Strand-specific single-cell methylomics reveals distinct modes of DNA demethylation dynamics during early mammalian development

Maya Sen[1†], Dylan Mooijman[1,2†], Jean-Charles Boisset[1], Alex Chialastri[3,4], Mina Popovic[5], Björn Heindryckx[5], Susana M. Chuva de Sousa Lopes[5,6], Siddharth S. Dey[3,4,7*] & Alexander van Oudenaarden[1*]

[1] Oncode Institute, Hubrecht Institute-KNAW (Royal Netherlands Academy of Arts and Sciences) and University Medical Center Utrecht, Utrecht, The Netherlands.

[2] Current address: Developmental Biology Unit, European Molecular Biology Laboratory, Heidelberg, Germany.

[3] Department of Chemical Engineering, University of California Santa Barbara, Santa Barbara, CA 93106, USA.

[4] Center for Bioengineering, University of California Santa Barbara, Santa Barbara, CA 93106, USA.

[5] Ghent-Fertility and Stem cell Team (G-FaST), Department of Reproductive Medicine, Ghent University Hospital, 9000 Ghent, Belgium.

[6] Department of Anatomy and Embryology, Leiden University Medical Center, 2333 ZC Leiden, The Netherlands.

[7] Neuroscience Research Institute, University of California Santa Barbara, Santa Barbara, CA 93106, USA.

† These authors contributed equally to this work.

*Correspondence should be addressed to A.v.O. (a.vanoudenaarden@hubrecht.eu) and S.S.D (sdey@ucsb.edu)

Abstract

DNA methylation (5mC) is central to cellular identity and the global erasure of 5mC from the parental genomes during preimplantation mammalian development is critical to reset the methylome of terminally differentiated gametes to the pluripotent cells in the blastocyst. While active and passive modes of demethylation have both been suggested to play a role in this process, the relative contribution of these two mechanisms to genome-wide 5mC erasure remains unclear. Here, we report a new high-throughput single-cell method (scMspJI-seq) that enables strand-specific quantification of 5mC, thereby allowing us to systematically probe the dynamics of global demethylation. First, when applied to hybrid mouse embryonic stem cells, we identified substantial cell-to-cell strand-specific 5mC heterogeneity, with a small group of cells displaying asymmetric

levels of 5mCpG between the two DNA strands of a chromosome suggesting loss of maintenance methylation. Next, using scMspJI-seq in preimplantation mouse embryos, we discovered that methylation maintenance is active till the 16-cell stage followed by passive demethylation in a fraction of cells within the early blastocyst at the 32-cell stage of development. Finally, we found that human preimplantation embryos qualitatively show temporally delayed yet similar demethylation dynamics as mouse preimplantation embryos. Collectively, these results demonstrate that scMspJI-seq is a sensitive and cost-effective method to map the strand-specific genome-wide patterns of 5mC in single cells, thereby enabling quantitative investigation of methylation dynamics in developmental systems.

In mammalian systems, DNA methylation (5-methylcytosine or 5mC) is a key epigenetic modification that is typically stably inherited from mother to daughter cells[1]. This property of 5mC plays an important role in facilitating the propagation of cellular identity through cell divisions and restricting the developmental potential of terminally differentiated cells[1,2]. Consequently, during preimplantation mammalian development, DNA methylation patterns on the terminally differentiated paternal sperm and maternal egg genomes are erased post-fertilization at a genome-wide scale to revert cellular memory towards an undifferentiated state in the blastocyst[3]. Therefore, understanding the mechanisms underlying global DNA demethylation dynamics is central to understanding the emergence of pluripotent cells during early development.

Removal of 5mC can proceed through two alternate mechanisms – passive and active demethylation. Methylated cytosines, within a CpG dinucleotide context are typically copied over to the newly synthesized DNA strands during genome replication by the maintenance methyltransferase, Dnmt1[4]. Passive demethylation relies on loss of 5mC through replicative dilution, in which inhibition of DNA methylation maintenance results in a reduction of 5mC levels after cell division and can be detected through asymmetric levels of 5mC on the two DNA strands of a chromosome. Alternatively, active mechanisms of 5mC erasure occur via conversion of 5mC to 5-hydroxymethylcytosine (5hmC) and other oxidized derivatives, that are not recognized by the DNA maintenance methylation machinery and are subsequently removed by base-excision repair pathways[5-7]. While early immunofluorescence-based studies revealed that the paternal genome undergoes active demethylation through conversion to 5hmC in the zygote, the maternal genome undergoes passive demethylation through the lack of Dnmt1 activity during replication[8-11]. Improving upon this coarse quantification of methylation dynamics, next-generation sequencing and mass spectroscopy based studies have recently revealed that the orthogonal regulation of

demethylation by active and passive mechanisms for the two parental genomes is not as distinct as suggested by these early studies[12-18]. However, the conclusions in these recent studies were partly based on bulk bisulfite-sequencing based methods that could not directly distinguish between active *vs.* passive demethylation, and therefore the relative contribution of these two mechanisms to 5mC reprogramming remains poorly understood.

To distinguish between these two mechanisms of demethylation requires strand-specific detection of 5mC in single cells. While asymmetric levels of 5mC between two DNA strands of a chromosome would indicate passive demethylation, the global loss of methylation coupled with symmetric levels of 5mC between two DNA strands would indirectly imply active demethylation (Fig. 1a). Bisulfite-based methods cannot distinguish between 5mC and 5hmC, and importantly when applied to bulk cellular populations, this approach cannot be used to quantify the relative levels of 5mC between two strands of a single chromosome[19]. Therefore, to identify the mechanisms regulating DNA demethylation dynamics, we developed a high-throughput method called scMspJI-seq to strand-specifically quantify 5mC on a genome-wide scale in single cells. Single cells are isolated into 384-well plates by fluorescence activated cell sorting or manual pipetting. Following cell lysis and protease treatment to remove chromatin, 5hmC sites in genome DNA (gDNA) are glucosylated using T4 phage β-glucosyltransferase (Fig. 1b). This modification blocks downstream detection of 5hmC and therefore, unlike bisulfite-based methods enables detection of only 5mC. Next, the restriction enzyme MspJI is added to the reaction mixture that recognizes ^{m}CNNR sites in the genome and creates double-stranded DNA breaks 16 bp downstream of the methylated cytosines leaving a 4-nucleotide 5' overhang[20]. Thereafter, double-stranded DNA adapters containing a 4-nucleotide 5' overhang are ligated to the fragmented gDNA molecules. These double-stranded DNA adapters, similar in design to those previously developed by us, contain a cell-specific barcode, a random 3 bp unique molecule identifier (UMI) to label individual 5mC sites on different alleles, a 5' Illumina adapter and a T7 promoter[21,22]. The ligated molecules are then amplified by *in vitro* transcription and used to prepare Illumina libraries as described previously (Fig. 1b)[21,22].

To validate the method, we first applied scMspJI-seq to single E14TG2a (E14) mouse embryonic stem cells (mES) cells. We found that 97.5% of the 5mCpG sites detected by scMspJI-seq in single cells overlapped with methylated sites observed in bulk bisulfite sequencing of E14 gDNA[23]. Further, as reported previously, we found that MspJI cuts gDNA 16 bp downstream of the methylated cytosine (Supplementary Fig. 1a)[20]. Most importantly, due to the maintenance activity of Dnmt1 in E14 cells, we observed similar levels of 5mC on both DNA strands of a chromosome in single cells, as expected (Supplementary Fig. 1b). To quantify the strand-specific

distribution of 5mCpG on each chromosome of a single cell, we defined a metric called as strand bias (denoted by *f*), which is the ratio of the number of 5mCpG sites detected on the plus strand divided by the total number of 5mCpG sites detected on both the plus and minus strands. Finally, to ensure that scMspJI-seq can detect differences in 5mCpG distribution between the two strands, and to confirm that the observed strand bias of 0.5 in E14 cells results from the maintenance activity of Dnmt1, we used CRISPR-Cas9 to knockout *Dnmt1*. We observed a dramatic increase in strand bias in E14 cells without *Dnmt1*, strongly suggesting that our new technology provides a sensitive readout of strand-specific methylation and the ability to distinguish between passive and active demethylation (Supplementary Fig. 1c).

