ARTICLES
https://doi.org/10.1038/s41588-017-0003-x
nature genetics

# Dynamic epigenomic landscapes during early lineage specification in mouse embryos

Yu Zhang[1], Yunlong Xiang[1,2], Qiangzong Yin[1], Zhenhai Du[1,2], Xu Peng[3], Qiujun Wang[1,2], Miguel Fidalgo[4], Weikun Xia[1,2], Yuanyuan Li[1], Zhen-ao Zhao[5], Wenhao Zhang[1,2], Jing Ma[1], Feng Xu[3,6], Jianlong Wang[4], Lei Li[5] and Wei Xie[1,2*]

**In mammals, all somatic development originates from lineage segregation in early embryos. However, the dynamics of transcriptomes and epigenomes acting in concert with initial cell fate commitment remains poorly characterized. Here we report a comprehensive investigation of transcriptomes and base-resolution methylomes for early lineages in peri- and postimplantation mouse embryos. We found allele-specific and lineage-specific de novo methylation at CG and CH sites that led to differential methylation between embryonic and extraembryonic lineages at promoters of lineage regulators, gene bodies, and DNA-methylation valleys. By using Hi-C experiments to define chromatin architecture across the same developmental period, we demonstrated that both global demethylation and remethylation in early development correlate with chromatin compartments. Dynamic local methylation was evident during gastrulation, which enabled the identification of putative regulatory elements. Finally, we found that de novo methylation patterning does not strictly require implantation. These data reveal dynamic transcriptomes, DNA methylomes, and 3D chromatin landscapes during the earliest stages of mammalian lineage specification.**

In mammals, early lineage specification in preimplantation and postimplantation embryonic development generates founder tissues for all subsequent somatic development[1]. The first lineage specification starts at the morula stage, when the inner cell mass (ICM) and the trophectoderm (TE) begin to segregate[2]. The ICM contains both cells of the epiblast lineage, which give rise to the entire fetus, and cells of the primitive endoderm lineage, which form visceral endoderm (VE) and parietal endoderm[3,4]. VE becomes the chief metabolic component of the visceral yolk sac, and parietal endoderm contributes to the transient parietal yolk sac[4]. TE contains progenitor cells for trophoblasts, which form the majority of the fetal-origin part of the placenta[3]. In mice, by embryonic day 6.5 (E6.5), the anterior epiblast gives rise to ectoderm, and the posterior proximal epiblast develops into the primitive streak, which then forms mesoderm and endoderm[5]. The resulting three germ layers contain virtually all progenitors for the future body plan[6].

Notably, early cell fate commitment is accompanied by extensive epigenetic reprogramming[2]. For example, drastic demethylation and remethylation of DNA take place during early embryogenesis[7]. DNA methylation plays critical roles in gene repression, genomic imprinting, and X chromosome inactivation[8]. Deficiency in DNA methyltransferases (DNMTs) often leads to lethality or sterility[7]. Interestingly, defects in extraembryonic tissues, which provide both nutrients and developmental cues for embryonic development, are frequently found in mice deficient in DNMTs[9–11]. A large portion of DNA methylation in gametes is removed during preimplantation development[7]. The methylome in postimplantation embryos then forms part of the epigenetic basis for the entire body plan. DNA methylome reprogramming in preimplantation embryos has been studied extensively[12–16]. However, because of the limited materials available and the difficulty of tissue isolation in early embryos, lineage-specific regulation of transcriptomes and epigenomes in peri- and postimplantation embryos is poorly characterized. Here we conducted a comprehensive analysis of transcriptomes and whole-genome DNA methylomes at single-base resolution for major lineages that arise before and after implantation. This analysis, together with Hi-C experiments probing higher-order chromatin structure during the same period, provides unprecedented spatiotemporal views for the establishment of the molecular architecture regulating early cell fate commitment and body plan in mammals.

## Results

**Mapping global transcriptomes and DNA methylomes during early lineage specification.** To study the transcriptional programs and epigenomes involved in early lineage segregation, we carefully dissected various tissues from peri- and postimplantation mouse embryos (DBA/2N male × C57BL/6N female mice) using methods described previously[17,18] (Fig. 1a, Supplementary Fig. 1a, Methods). These included ICM (E3.5 and E4.0), mural TE (E3.5), VE (E5.5 and E6.5), epiblast (E5.5 and E6.5), ectoderm (E7.5), endoderm (E7.5), mesoderm (E7.5), and primitive streak (PS) (E7.5). We chose TE and VE as representatives of extraembryonic tissues. Analysis of lineage-marker genes and transcriptomes strongly supported the correct identities of these tissues and dynamic transcription landscapes in early lineages (Supplementary Fig. 1b–e, Supplementary Table 1). E3.5 ICM and E4.0 ICM were grouped together and were distantly connected with all ICM-derived tissues from mice at E5.5 to E7.5 (Supplementary Fig. 1e). Both the marker genes and the global transcriptome of endoderm were similar to those of E6.5 VE

[1]Center for Stem Cell Biology and Regenerative Medicine, MOE Key Laboratory of Bioinformatics, School of Life Sciences, Tsinghua University, Beijing, China. [2]THU-PKU Center for Life Sciences, Tsinghua University, Beijing, China. [3]Institute of Molecular and Cell Biology, Agency for Science, Technology and Research (A*STAR), Singapore, Singapore. [4]Department of Cell, Developmental and Regenerative Biology, and Black Family Stem Cell Institute, Icahn School of Medicine at Mount Sinai, New York, NY, USA. [5]State Key Laboratory of Stem Cell and Reproductive Biology, Institute of Zoology, Chinese Academy of Sciences, Beijing, China. [6]Singapore Institute for Clinical Sciences, A*STAR, Singapore, Singapore. Yu Zhang, Yunlong Xiang, Qiangzong Yin, and Zhenhai Du contributed equally to this work. *e-mail: xiewei121@tsinghua.edu.cn

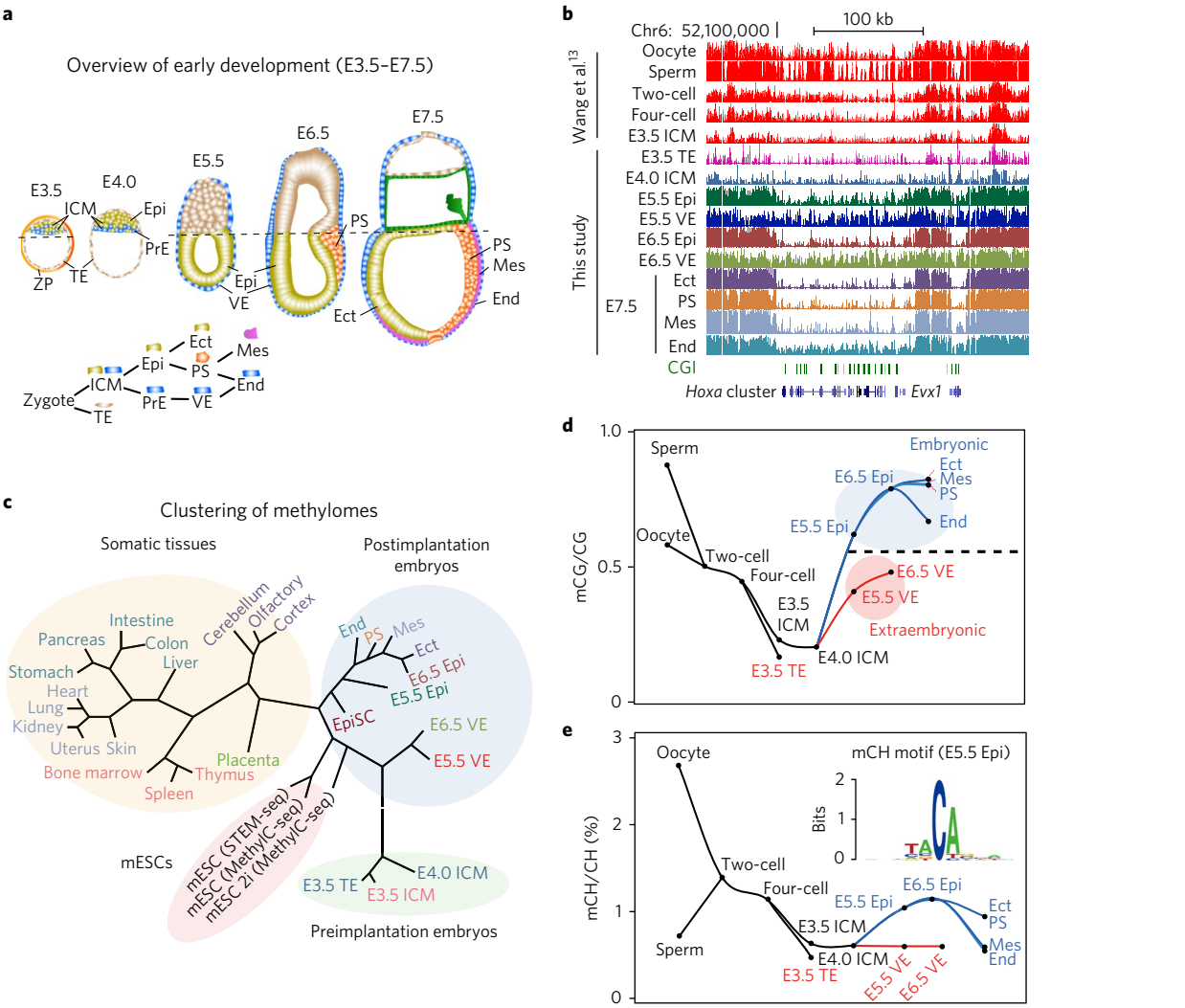


**Fig. 1 | The dynamics of lineage-specific DNA methylomes in early embryos. a**, An overview of mouse embryonic development from E3.5 to E7.5. The lineage relationship among various tissues is shown at the bottom. ICM, inner cell mass; TE, trophectoderm; ZP, zona pellucida; PrE, primitive endoderm; VE, visceral endoderm; Epi, epiblast; Ect, ectoderm; End, endoderm; Mes, mesoderm; PS, primitive streak. **b**, A UCSC Genome Browser snapshot showing lineage-specific DNA-methylation landscapes from oocyte/sperm to E7.5 embryo. MethylC-seq data from sperm/oocyte to E3.5 ICM were published previously[13]. **c**, A dendrogram showing clustering of mouse methylomes (1-kb bin for the entire genome) for early lineages, mESCs (this study or a previous study[22]), epiblast stem cells (EpiSC[54]), and somatic tissues[21]. **d**, The dynamics of lineage-specific CG methylation (mCG) from oocyte/sperm to E7.5 embryo. **e**, An analysis similar to that in **d** but for CH methylation (mCH). The sequence motif for CH methylation in E5.5 epiblast is shown as an inset. Analyses were done using pooled data from 2–3 biological replicates.

(Supplementary Fig. 1c–e); this is in line with the notion that VE also contributes to the endoderm lineage[19]. Taken together, these data demonstrate the high quality of the early tissues we isolated and lineage-specific transcription landscapes in early development.

