METHODS AND RESOURCES

5-Formylcytosine landscapes of human preimplantation embryos at single-cell resolution

Yun Gao[1,2☯], Lin Li[3☯], Peng Yuan[1,4☯], Fan Zhai[1,4☯], Yixin Ren[1,4], Liying Yan[1,4,5], Rong Li[1,4], Ying Lian[1,4], Xiaohui Zhu[1,4], Xinglong Wu[1,2,6,7], Kehkooi Kee[8], Lu Wen[1,2], Jie Qiao[1,4,5,6]*, Fuchou Tang[1,2,6]*

1 Beijing Advanced Innovation Center for Genomics, Department of Obstetrics and Gynecology, School of Life Sciences, Third Hospital, Peking University, Beijing, China, **2** Biomedical Pioneering Innovaiton Center, Ministry of Education Key Laboratory of Cell Proliferation and Differentiation, Peking University, Beijing, China, **3** Guangdong Provincial Key Laboratory of Proteomics, Department of Pathophysiology, School of Basic Medical Sciences, Southern Medical University, Guangzhou, China, **4** Key Laboratory of Assisted Reproduction, Ministry of Education, Beijing, China, **5** Beijing Key Laboratory of Reproductive Endocrinology and Assisted Reproductive Technology, Beijing, China, **6** Peking-Tsinghua Center for Life Sciences, Peking University, Beijing, China, **7** Academy for Advanced Interdisciplinary Studies, Peking University, Beijing, China, **8** Center for Stem Cell Biology and Regenerative Medicine, Department of Basic Medical Sciences, School of Medicine, Tsinghua University, Beijing, China

☯ These authors contributed equally to this work.
* jie.qiao@263.net (J.Q.); tangfuchou@pku.edu.cn (F.T.)

Abstract

Epigenetic dynamics, such as DNA methylation and chromatin accessibility, have been extensively explored in human preimplantation embryos. However, the active demethylation process during this crucial period remains largely unexplored. In this study, we use single-cell chemical-labeling-enabled C-to-T conversion sequencing (CLEVER-seq) to quantify the DNA 5-formylcytosine (5fC) levels of human preimplantation embryos. We find that 5-formylcytosine phosphate guanine (5fCpG) exhibits genomic element-specific distribution features and is enriched in L1 and endogenous retrovirus-K (ERVK), the subfamilies of repeat elements long interspersed nuclear elements (LINEs) and long terminal repeats (LTRs), respectively. Unlike in mice, paired pronuclei in the same zygote present variable difference of 5fCpG levels, although the male pronuclei experience stronger global demethylation. The nucleosome-occupied regions show a higher 5fCpG level compared with nucleosome-depleted ones, suggesting the role of 5fC in organizing nucleosome position. Collectively, our work offers a valuable resource for ten-eleven translocation protein family (TET)-dependent active demethylation-related study during human early embryonic development.

Introduction

Epigenetic regulation is crucial for early embryonic development to control the expression of key regulators to complete the reprogramming process [1–4]. The dynamics of DNA methylation and chromatin accessibility have been extensively analyzed in human preimplantation

OPEN ACCESS

Citation: Gao Y, Li L, Yuan P, Zhai F, Ren Y, Yan L, et al. (2020) 5-Formylcytosine landscapes of human preimplantation embryos at single-cell resolution. PLoS Biol 18(7): e3000799. https://doi.org/10.1371/journal.pbio.3000799

Academic Editor: Christa Buecker, Max F Perutz Laboratories Center of Molecular Biology, AUSTRIA

Received: January 26, 2020

Accepted: July 13, 2020

Published: July 30, 2020

Data Availability Statement: The raw data of CLEVER-seq are deposited at the Genome Sequence Archive for Human (GSA-Human) under the accession code HRA000194, while the processed data are deposited at the Gene Expression Omnibus (GEO) under the accession code GSE124035.

Funding: This project was supported by Beijing Municipal Science & Technology Commission (Z181100001318001 to F.T.). This project was also supported by grants from National Key R&D

Program of China (2018YFC1003100 to L.W., 2018YFA0107601 and 2017YFA0102702 to F.T.). This project was also supported by grants from the National Natural Science Foundation of China (81521002 to J.Q. and F.T., 81730038 and 31871482 to J.Q., 31571544 and 31871447 to L. Y., 31625018 to F.T.). The work was also supported by the Beijing Advanced Innovation Center for Genomics at Peking University (F.T. and J.Q.). The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

Competing interests: The authors have declared that no competing interests exist.

Abbreviations: bp, base pairs; CGI, CpG island; ChIP-seq, chromatin immunoprecipitation sequencing; CLEVER-seq, chemical-labeling-enabled C-to-T conversion sequencing; CNV, copy number variation; CpG, cytosine phosphate guanine; CTCF, CCCTC-binding factor; dbSNP, database of single nucleotide polymorphism; DHS, DNase I hypersensitive sites; ERVK, endogenous retrovirus-K; GEO, Gene Expression Omnibus; HB, Holm–Bonferroni; HCP, high-density CpG promoter; hESC, human embryonic stem cell; ICM, inner cell mass; ICP, intermediated-density CpG promoter; ICSI, intracytoplasmic sperm injection; IVF, in vitro fertilisation; LCP, low-density CpG promoter; LINE, long interspersed nuclear element; LTR, long terminal repeat; MALBAC, multiple annealing- and looping-based amplification cycles; MIR, mammalian-wide interspersed repeat; NDR, nucleosome-depleted region; PCA, principal component analysis; PCR, polymerase chain reaction; RPKM, reads per kilobase of transcript per million mapped reads; scCOOL-seq, chromatin overall omic-scale landscape sequencing; SINE, short interspersed nuclear element; SNP, single nucleotide polymorphism; SVA, SINE/variable number of tandem repeats/Alu; TE, trophectoderm; TET, ten-eleven translocation protein family; t-SNE, t-distributed stochastic neighbor embedding; TTS, transcriptional termination site; UTR, untranslated region; 5caC, 5-carboxylcytosine; 5fC, 5-formylcytosine; 5fCpG, 5-formylcytosine phosphate guanine; 5hmC, 5-hydroxymethylcytosine; 5mC, 5-methylcytosine.

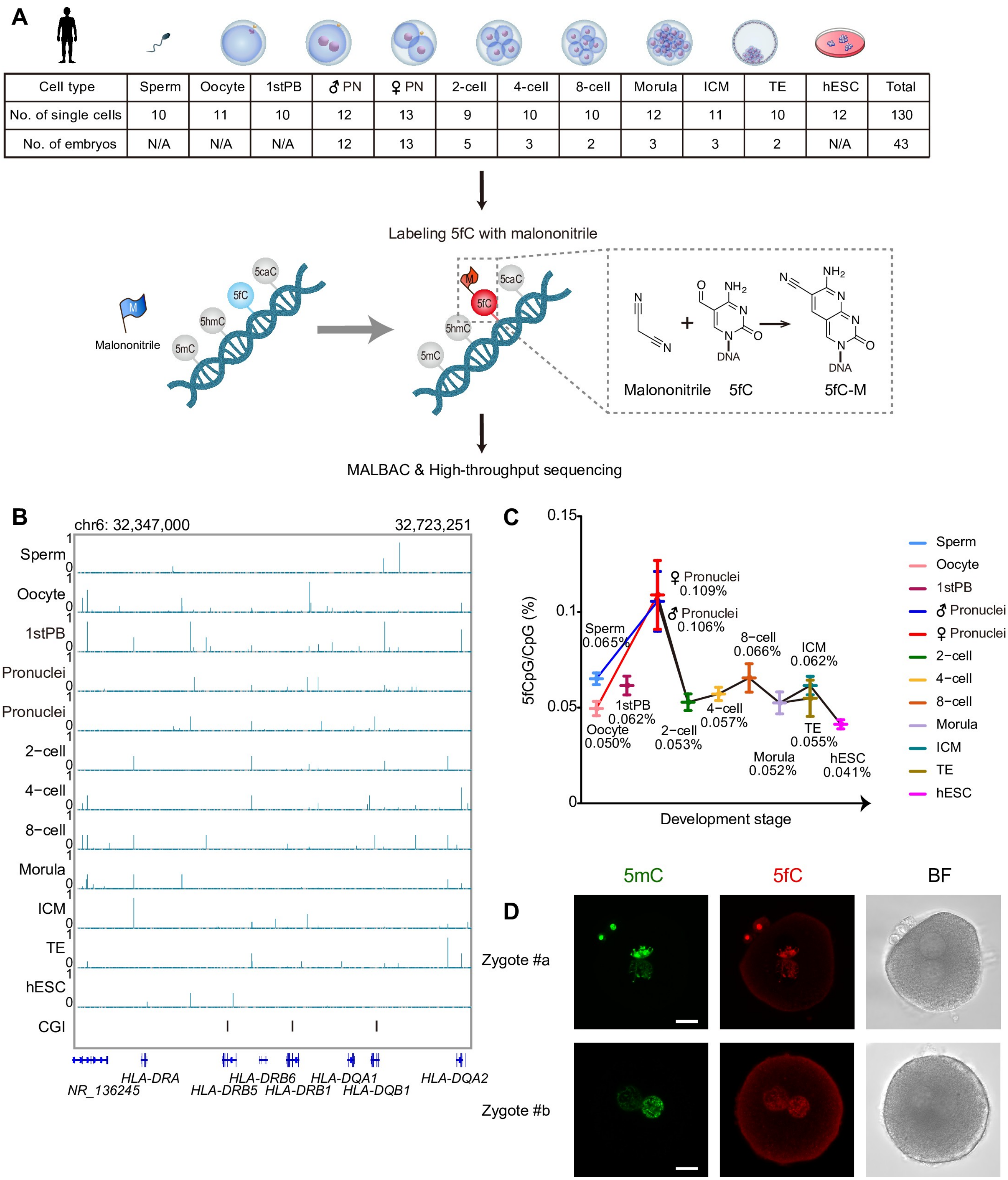
embryos [5–11]. Human early embryos go through waves of global demethylation after fertilization [12]. Active demethylation is mediated by the ten-eleven translocation protein family (TET) to produce a series of oxidized derivatives of 5-methylcytosine (5mC), including 5-hydroxymethylcytosine (5hmC), 5-formylcytosine (5fC), and 5-carboxylcytosine (5caC) [13–17]. As an essential active demethylation status, 5fC can organize the nucleosome position and be occupied by corresponding regulatory proteins [18–20]. Additionally, the production of 5fC in promoters and tissue-specific enhancers is usually associated with the up-regulation of gene expression [18,21,22]. However, the genome-wide 5fC dynamics during human preimplantation development are largely unknown. Here, we performed single-cell chemical-labeling-enabled C-to-T conversion sequencing (CLEVER-seq) to dissect the 5fC landscapes of human early embryos to provide new insights into the potential function of active demethylation in this process.

Results

Genome-wide identification of 5fC during human early embryonic development

5fC is a relatively rare DNA modification in the genome, with only 20 to 200 ppm of cytosines [15,16,23–26]. To map 5fC in a single cell's genome, we applied CLEVER-seq to human gametes, the first polar bodies, human preimplantation embryos at six key developmental stages (zygotes, 2-cell, 4-cell, 8-cell embryos, morulae, and blastocysts) and human embryonic stem cells (hESCs). In total, we obtained 130 euploid individual cells without copy number variations (CNVs) for further analysis (Fig 1A, S1A Fig and S1 Table). In CLEVER-seq, malononitrile is a key chemical that can specifically label 5fC with high efficiency [22]. The synthesized model DNA containing 5fC modifications was spiked into each sample to quantify the efficiency of malononitrile labeling (S2 Table). The average conversion rate was 79.6% in all 130 malononitrile-treated single-cell samples (S1B Fig). With approximately 5× sequencing depth for each malononitrile-treated single-cell sample, the average number of clean reads of each sample was 117.8 million with a 74.6% averaged mapping rate (S1C Fig). On average, 9.0 million, 6.4 million, and 5.0 million unique cytosine phosphate guanine (CpG) sites were covered in each malononitrile-treated sample at $\geq 1\times$, $\geq 3\times$, and $\geq 5\times$ coverage, respectively (S1D and S1E Fig).

