



ARTICLE

Distinct enhancer signatures in the mouse gastrula delineate progressive cell fate continuum during embryo development

Xianfa Yang¹, Boqiang Hu², Jiaoyang Liao¹, Yunbo Qiao³, Yingying Chen¹, Yun Qian¹, Su Feng¹, Fang Yu¹, Ji Dong², Yu Hou², He Xu¹, Ran Wang¹, Guangdun Peng^{1,4,5,6}, Jinsong Li^{1,7}, Fuchou Tang^{2,8,9} and Naihe Jing^{1,6,7}

Primary germ layers have the potential to form all tissues in the mature organism, and their formation during gastrulation requires precise epigenetic modulation of both proximal and distal regulatory elements. Previous studies indicated that spatial and temporal patterns of gene expression in the gastrula predispose individual regions to distinct cell fates. However, the underlying epigenetic mechanisms remain largely unexplored. Here, we profile the spatiotemporal landscape of the epigenome and transcriptome of the mouse gastrula. We reveal the asynchronous dynamics of proximal chromatin states during germ layer formation as well as unique gastrula-specific epigenomic features of regulatory elements, which have strong usage turnover dynamics and clear germ layer-specific signatures. Importantly, we also find that enhancers around organogenetic genes, which are weakly expressed at the gastrulation stage, are frequently pre-marked by histone H3 lysine 27 acetylation (H3K27ac) in the gastrula. By using the transgenic mice and genome editing system, we demonstrate that a pre-marked enhancer, which is located in the intron of a brain-specific gene *2510009E07Rik*, exhibits specific enhancer activity in the ectoderm and future brain tissue, and also executes important function during mouse neural differentiation. Taken together, our study provides the comprehensive epigenetic information for embryonic patterning during mouse gastrulation, demonstrates the importance of gastrula pre-marked enhancers in regulating the correct development of the mouse embryo, and thus broadens the current understanding of mammalian embryonic development and related diseases.

Cell Research (2019) 29:911–926; <https://doi.org/10.1038/s41422-019-0234-8>

INTRODUCTION

During mammalian embryo development, one fertilized zygote will undergo multiple rounds of cell fate commitment to form diverse cell types.¹ The gastrulation, which occurs at the early post-implantation stage, is one of the most crucial biological processes for the generation of the primary germ layers.^{2–4} During gastrulation, the pluripotent epiblast develops into the ectoderm, mesoderm, and endoderm, which set up the future body plan.³ Previous knowledge has revealed that regionalized cell distribution in the gastrula accounts for the formation of primary germ layers.^{2,4,5} Specifically, anterior epiblast cells develop into ectodermal cells, which give rise to the future nervous system and skin⁶; posterior epiblast cells develop into mesendoderm cells, which give rise to visceral organs⁷; while epiblast cells, which traverse the primitive streak and then migrate anteriorly, develop into mesodermal cells and mainly give rise to heart and blood cells.⁸

Multiple layers of molecular mechanisms including epigenetic regulation⁹ and signaling pathways¹⁰ are involved to orchestrate

the precise embryonic development. Recently, a number of studies have determined the dynamics of epigenetic programming during different stages of the mouse embryo development. Epigenomic landscape studies have revealed the features of histone H3 lysine 4 tri-methylation (H3K4me3) and H3 lysine 27 tri-methylation (H3K27me3) re-establishment and the enrichment of H3K4me3 broad domain at key cell identity-related genes,^{11,12} determined the dynamics of chromatin accessibility,¹³ chromatin structures,¹⁴ and histone H3 lysine 9 tri-methylation (H3K9me3) reprogramming^{15–17} in pre- and early post-implantation embryos. Other studies also provided the global chromatin state landscape during late embryo development.^{18,19} These findings together highlight the importance of epigenetic regulations in early mouse embryonic patterning. However, the knowledge of epigenetic dynamics during mouse gastrulation particularly for the generation of spatially-restricted germ layers remains largely elusive.

Recently, we and others have reported the distinct distributions of H3K27me3 and DNA methylation at a range of

¹State Key Laboratory of Cell Biology, CAS Center for Excellence in Molecular Cell Science, Shanghai Institute of Biochemistry and Cell Biology, Chinese Academy of Sciences, University of Chinese Academy of Sciences, 200031 Shanghai, China; ²Beijing Advanced Innovation Center for Genomics, College of Life Sciences, Peking University, 100871 Beijing, China; ³Precise Genome Engineering Center, School of Life Sciences, Guangzhou University, Guangzhou 510006 Guangdong, China; ⁴CAS Key Laboratory of Regenerative Biology and Guangdong Provincial Key Laboratory of Stem Cell and Regenerative Medicine, Guangzhou Institutes of Biomedicine and Health, Chinese Academy of Sciences, Guangzhou 510530 Guangdong, China; ⁵Guangzhou Regenerative Medicine and Health Guangdong Laboratory (GRMH-GDL), Guangzhou 510005 Guangdong, China; ⁶Institute for Stem Cell and Regeneration, Chinese Academy of Sciences, 100101 Beijing, China; ⁷School of Life Science and Technology, ShanghaiTech University, 201210 Shanghai, China; ⁸Ministry of Education Key Laboratory of Cell Proliferation and Differentiation, 100871 Beijing, China and ⁹Peking-Tsinghua Center for Life Sciences, Peking University, 100871 Beijing, China

Correspondence: Yunbo Qiao (y bqiao@gzhu.edu.cn) or Jinsong Li (jsli@sibcb.ac.cn) or Fuchou Tang (tangfuchou@pku.edu.cn) or Naihe Jing (njing@sibcb.ac.cn)

These authors contributed equally: Xianfa Yang, Boqiang Hu, Jiaoyang Liao, Yunbo Qiao

Received: 6 January 2019 Accepted: 29 August 2019

Published online: 7 October 2019

development-related genes in intraembryonic and extraembryonic cells,^{20,21} thereby explaining the differential developmental plasticity of these two types of cells. Studies have also revealed that epigenetic decommissioning, priming, or activation of regulatory elements, especially enhancers, play critical roles in the regulation of stem cell fate by modulating target gene expression or by mediating the interactions between transcription factors (TFs).^{22,23} For example, epigenetic priming of enhancers by histone H3 lysine 4 mono-methylation (H3K4me1), which facilitates the binding of pioneer factors, signifies developmental competence during human endodermal lineage diversification.²⁴ Histone H3 lysine 27 acetylation (H3K27ac)-marked cis-regulatory elements in pre-implantation embryos tend to reside near zygotic genome activation-related genes,¹² while rapid and pervasive enhancer usage turnover happens during mouse organ development.^{18,19} However, there still remains a knowledge gap concerning the epigenetic basis for the correct development of the primary germ layers during gastrulation.

To determine the dynamics of epigenetic programming and to specify the epigenetic mechanisms of primary germ layers' formation during mouse gastrulation, we collected sub-regions of the mouse gastrula for transcriptome and epigenome profiling. First, we revealed the asynchronous features of both proximal and distal chromatin state transitions among the three primary germ layers. Then, we identified that many key organogenetic gene-related enhancers are frequently pre-marked by H3K27ac at the gastrula stage. The epigenetic pre-patterning of organ development-related enhancers ensures the correct development of each germ layer towards a specific organ tissue. Finally, we uncovered the essential roles of an ectoderm- and brain-specific enhancer, *Ect2*, during mouse neural differentiation. Our study provides the first comprehensive epigenetic landscape of the mouse gastrula, reveals the *in vivo* epigenetic basis of primary germ layers' generation and future organ development, and bridges the knowledge gap of epigenetic information between pre-implantation and organogenetic embryos.

RESULTS

Epigenetic profiling of the mouse gastrula

We have previously revealed the epigenetic dynamics of DNA methylation and H3K27me3 during mouse early embryo development from the pre-streak (embryonic day 6.5, E6.5), mid-streak (E7.0), and late-streak (E7.5) stages, on both epiblasts (Epi, anterior epiblast (A), posterior epiblast (P), as well as anterior mesoderm part (AM), were separated at corresponding stages) and extra-embryonic (ExE) parts.²⁰ Here, we investigated the dynamics of H3K4me1, H3K4me3, H3K9ac, and H3K27ac modifications as well as the transcriptome in the mouse gastrula (Fig. 1a). Dissected tissues from a single embryo of each stage were subjected to RNA-seq for transcriptome profiling. Pooled samples of corresponding tissues (~10,000 cells per sample) were subjected to limited cell number chromatin immunoprecipitation followed by sequencing (lcChIP-seq)¹⁷ (Fig. 1a). All data have been prepared with two highly reproducible biological replicates (Supplementary information, Fig. S1 and Table S1). Using this strategy, we were able to trace the transcriptomic and epigenomic trajectory of ectodermal, mesendodermal, mesodermal, and extraembryonic lineages.

We first analyzed gene expression dynamics during the development of each lineage (Fig. 1b–e; Supplementary information, Table S2) and found that expression of ectodermal genes (e.g., *Tal2* and *Lhx5*), mesendodermal genes (e.g., *Cdx1* and *Wnt3a*), mesodermal genes (e.g., *Hand1* and *Myl2*), and placenta-related genes (e.g., *Etv2* and *Plac1*) was specifically up-regulated during the development of anterior, posterior, anterior mesodermal, and extraembryonic tissues, respectively (Fig. 1b–e). The formation of the primitive streak and region-specific distribution of morphogen, such as BMP4 and *Wnt3a* in the gastrula,²⁵ have

been reported to be crucial for the development of mouse embryos and related cell fate commitment. We found that both epigenomic pattern and transcriptomic profile of the primitive streak marker gene *T*, and morphogen *Bmp4* and *Wnt3a* accurately reflect the regional distribution of these factors in the gastrula (Supplementary information, Fig. S2a–c). These results robustly demonstrate the biological fidelity and validity of our sample collection strategy, which faithfully recapitulated the lineage diversification process during mouse gastrulation. We also analyzed the correlation between gene expression and epigenomic modifications at various ranges away from transcription start sites (TSSs) (Fig. 1b–e). As expected, epigenetic modifications in proximal regions (TSS \pm 2 kb) were more highly correlated with gene expression dynamics than those in the distal regions. In addition, we observed a relatively high positive correlation between gene expression and representative enhancer-related modifications (H3K4me1/H3K27ac) at both flanking (\pm 2–10 kb) and distal regions (\pm 10–100 kb) in intraembryonic lineages (Fig. 1b–d). These observations suggest that the activities of distal enhancers play important regulatory roles in gastrulation.

To systematically analyze the global dynamics of epigenetic modifications in the mouse gastrula, we counted the number of modified regions, which was calculated as the sum-up of 200-bp bins with corresponding modification signals detected by ChromHMM. We found that histone modifications underwent extensive changes in both intra- and extra-embryonic lineages (Fig. 1f–i; Supplementary information, Fig. S2d), suggesting a dramatic genome-wide remodeling of the epigenomic landscape during gastrulation. In general, histone modifications of the three intraembryonic lineages exhibited similar patterns, with fewer gain than loss of modified regions from stages E6.5 to E7.0 and more gain than loss of modified regions from stages E7.0 to E7.5 (Fig. 1f–h). We also found that H3K4me1, but not H3K4me3, exhibited the most dramatic changes during gastrulation. Given the requirement for rapid and pervasive transcriptional remodeling during lineage specification at gastrulation stage²⁶ (Fig. 1b–e), higher dynamics of H3K4me1 strongly suggests that enhancer-related regulatory mechanisms²⁷ are widely involved in the regulation of the lineage commitment during gastrulation.