Next, to examine the dynamics of DNA methylation during early lineage commitment, we developed a low-input method for genome-wide DNA-methylation profiling: STEM-seq (small-scale TELP-enabled methylome sequencing) (Supplementary Fig. 2a). This approach reduces DNA loss because bisulfite conversion is performed before TELP-mediated DNA amplification, a highly sensitive library-preparation method[20]. Our data showed that STEM-seq could accurately determine DNA methylomes with as little as 10 ng of genomic DNA, or 500 cells (Supplementary Fig. 2b–e). Next, we profiled high-depth methylomes for early lineages (with 2–3 replicates) by STEM-seq (sequencing information is provided in Supplementary Table 2). We first focused on CG methylation. The methylome data generally showed excellent replicate reproducibility (Supplementary Fig. 3a) and genome coverage for CG sites (Supplementary Fig. 3b,c). A global view of methylomes revealed large hypomethylated regions around the *Hoxa* gene cluster, as expected[21] (Fig. 1b). We also observed dynamic DNA methylation near developmentally regulated genes including *Hnf4a* (VE/endoderm marker), *Pou5f1* (also known as *Oct4*), and *Tdgf1* (epiblast markers), which are reciprocally methylated in epiblast or VE (Supplementary Fig. 3d). Notably, these promoters showed intermediate levels of methylation in endoderm, consistent with the mixed origin of endoderm from both epiblast and VE[19]. We then investigated whether the global CG methylome of each tissue reflects its spatiotemporal relationship by conducting a hierarchical clustering analysis of methylomes for early embryos, as well as for somatic tissues[21] and mouse embryonic stem cells (mESCs)[22] (Fig. 1c). We found that E3.5 ICM, E3.5 TE, and E4.0 ICM, which are all hypomethylated, clustered together away from all other lineages (Fig. 1c). The methylomes of endoderm, ectoderm, and mesoderm were much closer to each other than to the methylomes of the derivative somatic tissues. These data suggest

that substantial epigenome drift occurs between embryonic progenitor and somatic tissue.

**Dynamic lineage-specific methylation at CG and CH sites.** The segregation of ICM and TE is the first lineage-specification event in embryos[2]. Genome-wide, we identified a total of 208 and 47 promoters that were hypermethylated in ICM and TE, respectively (Supplementary Table 3, Methods). The majority of genes that were differentially expressed between ICM and TE did not show differences in promoter methylation (Supplementary Fig. 3e). For example, both *Pou5f1* and *Tdgf1* were expressed at high levels in ICM but not in TE, yet their promoters remained unmethylated in both lineages (Supplementary Fig. 3d). However, both promoters are methylated in TE-derived placenta[21] (Supplementary Fig. 3d), which indicates that DNA methylation is involved in maintaining these lineage regulators but not in initially silencing them. We then asked how de novo methylation occurs in concert with the specification of epiblast and VE. From E4.0 to E6.5, DNA methylation increased considerably genome-wide in epiblast, but it increased to a lesser extent in VE (Fig. 1d, Supplementary Fig. 3f). This was accompanied by epiblast-specific sharp upregulation of *Dnmt3a*, *Dnmt3b*, and *Dnmt3l* at E5.5, with *Dnmt3l* likely undergoing autorepression through promoter methylation at E6.5[23] (Supplementary Fig. 3g,h). In addition to CG methylation, CH methylation was relatively enriched in oocytes but was barely detected in sperm and after the four-cell stage, although it reappeared in E5.5 epiblast (Fig. 1e). Unlike CG methylation, which showed further increases, CH methylation decreased from E6.5 to E7.5 (Fig. 1d, e). This is consistent with the reduced expression of *Dnmt*-family genes (Supplementary Fig. 3g) and the fact that CH methylation cannot be maintained by DNMT1[24]. CH methylation in early embryos preferentially occurs in TACAG sequences (Fig. 1e and data not shown), similar to what is observed in embryonic stem cells[24]. In sum, these data indicate that lineage-specific de novo methylation of CG and CH sites correlates with the activities of DNMT proteins. Because CH methylation levels were much lower than CG methylation levels (Fig. 1e), we focused mainly on CG methylation in subsequent analyses, unless otherwise noted.

**Allele-specific de novo methylation highlights conserved gene body methylation.** In preimplantation development, the two parental genomes undergo differential demethylation[7]. We asked whether

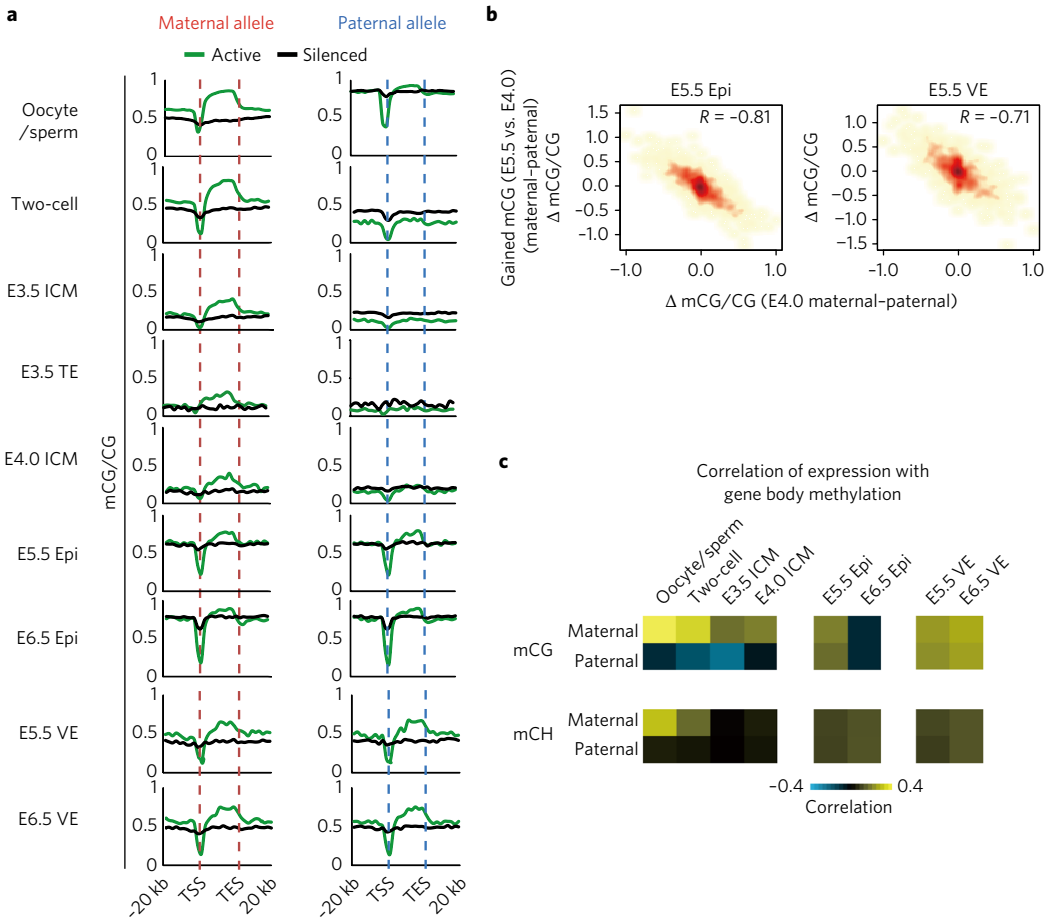


**Fig. 2 | Allelic DNA methylome reprogramming during early lineage specification. a**, The average allelic methylation levels near active and silenced genes at various stages for both embryonic and extraembryonic tissues. Considering the maternal inheritance of oocyte RNA, we used stringent criteria, selecting only genes with FPKM (fragments per kilobase of transcript per million mapped reads) values greater than 10 or less than 1 at all stages (from MII oocytes to E6.5 embryos) as constantly active or silenced genes, respectively. For MII oocyte and sperm, which have silent genomes, high and low FPKM values reflect chromatin states or transcription activities at previous developmental stages. Epi, epiblast; TSS, transcription start site; TES, transcription end site. **b**, Scatterplots comparing allelic differences in de novo methylation acquisition in E5.5 Epi and E5.5 VE and allelic differences in starting methylation in E4.0 ICM. **c**, Heat maps showing the Spearman correlation for expression versus gene body DNA methylation levels at each stage for CG and CH methylation on each allele. Analyses were done using pooled data from 2–3 biological replicates.

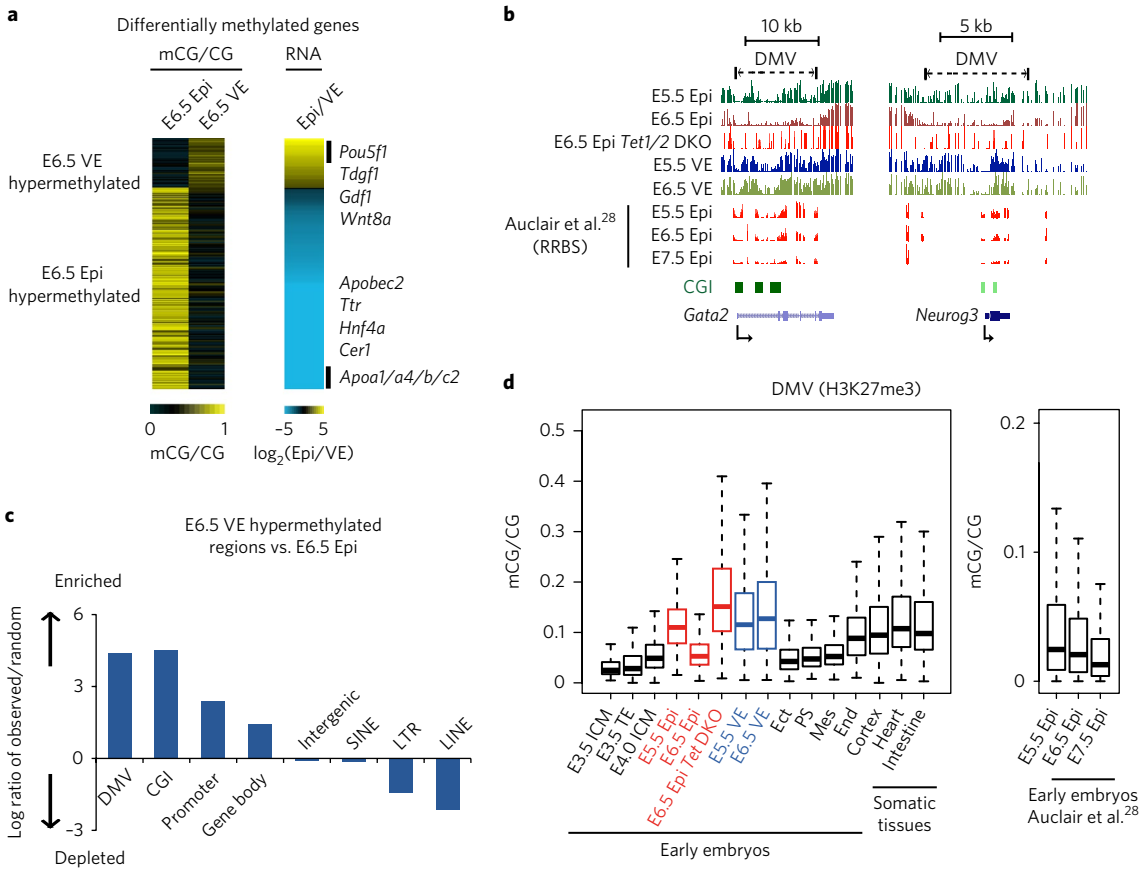


**Fig. 3 | Dynamic DNA methylation at promoters and DNA methylation valleys during lineage specification. a**, A heat map showing the promoter methylation levels (left) and gene expression ratios (right) for differentially methylated genes in epiblast (Epi) and VE (E6.5) (pooled replicates; $n$ = 2–3). Only genes with corresponding expression changes ('methylation effectors') are included. **b**, UCSC Genome Browser snapshots showing DNA methylation levels around *Gata2* and *Neurog3*. The reduced-representation bisulfite sequencing (RRBS) data from a previous study[28] are also shown. **c**, The enrichment (log ratio of observed/random) of E6.5 VE hypermethylated regions (versus E6.5 epiblast) in various classes of genomic elements. Random regions that matched the sizes of individual hypermethylated regions were used as controls. **d**, Box plots showing CG methylation levels of H3K27me3-marked DMVs (combining all DMVs in five lineages) in early embryos and somatic tissues. The red and blue boxes indicate epiblast and VE, respectively; black boxes represent various tissues as defined along the $x$-axis. Results of similar analyses using RRBS data for epiblast from E5.5 to E7.5 are also shown[28]. Ect, ectoderm; Mes, mesoderm; End, endoderm. Box plots show the median of each dataset (center line), the 25th and 75th percentiles (bottom and top edges, respectively), and 1.5 times the interquartile range (whiskers).