During preimplantation development, the maternal and paternal genomes display dramatically different 5mC erasure dynamics, and therefore we next wanted to test our ability to quantify strand-specific 5mC at the resolution of individual alleles. As the single-cell measurements in E14 cells did not provide allele-specific detection of 5mC for each chromosome, we applied scMspJI-seq to hybrid serum grown mES cells (CAST/EiJ x 129/Sv background)[22]. While the majority of cells displayed methylation maintenance as expected, we surprisingly observed a small population of cells that showed strong 5mC strand bias (Fig. 2). For example, cell 562 displayed similar levels of 5mCpG on the two DNA strands of chromosomes across both alleles (Fig. 2a), whereas cell 216 showed substantially different levels of 5mC on each DNA strand of a chromosome (Fig. 2b). Pearson correlation coefficient (*r*) between the plus and minus strands of individual cells show that while a majority of cells displayed high correlation, a small subset of cells were weakly correlated, suggesting loss of methylation maintenance in these cells (Fig. 2c). Allele-specific 5mCpG strand bias further revealed the existence of two epigenetically distinct population of mES cells (Fig. 2d). Taken together with the E14 cells, these results highlight that in the absence of allele-specific measurements, strand-specific 5mC quantification is averaged across both alleles, potentially obscuring a detailed view of the methylation status of the genome. Finally, we find that these two distinct 5mC strand bias patterns are also observed at a sub-chromosomal resolution, suggesting this is genome-wide phenomenon that potentially arises from differential methylation maintenance between individual mES cells (Fig. 2e).

To validate this cell-to-cell heterogeneity in 5mC strand bias, we reanalyzed data from a recent study that quantified 5mC in single cells using bisulfite sequencing[24]. While this method is low throughput, single-cell bisulfite sequencing can potentially also be used to infer strand-specific 5mC[25]. In agreement with our findings using scMspJI-seq, reanalysis of the published dataset also revealed hybrid mES cells with similar levels of 5mC on the plus and minus strands, and a small fraction of cells with substantially different levels of 5mC on the two strands of a

chromosome (Fig. 3). These results validate our previous observation of two distinct mES cell populations with and without 5mC strand bias (Fig. 2).

After establishing this new method, we next used scMspJI-seq to gain a deeper understanding of the 5mC erasure dynamics during preimplantation mouse development as the mechanistic details regulating this genome-wide reprogramming remains unclear from previous work. Early immunofluorescence-based studies showed that 5mC marks on the paternal genome are converted to 5hmC in the zygote[8-11]. As 5hmC is not maintained through cell division, and can be further oxidized to be removed by cytidine deaminase and base excision repair pathways, the paternal genome is effectively demethylated from the 1-cell to early blastocyst stage (approximately E3.5 or 32-cell stage) of development[7]. These same studies also reported that the maternal genome retains 5mC in the zygote[8-11]. This observation together with reports that Dnmt1 is primarily cytoplasmic during these early cell divisions, indirectly suggested that the maternal genome is passively demethylated through a lack of maintenance methylation[26-29]. However, later studies showed the existence of two isoforms of Dnmt1, with the lowly abundant Dnmt1s isoform present in the nucleus of blastomeres[30-32]. Thus, it remains unclear the extent to which the maternal genome is passively demethylated during these early stages. Further, more recently, bulk 5mC and 5hmC sequencing during these early stages have shown that the maternal genome also carries 5hmC marks, suggesting that the maternal genome also undergoes partial active demethylation[13]. As the mechanisms underlying this critical process of 5mC erasure during embryonic development remains unclear, we used strand-specific detection of 5mC in single cells to probe the dynamics of demethylation more closely.

We performed scMspJI-seq on hybrid mouse embryos (CAST/EiJ x C57BL/6 background) from the 2- to 32-cell stage of development. In contrast to previous studies that suggested passive demethylation of the maternal genome due to cytoplasmic localization of Dnmt1, experiments in 2-cell hybrid mouse embryos surprisingly revealed that 5mCpG on the maternal genome shows a tight strand bias distribution centered around 0.5, implying similar amounts of the mark of both DNA strands and that Dnmt1-mediated methylation maintenance is active at this stage (Fig. 4a and Supplementary Fig. 2a). To ensure that this lack of strand bias in the maternal genome at the 2-cell stage is not a technical artifact or a consequence of high *de novo* methylation activity of Dnmt3a/3b, we quantified the levels of 5mCpA, the most abundant non-CpG methylation, in these cells. Non-CpG methylation is not a substrate for Dnmt1 and is deposited on the genome as a result of the activity of the *de novo* methyltransferases, Dnmt3a and Dnmt3b[33-35]. In the 2-cell embryos, we found that 5mCpA on the maternal genome showed a bimodal pattern of strand bias distribution, suggesting that the lack of strand bias observed for 5mCpG is possibly a result of the

maintenance activity of Dnmt1 and not a consequence of high *de novo* methylation rates by Dnmt3a/3b (Fig. 4b and Supplementary Fig. 2b). Further, we have previously shown that bimodal strand bias distributions for 5hmC in 2-cell mouse embryos arises from the slow kinetics of Tet activity and can be used to identify sister cells[21,36]. This is because 5hmC is not maintained through cell divisions and new DNA strands have lower levels of 5hmC than older strands, resulting in sister cells exhibiting anti-correlated strand bias patterns over all the chromosomes in a cell. Similarly, as 5mCpA is not maintained through cell division, we found that the strong anti-correlation in 5mCpA between chromosomes of single cells can be used to identify sister cells (Supplementary Fig. 2c,d). These results further imply that at the 2-cell stage of development the kinetics of *de novo* methylation by Dnmt3a and Dnmt3b is slow (Fig. 4b). Taken together, these experiments provide preliminary evidence that the similar levels of 5mCpG found on both DNA strands of chromosomes in 2-cell blastomeres is a result of Dnmt1 maintenance activity.

Quantifying the dynamics of demethylation beyond the 2-cell stage, we observed for both the maternal and paternal genomes that a majority of chromosomes displayed no significant 5mCpG strand bias up to the 16-cell stage (Fig. 4a and Supplementary Fig. 2a). Surprisingly, beyond the 16-cell stage, we observed a widening of the 5mCpG strand bias distribution, suggesting reduced Dnmt1 maintenance activity (Fig. 4a and Supplementary Fig. 2a). These experiments suggest two distinct phases during preimplantation mouse development – an initial period of Dnmt1-mediated maintenance methylation followed by passive demethylation. Finally, we observed that the 5mCpG strand bias distribution at the 32-cell stage is trimodal. Performing *k*-means clustering on the 5mCpG strand bias in these single cells identified two distinct groups of cells as inferred by the mean silhouette scores – a population with no strand bias and another population with a bimodal strand bias distribution (Fig. 4c,d). Further, within the bimodal population, we observed pairs of cells for which all chromosomes were strongly anti-correlated, suggesting that these pairs are sister cells (Fig. 4e and Supplementary Fig. 2e). These observations reveal the existence of significant cell-to-cell heterogeneity in the genome-wide methylome landscapes of cells within the early blastocyst. Taken together, these results suggest maintenance methylation is active till the 16-cell stage and that from the 16- to 32-cell stage, a fraction of cells within the embryo show strong 5mCpG strand bias and undergo passive demethylation.

