Genome Biology

RESEARCH Open Access

A DNA methylation state transition model reveals the programmed epigenetic heterogeneity in human pre-implantation embryos

Chengchen Zhao[1†], Naiqian Zhang[2†], Yalin Zhang[1,3†], Nuermaimaiti Tuersunjiang[1], Shaorong Gao[1,3], Wenqiang Liu[1,3*] and Yong Zhang[1*]

* Correspondence: liuwenqiang@tongji.edu.cn; yzhang@tongji.edu.cn
†Chengchen Zhao, Naiqian Zhang and Yalin Zhang contributed equally to this work.
[1]Institute for Regenerative Medicine, Shanghai East Hospital, Shanghai Key Laboratory of Signaling and Disease Research, Frontier Science Center for Stem Cell Research, School of Life Science and Technology, Tongji University, Shanghai 200092, China
Full list of author information is available at the end of the article

Abstract

Background: During mammalian early embryogenesis, expression and epigenetic heterogeneity emerge before the first cell fate determination, but the programs causing such determinate heterogeneity are largely unexplored.

Results: Here, we present MethylTransition, a novel DNA methylation state transition model, for characterizing methylation changes during one or a few cell cycles at single-cell resolution. MethylTransition involves the creation of a transition matrix comprising three parameters that represent the probabilities of DNA methylation-modifying activities in order to link the methylation states before and after a cell cycle. We apply MethylTransition to single-cell DNA methylome data from human pre-implantation embryogenesis and elucidate that the DNA methylation heterogeneity that emerges at promoters during this process is largely an intrinsic output of a program with unique probabilities of DNA methylation-modifying activities. Moreover, we experimentally validate the effect of the initial DNA methylation on expression heterogeneity in pre-implantation mouse embryos.

Conclusions: Our study reveals the programmed DNA methylation heterogeneity during human pre-implantation embryogenesis through a novel mathematical model and provides valuable clues for identifying the driving factors of the first cell fate determination during this process.

Keywords: DNA methylation, Heterogeneity, First cell fate determination

Background

During mammalian pre-implantation embryogenesis, gene transcription regulation undergoes dramatic reprogramming [1–3], and expression heterogeneity emerges among cells within the same embryo before the first cell fate determination, i.e., the separation between the inner cell mass (ICM) and trophectoderm (TE) [4, 5]. A few reports have suggested that such heterogeneity could affect the first cell fate determination. For example, expression heterogeneity of *Carm1* in the 4-cell stage mouse embryo has been reported to cause later imbalanced expression of *Sox21*, which is critical in the determination of cell differentiation into ICM or TE cells [6, 7]. Two recent studies have indicated that the cause of *Carm1* expression heterogeneity can be traced back to the 2-cell stage and is associated with the expression heterogeneity of two long non-coding RNAs, *LincGET* [8] and *Neat1* [9]. Considering the robustness of early embryogenesis, such expression heterogeneity has been suggested to be the determinate result of a programmed process [10, 11], i.e., a set of cellular instructions. However, to the best of our knowledge, the suggested program causing such heterogeneity is largely unexplored. In addition to gene expression levels, DNA methylation levels have been reported to be heterogeneous among cells in pre-implantation embryos [12], likely as a result of a programmed process. As DNA methylation plays important roles in gene transcriptional regulation, the expression heterogeneity that emerges during early embryogenesis could at least partially be explained by DNA methylation. In contrast to gene expression, DNA methylation exists in a binary state for each CpG dyad, and the transition of its state during each cell cycle can be described quantitatively [13]. In addition, DNA methylation can be measured at a single CpG resolution in single cells [12, 14–16]. Therefore, it is promising to elucidate the set of instructions causing such epigenetic heterogeneity during early embryogenesis by building a mathematical model.

Several mathematical models have been built to quantitatively describe the DNA methylation state transition across cell cycles [17–24]. Two models built in the early 1990s specified that the DNA methylation state in a CpG dyad can be one of three states, unmethylated, hemi-methylated, and methylated, and that the type of DNA methylation state transition can be either de novo methylation or methylation maintenance [17, 18]. Genereux et al. and Sontag et al. improved the models by considering de novo methylation on different DNA strands separately [19, 20]. Fu et al. and Arand et al. considered the discriminates among the DNA methyltransferases (DNMTs) based on hairpin bisulfite sequencing (hairpin-BS-seq) data [21, 22]. McGovern et al. and von Meyenn et al. further incorporated 5-hydroxymethylcytosine (5hmC) into models to reflect the contribution of active demethylation to the DNA methylation state transition [23, 24]. Recently, Busto-Moner et al. enabled genome-wide quantification of subcell cycle kinetics of methylation maintenance using replication-associated bisulfite sequencing (Repli-BS-seq) data [25]. Researchers have proposed two types of methods to estimate the parameters of the models. One type of method relies on the acquisition of a DNA methylation equilibrium state. For example, Genereux et al. applied a maximum likelihood to estimate the parameters when the equilibrium state was reached [19], and Sontag et al. used a Markov chain model with a defined steady state [20]. The other type relies on the availability of specific genomics data. For example, Arand et al. calculated the efficiencies of DNMTs, i.e., the parameters of the transition model, based on hairpin-BS-seq data upon knockout (KO) of each DNMT [22]. von Meyenn et al. applied partial differential equations to estimate the parameters by using reduced representation bisulfite sequencing (RRBS) data, hairpin-BS-seq data, and TET-assisted bisulfite sequencing (TAB-seq) data [24]. Busto-Moner et al. used maximum likelihood estimation (MLE) to estimate the parameters quantifying the remethylation kinetics of nascent DNA post-replication relied on Repli-BS-seq data [25].

Although several mathematical models are available to describe the DNA methylation state transition, none of them can be applied to study DNA methylation heterogeneity during mammalian early embryogenesis using single-cell DNA methylome data, mainly for three reasons. First, the DNA methylation level changes dramatically across each cell cycle during early embryogenesis, demonstrating that the DNA methylation level is far from an equilibrium state. Therefore, models that incorporate parameter estimation methods based on an equilibrium state are not suitable. Second, active removal of DNA methylation is prevalent during mammalian early embryogenesis [1–3]; thus, models that do not consider active demethylation cannot be applied to this process. Third, due to technical limitations, technologies for profiling 5-methylcytosine (5mC) and 5hmC states together within the same cell are still not available, so models requiring both 5mC and 5hmC information are not suitable. Overall, a new DNA methylation state transition model is needed to study DNA methylation heterogeneity during mammalian early embryogenesis.

In this study, we developed a DNA methylation state transition model, MethylTransition, for characterizing methylation changes during one or a few cell cycles at single-cell resolution. MethylTransition introduces a methylation state ratio vector with 5 discrete states to describe the overall pattern of DNA methylation states for a given cell. To link the two methylation state ratio vectors before and after a cell cycle, MethylTransition uses a transition matrix comprising 3 parameters separately representing the probabilities of DNA methylation maintenance, active demethylation, and de novo methylation. The model estimates the parameters via a matrix approximation strategy with the Newton-Raphson method. Unlike previous models, MethylTransition does not need an equilibrium state or a 5hmC profile. Therefore, it is suitable for the elucidation of DNA methylation heterogeneity during mammalian early embryogenesis based on single-cell bisulfite sequencing (scBS-seq) data. We applied MethylTransition to scBS-seq data during human pre-implantation embryogenesis, and we found that the DNA methylation heterogeneity that emerges is largely determined by the initial DNA methylation state at the zygote stage (i.e., the 1-cell stage) and by a set of instructions represented by the model parameters. We further discussed the potential regulatory effects of DNA methylation heterogeneity in early embryos. The source code for MethylTransition can be found at https://github.com/TongjiZhanglab/MethylTransition [26].

Results

Computational framework of MethylTransition

To quantitatively characterize the DNA methylation state transition across a single cell cycle based on scBS-seq data, we developed MethylTransition, a probabilistic model with robustly estimated parameters. MethylTransition relies on the assumption that changes in the DNA methylation state at a CpG site across a single cell cycle occur in three steps: passive demethylation during DNA replication, active DNA methylation alteration by DNA methylation-modifying enzymes, and DNA methylation state combination during the combination of non-sister chromatids (Fig. 1a). A CpG site contains two CpG dyads in a diploid cell, while a CpG dyad contains two complemented CpG dinucleotides (Additional file 1: Fig. S1a). To illustrate the state transition probability at each step, we used *1* as a symbol to represent methylated CpG dinucleotides and *0* to

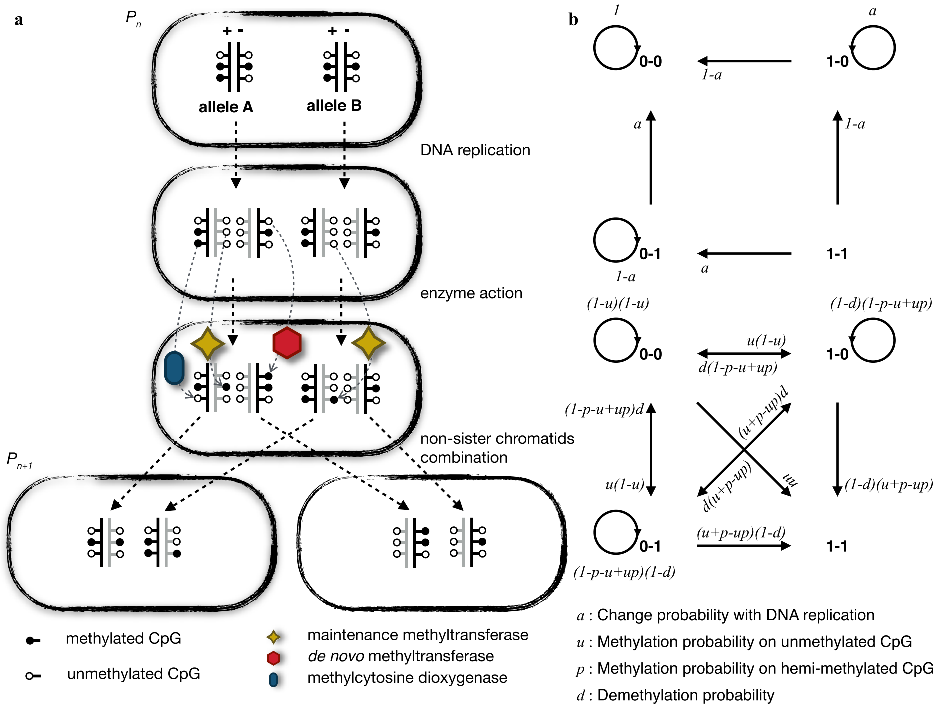
a
P_n
+ -
+ -
allele A
allele B
DNA replication
enzyme action
non-sister chromatids combination
P_{n+1}
methylated CpG
unmethylated CpG
maintenance methyltransferase
de novo methyltransferase
methylcytosine dioxygenase
b
1
0-0
1-a
1-0
a
a
1-a
0-1
a
1-1
1-a
(1-u)(1-u)
0-0
u(1-u)
d(1-p-u+up)
1-0
(1-d)(1-p-u+up)
(1-p-u+up)d
u(1-u)
(u+p-up)d
d(u+p-up)
uu
(1-d)(u+p-up)
0-1
(u+p-up)(1-d)
1-1
(1-p-u+up)(1-d)
a : Change probability with DNA replication
u : Methylation probability on unmethylated CpG
p : Methylation probability on hemi-methylated CpG
d : Demethylation probability