the two parental alleles are also subjected to distinct de novo methylation in postimplantation embryos. We first validated our allele-specific analyses with methylome data from gametes[13] and imprinted loci (Supplementary Fig. 4a,b, Methods). Notably, in E3.5 ICM, the maternal genome, but not the paternal genome, was hypermethylated in gene bodies, showing an oocyte-like methylome pattern[25] (Fig. 2a). The parental methylomes quickly became symmetrical by E5.5 as a result of allele-specific acquisition of DNA methylation that was anticorrelated with the starting levels of DNA methylation in E4.0 ICM (Fig. 2a,b). CG and CH methylation both occurred preferentially in active gene bodies in VE and epiblast from E5.5 to E6.5 (Fig. 2a,c). In E6.5 epiblast, the gene body CG methylation pattern became attenuated, and it is likely that this was a result of the saturation effect. These data show both allele-specific methylation and conserved gene body methylation in postimplantation embryos.

**Differential methylation of promoters and DNA methylation valleys between epiblast and VE.** We then asked whether the distinct methylomes in embryonic and extraembryonic tissues might regulate lineage-specific transcription programs. Because the parental methylomes became similar after E5.5, we conducted the analyses without separating the alleles. We identified promoters that were hypermethylated in E6.5 epiblast ($n$=2,936) or VE ($n$=242) (Supplementary Table 3). Of the corresponding genes, only a small fraction (6.4% and 20.2%, respectively) showed consistent changes in expression (threefold downregulation in hypermethylated tissues), and we considered these as possible 'DNA methylation effectors' (Supplementary Table 3). Genes that were specifically hypermethylated in epiblast included many VE markers, such as members of the apolipoprotein family *Apoa1*, *Apoa4*, *Apoa5*, *Apob*, and *Apoc2* (Fig. 3a). In contrast, several key epiblast marker genes such as *Pou5f1*, *Nanog*, and *Tdgf1* were hypermethylated in VE (Fig. 3a). Aside from these methylation effectors, the rest of the differentially methylated genes were largely silenced in both epiblast and VE (Supplementary Fig. 5a). Among these genes, those hypermethylated in VE were strongly enriched for developmental genes ($P=5.83\times10^{-5}$) and transcription factors ($P=7.41\times10^{-7}$), including the *Hox* genes (*Hoxb2*, *Hoxb3*, and *Hoxd12*), *Nkx2-5*, *Nkx2-6*, *Prdm14*, and *Hand1*. We did not observe this for genes that were hypermethylated in epiblast, which were overwhelmingly enriched for the olfactory receptor gene family ($P << 0.001$; fold enrichment, 4.72) (Supplementary Fig. 5a). Thus, DNA methylation is likely to be engaged in reciprocal gene silencing of lineage regulators (or future regulators) between epiblast and VE.

We found it intriguing that the promoters of developmental genes were preferentially methylated in VE. Careful examination revealed that these hypermethylated regions extended beyond promoters (Fig. 3b). Previously, we and others found that developmental genes tend to reside in large domains of hypomethylated regions, termed DNA methylation valleys (DMVs)[26] or DNA methylation canyons[27]. Using a previously described approach[26], we identified 842–900 DMVs in E6.5 epiblast, ectoderm, PS, mesoderm, and endoderm (Methods). We were not able to call DMVs in other lineages that were globally hypomethylated. Indeed, DMVs in early embryos were similarly enriched for developmental genes and Polycomb targets (Supplementary Fig. 5b,c). By identifying all hypermethylated regions in E6.5 VE and comparing them with those in E6.5 epiblast (Methods), we confirmed that promoters and CpG islands (CGIs) were preferentially methylated in VE (Fig. 3c). In epiblast, DMVs with trimethylation of histone H3 at Lys27 (H3K27me3) also gained partial DNA methylation at E5.5. Unlike those in VE, these DMVs quickly lost DNA methylation at E6.5 and remained relatively hypomethylated in somatic tissues (Fig. 3b,d, Supplementary Fig. 5d). Similar patterns were observed for an epiblast methylome dataset generated via reduced-representation bisulfite sequencing[28] (Fig. 3b,d). The changes of DNA methylation were most evident for non-CGI regions in DMVs, but they were also found in CGIs (Supplementary Fig. 5e). In fact, CGIs in DMVs were preferentially methylated in VE compared with other CGIs (Supplementary Fig. 5f). Because the DNA methylation oxidase genes *Tet1* and *Tet2* were expressed at high levels in peri- and/or postimplantation embryos (Supplementary Fig. 5g), we asked whether they are involved in demethylation of DMVs. To explore this, we generated *Tet1*/*Tet2* double-knockout (DKO) mice (by crossing *Tet1*-knockout mice and *Tet2*-knockout mice; Methods) and isolated E6.5 epiblast for STEM-seq analysis. Indeed, DMVs from *Tet1*/*Tet2* DKO mice showed increased DNA methylation compared with that in wild-type mice (Fig. 3b,d), indicating that DMVs undergo TET-mediated demethylation in epiblast at E6.5. The active demethylation of DMVs raises the possibility that perhaps the hypomethylation of DMVs is important for maintenance of the transcription plasticity of the associated developmental genes.

**Lineage-specific methylation is associated with chromatin higher-order structure.** The differential methylation between VE and epiblast was not limited to promoters and DMVs. A chromosome-wide view showed that such differences also existed in much larger regions (Fig. 4a). For example, whereas epiblast showed relatively even methylation across the chromosome, VE showed megabase-sized hypomethylated domains (Fig. 4a), a feature that resembled partially methylated domains (PMDs) in placenta[29]. Chromatin is known to be spatially organized into two types of large compartments, A and B, which show preferential physical interaction within each class but not between classes[30]. Compartments A and B generally match open chromatin domains with high gene densities and closed chromatin domains with low gene densities, respectively[30]. We asked whether the PMDs in VE correlate with such chromatin compartments. Using sisHi-C, a low-input Hi-C method[31] (Methods), we investigated higher-order chromatin organization for E3.5 ICM[31], E6.5 epiblast, E6.5 VE, and E7.5 ectoderm (Supplementary Table 2). We found that the three-dimensional chromatin interaction patterns were globally similar to one another in early lineages, as well as to those in mESCs[32] (Fig. 4b). This was also true for 'topological domains' (Fig. 4b) defined by directionality index[32] (Supplementary Fig. 6a,b), $P(s)$ curves (which reflect the relationship of genomic distances and chromatin interaction frequencies) (Supplementary Fig. 6c), and chromatin compartments (Fig. 4a). These data indicate that higher-order chromatin structure is established as early as in ICM and is largely conserved from E3.5 to E7.5. We then sought to identify all PMDs and highly methylated domains (HMDs) in E6.5 VE (Methods). Indeed, we found that HMDs and PMDs in VE correlated with chromatin compartments A and B, respectively (Fig. 4a,c). One interesting question is whether the higher-order chromatin structure modulates DNA methylation, or vice versa. As chromatin organization is already established in ICM (Fig. 4a,b), where the genome is globally hypomethylated, it is unlikely that DNA methylation regulates chromatin compartments. To test whether the preferential DNA methylation in compartment A in VE was simply due to higher transcriptional activities, we examined DNA-methylation levels in active gene bodies, inactive gene bodies, and intergenic regions in each compartment. In VE, active gene bodies were preferentially methylated in compartments A and B, which is in line with gene-body-dependent DNA methylation. However, inactive gene bodies and intergenic regions showed considerable levels of DNA methylation only in compartment A, and not in compartment B (Supplementary Fig. 7a), which suggests that compartment-correlated DNA methylation in VE may be independent of transcription.

Notably, all regions in epiblast seemed to acquire similar levels of DNA methylation in compartments A and B (Supplementary Fig. 7a). It is unclear why compartment-specific methylation was absent in epiblast. One possibility is that chromatin in compartment A is more accessible for DNMTs, but in epiblast excessive DNMT machinery leads to equal methylation in compartment B. Because CH methylation occurred at comparatively lower levels that were far from saturation, we asked whether CH methylation might be correlated with chromatin compartments in both lineages. Indeed, unlike CG methylation, CH methylation occurred preferentially in compartment A in both epiblast and VE (Supplementary Fig. 7b). As a result, CG and CH methylation were highly correlated in both E5.5 VE ($R=0.83$) and E6.5 VE ($R=0.80$), but showed weaker correlation in E5.5 epiblast ($R=0.37$) and virtually no correlation in E6.5 epiblast ($R=-0.02$) (Supplementary Fig. 7c). Taken together, our data indicate that lineage-specific de novo methylation correlates with chromatin compartment and differential expression of *Dnmt* genes.

**Paternal demethylation in preimplantation embryos correlates with chromatin compartment.** Given that de novo methylation is associated with chromatin higher-order structure, we asked whether this is also true for genome demethylation in preimplantation embryos. Surprisingly, we found that compartment A, but not compartment B, was preferentially demethylated on the paternal genome (Fig. 4d,e). This compartment-specific demethylation also explains the differential background methylation levels near active and inactive genes on the paternal genome in preimplantation embryos (Fig. 2a, Supplementary Fig. 7d). To determine whether such demethylation depends on TET3, a methylcytosine oxidase that preferentially demethylates the paternal genome[33], we analyzed a published methylome comparing wild-type and *Tet3*-knockout zygotes[34]. Although TET3 indeed showed a preference for compartment A (Supplementary Fig. 7e), its effect seemed to be moderate, thus indicating the presence of additional regulators for compartment-specific demethylation[35]. By contrast, the demethylation on the maternal allele seemed to be relatively uniform, enabling the inheritance of an oocyte methylome pattern to blastocysts (Fig. 4d,e). The allele-specific compartment-correlated methylome of ICM was clearly different from those of mESCs (Supplementary Fig. 7f). Methylomes of both primed and naive (cultured in 2i medium) mESCs showed little correlation with chromatin compartment. In sum, these data demonstrate that both demethylation and de novo methylation are associated with chromatin higher-order structure.