Finally, to conclusively demonstrate that the absence of 5mCpG strand bias up to the 16-cell stage arises from Dnmt1 mediated maintenance methylation, we performed bulk hairpin bisulfite sequencing. A hallmark of Dnmt1 mediated methylation is that both cytosines in a CpG dyad are symmetrically methylated and therefore we performed bulk hairpin bisulfite sequencing

that enables interrogation of the methylation status of CpG dyads[37]. We observed that the fraction of symmetrically methylated CpG dyads in the genome is high up to the 16-cell stage, with a dramatic reduction at the 32-cell stage (that is matched by an increase in hemi-methylated CpG dyads at this stage), thereby demonstrating that maintenance methylation is active initially and is followed by passive demethylation at the 32-cell stage (Fig. 4f).

We finally extended scMspJI-seq to explore the dynamics of global demethylation in human preimplantation embryos, ranging from developmental day 2 to 7. Studies in human preimplantation embryos have shown temporally slower, yet similar developmental dynamics to mouse embryos[38]. Despite lacking allelic information, our results suggest that the mouse and human 5mCpG demethylation dynamics are similar, with an initial phase till the 16-cell stage displaying a tight 5mCpG strand bias distribution centered around 0.5, followed by an increase in strand bias in a small fraction of cells from the 32- to 128-cell stage (Fig. 5a and Supplementary Fig. 3a). This is consistent with previous immunostainings in human preimplantation embryos that show a decrease in DNMT1 protein levels between Day 5 and Day 6 blastocysts[39,40]. Further, 5mCpA strand-bias distributions of human preimplantation embryos appear to be similar to the trend observed in mouse embryos with a majority of cells till the 16-cell stage displaying 5mCpA strand bias (Fig. 5b and Supplementary Fig. 3b). Finally, upon closer inspection of 5mCpA strand bias per cell, we observed three sister pairs in Day 3 embryos with a mirrored pattern of strand bias along the entire genome (Supplementary Fig. 3c).

In summary, we have developed a new high-throughput strand-specific method that enables us to quantify 5mC on a genome-wide scale in single cells. When applied to serum grown mES cells, we found substantial cell-to-cell variability in strand-specific 5mC landscapes, revealing the existence of chromosome-wide heterogeneity in the methylome of mES cells. Reanalysis of a previous single-cell bisulfite sequencing study further confirmed these results[24]. Furthermore, in addition to exploring strand-specific 5mC heterogeneity in single cells, scMspJI-seq also enables systematic investigation of the mechanisms regulating demethylation dynamics. In preimplantation mouse embryos, we surprisingly discovered two distinct phases of methylation dynamics – an initial phase till the 16-cell stage where methylation maintenance is active, followed by loss of maintenance in a fraction of cells within the early blastocyst at the 32-cell stage. These results further highlight the presence of strand-specific 5mC heterogeneity between individual cells during early mammalian development. In future, we plan to explore how this genome-wide heterogeneity in the methylome regulates lineage commitment during development. Finally, despite the reduced resolution due to lack of allelic information, we found similar demethylation dynamics in preimplantation human embryos. Thus, scMspJI-seq presents a new single-cell

241 strand-specific technology that potentially can be used to probe the dynamics of methylation
242 during development, cancer progression, aging and in other biological systems.

METHODS

Cell culture

E14tg2a mouse embryonic stem cells were obtained from American Type Culture Collection (ATCC CRL-182) and the hybrid 129/Sv:CAST/EiJ mouse embryonic stem cells were obtained from Jop Kind's group (Hubrecht Institute). Both lines were tested for mycoplasma contamination. Cells were grown on 0.1% gelatin in ES cell culture media; DMEM (1x) high glucose + glutamax (Gibco), supplemented with 10% FCS (Greiner) 100 µM β-mercaptoethanol (Sigma), 100 µM Non-essential amino acids (Gibco), 50 µg/mL Pen/Strep (Gibco) and 1000 U/mL ESGRO mLIF (Millipore). Cells were split every 2 days and media changed every day. Cells were harvested before FACS by washing 3 times with 1x PBS with calcium and magnesium and incubated with 0.05% Trypsin (Life Technologies). Cell were resuspended in ES culture media and cell clumps were removed by passing the cells through a BD Falcon 5 mL polystyrene tube with a filter top.

Crispr-Cas9 Dnmt1 knockout

Six gRNA sequences targeting three exons of mouse Dnmt1 were used as described previously[41]. Phosphorylated BbsI compatible restriction overhangs were added to gRNA top and bottom oligos and resuspended at 100 µM in nuclease-free water. Annealing of the oligos was performed in 1x ligation buffer (NEB) using the following program: 97°C for 5 minutes, ramp down by 1°C per 1 minute to 20°C. The pX330 CRISPR-Cas9-GFP gRNA plasmid was a kind gift from Eva van Rooij and mixed with 0.1 µM gRNA oligo. The reaction was simultaneously digested with BbsI (NEB) and ligated with T4 DNA ligase (NEB) overnight at 16°C. Ligation reactions were transformed into DH5α competent cells and subsequently sequenced using Sanger dideoxy sequencing to confirm the correct insert. All six pX300-gRNA plasmids were pooled and 1 µg was transfected into 2 million E14tg2a cells using Lipofectamine (Life Technologies). A separate pX300 empty vector was also transfected into E14tg2a to serve as a negative control. Two days later, single GFP positive cells were sorted into 384-well plates (BioRad) and subjected to scMspJI-seq.

Preimplantation mouse embryo isolation

CAST/EiJ x C57BL/6 hybrid mouse embryos were obtained from four 3-month-old superovulated B6 mothers (injected with pregnant mare serum gonadotropin (PMSG) and human chorionic gonadotropin (HCG) 22 h later) and incubated in M16 medium at 37°C and 5% CO_2. Cells were isolated using hyaluronic acid (Sigma) and trypsin (Life Technologies) and manually deposited into 384-well plates containing lysis buffer and Vapor-lock. Plates were subsequently centrifuged

277 at 1,000 rpm for 1 minute to ensure that cells reach the aqueous phase and then subjected to
278 scMspJI-seq. All animal experiments were approved by the Royal Netherlands Academy of Arts
279 and Sciences and were performed according to the animal experimentation guidelines of the
280 KNAW.
281
282 **Preimplantation human embryo isolation**
283 Supernumerary cryopreserved human embryos were obtained for research from patients
284 undergoing *in vitro* fertilization (IVF) using standard clinical protocols, at the Department for
285 Reproductive Medicine, Ghent University Hospital. Cleavage stage embryos, cryopreserved on
286 day 2 or 3 of development, were warmed using EmbryoThaw™ media (Fertipro, Belgium), as
287 outlined by the manufacturer. Blastocyst stage embryos, vitrified on day 5 or 6 of development,
288 were warmed using the Vitrification Thaw kit (Irvine Scientific, Netherlands), as described[42].
289 Embryos were transferred to either Cook Cleavage or Cook Blastocyst Medium (COOK, Ireland)
290 depending on their developmental stage, and cultured in 20 µL medium droplets under mineral
291 oil (Irvine Scientific, Netherlands) at 37 °C, 6% CO_2 and 5% O_2. When required, embryos were
292 briefly treated with Acidic Tyrode's Solution (Sigma-Aldrich, Belgium) for removal of the zona
293 pellucida. All embryos were washed and subsequently dissociated by gentle mechanical
294 dissociation in TrypLE Express Enzyme (Life Technologies, Belgium) using glass capillaries.
295 Single blastomeres were washed and manually deposited into 384-well plates containing lysis
296 buffer and Vapor-Lock. Plates were subsequently centrifuged at 1,000 rpm for 1 minute and
297 stored at -80 °C until further processing. This study was approved by the Ghent University
298 Institutional Review Board (EC2015/1114) and the Belgian Federal Commission for medical and
299 scientific research on embryos in vitro (ADV_060_UZGent). All embryos were donated following
300 patients' written informed consent.
301
302 **scMspJI-seq**
303 Prior to FACS or manual isolation of single cells, 384-well plates (BioRad) are prepared as follows:
304 4 µL of Vapor-Lock (Qiagen) is manually added to each well using a multichannel pipette followed
305 by 2 µL of lysis buffer (0.2 µL of 25 µg/µL Qiagen Protease, 0.2 µL of 10x NEB Buffer 4 and 1.6
306 µL of nuclease-free water) using the Nanodrop II liquid-handling robot (BioNex Solutions). All
307 downstream dispensing steps are performed using the liquid-handling robot. After spinning down
308 the 384-well plates, single cells are deposited into each well of the plate and incubated at 50°C
309 for 15 hours, 75°C for 20 minutes and 80°C for 5 minutes. 5hmC sites in the genome are then
310 glucosylated to block downstream recognition by MspJI by dispensing 0.5 µL of the following

