Dynamic epigenomic landscapes during early lineage specification in mouse embryos

Yu Zhang¹, Yunlong Xiang¹,², Qiangzong Yin¹, Zhenhai Du¹,², Xu Peng³, Qijun Wang¹,², Miguel Fidalgo⁴, Weikun Xia¹,², Yuanyuan Li¹, Zhen-ao Zhao⁵, Wenhao Zhang¹,², Jing Ma¹, Feng Xu³,⁶, Jianlong Wang⁴, Lei Li⁵ and Wei Xie¹,²*  

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 methylation patterns 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.
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. 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. 2a). Next, we profiled high-depth methylomes for early lineages (with 2–3 replicates) by STEM-seq (Supplementary Table 2). We first focused on CG methylation. The methylome data generally showed excellent replication 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. 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. 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 and mouse embryonic stem cells (mESCs) (Supplementary Fig. 1c). We found that E3.5 ICM, E3.5 TE, and E4.0 ICM, which are all hypomethylated, clustered together away from 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 pro-
genitor 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. 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 placenta21 (Supplementary Fig. 3d), which indicates that DNA methylation is involved in maintaining these lineage regulators but not in its silencing. 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 auto-regression through promoter methylation at E6.521 (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 DNMT122. CH methylation in early embryos preferentially occurs in TACAG sequences (Fig. 1e and data not shown), similar to what is observed in embryonic stem cells24. 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 demethylation21. We asked whether
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 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 (Fig. 2a). The parental methylomes quickly became symmetrical between epiblast and VE. AUCLAIR ET AL.28

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 study28 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 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 shown28.

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).

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 × 10−3) and transcription factors (P = 7.41 × 10−7), including the Hox genes (Hoxb2, Hoxb3, and Hoxd12), Nkx2-5, Nkx2-6, Pdmd4, 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) or DNA methylation canyons. Using a previously described approach, 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 (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 oxidation 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. 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. Compartments A and B generally match open chromatin domains with high gene densities and closed chromatin domains with low gene densities, respectively. We asked whether the PMDs in VE correlate with such chromatin compartments. Using sHi-C, a low-input Hi-C method (Methods), we investigated higher-order chromatin organization for E3.5 ICM, 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 (Fig. 4b). This was also true for ‘topological domains’ (Fig. 4b) defined by directional index (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, we analyzed a published methylome comparing wild-type and Tet3 knockout zygotes. 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. By contrast, the demethylation on the maternal allele seemed to be relatively uniform, enabling the inheritance of an oocyte methylome pattern to blastocysts. The allele-specific compartment-correlated methylation 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. We therefore sought to identify UMRs and LMRs in early embryos as previously described. In total, we identified 17,204–17,898 UMRs and 24,039–32,019 LMRs in ectoderm, PS, mesoderm, endoderm, and...
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. Furthermore, large fractions of UMRs and LMRs in E6.5 epiblast (94% and 58%, respectively) overlapped with DNase hypersensitivity sites in mESCs (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 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. 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 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\(^6\), 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\(^4\). 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\(^{34–36}\). 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\(^{38–40}\), 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 Dmnt3b (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 methyleome. Notably, mouse embryos can grow through the early stages of organogenesis in vitro\(^7\). Thus, we isolated E4.0 embryos in vivo; cultured them in vitro using established protocols\(^{33–35}\); 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 Dmnt3b (Supplementary Fig. 9d). These data indicate the presence of a default and progressive methylation-patterning process that is accelerated by dramatic upregulation of Dmnt genes (especially Dmnt3b) 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\(^8\). 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 methylation processes during the initial cell fate commitment. Lineage-specific methylation was particularly evident for embryonic and extraembryonic tissues. It is tempting to speculate that such differential methylation 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 methylene 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\(^9\). However, gene body methylation is relatively transient in embryonic tissues, as regions beyond active gene bodies also become demethylated 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\(^6\), 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\(^34–36\). 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\(^9\). 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.
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 CBZ medium at defined times. Immunosurgery was performed as reported previously26 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 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. 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 methods26,27. 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 decidual, 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 a whole. To dissect the germ layers from the 5-cell embryo, we coiled the endoderm similarly as for the VE. Next, glass needles were inserted parallel to the PS to cut off both mesoderm wings. Finally, the I-shaped PS was cut off from the lateral side of the ectoderm where the mesendoderm attached, and the rest was collected as ectoderm.