The highly efficient single-cell genome amplification by multiple annealing- and looping-based amplification cycles (MALBAC) enabled us to detect as many 5fC sites as possible [27]. We focused on the 5fC on CpG sites named as 5fCpG. The 5fCpG calling scheme was basically according to our previous study to keep the 5fCpG sites calling consistent [22]. The stringent 5fC candidate pool was established by selecting CpG sites with $\geq 65\%$ C-to-T ratio in at least two single-cell samples with malononitrile treatment, which was expected to remove potential polymerase chain reaction (PCR) errors. In the meantime, a "background pool" was defined by selecting sites with $\geq 65\%$ C-to-T ratio in at least two untreated (negative control) single-cell samples, which was used to subtract additional background noises. Known SNP from dbSNP database hg19 v135 were removed from the candidate sites. The binomial test were used to identify 5fCpG candidate sites from our sequencing data, and CpG sites with Holm–Bonferroni (HB) method-adjusted $p < 0.01$ were selected as 5fCpG sites. A total of 24,830 to 45,398 5fCpG sites were identified from merged samples for each developmental stage analyzed (S1F Fig and Fig 1B). A total of 8,033 to 393,602 CpG sites were covered across all cells in each developmental stage at $\geq 3\times$ sequencing depth, and 0.17% to 0.89% of them were 5fCpG sites by developmental stage aggregrated analysis (S1G Fig). That is, 0.17–0.89% of the CpG sites covered in every indivual cell had formyl modification detected in at least one individual



Cell type	Sperm	Oocyte	1stPB	♂ PN	♀ PN	2-cell	4-cell	8-cell	Morula	ICM	TE	hESC	Total
No. of single cells	10	11	10	12	13	9	10	10	12	11	10	12	130
No. of embryos	N/A	N/A	N/A	12	13	5	3	2	3	3	2	N/A	43

Fig 1. The DNA formylation dynamic of human early embryos and hESC. (A) The flowchart of CLEVER-seq and sample information. The number of sperm cells, oocytes, the first polar bodies, male pronuclei, female pronuclei, blastomeres of two-cell, four-cell, eight-cell, morula, ICM and TE, and hESCs without CNV are listed

in the top table, as well as the number of embryos. The sampling time of pronucleus at zygote stage is 16 to 18 h after ICSI. (B) The UCSC browse view showing the 5fCpG sites distribution across the human early embryo developmental stages. (C) The line chart showing 5fCpG level throughout the human early embryonic development. The 5fCpG percentage was calculated by the number of 5fCpG sites divided by the sum of 5fCpG sites and unmodified CpG sites. The center is the mean of 5fCpG level and error bars are SEM. The sample size are listed in Fig 1A. The numerical data is listed in S1 Table. (D) The representative immunostaining image of 5mC and 5fC in human zygotes. The scale bar represents 20 μm. BF, bright field; CGI, CpG island; CLEVER-seq, chemical-labeling-enabled C-to-T conversion sequencing; CNV, copy number variation; CpG, cytosine phosphate guanine; hESC, human embryonic stem cell; ICM, inner cell mass; ICSI, intracytoplasmic sperm injection; MALBAC, multiple annealing- and looping-based amplification cycles; SEM, standard error of the mean; TE, trophectoderm; UCSC, University of California, Santa Cruz; 5fCpG, 5-formylcytosine phosphate guanine; 5mC, 5-methylcytosine.

https://doi.org/10.1371/journal.pbio.3000799.g001

cell of the same developmental stages. The percentage of 5fCpG in the genome of individual sperm was higher than that in the genome of individual oocyte, which was 0.065% and 0.050%, respectively (Fig 1C). The first polar body had a higher 5fCpG level (0.062%) than the oocyte, indicating the different extent of active demethylation occurred during oocyte maturation. After fertilization, the 5fCpG content drastically increased in pronuclei. At 16 to 18 h after fertilization, the male and female pronuclei presented comparable 5fCpG levels, 0.106% and 0.109%. Immunostaining of human zygotes using 5fC antibody also proved the existence of 5fC (Fig 1D). From the zygote to the 2-cell stage, the 5fCpG level decreased to 0.053% (Fig 1C). During the cleavage stages, a higher 5fCpG percentage was observed at the 8-cell stage (0.066%). The 5fCpG level of pluripotent hESCs (0.041%) was lower than that of the pluripotent inner cell mass (ICM, 0.062%). These results indicated that human early embryos exhibited extensive 5fC dynamics due to active demethylation.

The distribution characteristics of 5fCpG in human early embryos

When we compared two consecutive developmental stages, most 5fCpG sites were newly generated, suggesting that the production of 5fCpG was highly dynamic during human early embryonic development (Fig 2A) [28]; 66%, 71%, and 67% of 5fCpG sites were newly generated in male pronuclei, female pronuclei, and 2-cell stages compared with those in sperm, oocyte and pronuclei stages, respectively. However, a certain proportion, 10% to 25% of 5fCpG, was inherited from the former developmental stage. Then we explored the relationship of newly generated 5fCpG in promoter regions and RNA expression of corresponsing genes (S2A Fig). Unlike in mice, we did not observe the phenomenon that 5fCpG production in promoters procedes the up-regulation of corresponding genes in the following human early embryo developmental stages [22]. Instead, the promoter 5fCpG-marked genes exhibited the trend of up-regulation in the current developmental stage, such as the oocytes, female pronuclei, 2-cell, 4-cell, 8-cell embryos, and TE when comparing to the former stage. Only 2,823 (10%) 5fCpG sites in hESCs were shared between ICM and hESCs, indicating drastically distinct active demethylation patterns between them (Fig 2A and S2B Fig). According to genomic annotation, over half of the 5fCpG sites were located in intergenic regions, and a steady portion of 5fCpG sites were situated in the intron (30% on average) and exon regions (10% on average) due to the relatively longer length of those elements (Fig 2B). However, relative enrichment analysis showed that 5fCpG sites were enriched in functional genomic regions, such as gene body regions, transcriptional termination sites (TTSs) and enhancers (S2C Fig), whereas these sites were depleted from intergenic regions, 5′ untranslated regions (UTRs), 3′UTRs, and CpG islands (CGIs). The enrichment of 5fCpG sites in promoters was variable at different developmental stages. In most stages except for the oocyte, 8-cell and trophectoderm (TE) stages, 5fCpG sites were depleted from the promoters. Then, we calculated the 5fCpG ratio in different genomic regions. In the intergenic region, the 5fCpG level was significantly higher than that in the intragenic region in the oocyte ($p = 3.7 \times 10^{-2}$), the first polar body ($p = 2.0 \times 10^{-2}$), four-cell ($p = 1.3 \times 10^{-2}$), ICM ($p = 1.8 \times 10^{-2}$), and hESC ($p = 2.2 \times 10^{-3}$)

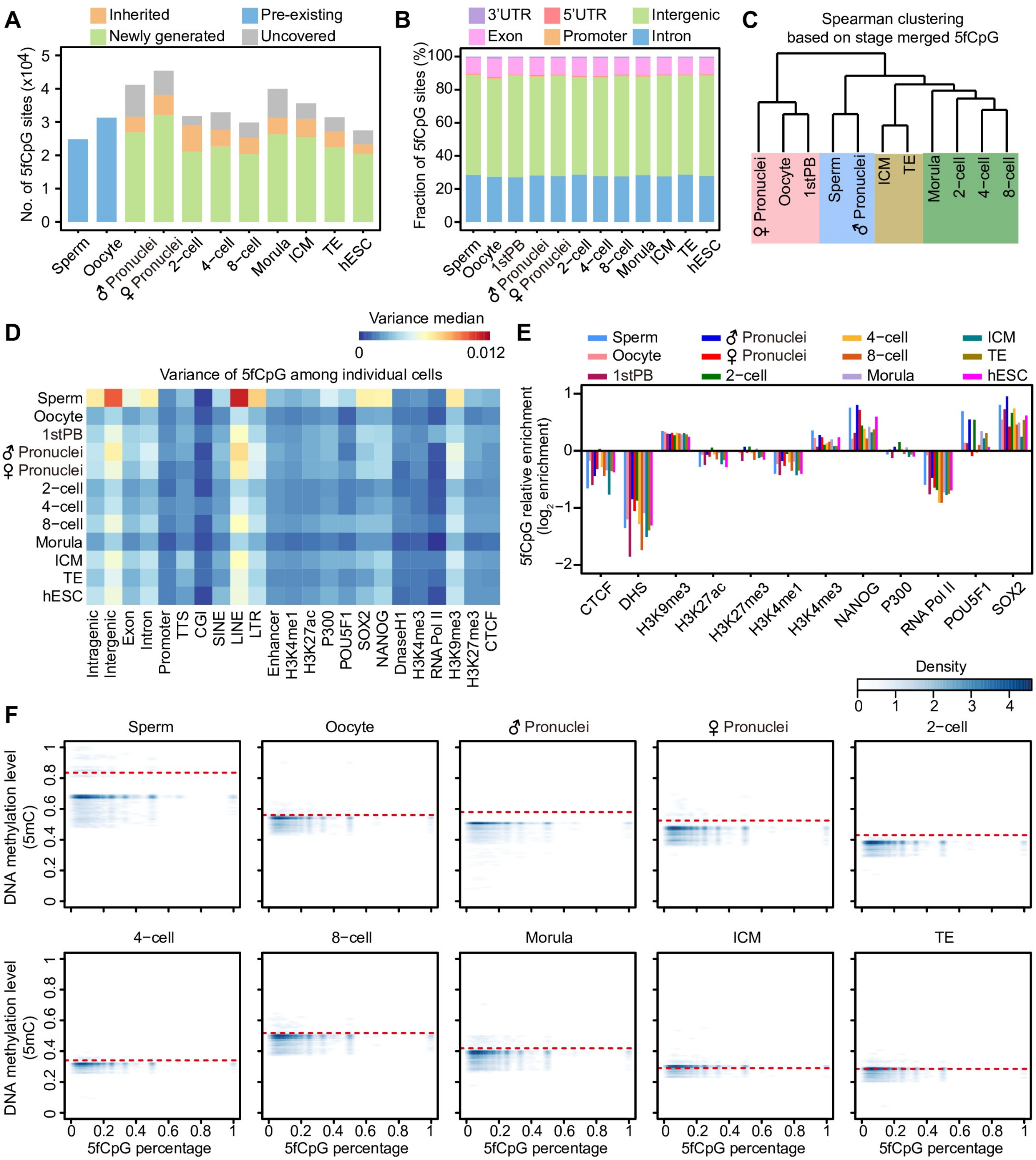


Fig 2. The production of 5fC in human early embryos. (A) The stacked bar plot showing the newly generated and inherited number of 5fCpG sites between two consecutive stages. The hESC was compared with ICM, whereas both ICM and TE were compared with morula. (B) The stacked bar plot showing the fraction of 5fCpG sites located in different genomic regions in each developmental stages. (C) Unsupervised clustering of stage merged 5fCpG sites calculated by Spearman correlation.

(D) Heat map showing the median variance of 5fCpG abundance in different genomic regions among individual cells. The variance is calculated by the 5fCpG distribution in 1-kb window. The color key from blue to red represents the value of median variance from low to high. The binding peak of histone, transcription factor, and DNase I hyper-sensitive sites are downloaded from GSE29611, GSE61475, and GSE32970, respectively. (E) Relative enrichment analysis of 5fCpG sites in distinct binding regions of transcription factor and histone as well as DNase I hyper-sensitive sites. The DHS are downloaded from GSE32970. The binding peaks of histone and transcription factor are downloaded from GSE29611, GSE61475. In (A–E), the sample size in these panels are listed in Fig 1A. In (A–B) and (D–E), the numerical data is listed in S1 Data. (F) Density plot showing the relationship of 5mC and 5fC in 5fCpG-marked 1-kb windows. The x axis shows the 5fCpG percentage in 1-kb window calculated by the number of 5fCpG divided by the sum of unmodified CpG and 5fCpG covered. The red dashed line denotes the mean of DNA methylation level calculated in 1-kb windows across the genome in each developmental stage. The DNA methylaiton data of human early embryos are from GSE81233. The numerical data is listed in S2 Data. CGI, CpG island; CpG, cytosine phosphate guanine; CTCF, CCCTC-binding factor; DHS, DNase I hypersensitive sites; hESC, human embryonic stem cell; ICM, inner cell mass; LINE, long interspersed nuclear element; LTR, long terminal repeat; SINE, short interspersed nuclear element; TE, trophectoderm; TTS, transcriptional termination site; UTR, untranslated region; 5fC, 5-formylcytosine; 5fCpG, 5-formylcytosine phosphate guanine.

https://doi.org/10.1371/journal.pbio.3000799.g002

stages (S2D Fig). Furthermore, exon, intron, and TTS regions showed higher 5fCpG levels than did other genomic elements (S3A Fig). In contrast, CGI, except for the oocyte stage, showed relatively lower 5fCpG levels that coincided with their low DNA methylation (5mC) levels (S3A Fig). When promoters were classified into high-density CpG promoters (HCPs), intermediated-density CpG promoters (ICPs), and low-density CpG promoters (LCPs) according to CpG densities, the 5fCpG abundance in HCPs, ICPs, and LCPs was similar in most stages (S3B Fig). In 2-cell stage, the 5fCpG level in HCPs was lower than that in ICPs and LPCs (Student *t* test $p = 5.0 \times 10^{-6}$ and $p = 6.9 \times 10^{-5}$, respectively). Also in male pronuclei and 4-cell stage, the 5fCpG levels in HCPs were lower than that in LCPs (Student *t* test $p = 3.0 \times 10^{-2}$ and $p = 4.5 \times 10^{-3}$, respectively).

