and detected by mass spectrometer set to multiple reactions monitoring (MRM) in the positive electrospray ionization mode. The nucleosides were quantified using the nucleoside to base ion mass transitions of 256 to 140 (5fC) and 228 to 112 (C) as previous reported (Song et al., 2013). Quantification and detection limits were determined by comparison with the standard curves obtained from nucleoside standards detected at the same time.

Single-cell CLEVER-seq

Single mouse embryonic stem cells, gametes and early embryos were picked into the lysis buffer (20 mM Tris, 2 mM EDTA, 20 mM KCl, and 0.3% Triton X-100) using a mouth pipette. Genomic DNAs were released after treatment with proteinase (QIAGEN) at 50°C for 3 hr. The single cell lysate was used to proceed to the next step or was freezen at −80°C for temporary storage. Trace amount of unmethylated lambda DNA and 5fC-containing model oligos (see Table S3 for details) were spiked into the lysis buffer. To label the 5fC adequately, 150 mM malononitrile (J&K) was added directly to the single cell lysate, which was allowed to react at 37°C for 20 hr with continuously shaking in a thermomixer (Eppendorf, 850 rpm).

To obtain the whole genome 5fC profiles from the single cells, MALBAC technique was applied to amplify the malononitrile-labeled genomic DNAs with minor modifications (Zong et al., 2012). Briefly, 11 rounds of quasi-linear pre-amplification and 15 cycles of exponential amplification were performed to obtain sufficient DNA for library construction. Sequencing libraries were constructed under the instruction of NEBNext Ultra DNA Library Prep Kit for Illumina (NEB), and the final libraries were amplified using the DeepVent DNA

polymerase (3′-5′ exo-, NEB). All of the quality-ensured libraries were sequenced on the HiSeq 2500 platform in 150 bp paired-end mode (sequenced by Novogene).

CLEVER-seq (reduced representative)

To demonstrate that 5fC level calculated using CLEVER-seq is consistent with the 5fC level obtained by sequencing bulk samples, we also combined CLEVER-seq with reduced representative sequencing. The labeled DNA was first digested with 9 U DpnII (NEB) at 37°C for 3h, and end polished and 3′adenylated with 5 U klenow polymerase (3′-5′ exo-, Fermentas). After heat inactivation, DNA fragments were ligated with sequencing adapters (NEB), supplemented with 30 U of highly concentrated T4 DNA ligase (Fermentas) and PCR enriched with 1.6 U Deep Vent DNA polymerase (3′-5′ exo-, NEB).

QUANTIFICATION AND STATISTICAL ANALYSIS

Immunofluorescence of 5mC and 5fC

Immunostaining was performed according to the protocol described previously (Guo et al., 2013) . Briefly, mouse early embryos were first fixed with 4% paraformaldehyde for 30 min at room temperature. After washed with PBST buffer, these cells were further permeabilized with 0.2% Triton X-100 (Sigma), and DNAs were denatured after 4N HCl treatment for 30 min at room temperature. After blocking with 0.1% PBS-BSA (Sigma) overnight at 4°C, cells were incubated with anti-5mC (1:500 dilution, Eurogentec) and anti-5fC (1:1000 dilution, active motif) antibodies overnight at 4°C. Alexa 594 goat anti-rabbit IgG (1:500 dilution, Invitrogen) or Cy5 640 goat anti-mouse IgG (1:500 dilution, Invitrogen) were used as the second antibodies. All images were acquired and analyzed using UltraVIEW VoX Spinning Disk Confocal Microscope (PerkinElmer).

Data processing

Sequencing Read Quality Control

All the de-multiplexed fastq reads were stripped to remove the low-quality bases and artificial sequences, including MALBAC primer and illumina adaptor sequences using trim galore (version: 0.3.3) (http://www.bioinformatics.babraham.ac.uk/projects/trim_galore/) and customized scripts.