Fig. 1 Computational framework of MethylTransition. **a** Schematic overview of DNA methylation changes during mitosis. MethylTransition relies on the assumption that changes in the DNA methylation state at a CpG site across a single cell cycle occur in three steps: passive demethylation during DNA replication, active DNA methylation by DNA methylation-modifying enzymes, and DNA methylation state combination during the combination of non-sister chromatids. The open circle represents a CpG dinucleotides with an unmethylated C, and the solid circle represents a CpG dinucleotides with a methylated C. In the first step, DNA replication, the newly synthetized strand (gray) is unmethylated. In the second step, enzyme action, the DNA methylation state in a CpG dyad can be changed by different enzymes. In the third step, non-sister chromatid combination, the observed DNA methylation state of a CpG site in a single cell is the combination of the DNA methylation states from both homologous chromosomes for diploid organisms. **b** Transition probabilities among DNA methylation states during DNA replication and enzyme action. We used *1* as a symbol to represent a CpG dinucleotides with a methylated C and *0* to represent a CpG dinucleotides with an unmethylated C. The DNA methylation state for a CpG dyad can be unmethylated (*0-0*), hemi-methylated (*0-1* or *1-0*), or fully methylated (*1-1*). The upper panel shows the transition probabilities among states during DNA replication, where *a* represents the probability of parental DNA being present in one of two daughter strands and is 0.5 for the symmetric division. The lower panel shows the transition probabilities among states during enzyme action, where *u* is the probability of de novo methylation on an unmethylated CpG dyad, *p* is the probability of methylation maintenance on a hemi-methylated CpG dyad, and *d* is the probability of active demethylation on a methylated or hemi-methylated CpG dyad

represent unmethylated CpG dinucleotides. Thus, the DNA methylation state for a CpG dyad can be one of the following types: unmethylated (*0-0*), hemi-methylated (*0-1* or *1-0*), or fully methylated (*1-1*). In the first step, i.e., DNA replication, MethylTransition assumes that the newly synthetized strand is unmethylated. The transition probabilities among DNA methylation states for a CpG dyad are displayed in the upper panel of Fig. 1b (see the “Methods” section for details), where *a* represents the probability of parental DNA in one of two daughter strands and is 0.5 for the symmetric division. In the second step, i.e., enzyme action, the DNA methylation state in a CpG dyad can be changed by different enzymes. According to the distinct roles of DNA methylation-modifying enzymes, MethylTransition denotes the probability of de novo methylation on an unmethylated CpG dyad as *u*, the probability of methylation maintenance on a hemi-methylated CpG dyad as *p*, and the probability of active demethylation on a methylated or hemi-methylated CpG dyad as d. The transition probabilities among DNA methylation states for a CpG dyad are shown in the lower panel of Fig. 1b (see the "Methods" for details). In the third step, i.e., non-sister chromatid combination, the observed DNA methylation state of a CpG site in a single cell is the combination of the DNA methylation states of both homologous chromosomes for diploid organisms. There are five possible methylation states at each CpG site: unmethylated (labeled S1), quarter-methylated (S2), half-methylated (S3), three-quarter-methylated (S4), and fully methylated (S5) (Additional file 1: Fig. S1a; see the "Methods" section for details). Then, MethylTransition generates a matrix to describe the transition probability among the five states in this step (see the "Methods" section for details). Overall, MethylTransition constructs a probability matrix (T_A) with introduced parameters to quantitatively characterize the DNA methylation state transition across a single cell cycle.

Within the framework of MethylTransition, a transition frequency matrix could theoretically be calculated by measuring the DNA methylation state of a CpG site before and after a cell cycle, and then the parameters u, d, and p could be estimated. However, due to the high dropout rate of public scBS-seq data, the percentage of CpG sites both sequenced in two scBS-seq datasets is generally less than 20% (Additional file 1: Fig. S1b), and it is difficult to accurately assign a methylation state to a CpG site based on scBS-seq data. To overcome the above difficulties, we propose two assumptions. First, the DNA methylation states of CpG sites in the promoter of a gene can be represented by the average DNA methylation level in this promoter. This assumption is largely true for the mammalian genome [27], and the various DNA methylation levels in promoters are much smaller than those in other regions (Additional file 1: Fig. S1c). Second, the activity levels of DNA methylation-modifying enzymes in genomic regions with similar chromatin status are largely consistent. Based on these two assumptions, for genes with similar chromatin status at their promoters, each gene's promoter can be assigned one of the five methylation states (S1 to S5) using scBS-seq data, and the methylation state ratio vector of the genes in each methylation state can be used to present the overall pattern of methylation states for those genes within a given cell (see the "Methods" section for details). Given two separate methylation state ratio vectors before and after a cell cycle, the transition frequency matrix T_B linking the two vectors can be calculated. Then, MethylTransition can estimate the parameters u, d, and p by minimizing the difference between the transition probability matrix T_A and the calculated transition frequency matrix T_B, while a cost function is constructed and optimized by using the Newton-Raphson method [28] (see "Methods" section for details). In addition to estimating the transition parameters with given initial and terminal DNA methylation states, MethylTransition can also calculate the methylation state ratio vector after a cell cycle with given initial DNA methylation states and a set of transition parameters. MethylTransition has been implemented as an R library (Additional file 1: Fig. S1d), and the source code has been released on GitHub (https://github.com/TongjiZhanglab/MethylTransition).

Performance evaluation of MethylTransition

To evaluate the performance of MethylTransition, we used three different strategies. First, we tested the robustness of parameter estimation with respect to cell-to-cell variation. We applied MethylTransition to all available high-quality scBS-seq data at different stages of human pre-implantation embryos [12], and the parameters were estimated for each pair of two cells from the adjacent stages. The variances of parameters estimated by using different cells from the same stage were small (Fig. 2a), confirming the robustness of MethylTransition in parameter estimation with respect to cell-to-cell variation.

Second, we examined the robustness of parameter estimation for different dropout rates of scBS-seq data. We applied MethylTransition to a pair of scBS-seq datasets from two adjacent human pre-implantation stages, the zygote stage and the 2-cell stage, with average promoter DNA methylation levels of 16.8% and 15.0%, respectively. When all 22,056 promoters with sequenced CpG sites in both datasets were used to construct the methylation state ratio vectors, the parameters were estimated as $u = 0.03$, $d = 0.23$, and $p = 0.76$. To evaluate the effect of dropout on parameter estimation, we randomly sampled fractions of the 22,056 promoters to represent dropout rates from 10 to 90% (with 10% as the interval), and the subsets of genes were used to construct the

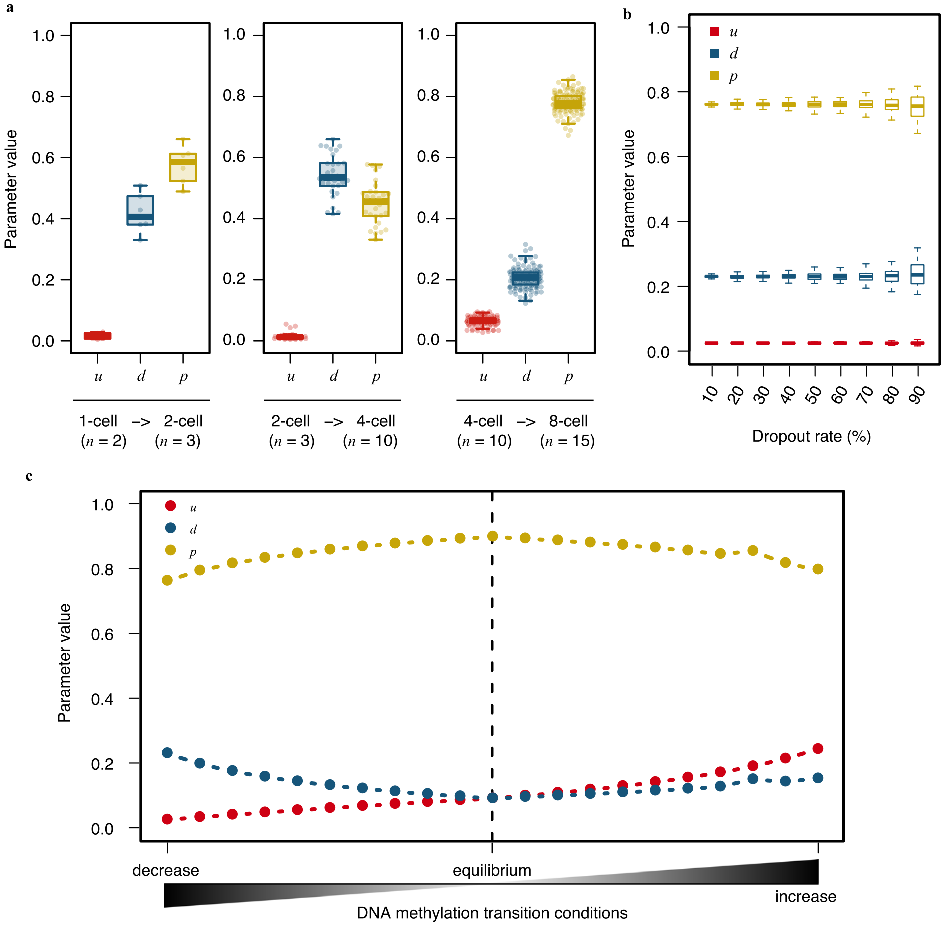

Fig. 2 Performance evaluation of MethylTransition. **a** Box plots showing the estimated parameter values from different cell pairs during human first three cell cycles. Each dot represents the parameter value estimated using a pair of scBS-seq data. All available high-quality published scBS-seq data (more than 20% of all CpG sites detected) were used. **b** Box plots showing the robustness of MethylTransition's parameter estimation for datasets with simulated dropout rates from 10 to 90%. **c** Graph showing the biological meaning of MethylTransition's parameter estimation for a series of simulated DNA methylation transition conditions (gradually changes from decreasing condition to equilibrium condition and then to increasing condition)

methylation state ratio vectors and then to estimate the parameters (see "Methods" section for details). Across the different dropout rates, the estimated parameters were consistent, with slightly increased standard deviations at high dropout rates (Fig. 2b). Our results demonstrate that the parameter estimation method of MethylTransition is robust even with a high dropout rate of scBS-seq data.