Then, we conducted unsupervised clustering of the embryos by 5fCpG sites (Fig 2C). The oocytes, the first polar bodies, and female pronuclei clustered together but were separated from other embryo stages. Sperm and male pronuclei also clustered together. Similarly, ICM and TE clustered together. The cleavage stages, from the 2-cell to morula stage, were clustered together. These results indicated the similarity and association of 5fCpG sites in neighboring developmental stages. To estimate the heterogeneity of the 5fCpG site distribution, we calculated the variance in the 5fCpG level in a consecutive 1-kb window in the genome among cells at each developmental stage [29]. The variance in sperm was highest, up to a median of 8.9×10^{-3} (S3C Fig). The variance in the morula stage was lowest, with a median of only 3.3×10^{-3}. Then t-distributed stochastic neighbor embedding (t-SNE) analysis of 5fCpG-marked windows was performed at the window size of 1 kb, 10 kb, 100 kb, 1 Mb, 10 Mb, respectively. The cells at sperm and male pronuclei stages clustered relatively closely at the windows of 1 kb, 10 kb, and 1 Mb, while cells of different developmental stages dispersedly distributed in t-SNE map (S3D Fig). Similarly, principal component analyses (PCAs) were also conducted based on 5fCpG-marked windows. For the 1-kb and 10-kb windows, the heterogeneity among single cells of male pronuclei and female pronuclei was stronger than that of other stages (S4 Fig). For the window of 100 kb, 1 Mb, 10 Mb, the heterogeneity among single cells of male pronuclei and sperm was more dominant than that of other stages. The 5fCpG sites in the intergenic region exhibited higher variance compared with those in the intragenic region (Fig 2D). The variance in 5fCpG sites located in repeat element long interspersed nuclear elements (LINEs) was higher than that in short interspersed nuclear elements (SINEs) and long terminal repeats (LTRs). The binding sites obtaind from chromatin immunoprecipitation sequencing (ChIP-seq) data of hESC [30,31] was used as presumed ones in human early embryos. Heterochromatin regions (H3K9me3 marked) showed higher variance in 5fCpG sites than did regions marked by H3K4me3 and H3K27me3. The 5fCpG sites in promoters, CGIs, and RNA polymerase II binding sites showed relatively lower variance, suggesting a conserved 5fC distribution in functionally essential genomic regions. According to the published ChIP-seq data of hESCs [30,31], 5fCpG sites were enriched in heterochromatin regions marked by H3K9me3 and euchromatin regions marked by H3K4me3 (Fig 2E). However, these sites were depleted from the binding sites of CCCTC-binding factor (CTCF), and DNase I hypersensitive sites (DHS) in most preimplantation developmental stages. Additionally, 5fCpG were enriched in the binding sites of pluripotent transcription factors, such as NANOG, POU5F1, and SOX2. To explore the relationship between 5fCpG and 5mC, we calculated the DNA methylation level in 5fCpG-marked 1-kb windows in each developmental stage of human early embryos. Majority of 5fCpG-marked regions had relatively low methylation levels compared to the average 5mC level in the current stage in sperm, male pronuclei, female pronuclei, and 2 -cell (Fig 2F). However, this low DNA methylation tendency of 5fCpG-marked bins began to decrease since 4-cell onwards.

5fCpG distribution on repeat elements

Repeat elements have been showed to participate in the regulation of early embryonic development [32]. Enrichment analysis showed that 5fCpG sites were enriched in LINEs and satellites compared with those in SINEs and LTRs (S5A Fig). Moreover, the 5fCpG levels in LINEs and the SINE/variable number of tandem repeats/Alu (SVA) were higher than that in other repeat elements (S5B Fig). For the subfamily of repeat elements, 5fCpG sites were relatively enriched in endogenous retrovirus-K (ERVK), L1 (except for two-cell stages) and SVA (except for oocyte stages; Fig 3A). The level of 5fCpG was higher in L1, the evolutionarily younger subfamily of LINEs, compared with that in L2, the evolutionarily older subfamily (Fig 3B and 3C). The subfamilies of SINEs, Alu, and mammalian-wide interspersed repeat (MIR), showed similar 5fCpG levels. The 5fCpG level in ERVK was also higher than that in other subfamilies of LTRs (Fig 3B and 3D). Then we explored the RNA expression and DNA methylation level on L1 and ERVK with 5fCpG modification. Compared with the randomly selected ones only with unmodified CpG, these 5fCpG-marked L1 showed lower expression level in most developmental stages (S5C Fig). In ERVK, the 5fCpG-marked ones exhibited lower expression level in 2-cell embryos, morula, and TE but higher expression level in oocytes and 4-cell embryos. The markedly high 5fCpG level in L1 and ERVK may be associated with the regulation of the transcription of these repeat elements. The 5fCpG-marked L1 and ERVK showed significantly higher DNA methylation levels compared to the randomly selected ones (S5D Fig), which suggested that those repeat element regions with higher DNA methylation modification had higher tendency to retain the active demethylation marks.

Comparison of 5fCpG in paternal and maternal genomes

The parental genomes undergo different extents of demethylation and remethylation during human embryonic development [12,33]. To determine the 5fCpG features of parental genomes, we compared the 5fCpG-marked regions defined in which 1-kb consecutive windows in genome had at least one 5fCpG sites in gametes and pronuclei. A total of 3,513 (10%) 5fCpG-marked regions in oocytes were shared with those in sperm (Fig 4A). Additionally, 7,948 (20%) 5fCpG sites in female pronuclei were shared with those in male pronuclei (Fig 4B, 4C and 4D). Enrichment analysis showed that in contrast to gamete- and pronuclei-shared regions, gamete-specific and pronuclei-specific 5fCpG-marked regions were enriched in promoters, exons, CGI, and enhancers (Fig 4E). The gamete-shared and pronuclei-shared 5fCpG-marked regions were enriched in LINEs and satellites (S5E Fig), especially in L1, and depleted in MIR compared with gamete-specific and pronuclei-specific regions (Fig 4F). These results suggested that 5fCpG was distinctly distributed on parental genomes.

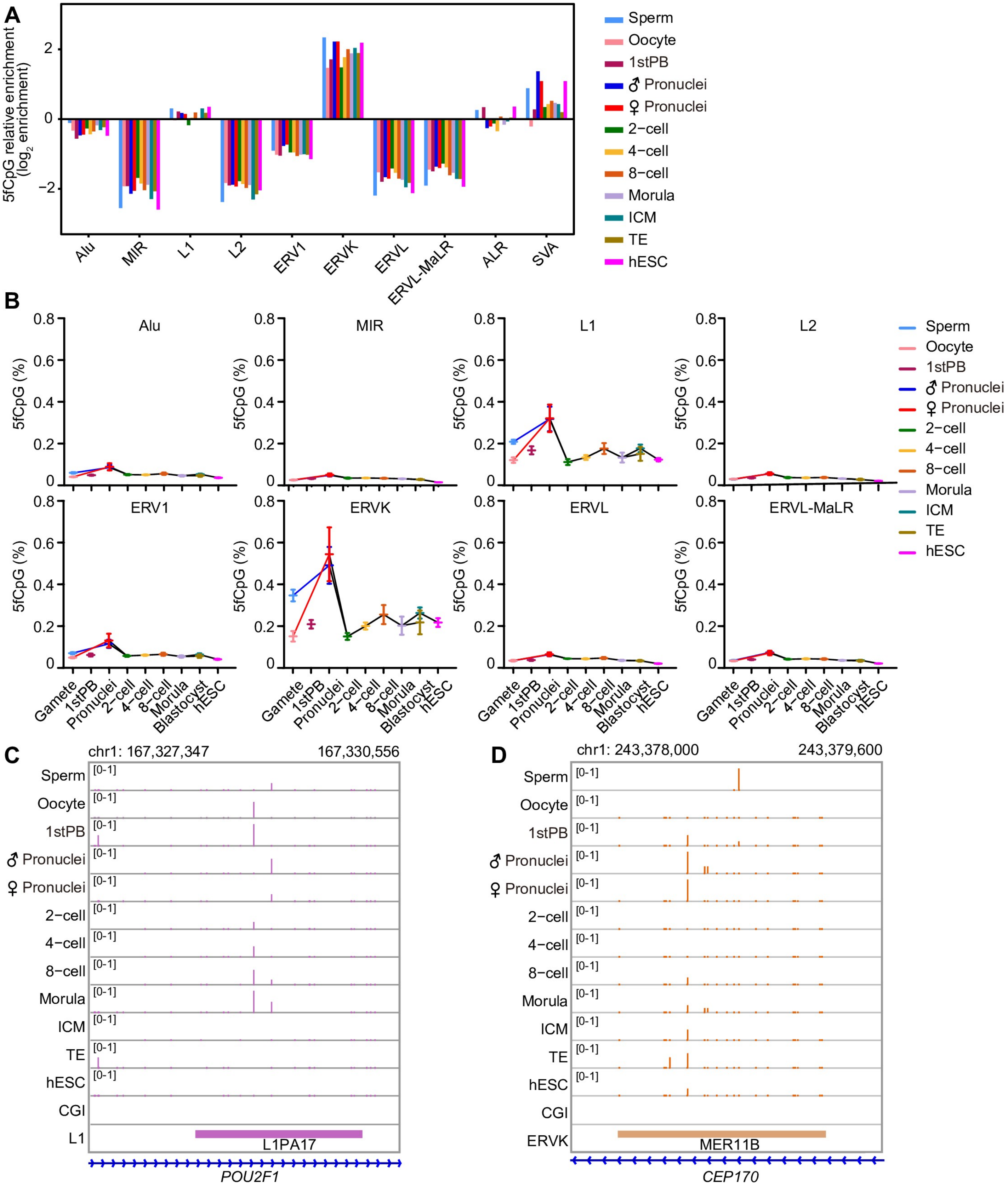

A
5fCpG relative enrichment
(log$_2$ enrichment)
2
0
−2
Alu
MIR
L1
L2
ERV1
ERVK
ERVL
ERVL-MaLR
ALR
SVA
Sperm
Oocyte
1stPB
♂ Pronuclei
♀ Pronuclei
2−cell
4−cell
8−cell
Morula
ICM
TE
hESC
B
Alu
MIR
L1
L2
ERV1
ERVK
ERVL
ERVL-MaLR
5fCpG (%)
0.8
0.6
0.4
0.2
0
Gamete
1stPB
Pronuclei
2−cell
4−cell
8−cell
Morula
Blastocyst
hESC
Sperm
Oocyte
1stPB
♂ Pronuclei
♀ Pronuclei
2−cell
4−cell
8−cell
Morula
ICM
TE
hESC
C
chr1: 167,327,347
167,330,556
Sperm
Oocyte
1stPB
♂ Pronuclei
♀ Pronuclei
2−cell
4−cell
8−cell
Morula
ICM
TE
hESC
CGI
L1
[0-1]
L1PA17
POU2F1
D
chr1: 243,378,000
243,379,600
Sperm
Oocyte
1stPB
♂ Pronuclei
♀ Pronuclei
2−cell
4−cell
8−cell
Morula
ICM
TE
hESC
CGI
ERVK
[0-1]
MER11B
CEP170

Fig 3. The 5fCpG distribution character on repeat elements. (A) Relative enrichment analysis of 5fCpG sites located in repeat elements and the subfamilies. (B) The line chart showing the 5fCpG level on subfamilies of repeat elements throughout the human early embryo developmental stages. The center is the mean of 5fCpG level and error bars are SEM. (A–B) The sample size in these panels are listed in Fig 1A, and the numerical data are listed in S1 Data. (C–D) The UCSC browser view of 5fCpG sites located in L1 (C) and ERVK (D). CGI, CpG island; ERVK, endogenous retrovirus-K; ERVL-MaLR, endogenous retrovirus-L, mammalian-apparent long-terminal repeat retrotransposon; hESC, human embryonic stem cell; ICM, inner cell mass; MIR, mammalian-wide interspersed repeat; PB, polar body; SEM, standard error of the mean; TE, trophectoderm; UCSC, University of California, Santa Cruz; 5fCpG, 5-formylcytosine phosphate guanine.

https://doi.org/10.1371/journal.pbio.3000799.g003

Variable differences of 5fCpG levels in paired pronuclei

In the pronuclei stage, the 5fCpG level is higher than that in other stages, which is similar in mice [22]. In mice, for paired pronuclei, the 5fCpG levels of male pronuclei are always higher than those of paired female pronuclei. However, in humans, paired pronuclei showed a distinct pattern. In total, we successfully obtained 10 paired pronuclei to quantify 5fCpG levels. Five pairs presented higher 5fCpG levels in male pronuclei, and the highest difference was up to 0.067% (Fig 5A and 5B). In contrast, in the remaining 5 pairs, female pronuclei had higher 5fCpG levels than did paired male pronuclei. That is, in some zygotes the 5fCpG levels of paternal gemone were higher than that of maternal genome, whereas in others it was opposite. To a certain extent, the differences in both intergenic and intragenic regions collectively contributed to the differences in the whole genome (Fig 5C). However, the difference in 5fCpG levels in enhancers and promoters between paired parental nuclei was nonsignificant. The most drastic differences in paired pronuclei resulted from the repeat elements, LINEs and ALRs (Fig 5D). In particular, the subfamilies L1 and ERVK showed the most significant differences. The variable differences of 5fCpG levels in paired pronuclei were completely different from those in mice, mainly resulting from the difference in 5fCpG on repeat elements.