Alignment

The remaining cleaned reads were then aligned to the mm9 mouse and hg19 human reference genome (downloaded from the UCSC Genome Browser) using the Bismark tool (version: 0.13) (http://www.bioinformatics.bbsrc.ac.uk/projects/bismark/) with the default parameters (Krueger and Andrews, 2011). Duplicated reads were removed using samtools (version: 0.1.18) (http://samtools.sourceforge.net/) (Li et al., 2009). The 137-bp spike-in DNA was used as an internal reference to calculate the C-to-T conversion rate for each CLEVER-seq experiment. When we estimated the CpG coverage in the merged groups of single cells, we used the highest coverage among the cells that captured the particular CpG site.

5fC site identification and filtering

To identify 5fC sites at single-cell level, we restricted our research space at CpG sites. For these sites, we counted the number of "C" bases as N_C and the number of "T" reads as N_T. Then, we used the binomial distribution having parameters N as the sequencing depth (N_C+N_T) and p as the C-to-T conversion rate from spike-in lambda DNA, to assess the probability of observing N_T by chance (Yu et al., 2012). CpG sites with p-value < 0.005 was selected and used for subsequent analysis. CpG sites with known SNPs (database, version: VCFv4.1) were discarded. To avoid potential C-to-T mutation caused by PCR errors, all CpG sites with $\geq$ 65% C-to-T conversion rate in at least two treated single cell samples were grouped as the candidate pool. To reduce the background noise, all CpG sites with $\geq$ 65% C-to-T conversion rate in at least 2 untreated control single cell samples were grouped as the background pool. For each labeled single cell sample, all CpG sites with at least 3 sequencing reads and a conversion rate of $\geq$ 65% were first intersected with the candidate pool, and then subtracted by the background pool, so as to give the 5fC sites for this single cell; all CpG sites with at least 3 reads and a mutation rate of $\leq$ 25% were identified as regular C. The 5fCpG/CpG ratio was calculated by dividing the number of 5fC sites by the number of regular Cs. The variance of 5fC distribution was calculated based on previous reported method (Smallwood et al., 2014) with minor revisions, we calculated the 5fC abundance averaged from 1-kb bins.

To evaluate false-negative detection rate (FDR), all CpG sites in untreated control cells with at least 3 reads and a mutation rate of $\geq$ 65% were intersected with the background pool and subtracted by the candidate pool, giving the false detected 5fC; all CpG sites with at least 3 reads and a mutation rate of $\leq$ 25% were identified as C sites. The false detection level was calculated by dividing the false detected 5fCs by regular Cs. The FDR was calculated by (False detected $5fCpG_{Control}/CpG_{Control}$)/($5fCpG_{Treated}/CpG_{Treated}$). The average FDR for all samples under the same C-to-T cutoff (65%) is 8.6%.

Genomic Annotations and External Datasets

Genomic element annotations were carried out using HOMER software (http://homer.salk.edu/homer/) (Heinz et al., 2010). High-density CpG promoter (HCP), intermediate-density CpG promoter (ICP), and low-density CpG promoter (LCP) annotations were all taken from the reference by (Mikkelsen et al., 2007) without any modifications. The methylation sequencing data of mouse early embryos were acquired from GSE34864 (Mikkelsen et al., 2007). RNA-seq data of mouse pre-implantation embryos were downloaded from GSE45719 (Deng et al., 2014). MAB-seq data of mESC were downloaded from (Wu et al., 2014). 5hmC and 5fC dataset of 2-cell embryo were downloaded from GSE56697 (Wang et al., 2014). H3K27ac of 2-cell embryo data were downloaded from GSE6639 (Wu et al., 2016). Human ES sequencing data is lifted over using (http://genome.ucsc.edu/cgi-bin/hgLiftOver) (Kent et al., 2003)

DATA AND SOFTWARE AVAILABILITY

Software

All software is freely or commercially available and is listed in the STAR Methods description and Key Resources Table.

Data Resources

The accession number for the CLEVER-seq sequencing data reported in this study is GEO: GSE84833.