reaction mixture: 0.1 μL of T4-BGT (NEB), 0.1 μL of UDP-Glucose (NEB), 0.05 μL of 10x NEB Buffer 4 and 0.25 μL of nuclease-free water. After incubation at 37°C for 16 hours, 0.5 μL the following reaction mixture is added: 0.1 μL of 25 μg/μL Qiagen Protease, 0.05 μL of 10x NEB Buffer 4 and 0.35 μL of nuclease-free water. The plate is then incubated at 50°C for 5 hours, 75°C for 20 minutes and 80°C for 5 minutes. Thereafter, gDNA is digested by the restriction enzyme MspJI by the addition of 0.5 μL of the following reaction mixture: 0.02 μL of MspJI (NEB), 0.12 μL of 30x enzyme activator solution (NEB), 0.05 μL of 10x NEB Buffer 4 and 0.31 μL of nuclease-free water. The digestion is performed at 37°C for 5 hours followed by heat inactivation of MspJI at 65°C for 20 minutes. Next, 0.2 μL of cell-specific double-stranded adaptors are added to individual wells and these adapters are ligated to the fragmented gDNA molecules by adding 0.8 μL of the following reaction mixture: 0.07 μL of T4 DNA ligase (NEB), 0.1 μL of T4 DNA ligase buffer (NEB), 0.3 μL of 10 mM ATP (NEB) and 0.33 μL of nuclease-free water. The ligation is performed at 16°C for 16 hours. Next, wells containing unique cell-specific adapters are pooled using a multichannel pipette and incubated with 0.8x Agencourt Ampure (Beckman Coulter) beads for 30 minutes, washed twice with 80% ethanol and resuspended in 6.4 μL of nuclease-free water. Thereafter, *in vitro* transcription and Illumina library preparation is performed as described previously in the scAba-seq protocol.

scMspJI-seq adapters

The double-stranded scMspJI-seq adapters are designed to contain a T7 promoter, 5' Illumina adapter, 3 bp UMI, 8 bp cell-specific barcode, and a random 4-nucleotide 5' overhang. The general design of the top and bottom strand is shown below:

Top oligo:

5' –CGATTGAGGCCGGTAATACGACTCACTATAGGGGTTCAGAGTTCTACAGTCCGACGATCNNN[8 bp cell-barcode] – 3'

Bottom oligo:

5' – NNNN[8 bp cell-barcode]NNNGATCGTCGGACTGTAGAACTCTGAACCCCTATAGTGAGTCGTATTACCGGCCTCAATCG – 3'

The sequence of the 8 bp cell-specific barcode is provided in Supplementary Table 1. The protocol for phosphorylating the bottom strand and for annealing the top and bottom strands to generate the double-stranded adapters is described previously in the scAba-seq protocol.

scMspJI-seq analysis pipeline

scMSpJI-seq libraries were sequenced on an Illumina NextSeq 500 platform. Reads containing the correct cell-specific barcode were mapped to the mouse (mm10) or human (hg19) genome using the Burrows-Wheeler Aligner (BWA) and filtered for uniquely mapping reads to the genome. Custom scripts written in Perl were then used to demultiplex the data, identify 5mC position, strand information, and remove PCR duplicates. Custom codes will be made available upon request.

Strand-specific scNMT-seq analysis pipeline

Bisulfite sequencing data from published scNMT libraries (GSE109262) [24] were processed as described previously[43]. The first nine bases of the raw reads were trimmed using Trim Galore (v0.5.0) and mapped using Bismark (v20) to the mouse genome (mm10) with the 129/CAST background. SNPs specific to 129/CAST mouse genome were prepared using SNPsplit (v0.3.2) and a list of known variant call files from the Mouse Genomes Project (http://www.sanger.ac.uk/resources/mouse/genomes/). After mapping with Bismark, duplicate sequences were removed and CpG methylation calls were extracted with strand-specific information. Further data analysis and visualization of the methylation calls used custom scripts that will be made available upon request.

Hairpin Bisulfite Sequencing

Hairpin bisulfite sequencing was performed on bulk mouse embryos samples (2- to 64-cell stage mouse embryos). The embryos were treated with protease (1 μL of 25 μg/μL Qiagen Protease, 1 μL of 10x NEB Buffer 4, and 8 μL of nuclease-free water). Then, 0.5 ng of genomic DNA was digested with 20 μL of MspI master mix (1 μL of MspI (NEB), 2 μL 10x NEB CutSmart Buffer in a total volume of 20 μL) and incubated at 37°C for 1 hour. After digestion, the fragmented genomic DNA was ligated with 1 μL of 10 μM phosphorylated hairpin oligo mix (1 μL of NEB T4 ligase, 1 μL of 10x NEB T4 Ligase buffer, 2 μL of 10mM ATP, 5 μL of nuclease-free water) and incubated overnight at 16°C. The hairpin oligo was prepared as follows: The oligo (G/iMe-dC/iMe-dC/G/iMe-dC/iMe-dC/GG/iMe-dC/GG/iMe-dC/AAG/iBiodT/GAAG/iMe-dC/iMe-dC/G/iMe-dC/iMe-dC/GG/iMe-dC/G) was resuspended in 100 μM of Low-TE. The hairpin oligo was then phosphorylated (1 μL of 100 μM hairpin oligo, 3 μL of 10x T4 Ligase Buffer, 1 μL T4 PNK and 5 μL of nuclease free water) and incubated at 37°C for an hour. Subsequently, the phosphorylated oligo was heated at 94°C and placed in ice water to generate the loop. For purification of the ligation mixture, Dynabeads™ M-280 Streptavidin beads were used following the recommended

manufacturer's protocol with the following changes: the bead-ligation mixture was incubated for 1 hour at RT on a rotator and a cold 10 mM Tris-HCl wash step was included. Subsequently, we performed bisulfite sequencing on the sample using the protocol described previously[44]. After sequencing the libraries on a Miseq 300 bp or NextSeq 500 75 bp pair-end run, we used HBS-tools and custom Perl scripts to analyze the methylated CpG dyads[45].

Data Availability

Accession code GEO: In preparation

Code Availability

Custom codes to detect strand-specific 5mC using scMspJI-seq will be made available upon request.