**Dynamic methylation identifies putative *cis*-regulatory elements during gastrulation.** Although the global methylome is largely established by E6.5 (Fig. 1d), we asked whether dynamic DNA methylation occurs at individual loci after that point. Previously, it

was shown that unmethylated regions (UMRs) and low-methylation regions (LMRs) preferentially mark *cis*-regulatory elements such as promoters and enhancers, respectively[36]. We therefore sought to identify UMRs and LMRs in early embryos as previously described[37]. In total, we identified 17,204–17,898 UMRs and 24,039–32,019 LMRs in ectoderm, PS, mesoderm, endoderm, and

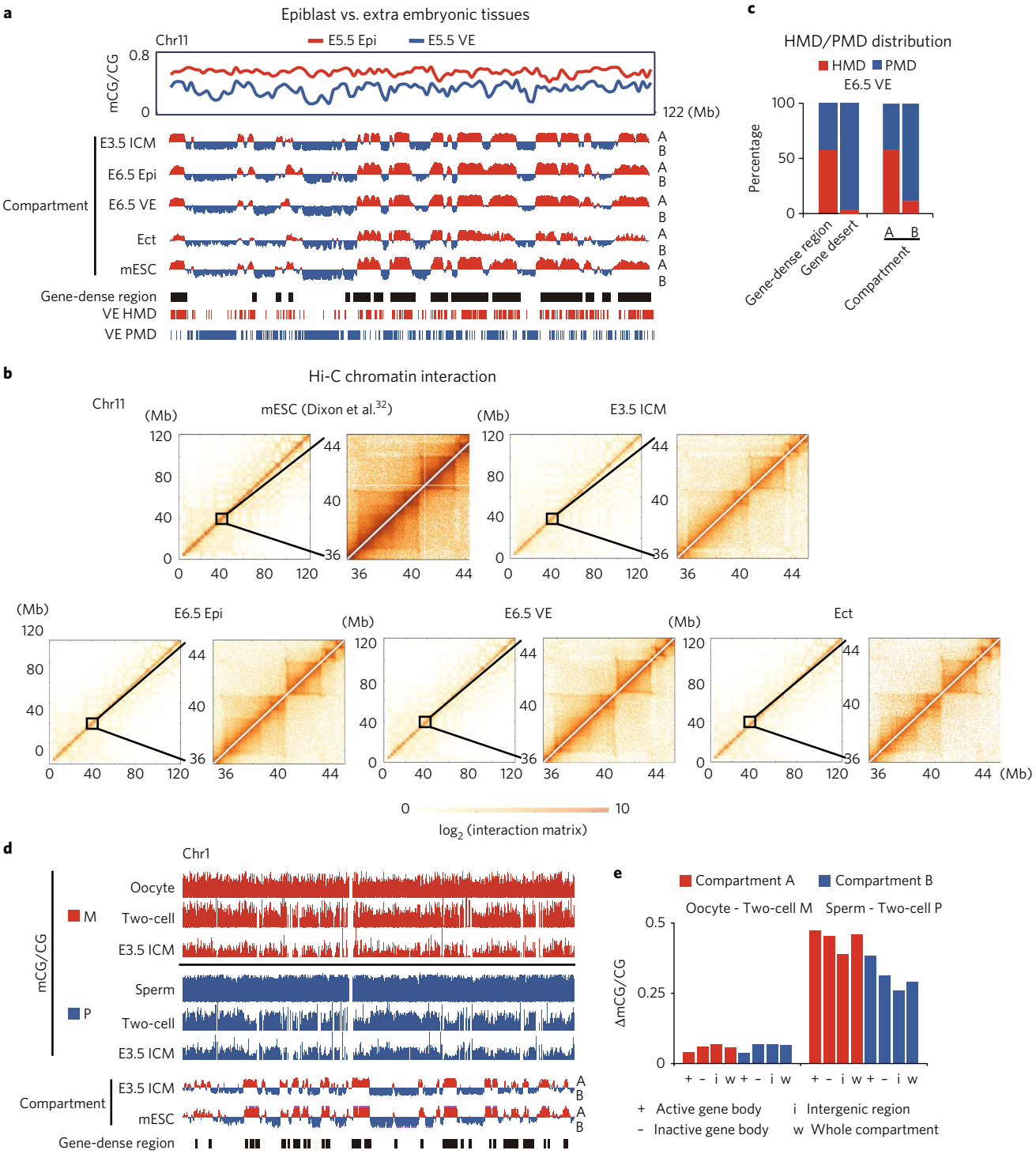


**Fig. 4 | Both demethylation and de novo methylation in early development are correlated with chromatin compartment. a**, Top, a chromosome-wide view of DNA methylation levels (1-Mb bin) for E5.5 epiblast (Epi) and E5.5 VE (replicates pooled; $n = 2$). Bottom, chromatin compartments determined by Hi-C (middle), gene-dense region, and HMD/PMD in VE (replicates pooled; $n = 2$). Ect, ectoderm. **b**, Heat maps showing chromatin interaction for mESCs[32], E3.5 ICM, E6.5 epiblast, E6.5 VE, and ectoderm (Ect) as determined by Hi-C. Zoomed-in views of highlighted regions are also shown. **c**, Bar charts showing the distribution of identified PMDs and HMDs in E6.5 VE in gene-dense/gene-desert regions (left) and in compartments A and B, respectively (right). **d**, A comparison of chromosome-wide CG methylomes between maternal (M; top) and paternal (P; bottom) alleles from gamete to E3.5 ICM[13]. Chromatin compartments and gene-dense regions are also shown at the bottom. **e**, Bar charts showing the average methylation loss from gametes to two-cell embryos on each allele in chromatin compartment A or B[13].

E6.5 epiblast (Supplementary Table 4). We did not carry out similar analyses in earlier lineages because of the difficulty of LMR/UMR calling in globally hypomethylated genomes. As validation, we found that the locations of UMRs were strongly enriched for promoters and were largely invariant among different lineages (Supplementary Fig. 8a). In contrast, LMRs were much more dynamic, indicating putative enhancers[36]. Furthermore, large fractions of UMRs and LMRs in E6.5 epiblast (94% and 58%, respectively) overlapped with DNase hypersensitivity sites in mESCs[38] (Supplementary Fig. 8b). The epiblast tissue-specific LMRs (tsLMRs) showed hypermethylation in *Tet1*/*Tet2* DKO mutant E6.5 epiblast (Fig. 5a), indicating the involvement of TET proteins in the demethylation of these

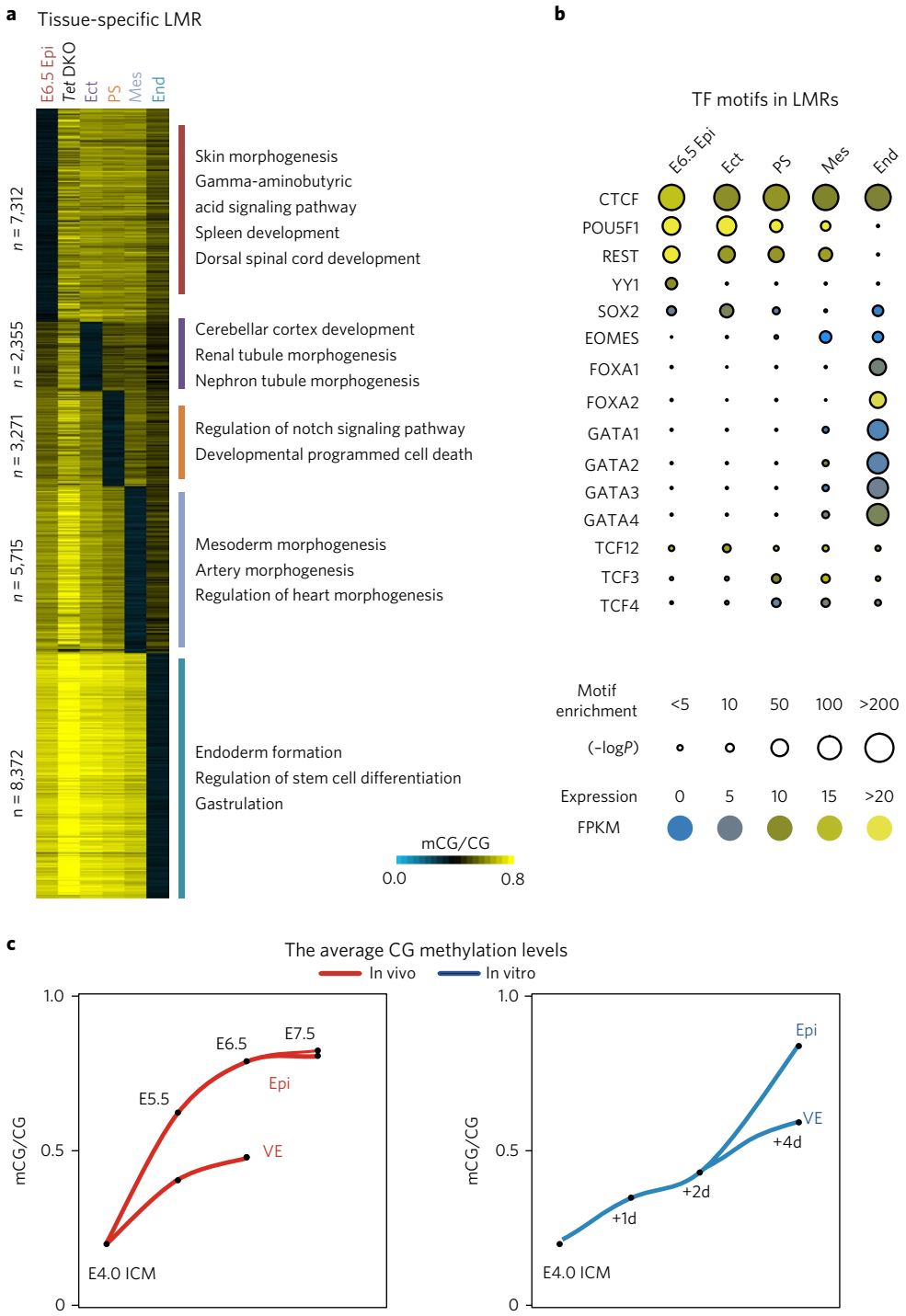


**Fig. 5 | Dynamic DNA-methylation reprogramming during gastrulation and in IVC embryos. a**, A heat map showing the average CG methylation levels in early lineages (including *Tet1*/*Tet2* DKO E6.5 epiblast (Epi)) for tsLMRs identified in this study (replicates pooled; $n = 2$–3). The GREAT analysis[39] results for tsLMRs in each tissue are also listed. Ect, ectoderm; Mes, mesoderm; End, endoderm. **b**, Motifs identified from LMRs in early lineages by HOMER[55]. As shown by the key at the bottom, motif enrichment is represented by the area of the circle, and the expression level of the corresponding transcription factor (TF) is color-coded. A pooled set of previously reported enhancers[46] was used as background. **c**, The global methylation levels (1-kb bin) for tissues isolated from in vivo and in vitro embryos (replicates pooled; $n = 2$).

putative enhancers. Using GREAT analysis[39], we found that tsLMRs (Supplementary Table 4) were preferentially located near genes involved in corresponding lineage specification (Fig. 5a). We then determined which regulators may function at LMRs by searching for their DNA motifs in these regions (Fig. 5b). For instance, the motif of POU5F1 was enriched in epiblast, ectoderm, and, to a lesser extent, PS LMRs, whereas SOX2 was enriched mainly in ectoderm LMRs. This is consistent with their expression patterns as determined in this study (Supplementary Fig. 1d) and in previous work[40–42] (it is worth noting that *Pou2f1* and *Pou3f1* are also weakly expressed at these stages (data not shown)). In fact, conditional depletion of *Pou5f1* in postimplantation embryos leads to deficient cell proliferation in PS[41]. FOXA1, FOXA2, and GATA4 (the motifs of GATA family members were highly similar; data not shown) were enriched in endoderm tsLMRs, consistent with their pivotal roles in endoderm differentiation[43–45]. Taken together, these results demonstrate that the dynamic DNA methylation at LMRs correlates with lineage identities during gastrulation.