By combining our data with the published dataset of H3K4me3 and H3K27me3 in pre-implantation embryos,¹¹ we detected a clear developmental trajectory from pre- to post-implantation stages (Supplementary information, Fig. S3a). Globally, the enrichment for H3K4me3 in the gastrula was highly different from that in the oocyte, but was similar to H3K4me3 marks in pre-implantation embryos (Supplementary information, Fig. S3b). Previous studies revealed that H3K4me3-marked broad domains are highly associated with the regulation of cell identity-related genes in pre-implantation embryos.^{11,12} However, the noticeable presence of H3K4me3 broad domains in pre-implantation embryos was substantially reduced in the gastrula where only medium H3K4me3 peaks distributed around the TSS regions (e.g., the *Nanog* promoter region) (Supplementary information, Fig. S3c, d).

Asynchronous cell fate commitment revealed by proximal chromatin state transition

To further illustrate the functions of histone modifications and related chromatin states for cell fate commitment during gastrulation, we trained a ten-state model²⁸ consisting of five active states (active TSSs, flanking regions of active TSSs, bi-acetylated TSSs, bi-acetylated enhancers, and active enhancers), three bivalent states (poised enhancers, bivalent enhancers, and bivalent TSSs), and two repressive states (repressed Polycomb state and quiescent chromatin) based on the combinatorial histone modification patterns observed in all analyzed samples (Fig. 2a). We found that the TSS-related chromatin regions exhibited minimal levels of DNA methylation but maintained a

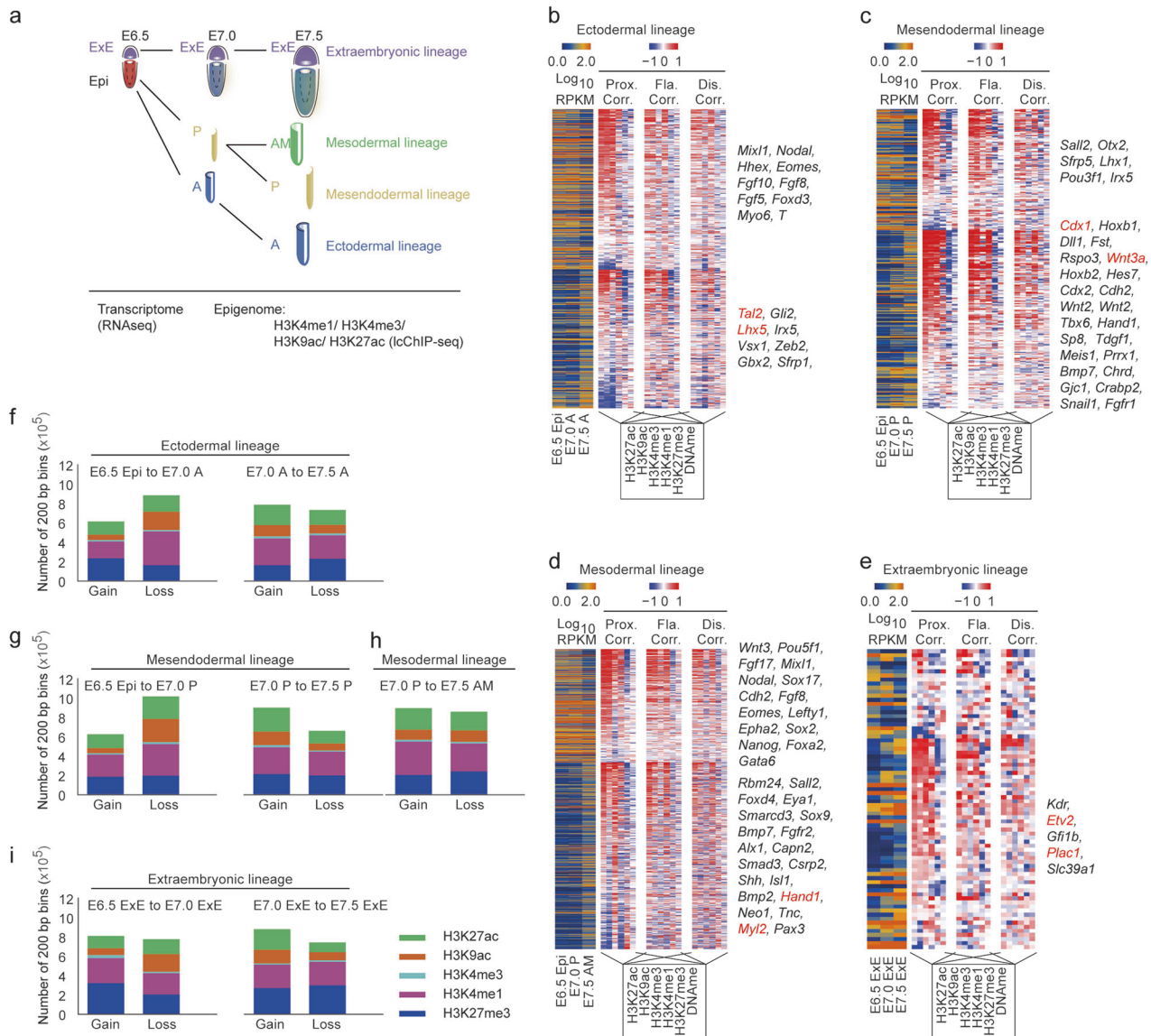
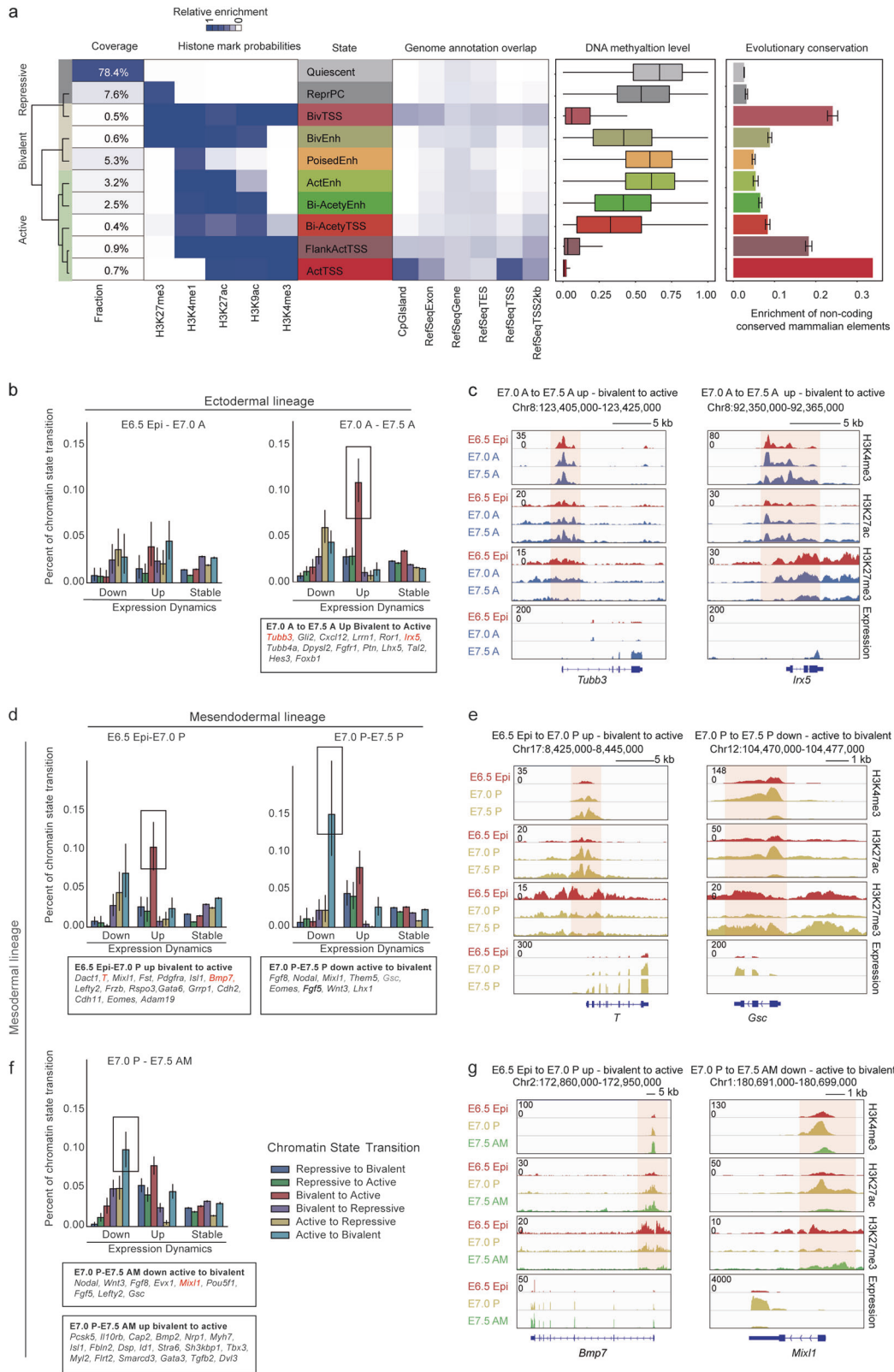


Fig. 1 Epigenomic profiling of the mouse gastrula. **a** An illustrative scheme describing the sample collection strategy for the study of mouse gastrula. **b–e** Heatmap indicating gene expression dynamics (down-regulation (top), up-regulation (bottom)), the corresponding correlation with each epigenetic modification in proximal (Prox, TSS \pm 2 kb), flanking (Fla, TSS \pm 2–10 kb), and distal regions (Dis, TSS \pm 10–100 kb) during the development of ectodermal (**b**), mesendodermal (**c**), mesodermal (**d**) and extraembryonic (**e**) lineages. Representative genes are listed on the right of each map. **f–i** Bar plot showing the number of 200-bp bins defined by ChromHMM with gain or loss of selected modification during the developmental process of the ectodermal (**f**), mesendodermal (**g**), mesodermal (**h**) and extraembryonic (**i**) lineages. H3K27me3 ChIP-seq and DNA methylation data for the mouse gastrula are from our previous report (GSE104243)

high degree of evolutionary conservation of DNA sequences (Fig. 2a). By contrast, enhancer-related chromatin regions, especially poised enhancers, displayed high levels of DNA methylation and low evolutionary conservation (Fig. 2a). Previous studies revealed that bivalent enhancers show the broad distribution of DNA methylation only in pluripotent cells but not in primary tissues,²⁹ therefore high levels of DNA methylation at bivalent enhancers in the gastrula indicate the high potential of epiblast cells for multi-lineage generation and future tissue development.

Next, we profiled the chromatin state dynamics of the proximal regions (TSS \pm 2 kb) and specific histone modification changes during gastrulation. By profiling the relationship between chromatin state in the proximal region and gene expression, we found that the up- and down-regulation of gene expression during intraembryonic lineage development are likely controlled

by the chromatin state transitions from bivalent to active and from active to bivalent, respectively (Fig. 2b–g). Surprisingly, we also found distinct patterns of chromatin state changes between the ectodermal lineage and the other two lineages. During ectoderm development, the bivalent-to-active transition, which is usually associated with gene up-regulation and stem cell fate specification,³⁰ was more prominent from the E7.0 A to E7.5 A stages than that from the E6.5 Epi to E7.0 A stages (Fig. 2b, c). In contrast, during mesendoderm and mesoderm development, bivalent-to-active transition became extrusive from as early as the E6.5 Epi to E7.0 P stages (Fig. 2d–g). These results indicate that the cell fate of the anterior epiblast might be committed much later than the posterior part, reflecting the asynchronous fate commitment between ectodermal and mesendodermal lineages in the context of epigenetic regulation. The delayed commitment of ectodermal cells may explain the notion that the epiblast stem cells are



developmentally similar to late-gastrula-stage ectoderm.³¹ Consistently, by analyzing the dynamics of H3K4me3 and H3K27me3 co-occupied domains, we also found the asynchronous feature between ectodermal and mesendodermal/mesodermal lineages (Supplementary information, Fig. S3e–g). Besides, we found that genes, such as *Gsc* and *Mixl1*, down-regulated in the

mesodermal lineage from E7.0 P to E7.5 P and in the mesodermal lineage from E7.0 P to E7.5 AM, were mainly regulated by active-to-bivalent transitions (Fig. 2d–g). These results indicate that mesendodermal/mesodermal cell fate specification occurs in the early stage of gastrulation and cell fate determination occurs thereafter.