Third, we evaluated whether the estimated parameters could represent the levels of corresponding DNA methylation-modifying activities. The dramatic DNA methylation decrease between the human zygote stage and 2-cell stage was characterized by a higher probability of active demethylation ($d = 0.23$) and a lower probability of methylation maintenance ($p = 0.76$) than that in the equilibrium condition ($d = 0.09$ and $p = 0.90$, respectively), consistent with the findings of previous studies [29]. To confirm the biological meanings of the estimated parameters, we gradually altered the methylation state ratio vector at the 2-cell stage to represent a series of DNA methylation transition conditions (from decrease to equilibrium, then to increase), and the altered vectors were used to estimate the parameters (see the "Methods" section for details). Consistent with our expectations, the probability of active demethylation (d) decreased dramatically from DNA methylation decreasing conditions to the equilibrium condition; this decrease in d was coupled with rapid increases in the probability of methylation maintenance (p) and the probability of de novo methylation (u), which clearly increased from the equilibrium condition to increasing conditions (Fig. 2c). To further evaluate the interpretability of parameters, we extended the application of MethylTransition to bulk cell-level BS-seq data with methylation enzymes knockout (see the "Methods" section for details). We collected the BS-seq data in wild-type and two types of enzyme-knockout mouse embryos (*Dnmt1*$^{-/-}$, *Dnmt3a/b*$^{-/-}$) at the E8.5 stage [30]. The *Dnmt1*$^{-/-}$ embryos displayed the smallest p, and the *Dnmt3a/b*$^{-/-}$ embryos showed the smallest u, which are in agreement with the known function of these DNMTs (Additional file 1: Fig. S2a). In addition, we collected the BS-seq data in wild-type and two types of enzyme-knockout zebrafish embryos (*Tet3*$^{-/-}$, *Tet1/2/3*$^{-/-}$) at 24 h post-fertilization (hpf) [31]. Both *Tet3*$^{-/-}$ embryos and *Tet1/2/3*$^{-/-}$ embryos showed smaller d than wild-type embryos, and *Tet1/2/3*$^{-/-}$ embryos showed the smallest d (Additional file 1: Fig. S2b). Our results confirmed that the estimated parameters can reflect the levels of DNA methylation-modifying activities.

Programmed DNA methylation heterogeneity in human pre-implantation embryos

To comprehensively investigate the cause of DNA methylation heterogeneity in early embryogenesis, we reanalyzed scBS-seq data in human pre-implantation embryos. We observed that 31.7~42.6% of the promoters showed DNA methylation state polymorphism among cells within a 4-cell stage embryo, and the percentage increased to 54.4~65.7% for 8-cell stage embryos, confirming that DNA methylation states are quite heterogeneous among cells in pre-implantation embryos [12]. To quantitively measure such heterogeneity, we used public scBS-seq data from a single embryo for each stage and calculated a heterogeneity score for each promoter based on the DNA methylation states from different cells in the same embryo (Additional file 1: Fig. S3; see the "Methods" section for details). When classifying the promoters into five classes based on their DNA methylation states in the zygote stage, we found that the DNA methylation heterogeneity in 4-cell and 8-cell stage embryos was significantly different among classes, while promoters with heavier DNA methylation states in the zygote stage tended to have higher DNA methylation heterogeneity in the 4-cell and 8-cell stages (Fig. 3a, b). Given such an association between the initial methylation states and the DNA methylation heterogeneity at later stages, we hypothesized that the DNA methylation heterogeneity in early embryos might be an intrinsic output of a programmed process following a set of cellular instructions.

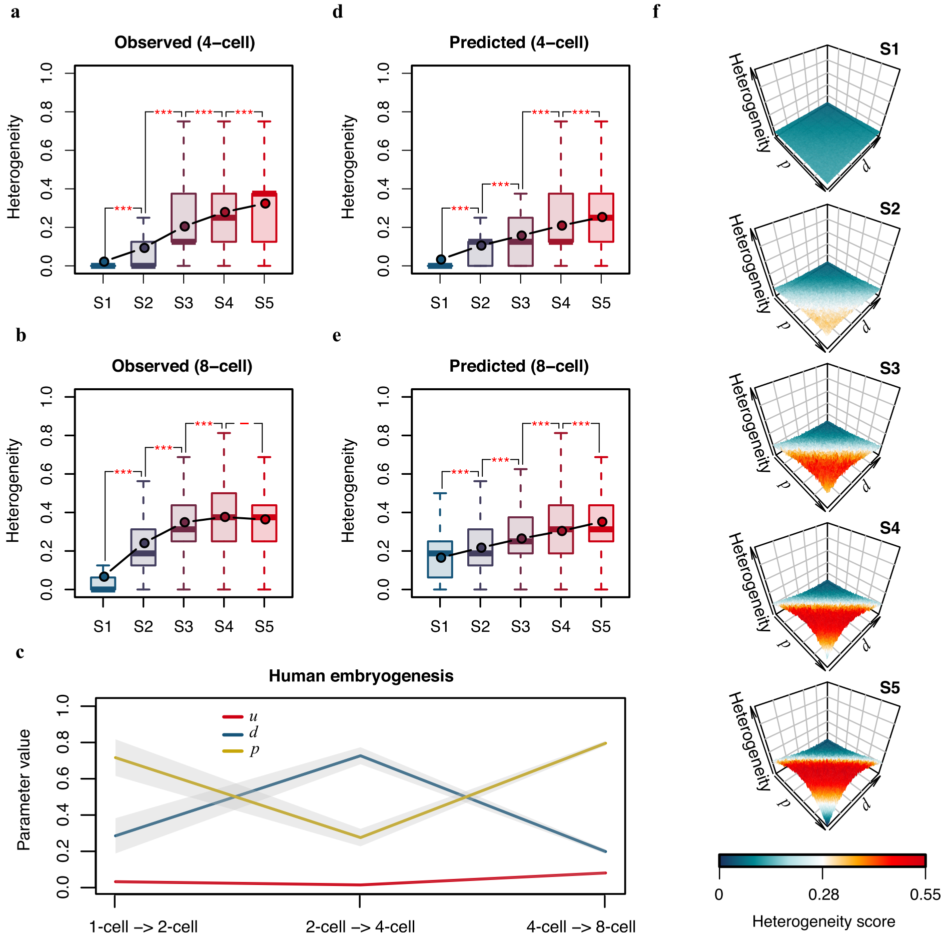

Fig. 3 Programmed DNA methylation heterogeneity in human pre-implantation embryos. **a**, **b** Box plots of the observed DNA methylation heterogeneity of promoters in the 4-cell stage (**a**) and 8-cell stage (**b**) during human embryogenesis. Promoters were classified into 5 classes based on their DNA methylation states in the zygote stage. For each box, the point indicates the average DNA methylation heterogeneity of the class. Student's *t* test was performed for comparisons between adjacent classes (****p* value < 0.001; ***p* value < 0.01; **p* value < 0.05; $^{-}$*p* value > 0.05). The embryo with the largest number of high-quality scBS-seq data (more than 20% of all CpG sites detected) at each stage were used to calculate the DNA methylation heterogeneity. **c** Estimated parameters of the first three cell cycles during human embryogenesis. The lines link the mean values of the parameters in each cell cycle. The gray shaded areas represent the 95% confidence intervals around the mean values. **d**, **e** Box plots of the predicted DNA methylation heterogeneity of promoters in the 4-cell stage (**d**) and 8-cell stage (**e**) during human embryogenesis. Promoters were classified into 5 classes based on their DNA methylation states in the zygote stage. For each box, the point indicates the average DNA methylation heterogeneity of the class. Student's *t* test was performed for comparisons between adjacent classes (****p* value < 0.001). **f** 3D plots of predicted DNA methylation heterogeneity scores at the 8-cell stage for simulated values of *p* and *d* for five classes of promoters. Here, the parameter *u* is a constant (*u* = 0.03). The colors represent the predicted heterogeneity scores

To test the above hypothesis, we applied MethylTransition to the first three cell cycles of human embryogenesis using scBS-seq data from a single embryo for each stage. To simplify the calculation, we used all genes' promoters to estimate the parameters. For each combination of cell pairs between embryos in adjacent stages, the parameters u, d, and p were estimated, with substantial robustness among combinations (Fig. 3c). The highest value of the parameter d, i.e., the probability of active demethylation, occurred during the transition between the 2-cell and 4-cell stages, while the highest value of the parameter p, i.e., the probability of methylation maintenance, occurred during the transition between the 4-cell and 8-cell stages. The parameter u, i.e., the probability of de novo methylation, was consistently low during this process. Given the initial methylation states of the promoters at the zygote stage and the estimated parameters of the first three cell cycles, MethylTransition was applied to predict the DNA methylation heterogeneity for each promoter in the 4-cell and 8-cell stages, respectively (Fig. 3d, e; see the "Methods" section for details). The promoters with heavier DNA methylation states in the zygote stage clearly displayed a higher predicted DNA methylation heterogeneity in the 4-cell and 8-cell stages, consistent with the trends in observed heterogeneity among classes with distinct initial methylation states. Such predictable heterogeneity demonstrates that the DNA methylation heterogeneity in early embryos is largely an intrinsic output of a programmed process with given initial methylation states and a few DNA methylation state transition parameters.

To investigate the impacts of initial methylation states and DNA methylation state transition parameters on this programmed heterogeneity, we calculated the predicted scores of DNA methylation heterogeneity in the 8-cell stage with simulated values of MethylTransition parameters for five classes of promoters with distinct initial methylation states (Fig. 3f). To simplify the simulation, we kept $u = 0.03$, which was supported by the absence of nuclear DNMT3a and DNMT3b during this process [32], and used the same values of p or d for all three cell cycles. We observed that for any given class, the predicted scores of DNA methylation heterogeneity varied dramatically for different combinations of p and d, while with a given set of p and d values, the predicted scores tended to be higher in classes with heavier initial methylation states. In a typical mammalian cell, most promoters are either unmethylated (S1) or fully methylated (S5); in addition, during a typical mammalian cell cycle, the probability of methylation maintenance (p) is very high, while the probability of active demethylation (d) is quite low [33, 34]. From the simulation, we observed low predicted heterogeneity (score < 0.2) in the S1 and S5 classes for combinations of high p (> 0.94) and low d (< 0.05) values, indicating that DNA methylation heterogeneity is unlikely to be introduced during the cell cycle in a typical mammalian cell. Unlike typical mammalian cells, the cells in early embryos display clearly higher d and lower p values, which can intrinsically cause initial methylation state depended on DNA methylation heterogeneity during early embryogenesis. Taken together, our results indicate that the largely programmed DNA methylation heterogeneity in early embryos is a natural result of the unique probabilities of methylation maintenance and active demethylation that occur during this process.