Next, we asked whether these LMRs in early lineages are retained in somatic tissues. Using published datasets[21,46], we found that tsLMRs from early embryos showed significant overlap with putative enhancers in E14.5 and somatic tissues (Supplementary Fig. 8c). However, the enrichment decreased as development proceeded, suggesting gradual drift of the epigenome. UMRs and LMRs enriched in early embryos but not in somatic tissues (Supplementary Table 5) included those at the promoters of pluripotency genes such as *Pou5f1*, *Nanog*, and *Tdgf1* (Supplementary Fig. 8d). Distal UMRs and LMRs specific for early embryos were preferentially located near many developmental regulator genes such as *Lin28a*, *Sall4*, and *Dnmt3b* (Supplementary Fig. 8d,e). Taken together, these data demonstrate that dynamic DNA methylation occurs at lineage-specific putative enhancers during gastrulation.

**Global lineage methylome patterning does not strictly require implantation.** Because de novo methylation is accompanied by the implantation of embryos, we asked whether implantation is required for establishment of the DNA methylome. Notably, mouse embryos can grow through the early stages of organogenesis in vitro[47]. Thus, we isolated E4.0 embryos in vivo; cultured them in vitro using established protocols[47,48]; and collected the embryos at days 1, 2 and 4 for STEM-seq and RNA-seq analyses. The in vitro–cultured (IVC) embryos developed more slowly than their in vivo counterparts, retaining a blastocyst-like shape after 2 d of culture (data not shown) and then adopting a postimplantation-embryo-like morphology by day 4 (Supplementary Fig. 9a). Despite the delayed development, de novo methylation occurred in embryos after 1 d of IVC culture (IVC + 1d) and in IVC + 2d embryos (Fig. 5c). For day 4 embryos, we segregated and collected epiblast-like and VE-like tissues (on the basis of morphology) (Methods). Lineage-marker analysis and global transcriptome clustering analysis showed that IVC + 4d epiblast and VE resembled E5.5 epiblast and VE in vivo (Supplementary Fig. 9b,c). In IVC + 4d VE, DNA methylation continued to increase at a relatively steady rate. However, the acquisition of DNA methylation in IVC + 4d epiblast was much faster (Fig. 5c) and was closely accompanied by sharp upregulation of *Dnmt3b* (Supplementary Fig. 9d). These data indicate the presence of a default and progressive methylation-patterning process that is accelerated by dramatic upregulation of *Dnmt* genes (especially *Dnmt3b*) preferentially in epiblast. We noted that the methylation patterns of IVC + 4d epiblast and IVC + 4d VE largely recapitulated those of their in vivo counterparts, both in a chromosome-wide analysis (Supplementary Fig. 10a) and in gene bodies (Supplementary Fig. 10b). We observed compartment-dependent methylation patterns in early-stage IVC embryos as well (especially IVC + 2d embryos) (Supplementary Fig. 10a). Notably, compared with their counterparts in vivo (both E5.5 and E6.5), IVC + 4d epiblast and VE showed higher global methylation overall (Fig. 5c). In addition, a detailed analysis showed that aberrant hypermethylation in IVC + 4d epiblast was located preferentially in DMVs and CGIs (Supplementary Fig. 10c), which raises the possibility that these regions are highly sensitive to environmental changes. Taken together, these data indicate that a similar mechanism may govern de novo methylation and lineage-specific methylation patterning both in vivo and in vitro, and that this mechanism does not strictly require implantation.

## Discussion

Lineage segregation during pre- and postimplantation development gives rise to the earliest fate-committed cell types and the founder tissues for complete body development. These events also provide models for studying cell fate determination from naive pluripotency to primed states for differentiation[49]. However, the transcription circuitry and epigenetic regulation in these processes in vivo remain poorly understood. Here, by using several complementary approaches with carefully dissected early lineages, we obtained a comprehensive view of transcriptome, methylome, and 3D chromatin organization during early lineage specification. Our work identified extensive stage-specific and lineage-specific patterning of DNA methylomes during the initial cell fate commitment. Lineage-specific methylomes were particularly evident for embryonic and extraembryonic tissues. It is tempting to speculate that such differential methylomes may provide an epigenetic barrier not only between embryonic and extraembryonic tissues in the fetus, but also between extraembryonic fetal tissues and maternal tissues. It is likely that methylome patterning is regulated by multiple factors. First, we found that transcription-dependent gene body methylation exists in both embryonic and extraembryonic lineages, which suggests that it is an evolutionarily conserved mechanism[50]. However, gene body methylation is relatively transient in embryonic tissues, as regions beyond active gene bodies also become methylated eventually, probably as a result of highly active DNMTs. Second, amid global de novo methylation, DMVs were unexpectedly demethylated in epiblast but not in VE, via a process that involves the TET proteins. As DMVs are preferentially located near promoters of developmental genes and transcription factors[26], the hypomethylation of DMVs may be essential to maintain the plasticity of developmental regulators for rapid response to signals. Finally, both demethylation and de novo methylation in early development were strongly correlated with higher-order chromatin structure. We speculate that chromatin higher structure may regulate the accessibility of DNMTs and regulators of demethylation, especially when their availability is limited. Importantly, compartment-wide PMDs are also a hallmark for cancer and immortalized cell lines[51,52]. It would be interesting to investigate whether the presence of PMDs in these cells is also attributable to downregulation of DNMTs. These data show that de novo methylation seems to be a pervasive process regulated by inherited methylation from previous stages, lineage-specific expression of DNA methylation machinery, gene activity, and 3D chromatin organization (Supplementary Fig. 10d). Finally, a recent study reported that it is likely that the differential methylation patterning between embryonic and extraembryonic tissues is driven by WNT and FGF signaling[53]. Taken together, our results provide an unprecedented view of transcription circuitry and epigenetic landscapes in early lineage specification. Investigation of this molecular architecture and its highly dynamic reprogramming should help researchers decipher the regulatory foundation for initial cell fate commitment and body plan in mammalian development.

## Methods

Methods, including statements of data availability and any associated accession codes and references, are available at https://doi.org/10.1038/s41588-017-0003-x.

Received: 11 July 2017; Accepted: 1 November 2017;
Published online: 4 December 2017