5fCpG levels in paired two-cell embryos

For paired blastomeres in two-cell embryos, we compared 5fCpG levels between two blastomeres from the same embryo. Three pairs of two-cell blastomeres showed different 5fCpG levels, and one pair showed similar levels (S6A Fig). Additionally, the difference in 5fCpG levels was maintained in enhancers, promoters, exons, and introns (S6B Fig). For subfamilies of repeat elements, these four pairs of blastomeres showed similar 5fCpG levels on Alu, MIR, L2, ERV1, ERVL, and ERVL-MaLR (S6C Fig). However, the difference in 5fCpG distribution was mainly exhibited in L1 and ERVK, subfamilies of LINEs, and LTRs, respectively, which resembled the situation in paired pronuclei (Fig 5D).

Regulatory elements enriched in 5fC-marked regions

Taking advantage of the DNase-seq data of hESCs as putative DHS of human pluripotent ICM at the blastocyst stage [34,35], we analyzed the chromatin state surrounding 5fC sites. Throughout all developmental stages we analyzed, compared with its flanking regions, the center of 5fC sites showed relatively less chromatin accessibility (Fig 6A). For the proximal 5fC sites (within TSS ± 2 kb), this phenomenon was even more prominent. However, the shore regions (approximately 1392 base pairs [bp] upstream or downstream of TSS) around the center showed strong DHS signals (Fig 6A). Using the chromatin overall omic-scale landscape sequencing (scCOOL-seq) data of human early embryos [7], we analyzed the 5fCpG levels on the nucleosome-occupied regions and nucleosome-depleted regions (NDRs; Fig 6B). Interestingly, the 5fCpG level in the nucleosome-occupied regions was higher than that in NDRs in most developmental stages, such as the oocyte ($p = 1.5 \times 10^{-2}$), two-cell ($p = 2.2 \times 10^{-2}$), four-cell ($p = 1.2 \times 10^{-2}$), ICM ($p = 1.3 \times 10^{-3}$), and hESC ($p = 5.9 \times 10^{-6}$) stages. This finding is

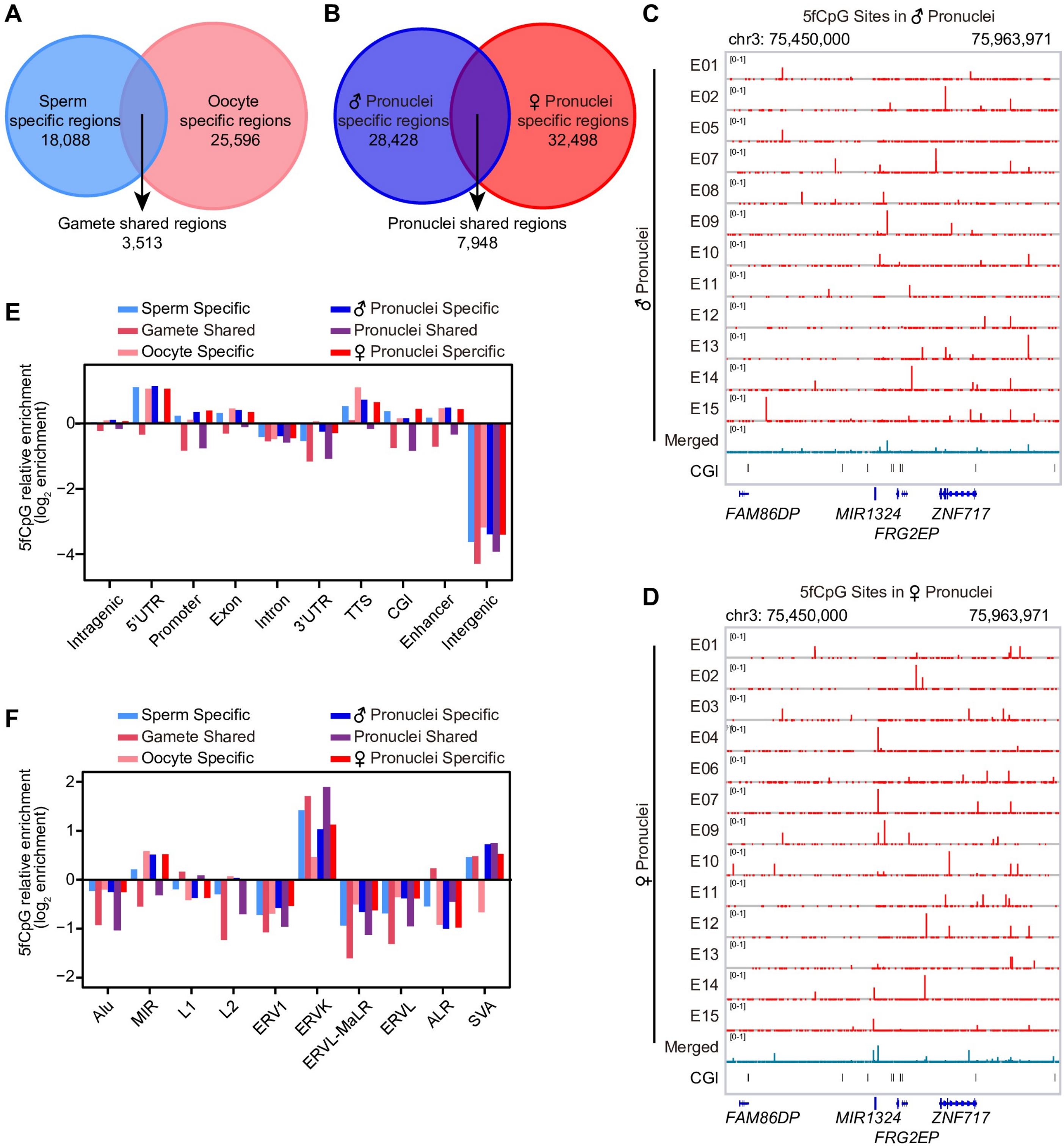


Fig 4. The features of 5fCpG sites distribution on paternal and maternal genome. (A) Venn diagram showing the overlap of 5fCpG-marked regions (1-kb window) between sperm and oocyte. The number of sperm-specific, oocyte-specific, and gamete-shared 5fCpG-marked regions are listed in the diagram. (B) Venn diagram showing the overlap of 5fCpG-marked regions (1-kb window) between male pronuclei and female pronuclei. The number of male pronuclei-specific, female pronuclei-specific, and pronuclei-shared 5fCpG-marked regions are listed in the diagram. (A–B) The sample size in these panels are listed in Fig 1A. (C–D) The UCSC browser view of 5fCpG sites in individual male pronucleus (C) and female pronucleus (D). (E) Relative enrichment of gamete-specific, gamete-shared, pronuclei-specific, and pronuclei-shared 5fCpG-marked regions (1-kb window) in distinct genomic region. (F) Relative enrichment of gamete-specific, gamete-

shared, pronuclei-specific, and pronuclei-shared 5fCpG-marked regions (1-kb window) in repeat elements and subfamilies. (E–F) The sample size in these panels are listed in Fig 1A, and the numerical data are listed in S1 Data. CGI, CpG island; MIR, mammalian-wide interspersed repeat; TTS, transcriptional termination site; UCSC, University of California, Santa Cruz; UTR, untranslated region; 5fCpG, 5-formylcytosine phosphate guanine.

https://doi.org/10.1371/journal.pbio.3000799.g004

compatible with previously published results showing that 5fCpG could interact with histones and organize nucleosome positioning [18,36].

According to previous studies, 5fCpG on tissue-specific enhancers is associated with the up-regulation of the expression of corresponding genes [18,37]. We performed motif analysis of distal and proximal 5fCpG-marked genomic regions to find out the enriched transcription factors in these regulatory regions. In the distal 5fCpG-marked regions (2 kb away from TSS), we found that the enhancer-associated protein EP300 that was expressed across the human preimplantation development, which acts as histone acetyltransferase to regulate transcription and is vital for embryonic development and cell proliferation [38,39], showed consistant enrichment in distal 5fCpG-marked regions (Fig 6C). The binding motif of EWSR1 involving in mitosis [40] and neuronal morphology [41] was strongly enriched in distal 5fCpG-marked regions with gradually increased expression across all stages analyzed. ZNF136, ZNF281, and ZNF675, the members of zinc-finger family, also exhibited significant enrichment in these regions. Besides, the binding motifs of pluripotency transcription factors, such as POU5F1 and SOX2, were significantly enriched. Moreover, the binding motifs of several transcription factors that determined cell fate during early lineage differentiation, such as SOX4, SOX15, and GATA6, were also enriched in these distal 5fC-marked regions. For the proximal 5fCpG-marked genomic regions (TSS ± 2 kb), the binding sequences of transcription factors regulating the cell cycle and mitosis, such as E2F3 and EGR1, were enriched (Fig 6D). Additionally, the binding motif of MAZ, a transcription factor regulating transcription complex formation which was expressed in all stages analyzed, was consistently enriched in the proximal 5fCpG-marked regions across human early embryonic development. Collectively, these analyses demonstrated that 5fCpG-marked regulatory regions exhibited the tendency of important transcription factors' binding during human early embryonic development.

Discussion

After fertilization, embryos undergo drastic epigenomic reprogramming, especially TET-mediated DNA demethylation. Here, we used single-cell CLEVER-seq to depict the DNA formylation profiles of human early embryos. During preimplantation development in humans, pronuclei (16–18 h after intracytoplasmic sperm injection [ICSI]) exhibited the highest 5fCpG levels. Male pronuclei and female pronuclei showed comparable 5fCpG levels, although the paternal genome experienced a much more extensive demethylation process from sperm to male pronuclei compared to maternal genome [12]. Furthermore, we analyzed 10 pairs of pronuclei from the same zygotes and found variable 5fCpG content in pronuclei. That is, in some zygotes, 5fCpG levels were higher in male pronuclei than in female ones, whereas in other zygotes it was the opposite. Different from mice, in 5 pairs of human pronuclei, the 5fCpG level in male pronuclei was higher than that in paired female pronuclei; the remaining 5 pairs showed the opposite phenomenon that female pronuclei had a higher 5fCpG level than did male pronuclei. Because *TDG* is extremely lowly expressed in human early embryos, it may not account for the inconsistent 5fCpG level in paired pronuclei [5,42]. Using immunostaining, our previous work showed that the 5hmC distribution in male and female pronuclei of human was also inconsistent: although majority of male pronuclei showed stronger 5hmC signals, still some female pronuclei exhibited higher fluorescent intensity of 5hmC compared with paired male one [5]. In mice zygotes, the 5fC and 5hmC level in paternal genome was always higher than that in maternal genome in the same embryo [22,43]. Different from mice,