Association of sequence and epigenetic features with DNA methylation heterogeneity

Although the DNA methylation heterogeneity in early embryos was largely predictable by MethylTransition with a few parameters, some promoters showed substantially higher or lower scores of observed methylation heterogeneity than predicted, especially promoters in the S1 and S2 classes (Fig. 4a, Additional file 1: Fig. S4a), indicating that additional features may have significant impacts on the DNA methylation heterogeneity of promoters in these classes. In order to investigate which sequence or epigenetic

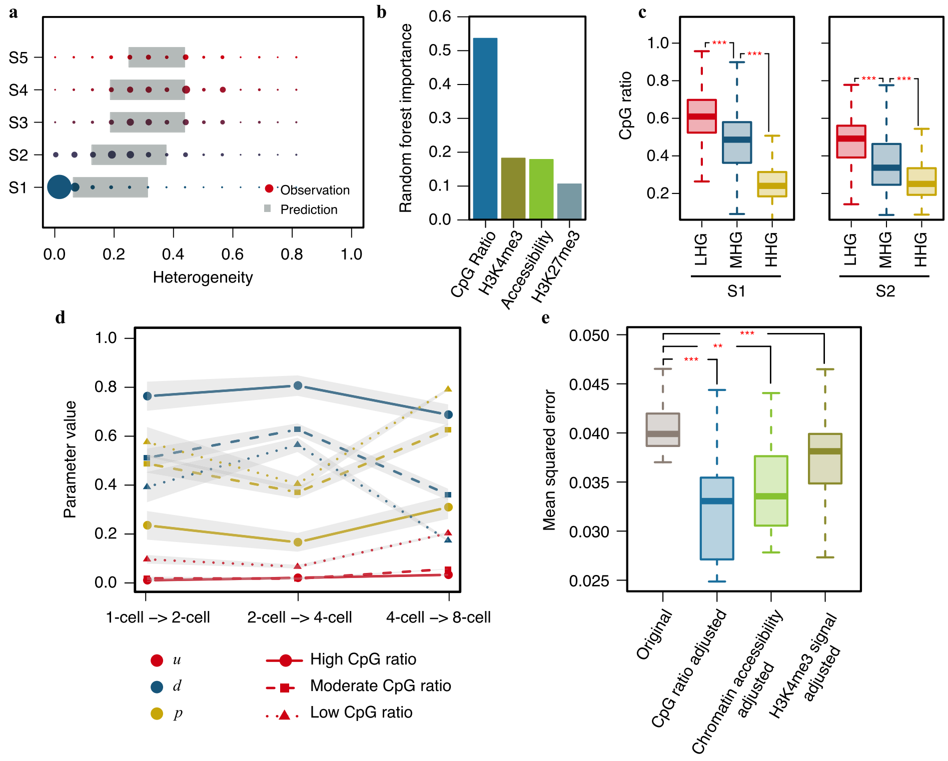

Fig. 4 Impacts of sequence and epigenetic features on DNA methylation heterogeneity. **a** Differences between the observed and predicted DNA methylation heterogeneity scores in human 8-cell stage embryos. Promoters were classified into 5 classes based on their DNA methylation states in the zygote stage. The size of each solid circle represents the number of promoters. Each gray box shows the 1st and 3rd quartiles of the predicted heterogeneity scores in each class. **b** Bar plot of the importance of sequence and epigenetic features for DNA methylation heterogeneity at the 8-cell stage, as determined by random forest analysis. The features include CpG ratio, chromatin accessibility, and H3K4me3 and H3K27me3 levels at the 8-cell stage. **c** Box plots of the differences in CpG ratio between different categories of promoters in the S1 and S2 classes. The promoters in the S1 and S2 classes were divided into three categories: higher-heterogeneity gene promoters (HHGs; with observed heterogeneity higher than the 3rd quantile of the predicted heterogeneity score of the class), model-predictable heterogeneity gene promoters (MHGs), and lower-heterogeneity gene promoters (LHGs; with observed heterogeneity lower than the 1st quantile of the predicted heterogeneity score of the class). Student's *t* test was performed for comparisons between adjacent categories (****p* value < 0.001). **d** Estimated parameters for different groups of promoters with distinct CpG ratios during the first three cell cycles of human embryogenesis. The promoters were grouped into three groups with high (≥ 0.6), moderate (between 0.4 and 0.6), or low (< 0.4) CpG ratios. The lines link the mean values of the parameters for each cell cycle. The gray shaded areas represent the 95% confidence intervals around the mean values. Red indicates the parameter *u*, blue indicates the parameter *d*, and yellow indicates the parameter *p*. **e** Box plots of the differences in the mean squared error between the original prediction obtained using uniform sets of *u*, *d*, and *p* for all promoters and three conditional predictions based separately on CpG ratio grouping, chromatin accessibility grouping, and H3K4me3 signal grouping. The mean squared error was calculated as the average squared difference between the predicted methylation heterogeneity score and the observed score. Student's *t* test was performed for comparisons between the original prediction and each of the conditional predictions (****p* value < 0.001; ***p* value < 0.01)

features may strongly associate with DNA methylation heterogeneity, we performed random forest analysis to identify relevant features from sequence patterns (CpG ratio), chromatin accessibility, and histone modification patterns (H3K4me3, H3K27me3) in human pre-implantation embryos (see the "Methods" section for details). Among these features, the CpG ratio showed much greater relevance to promoter DNA methylation heterogeneity than other features. Both chromatin accessibility and H3K4me3 signals showed moderate relevance (Fig. 4b). For the S1 and S2 classes, the promoters with lower heterogeneity scores than predicted showed significantly higher CpG ratio, chromatin accessibility, and H3K4me3 signals than those with heterogeneity scores similar to the predicted scores (Fig. 4c, Additional file 1: Fig. S4b, c), suggesting that the limited performance of MethylTransition for the S1 and S2 classes is likely due to these sequence and epigenetic features. To examine whether DNA methylation state transition parameters are variable among promoters with distinct CpG ratios, we divided promoters into three groups with high (≥ 0.6), moderate (between 0.4 and 0.6), or low (< 0.4) CpG ratio and applied MethylTransition to the first three cell cycles of human embryogenesis based on the three groups of promoters separately. The sets of the estimated parameters *u*, *d*, and *p* were different among groups of promoters with distinct CpG ratios (Fig. 4d). The group of promoters with a high CpG ratio showed lower values of *u* and *p*, together with higher values of *d*, than the other two groups, consistent with the findings of previous studies that CpG islands tend to be protected from DNA methylation [35]. We further predicted the DNA methylation heterogeneity for each promoter in the 8-cell stage using parameter values estimated based on CpG ratio grouping, and the predicted heterogeneity showed a significantly smaller mean squared error to the observed than the original prediction obtained using uniform *u*, *d*, and *p* values for all promoters (Fig. 4e). In addition to assessing DNA methylation state transition parameters among promoters with distinct CpG ratio, we also calculated the parameters among promoters with distinct chromatin accessibility or distinct H3K4me3 signals (Additional file 1: Fig. S4d, e). The heterogeneity predicted using parameters estimated based on chromatin accessibility or H3K4me3 signal grouping also showed a significantly smaller mean squared error to the observed than the original prediction (Fig. 4e). Taken together, our results demonstrate that the features of CpG ratio, chromatin accessibility, and distinct H3K4me3 signals are strongly associated with DNA methylation heterogeneity by diverse DNA methylation state transition parameters.

Potential regulatory effects of DNA methylation heterogeneity

As DNA methylation at promoters plays important roles in gene transcriptional regulation, we hypothesized that the largely programmed DNA methylation heterogeneity at promoters can affect the heterogeneity of gene expression. To test this hypothesis, we first calculated the correlation between promoter DNA methylation heterogeneity scores and expression heterogeneity scores in human 8-cell stage embryos (see the "Methods" section for details), and the two types of heterogeneity scores indeed showed a positive correlation (cor = 0.18; Fig. 5a). Next, we investigated the temporal association between promoter DNA methylation heterogeneity and expression heterogeneity. The genes with high DNA methylation heterogeneity promoters at the 4-cell stage showed significantly larger expression heterogeneity increasement from that stage to

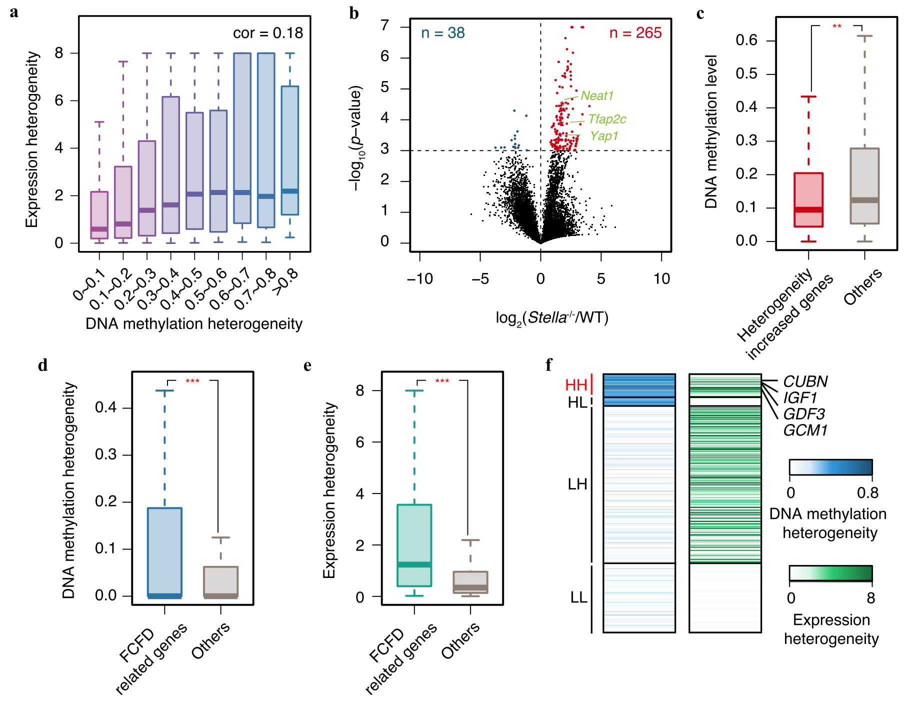

Fig. 5 Potential regulatory effects of DNA methylation heterogeneity. **a** Box plots of the relationship between DNA methylation heterogeneity and expression heterogeneity in human 8-cell stage embryos. The Pearson correlation coefficient of the two heterogeneity scores was 0.18. **b** Volcano plot comparing the gene expression heterogeneity between mouse *Stella*$^{-/-}$ embryos and mouse wild-type (WT) embryos at the 8-cell stage. The *x*-axis represents the $\log_2$-transformed fold change, and the *y*-axis represents the $\log_{10}$-transformed *p* value. A total of 265 genes had significantly higher expression heterogeneity scores in *Stella*$^{-/-}$ embryos than in WT embryos, while 38 genes had significantly lower expression heterogeneity scores in *Stella*$^{-/-}$ embryos than in WT embryos. Three genes reported to be related to cell fate determination during mouse embryogenesis are indicated in green. **c** Box plots comparing the DNA methylation levels in mouse zygotes between genes with increased expression heterogeneity and other genes. Student's *t* test was performed (**p value < 0.01). **d**, **e** Box plots of the DNA methylation heterogeneity (**d**) and expression heterogeneity (**e**) in human 8-cell stage embryos between first cell fate determination (FCFD)-related genes and other genes. Student's *t* test was performed for comparisons (***p value < 0.001). **f** Heatmap of the DNA methylation heterogeneity scores (blue) and expression heterogeneity scores (green) for FCFD-related genes. The FCFD-related genes were classified into four groups based on their DNA methylation and expression heterogeneity scores at the 8-cell stage (HH: DNA methylation heterogeneity > 0.3 and expression heterogeneity > 0.5; HL: DNA methylation heterogeneity > 0.3 and expression heterogeneity ≤ 0.5; LH: DNA methylation heterogeneity ≤ 0.3 and expression heterogeneity > 0.5; and LL: DNA methylation heterogeneity ≤ 0.3 and expression heterogeneity ≤ 0.5). Within the HH group, three reported ICM markers (*GDF3*, *CUBN*, and *IGF1*) and a TE marker (*GCM1*) are indicated