## References


1. Rossant, J. & Tam, P. P. Emerging asymmetry and embryonic patterning in early mouse development. *Dev. Cell* **7**, 155–164 (2004).
2. Zernicka-Goetz, M., Morris, S. A. & Bruce, A. W. Making a firm decision: multifaceted regulation of cell fate in the early mouse embryo. *Nat. Rev. Genet.* **10**, 467–477 (2009).
3. Rossant, J. & Tam, P. P. Blastocyst lineage formation, early embryonic asymmetries and axis patterning in the mouse. *Development* **136**, 701–713 (2009).
4. Bielinska, M., Narita, N. & Wilson, D. B. Distinct roles for visceral endoderm during embryonic mouse development. *Int. J. Dev. Biol.* **43**, 183–205 (1999).
5. Arnold, S. J. & Robertson, E. J. Making a commitment: cell lineage allocation and axis patterning in the early mouse embryo. *Nat. Rev. Mol. Cell Biol.* **10**, 91–103 (2009).
6. Lawson, K. A., Meneses, J. J. & Pedersen, R. A. Clonal analysis of epiblast fate during germ layer formation in the mouse embryo. *Development* **113**, 891–911 (1991).
7. Smith, Z. D. & Meissner, A. DNA methylation: roles in mammalian development. *Nat. Rev. Genet.* **14**, 204–220 (2013).
8. Bird, A. DNA methylation patterns and epigenetic memory. *Genes Dev.* **16**, 6–21 (2002).
9. Bourc'his, D., Xu, G. L., Lin, C. S., Bollman, B. & Bestor, T. H. Dnmt3L and the establishment of maternal genomic imprints. *Science* **294**, 2536–2539 (2001).
10. Branco, M. R. et al. Maternal DNA methylation regulates early trophoblast development. *Dev. Cell* **36**, 152–163 (2016).
11. McGraw, S. et al. Loss of DNMT1o disrupts imprinted X chromosome inactivation and accentuates placental defects in females. *PLoS Genet.* **9**, e1003873 (2013).
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. Guo, H. et al. The DNA methylation landscape of human early embryos. *Nature* **511**, 606–610 (2014).
15. Smith, Z. D. et al. DNA methylation dynamics of the human preimplantation embryo. *Nature* **511**, 611–615 (2014).
16. Gao, F. et al. De novo DNA methylation during monkey pre-implantation embryogenesis. *Cell Res.* **27**, 526–539 (2017).
17. Nagy, A., Gertsenstein, M., Vintersten, K. & Behringer, R. Separating postimplantation germ layers. *CSH Protoc.* http://dx.doi.org/10.1101/pdb.prot4368 (2006).
18. Beddington, R. S. P. Isolation, culture and manipulation of post-implantation mouse embryos. In: M. Monk ed. *Mammalian Development: A Practical Approach* (pp. 43–69. IRL Press, Oxford, UK, 1987).
19. Kwon, G. S., Viotti, M. & Hadjantonakis, A. K. The endoderm of the mouse embryo arises by dynamic widespread intercalation of embryonic and extraembryonic lineages. *Dev. Cell* **15**, 509–520 (2008).
20. Peng, X. et al. TELP, a sensitive and versatile library construction method for next-generation sequencing. *Nucleic Acids Res.* **43**, e35 (2015).
21. Hon, G. C. et al. Epigenetic memory at embryonic enhancers identified in DNA methylation maps from adult mouse tissues. *Nat. Genet.* **45**, 1198–1206 (2013).
22. 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).
23. Hu, Y. G. et al. Regulation of DNA methylation activity through *Dnmt3L* promoter methylation by Dnmt3 enzymes in embryonic development. *Hum. Mol. Genet.* **17**, 2654–2664 (2008).
24. He, Y. & Ecker, J. R. Non-CG methylation in the human genome. *Annu. Rev. Genomics Hum. Genet.* **16**, 55–77 (2015).
25. Pastor, W. A. et al. Naive human pluripotent cells feature a methylation landscape devoid of blastocyst or germline memory. *Cell Stem Cell* **18**, 323–329 (2016).
26. Xie, W. et al. Epigenomic analysis of multilineage differentiation of human embryonic stem cells. *Cell* **153**, 1134–1148 (2013).
27. Jeong, M. et al. Large conserved domains of low DNA methylation maintained by Dnmt3a. *Nat. Genet.* **46**, 17–23 (2014).
28. Auclair, G., Guibert, S., Bender, A. & Weber, M. Ontogeny of CpG island methylation and specificity of DNMT3 methyltransferases during embryonic development in the mouse. *Genome Biol.* **15**, 545 (2014).
29. Schroeder, D. I. et al. Early developmental and evolutionary origins of gene body DNA methylation patterns in mammalian placentas. *PLoS Genet.* **11**, e1005442 (2015).
30. Lieberman-Aiden, E. et al. Comprehensive mapping of long-range interactions reveals folding principles of the human genome. *Science* **326**, 289–293 (2009).
31. Du, Z. et al. Allelic reprogramming of 3D chromatin architecture during early mammalian development. *Nature* **547**, 232–235 (2017).
32. Dixon, J. R. et al. Topological domains in mammalian genomes identified by analysis of chromatin interactions. *Nature* **485**, 376–380 (2012).
33. Gu, T. P. et al. The role of Tet3 DNA dioxygenase in epigenetic reprogramming by oocytes. *Nature* **477**, 606–610 (2011).
34. Peat, J. R. et al. Genome-wide bisulfite sequencing in zygotes identifies demethylation targets and maps the contribution of TET3 oxidation. *Cell Reports* **9**, 1990–2000 (2014).
35. Amouroux, R. et al. De novo DNA methylation drives 5hmC accumulation in mouse zygotes. *Nat. Cell Biol.* **18**, 225–233 (2016).
36. Stadler, M. B. et al. DNA-binding factors shape the mouse methylome at distal regulatory regions. *Nature* **480**, 490–495 (2011).
37. Burger, L., Gaidatzis, D., Schübeler, D. & Stadler, M. B. Identification of active regulatory regions from DNA methylation data. *Nucleic Acids Res.* **41**, e155 (2013).
38. Vierstra, J. et al. Mouse regulatory DNA landscapes reveal global principles of *cis*-regulatory evolution. *Science* **346**, 1007–1012 (2014).
39. McLean, C. Y. et al. GREAT improves functional interpretation of *cis*-regulatory regions. *Nat. Biotechnol.* **28**, 495–501 (2010).
40. Schöler, H. R., Dressler, G. R., Balling, R., Rohdewohld, H. & Gruss, P. Oct-4: a germline-specific transcription factor mapping to the mouse t-complex. *EMBO J.* **9**, 2185–2195 (1990).
41. DeVeale, B. et al. Oct4 is required ~E7.5 for proliferation in the primitive streak. *PLoS Genet.* **9**, e1003957 (2013).
42. Iwafuchi-Doi, M. et al. Transcriptional regulatory networks in epiblast cells and during anterior neural plate development as modeled in epiblast stem cells. *Development* **139**, 3926–3937 (2012).
43. Ang, S. L. et al. The formation and maintenance of the definitive endoderm lineage in the mouse: involvement of HNF3/forkhead proteins. *Development* **119**, 1301–1315 (1993).
44. Bossard, P. & Zaret, K. S. GATA transcription factors as potentiators of gut endoderm differentiation. *Development* **125**, 4909–4917 (1998).
45. Kuo, C. T. et al. GATA4 transcription factor is required for ventral morphogenesis and heart tube formation. *Genes Dev.* **11**, 1048–1060 (1997).
46. Shen, Y. et al. A map of the *cis*-regulatory sequences in the mouse genome. *Nature* **488**, 116–120 (2012).
47. Libbus, B. L. & Hsu, Y. C. Sequential development and tissue organization in whole mouse embryos cultured from blastocyst to early somite stage. *Anat. Rec.* **197**, 317–329 (1980).
48. Morris, S. A. et al. Dynamics of anterior-posterior axis formation in the developing mouse embryo. *Nat. Commun.* **3**, 673 (2012).
49. Kalkan, T. & Smith, A. Mapping the route from naive pluripotency to lineage specification. *Philos. Trans. R. Soc. Lond. B Biol. Sci.* **369**, 20130540 (2014).
50. Baubec, T. et al. Genomic profiling of DNA methyltransferases reveals a role for DNMT3B in genic methylation. *Nature* **520**, 243–247 (2015).
51. Berman, B. P. et al. Regions of focal DNA hypermethylation and long-range hypomethylation in colorectal cancer coincide with nuclear lamina-associated domains. *Nat. Genet.* **44**, 40–46 (2011).
52. Lister, R. et al. Human DNA methylomes at base resolution show widespread epigenomic differences. *Nature* **462**, 315–322 (2009).
53. Smith, Z. D. et al. Epigenetic restriction of extraembryonic lineages mirrors the somatic transition to cancer. *Nature* **549**, 543–547 (2017).
54. Zylicz, J. J. et al. Chromatin dynamics and the role of G9a in gene regulation and enhancer silencing during early mouse development. *eLife* **4**, e09571 (2015).
55. Heinz, S. et al. Simple combinations of lineage-determining transcription factors prime *cis*-regulatory elements required for macrophage and B cell identities. *Mol. Cell* **38**, 576–589 (2010).


## Acknowledgements

We are grateful to members of the Xie laboratory for helpful comments during preparation of the manuscript. We thank J. Na for critical reading of the manuscript. This work was supported by the National Key R&D Program of China (2016YFC0900301 to W. Xie; 2017YFC1001401 to L.L.), the National Basic Research Program of China (2015CB856201 to W. Xie), the National Natural Science Foundation of China (31422031 to W. Xie), the THU-PKU Center for Life Sciences (W. Xie), Beijing Advanced Innovation Center for Structural Biology (W. Xie), and the Biomedical Research Council of A*STAR (Agency for Science, Technology and Research), Singapore (F.X.). W. Xie is a Howard Hughes Medical Institute (HHMI) International Research Scholar. J.W. was funded by grants from the NIH (R01GM095942 and R21HD087722) and the Empire State Stem Cell Fund through

the New York State Department of Health (NYSTEM) (C028103 and C028121), and is a recipient of an Irma T. Hirschl and Weill-Caulier Trusts Career Scientist Award.

## Author contributions

Y.Z. and Q.Y. developed and conducted STEM-seq experiments. Y.X. dissected mouse tissues from embryos in vivo, carried out in vitro culture of embryos, and conducted RNA-seq. Z.D. conducted Hi-C experiments and related analysis. Z.Z. and L.L. advised on embryo lineage dissection. X.P. and F.X. advised on the development of STEM-seq. Y.L. and Q.W. conducted high-throughput sequencing. Y.Z. and Y.X. carried out data analysis. Q.W., W.Z., and W. Xia helped with the generation of *Tet1/2* double-knockout mice. J.M., M.F., and J.W. helped with various experiments and/or advised the project. Y.Z. and W. Xie wrote the manuscript.

## Competing interests

A patent for STEM-seq has been filed (2014104662612 China and PCT/CN2015/088680).

## Additional information

**Supplementary information** is available for this paper at https://doi.org/10.1038/s41588-017-0003-x.

**Reprints and permissions information** is available at www.nature.com/reprints.

**Correspondence and requests for materials** should be addressed to.W. Xie.

**Publisher's note**: Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

## Methods

**Embryo collection.** For collection of E3.5 and E4.0 tissues, 6-week-old C57BL/6N female mice were injected with pregnant mare serum gonadotropin followed by human chorionic gonadotropin before being mated with DBA/2N male mice. The first day that a vaginal plug was observed was considered as E0.5.

Fertilized embryos were flushed out from the uterus with HEPES-buffered CZB medium at defined times. Immunosurgery was performed as reported previously[56] to remove TE and isolate ICM. Briefly, after pronase treatment to remove the zona pellucida, blastocysts were incubated with DMEM containing rabbit anti-mouse serum (1:10) for 30 min and then washed three times in DMEM plus 10% FBS. The resulting embryos were exposed to guinea pig complement (1:5 in DMEM) for 10 min, washed three times, and then pipetted under microscopy to carefully remove TE cells. We separated TE from blastocysts by manual bisection to collect the opposite part of ICM as described previously[57]. The derivatives of TE at later stages were not investigated because of the difficulty of cleanly separating them from maternal tissues after embryo implantation.

E5.5–E7.5 tissues were collected via previously described methods[17,58]. Briefly, female mice were mated naturally, and the first day that the vaginal plug was observed was considered as E0.5. After embryos were dissected from uterus and decidua, they were transferred into a dish containing DMEM plus 10% FBS to remove the Reichert's membrane using syringe needles. Embryonic regions were separated from extraembryonic tissues and transferred into pancreatic and trypsin enzyme solution at room temperature for 2–10 min. For E5.5 and E6.5 embryos, we obtained VE by gently sucking the embryonic part into a capillary pipet two or three times, which detached the VE from the embryo and isolated the rest of the embryonic region as epiblast. To dissect the three germ layers from E7.5 embryo, first we collected the endoderm similarly as for the VE. Next, glass needles were inserted parallel to the PS to cut off both mesoderm wings. Finally, the J-shaped PS was cut off from the lateral side of the ectoderm where the mesoderm attached, and the rest was collected as ectoderm.

*Tet1*$^{+/-}$ mice (B6;129S4-*Tet1*$^{tm1.1Jae}$/J) and *Tet2*$^{-/-}$ mice (B6;129S-*Tet2*$^{tm1.1Iaai}$/J) were purchased from The Jackson Laboratory. After mating *Tet1*$^{+/-}$;*Tet2*$^{-/-}$ heterozygotes, we collected E6.5 epiblast from embryonic regions from *Tet1*/*Tet2* DKO embryos as described above. Extraembryonic regions were used for genotyping.

In vitro culture of mouse embryos was carried out as previously described[48,59]. Briefly, 6-week-old C57BL/6N female mice were injected with hormone and mated with DBA/2N male mice. E4.0 embryos were flushed out of the uterus with HEPES-buffered CZB medium and cultured for 4 d in a 35-mm Falcon plastic dish that contained 2 ml of CMRL 1066 supplemented with 1 mM glutamine, 1 mM sodium pyruvate, and 20% FBS. As described previously, the embryonic region was cut off with a needle and subjected to trypsin and pancreatic enzyme digestion followed by mechanical dissection to separate epiblast from VE.

**STEM-seq library preparation and sequencing.** The detailed STEM-seq procedure is described as below.

(1) Early lineage samples were lysed with 10 µl of lysis buffer (10 mM Tris-HCl, pH 7.4, 10 mM NaCl, 3 mM $MgCl_2$, 0.1 mM EDTA, pH 8.0, NP-40 0.5%) and 1 µl of protease K (Roche; 10910000) for at least 3 h at 55 °C. The reaction was then heat-inactivated for 1 h at 72 °C. After lysis, spike-in λ-DNA (Promega; D150A) was added at a mass ratio of 1/200. The reaction (20 µl) was then treated with 1 µl of dsDNA Fragmentase (NEB; M0348AA) for 30 min.

(2) The digested DNA was directly treated with bisulfite conversion reagent in a 140-µl reaction with the EpiTect Fast Bisulfite Conversion Kit (Qiagen; 59824) according to a modified protocol: denature for 8 min at 95 °C, incubate at 60 °C for 25 min, and repeat the procedure.