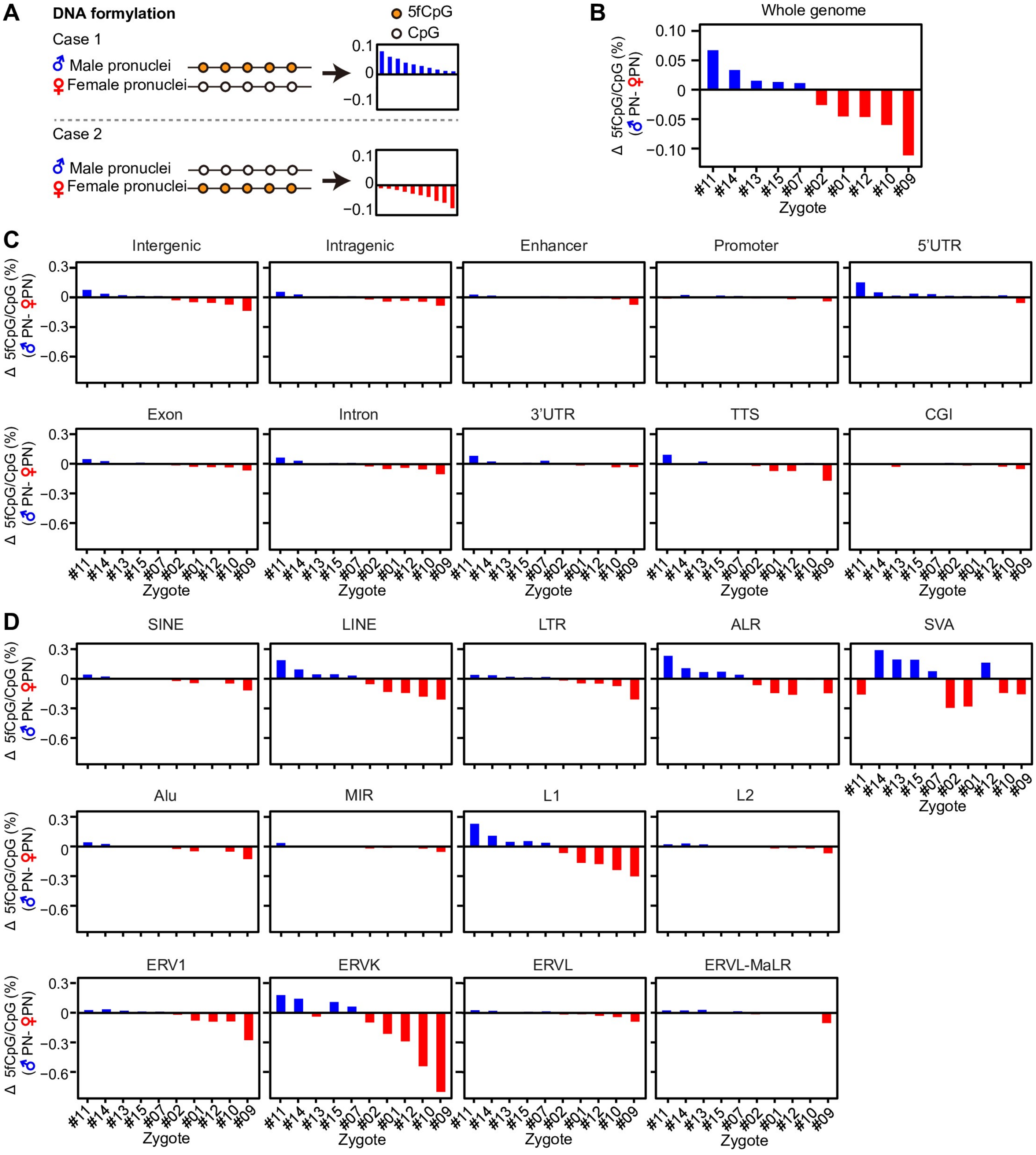


Fig 5. The asymmetric 5fCpG distribution in paired pronucleus. (A) The illustration diagram showing the two cases of 5fCpG level in paired pronucleus. The bar plot is used to show the difference of 5fCpG level in male pronuclei minus that in corresponding female one. The bar is in blue if the male pronuclei has higher 5fCpG level than female one (difference > 0), otherwise it is in red. (B) The bar plot showing the difference of 5fCpG level between paired pronucleus in whole genome ($n = 10$). The difference is calculated by 5fCpG level in male pronuclei minus that in corresponding female one. The order of paired pronuclei is ranked by the value of

difference from high to low. (C) The bar plot showing the difference of 5fCpG level between paired pronucleus in distinct genomic region (n = 10). The pronuclei are ranked by the order showed in Fig 5B. (D) The bar plot showing the difference of 5fCpG level between paired pronuclei in distinct repeat elements (n = 10). The pronuclei are ranked by the order showed in Fig 5B. (B–D) The numerical data are listed in S1 Data. ALR, alpha-satellite repeat; CGI, CpG island; CpG, cytosine phosphate guanine; ERV1, endogenous retrovirus-1; ERVK, endogenous retrovirus-K; ERVL, endogenous retrovirus-L; ERVL-MaLR, endogenous retrovirus-L, mammalian-apparent long-terminal repeat retrotransposon; LINE, long interspersed nuclear element; LTR, long terminal repeat; MIR, mammalian-wide interspersed repeat; SINE, short interspersed nuclear element; SVA, SINE/variable number of tandem repeats/Alu; TTS, transcriptional termination site; UTR, untranslated region; 5fCpG, 5-formylcytosine phosphate guanine.

https://doi.org/10.1371/journal.pbio.3000799.g005

the active demethylation extent between parental genome in human zygotes may present in an embryo-specific manner due to the complicated genetic background of human embryos. The obvious difference in 5fCpG in paired pronuclei came from the 5fCpG in repeat element L1 and ERVK, the subfamilies of LINEs and LTRs. In addition to pronuclei, early embryos in other stages showed relatively higher 5fCpG abundance in L1 and ERVK. 5fC could reduce the specificity of the substrate to RNA polymerase II; thus, the high level of 5fC in those repeat elements may reduce their transcription to improve genome stability [44,45].

The formyl group of 5fC has high activity to react with the amine residues of protein to form the Schiff base complex [18]. The transcription regulators selectively binding to 5fC had been identified by proteomics screening, such as FOXK1, FOXP1, and FOXI3 [19]. In human embryos, we also found that the binding motifs of essential transcription factors, such as POU5F1 and SOX2 to maintain pluripotency and E2F3 and EGR1 to regulate the cell cycle, were significantly enriched in 5fCpG-marked regulatory regions. In addition to transcription factors, the lysines of histone proteins can interact with the formyl group of 5fC [36,46]. Consistent with the above data, our data showed that the 5fCpG level in nucleosome-occupied regions was higher than that in nucleosome-depleted regions in many developmental stages. The function of 5fC in organizing nucleosome position and altering the structure of the DNA helix could contribute to transcription control [18,36,47]. Increasing evidence shows that 5fC does not only serve as an intermediate of active DNA demethylation but also participates in functional regulations [48]. In the future, functional work is needed to determine the comprehensive roles of 5fC in vivo.

For a better understanding of the complicated active DNA demethylation process after fertilization, the features of 5hmC and 5caC also remain to be investigated, which rely on the improvement of techniques to accurately quantify these DNA modifications [49–52]. In summary, our work presents the 5fC landscapes of human early embryos, which provides new insights into the complex regulation network during human preimplantation development.

Methods

Ethics statement

The research project was approved by the Reproductive Study Ethics Committee of Peking University Third Hospital (research license number 2017SZ-081), and the study strictly followed the guidelines set forth by the Ethics Committee. Oocytes were voluntarily donated by female donors at the Center for Reproductive Medicine, Peking University Third Hospital, with signed written informed consent. Sperm donation was provided by a healthy man with proven fertility who signed the written informed consent.

Standard in vitro fertilisation (IVF) protocols for the gametes and embryo collection, thawing, and culture were executed as previously described [53,54].

Collection of human gamete and early embryos

Gametes and the first polar body. After several rounds of washing with HTF medium (Life global), the swim-up sperm were collected for further processing. Motile single sperm

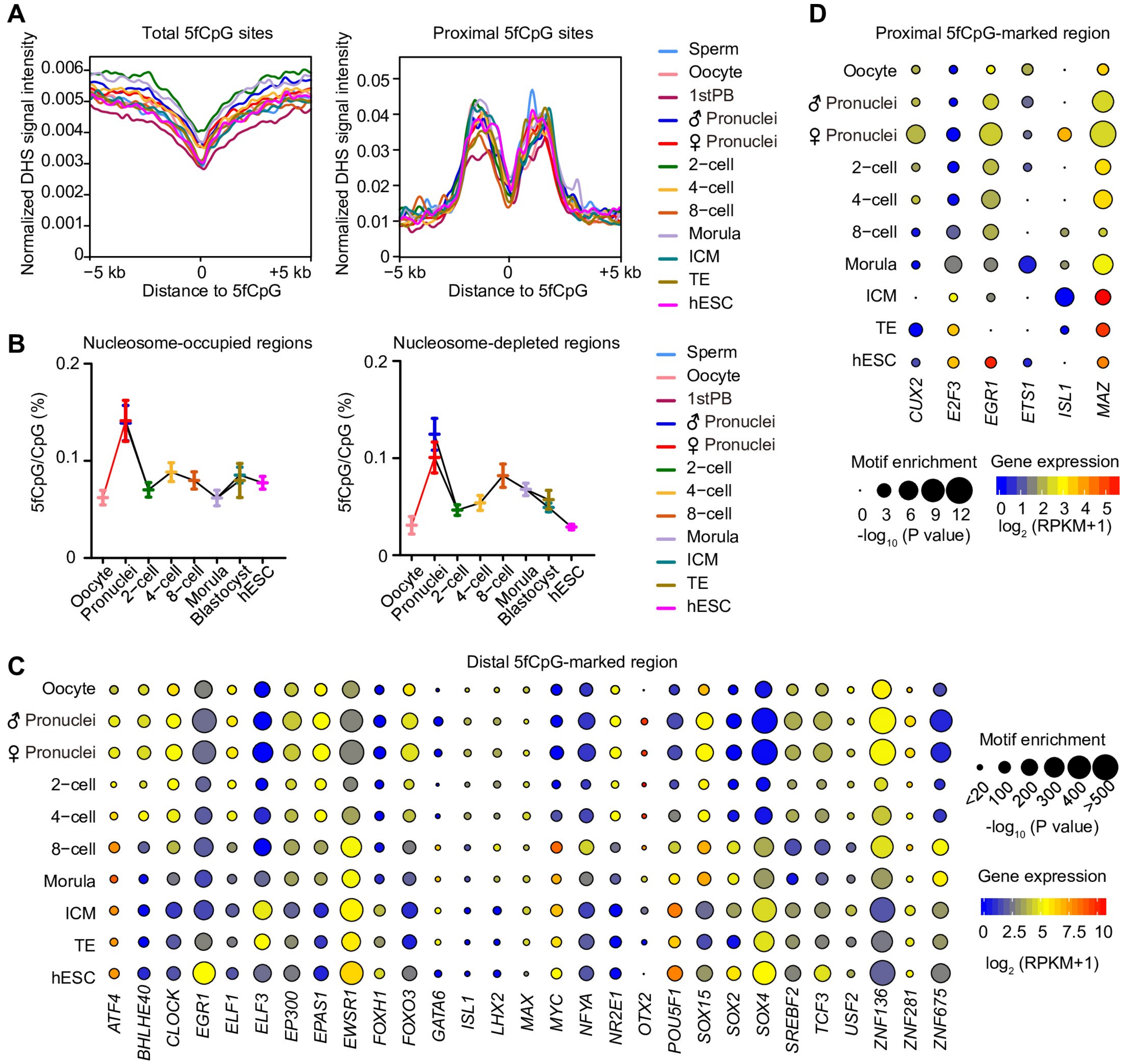


Fig 6. The regulatory network of 5fC-mark region. (A) The normalized DHS signal of hESC at the center of all 5fCpG sites and its flanking regions (left panel) as well as at the center of proximal 5fCpG sites (TSS ± 2 kb) and its flanking regions (right panel) in each development stage are shown. The DHS signal of hESC is downloaded from GSE32970. (B) The line chart showing the 5fCpG level in NORs (left panel) and in NDRs (right panel). The 5fCpG level was calculated by the number of 5fCpG sites divided by the sum number of 5fCpG and unmodified CpG sites. scCOOL-seq of human preimplantation embryos is from GSE100272. The center is the mean of 5fCpG level, and error bars are SEM. (C) Motif analysis of 5fCpG-marked distal regions (2 kb away from TSS). Only motif with $P \leq 10^{-20}$ and RPKM ≥ 1 in at least one stage are shown in the diagram. The significance was calculated by the binomial test in HOMER by default of motif enrichment. The color of the circle from blue to red indicates the expression level in each stage from low to high. The scRNA-seq data of human early embryos and ESCs are from GSE36552. The size of the circle indicates the $-\log_{10}(P \text{ value})$. (D) Motif analysis of 5fCpG-marked proximal regions (TSS ± 2 kb). Only motif with $P \leq 10^{-5}$ and RPKM ≥ 1 in at least one stage are shown in the diagram. The significance was calculated by the binomial test in HOMER by default of motif enrichment. The color of the circle from blue to red indicates the expression level in each stage from low to high. The scRNA-seq data of human early embryos, and ESCs are from GSE36552. The size of the circle indicates the $-\log_{10} P$ value. (A–D) The sample size in these panels are listed in Fig 1A. (B–D) The numerical data are listed in S1 Data. CpG, cytosine phosphate guanine; DHS, DNase I hypersensitive sites; ESC, embryonic stem cell; hESC, human embryonic stem cell; HOMER, hypergeometric optimization of motif enrichment; NDR, nucleosome-depleted region; NOR, nucleosome-occupied region; scCOOL-seq, chromatin overall omic-scale landscape

sequencing; RPKM, reads per kilobase of transcript per million mapped reads; scRNA-seq, single-cell RNA sequencing; SEM, standard error of the mean; TSS, transcription start site; 5fC, 5-formylcytosine; 5fCpG, 5-formylcytosine phosphate guanine.

https://doi.org/10.1371/journal.pbio.3000799.g006

with normal morphology was collected via micromanipulation. For the metaphase II oocyte, the granulosa cells were removed with the treatment of hyaluronidase (Sigma). The first polar bodies were collected by a micropipette with laser-assisted biopsy. The metaphase II oocytes without the first polar body were transferred into an acidic solution drop (1 μL 36% HCl diluted in 1 mL DPBS) to remove the zona pellucida. The oocytes were then washed several times in DPBS with 0.1% HSA by gentle pipetting to eliminate any somatic contamination.