the 8-cell stage than other genes (Additional file 1: Fig. S5a). Differently, the genes with high expression heterogeneity at the 4-cell stage showed similar promoter DNA methylation heterogeneity increasement from that stage to the 8-cell stage with other genes (Additional file 1: Fig. S5b), which implied that DNA methylation heterogeneity at promoters might affect the heterogeneity of the gene expression. Due to restrictions preventing the editing of human embryos, we then used mouse embryos to further clarify the effects of DNA methylation heterogeneity on expression heterogeneity. Similar to the case in human embryos, the DNA methylation heterogeneity that emerged during mouse early embryogenesis was also largely determined by the initial DNA methylation state at the zygote stage (Additional file 1: Fig. S6). In this study, we generated single-cell RNA-seq data for cells in the 8-cell stage from *Stella*$^{-/-}$ mouse embryos, which have been reported to have much higher DNA methylation levels than normal embryos [36]. Upon comparing the expression heterogeneity scores of *Stella*$^{-/-}$ mouse embryos to those of normal embryos, we identified 6.97-fold more genes with significantly elevated heterogeneity scores in *Stella*$^{-/-}$ embryos than genes with reduced scores including genes encoding reported lineage-determining factors of ICM and TE (*Neat1* [9], *Tfap2c* [37], *Yap1* [38, 39]) (Fig. 5b). The genes with elevated expression heterogeneity scores in *Stella*$^{-/-}$ embryos showed significantly lower initial DNA methylation levels at promoters in the zygote stage in normal mouse embryos than other genes (Fig. 5c), indicating that increases in initial DNA methylation levels can increase the expression heterogeneity in later stages, especially for genes with initially minimally methylated promoters. Considering that promoters with heavier DNA methylation in the zygote stage tended to have higher DNA methylation heterogeneity in later stages, the increased expression heterogeneity in *Stella*$^{-/-}$ embryos might be mediated by the largely programmed DNA methylation heterogeneity at promoters.

We further hypothesized that the largely programmed DNA methylation heterogeneity at promoters might contribute to the first cell fate determination. To test this hypothesis, among 7471 expressed genes in ICM or TE, we defined 899 ICM- or TE-specific genes as the list of genes related to the first cell fate determination in humans (see the "Methods" section for details). At the 8-cell stage, that list of genes showed significantly higher DNA methylation heterogeneity at promoters, together with higher expression heterogeneity, than other expressed genes (Fig. 5d, e), indicating that the separation between the ICM and TE could at least partially be explained by the emerged expression heterogeneity and by largely programmed DNA methylation heterogeneity at promoters. Among the list of genes related to the first cell fate determination, 64 genes showed both high DNA methylation heterogeneity at promoters and expression heterogeneity at the 8-cell stage (Fig. 5f and Additional file 2: Table S1), including the known ICM marker genes *GDF3* [40], *CUBN* [41], and *IGF1* [42] and the known TE marker gene *GCM1* [43]. The clear association between DNA methylation heterogeneity and the first cell fate determination suggests that consideration of DNA methylation heterogeneity may provide clues for identifying driving factors of driving cell fate determination during human pre-implantation embryogenesis.

Discussion

Although expression and epigenetic heterogeneity occurring before the first cell fate determination in mammalian early embryogenesis has been suggested to be a determinate result of a program [10–12] (i.e., a process with initial states and a set of cellular instructions), to the best of our knowledge, no studies have been designed to illustrate such a program. In this study, we took advantage of the quantitative features of scBS-seq data to elucidate the set of instructions causing DNA methylation heterogeneity during early embryogenesis by building MethylTransition, a mathematical model that enables characterization of the DNA methylation state transition during single or a few cell cycles. The results of this study clearly demonstrate that DNA methylation heterogeneity at promoters in early mammalian embryos is largely an intrinsic output of a programmed process with given initial methylation states and a few DNA methylation state transition parameters. It should be clarified that the cell cycle itself does not naturally result in DNA methylation heterogeneity. Calculations based on MethylTransition showed that DNA methylation heterogeneity is unlikely to be introduced during the cell cycle in typical adult mammalian cells, which have different initial methylation state distributions and DNA methylation state transition parameters than cells in early mammalian embryos. This study illustrates that the largely programmed DNA methylation heterogeneity in early mammalian embryos is a determinate result of the unique probabilities of methylation maintenance and active demethylation that occur during early embryogenesis.

The process of the first cell fate determination in early mammalian embryos is associated with distinct expression and epigenetic patterns between ICM and TE cells [6, 7]. This study showed a clear association between DNA methylation heterogeneity and the first cell fate determination, and the results suggest that the expression heterogeneity that emerges before the first cell fate determination could be at least partially caused by largely programmed DNA methylation heterogeneity. In addition to DNA methylation, multiple layers of factors participate in transcriptional regulation. A previous model study based on mouse embryogenesis data proposed that the first cell fate determination-related expression heterogeneity is initiated by random segregation at each cleavage division and strengthened by competing transcriptional circuits [5]. In human embryos, the major phase of zygotic genome activation begins at the 8-cell stage, much later than in mouse embryos (in which it occurs at the 2-cell stage) [44], indicating that humans have a larger time window for the accumulation of DNA methylation heterogeneity before the large-scale transcription begins. As a supplement to that model study, this study suggests that the initial source of expression heterogeneity in human embryos could be at least partially mediated by the largely programmed DNA methylation heterogeneity at promoters. We suspect that consideration of DNA methylation heterogeneity could provide valuable clues for identifying driving factors of the first cell fate determination during human pre-implantation embryogenesis.

Unlike other mathematical models, MethylTransition was designed to characterize the methylation state changes that occur during one or a few cell cycles at single-cell resolution, and its application could be extended to other biological processes with dramatic DNA methylation changes, such as oogenesis [45]. Despite its unique features, MethylTransition has two technical limitations. First, due to the high dropout rates of public scBS-seq data, MethylTransition relies on the assumption that the activity levels of DNA methylation-modifying enzymes in a cell are constant across the genome, which is inconsistent with numerous reports that both sequence and epigenetic features affect the activity of these enzymes [25, 46–49]. This limitation could be partially resolved by applying this framework to a subset of promoters with similar sequence or epigenetic features. Second, MethylTransition assumes that de novo methylation, methylation maintenance, and active demethylation are three independent events. However, this may not always be true, as crosstalk between de novo methylation and methylation maintenance has been reported [50–52]. Those studies indicated that the functions of the DNA methylation-modifying enzymes are complex and not independent. Furthermore, some factors, for example, Stella, have distinct roles in the modulation of DNA methylation at different mouse development stages [36, 53]. Although the parameters of MethylTransition integrated the function of various enzymes to represent the overall levels of DNA methylation-modifying activities, incorporating the crosstalk of those activities may further improve the interpretability of this model. Future versions of MethylTransition could be extended to consider the conditional probabilities among the three DNA methylation-modifying activities.

Conclusions

To quantitatively characterize the DNA methylation state transition during mammalian early embryogenesis, we presented MethylTransition, a novel DNA probabilistic model for characterizing methylation changes during one or a few cell cycles at single-cell resolution. By applying MethylTransition to scBS-seq data during human pre-implantation embryogenesis, we elucidated that the DNA methylation heterogeneity that emerges at promoters is a determinate result of the unique probabilities of methylation maintenance and active demethylation that occur during early embryogenesis. This study further suggested that consideration of programmed DNA methylation heterogeneity provides valuable clues for identifying driving factors of the first cell fate determination.

Methods

Computational framework of MethylTransition

We used *1* as a symbol to represent methylated CpG dinucleotides and *0* to represent unmethylated CpG dinucleotides. Thus, the DNA methylation state for a CpG dyad in one chromosome could be one of the following types: unmethylated (*0-0*), hemi-methylated (*0-1* or *1-0*), or fully methylated (*1-1*). The calculated ratios of CpG dyads in these four methylation states are x_1, x_2, x_3, and x_4, and the ratio vector M_s is:

$$M_s = \begin{pmatrix} x_1 \\ x_2 \\ x_3 \\ x_4 \end{pmatrix} \tag{1}$$

MethylTransition relies on the assumption that changes in the DNA methylation state at a CpG site across a single cell cycle occur in three steps. In the first step, subsequent to the passive demethylation of DNA with DNA replication in the cell, the ratio vector M_s changes to M_r, which can be represented as:

$$T_{sr} = \begin{pmatrix} 1 & a & (1-a) & 0 \\ 0 & (1-a) & 0 & (1-a) \\ 0 & 0 & a & a \\ 0 & 0 & 0 & 0 \end{pmatrix} \tag{2}$$

$$M_r = T_{sr} \cdot M_s \tag{3}$$

where a represents the probability of parental DNA in one of two daughter strands and is 0.5 for the symmetric division. M_r is the ratio vector after DNA replication.

In the second step, the DNA methylation state in a CpG dyad can be altered by DNMTs or by DNA demethylases. We denoted the probability of maintaining methylation at a hemi-methylated CpG site as p, the probability of de novo methylation at a CpG site as u, and the probability of active demethylation at a methylated CpG site as d. The change in DNA methylation under the actions of various enzymes can be expressed as:

$$T_{re} = \begin{pmatrix} (1-u)\times(1-u) & (1-u)\times(1-p)\times d & d\times(1-u)\times(1-p) & d\times d \\ u\times(1-u) & (1-p)\times(1-u)\times(1-d) & d\times(u+p-u\times p) & (1-d)\times d \\ (1-u)\times u & (u+p-u\times p)\times d & (1-d)\times(1-u)\times(1-p) & d\times(1-d) \\ u\times u & (u+p-u\times p)\times(1-d) & (1-d)\times(u+p-u\times p) & (1-d)\times(1-d) \end{pmatrix} \tag{4}$$

$$M_e = T_{re}\cdot M_r \tag{5}$$

where M_e represents the ratio vector after actions have been exerted by various enzymes.