(3) The converted DNA was subjected to column purification and desulfonation on MinElute DNA spin columns (Qiagen; 59824) with carrier RNA (Qiagen; 59824) according to the manufacturer's instructions. The purified DNA was eluted in 30 µl of elution buffer.

(4) The converted DNA was then subjected to TELP library preparation as previously described[20]. Specifically, 28 µl of purified converted DNA was mixed with 1 µl of 1 mM dCTP and 1 µl of ExTaq buffer, incubated at 95 °C for 1 min, and then quickly cooled on ice. 1 µl of terminal transferase (NEB; M0315L) was then added to the mix, and the reaction was incubated at 37 °C for 30 min, after which 1 µl of 1 mM dATP was added and the mixture was incubated at 37 °C for 5 min before being inactivated at 80 °C for 20 min. This was followed by DNA extension with a poly-G-containing primer. Specifically, the previous reaction (~30 µl) was added to a 30-µl reaction mixture (6.2 µl of $ddH_20$, 12 µl of 5 × KAPA buffer A, 5 µl of 2.5 mM dNTP, 6 µl of 2 µM poly G primer, and 0.8 µl of KAPA 2 G polymerase (KE5507)). This was followed by a PCR reaction: 95 °C for 3 min, (47 °C for 1 min, 68 °C for 2 min) × 16 cycles, 72 °C for 10 min, and a pause at 4 °C until the next step. The mixture was digested with 2 µl of Exonuclease I (NEB; M0293L) and 6 µl of Exonuclease I buffer at 37 °C for 50 min and was inactivated at 80 °C for 10 min. A one-third volume (23 µl) of 4 × B&W buffer (40 mM Tris-HCl, pH 8.0, 2 mM EDTA, 4 M NaCl) was then added to the 69 µl of digested mixture. Each reaction system was supplemented with 10 µl of prewashed streptavidin beads (prewashed with 1 × B&W buffer three times and resuspended with 10 µl of 1 × B&W buffer). The mixture with beads was mixed in a Thermomixer (Eppendorf) at 1,400 r.p.m. (5 s on, 10 s off) at 23 °C for 30 min, and beads were washed once with 120 µl of 1 × B&W buffer and three times with 120 µl of EBT buffer (10 mM Tris-HCl, pH 8.0, 0.02% Triton X-100). The washed beads were resuspended with 20 µl of ligation reaction, including 8.4 µl of EB buffer (10 mM Tris-HCl, pH 8.0), 0.6 µl of 10 µM TA adaptor, 10 µl of 2 × Quick ligase buffer, and 1 µl of Quick ligase (NEB; M2200L). The ligation mixture was rotated at 4 °C overnight, then moved to room temperature for 10 min and washed with 120 µl of EBT buffer three times. DNA was eluted in 35.5 µl of $H_2O$ in a Thermomixer at 66 °C (1,400 r.p.m., 5 s on, 10 s off for 30 min). The eluted 35.5 µl of DNA was added to the 14.5-µl reaction mixture (5 µl of 2.5 mM dNTP, 5 µl of 10 × ExTaq buffer, 0.5 µl of ExTaq (RR006), 2 µl of 20 µM P1_FL and 2 µl of 20 µM index primer). This was followed by a PCR reaction: 95 °C for 3 min, (95 °C for 30 s, 58 °C for 30 s, 72 °C for 30 s) × 12 cycles, 72 °C for 10 min, and a pause at 4 °C until the next step. The resulting libraries were size-selected with AMPure XP according to the manufacturer's standard protocol and subjected to deep sequencing.

**RNA-seq library preparation and sequencing.** Total RNAs from various lineages isolated from E5.5–E7.5 embryos were extracted with the RNeasy Plus micro kit (Qiagen; 74034) according to the manufacturer's protocol. For ICM and TE, cells were directly lysed in hypotonic lysis buffer without RNA extraction (Amresco; M334). The cDNA libraries were then generated via the Smart-seq2 method[60]. After reverse transcription reaction with oligo-dT primers and preamplification, cDNAs were sheared by Covaris and subjected to Illumina TruSeq library preparation. All libraries were sequenced on an Illumina HiSeq 1500 according to the manufacturer's instructions.

**Generation of sisHi-C library and sequencing.** sisHi-C libraries were produced as described[31]. Briefly, samples were cross-linked with 1% formaldehyde at room temperature for 10 min. Formaldehyde was quenched with glycine for 10 min at room temperature. After being washed with 1 × PBS, the embryos were lysed on ice and the chromatin was solubilized with 0.5% SDS. The nuclei were digested with MboI at 37 °C overnight. After fill-in with biotin-14-dCTP, the fragments were ligated in a small volume. Reversal of cross-linking, DNA purification, and sonication were done sequentially. The biotin-labeled DNA was then pulled down with Dynabeads MyOne Streptavidin C1 (Life Technology). The fragments that included a ligation junction were subjected to Illumina library preparation. Fourteen cycles of PCR amplification were performed with Extaq (Takara), and the products were purified and size-selected with AMPure XP beads. All libraries were sequenced on an Illumina HiSeq 1500 according to the manufacturer's instructions.

**STEM-seq data processing.** All STEM-seq datasets were mapped to the mm9 reference genome by BSSeeker2[61]. Because STEM-seq libraries contain poly C in the ends of reads, we used scripts to remove poly G from the beginning of read 2 for paired-end mapping. Alignments were performed with the following parameters in addition to the default parameters: --bt2-p 8 --XS 0.2,3 --a CCCCCC --m 4. Multi-mapped reads and PCR duplicates were removed. We also removed the reads marked by BSSeeker2 as unconverted (--XS 0.2, 3) and reads with mapped region lengths shorter than 30 bp. After validating the reproducibility between replicates, we pooled data from replicates for subsequent analyses.

**Quantification of CG and CH methylation.** For each CG site, the methylation level was calculated as the total methylated counts (combining Watson and Crick strands) divided by the total counts across all reads covering that CG. Because the CH site is usually asymmetrical, CH methylation was calculated separately for each strand. The bisulfite conversion error rate was subtracted from the CG or CH methylation level. If the methylation value was less than the error rate, the methylation value for that site was set as 0.

**Allele assignment of sequencing reads.** To generate strain-specific genomes by considering SNP information, we downloaded SNP tables for the DBA/2J and C57BL/6N strains from the Sanger Institute Mouse Genome Project. We generated DBA/2J and C57BL/6N genomes by substituting corresponding bases from the mm9 genome. Please note that because we used the DBA/2N strain instead of the DBA/2J strain, we verified the identity of the strain by sequencing its genome. The genomes of DBA/2N and DBA/2J are very similar, and 99.4% of SNPs identified in the DBA/2N strain (compared with the reference genome) were the same as those found in the DBA/2J strain identified by the Sanger Institute.

To minimize the mapping bias introduced by the two parental alleles, we aligned all STEM-seq reads to the genomes of the C57BL and DBA strains separately with BSseeker2.8, using the following parameters: --bt2-p 8 --XS 0.2,3 --a CCCCCC --m 4. SNP information from both reads in the pair was summed and used. If the SNP contained a cytosine, its bisulfite-converted form (T) was also considered. SNPs that became non-informative (i.e., could not be distinguished from the opposite allele) after bisulfite conversion were discarded. When multiple SNPs were present in a read (or a read pair), the parental origin was determined by votes from all SNPs, and the read was assigned to the allele that received at least two-thirds of the total votes.

**RNA-seq data processing.** RNA-seq reads were mapped to the mm9 reference genome by TopHat (version 2.0.11)[62]. Cufflinks (version 2.0.2)[62] was used to calculate the gene expression levels, with the refFlat database from the UCSC Genome Browser used as reference.

**Hi-C data processing.** Sequencing reads were mapped, processed, and iteratively corrected with HiC-Pro as described previously[63,64]. Briefly, the read pairs were mapped to the mm9 reference genome in a two-step approach with bowtie2[65]. Then the invalid read pairs including dangling ends, self-circle ligation, and duplicates were discarded. The genome was divided into bins of specific lengths to generate the contact maps. We used 100-kb and 40-kb bins to investigate global chromatin contacts and local domain contacts, respectively. Hi-C interaction heat maps were generated with the normalized interaction maps with HiCPlotter[66]. We carried out A/B compartment segmentation with a 100-kb interaction matrix using a previously described method[30]. After validating the reproducibility between replicates for each cell type, we pooled data from replicates for subsequent analyses.

**Validation of STEM-seq and RNA-seq datasets.** To compare MethylC-seq and STEM-seq or to make comparisons between STEM-seq replicates, we calculated the average methylation values for 2-kb bins across the entire genome. Bins that had values in both samples were selected, and the Pearson correlation was calculated between samples or replicates. For RNA-seq samples, the Spearman correlation coefficients were calculated for FPKM values across all genes in the genome between replicates.

**Hierarchical clustering of DNA methylomes.** The average methylation value was calculated in a 1-kb window for the entire genome for each tissue/cell type. Hierarchical clustering was done with Cluster 3.0[67] with the parameter --e 2 (Pearson correlation). Java Treeview was used to visualize the clustering result. The methylomes of somatic tissues were obtained from a previous study[21].

**Identification of differentially methylated CG sites.** CG sites covered by at least five reads were selected[13]. Two-tailed Fisher's exact test was performed to evaluate the significance of differentially methylated CG sites between two stages. Only CG sites with $P < 0.05$ and changes in CG methylation levels between two stages greater than 0.2 were identified and used for downstream analyses.

**CH methylation motif analysis.** The CH sites covered by at least ten reads were sorted by their methylation levels. The top 5,000 sites were selected, and sequences within ±5 bp around the CH sites were subjected to a motif analysis with Weblogo3.0[68].

**Identification of differentially methylated promoters between E3.5 ICM and TE, and between E6.5 Epi and VE.** First we calculated the methylation levels between two samples for each gene promoter (transcription start site within 500 bp). Those genes with promoter methylation levels greater than 0.35 in one sample and twofold greater than those in the other sample were identified as differentially methylated promoters.

**Analysis of differentially methylated regions between VE and epiblast.** We first identified differentially methylated CG sites between VE and epiblasts as described above. Then we identified differentially methylated bins (2-kb) containing at least three differentially methylated CG sites. These bins were further merged into differentially methylated regions if they were no more than 2 kb away. To determine the genomic distribution of hypermethylated regions, we segmented the genome into transcription start sites, exons, introns, transcription end sites, and intergenic regions using annotations combining the RefSeq, UCSC Known Gene, Ensemble, and GENCODE databases. To assess the significance of hypermethylated regions falling into a certain category, we generated a set of random regions with lengths equal to those of each individual hypermethylated region. The numbers of regions that fell into each category were calculated, and the significance was computed as the log ratio of observed numbers divided by those for random regions.

**Analysis of DMVs.** The DMVs were identified as previously described[26]. DAVID[69] was used for Gene Ontology analysis for DMV genes.

**Identification of gene-dense regions and gene deserts.** The genome was split into 1-Mb bins, and genes located in each bin were counted. Gene-dense or gene-desert regions were identified as those with more than ten genes or no more than one gene in each bin, respectively.