Human preimplantation embryos. The assessment of embryos development and the scheduling of embryo collection were performed according to a previous study [55]. Following oocyte retrieval, oocytes were fertilized via ICSI and signs of fertilization were checked on the following day. Male and female pronuclei of zygotes were collected through careful micromanipulation after broken the zona pellucida and the cytoplasmic membrane. The 2-cell and 4-cell embryos were collected at 27 h and 48 h post fertilization, respectively. The 8-cell and morula stage embryos were obtained at Day 3 and Day 4 according to their morphology, respectively. At Days 5 and 6, the ICM and TE of blastocysts were isolated with laser-assisted micromanipulation. Embryos with developmental retardation were excluded from this study. The embryos were treated with acidic solution (1 μL 36% HCl diluted in 1 mL DPBS) to remove the zona pellucida. The embryos without zona pellucida were then washed several times by gentle pipetting in DPBS with 0.1% HSA. Subsequently, a mixture of Accutase (Sigma) and 0.25% trypsin at the ratio of 1:1 was used to digest the embryo into single cells in humidified incubator for 20 to 50 min with 5% CO_2 at 37˚C. Then single blastomeres were carefully washed in DPBS with 0.1% HSA before being transferred into lysis buffer.

hESC culture

The hESCs (H9) were purchased from WiCell Institute. hESCs were cultured on the hESC--Qualified Matrix (Corning) in a feeder-free condition. And the complete E8 culture medium were used in cell culture. With regular passaging, we collected the colonies of hESCs and digested them into single-cell suspension using Accutase (Sigma) at 37˚C for 1 h. Then the single cells of hESCs were used for CLEVER-seq library construction.

CLEVER-seq library construction

The CLEVER-seq experiment was carried out according to the previous study with slight modification [22]. The single cells of human early embryo were transferred into lysis buffer (20 mM Tris, 2 mM EDTA, 20 mM KCl, and 0.3% TritonX-100, 1 mg/mL QIAGEN protease) by mouth pipette. After centrifugation (1 min at 7,000 r.p.m., 4˚C), the sample was incubated at 50˚C for 3 h to release the genomic DNA followed by 70˚C for 30 min to inactive the protease. Then trace mount of spike-in DNA contain 5fC modification was added into sample. In order to label the 5fC, 150 mM malononitrile (J&K) was used. Mineral oil (15 μL) was added onto the surface of the sample to prevent evaporation. The sample was incubated at 37˚C for 20 h with constant shaking at 850 r.p.m. in Thermomixer (Eppendorf). Then the single cell genome was amplified by MALBAC (Yikon Genomics) according to the user guide. After amplification, the genomic DNA was fragment into approximately 300 by covaris S2. The library construction was performed using KAPA Hyper Prep Kit (Kapa Biosystems). The libraries were sequenced on Illumina Hiseq 4000 platform in pair-end 150 bp model (Novogene).

Genomic DNA sequencing

In order to differentiate the male and female pronucleus, the genomic DNA was sequenced from the peripheral blood and sperm pellet of couple donators to call SNP. The genomic DNA was extracted using DNesay Blood and Tissue Kit (Qiagen). About 500 ng DNA was fragment into 300 bp by by covaris S2. The libraries were constructed using KAPA Hyper Prep Kit (Kapa Biosystems). And the libraries were sequenced the same way as CLEVER-seq libraries.

Immunostaining of 5fC and 5mC in human zygote

The immunostaining was carried out according to previous study with slight modification [5]. Briefly, the zona pellucida of human zygotes was removed by treating with acidic solution (1 μL 36% HCl diluted in 1 mL DPBS). After several times of washing in DPBS, the sample was fixed 4% paraformaldehyde for 30 min at room temperature. After washing with DPBS containing 0.1% HSA, the membrane was permeated in 1% Triton X-100 at room temperature for 15 min. For detection of 5fC and 5mC, the DNA of embryos was denatured by 4N HCl at room temperature for 20 min and then neutralized for 10 min with 100 mM Tris-HCl. The sample was blocked in 1% goat serum at room temperature for 1.5 h. The primary antibody of 5fC (Active Motif, Cat#61223) and 5mC (Eurogentec, Cat#BI-MECY-1000) was diluted at the ratio of 1:500. The sample was incubated in the diluted antibody buffer for 3 h at room temperature. After washing, the sample was incubated with Alexa Fluor 594 goat anti-rabbit IgG (1:500, Invitrogen, Cat#A-11012) and Alexa Fluor 488 goat anti-mouse IgG (1:500, Invitrogen, Cat#A-11001) for 1 h at room temperature. All the images were captured and analysed using Nikon A1R high-speed laser confocal microscopy.

Reads quality control and alignment

We used Trim Galore (version 0.3.3) to remove low quality bases, adaptor, and MALBAC primer sequences. Then, reads passed quality control were mapped to human reference genome (version: hg19) using Bismark (version 0.13.0) [56]. Then PCR duplicates were removed with the command 'samtools rmdup' (version 0.1.18) [57]. A 137-bp model DNA contain 5fC modification (model 1 or model 2; information are provided in S2 Table), and lambda were added in same library to estimate the C-to-T conversion rate and random C-to-T conversion rate in each single cell experiment.

CNV detection with CLEVER-seq data

First, we selected euploid cells for downstream analysis. Briefly, read counts of 1 Mb nonoverlapping windows were used to deduce CNV by R package HMMcopy [58] with GC and mappability correction.

Gender estimation of pronuclei

With read counts of sex chromosomes, we can estimate gender of each embryo. In this way, we can unambiguously distinguish between male PN and female PN in paired pronuclei having chromosomes XY, also male PN from unpaired pronuclei having chromosomes XY. As for pronuclei having chromosomes XX, we used SNPs identified from parental genomic DNA to distinguish between male PN and female PN. Using an established pipline [12,59], parental genomic data were cleaned with Trim Galore (version 0.3.3) and mapped to the human reference genome hg19 with bwa (version 0.7.12) [60]. PCR duplications were removed with Picard (version 1.126); then SNPs were identified with GATK HaplotypeCaller [61].

5fCpG site identification and 5fCpG level estimation

Following a previous study [22], we restricted 5fC at CpG sites without known SNPs (dbSNP database hg19 version 135). All CpG sites with C-to-T conversion rate ≥0.65 in at least two treated cells were grouped as the candidate pool to avoid PCR amplification errors, whereas CpG sites with C-to-T conversion rate ≥0.65 in control cells were grouped as the candidate noise pool to avoid background noises. For each treated single cell, CpG sites covered ≥3 times with C-to-T conversion rate ≥0.65, in the candidate pool but not in the noise pool, were identified as candidate 5fCpG sites in this single cell. To calculate the possibility of observing 5fCpG by chance, we used binomial distribution estimation with 'dbinom' in R, with C-to-T conversion rate estimated from lambda DNA, according to previous study [62]. For each site, the number of reads supporting 'T' were denoted as 'NT' and the number of reads supporting 'C' were denoted as 'NC', and the sequencing depth was NT+NC. 5fCpG sites with Holm–Bonferroni method-adjusted $p < 0.01$ were retained as final 5fCpG sites in this single cell for further analysis. And CpG sites covered ≥3 times with C-to-T conversion rate ≤0.25 were identified as unmodified C sites. In order to model false positives, the negative control single cells (those not treated with malononitrile) were also analyzed with the same procedures. The number of called 5fCpG sites were divided by the number of total sequenced CpG sites (defined as '5fCpG abundance') for both the treated and untreated samples. And false-positive detection rate was calculated by dividing the 5fCpG abundance of the untreated samples by that of the treated samples. At the C-to-T cut-off of 0.65, false-positive detection rate of 8% was observed in hESC samples.

For random C-to-T rate of lambda DNA, we selected the C sites covered ≥4 times in an individual cell sample to avoid sequencing errors and counted the number of T readout on these C sites. The random C-to-T rate was calculated by the number of T readouts divided by the total number of these C sites sequenced. Based on labmda DNA, the average false positive rate in 130 treatment sample was 1.15%.

The cluster and the variance of 5fCpG level were calculated based on 1-kb windows across the genome, and only windows with 5fCpG sites identified in at least one stage were retained for this analysis. The motif enrichment analysis was done with HOMER (version 4.10.3) based on regions 100 bp upstream and 100 bp downstream of 5fCpG sites identified in each stage.

The global or genomic region 5fCpG ratio was estimated by the number of 5fCpG sites divided by the sum of the number of 5fCpG and unmodified C sites. The genomic locations of exons, introns, CGIs, 5′UTR, and 3′UTR repeat elements and their subfamilies were downloaded from the UCSC genome browser. Promoters were regions 1-kb upstream and 0.5-kb downstream of TSS, which were grouped into three classes based on CpG dengsity, that is, HCP, ICP, and LCP [63]. The significance of 5fCpG level between genomic regions was calculated by two-tailed Student *t* test.

The published data sets used in this study were downloaded from the Gene Expression Omnibus (GEO) with the following accession numbers: GSE100272 (scCOOL-seq of human preimplantation embryos), GSE36552 (scRNA-seq data of human preimplantation embryos), GSE32970 (DNase-seq peak and signal of human ESCs), GSE29611 (ChIP-seq of human ESCs), GSE61475 (transcription factor binding sites in human ESCs).

DNA methylation levels and RNA expression levels of 5fCpG marked repeats

We used read counts located in each 5fC marked repeat region and RPKM (reads per kilobase of transcript per million mapped reads) method to estimate the expression levels of repeat elements. As a control, we randomly selected the same number of L1 or ERVK regions with only unmodified CpG covered to calculate the corresponding DNA methylation levels or RNA expression levels. The repeat annotation is downloaded from hg19 Repeat Masker in UCSC. The RNA-seq data are from GSE36552, and the DNA methylation data are from GSE81233.

t-SNE and PCA analysis of 5fCpG sites

5fCpG percentage in 1-kb, 10-kb, 100-kb, 1-Mb, or 10-Mb windows were calculated by the number of 5fCpG divided by the sum of unmodified CpG and 5fCpG. Only the windows containing 5fCpG modification at least in one single cell were considered when performing dimensionality reduction analysis. PCA was performed by ‘pcaMethods’ package in R, and tSNE analysis was performed by ‘tsne’ package in R.

Code availability

All the data analysis was conducted in perl, Python, and R language. All the computational code used in this study are available upon contacting with corresponding authors.