In the third step, for diploid organisms, the observed DNA methylation state of a CpG site in a single cell is the combination of the DNA methylation states from both homologous chromosomes. Such combination results in five possible DNA methylation states at each CpG site: unmethylated (labeled S1), quarter-methylated (S2), half-methylated (S3), three-quarter-methylated (S4), and fully methylated (S5). These states can be experimentally measured, and the ratios of CpG sites in these five methylation states (S_1, S_2, S_3, S_4, and S_5) can be calculated. Thus, the DNA methylation state changes from mother cell to daughter cell are represented by:

$$T_A = \begin{pmatrix} x_{11} & x_{12} & x_{13} & x_{14} & x_{15} \\ x_{21} & x_{22} & x_{23} & x_{24} & x_{25} \\ x_{31} & x_{32} & x_{33} & x_{34} & x_{35} \\ x_{41} & x_{42} & x_{43} & x_{44} & x_{45} \\ x_{51} & x_{52} & x_{53} & x_{54} & x_{55} \end{pmatrix} \tag{6}$$

$$\begin{pmatrix} {S_1}' \\ {S_2}' \\ {S_3}' \\ {S_4}' \\ S_5 \end{pmatrix} = T_A\cdot \begin{pmatrix} S_1 \\ S_2 \\ S_3 \\ S_4 \\ S_5 \end{pmatrix} \tag{7}$$

where ${S_1}'$, ${S_2}'$, ${S_3}'$, ${S_4}'$, and ${S_5}'$ represent the ratios of CpG sites in the five DNA methylation states in a daughter cell, and x_{ij} indicates the probability that the DNA methylation states are changed from state j in the mother cell to state i in the daughter cell. The element in the ith row and the jth column of the probability matrix $T_{re}\cdot T_{sr}$ described above is denoted as t_{ij}. The relationships between x_{ij} and t_{ij} are calculated as follows:

$$\begin{gathered}
x_{11} = t_{11} \times t_{11} \\
x_{12} = \frac{1}{4} \times (t_{11} \times t_{12} + t_{11} \times t_{13} + t_{12} \times t_{11} + t_{13} \times t_{11}) \\
x_{13} = \frac{1}{6} \times (t_{11} \times t_{14} + t_{12} \times t_{12} + t_{12} \times t_{13} + t_{13} \times t_{12} + t_{13} \times t_{13} + t_{14} \times t_{11}) \\
x_{14} = \frac{1}{4} \times (t_{12} \times t_{14} + t_{13} \times t_{14} + t_{14} \times t_{12} + t_{14} \times t_{13}) \\
x_{15} = t_{14} \times t_{14} \\
\ldots\ldots \\
x_{51} = t_{41} \times t_{41} \\
x_{52} = \frac{1}{4} \times (t_{41} \times t_{42} + t_{41} \times t_{43} + t_{42} \times t_{41} + t_{43} \times t_{41}) \\
x_{53} = \frac{1}{6} \times (t_{41} \times t_{44} + t_{42} \times t_{42} + t_{42} \times t_{43} + t_{43} \times t_{42} + t_{43} \times t_{43} + t_{44} \times t_{41}) \\
x_{54} = \frac{1}{4} \times (t_{42} \times t_{44} + t_{43} \times t_{44} + t_{44} \times t_{42} + t_{44} \times t_{43}) \\
x_{55} = t_{44} \times t_{44}
\end{gathered} \tag{8}$$

Based on the above three steps, the changes in DNA methylation states during one cell cycle can be represented by the above probability matrix T_A with a few parameters. Then scBS-seq data are used to infer the ratio of each DNA methylation state within a single cell. To imitate the DNA methylation state transition before and after the cell cycle, the state transition frequency matrix (T_B) is calculated using scBS-seq data of a cell from the early stage (mother cell) and a cell from the adjacent late stage (daughter cell) as follows:

$$T_B = \begin{pmatrix} o_{11} & o_{12} & o_{13} & o_{14} & o_{15} \\ o_{21} & o_{22} & o_{23} & o_{24} & o_{25} \\ o_{31} & o_{32} & o_{33} & o_{34} & o_{35} \\ o_{41} & o_{42} & o_{43} & o_{44} & o_{45} \\ o_{51} & o_{52} & o_{53} & o_{54} & o_{55} \end{pmatrix} \tag{9}$$

where o_{ij} is the observed frequency at which the DNA methylation state is changed from state j in the mother cell to state i in the daughter cell.

MethylTransition estimates the parameters u, d, and p by minimizing the difference between the transition probability matrix T_A and the calculated transition frequency matrix T_B. We defined a cost function as the square of the Euclidean distance between T_A and T_B [28]:

$$J(u, d, p) = \sum_{i=1}^{5} \sum_{j=1}^{5} \left(o_{i,j} - x_{i,j}\right)^2 \tag{10}$$

The best parameter was obtained by minimizing $J(u, d, p)$. Although it is unrealistic to expect an algorithm to find global minima, many numerical optimization techniques can be applied to find local minima. According to the definition of probability, the parameters u, d, and p range from 0 to 1. Considering the reliability and efficiency of the algorithm, we used the Newton-Raphson method as the optimization algorithm.

$$\left(\hat{u}, \hat{d}, \hat{p}\right) = \mathrm{argmin}_{u,d,p} J(u, d, p) \tag{11}$$

As the estimates of parameters with different initial values in the Newton-Raphson method are distributed around the true value, MethylTransition randomly selects three numbers in the range of 0 to 1 as the initial values of u, d, and p to estimate the parameters and repeats the analysis 100 times. The parameters that minimize the cost function value are prioritized.

Calculation of the DNA methylation state ratio vector

We assigned a DNA methylation state S to each gene promoter region (± 2 kb around the transcription start site (TSS)) for a given cell, as follows:

$$S = \begin{cases} \text{S1, if } m \leq 0.125 \\ \text{S2, if } 0.125 \leq m < 0.375 \\ \text{S3, if } 0.375 \leq m < 0.625 \\ \text{S4, if } 0.625 \leq m < 0.875 \\ \text{S5, if } m \geq 0.875 \\ \text{NA, if } m \text{ is not available} \end{cases} \tag{12}$$

where m is the average methylation level of the promoter. To avoid the missing value problem of scBS-seq data, the promoters with state as NA are removed from the parameter estimation. The DNA methylation state ratio vector contains 5 items representing the fractions of the promoters assigned as S1, S2, S3, S4, and S5, and the vector is used to present the overall pattern of methylation states for a given cell.

Performance evaluation for the robustness of subsampling

We calculated the DNA methylation state vectors of two adjacent human preimplantation stages based on two scBS-seq datasets (GSM2481558 and GSM2481533). To evaluate the effects of dropout on parameter estimation, we used a bootstrapping approach. Briefly, we randomly sampled fractions of 22,056 promoters to represent the dropout rates from 10 to 90% (with 10% as the interval) with 100 rounds of replacements for each dropout rate. In each sampling, the selected promoters were used to construct a methylation state ratio vector and a transition frequency matrix, which were used to estimate a set of parameters.

Performance evaluation for biological meanings

We used the original transition frequency matrix calculated based on the two scBS-seq datasets (GSM2481558 and GSM2481533) as the decreased methylation transition condition, and we used an identity matrix to represent the equilibrium condition. To simulate a gradual change in transition conditions from decreasing to equilibrium, we generated 9 transition frequency matrixes to represent the linear switch from the original transition frequency matrix to an identical matrix with equal intervals. We further transposed and scaled the original transition frequency matrix as the increased transition condition. To simulate a gradual change in transition conditions from equilibrium to increasing, we generated 9 transition frequency matrixes to represent the linear switch from the identical matrix to the transposed and scaled matrix with equal intervals. We estimated the sets of parameters based on the original transition frequency matrix, the identical matrix, the transposed and scaled matrix, and 18 generated transition frequency matrixes.

Extension of MethylTransition on bulk-cell BS-seq data

Each high coverage CpG site (coverage > 4) in a given bulk cell BS-seq data was assigned a DNA methylation state S, as follows:

$$S = \begin{cases} \text{S1, if } m_b \leq 0.125 \\ \text{S2, if } 0.125 \leq m_b < 0.375 \\ \text{S3, if } 0.375 \leq m_b < 0.625 \\ \text{S4, if } 0.625 \leq m_b < 0.875 \\ \text{S5, if } m_b \geq 0.875 \\ \text{NA, if } m_b \text{ is not available} \end{cases} \tag{13}$$

where m_b is the methylation level of the CpG site. The sites with state as NA are removed from the parameter estimation. The transition frequency matrix was calculated using two paired BS-seq data. We assigned an approximate number of cell cycles based on the knowledge of development timing ($n = 5$ for 8-cell to E8.5 mouse embryo and $n = 10$ for epiboly to 24 hpf zebrafish embryo). The logarithm transformation of the transition frequency matrix was used to estimate the parameters during multiple cell cycles, with the assumption that the parameters of each cell cycle are consistent.

Calculation of DNA methylation heterogeneity and expression heterogeneity

The DNA methylation heterogeneity of a given promoter in a pre-implantation embryo represents the promoter's polymorphism of methylation states across cells in the same embryo. For each promoter, we calculated the proportions of cells belonging to each of the five methylation states. Then, we calculated the inequality of the methylation state distribution combined with the Gini index to quantitively measure the heterogeneity (Additional file 1: Fig. S3). If x_i is the proportion of the methylation state i, and there are n states ($n = 5$), then the DNA methylation heterogeneity H_{methyl} is given by:

$$H_{\text{methyl}} = 1 - \frac{\sum_{i=1}^{n}\sum_{j=1}^{n}\left|x_i - x_j\right|}{2n\sum_{i=1}^{n}x_i} \tag{14}$$

The expression heterogeneity of a given gene in a pre-implantation embryo represents the polymorphism of gene expression across cells. For each gene, we calculated the expression heterogeneity H_{expr} as the squared coefficient of variation (CV^2) of the expression level across the cells in the same embryo.

Prediction of DNA methylation heterogeneity

With given parameters (u, d, p) and an initial DNA methylation state ratio vector (M_i), the DNA methylation transition matrix ($T_{(u,d,p)}$) and the terminal DNA methylation state ratio vector (M_t) can be calculated by the equations mentioned in the "Computational framework of MethylTransition" section. For each promoter, the probability that its DNA methylation state at the zygote stage will change to any of the five states in the 2-cell stage can be calculated. Then, a random sampling method (based on the R function *sample()*) was used to assign each promoter a state in each of the 2-cell stage cells based on the calculated probability and the initial DNA methylation state in the zygote stage cell. For different cells in the same stage, different random seeds were used. The procedures for the prediction of the state transition from the 2-cell stage to the 4-cell stage and those for prediction of the state transition from the 4-cell stage to the 8-cell stage were the same. For each promoter, the DNA methylation heterogeneity H_{methyl} was calculated as mentioned above with the predicted DNA methylation states for each cell in 4-cell or 8-cell stage embryos.

Identification of relevant features of DNA methylation heterogeneity

The CpG ratio was calculated for each promoter, as previously described [54]. The chromatin accessibility on each promoter was also calculated using a published assay for transposase-accessible chromatin using sequencing (ATAC-seq) data for human 8-cell stage embryos [55]. The H3K4me3 and H3K27me3 signals on each promoter were derived from published ChIP-seq data from human 8-cell stage embryos [56]. Based on the differences between the observed and predicted DNA methylation heterogeneity scores for the gene promoters, the promoters in classes S1 and S2 were divided into three categories: higher-heterogeneity gene promoters (HHGs; with observed heterogeneity higher than the 3rd quantile of the predicted heterogeneity score of the class), model-predictable heterogeneity gene promoters (MHGs), and lower-heterogeneity gene promoters (LHGs; with observed heterogeneity lower than the 1st quantile of the predicted heterogeneity score of the class). With the values for the above sequence and epigenetic features and the category labels, we performed random forest analysis using the *RandomForestRegressor* function in the Python package *sklearn* [57] to evaluate the importance of each feature to the DNA methylation heterogeneity.