Fig. 2 Proximal chromatin state transition reflects asynchronous cell fate commitment of intraembryonic lineages. **a** Average genomic coverage, histone mark probabilities, chromatin state definitions, genomic annotation overlap, average DNA methylation level and enrichment of non-coding mammalian conserved elements. **b, d, f** Bar plots showing the ratio of 200-bp bins with respective chromatin state changes along with gene expression changes during the development of each intraembryonic lineage at the promoter regions (ectodermal lineage (**b**); mesendodermal lineage (**d**); mesodermal lineage (**f**)). The promoter regions (TSS \pm 2 kb) were divided into twenty 200-bp bins, and bins with the same chromatin state change were summed to calculate the percentage. Representative chromatin state changes are highlighted with rectangles and related genes are also listed. **c, e, g** Representative genome browser snapshots of H3K4me3, H3K27ac and H3K27me3 at the loci of representative genes (highlighted in red in **b, d, f**). H3K27me3 ChIP-seq and DNA methylation data for the mouse gastrula are from our previous report (GSE104243)

Lineage-specific dynamics of enhancers drives embryonic patterning during mouse gastrulation

The tightly regulated activity of enhancers is crucial for stem cell differentiation,³² mammalian development,³³ and disease pathology.³⁴ Based on analyses of chromatin states (Fig. 2a), we profiled 2,504,271 bins with candidate enhancer activities in the mouse gastrula (Fig. 3a). We further clustered all enhancer bins into 11 clusters based on their distribution specificity among all samples (Fig. 3a). Using GREAT analysis, we found that lineage-specific enhancer bins were enriched near genes with coordinated biological functions and phenotypic changes (Fig. 3b, c; Supplementary information, Table S3). For example, genes around extraembryonic-specific group 1 enhancer bins (G1: cluster 2) are related to trophectodermal cell differentiation and placental development, and knockout of these genes leads to small placenta; genes neighboring posterior-specific group 2 enhancer bins (G2: clusters 4 and 5) are related to cardiac cell fate specification, hematopoietic progenitor cell specification and other processes in mesodermal development, and knockout of these genes leads to mesocardium morphology abnormality; genes neighboring anterior-specific group 3 enhancer bins (G3: cluster 6) are related to brain and neural development, and knockout of these genes leads to abnormal thalamus morphology; and genes neighboring intraembryonic-specific group 4 enhancers (G4: clusters 9 and 10) are involved in the regulation of gastrulation and embryonic development, and knockout of these genes leads to abnormalities in gastrulation (Fig. 3b, c). We also found the enrichment of TF-binding motifs in extraembryonic and intraembryonic enhancer bins (Fig. 3d; Supplementary information, Table S3). For example, binding motifs for Cdx2 and USF1/2, which are required for extraembryonic development,^{35,36} were enriched in extraembryonic-specific enhancer bins, and motifs for essential developmental TFs, such as Zic1/2,³⁷ Oct4³⁸ and TCF12,³⁹ were enriched in intraembryonic-specific enhancer bins (Fig. 3d).

Furthermore, we analyzed the rapid and pervasive changes in enhancer activities during the development of intraembryonic lineages (Supplementary information, Table S4). We found that ectodermal cells in E7.5 A shared a large proportion of enhancers with E6.5 epiblast cells, whereas mesendodermal or mesodermal enhancers were predominantly generated at the E7.0 or E7.5 stage (Fig. 4a, f, g). In the ectodermal lineage, many ectodermal enhancer bins were maintained in the active state, such as enhancers residing near *Otx2* (Fig. 4a). During the E6.5 Epi to E7.0 A stage, there was only a mild shift from the quiescent to active or poised enhancer bins, whereas during the E7.0 A to E7.5 A stage, a large number of enhancer bins shifted from quiescent to active, poised or bivalent states (Fig. 4a). Further analyses of enhancer-neighboring genes highlighted the importance of anterior-specific enhancers in ectoderm fate specification and brain development (Fig. 4a). For example, one of the enhancers, which we named as Ect1 here and is located upstream of the *Tal2* TSS, exhibited anterior-specific H3K27ac enrichment in E7.5 A (Fig. 4a, b) and intensive interactions with nearby chromatin⁴⁰ (Fig. 4c). Transgenic reporter assay showed that Ect1 was specifically activated in the anterior region of E7.5 embryos (Fig. 4d) and remained active until E9.5 in embryonic forebrain region (Fig. 4e).

For the mesendodermal lineage, we found that most mesendodermal enhancer bins were in quiescent states in E6.5 Epi, and many enhancer bins in E7.0 P became silenced at E7.5 P stage (e.g., *Fgf10*- and *Bmp10*-related enhancers), and more than half of E7.5 P active or poised enhancer bins (such as enhancers near *Mesp2* and *Hand2* genes) were de novo generated (Fig. 4f). The difference in enhancer usage at E7.0 P and E7.5 P indicates that cells in these two stages likely execute distinct developmental functions during gastrulation. As expected, the enhancers with E7.0 P active/poised but E7.5 P quiescent state were usually related to heart growth and cyclin-dependent kinase activity, whereas enhancers with E7.0 P quiescent but E7.5 P active/poised state were usually responsible for nephron epithelium morphogenesis and muscle cell proliferation (Fig. 4f).

In the mesodermal lineage, we also found that most E7.5 AM enriched enhancer bins, which are usually located near genes responsible for heart development, were quiescent at E6.5 Epi and E7.0 P stages (Fig. 4g). These remarkable distinctions of enhancer usage turnover pattern further confirm the asynchronous commitment of different cell lineages in intraembryonic tissues (Fig. 2b–g) and emphasize the distinct functions of lineage-specific enhancers for gastrulation.

Identification of gastrula-specific enhancers

Next, we probed the existence of gastrula-specific enhancer bins by comparison with data from fetal and postnatal mouse tissue samples in the Encyclopedia of DNA Elements (ENCODE) project⁴¹ (Supplementary information, Table S5) and found 491,119 gastrula-specific enhancer bins (Fig. 5a). These gastrula-specific enhancer bins were enriched near genes that are related to embryonic pattern formation, gastrulation and primary germ layer specification, and are highly expressed at the gastrulation stage, such as *Sox2*, *Eomes* and *Mesp1* (Fig. 5b–e). We also identified an enhancer, which we named as Mes1 and is located in the intron region of the *Aox2* gene, exhibited enrichment of H3K27ac in E7.5 AM but not in E11.5 or P7 heart sample (Fig. 5f). Enhancer reporter assays further validated the activity of Mes1 in E7.5 anterior mesodermal cells as well as E8.5 early primitive heart tube and artery system (Fig. 5g), indicating a potential role of the Mes1 enhancer in mouse cardiac mesoderm progenitor cell development.

Vital organ development-related enhancers are epigenetically pre-marked by H3K27ac in the gastrula

Given the important role of gastrulation in establishing the future body plan, we analyzed the dynamics of lineage-specific enhancers across early and late mouse development by comparing our dataset with the published H3K27ac datasets in the 8-cell embryo¹² and in E11.5, E14.5, E17.5, P0, P7, P21, and P56 mouse brain, liver and heart samples¹⁹ (Supplementary information, Table S6). We found that most of the active enhancer bins in the gastrula were in quiescent states in the 8-cell embryo stage (Fig. 6a–c), whereas many gastrula enhancer bins were active in organogenic embryos with constitutive or dynamic patterns (Fig. 6a–c). Further analyses revealed that these enhancer bins were usually located near crucial organogenetic genes, which exhibited higher expression in later developmental stages