References

1. Cedar, H. & Bergman, Y. Programming of DNA Methylation Patterns. *Annu. Rev. Biochem.* **81,** 97–117 (2012).
2. Dirk, S. Function and information content of DNA methylation. *Nature* **517,** 321–326 (2015).
3. Hackett, J. A. & Surani, M. A. DNA methylation dynamics during the mammalian life cycle. *Philos. Trans. R. Soc. Lond., B, Biol. Sci.* **368,** 20110328 (2013).
4. Leonhardt, H., Page, A. W., Weier, H.-U. & Bestor, T. H. A targeting sequence directs DNA methyltransferase to sites of DNA replication in mammalian nuclei. *Cell* **71,** 865–873 (1992).
5. Tahiliani, M. *et al.* Conversion of 5-Methylcytosine to 5-Hydroxymethylcytosine in Mammalian DNA by MLL Partner TET1. *Science* **324,** 930–935 (2009).
6. Ito, S. *et al.* Tet proteins can convert 5-methylcytosine to 5-formylcytosine and 5-carboxylcytosine. *Science* **333,** 1300–1303 (2011).
7. Wu, X. & Zhang, Y. TET-mediated active DNA demethylation: mechanism, function and beyond. *Nat. Rev. Genet.* **18,** 517–534 (2017).
8. Wossidlo, M. *et al.* 5-Hydroxymethylcytosine in the mammalian zygote is linked with epigenetic reprogramming. *Nat. Commun.* **2,** 241 (2011).
9. Mayer, W., Niveleau, A., Walter, J., Fundele, R. & Haaf, T. Demethylation of the zygotic paternal genome. *Nature* **403,** 501–502 (2000).
10. Inoue, A. & Zhang, Y. Replication-dependent loss of 5-hydroxymethylcytosine in mouse preimplantation embryos. *Science* **334,** 194–194 (2011).
11. Iqbal, K., Jin, S.-G., Pfeifer, G. P. & Szabo, P. E. Reprogramming of the paternal genome upon fertilization involves genome-wide oxidation of 5-methylcytosine. *Proc. Natl. Acad. Sci. U.S.A.* **108,** 3642–3647 (2011).
12. Smith, Z. D. *et al.* A unique regulatory phase of DNA methylation in the early mammalian embryo. *Nature* **484,** 339–344 (2012).
13. Wang, L. *et al.* Programming and Inheritance of Parental DNA Methylomes in Mammals. *Cell* **157,** 979–991 (2014).
14. Arand, J. *et al.* Selective impairment of methylation maintenance is the major cause of DNA methylation reprogramming in the early embryo. *Epigenetics Chromatin* **8,** 1 (2015).
15. Okamoto, Y. *et al.* DNA methylation dynamics in mouse preimplantation embryos revealed by mass spectrometry. *Sci. Rep.* **6,** 19134 (2016).
16. Guo, H. The DNA methylation landscape of human early embryos. *Nature* **511,** 606–610 (2014).

17. Smith, Z. D. *et al.* DNA methylation dynamics of the human preimplantation embryo. *Nature* **511,** 611–615 (2014).
18. Okae, H. *et al.* Genome-Wide Analysis of DNA Methylation Dynamics during Early Human Development. *PLoS Genet.* **10,** e1004868 (2014).
19. Huang, Y. *et al.* The Behaviour of 5-Hydroxymethylcytosine in Bisulfite Sequencing. *PLoS One* **5,** e8888 (2010).
20. Cohen-Karni, D. *et al.* The MspJI family of modification-dependent restriction endonucleases for epigenetic studies. *Proc. Natl. Acad. Sci. U.S.A.* **108,** 11040–11045 (2011).
21. Mooijman, D., Dey, S. S., Boisset, J. C., Crosetto, N. & van Oudenaarden, A. Single-cell 5hmC sequencing reveals chromosome-wide cell-to-cell variability and enables lineage reconstruction. *Nat. Biotechnol.* **34,** 852–856 (2016).
22. Rooijers, K. *et al.* Simultaneous quantification of protein-DNA contacts and transcriptomes in single cells. *Nat. Biotechnol.* **37,** 766–772 (2019).
23. Habibi, E. *et al.* Whole-Genome Bisulfite Sequencing of Two Distinct Interconvertible DNA Methylomes of Mouse Embryonic Stem Cells. *Cell Stem cell* **13,** 360–369 (2013).
24. Clark, S. J. *et al.* scNMT-seq enables joint profiling of chromatin accessibility DNA methylation and transcription in single cells. *Nat. Commun.* **9,** 781 (2018).
25. Lister, R. *et al.* Human DNA methylomes at base resolution show widespread epigenomic differences. *Nature* **462,** 315–322 (2009).
26. Carlson, L. L., Page, A. W. & Bestor, T. H. Properties and localization of DNA methyltransferase in preimplantation mouse embryos: implications for genomic imprinting. *Genes Dev.* **6,** 2536–2541 (1992).
27. Cardoso, M. C. & Leonhardt, H. DNA Methyltransferase Is Actively Retained in the Cytoplasm during Early Development. *J. Cell Biol.* **147,** 25–32 (1999).
28. Howell, C. Y. *et al.* Genomic Imprinting Disrupted by a Maternal Effect Mutation in the Dnmt1 Gene. *Cell* **104,** 829–838 (2001).
29. Ratnam, S. *et al.* Dynamics of Dnmt1 Methyltransferase Expression and Intracellular Localization during Oogenesis and Preimplantation Development. *Dev. Biol.* **245,** 304–314 (2002).
30. Cirio, M. C. *et al.* Preimplantation expression of the somatic form of Dnmt1 suggests a role in the inheritance of genomic imprints. *BMC Dev. Biol.* **8,** 9 (2008).
31. Kurihara, Y. *et al.* Maintenance of genomic methylation patterns during preimplantation development requires the somatic form of DNA methyltransferase 1. *Dev. Biol.* **313,** 335–346 (2008).

32. Hirasawa, R. *et al.* Maternal and zygotic Dnmt1 are necessary and sufficient for the maintenance of DNA methylation imprints during preimplantation development. *Genes Dev.* **22,** 1607–1616 (2008).
33. Ramsahoye, B. H. *et al.* Non-CpG methylation is prevalent in embryonic stem cells and may be mediated by DNA methyltransferase 3a. *Proc. Natl. Acad. Sci. U.S.A.* **97,** 5237–5242 (2000).
34. Ziller, M. J. *et al.* Genomic Distribution and Inter-Sample Variation of Non-CpG Methylation across Human Cell Types. *PLoS Genet.* **7,** e1002389 (2011).
35. Arand, J. *et al.* In vivo control of CpG and non-CpG DNA methylation by DNA methyltransferases. *PLoS Genet.* **8,** e1002750 (2012).
36. Wangsanuwat, C., Aldeguer, J. F., Rivron, N. C. & Dey, S. S. A probabilistic framework for cellular lineage reconstruction using single-cell 5-hydroxymethylcytosine sequencing. *bioRxiv* doi:https://doi.org/10.1101/739300
37. Zhao, L. *et al.* The dynamics of DNA methylation fidelity during mouse embryonic stem cell self-renewal and differentiation. *Genome Res.* **24,** 1296–1307 (2014).
38. Iurlaro, M., Meyenn, von, F. & Reik, W. DNA methylation homeostasis in human and mouse development. *Curr. Opin. Genet. Dev.* **43,** 101–109 (2017).
39. Petrussa, L., Van de Velde, H. & De Rycke, M. Dynamic regulation of DNA methyltransferases in human oocytes and preimplantation embryos after assisted reproductive technologies. *Mol. Hum. Reprod.* **20,** 861–874 (2014).
40. Petrussa, L., Van de Velde, H. & De Rycke, M. Similar kinetics for 5-methylcytosine and 5-hydroxymethylcytosine during human preimplantation development in vitro. *Mol. Reprod. Dev.* **83,** 594–605 (2016).
41. Sanjana, N. E., Shalem, O. & Zhang, F. Improved vectors and genome-wide libraries for CRISPR screening. *Nat. Methods* **11,** 783–784 (2014).
42. Van Landuyt, L. *et al.* Closed blastocyst vitrification of biopsied embryos: evaluation of 100 consecutive warming cycles. *Hum. Reprod.* **26,** 316–322 (2011).
43. Smallwood, S. A. *et al.* Single-cell genome-wide bisulfite sequencing for assessing epigenetic heterogeneity. *Nat. Methods* **11,** 817–820 (2014).
44. Clark, S. J. *et al.* Genome-wide base-resolution mapping of DNA methylation in single cells using single-cell bisulfite sequencing (scBS-seq). *Nat. Protoc.* **12,** 534–547 (2017).
45. Sun, M.-A., Velmurugan, K. R., Keimig, D. & Xie, H. HBS-Tools for Hairpin Bisulfite Sequencing Data Processing and Analysis. *Adv. Bioinformatics* **2015,** 760423 (2015).