Identification of human first cell fate determination-related genes

We first excluded genes that were not expressed in blastocysts (genes with FPKM values less than 1 in ICM and TE cells). We then identified differentially expressed genes between ICM and TE cells using DESeq2 with a threshold of p value < 0.05 and fold change > 4. These differentially expressed genes were considered as the genes related to human first cell fate determination.

Single-cell RNA-seq library generation and sequencing of *Stella*$^{-/-}$ mouse embryos

Stella$^{-/-}$ mice have been described previously [36]. RNA was isolated from each single cell of two 8-cell stage *Stella*$^{-/-}$ embryos with the Smart-seq2 protocol. To generate RNA sequencing libraries, a KAPA Hyper Prep Kit was used following the manufacturer's instructions. The adapters used were from a KAPA Single-Indexed Adapter Kit. Paired-end 150-bp sequencing was further performed on a HiSeq X Ten platform (Illumina) at Novogene. The animal experimental procedures were performed according to the Tongji University Guide for the use of laboratory animals.

Supplementary Information

Supplementary information accompanies this paper at https://doi.org/10.1186/s13059-020-02189-8.

Additional file 1: Fig. S1. Computational framework of MethylTransition. **Fig. S2.** Estimated parameters in DNA methylation enzyme knockout samples. **Fig. S3.** Schematic workflow of the calculation of DNA methylation heterogeneity. **Fig. S4.** Impacts of epigenetic features on DNA methylation heterogeneity. **Fig. S5.** The temporal association between promoter DNA methylation heterogeneity and expression heterogeneity. **Fig. S6.** Programmed DNA methylation heterogeneity in mouse pre-implantation embryos.

Additional file 2: Table S1. A total of 64 genes showed both high DNA methylation heterogeneity at promoters and expression heterogeneity in human 8-cell stage embryos.

Additional file 3. Review history.

Acknowledgements

We thank Dr. Bing Zhu for providing the *Stella*$^{-/-}$ mice.

Peer review information

Kevin Pang was the primary editor of this article and managed its editorial process and peer review in collaboration with the rest of the editorial team.

Review history

The review history is available as Additional file 3.

Authors' contributions

C.Z., N.Z., and Y.Z. (Yong Zhang) designed the study, built the model and wrote the manuscript. C.Z. analyzed the data and implemented the software package. Y.Z. (Yalin Zhang) and W.L performed the experiments under the supervision of S.G. N.T. contributed to the manuscript writing. All authors approved the final version of the manuscript.

Funding

This work was supported by the National Key Research and Development Program of China (2017YFA0102600, 2016YFA0100400), National Natural Science Foundation of China (32030022, 31970642, 31721003, 31900491, 31601079, 62072277), Program of Shanghai Academic Research Leader (17XD1403600), National Program for Support of Top-notch Young Professionals, China Postdoctoral Science Foundation (2018 M642073), Major Program of Development Fund for Shanghai Zhangjiang National Innovation Demonstration Zone (ZJ2018-ZD-004), Shanghai Municipal Science and Technology Major Project (2018SHZDZX01), LCNBI, ZJLab, and Excellent Youth Training Programs of Health Department of Shanghai Municipal Government (17QA1402700).

Availability of data and materials

The raw sequence data generated in this manuscript have been deposited in the ArrayExpress database at EMBL-EBI (https://www.ebi.ac.uk/arrayexpress) under accession number E-MTAB-9360 [58, 59] and the Genome Sequence Archive (GSA) in BIG Data Center (https://bigd.big.ac.cn/gsa) with accession number CRA002264 [60]. The public scBS-seq data of human early embryos used in this paper are available at Gene Expression Omnibus (GEO) under accession GSE81233 [12]. The public COOL-seq data of human early embryos used in this paper are available at GEO under accession GSE100272 [16]. The public scRNA-seq data of human early embryos used in this paper are available at GEO under accession GSE36552 [61] and GSE71318 [62]. The public ChIP-seq data of human early embryos used in this paper are available at GEO under accession GSE124718 [56]. The public ATAC-seq data of human early embryos used in this paper are available at GEO under accession GSE101571 [55]. The public COOL-seq data of mouse early embryos used in this paper are available at GEO under accession GSE78140 [15]. The public scRNA-seq data of mouse early embryos used in this paper are available at GEO under accession GSE45719 [63], GSE65160 [64], and GSE98150 [65]. The public BS-seq data of mouse embryos with methylation enzyme knockout used in this paper are available at GEO under accession GSE130735 [30]. The public BS-seq data of zebrafish embryos with methylation enzyme knockout used in this paper are available at GEO under accession GSE68087 [66]. MethylTransition is an open-source R package under GNU General Public License v3.0, and it is available at https://github.com/TongjiZhanglab/MethylTransition [26] and https://doi.org/10.5281/zenodo.4075746 [67]. The data analysis in this manuscript was conducted using several custom scripts, all available at https://github.com/TongjiZhanglab/MethylTransition_DataAnalysis.

Ethics approval and consent to participate

The animal experimental procedures were performed according to the Tongji University Guide for the use of laboratory animals under the approval of the Ethics Committee of Experimental Animal Center of Tongji University (approval number: TJLAC-020-141).

Consent for publication

Not applicable

Competing interests

The authors declare that they have no competing interests.

Author details

[1]Institute for Regenerative Medicine, Shanghai East Hospital, Shanghai Key Laboratory of Signaling and Disease Research, Frontier Science Center for Stem Cell Research, School of Life Science and Technology, Tongji University, Shanghai 200092, China. [2]School of Mathematics and Statistics, Shandong University at Weihai, Weihai 264209, China. [3]Clinical and Translational Research Center of Shanghai First Maternity and Infant Hospital, Tongji University, Shanghai 200092, China.

Received: 21 April 2020 Accepted: 27 October 2020
Published online: 16 November 2020

References

1. Bird A. DNA methylation patterns and epigenetic memory. Genes Dev. 2002;16:6–21.
2. Hackett JA, Surani MA. DNA methylation dynamics during the mammalian life cycle. Philos Trans R Soc Lond Ser B Biol Sci. 2013;368:20110328.
3. Guo H, Zhu P, Yan L, Li R, Hu B, Lian Y, Yan J, Ren X, Lin S, Li J, et al. The DNA methylation landscape of human early embryos. Nature. 2014;511:606–10.

4. Biase FH, Cao X, Zhong S. Cell fate inclination within 2-cell and 4-cell mouse embryos revealed by single-cell RNA sequencing. Genome Res. 2014;24:1787–96.
5. Shi J, Chen Q, Li X, Zheng X, Zhang Y, Qiao J, Tang F, Tao Y, Zhou Q, Duan E. Dynamic transcriptional symmetry-breaking in pre-implantation mammalian embryo development revealed by single-cell RNA-seq. Development. 2015;142:3468–77.
6. Torres-Padilla ME, Parfitt DE, Kouzarides T, Zernicka-Goetz M. Histone arginine methylation regulates pluripotency in the early mouse embryo. Nature. 2007;445:214–8.
7. Goolam M, Scialdone A, Graham SJL, Macaulay IC, Jedrusik A, Hupalowska A, Voet T, Marioni JC, Zernicka-Goetz M. Heterogeneity in Oct4 and Sox2 targets biases cell fate in 4-cell mouse embryos. Cell. 2016;165:61–74.
8. Wang J, Wang L, Feng G, Wang Y, Li Y, Li X, Liu C, Jiao G, Huang C, Shi J, et al. Asymmetric expression of LincGET biases cell fate in two-cell mouse embryos. Cell. 2018;175:1887–901 e1818.
9. Hupalowska A, Jedrusik A, Zhu M, Bedford MT, Glover DM, Zernicka-Goetz M. CARM1 and paraspeckles regulate pre-implantation mouse embryo development. Cell. 2018;175:1902–16 e1913.
10. Vassena R, Boue S, Gonzalez-Roca E, Aran B, Auer H, Veiga A, Izpisua Belmonte JC. Waves of early transcriptional activation and pluripotency program initiation during human preimplantation development. Development. 2011;138: 3699–709.
11. Xue Z, Huang K, Cai C, Cai L, Jiang CY, Feng Y, Liu Z, Zeng Q, Cheng L, Sun YE, et al. Genetic programs in human and mouse early embryos revealed by single-cell RNA sequencing. Nature. 2013;500:593–7.
12. Zhu P, Guo H, Ren Y, Hou Y, Dong J, Li R, Lian Y, Fan X, Hu B, Gao Y, et al. Single-cell DNA methylome sequencing of human preimplantation embryos. Nat Genet. 2018;50:12–9.
13. Chen ZX, Riggs AD. Maintenance and regulation of DNA methylation patterns in mammals. Biochem Cell Biol. 2005;83:438–48.
14. Smallwood SA, Lee HJ, Angermueller C, Krueger F, Saadeh H, Peat J, Andrews SR, Stegle O, Reik W, Kelsey G. Single-cell genome-wide bisulfite sequencing for assessing epigenetic heterogeneity. Nat Methods. 2014;11:817–20.
15. Guo F, Li L, Li J, Wu X, Hu B, Zhu P, Wen L, Tang F. Single-cell multi-omics sequencing of mouse early embryos and embryonic stem cells. Cell Res. 2017;27:967–88.
16. Li L, Guo F, Gao Y, Ren Y, Yuan P, Yan L, Li R, Lian Y, Li J, Hu B, et al. Single-cell multi-omics sequencing of human early embryos. Nat Cell Biol. 2018;20:847–58.
17. Otto SP, Walbot V. DNA methylation in eukaryotes: kinetics of demethylation and de novo methylation during the life cycle. Genetics. 1990;124:429–37.
18. Pfeifer GP, Steigerwald SD, Hansen RS, Gartler SM, Riggs AD. Polymerase chain reaction-aided genomic sequencing of an X chromosome-linked CpG island: methylation patterns suggest clonal inheritance, CpG site autonomy, and an explanation of activity state stability. Proc Natl Acad Sci U S A. 1990;87:8252–6.
19. Genereux DP, Miner BE, Bergstrom CT, Laird CD. A population-epigenetic model to infer site-specific methylation rates from double-stranded DNA methylation patterns. Proc Natl Acad Sci U S A. 2005;102:5802–7.
20. Sontag LB, Lorincz MC, Georg Luebeck E. Dynamics, stability and inheritance of somatic DNA methylation imprints. J Theor Biol. 2006;242:890–9.
21. Fu AQ, Genereux DP, Stoger R, Burden AF, Laird CD, Stephens M. Statistical inference of in vivo properties of human DNA methyltransferases from double-stranded methylation patterns. Plos One. 2012;7:e32225.
22. Arand J, Spieler D, Karius T, Branco MR, Meilinger D, Meissner A, Jenuwein T, Xu G, Leonhardt H, Wolf V, Walter J. In vivo control of CpG and non-CpG DNA methylation by DNA methyltransferases. Plos Genet. 2012;8:e1002750.
23. McGovern AP, Powell BE, Chevassut TJ. A dynamic multi-compartmental model of DNA methylation with demonstrable predictive value in hematological malignancies. J Theor Biol. 2012;310:14–20.
24. von Meyenn F, Iurlaro M, Habibi E, Liu NQ, Salehzadeh-Yazdi A, Santos F, Petrini E, Milagre I, Yu M, Xie Z, et al. Impairment of DNA methylation maintenance is the main cause of global demethylation in naive embryonic stem cells. Mol Cell. 2016;62:848–61.
25. Busto-Moner L, Morival J, Ren H, Fahim A, Reitz Z, Downing TL, Read EL. Stochastic modeling reveals kinetic heterogeneity in post-replication DNA methylation. PLoS Comput Biol. 2020;16:e1007195.
26. Zhao C, Zhang N, Zhang Y, Tuersunjiang N, Gao S, Liu W, Zhang Y. A DNA methylation state transition model reveals the programmed epigenetic heterogeneity in human pre-implantation embryos. GitHub. https://github.com/TongjiZhanglab/MethylTransition. Accessed 1 Nov 2020.
27. Hodges E, Smith AD, Kendall J, Xuan Z, Ravi K, Rooks M, Zhang MQ, Ye K, Bhattacharjee A, Brizuela L, et al. High definition profiling of mammalian DNA methylation by array capture and single molecule bisulfite sequencing. Genome Res. 2009;19:1593–605.
28. Lee DD, Seung HS. Algorithms for non-negative matrix factorization. In: Proceedings of the 13th International Conference on Neural Information Processing Systems. Denver: MIT Press; 2000. p. 535–41.
29. Greenberg MVC, Bourc'his D. The diverse roles of DNA methylation in mammalian development and disease. Nat Rev Mol Cell Biol. 2019;20:590–607.
30. Dahlet T, Argueso Lleida A, Al Adhami H, Dumas M, Bender A, Ngondo RP, Tanguy M, Vallet J, Auclair G, Bardet AF, Weber M. Genome-wide analysis in the mouse embryo reveals the importance of DNA methylation for transcription integrity. Nat Commun. 2020;11:3153.
31. Bogdanovic O, Smits AH, de la Calle ME, Tena JJ, Ford E, Williams R, Senanayake U, Schultz MD, Hontelez S, van Kruijsbergen I, et al. Active DNA demethylation at enhancers during the vertebrate phylotypic period. Nat Genet. 2016;48:417–26.
32. 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. 2014;20:861–74.
33. Blomen VA, Boonstra J. Stable transmission of reversible modifications: maintenance of epigenetic information through the cell cycle. Cell Mol Life Sci. 2011;68:27–44.
34. Suzuki MM, Bird A. DNA methylation landscapes: provocative insights from epigenomics. Nat Rev Genet. 2008;9:465–76.
35. Smallwood SA, Kelsey G. De novo DNA methylation: a germ cell perspective. Trends Genet. 2012;28:33–42.
36. Nakamura T, Arai Y, Umehara H, Masuhara M, Kimura T, Taniguchi H, Sekimoto T, Ikawa M, Yoneda Y, Okabe M, et al. PGC7/Stella protects against DNA demethylation in early embryogenesis. Nat Cell Biol. 2007;9:64–71.
37. Paul S, Knott JG. Epigenetic control of cell fate in mouse blastocysts: the role of covalent histone modifications and chromatin remodeling. Mol Reprod Dev. 2014;81:171–82.