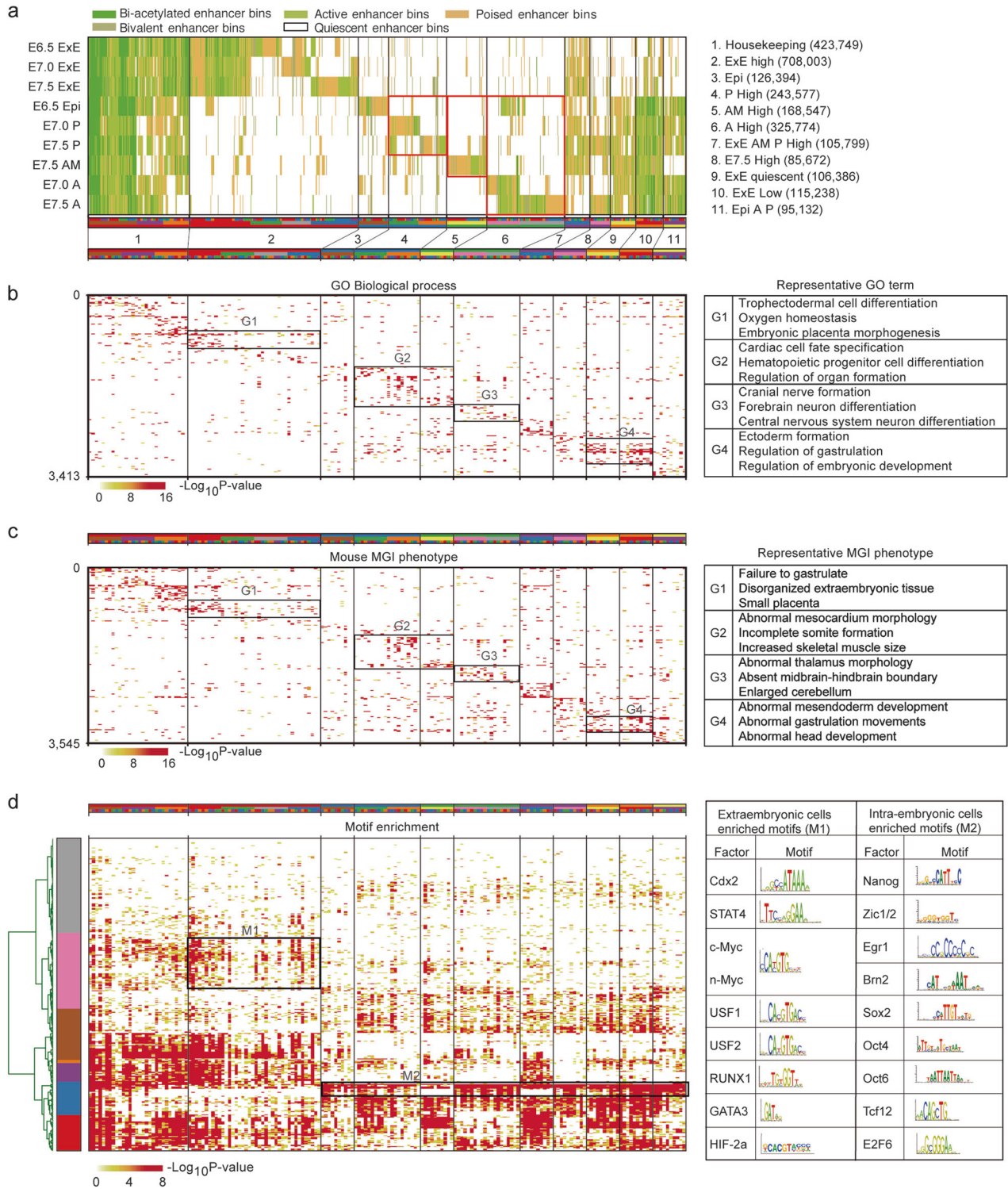
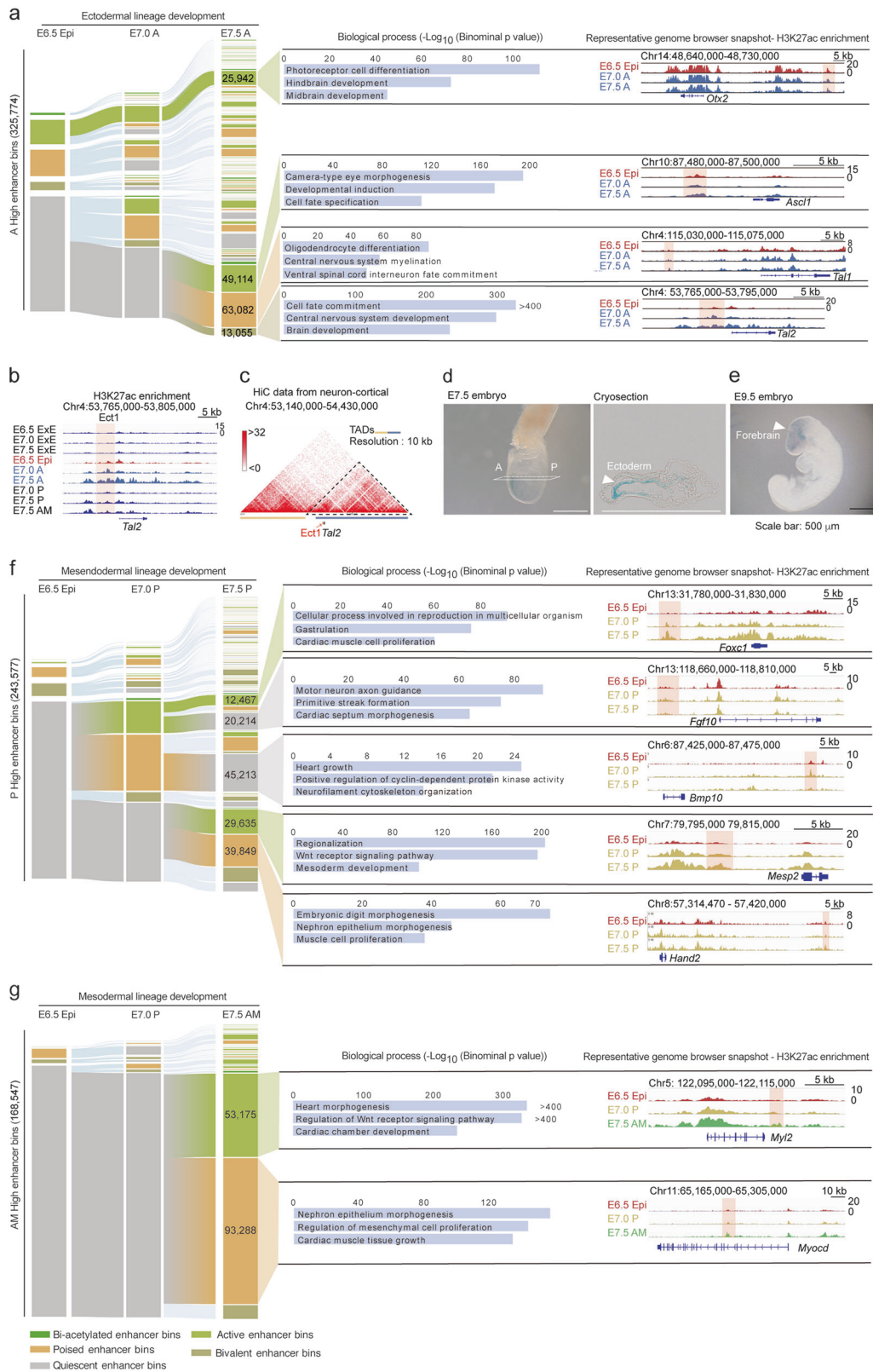


Fig. 3 Enhancer profiling of the mouse gastrula. **a** Enhancer modules based on k-means clustering of 2.5 million enhancer bins (200-bp) classified as bi-acetylated enhancer bins (dark green), active enhancer bins (green), poised enhancer bins (dark yellow) and bivalent enhancer bins (moss green) and quiescent enhancer bins across 9 mouse gastrula samples. All enhancer bins were grouped into 11 groups, 18 modules and 180 sub-modules for further functional annotation. The names of each group and module are shown. **b–d** Gene ontology (GO) analysis (**b**), mouse phenotypes (**c**) and motif enrichment (**d**) of each sub-module for enhancer bin neighboring genes. Rectangles pinpoint enrichments for selected modules with representative functions, phenotypes, and motifs. Representative GO terms, phenotypes, and motifs are listed on the right



(Fig. 6a–c; Supplementary information, Fig. S4a–h). For example, *Arx*, which plays important roles in forebrain development, has been reported to be regulated by hs122 and hs123 enhancers in the dorsal brain region.⁴² Double knockout of both enhancers, as well as single knockout of hs122 but not hs123, leads to structural and morphological defects in mouse brain.⁴² We found that, even

though there was no apparent expression of *Arx* in the gastrula, the hs122 but not hs123 enhancer was already pre-marked by H3K27ac and exhibited enhancer activity in some ectodermal cells of the gastrula (Fig. 6a). This result indicates that biologically significant, organogenesis-related enhancers are preferentially pre-marked by H3K27ac in the gastrulation stage. Similarly,

Fig. 4 Lineage-specific dynamics of enhancers drives embryonic patterning during gastrulation. **a** Alluvial plot depicting the global activity dynamics of enhancer bins during the development of ectodermal lineage. Each line represents a 200-bp bin defined based on the ChromHMM categories. Numbers of representative state enhancer bins are shown. The representative biological process terms identified using GREAT analysis are listed. Bar plots indicate the significance of corresponding terms. The representative genome browser snapshots are also shown. **b, c** Representative genome browser snapshots of H3K27ac in all samples (**b**) and chromatin interaction frequency heatmap at 10 kb resolution in mouse cortical neurons (**c**) at the range of *Tal2* locus. The *Tal2*-related enhancer Ect1 (highlighted with a red rectangle, **b**) is located in the topologically associating domain (TAD). Chromatin conformation capture HiC data are from a published dataset (**c**).³⁹ **d, e** Ect1 enhancer activity reporter expression (blue staining) in representative transgenic mouse gastrula (E7.5 stage, **d**) and organogenic embryo (E9.5 stage, **e**). A cryosection slide of the stained E7.5 embryo is also shown (**d**). All experiments were performed at least twice. **f, g** Alluvial plots depicting the global activity dynamics of enhancer bins during the development of mesendodermal (**f**) and mesodermal (**g**) lineages. Each line represents a 200-bp bin defined based on the ChromHMM categories. Numbers of representative state enhancer bins are shown. The representative biological process terms identified by GREAT analysis are listed. Bar plots indicate the significance of corresponding terms. The genome browser snapshots of representative genes are also shown

enhancers (we named as Men1, Men2, Men3, and Ect3) located nearby of *Nkx2-5*, *Notch1*, *Lbh*, or *Pax6*, which are responsible for heart/limb or brain development, exhibited epigenetic pre-marking and clear enhancer activities in future organs (Fig. 6b, c; Supplementary information, Fig. S4d, h). These data illustrate that the enhancers of genes essential for fetal organ development are often pre-marked by active epigenetic modifications in the gastrula, even though the corresponding genes are weakly expressed in this stage; and suggest that this region-specific pre-marking of enhancers in the gastrula may ensure the correct development of primary germ layers for organ development later in mouse embryogenesis.

A brain activated enhancer with ectoderm specificity plays important roles in the neural differentiation process

To further explore the biological significance of gastrula pre-marked enhancers, we identified a novel enhancer, which we named as Ect2 and is located in the intronic region of protein-coding gene *2510009E07Rik* (Fig. 7a). The genomic region of Ect2 showed enrichment of ATAC-seq signal²¹ and strong ectoderm-specific enrichment of H3K27ac in the gastrula (Fig. 7a; Supplementary information, Fig. S5a). The region maintained the epigenetic pattern in brain tissues till postnatal stages (Fig. 7a), and also exhibited intensive interactions with nearby genomic regions that were located in the same topologically associated domain (TAD)⁴⁰ (Fig. 7b). Moreover, enhancer reporter assays demonstrated that Ect2 exhibited strong enhancer activity in ectodermal cells of the gastrula and brain region-specific activity in E11.5 organogenic embryos (Fig. 7c). This result strongly implies the potential role of Ect2 enhancer in regulating mouse brain development. To address this, we took advantage of the CRISPR-Cas9-based genome editing system to delete the Ect2 enhancer fragment in mouse embryonic stem cells (Fig. 7d). Multiple single cell-derived Ect2-knockout cell lines were generated, and independent knockout cell lines were further subjected to neural differentiation as previously described⁴³ (Fig. 7d). We found that the expression of several neural progenitor markers (e.g., *Otx2* and *Nestin*) and neuronal markers (e.g., *Map2*, *Tuj1*, and *NeuN*) was significantly compromised at both mRNA and protein levels compared to the wild-type cells after 8 days of differentiation (Fig. 7e, f). We also profiled the transcriptome of Ect2-knockout cells on day 0 and day 8 of differentiation, and found that the knockout of Ect2 specifically affected the transcriptomic state of the differentiated cells but not of pluripotent stem cells (Supplementary information, Fig. S5b, c). Analysis of differentially expressed genes (DEGs) revealed that knockout of Ect2 biased the cells towards endodermal fate and impaired neural differentiation (Fig. 7g; Supplementary information, Fig. S5d, and Table S7). To further identify targets of Ect2, which strongly interacts with neighboring chromatin (Fig. 7b), we profiled the expression changes of genes within 20 Mb of Ect2 between wild-type and Ect2-knockout cells (Fig. 7h). We observed significant down-regulation of *2510009E07Rik* ($P < 0.01$), *Map3k13* ($P < 0.05$), and

Masp1 ($P < 0.01$) (Fig. 7h; Supplementary information, Fig. S5e). Considering the relatively low expression of *Map3k13* and *Masp1* and high expression of *2510009E07Rik* during neural differentiation and in the central nervous system (Supplementary information, Fig. S5e–g), we inferred that *2510009E07Rik*, a protein-coding gene with no reported function, might be a direct target of Ect2. Indeed, the expression level of *2510009E07Rik* was down-regulated at the late stage of neural differentiation (Supplementary information, Fig. S5h), specific knockdown of *2510009E07Rik* closely mimicked the transcriptomic state of Ect2-knockout cells (Fig. 7i). Moreover, re-expression of *2510009E07Rik* in Ect2-knockout cells partially rescued the neural differentiation abnormalities (Supplementary information, Fig. S5i). Therefore, we concluded that the enhancer Ect2, which acts as an active enhancer in both ectoderm and fetal brain tissues, plays important roles in neural development by regulating the expression of *2510009E07Rik*. These results clearly illustrate the importance of gastrula pre-activated enhancers in mammalian development.