Acknowledgements

We would like to thank members of the van Oudenaarden and Dey groups for constructive feedback. We thank patients of the Department for Reproductive Medicine, Ghent University Hospital, for donating their cryopreserved embryos for this study. We acknowledge support for the computational work from the Center for Scientific Computing at the California NanoSystems Institute (CNSI) and Materials Research Laboratory (MRL) at UCSB: an NSF MRSEC (DMR-1720256) and NSF CNS-1725797. We thank Ferring Pharmaceuticals (Aalst, Belgium) for an unrestricted educational grant. This work was supported by a Fonds Wetenschappelijk Onderzoek – Vlaanderen (FWO, Research Foundation – Flanders; G051516N) grant to B.H, the De Snoo-van't Hoogerhuijs Stichting to S.C.d.S.L., an European Research Council Advanced grant (ERC-AdG 742225-IntScOmics) and a Nederlandse Organisatie voor Wetenschappelijk Onderzoek (NWO) TOP award (NWO-CW 714.016.001) to A.v.O., and an UC Cancer Research Coordinating Committee (CTN-19-585462) grant and an NIH R01HD099517 grant to S.S.D. This work is part of the Oncode Institute which is partly financed by the Dutch Cancer Society.

Author Contributions

M.S., D.M., S.S.D. and A.v.O. designed the study. S.S.D. and D.M. developed the method. S.S.D., M.S., D.M. performed experiments. J.-C.B. and D.M. isolated preimplantation mouse embryos. M.P., S.C.d.S.L. isolated preimplantation human embryos. M.S., D.M., J.-C.B., A.C., S.S.D. and A.v.O. analyzed the data. S.S.D., M.S., D.M. and A.v.O. wrote the manuscript.

Competing Interests

The authors declare no competing financial interests.

Figures

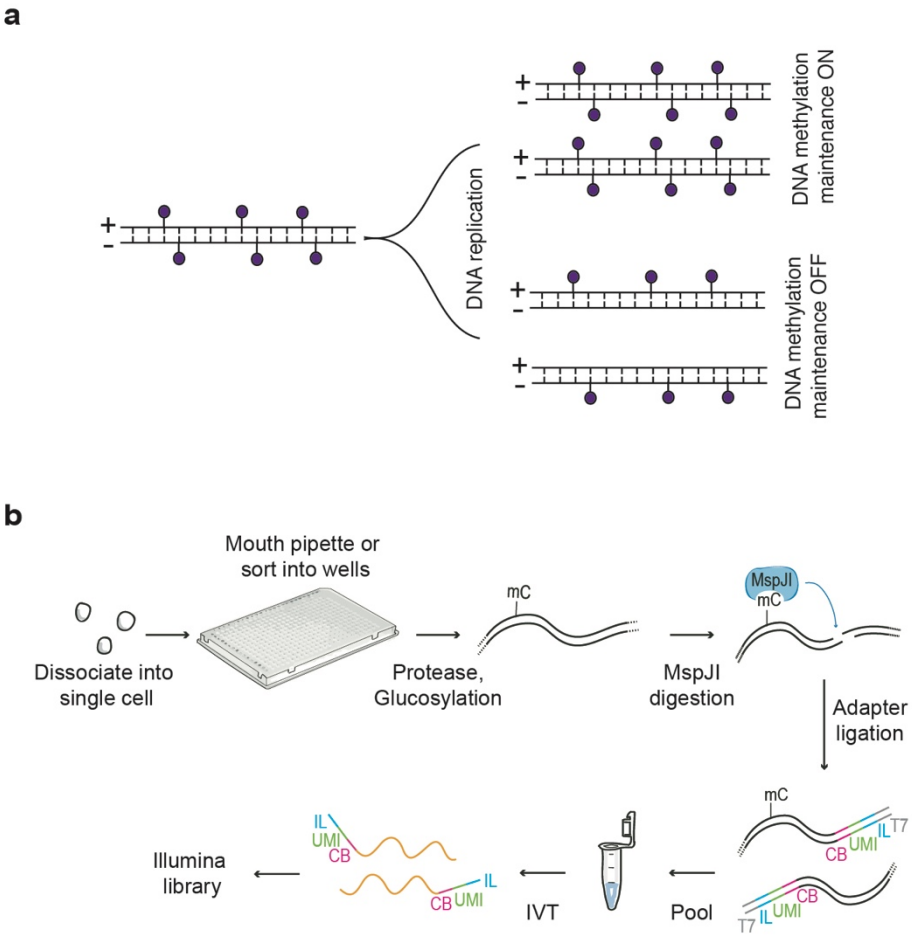


Figure 1 | Schematic of scMspJI-seq. (**a**) DNA methylation maintenance can be probed using strand-specific quantification of 5mC in single cells. Cells displaying methylation maintenance have symmetric levels of 5mCpG on both DNA strands of a chromosome while loss of methylation maintenance results in asymmetric levels of 5mCpG between the two DNA strands. (**b**) Single cells isolated by FACS or manual pipetting are deposited into 384-well plates and lysed. Following protease treatment to strip off chromatin and blocking of 5hmC sites by glucosylation, MspJI is used to recognize 5mC sites and cut gDNA 16 bp downstream of the methylated cytosine. After ligating double-stranded adapters – containing a cell-specific barcode (CB, pink), a random 3 bp unique molecule identifier to label individual 5mC sites on different alleles (UMI, green), 5' Illumina adapter (IL, blue) and T7 promoter (T7, gray) – to the fragmented gDNA, molecules from all single cells are pooled and amplified by *in vitro* transcription. The amplified RNA molecules are used to prepare scMspJI-seq libraries and sequenced on an Illumina platform.