Tet1−/− mice (B6;129S4-Tet1tm1Jw) and Tet2−/− mice (B6;129S-‐Tet2tm2Jw) were purchased from the Jackson Laboratory. After mating Tet1−/−; Tet2−/− heterozygotes, we collected E6.5 epithlast 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 described27,28. 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 CBZ 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 epithast 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 MgCl2, 0.1 mM EDTA, pH 8.0, NP-40 0.5%) for at least 30 min at 37 °C. The reaction was 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; M0348A) for 30 min.

(2) The digested DNA was directly treated with bisulfite conversion reagent in a 1:40 ratio reaction with the EpiTect Fast Bisulfite Conversion Kit (Qiagen; 59824) with carrier RNA. After reverse transcription reaction with oligo-dT primers and preamplification, cDNAs were sheared by Covaris and subjected to Illumina TrueSeq library preparation. cDNAs were sequenced on an Illumina HiSeq 1500 according to the manufacturer’s instructions.

(3) 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. After embryos were dissected from uterus and decidual, 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 a whole. To dissect the germ layers from the 5-cell embryo, we coiled the endoderm similarly as for the VE. Next, glass needles were inserted parallel to the PS to cut off both mesoderm wings. Finally, the I-shaped PS was cut off from the lateral side of the ectoderm where the mesendoderm attached, and the rest was collected as ectoderm.

Tet1−/− mice (B6;129S4-Tet1tm1Jw) and Tet2−/− mice (B6;129S-‐Tet2tm2Jw) were purchased from the Jackson Laboratory. After mating Tet1−/−; Tet2−/− heterozygotes, we collected E6.5 epithlast from embryonic regions from Tet1/Tet2 DKO embryos as described above. Extraembryonic regions were used for genotyping.

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 ICN 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 method29. After reverse transcription reaction with oligo-dT primers and preamplification, cDNAs were sequenced on an Illumina TrueSeq library preparation. All libraries were sequenced on an Illumina HiSeq 1500 according to the manufacturer’s instructions.

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 BSeeaker2.8, using the following parameters: -b 2 -p 8 -X 0.2 -a 0.95 -X CCCCCC -m 4. Multi-mapped reads and PCR duplicates were removed. We also removed the reads marked by BSeeaker2 as unconverted (-X 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.

STEM-seq data processing. All STEM-seq datasets were mapped to the mm9 reference genome by BSeeaker2. Because STEM-seq libraries contain poly C in the ends of reads, we used scripts to remove poly C from the beginning of read 2 for paired-end mapping. Alignments were performed with the following parameters in the default parameters: -b 2 -p 8 -X 0.2 -a 0.95 -X CCCCCC -m 4. Multi-mapped reads and PCR duplicates were removed. We also removed the reads marked by BSeeaker2 as unconverted (-X 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 BSeeaker2.8, using the following parameters: -b 2 -p 8 -X 0.2 -a 0.95 -X CCCCCC -m 4. SNPs 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.

© 2017 Nature America Inc., part of Springer Nature. All rights reserved.
Hi-C data processing. Sequencing reads were mapped, processed, and iteratively corrected with HiC-Pro as described previously. Briefly, the raw reads were mapped to the mm9 reference genome in a two-step approach with bowtie2. 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 local chromatin contacts and local domain contacts, respectively. Hi-C interaction heat maps were generated with the normalized interaction maps with HiCPlotter. We carried out A/B compartment segmentation with a 100-kb interaction matrix using a previously described method. 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 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.

Identification of differentially methylated CG sites. CG sites covered by at least five reads were selected. 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 Weblogo 3.0.

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 described previously. DAVID 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 PDMs and HMDs for VE. The PDMs and HMDs in VE were identified as previously described. 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 PDM 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 PDMs and HMDs, respectively. We also excluded the promoter regions (±2.5 kb) for PDMs.

Identification of allatively 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/6J and DBA/2 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 Hseq-count. 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 < 10−5.

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. 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 (40kb) 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, 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 (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, 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; 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. 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. HOLDER 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 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 (28). Regions overlapping annotated promoters (ReSeq) (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.

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 ssHi-C datasets, are available through the Gene Expression Omnibus (GEO) under accession GSE76505.

References
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.
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**

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**

11. **Description of research animals**

Provide details on animals and/or animal-derived materials used in the study.

See Supplementary Information, "Embryo collection" subsection

12. **Description of human research participants**

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

NA