DISCUSSION

Lineage specification during mouse embryogenesis is one of the most important topics for biologists, due to the diversification of cell types comprising the whole body. The establishment of stem cells with different degrees of pluripotency^{44,45} and specific primary germ layer progenitor cells^{46–49} has greatly accelerated the understanding of these processes. However, the molecular mechanisms of lineage specification in vivo remain poorly understood. With the advent of high-throughput sequencing technologies, many studies have been performed to profile the epigenetic landscape of early mammalian embryos.^{11–14,50,51} However, the epigenetic profile for histone modifications and its underlying biological significance during mouse primary germ layer formation remain largely unknown. In this study, by using limited cell number epigenome profiling technologies and careful embryo dissection, we present a comprehensive and lineage-specific epigenomic landscape of the mouse gastrula from the pre-streak to the late-streak stages. By applying multiple bioinformatics and experimental analyses, we reveal the asynchronous features of cell fate commitment in the gastrula and identify the existence of gastrula-specific enhancers. We further determine and validate the epigenetic basis for the primary germ layers' formation and related organ development through the discovery of epigenetically pre-marked enhancers in the gastrula. Taken together, our study provides a comprehensive interpretation of epigenetic regulatory mechanisms in mouse embryogenesis and greatly broadens the current understanding of mammalian development.

Chromatin state transition has been reported to be fully involved in cell fate commitment.⁵² Here, we comprehensively profile both TSS proximal and distal chromatin state transitions in the gastrula. We reveal the lineage-specific dynamics of chromatin state transitions in different primary germ layers. Our analyses of

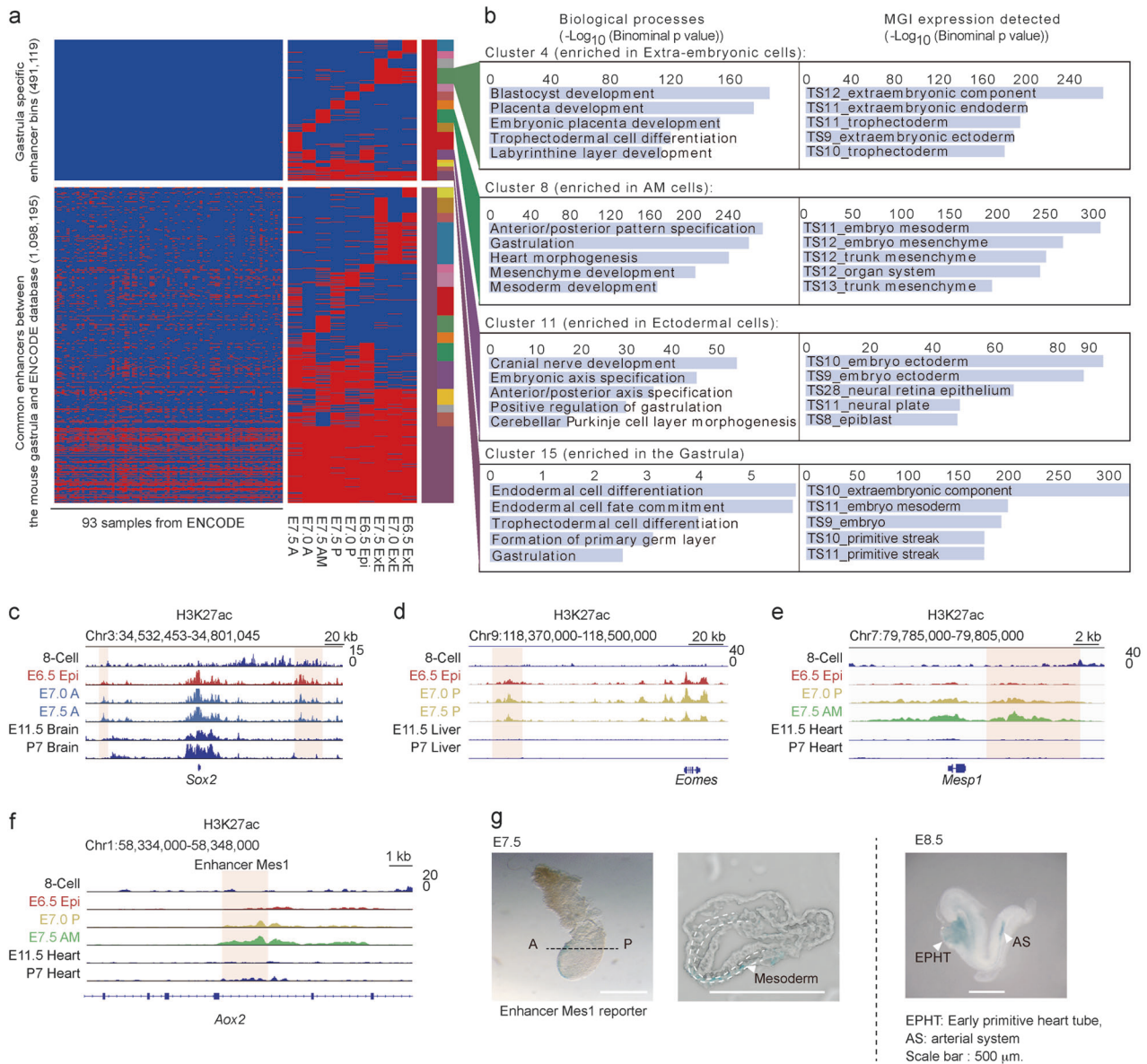


Fig. 5 Identification of gastrula-specific enhancers. **a** Heatmap showing a large number of gastrula-specific enhancer bins based on comparison with the published dataset from ENCODE. The dataset used is also listed in Supplementary information, Table S4. **b** Representative biological process terms and expression stage information for selected enhancer bin clusters based on GREAT analyses. **c–e** Representative genome browser snapshots of H3K27ac for gastrula-specific enhancers during development of the ectoderm (**c**), mesoderm (**d**), and mesoderm (**e**). **f** Representative genome browser snapshot of H3K27ac for Mes1 enhancer in the gastrula and later embryonic samples. **g** Enhancer activity reporter expression (blue staining) in representative transgenic mouse gastrula (E7.5) and E8.5 stages. A cryosection slide of the stained E7.5 embryo is also shown. H3K27ac data of E11.5 and P7 (postnatal day 7) embryo samples are from a published dataset¹⁹

proximal chromatin state transition reveal that the bivalent-to-active transition has already become prominent from stage E6.5 to stage E7.0 in mesodermal and mesodermal lineages, while this becomes dominant from stage E7.0 to stage E7.5 in ectoderm lineage. Further investigation identifies that the chromatin state transition pattern is highly related to lineage development-related genes (Fig. 2b–f), suggesting the occurrence of lineage commitment. The study of distal regulatory elements shows that E7.5 ectoderm cells (E7.5 A) still share a large number of enhancers with E6.5 Epi cells (Fig. 4a), indicating similar enhancer signature between ectoderm cells and E6.5 epiblast cells; in contrast, most enhancers in mesodermal and mesodermal cells are de novo generated at the late gastrula stages (E7.0–E7.5) (Fig. 4f, g), suggesting that distinct enhancer usage patterns exist between

mesodermal/mesodermal cells and E6.5 epiblast cells. Bioinformatic analyses reveal a strong correlation between these enhancers and lineage development. Thus, the temporal distinction of chromatin state transitions at both TSS proximal and distal regions between mesoderm/mesoderm and ectoderm strongly indicates that cell fate commitment during gastrulation is asynchronous. These findings not only support the current notion that epiblast stem cells are more similar to ectoderm,³¹ but also address the discrepancy in molecular mechanisms for cell fate commitment, which requires further exploration.

Developmental competence is a key concept in mammalian development, describing that lineage intermediate cells possess the ability to respond to inductive signals with precise timing.²⁴ The developmental competence established in primary germ

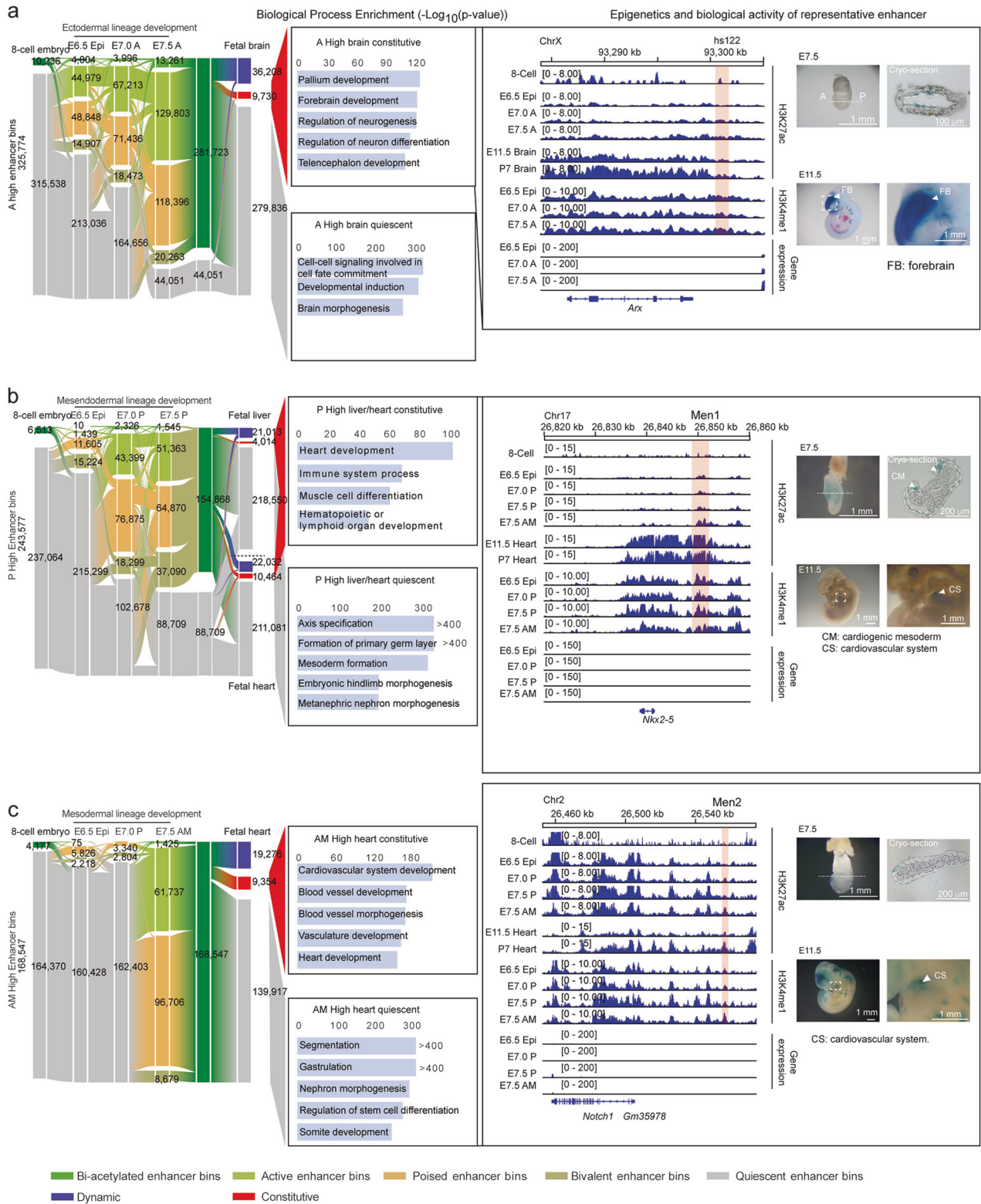
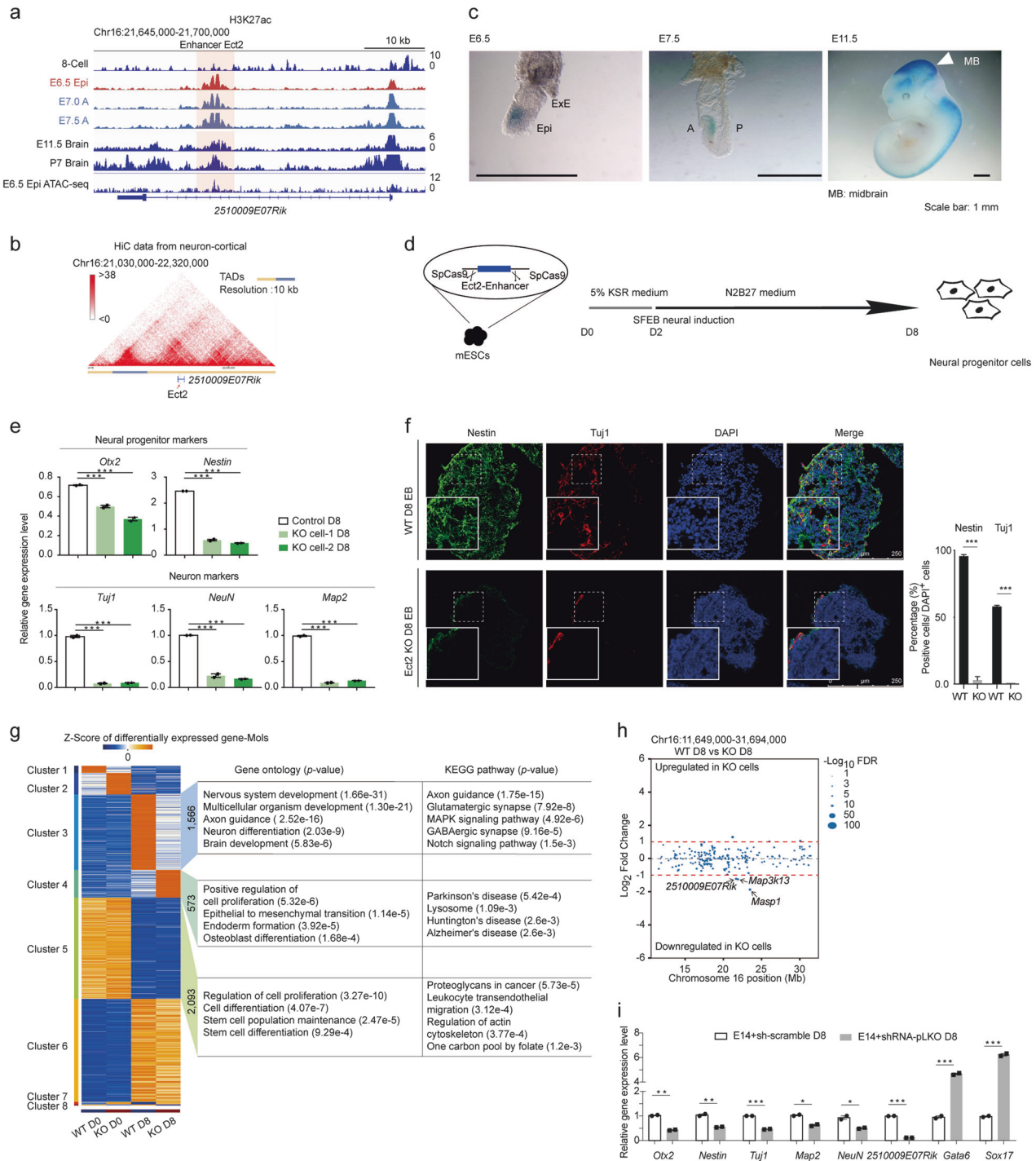