**Identification of PMDs and HMDs for VE.** The PMDs and HMDs in VE were identified as previously described[52]. We calculated the average methylation level for each 10-kb bin, and included only bins with at least 20 CpGs. Because of the different global methylation levels for different cell types, we used different cutoffs for PMD and HMD identification. Specifically, hypomethylated bins (mCG/CG ≤ 0.3 for E5.5 VE and mCG/CG ≤ 0.4 for E6.5 VE) and hypermethylated bins (mCG/CG ≥ 0.6 for E5.5 VE and mCG/CG ≥ 0.7 for E6.5 VE) were identified and merged into PMDs and HMDs, respectively. We also excluded the promoter regions (±2.5 kb) for PMDs.

**Identification of allelically expressed genes.** To minimize the mapping bias introduced by the sequence differences between the two parental alleles, we aligned all sequencing reads to the genomes of the C57BL/6N and DBA/2J strains (mm9) separately. We examined all SNPs with high-quality base-calling (Phred score ≥ 30) and assigned each read to its parental origins. Only SNP information from both paired reads was retained. If multiple SNPs were present in a read, we determined the parental origin that received at least two-thirds of the total votes from all SNPs. The assigned reads mapped to exons were quantified by Htseq-count[70]. Allele-specific genes were identified on the basis of at least threefold change between the numbers of reads assigned to maternal or paternal alleles with $P < 10^{-3}$.

**Identification of topologically associated domains.** We used a directionality index and a hidden Markov model (HMM) to identify topologically associating domains (TADs) as previously described[32]. We used a 40-kb bin resolution and 2-Mb window size to calculate the directionality index score. We defined TAD boundaries as the middle bin (40 kb) between two consecutive TADs identified by HMM with distances of no more than 400 kb.

**Identification of compartments A and B.** Compartments A and B were identified as described previously[30], with several modifications. For each stage, we used normalized 100-kb interaction matrices in this analysis. Bins that had no interactions with any other bins were removed, and the expected interaction matrices were generated via a previously described window sliding approach[71] (bin size, 400 kb; step size, 100 kb). The resulting correlation matrices were subjected to principal component analysis. Principal component 1 of the correlation matrix and the gene density of genome mm9 were used to generate compartments A and B.

***P*(*s*) curve analysis.** The $P(s)$ curve was calculated as previously described[72], using 100-kb-resolution normalized interaction matrices. First, we used 1.15 as an increasing factor to generate logarithmically spaced bins (100 kb, 100 kb × 1.15, 100 kb × $1.15^2$, and so on). Next, for each bin we counted all the numbers of interactions in the corresponding distances. To calculate the probability ($P(s)$), we divided the total numbers of interactions generated in the last step for each bin by the total number of possible region pairs. Finally, the $P(s)$ values were normalized to enable the sum over the range of the distances to be 1.

**Analyses of LMRs, UMRs and tissue-specific LMRs.** The methylomes of E6.5 epiblast and E7.5 germ layer samples were segmented with an HMM as previously described[36,37]. UMRs, LMRs, and fully methylated regions were identified accordingly. LMRs that were unique to a lineage were identified as tsLMRs. The functional enrichment for genes near tsLMRs was analyzed with the GREAT tool[39]. HOMER[55] was used to identify potential transcription factor motifs in LMRs.

**Identification of early embryo enriched UMRs/LMRs and their predicted target genes.** We first combined all UMRs and LMRs identified in the five early lineages (E6.5 epiblast, E7.5 ectoderm, E7.5 PS, E7.5 mesoderm, and E7.5 endoderm). We then selected those regions with lower methylation levels in early lineages (average mCG/CG ≤ 0.4) and higher methylation levels (mCG/CG ≥ 0.5) in at least two-thirds of total somatic tissues (≥8). Regions overlapping annotated promoters (RefSeq) (within 2.5 kb) were identified as promoter UMRs/LMRs, and the rest were classified as distal UMRs/LMRs. To identify the possible gene targets of distal UMRs/LMRs, we examined all genes within 200 kb of each UMR/LMR and calculated the Spearman correlation between methylation values and expression levels for each UMR/LMR–gene pair across all early lineages and somatic tissues. UMR/LMR–gene pairs that showed strong negative correlation ($R < -0.4$) were selected for downstream analysis as previously described[73].

**Life Sciences Reporting Summary.** Further information on experimental design is available in the Life Sciences Reporting Summary.

**Data availability.** All sequencing data, including the STEM-seq, MethylC-seq, RNA-seq, and sisHi-C datasets, are available through the Gene Expression Omnibus (GEO) under accession GSE76505.

## References


56. Solter, D. & Knowles, B. B. Immunosurgery of mouse blastocyst. *Proc. Natl. Acad. Sci. USA* **72**, 5099–5102 (1975).
57. Ohnishi, Y. et al. Cell-to-cell expression variability followed by signal reinforcement progressively segregates early mouse lineages. *Nat. Cell Biol.* **16**, 27–37 (2014).
58. Harrison, S. M., Dunwoodie, S. L., Arkell, R. M., Lehrach, H. & Beddington, R. S. Isolation of novel tissue-specific genes from cDNA libraries representing the individual tissue constituents of the gastrulating mouse embryo. *Development* **121**, 2479–2489 (1995).
59. Libbus, B. L. & Hsu, Y. C. Changes in S-phase associated with differentiation of mouse embryos in culture from blastocyst to early somite stage. *Anat. Embryol. (Berl.)* **159**, 235–244 (1980).

60. Picelli, S. et al. Full-length RNA-seq from single cells using Smart-seq2. *Nat. Protoc.* **9**, 171–181 (2014).
61. Guo, W. et al. BS-Seeker2: a versatile aligning pipeline for bisulfite sequencing data. *BMC Genomics* **14**, 774 (2013).
62. Trapnell, C. et al. Differential gene and transcript expression analysis of RNA-seq experiments with TopHat and Cufflinks. *Nat. Protoc.* **7**, 562–578 (2012).
63. Servant, N. et al. HiC-Pro: an optimized and flexible pipeline for Hi-C data processing. *Genome Biol.* **16**, 259 (2015).
64. Imakaev, M. et al. Iterative correction of Hi-C data reveals hallmarks of chromosome organization. *Nat. Methods* **9**, 999–1003 (2012).
65. Langmead, B. & Salzberg, S. L. Fast gapped-read alignment with Bowtie 2. *Nat. Methods* **9**, 357–359 (2012).
66. Akdemir, K. C. & Chin, L. HiCPlotter integrates genomic data with interaction matrices. *Genome Biol.* **16**, 198 (2015).
67. de Hoon, M. J., Imoto, S., Nolan, J. & Miyano, S. Open source clustering software. *Bioinformatics* **20**, 1453–1454 (2004).
68. Crooks, G. E., Hon, G., Chandonia, J. M. & Brenner, S. E. WebLogo: a sequence logo generator. *Genome Res.* **14**, 1188–1190 (2004).
69. Huang, W., Sherman, B. T. & Lempicki, R. A. Bioinformatics enrichment tools: paths toward the comprehensive functional analysis of large gene lists. *Nucleic Acids Res.* **37**, 1–13 (2009).
70. Anders, S., Pyl, P. T. & Huber, W. HTSeq—a Python framework to work with high-throughput sequencing data. *Bioinformatics* **31**, 166–169 (2015).
71. Dixon, J. R. et al. Chromatin architecture reorganization during stem cell differentiation. *Nature* **518**, 331–336 (2015).
72. Naumova, N. et al. Organization of the mitotic chromosome. *Science* **342**, 948–953 (2013).
73. Bell, R. E. et al. Enhancer methylation dynamics contribute to cancer plasticity and patient mortality. *Genome Res.* **26**, 601–611 (2016).

nature research

Corresponding author(s): Wei Xie

☐ Initial submission ☐ Revised version ☒ Final submission

# Life Sciences Reporting Summary

Nature Research wishes to improve the reproducibility of the work that we publish. This form is intended for publication with all accepted life science papers and provides structure for consistency and transparency in reporting. Every life science submission will use this form; some list items might not apply to an individual manuscript, but all fields must be completed for clarity.

For further information on the points included in this form, see Reporting Life Sciences Research. For further information on Nature Research policies, including our data availability policy, see Authors & Referees and the Editorial Policy Checklist.

## Experimental design

1. Sample size

   Describe how sample size was determined. NA

2. Data exclusions

   Describe any data exclusions. NA

3. Replication

   Describe whether the experimental findings were reliably reproduced. The findings were reproduced in two biological replicates.

4. Randomization

   Describe how samples/organisms/participants were allocated into experimental groups. NA

5. Blinding

   Describe whether the investigators were blinded to group allocation during data collection and/or analysis. NA

Note: all studies involving animals and/or human research participants must disclose whether blinding and randomization were used.

6. Statistical parameters

For all figures and tables that use statistical methods, confirm that the following items are present in relevant figure legends (or in the Methods section if additional space is needed).

| n/a | Confirmed |
|---|---|
| ☐ | ☒ The exact sample size (*n*) for each experimental group/condition, given as a discrete number and unit of measurement (animals, litters, cultures, etc.) |
| ☐ | ☒ A description of how samples were collected, noting whether measurements were taken from distinct samples or whether the same sample was measured repeatedly |
| ☐ | ☒ A statement indicating how many times each experiment was replicated |
| ☐ | ☒ The statistical test(s) used and whether they are one- or two-sided (note: only common tests should be described solely by name; more complex techniques should be described in the Methods section) |
| ☐ | ☒ A description of any assumptions or corrections, such as an adjustment for multiple comparisons |
| ☐ | ☒ The test results (e.g. *P* values) given as exact values whenever possible and with confidence intervals noted |
| ☐ | ☒ A clear description of statistics including central tendency (e.g. median, mean) and variation (e.g. standard deviation, interquartile range) |
| ☐ | ☒ Clearly defined error bars |

*See the web collection on statistics for biologists for further resources and guidance.*

## Software

Policy information about availability of computer code

7. Software

| | |
|---|---|
| Describe the software used to analyze the data in this study. | See Supplementary Information, "Data Analyses" subsection |

For manuscripts utilizing custom algorithms or software that are central to the paper but not yet described in the published literature, software must be made available to editors and reviewers upon request. We strongly encourage code deposition in a community repository (e.g. GitHub). *Nature Methods* guidance for providing algorithms and software for publication provides further information on this topic.

## Materials and reagents

Policy information about availability of materials

8. Materials availability

| | |
|---|---|
| Indicate whether there are restrictions on availability of unique materials or if these materials are only available for distribution by a for-profit company. | NA |

9. Antibodies

| | |
|---|---|
| Describe the antibodies used and how they were validated for use in the system under study (i.e. assay and species). | NA |

10. Eukaryotic cell lines

| | |
|---|---|
| a. State the source of each eukaryotic cell line used. | NA |
| b. Describe the method of cell line authentication used. | NA |
| c. Report whether the cell lines were tested for mycoplasma contamination. | NA |
| d. If any of the cell lines used are listed in the database of commonly misidentified cell lines maintained by ICLAC, provide a scientific rationale for their use. | No |

## Animals and human research participants

Policy information about studies involving animals; when reporting animal research, follow the ARRIVE guidelines

11. Description of research animals

| | |
|---|---|
| Provide details on animals and/or animal-derived materials used in the study. | See Supplementary Information, "Embryo collection" subsection |

Policy information about studies involving human research participants

12. Description of human research participants

| | |
|---|---|
| Describe the covariate-relevant population characteristics of the human research participants. | NA |