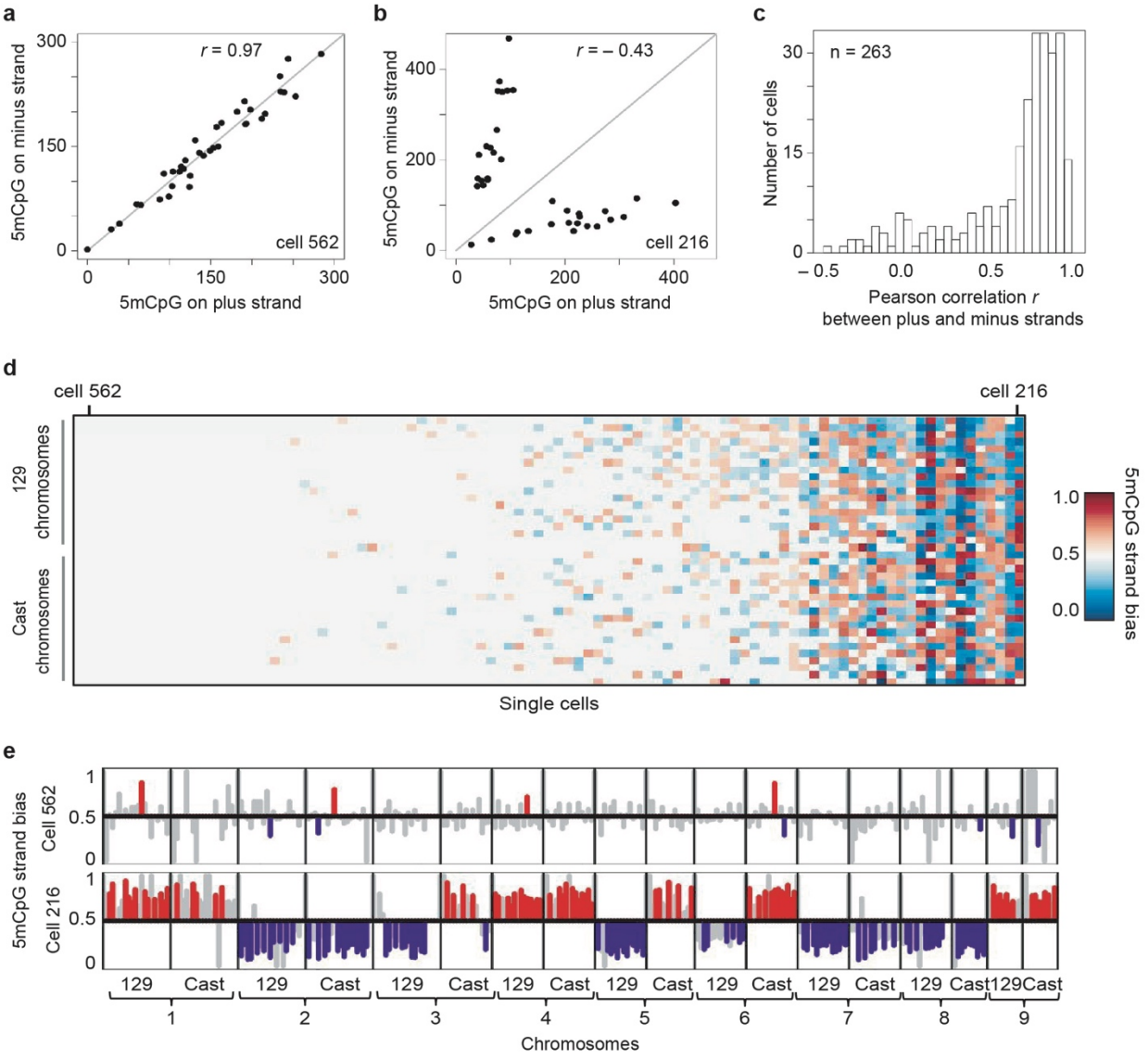


Figure 2 | Cell-to-cell heterogeneity in genome-wide strand-specific methylome landscapes in mES cells. (**a**) An example of a mES cell (cell #216) that shows similar amounts of 5mCpG on both the plus and the minus strand of each chromosome. (**b**) Another mES cell (cell #526) with asymmetric amounts of 5mCpG between the plus and the minus strand of each chromosome. (**c**) Histogram of Pearson correlations between the 5mCpG levels on the plus and the minus stand over all chromosomes in a cell show that while a majority of cells have similar amounts of 5mCpG on both strands (high Pearson correlation), a small fraction of cells display unequal levels of 5mCpG between the two strands of each chromosome (low Pearson correlation). (**d**) Ordered heatmap showing 5mCpG strand bias per chromosome for the maternal and paternal alleles in individual mES cells. (**e**) 5mCpG strand bias of cell #526 (*top*) and cell #216 (*bottom*) for 10 MB bins along the first 9 chromosomes are shown with statistically significant

(*P* < 0.05, likelihood ratio test) strand biases towards the plus and minus strands shown in red and blue, respectively. Strand biases of bins that are not statistically significant are shown in gray (*P* > 0.05, likelihood ratio test).

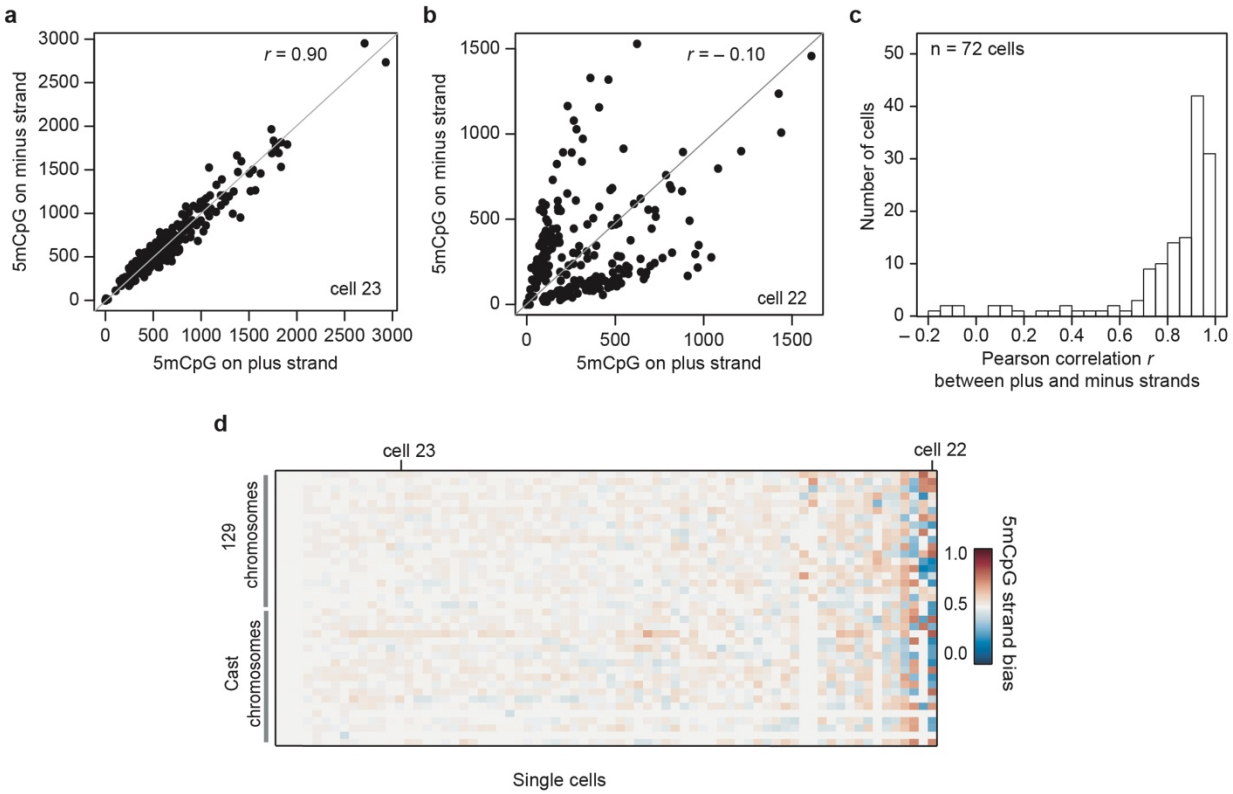
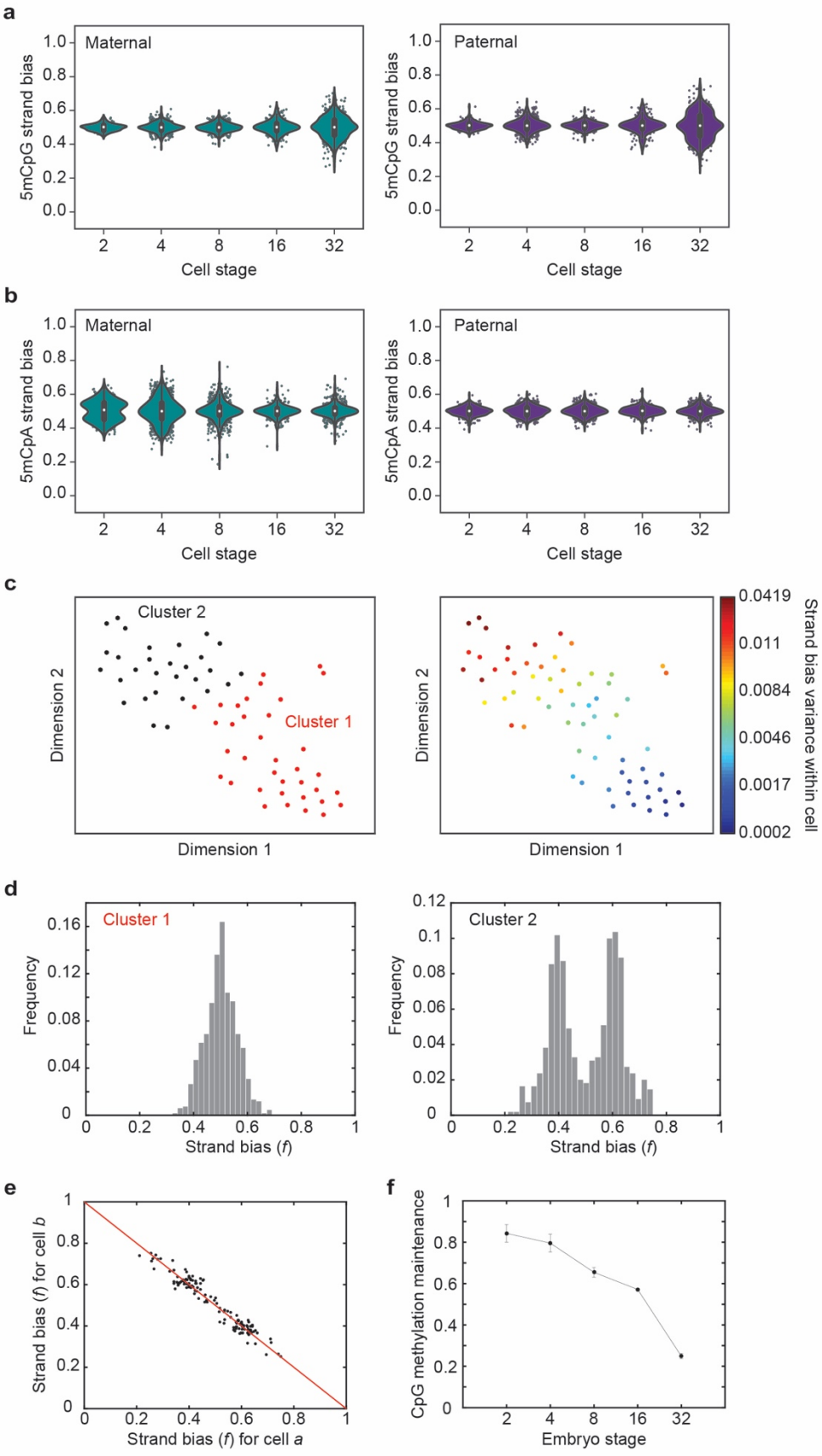