Fig. 6 Vital organ development-related enhancers are pre-marked by H3K27ac in the gastrula. **a–c** Alluvial plots revealing the global dynamics of enhancer bins during the development from the 8-cell stage embryo to primary germ layers in the gastrula and related future organs. Ectodermal lineage and fetal brain (**a**); mesendodermal lineage and fetal liver or fetal heart (**b**); mesodermal lineage and fetal heart (**c**). Numbers of enhancer bins for each state are shown. The related biological process terms of neighboring genes of selected enhancer clusters are shown in the middle panel. Genome browser snapshots of H3K27ac and H3K4me1 around representative genes are shown. Enhancer regions with relevant dynamics are highlighted with red rectangles and enhancer activity reporter expression (blue staining) for highlighted enhancer regions is also shown on the right. Enhancers for *Nkx2-5* and *Notch1* were named Men1 and Men2, respectively. H3K27ac data of E11.5 and P7 embryo samples are from a published dataset¹⁹



layers has been reported to be crucial for future body development.⁵³ Studies have revealed the important roles of the interaction among extrinsic signals, pioneer factors and chromatin states in mediating the formation of developmental competence and related mammalian development.⁵⁴ For example, enhancers primed with H3K4me1 mediate the formation of developmental competence and further facilitate the binding of pioneer factors in endodermal progenitor cells.²⁴ However, it remained elusive whether tissue-specific enhancers are also epigenetically pre-marked in vivo. Here, we profile the enhancer turnover continuum during mouse embryogenesis by combining our data with published dataset.^{12,19} Surprisingly, we find that enhancers,

residing near key organ development-related genes, are frequently pre-marked by H3K27ac in the gastrula (such as *Men1*, *Men2*, *Men3*, *Ect2*, and *Ect3*) (Fig. 6a–c). Similarly, studies based on DNA methylation profiling of the post-implantation embryos have also revealed that a number of developmental gene-related enhancers in E14.5 embryos and somatic tissues are hypomethylated or lowly methylated in the mouse gastrula.^{51,55} These results indicate that the competence for organ development has been established in the gastrula through epigenetically pre-marking. This may propose a general principle for epigenetic regulation of developmental competence in mouse embryo development. In the future, further exploration of the interactions

Fig. 7 A brain activated enhancer with ectoderm specificity plays important roles in neural differentiation. **a** A representative genome browser snapshot of H3K27ac enrichment and chromatin accessibility at the range of Ect2 enhancer from the 8-cell stage to P7 brain. H3K27ac data of E11.5 and P7 embryo samples are from a published dataset.¹⁹ Chromatin accessibility data of E6.5 epiblast is from a published dataset.²¹ **b** Interaction frequency heatmap at 10 kb resolution in mouse cortical neurons around the Ect2 enhancer. The Ect2 enhancer is located in the TAD. Chromatin conformation capture HiC data are from a published dataset.³⁹ **c** Ect2 enhancer activity reporter expression in representative transgenic mouse gastrula (E6.5 and E7.5 stages) and organogenetic mouse embryo (E11.5) (blue staining). All experiments were performed at least twice. **d** Scheme showing the workflow for Ect2 enhancer functional validation. **e** Bar plots showing the relative expression levels of neural-related genes in both wild-type (WT) cells (control) and Ect2 enhancer knockout cells (KO cell-1 and KO cell-2) on day 8 of differentiation. Quantitative gene expression level was normalized to the expression level of *Gapdh*. Results are shown as means \pm SEM. Student's *t*-test, ****P* < 0.001. **f** Immunocytochemistry analysis of Tuj1 and Nestin protein levels in mouse embryoid bodies on day 8. Quantification of the data is shown on the right. **g** Heatmap showing differentially expressed genes (DEGs) of WT cells and Ect2-knockout cells on day 0 and day 8 during mouse neural differentiation. Gene counts of each cluster, representative GO terms, and KEGG pathways are shown on the right. The color key from blue to red indicates low to high gene expression level. **h** Dot plot indicating RNA-seq results of all genes within a 20 Mb window around the Ect2 enhancer. Each dot represents one gene, and the size of each dot represents the significance for gene expression change. The red dotted line shows the threshold for DEGs. **i** Bar plot indicating gene expression pattern after the knockdown of the *251009E07Rik* gene in wild-type cells. Quantitative gene expression level was normalized to the expression level of *Gapdh*. Results are shown as means \pm SEM. Student's *t* test, **P* < 0.05, ***P* < 0.01. All experiments were performed at least three times

among extrinsic signals, pioneer factors, DNA methylation, and chromatin states at these pre-marked enhancers will provide more comprehensive explanations for developmental competence.

A recent study reported the accessibility of the regulatory genome before gastrulation of the *Drosophila* embryo by using a single-cell combinatorial indexing assay for transposase accessible chromatin with sequencing (sci-ATAC-seq).⁵⁶ In this study, we have reasonably used an embryo dissection strategy^{15,21,51} to profile the epigenomic landscape of the mouse gastrula based on current knowledge of mouse embryo pattern and current histone modification profiling technologies. In the future, with continued improvement in technologies,⁵⁷ such as the uliCUT&RUN technology,⁵⁸ epigenomic profiling of histone modifications using single cells from the gastrula will provide more thorough explanations not only for the formation of primary germ layers but also for the heterogeneities that exist at this stage.

MATERIALS AND METHODS

Animal use, embryo staging and collection

All animal experiments were performed under the guidelines of the Animal Ethical Committee of the Institute of Biochemistry and Cell Biology, Shanghai Institutes for Biological Sciences, Chinese Academy of Sciences. Plugged female mice were picked after mating and counted as embryonic day 0.5 (E0.5). Mice were euthanized when embryos developed into nominal day 6.5, 7.0 or 7.5. Embryos were removed from the implantation site as described previously.⁵⁹ Embryos were then transferred to petri dishes with phosphate buffered saline (PBS). The removal of decidua and Reichert's membrane was performed using sharpened surgeon tweezers. The developmental stage and morphology of embryos were reconfirmed according to the Theiler stage methodology. The acquired embryos were then dissected into essential parts for the following experiments.

Transcript preparation, quantitative real-time PCR, cDNA sequencing library construction and sequencing

For in vitro culture cells, bulk cells were collected, and RNAs were extracted using Trizol (Invitrogen, 15596018) and precipitated using isopropanol. RNAs were then reverse transcribed into cDNA by using FastQuant RT Super Mix (Tiangen, KR108). For embryo samples, cDNA pools were prepared following a previous protocol.⁶⁰ Quantitative real-time PCR was performed using diluted acquired cDNA with Stormstar SYBR green qPCR master mix (DBI Bioscience, DBI-143). cDNA sequencing libraries were prepared by using NEBNext Ultra II RNA Library Prep kit (NEB, E7775L) or Nextera XT DNA Library Preparation Kit (FC-131-1024) according to the manufacturer's instructions. Paired-end 125 bp sequencing was performed on a HiSeq2500 platform at the Berry

Genomics Co., Ltd (Beijing, China) and Novogene Company (Beijing, China).

Embryo ChIP sample preparation and ChIP library construction

Essential parts of embryos were collected and pooled to reach a scale of 10,000 cells per sample. The basic workflow for sample collection and chromatin immunoprecipitation (ChIP) was as previously reported.²⁰ Noticeably, for the experiment of H3K9ac and H3K27ac, 20 mM sodium butyrate was added to inhibit histone deacetylase, and in all experiments, proteinase inhibitor cocktail (Roche, 11697498001) and PMSF were added to inhibit protein degradation. In sum, we performed the ChIP assay for histone modifications with specific antibodies against H3K4me3 (Abcam, 8580), H3K4me1 (Abcam, 8895), H3K27ac (Active Motif, 39133), and H3K9ac (Millipore, 06-942). ChIP libraries were prepared using NEBNext Ultra DNA library preparation kit (NEB, E7370) according to the manufacturer's instructions.

Sequencing data quality control

All the de-multiplexed sequencing reads were first cleaned to remove any artificial sequences, such as sequencing adapters introduced during the experimental process, and reads with >10% low-quality bases were also discarded.