Supporting information

S1 Fig. Quality control of CLEVER-seq library and the summary of the number of 5fCpG sites. (A) Representative copy number variation plot showing that only euploid cell were retained for further analysis in this study. (B) Box plot showing the C-to-T conversion rate of spike-in oligo DNA which contained a 5fC modification in each stages. The box indicates the median, 25% quartile, and 75% quartile. The whisker represents the 1.5 times of IQR. (C) Box plot showing the mapping rate of 130 libraries in each stages. The box indicates the median, 25% quartile, and 75% quartile. The whisker represents the 1.5 times of IQR. (D) Box plot showing the CpG coverage at 1× (blue), 3× (red), 5× (green) depth of libraries in different stages. The box indicates the median, 25% quartile, and 75% quartile. The whisker represents the 1.5 times of IQR. (E) Histogram showing the number of covered unmodified CpG sites in each stages at 3× depth. (B–E) The sample size in these panels are listed in Fig 1A; the numerical data are listed in S1 Table. (F) The table showing the number of 5fCpG sites identified in each single cell and the number of stage-merged 5fCpG sites. (G) The table summarizing the number of CpG sites covered by all cells in each stage at 3× depth, the number as well as the ratio of 5fCpG sites in all covered CpG sites. CLEVER-seq, chemical-labeling-enabled C-to-T conversion sequencing; CpG, chemical-labeling-enabled C-to-T conversion sequencing; IQR, interquartile range; 5fC, 5-formylcytosine; 5fCpG, 5-formylcytosine phosphate guanine.
(TIF)

S2 Fig. The 5fCpG enrichment on genomic region and comparison between ICM and hESC. (A) The normalized fold changes of average gene expression levels between two consecutive stages for genes with 5fCpG in promoters (TSS ± 2 kb). The hESC was compared with ICM, whereas both ICM and TE were compared with morula.The color from blue to red indicate the values of $\log_2$ of expression level fold change between two consecutive stages from low to high. The scRNA-seq data of human early embryos are from GSE36552. (B) Venn diagram showing the 5fCpG site overlap between ICM and hESC. The number of ICM-specific, hESC-specific, and their shared 5fCpG sites are listed in the diagram. (C) Bar plot showing the relative enrichment of 5fCpG sites on different genomic region. (D) Line chart showing the 5fCpG level in intergenic and intragenic region across human early embryo developmental stages. The center is the mean of 5fCpG level and error bars are SEM. (A–D) The sample size in these panels are listed in Fig 1A. (A, C, and D) The numerical data is listed in S3 Data.

hESC, human embryonic stem cell; ICM, inner cell mass; scRNA-seq, single-cell RNA sequencing; SEM, standard error of the mean; TE, trophectoderm; 5fCpG, 5-formylcytosine phosphate guanine.
(TIF)

S3 Fig. The 5fCpG dynamic in genomic regions during human early embryonic development. (A) The line chart showing the 5fCpG level in distinct genomic regions of human early embryo and hESC. The center is the mean of 5fCpG level, and error bars are SEM. (B) The line chart showing the 5fCpG level in three types of promoter, HCP (left panel), ICP (middle panel), and LCP (right panel) according to the CpG density. The center is the mean of 5fCpG level, and error bars are SEM. (A–B) The numerical data are listed in S3 Data. (C) Box plot showing the variance of 5fCpG abundance during human early embryonic development. The variance was calculated in 1-kb window. The box indicates the median, 25% quartile, and 75% quartile. The whisker represents the 1.5 times of IQR. The numerical data are listed in S4 Data. (A–C) The sample size in these panels are listed in Fig 1A. (D) t-SNE analysis of 5fCpG-marked windows at different window sizes. CpG, cytosine phosphate guanine; HCP, high-density CpG promoter; hESC, human embryonic stem cell; ICP, intermediated-density CpG promoter; IQR, interquartile range; LCP, low-density CpG promoter; SEM, standard error of the mean; t-SNE, t-distributed stochastic neighbor embedding; 5fCpG, 5-formylcytosine phosphate guanine.
(TIF)

S4 Fig. PCA of 5fCpG-marked genomic windows at the windows size of 1 kb, 10 kb, 100 kb, 1 Mb, 10 Mb. PCA, principal component analysis; 5fCpG, 5-formylcytosine phosphate guanine.
(TIF)

S5 Fig. 5fCpG distribution and enrichment on repeat elements. (A) Enrichment analysis of 5fCpG site on different repeat elements. (B) The line chart showing the 5fCpG level on distinct repeat elements. The center is the mean of 5fCpG level and error bars are SEM. (C) The line char showing the expression level of L1 (left panel) and ERVK (right panel) with 5fCpG modification or randomly seletcted ones only with unmodified CpG. The significance based on Student t-test is denoted. The scRNA-seq data of human early embryos are from GSE36552. (D) The line char showing the DNA methylation levels of L1 (left panel) and ERVK (right panel) with 5fCpG modification or randomly seletcted ones only with unmodified CpG. The significance based on Student *t* test is denoted.The DNA methylaiton data of human early embryos are from GSE81233. (E) Enrichment analysis of gamete-specific, gamete-shared, pronuclei-specific, and pronuclei-shared 5fCpG-marked regions (1-kb window) in repeat elements. (A–E) The sample size in these panels are listed in Fig 1A, and the numerical data are listed in S3 Data. CpG, cytosine phosphate guanine; ERVK, endogenous retrovirus-K; SEM, standard error of the mean; scRNA-seq, single-cell RNA sequencing; 5fCpG, 5-formylcytosine phosphate guanine.
(TIF)

S6 Fig. The 5fCpG level in blastomeres of paired two-cell. (A) The 5fCpG level in blastomeres of paired two-cell in whole genome ($n = 4$). The dots in red and blue represent the mean of 5fCpG level in two blastomeres, respectively. (B) The line chart showing the 5fCpG level of paired blastomeres in two-cell stage in different genomic regions ($n = 4$). The dots represent the mean of 5fCpG level. (C) The line chart showing the 5fCpG level of paired blastomeres in two-cell stage in different repeat elements and the subfamilies of them ($n = 4$). The dots represent the mean of 5fCpG level. (A–C) The numerical data is listed in S3 Data. 5fCpG,

5-formylcytosine phosphate guanine.
(TIF)

S1 Table. Basic summary of human preimplantation embryo sample information.
(XLSX)

S2 Table. The sequence of model DNA.
(XLSX)

S1 Data. The individual numerical values in Fig 2A and 2B, Fig 2D and 2E, Fig 3A and 3B, Fig 4E and 4F, Fig 5B–5D, Fig 6B–6D.
(XLS)

S2 Data. The individual numerical values in Fig 2F.
(ZIP)

S3 Data. The individual numerical values in S2A Fig, S2C and S2D Fig, S3A and S3B Fig, S5A–S5E Fig, S6A–S6C Fig.
(XLS)

S4 Data. The individual numerical values in S3C Fig.
(ZIP)

Acknowledgments

We feel very grateful to Prof. C. Yi for generously providing model DNA containing 5fC modification as spike-in. We thank Dr. C. Zhu and Mr. B. He for consulting the bioinformatics analysis, Dr. H. Zeng and Mr. D. Bai for their experimental suggestion. Some parts of the bioinformatics analyses were performed on the Computing Platform at the Center for Life Science.

Author Contributions

Formal analysis: Yun Gao, Lin Li.

Investigation: Yun Gao, Peng Yuan, Fan Zhai, Yixin Ren, Liying Yan, Rong Li, Ying Lian, Xiaohui Zhu, Xinglong Wu, Lu Wen.

Project administration: Jie Qiao, Fuchou Tang.

Resources: Kehkooi Kee.

Supervision: Jie Qiao, Fuchou Tang.

Writing – original draft: Yun Gao, Lin Li, Peng Yuan, Fuchou Tang.

Writing – review & editing: Yun Gao, Lin Li, Peng Yuan, Fuchou Tang.

References

1. Hackett JA, Surani MA. Beyond DNA: programming and inheritance of parental methylomes. Cell. 2013; 153(4):737–9. https://doi.org/10.1016/j.cell.2013.04.044 PMID: 23663772
2. Saitou M, Kagiwada S, Kurimoto K. Epigenetic reprogramming in mouse pre-implantation development and primordial germ cells. Development. 2012; 139(1):15–31. https://doi.org/10.1242/dev.050849 PMID: 22147951
3. Burton A, Torres-Padilla ME. Chromatin dynamics in the regulation of cell fate allocation during early embryogenesis. Nature reviews Molecular cell biology. 2014; 15(11):723–34. https://doi.org/10.1038/nrm3885 PMID: 25303116

4. Li E. Chromatin modification and epigenetic reprogramming in mammalian development. Nature reviews Genetics. 2002; 3(9):662–73. https://doi.org/10.1038/nrg887 PMID: 12209141

5. Guo H, Zhu P, Yan L, Li R, Hu B, Lian Y, et al. The DNA methylation landscape of human early embryos. Nature. 2014; 511(7511):606–10. https://doi.org/10.1038/nature13544 PMID: 25079557

6. Smith ZD, Chan MM, Humm KC, Karnik R, Mekhoubad S, Regev A, et al. DNA methylation dynamics of the human preimplantation embryo. Nature. 2014; 511(7511):611–5. https://doi.org/10.1038/nature13581 PMID: 25079558

7. Li L, Guo F, Gao Y, Ren Y, Yuan P, Yan L, et al. Single-cell multi-omics sequencing of human early embryos. Nature cell biology. 2018.

8. Wu J, Xu J, Liu B, Yao G, Wang P, Lin Z, et al. Chromatin analysis in human early development reveals epigenetic transition during ZGA. Nature. 2018; 557(7704):256–60. https://doi.org/10.1038/s41586-018-0080-8 PMID: 29720659

9. Gao L, Wu K, Liu Z, Yao X, Yuan S, Tao W, et al. Chromatin Accessibility Landscape in Human Early Embryos and Its Association with Evolution. Cell. 2018; 173(1):248–59 e15. https://doi.org/10.1016/j.cell.2018.02.028 PMID: 29526463

10. Okae H, Chiba H, Hiura H, Hamada H, Sato A, Utsunomiya T, et al. Genome-wide analysis of DNA methylation dynamics during early human development. PLoS Genet. 2014; 10(12):e1004868. https://doi.org/10.1371/journal.pgen.1004868 PMID: 25501653

11. Fulka H, Mrazek M, Tepla O, Fulka J Jr., DNA methylation pattern in human zygotes and developing embryos. Reproduction. 2004; 128(6):703–8. https://doi.org/10.1530/rep.1.00217 PMID: 15579587

12. Zhu P, Guo H, Ren Y, Hou Y, Dong J, Li R, et al. Single-cell DNA methylome sequencing of human preimplantation embryos. Nature genetics. 2018; 50(1):12–9. https://doi.org/10.1038/s41588-017-0007-6 PMID: 29255258

13. Tahiliani M, Koh KP, Shen Y, Pastor WA, Bandukwala H, Brudno Y, et al. Conversion of 5-methylcytosine to 5-hydroxymethylcytosine in mammalian DNA by MLL partner TET1. Science. 2009; 324 (5929):930–5. https://doi.org/10.1126/science.1170116 PMID: 19372391

14. Ito S, D'Alessio AC, Taranova OV, Hong K, Sowers LC, Zhang Y. Role of Tet proteins in 5mC to 5hmC conversion, ES-cell self-renewal and inner cell mass specification. Nature. 2010; 466(7310):1129–33. https://doi.org/10.1038/nature09303 PMID: 20639862

15. Globisch D, Munzel M, Muller M, Michalakis S, Wagner M, Koch S, et al. Tissue distribution of 5-hydroxymethylcytosine and search for active demethylation intermediates. PLoS ONE. 2010; 5(12):e15367. https://doi.org/10.1371/journal.pone.0015367 PMID: 21203455

16. Ito S, Shen L, Dai Q, Wu SC, Collins LB, Swenberg JA, et al. Tet proteins can convert 5-methylcytosine to 5-formylcytosine and 5-carboxylcytosine. Science. 2011; 333(6047):1300–3. https://doi.org/10.1126/science.1210597 PMID: 21778364

17. Williams K, Christensen J, Helin K. DNA methylation: TET proteins-guardians of CpG islands? EMBO reports. 2011; 13(1):28–35. https://doi.org/10.1038/embor.2011.233 PMID: 22157888

18. Raiber EA, Portella G, Martinez Cuesta S, Hardisty R, Murat P, Li Z, et al. 5-Formylcytosine organizes nucleosomes and forms Schiff base interactions with histones in mouse embryonic stem cells. Nature chemistry. 2018; 10(12):1258–66. https://doi.org/10.1038/s41557-018-0149-x PMID: 30349137

19. Iurlaro M, Ficz G, Oxley D, Raiber EA, Bachman M, Booth MJ, et al. A screen for hydroxymethylcytosine and formylcytosine binding proteins suggests functions in transcription and chromatin regulation. Genome biology. 2013; 14(10):R119. https://doi.org/10.1186/gb-2013-14-10-r119 PMID: 24156278