Figure 3 | Variability in strand-specific 5mCpG profiles in mES cells. (**a**) A representative mES cell (cell #23) with similar amounts of 5mCpG within 10 MB bins on both DNA strands. (**b**) Another representative mES cell (cell #22) with unequal amounts of 5mCpG between the two DNA strands for 10 MB bins. (**c**) Histogram of Pearson correlations between the 5mCpG levels on the plus and the minus stand over the entire genome (10 MB) in a cell. (**d**) Ordered heatmap showing 5mCpG strand bias per chromosome for maternal and paternal alleles in individual mES cells (n=72). The results in this figure is based on strand-specific reanalysis of single-cell bisulfite sequencing data obtained from previous work by Clark *et al.* [24]



552

Figure 4 | DNA demethylation dynamics in preimplantation mouse embryos. (**a**) Violin plots of 5mCpG strand bias for both the maternal (*left*) and paternal (*right*) genome show a tight distribution centered around $f = 0.5$ till the 16-cell stage and a wider distribution at the 32-cell stage of development. (**b**) For the maternal genome (*left*), 5mCpA strand bias show a bimodal distribution at the 2-cell stage that moves towards a tight unimodal distribution by the 32-cell stage of development. The paternal genome (*right*) shows a unimodal distribution centered at $f = 0.5$ throughout preimplantation development till the 32-cell stage. (**c**) t-SNE map displaying 2 cluster of single cells. These clusters were identified by *k*-means clustering on the 5mCpG strand bias for all paternal chromosomes (*left*). The *right* panel shows the strand bias variance within each cell superimposed on the t-SNE map. (**d**) The two clusters shown in panel **c** display dramatically different 5mCpG strand bias distributions – one cluster (*left*) shows a unimodal distribution while the other cluster (*right*) shows a bimodal distribution implying loss of methylation maintenance. (**e**) Strand bias of chromosomes between anti-correlated cell pairs suggesting that these pairs are sister cells. (**f**) Bulk hairpin bisulfite sequencing reveals that the fraction of CpG dyads that are symmetrically methylated drops substantially from the 16- to 32-cell stage of development.

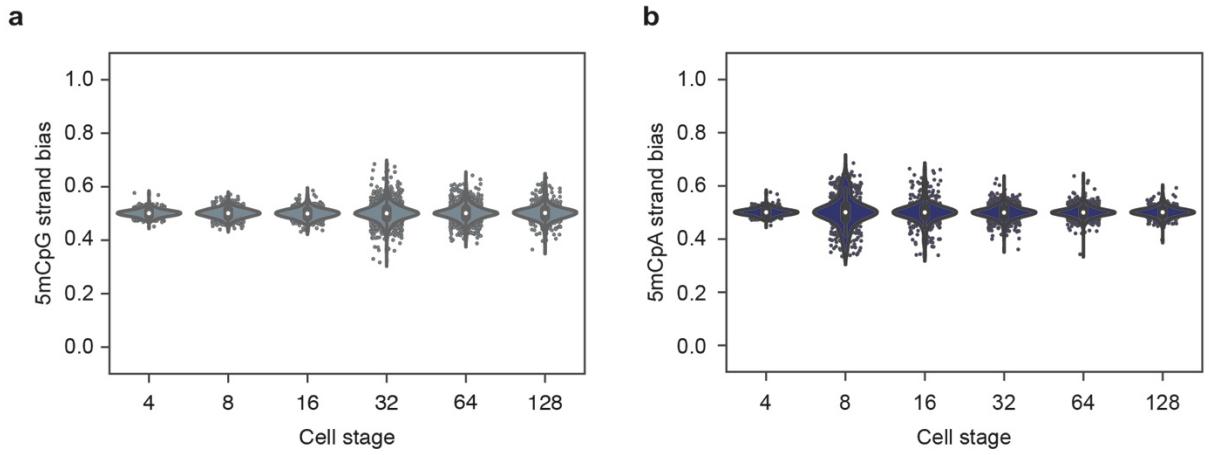


Figure 5 | DNA demethylation dynamics in preimplantation human embryos. (**a**) Violin plots showing 5mCpG strand bias from the 4- to 128-cell stage of human embryogenesis. In the absence of allele specific information, the strand bias represents an average over both alleles. Similar to mouse embryos, human embryos initially show no 5mCpG strand followed by an increase at the 16-cell stage of embryogenesis. (**b**) Violin plots showing 5mCpA strand bias from the 4- to 128-cell stage of human embryogenesis. 5mCpA strand bias dynamics in human embryos is similar to that observed in mouse embryos in Figure 4b.