RNA-seq data processing

For the RNA-seq data, the annotation of the transcriptome was defined by combining the UCSC mm10 RefSeq genes and the GENCODE vM5 lncRNAs. Then the cleaned reads were mapped using TopHat (version 2.0.9)⁶¹ with default parameters. HTSeq⁶² was used to count the number of reads mapped to each annotated gene based on the mapping results. These results were further used for calculation of DEGs using DESeq2⁶³ package in Bio-conductor. Furthermore, for the quantification of RNA expression, the gene expression levels (RPKMs, reads per kilobase per million reads) of each sample were calculated based on the number of reads for each gene counted by HTSeq:

$$\text{RPKM}_{\text{gen},s} = \frac{\text{Counts}_{\text{gen},s}}{10^9 \times \text{Length}_{\text{gen}} \times \sum_{\text{gen}} \text{Counts}_{\text{gen},s}},$$

where gen for genes and s for samples.

Genes with variation above 1 among all samples were selected and the ComBat function in *sva* package of Bio-conductor⁶⁴ was used to remove the systematic error introduced by different sequencing batches. The normalized RPKMs after removing batch effects were used for the subsequent analyses, including clustering and combinational analyses with ChIP-seq results.

ChIP-seq data processing

The whole ChIP-seq analysis pipeline was available in https://github.com/kundajelab/chipseq_pipeline. In brief, for the ChIP-seq data, the trimmed reads were mapped to the mouse genome (mm10 assembly) using BWA aligner (version 0.7.5a-r405)⁶⁵ with the options “-i 15 -q 10 -t 4” and the density tracks and peaks were called by MACS2(v2.10)⁶⁶ based on the mapping results. The density track of interval *i* in each sample was further normalized by the “macs2 bdgcmp” command with the control sample.

Especially, for all biological replicates, the peaks were called using the Irreproducibility Discovery Rate (IDR) framework developed by Kundaje et al. (<https://sites.google.com/site/anshulkundaje/projects/idr>)⁶⁷ and the densities were calculated using the merged results for samples that passed the IDR framework.

For the broad domain analysis of H3K4me3, the MACS2(v2.10) was again used to call the broad peaks using the options:

–broad –f BED –g mm –n chip_output –B

Then based on the broad peaks called by MACS2, we classified the H3K4me3-marked domains into four groups based on the segment lengths at their promoter regions.¹¹ To specify, promoters overlapping with segments > 5000 bp were defined as broad H3K4me3 domain genes; those overlapping with 1000 bp < segments ≤ 5000 bp were defined as median H3K4me3 domain genes; those overlapping with 200 bp < segments ≤ 1000 bp were defined as narrow H3K4me3 domain genes; those overlapping with segments ≤ 200 bp were defined as null H3K4me3 domain genes. Stage-specific broad domain genes were defined as the broad domains that only occurred at that specific developmental stage.

For the H3K4me3 and H3K27me3 domain analysis, we called the H3K4me3 and H3K27me3 co-occupied domains across all the samples using the R package chromstaR in a joint multivariate analysis, and the parameters were set as: bin size = 1,000 bp; step size = 200 bp; mode = combinatorial. The state transitions along the developmental lineages were then plotted through alluvial diagrams in R package ggalluvial.

The DNA methylation analysis pipeline

The trimmed reads were mapped to mm10 mouse genome following our modified bisulfite sequencing pipeline.⁶⁸ PCR duplicates were removed using samtools and DNA methylation level of every cytosine was calculated based on our previous publication.⁶⁸ The pairwise comparisons of consecutive samples on a single CpG level were calculated using a b-binomial model and the b difference distribution.

The relationship between the epigenetic modifications and gene expression

The epigenetic modifications that we took into consideration include H3K27ac, H3K9ac, H3K4me3, H3K4me1, H3K27me3, and DNA methylation. The epigenetic signals of the histone markers for the ±2 kb TSS region, ±2–10 kb from TSS and ±10–100 kb from TSS were calculated by averaging the signal density around the promoter region, and DNA methylation level was calculated by averaging the methylation level of the single bases with the sequencing depth above 3 at respective ranges from TSS.

Then we examined the relationship between epigenetic modification and gene expression for each gene and each marker. In principle, the Pearson correlation value calculated using the average epigenetic track density and the gene expression level for a given gene in different samples were used to represent this relationship. However, Pearson correlation will usually be influenced when the expression level of genes or the epigenetic track density is stable between different stages. For example, if the RPKM and DNA methylation level of ±2 kb around TSS for a given

gene are like below in three continuous stages:

$$\text{RPKM} = [0.13, 0.23, 0.33], \text{DNA_Meth} = [0.11, 0.13, 0.15]$$

the Pearson correlation will be 1.0 in this case. However, as the RPKM will present a high level variation when the mean RPKM is low, a 0.1 difference between two continuous stages may not be large enough to get rid of the effect of technical errors. The same thing also influenced the DNA methylation levels. Therefore, in order to get a more accurate index for this relationship, the random jitter with Gaussian distribution was introduced into the input vector and the Pearson correlation was calculated in a bootstrap strategy. The average and standard deviation of the TSS signal for the Gaussian distribution for each epigenetic marker *m* was calculated using TSS signal in all samples and all genes:

$$\mu_m = \frac{1}{n_{\text{gen}} \times n_s} \sum_{n_{\text{gen}}}^i \sum_{n_s}^j \text{signal}_{m,i,j}$$

$$\sigma_m = \sqrt{\frac{1}{N} \sum_{n_{\text{gen}}}^i \sum_{n_s}^j (\text{signal}_{m,i,j} - \mu_m)^2}$$

However, as the standard deviation of RPKM would change with the different average level of RPKM, a k-nearest neighbor strategy was used to estimate the standard deviation of RPKM. We first calculated the average expression level and the standard deviation for each gene among all samples:

$$\mu_{\text{RPKM,gen}} = \frac{1}{n_s} \sum_{n_s}^i \text{RPKM}_{\text{gen},i}$$

$$\sigma'_{\text{RPKM,gen}} = \sqrt{\frac{1}{n_s} \sum_{n_s}^i (\text{RPKM}_{\text{gen},i} - \mu_{\text{RPKM,gen}})^2}$$

Then all ($\mu_{\text{RPKM,gen}}$, $\sigma_{\text{RPKM,gen}}$) pairs for different genes were ordered according to $\mu_{\text{RPKM,gen}}$. For a given gene, the average for the Gaussian distribution of the jitter is $\mu_{\text{RPKM,gen}}$ and the standard deviation is the average value of the standard deviation for the 20 genes nearest to the corresponding average RPKM value.

$$\sigma_{\text{RPKM,gen}} = \frac{1}{20} \sum_{i=-10}^{10} \sigma'_{\text{RPKM,gen,rank}+i}$$

Then the recalibrated Pearson correlation using the bootstrap strategy is:

$$R_{\text{recal}}(\text{gen}, m) = \frac{\sum_0^{n=100} \text{PearsonR}(\text{signal}_{\text{jitter}_m, \text{gen}}, \text{RPKM}_{\text{jitter}_m, \text{gen}})}{100}$$

where

$$\text{signal_jitter}_{m, \text{gen}} = N(\mu_m, \sigma_m)$$

$$\text{RPKM_jitter}_{m, \text{gen}} = N(\mu_{\text{RPKM,gen}}, \sigma_{\text{RPKM,gen}})$$

ChromHMM analysis

Chromatin states were identified and characterized using ChromHMM(v1.11).⁶⁹ The alignment files of H3K27ac, H3K9ac, H3K4me3, H3K4me1 and H3K27me3 modifications across 9 samples were binned into 200-bp bins using the BinarizeBam command, with the input alignment file as the control. Next, we trained the model with 8, 10, 12, 15 and 18 emission states at 200-bp resolution and default parameters using the LearnModel command. The 15-state result was selected, and we then merged

states that shared similar patterns with other states and finally obtained 10 states (Fig. 2a). The enrichment level of non-coding mammalian-conserved elements (GERP) for each state was annotated using the files downloaded from http://mendel.stanford.edu/SidowLab/downloads/gerp/mm9.GERP_scores.tar.gz and then lifted over to mm10. To describe the chromatin states in a more intuitive way, the emission matrix for the 9 states (except Quiescent state) was clustered using R command:

```
hcl_state <- hclust(as.dist((1 - cor(t(mat_Emission), method
= "pearson"))/2), method = "median")
```

Clustering tree was then cut into three sub-trees. These three resulting sub-trees were then concluded as repressive state, bivalent state and active state (Fig. 2a), which were further used in the following analyses.

Dynamics of enhancers

We analyzed the dynamics of enhancers using the results of ChromHMM. Four states among the 10 states were assumed to imply enhancers, including bivalent enhancer bins, poised enhancer bins, active enhancer bins and bi-acetylated enhancer bins. The matrix used 1 for "Bivalent enhancer bin", 2 for "Poised enhancer bin", 3 for "Active enhancer bin" and 4 for "Bi-acetylated enhancer bin", with 9 samples for columns and 2.5 million 200-bp tiles for rows, which have at least one enhancer bin that existed using the segment results of ChromHMM (Fig. 3a). We used a two-step K-means clustering strategy to cluster this matrix. First, the 2.5 million of 200-bp tiles were clustered into 18 modules with the parameters:

```
k = 18, maxIterations = 10, runs = 30, initializationMode = "random", seed
= 50, initializationSteps = 5, epilison = 1e-4
```

by using the `mllib` package of `pyspark` engine (v1.6.2) and reordered the position of the 18 modules manually. Second, for each cluster, we used K-means (`scikit-learn` v0.18.1) with parameter (`k=10`, `random_state=0`) to divide 1 group into 10 sub-modules, and then determine the order for the 10 sub-modules by clustering the center of the resulting 10 sub-modules using ward distance (`scipy` v0.18.1). Finally, the regions of the resulting $18 \times 10 = 180$ sub-modules were extracted and used for the enrichment analyses of proximity to gene members of a catalog of gene sets (Biological Process (BP) of Gene Ontology (GO), phenotypes of Mouse Genotype Informatics (MGI) and MGI expressed detection) by GREAT.⁷⁰ The GREAT web API was used to automatically submit region descriptions and retrieve results for subsequent parsing. We restricted ourselves to the interpretation of results with an enrichment ratio of at least 2 and showed the results of binomial FDR *P*-values in the correspondent figure. Details of enhancer dynamics during the development of each lineage in the gastrula have been deposited in the Gene Expression Omnibus (GEO) under accession number GSE98101.

Identification of gastrula-specific enhancer bins in the mouse gastrula

The 0/1 binarized 200-bp intervals generated by ChromHMM for samples in the gastrula were compared with H3K27ac narrow peak file downloaded from ENCODE (See Supplementary information, Table S5). The downloaded H3K27ac peak intervals were merged using "bedtools multiinter" command and then joined with our data using "bedtools intersect". If there were any peaks in ENCODE dataset that were intersected with a 200-bp interval in the gastrula sample, it would be considered as "common H3K27ac peak" for this interval, while gastrula-specific enhancer bins were defined as H3K27ac peaks in the gastrula samples that did not overlap with any peaks in the ENCODE data (Fig. 5a).

Mouse enhancer activity reporter construction

Hsp68 minimal promoter driving LacZ reporter plasmid (pHsp68-LacZ plasmid) was kindly given by Prof. Bin Zhou (Shanghai Institute of Biochemistry and Cell Biology) as a gift. The plasmid was firstly linearized by *Hind*III (FastDigest, FD0504) at 37 °C for 2 h and purified (Qiagen, 28704). Enhancer fragments were cloned from mouse genomic DNA by using KOD neo polymerase system (Toyobo, KFX-201), and then ligated to linearized pHsp68-LacZ plasmid by seamless cloning system (Beyotime, D7010M). Acquired plasmid was further digested with *Not*I (FastDigest, FD0596) and *Xho*I (FastDigest, FD0694) followed by purification to obtain linearized fragment for transgenic mouse generation. The primers used here were included in Supplementary information, Table S8.