20. Song CX, He C. Potential functional roles of DNA demethylation intermediates. Trends in biochemical sciences. 2013; 38(10):480–4. https://doi.org/10.1016/j.tibs.2013.07.003 PMID: 23932479

21. Iurlaro M, McInroy GR, Burgess HE, Dean W, Raiber EA, Bachman M, et al. In vivo genome-wide profiling reveals a tissue-specific role for 5-formylcytosine. Genome biology. 2016; 17(1):141. https://doi.org/10.1186/s13059-016-1001-5 PMID: 27356509

22. Zhu C, Gao Y, Guo H, Xia B, Song J, Wu X, et al. Single-Cell 5-Formylcytosine Landscapes of Mammalian Early Embryos and ESCs at Single-Base Resolution. Cell stem cell. 2017; 20(5):720–31 e5. https://doi.org/10.1016/j.stem.2017.02.013 PMID: 28343982

23. Pfaffeneder T, Hackner B, Truss M, Munzel M, Muller M, Deiml CA, et al. The discovery of 5-formylcytosine in embryonic stem cell DNA. Angewandte Chemie. 2011; 50(31):7008–12. https://doi.org/10.1002/anie.201103899 PMID: 21721093

24. Xia B, Han D, Lu X, Sun Z, Zhou A, Yin Q, et al. Bisulfite-free, base-resolution analysis of 5-formylcytosine at the genome scale. Nature methods. 2015; 12(11):1047–50. https://doi.org/10.1038/nmeth.3569 PMID: 26344045

25. Booth MJ, Marsico G, Bachman M, Beraldi D, Balasubramanian S. Quantitative sequencing of 5-formylcytosine in DNA at single-base resolution. Nature chemistry. 2014; 6(5):435–40. https://doi.org/10.1038/nchem.1893 PMID: 24755596

26. Wagner M, Steinbacher J, Kraus TF, Michalakis S, Hackner B, Pfaffeneder T, et al. Age-dependent levels of 5-methyl-, 5-hydroxymethyl-, and 5-formylcytosine in human and mouse brain tissues. Angewandte Chemie. 2015; 54(42):12511–4. https://doi.org/10.1002/anie.201502722 PMID: 26137924

27. Zong C, Lu S, Chapman AR, Xie XS. Genome-wide detection of single-nucleotide and copy-number variations of a single human cell. Science. 2012; 338(6114):1622–6. https://doi.org/10.1126/science.1229164 PMID: 23258894

28. Su M, Kirchner A, Stazzoni S, Muller M, Wagner M, Schroder A, et al. 5-Formylcytosine Could Be a Semipermanent Base in Specific Genome Sites. Angewandte Chemie. 2016; 55(39):11797–800. https://doi.org/10.1002/anie.201605994 PMID: 27561097

29. Smallwood SA, Lee HJ, Angermueller C, Krueger F, Saadeh H, Peat J, et al. Single-cell genome-wide bisulfite sequencing for assessing epigenetic heterogeneity. Nature methods. 2014; 11(8):817–20. https://doi.org/10.1038/nmeth.3035 PMID: 25042786

30. Consortium EP. An integrated encyclopedia of DNA elements in the human genome. Nature. 2012; 489 (7414):57–74. https://doi.org/10.1038/nature11247 PMID: 22955616

31. Tsankov AM, Gu H, Akopian V, Ziller MJ, Donaghey J, Amit I, et al. Transcription factor binding dynamics during human ES cell differentiation. Nature. 2015; 518(7539):344–9. https://doi.org/10.1038/nature14233 PMID: 25693565

32. Percharde M, Lin CJ, Yin Y, Guan J, Peixoto GA, Bulut-Karslioglu A, et al. A LINE1-Nucleolin Partnership Regulates Early Development and ESC Identity. Cell. 2018; 174(2):391–405 e19. https://doi.org/10.1016/j.cell.2018.05.043 PMID: 29937225

33. Fang F, Hodges E, Molaro A, Dean M, Hannon GJ, Smith AD. Genomic landscape of human allele-specific DNA methylation. Proceedings of the National Academy of Sciences of the United States of America. 2012; 109(19):7332–7. https://doi.org/10.1073/pnas.1201310109 PMID: 22523239

34. Thurman RE, Rynes E, Humbert R, Vierstra J, Maurano MT, Haugen E, et al. The accessible chromatin landscape of the human genome. Nature. 2012; 489(7414):75–82. https://doi.org/10.1038/nature11232 PMID: 22955617

35. Consortium EP. A user's guide to the encyclopedia of DNA elements (ENCODE). PLoS Biol. 2011; 9 (4):e1001046. https://doi.org/10.1371/journal.pbio.1001046 PMID: 21526222

36. Li F, Zhang Y, Bai J, Greenberg MM, Xi Z, Zhou C. 5-Formylcytosine Yields DNA-Protein Cross-Links in Nucleosome Core Particles. Journal of the American Chemical Society. 2017; 139(31):10617–20. https://doi.org/10.1021/jacs.7b05495 PMID: 28742335

37. Raiber EA, Beraldi D, Ficz G, Burgess HE, Branco MR, Murat P, et al. Genome-wide distribution of 5-formylcytosine in embryonic stem cells is associated with transcription and depends on thymine DNA glycosylase. Genome biology. 2012; 13(8):R69. https://doi.org/10.1186/gb-2012-13-8-r69 PMID: 22902005

38. Yao TP, Oh SP, Fuchs M, Zhou ND, Ch'ng LE, Newsome D, et al. Gene dosage-dependent embryonic development and proliferation defects in mice lacking the transcriptional integrator p300. Cell. 1998; 93 (3):361–72. https://doi.org/10.1016/s0092-8674(00)81165-4 PMID: 9590171

39. Visel A, Blow MJ, Li ZR, Zhang T, Akiyama JA, Holt A, et al. ChIP-seq accurately predicts tissue-specific activity of enhancers. Nature. 2009; 457(7231):854–8. https://doi.org/10.1038/nature07730 PMID: 19212405

40. Embree LJ, Azuma M, Hickstein DD. Ewing sarcoma fusion protein EWSR1/FLI1 interacts with EWSR1 leading to mitotic defects in zebrafish embryos and human cell lines. Cancer research. 2009; 69 (10):4363–71. https://doi.org/10.1158/0008-5472.CAN-08-3229 PMID: 19417137

41. Yoon Y, Park H, Kim S, Nguyen PT, Hyeon SJ, Chung S, et al. Genetic Ablation of EWS RNA Binding Protein 1 (EWSR1) Leads to Neuroanatomical Changes and Motor Dysfunction in Mice. Exp Neurobiol. 2018; 27(2):103–11. https://doi.org/10.5607/en.2018.27.2.103 PMID: 29731676

42. Yan L, Yang M, Guo H, Yang L, Wu J, Li R, et al. Single-cell RNA-Seq profiling of human preimplantation embryos and embryonic stem cells. Nature structural & molecular biology. 2013; 20(9):1131–9.

43. Amouroux R, Nashun B, Shirane K, Nakagawa S, Hill PW, D'Souza Z, et al. De novo DNA methylation drives 5hmC accumulation in mouse zygotes. Nature cell biology. 2016; 18(2):225–33. https://doi.org/10.1038/ncb3296 PMID: 26751286

44. Kellinger MW, Song CX, Chong J, Lu XY, He C, Wang D. 5-formylcytosine and 5-carboxylcytosine reduce the rate and substrate specificity of RNA polymerase II transcription. Nature structural & molecular biology. 2012; 19(8):831–3.

45. Kitsera N, Allgayer J, Parsa E, Geier N, Rossa M, Carell T, et al. Functional impacts of 5-hydroxymethylcytosine, 5-formylcytosine, and 5-carboxycytosine at a single hemi-modified CpG dinucleotide in a gene promoter. Nucleic acids research. 2017; 45(19):11033–42. https://doi.org/10.1093/nar/gkx718 PMID: 28977475

46. Ji S, Shao H, Han Q, Seiler CL, Tretyakova NY. Reversible DNA-Protein Cross-Linking at Epigenetic DNA Marks. Angewandte Chemie. 2017; 56(45):14130–4. https://doi.org/10.1002/anie.201708286 PMID: 28898504

47. Raiber EA, Murat P, Chirgadze DY, Beraldi D, Luisi BF, Balasubramanian S. 5-Formylcytosine alters the structure of the DNA double helix. Nature structural & molecular biology. 2015; 22(1):44–9.

48. Bachman M, Uribe-Lewis S, Yang X, Burgess HE, Iurlaro M, Reik W, et al. 5-Formylcytosine can be a stable DNA modification in mammals. Nature chemical biology. 2015; 11(8):555–7. https://doi.org/10.1038/nchembio.1848 PMID: 26098680

49. Mooijman D, Dey SS, Boisset JC, Crosetto N, van Oudenaarden A. Single-cell 5hmC sequencing reveals chromosome-wide cell-to-cell variability and enables lineage reconstruction. Nature biotechnology. 2016; 34(8):852–6. https://doi.org/10.1038/nbt.3598 PMID: 27347753

50. Zeng H, He B, Xia B, Bai D, Lu X, Cai J, et al. Bisulfite-Free, Nanoscale Analysis of 5-Hydroxymethylcytosine at Single Base Resolution. Journal of the American Chemical Society. 2018; 140(41):13190–4. https://doi.org/10.1021/jacs.8b08297 PMID: 30278133

51. Schutsky EK, DeNizio JE, Hu P, Liu MY, Nabel CS, Fabyanic EB, et al. Nondestructive, base-resolution sequencing of 5-hydroxymethylcytosine using a DNA deaminase. Nature biotechnology. 2018.

52. Liu Y, Siejka-Zielinska P, Velikova G, Bi Y, Yuan F, Tomkova M, et al. Bisulfite-free direct detection of 5-methylcytosine and 5-hydroxymethylcytosine at base resolution. Nature biotechnology. 2019.

53. Chian RC, Lim JH, Tan SL. State of the art in in-vitro oocyte maturation. Current opinion in obstetrics & gynecology. 2004; 16(3):211–9.

54. Sathananthan AH, Osianlis T. Human embryo culture and assessment for the derivation of embryonic stem cells (ESC). Methods in molecular biology. 2010; 584:1–20. https://doi.org/10.1007/978-1-60761-369-5_1 PMID: 19907969

55. Niakan KK, Han J, Pedersen RA, Simon C, Pera RA. Human pre-implantation embryo development. Development. 2012; 139(5):829–41. https://doi.org/10.1242/dev.060426 PMID: 22318624

56. Krueger F, Andrews SR. Bismark: a flexible aligner and methylation caller for Bisulfite-Seq applications. Bioinformatics. 2011; 27(11):1571–2. https://doi.org/10.1093/bioinformatics/btr167 PMID: 21493656

57. Li H, Handsaker B, Wysoker A, Fennell T, Ruan J, Homer N, et al. The Sequence Alignment/Map format and SAMtools. Bioinformatics. 2009; 25(16):2078–9. https://doi.org/10.1093/bioinformatics/btp352 PMID: 19505943

58. Ha G, Roth A, Lai D, Bashashati A, Ding J, Goya R, et al. Integrative analysis of genome-wide loss of heterozygosity and monoallelic expression at nucleotide resolution reveals disrupted pathways in triple-negative breast cancer. Genome research. 2012; 22(10):1995–2007. https://doi.org/10.1101/gr.137570.112 PMID: 22637570

59. Li L, Guo F, Gao Y, Ren Y, Yuan P, Yan L, et al. Single-cell multi-omics sequencing of human early embryos. Nature cell biology. 2018.

60. Li H, Durbin R. Fast and accurate short read alignment with Burrows-Wheeler transform. Bioinformatics. 2009; 25(14):1754–60. https://doi.org/10.1093/bioinformatics/btp324 PMID: 19451168

61. Van der Auwera GA, Carneiro MO, Hartl C, Poplin R, Del Angel G, Levy-Moonshine A, et al. From FastQ data to high confidence variant calls: the Genome Analysis Toolkit best practices pipeline. Current protocols in bioinformatics. 2013; 43:11 0 1–33.

62. Yu M, Hon GC, Szulwach KE, Song CX, Zhang L, Kim A, et al. Base-resolution analysis of 5-hydroxymethylcytosine in the mammalian genome. Cell. 2012; 149(6):1368–80. https://doi.org/10.1016/j.cell.2012.04.027 PMID: 22608086

63. Xie W, Schultz MD, Lister R, Hou Z, Rajagopal N, Ray P, et al. Epigenomic analysis of multilineage differentiation of human embryonic stem cells. Cell. 2013; 153(5):1134–48. https://doi.org/10.1016/j.cell.2013.04.022 PMID: 23664764