38. Kohri N, Akizawa H, Iisaka S, Bai H, Yanagawa Y, Takahashi M, Komatsu M, Kawai M, Nagano M, Kawahara M. Trophectoderm regeneration to support full-term development in the inner cell mass isolated from bovine blastocyst. J Biol Chem. 2019;294:19209–23.
39. Chi F, Sharpley MS, Nagaraj R, Roy SS, Banerjee U. Glycolysis-independent glucose metabolism distinguishes TE from ICM fate during mammalian embryogenesis. Dev Cell. 2020;53:9–26 e24.
40. Levine AJ, Brivanlou AH. GDF3, a BMP inhibitor, regulates cell fate in stem cells and early embryos. Development. 2006; 133:209–16.
41. Hermitte S, Chazaud C. Primitive endoderm differentiation: from specification to epithelium formation. Philos Trans R Soc Lond Ser B Biol Sci. 2014;369. https://doi.org/10.1098/rstb.2013.0537.
42. Wamaitha SE, Grybel KJ, Alanis-Lobato G, Gerri C, Ogushi S, McCarthy A, Mahadevaiah SK, Healy L, Lea RA, Molina-Arcas M, et al. IGF1-mediated human embryonic stem cell self-renewal recapitulates the embryonic niche. Nat Commun. 2020;11:764.
43. Assou S, Boumela I, Haouzi D, Monzo C, Dechaud H, Kadoch IJ, Hamamah S. Transcriptome analysis during human trophectoderm specification suggests new roles of metabolic and epigenetic genes. Plos One. 2012;7:e39306.
44. Schulz KN, Harrison MM. Mechanisms regulating zygotic genome activation. Nat Rev Genet. 2019;20:221–34.
45. Gu C, Liu S, Wu Q, Zhang L, Guo F. Integrative single-cell analysis of transcriptome, DNA methylome and chromatin accessibility in mouse oocytes. Cell Res. 2019;29:110–23.
46. Chen Z, Yin Q, Inoue A, Zhang C, Zhang Y. Allelic H3K27me3 to allelic DNA methylation switch maintains noncanonical imprinting in extraembryonic cells. Sci Adv. 2019;5:eaay7246.
47. Rose NR, Klose RJ. Understanding the relationship between DNA methylation and histone lysine methylation. Biochim Biophys Acta. 1839;2014:1362–72.
48. Dhayalan A, Rajavelu A, Rathert P, Tamas R, Jurkowska RZ, Ragozin S, Jeltsch A. The Dnmt3a PWWP domain reads histone 3 lysine 36 trimethylation and guides DNA methylation. J Biol Chem. 2010;285:26114–20.
49. Li Y, Zheng H, Wang Q, Zhou C, Wei L, Liu X, Zhang W, Zhang Y, Du Z, Wang X, Xie W. Genome-wide analyses reveal a role of Polycomb in promoting hypomethylation of DNA methylation valleys. Genome Biol. 2018;19:18.
50. Fatemi M, Hermann A, Gowher H, Jeltsch A. Dnmt3a and Dnmt1 functionally cooperate during de novo methylation of DNA. Eur J Biochem. 2002;269:4981–4.
51. Fatemi M, Hermann A, Pradhan S, Jeltsch A. The activity of the murine DNA methyltransferase Dnmt1 is controlled by interaction of the catalytic domain with the N-terminal part of the enzyme leading to an allosteric activation of the enzyme after binding to methylated DNA. J Mol Biol. 2001;309:1189–99.
52. Gujar H, Weisenberger DJ, Liang G. The roles of human DNA methyltransferases and their isoforms in shaping the epigenome. Genes (Basel). 2019;10. https://doi.org/10.3390/genes10020172.
53. Han L, Ren C, Zhang J, Shu W, Wang Q. Differential roles of Stella in the modulation of DNA methylation during oocyte and zygotic development. Cell Discov. 2019;5:9.
54. Weber M, Hellmann I, Stadler MB, Ramos L, Paabo S, Rebhan M, Schubeler D. Distribution, silencing potential and evolutionary impact of promoter DNA methylation in the human genome. Nat Genet. 2007;39:457–66.
55. Wu J, Xu J, Liu B, Yao G, Wang P, Lin Z, Huang B, Wang X, Li T, Shi S, et al. Chromatin analysis in human early development reveals epigenetic transition during ZGA. Nature. 2018;557:256–60.
56. Xia W, Xu J, Yu G, Yao G, Xu K, Ma X, Zhang N, Liu B, Li T, Lin Z, et al. Resetting histone modifications during human parental-to-zygotic transition. Science. 2019;365:353–60.
57. Pedregosa F, Varoquaux G, Gramfort A, Michel V, Thirion B, Grisel O, Blondel M, Prettenhofer P, Weiss R, Dubourg V. Scikit-learn: machine learning in Python. J Mach Learn Res. 2011;12:2825–30.
58. Zhao C, Zhang N, Zhang Y, Tuersunjiang N, Gao S, Liu W, Zhang Y. A DNA methylation state transition model reveals the programmed epigenetic heterogeneity in human pre-implantation embryos. Datasets. ArrayExpress. https://www.ebi.ac.uk/arrayexpress/experiments/E-MTAB-9360. Accessed 1 Nov 2020.
59. Zhao C, Zhang N, Zhang Y, Tuersunjiang N, Gao S, Liu W, Zhang Y. A DNA methylation state transition model reveals the programmed epigenetic heterogeneity in human pre-implantation embryos. Datasets. ENA. https://www.ebi.ac.uk/ena/browser/view/PRJEB39418. Accessed 1 Nov 2020.
60. Zhao C, Zhang N, Zhang Y, Tuersunjiang N, Gao S, Liu W, Zhang Y. A DNA methylation state transition model reveals the programmed epigenetic heterogeneity in human pre-implantation embryos. Datasets. GSA. https://bigd.big.ac.cn/gsa/browse/CRA002264. Accessed 1 Nov 2020.
61. Yan L, Yang M, Guo H, Yang L, Wu J, Li R, Liu P, Lian Y, Zheng X, Yan J, et al. Single-cell RNA-Seq profiling of human preimplantation embryos and embryonic stem cells. Nat Struct Mol Biol. 2013;20:1131–9.
62. Dang Y, Yan L, Hu B, Fan X, Ren Y, Li R, Lian Y, Yan J, Li Q, Zhang Y, et al. Tracing the expression of circular RNAs in human pre-implantation embryos. Genome Biol. 2016;17:130.
63. Deng Q, Ramskold D, Reinius B, Sandberg R. Single-cell RNA-seq reveals dynamic, random monoallelic gene expression in mammalian cells. Science. 2014;343:193–6.
64. Yang L, Ma ZC, Cao C, Zhang YH, Wu XL, Lee R, Hu BQ, Wen L, Ge H, Huang YY, et al. MR-seq: measuring a single cell's transcriptome repeatedly by RNA-seq. Sci Bull. 2017;62:391–8.
65. Wang C, Liu X, Gao Y, Yang L, Li C, Liu W, Chen C, Kou X, Zhao Y, Chen J, et al. Reprogramming of H3K9me3-dependent heterochromatin during mammalian embryo development. Nat Cell Biol. 2018;20:620–31.
66. Bogdanovic O, Smits AH, Mustienes ED, Tena JJ, Ford E, Williams R, Senanayake U, Schultz MD, Hontelez S, van Kruijsbergen I, et al. Active DNA demethylation at enhancers during the vertebrate phylotypic period. Nat Genet. 2016;48:417.
67. Zhao C, Zhang N, Zhang Y, Tuersunjiang N, Gao S, Liu W, Zhang Y. A DNA methylation state transition model reveals the programmed epigenetic heterogeneity in human pre-implantation embryos. Zenodo. https://doi.org/10.5281/zenodo.4075746. Accessed 1 Nov 2020.

Publisher's Note