Transgenic mouse acquisition and LacZ staining

Generally, we used zygote pronuclear microinjection to produce LacZ transgenic mice. Egg donor female mice (3-week-old C57BL/6) were superovulated by injecting PMSG (pregnant mare's serum gonadotropin) and HCG (human chorionic gonadotropin) prior to mating. Zygotes were harvested about 18 h after mating. Purified DNA was diluted with injection buffer (10 mM Tris, pH 7.5, 0.1 mM EDTA) to a final concentration of 3–4 ng/μL and used for pronuclear injections of PN4 stage zygotes with a FemtoJet Microinjection System (Eppendorf). The injected embryos were cultured to the 2-cell stage in KSOM medium with amino acids at 37 °C under 5% CO₂, and then transferred to the oviduct of pseudo-pregnant ICR females and marked as 0.5 dpc. Embryos were collected at the appropriate stage for LacZ staining. Lac staining was performed as previously reported.⁷¹

CRISPR-Cas9-mediated enhancer fragment deletion

Small guided RNAs were designed by using the online tool Chopchop (<http://chopchop.cbu.uib.no/>). The plasmid preparation scheme and knockout cell line generation was performed as previously reported.⁷² The potential off-target sites were also tested based on the website's indication.

shRNA design

shRNAs for knocking down the expression of *2510009E07Rik* were designed by using online tool GPP Web Portal (<https://portals.broadinstitute.org/gpp/public/>). Replicated shRNAs were selected. Scramble sequence for each shRNA was designed using online tool Create Scrambled Sequence (<https://www.genscript.com/tools/create-scrambled-sequence>).

Mouse embryonic stem cell culture and neural differentiation

Mouse embryonic stem cell E14Tg2A was cultured in standard medium supplemented with 3 μM ChIR99021, 1 μM PD0325901 and 10 ng/mL mouse leukemia inhibitory factor (mLIF, Millipore). The neural differentiation was performed as previously reported.⁴³

Data availability

The gene expression and histone modification data are deposited in NCBI GEO under the accession number GSE98101.

ACKNOWLEDGEMENTS

We thank Drs Patrick Tam, Chi-chung Hui, Bing Zhu and Dangsheng Li for helpful advice and discussion. We thank Dr Bin Zhou for the generous gift of the pHsp68-LacZ plasmid. This work was supported in part by the National Key Basic Research and Development Program of China (2018YFA0800100, 2018YFA0108000, 2018YFA0107200, 2017YFA0102700, 2014CB964804, 2018YFA0107601 and 2015CB964500), the Strategic Priority Research Program of the Chinese Academy of Sciences (XDA16020501 and XDA16020404), the Genome Tagging Project, and the National Natural Science Foundation of China (31900454, 31501178, 31430058, 31571513, 31630043, 31871456, 91519314, 31661143042, 81561138005, 31625018, 81521002, 31530048, 81672117 and 31730062), and Shanghai Municipal Commission for Science and Technology (17JC1400900 and 17411954900).

AUTHOR CONTRIBUTIONS

Y. Qiao, J.L., F.T., and N.J. conceived and supervised the project. X.Y., J.L., Y.C., Y. Qian, S.F., F.Y., H.X., and Y.H. performed the experiments. B.H., J.D., Y.H., and R.W. performed the computational analyses. X.Y., B.H., Y. Qiao, G.P., F.T., and N.J. wrote the paper with the help from all other authors.

ADDITIONAL INFORMATION

Supplementary information accompanies this paper at <https://doi.org/10.1038/s41422-019-0234-8>.

Competing interests: The authors declare no competing interests.

REFERENCES

- Wolpert, L., Tickle, C. & Martinez-Arias, A. *Principles of Development*. Fifth edn (Oxford, UK: Oxford University Press, 2015).
- 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).
- Tam, P. P. & Behringer, R. R. Mouse gastrulation: the formation of a mammalian body plan. *Mech. Dev.* **68**, 3–25 (1997).
- Downs, K. M. & Davies, T. Staging of gastrulating mouse embryos by morphological landmarks in the dissecting microscope. *Development* **118**, 1255–1266 (1993).
- Takaoka, K. & Hamada, H. Cell fate decisions and axis determination in the early mouse embryo. *Development* **139**, 3–14 (2012).
- Benitah, S. A. & Frye, M. Stem cells in ectodermal development. *J. Mol. Med.* **90**, 783–790 (2012).
- Tada, S. et al. Characterization of mesendoderm: a diverging point of the definitive endoderm and mesoderm in embryonic stem cell differentiation culture. *Development* **132**, 4363–4374 (2005).
- Brand, T. Heart development: molecular insights into cardiac specification and early morphogenesis. *Dev. Biol.* **258**, 1–19 (2003).
- Atsli, Y. & Stunnenberg, H. G. The interplay of epigenetic marks during stem cell differentiation and development. *Nat. Rev. Genet.* **18**, 643–658 (2017).
- Liu, C., Peng, G. & Jing, N. TGF- β signaling pathway in early mouse development and embryonic stem cells. *Acta. Biochim. Biophys. Sin.* **50**, 68–73 (2018).
- Liu, X. et al. Distinct features of H3K4me3 and H3K27me3 chromatin domains in pre-implantation embryos. *Nature* **537**, 558–562 (2016).
- Dahl, J. A. et al. Broad histone H3K4me3 domains in mouse oocytes modulate maternal-to-zygotic transition. *Nature* **537**, 548–552 (2016).
- Wu, J. et al. The landscape of accessible chromatin in mammalian preimplantation embryos. *Nature* **534**, 652–657 (2016).
- Ke, Y. et al. 3D chromatin structures of mature gametes and structural reprogramming during mammalian embryogenesis. *Cell* **170**, 367–381 (2017).
- Wang, C. et al. Reprogramming of H3K9me3-dependent heterochromatin during mammalian embryo development. *Nat. Cell. Biol.* **20**, 620–631 (2018).
- Zylicz, J. J. et al. G9a regulates temporal preimplantation developmental program and lineage segregation in blastocyst. *Elife* **7**, pii: e33361 (2018).
- Zylicz, J. J. et al. Chromatin dynamics and the role of G9a in gene regulation and enhancer silencing during early mouse development. *Elife* **4**, pii: e09571 (2015).
- Gorkin, D. U. et al. Systematic mapping of chromatin state landscapes during mouse development. *BioRxiv* 2017. <https://doi.org/10.1101/166652>
- Nord, A. S. et al. Rapid and pervasive changes in genome-wide enhancer usage during mammalian development. *Cell* **155**, 1521–1531 (2013).
- Yang, X. et al. Silencing of developmental genes by H3K27me3 and DNA methylation reflects the discrepant plasticity of embryonic and extraembryonic lineages. *Cell. Res.* **28**, 593–596 (2018).
- Smith, Z. D. et al. Epigenetic restriction of extraembryonic lineages mirrors the somatic transition to cancer. *Nature* **549**, 543–547 (2017).
- Adam, R. C. et al. Pioneer factors govern super-enhancer dynamics in stem cell plasticity and lineage choice. *Nature* **521**, 366–370 (2015).
- Ong, C. T. & Corces, V. G. Enhancers: emerging roles in cell fate specification. *EMBO Rep.* **13**, 423–430 (2012).
- Wang, A. et al. Epigenetic priming of enhancers predicts developmental competence of hESC-derived endodermal lineage intermediates. *Cell Stem Cell* **16**, 386–399 (2015).
- Tam, P. P. & Loebel, D. A. Gene function in mouse embryogenesis: get set for gastrulation. *Nat. Rev. Genet.* **8**, 368–381 (2007).
- Peng, G. et al. Spatial transcriptome for the molecular annotation of lineage fates and cell identity in mid-gastrula mouse embryo. *Dev. Cell.* **36**, 681–697 (2016).
- Creyghton, M. P. et al. Histone H3K27ac separates active from poised enhancers and predicts developmental state. *Proc. Natl. Acad. Sci. USA* **107**, 21931–21936 (2010).
- Roadmap Epigenomics, C. et al. Integrative analysis of 111 reference human epigenomes. *Nature* **518**, 317–330 (2015).
- Charlet, J. et al. Bivalent regions of cytosine methylation and H3K27 acetylation suggest an active role for DNA methylation at enhancers. *Mol. Cell* **62**, 422–431 (2016).
- De Gobbi, M. et al. Generation of bivalent chromatin domains during cell fate decisions. *Epigenetics Chromatin* **4**, 9 (2011).
- Kojima, Y. et al. The transcriptional and functional properties of mouse epiblast stem cells resemble the anterior primitive streak. *Cell Stem Cell* **14**, 107–120 (2014).
- Whyte, W. A. et al. Enhancer decommissioning by LSD1 during embryonic stem cell differentiation. *Nature* **482**, 221–225 (2012).
- Osterwalder, M. et al. Enhancer redundancy provides phenotypic robustness in mammalian development. *Nature* **554**, 239–243 (2018).
- Shen, H. et al. Suppression of enhancer overactivation by a RACK7-histone demethylase complex. *Cell* **165**, 331–342 (2016).
- Strumpf, D. et al. Cdx2 is required for correct cell fate specification and differentiation of trophoblast in the mouse blastocyst. *Development* **132**, 2093–2102 (2005).
- Jiang, B. & Mendelson, C. R. USF1 and USF2 mediate inhibition of human trophoblast differentiation and CYP19 gene expression by Mash-2 and hypoxia. *Mol. Cell. Biol.* **23**, 6117–6128 (2003).
- Nagai, T. et al. The expression of the mouse *Zic1*, *Zic2*, and *Zic3* gene suggests an essential role for *Zic* genes in body pattern formation. *Dev. Biol.* **182**, 299–313 (1997).
- Wu, G. & Scholer, H. R. Role of Oct4 in the early embryo development. *Cell. Regen.* **3**, 7 (2014).
- Mesman, S. & Smidt, M. P. Tcf12 is involved in early cell-fate determination and subset specification of midbrain dopamine neurons. *Front. Mol. Neurosci.* **10**, 353 (2017).
- Bonev, B. et al. Multiscale 3D genome rewiring during mouse neural development. *Cell* **171**, 557–572 (2017).
- Yue, F. et al. A comparative encyclopedia of DNA elements in the mouse genome. *Nature* **515**, 355–364 (2014).
- Dickel, D. E. et al. Ultraconserved enhancers are required for normal development. *Cell* **172**, 491–499 (2018).
- Zhang, K. et al. Distinct functions of BMP4 during different stages of mouse ES cell neural commitment. *Development* **137**, 2095–2105 (2010).
- Nichols, J. & Smith, A. Naive and primed pluripotent states. *Cell Stem Cell* **4**, 487–492 (2009).
- Smith, A. Formative pluripotency: the executive phase in a developmental continuum. *Development* **144**, 365–373 (2017).
- Kubo, A. et al. Development of definitive endoderm from embryonic stem cells in culture. *Development* **131**, 1651–1662 (2004).
- Cheng, X. et al. Self-renewing endodermal progenitor lines generated from human pluripotent stem cells. *Cell Stem Cell* **10**, 371–384 (2012).
- Kumar, N. et al. Generation of an expandable intermediate mesoderm restricted progenitor cell line from human pluripotent stem cells. *Elife* **4**, pii: e08413 (2015).
- Liu, C. et al. Suppressing nodal signaling activity predisposes ectodermal differentiation of epiblast stem cells. *Stem Cell Rep.* **11**, 43–57 (2018).
- Du, Z. et al. Allelic reprogramming of 3D chromatin architecture during early mammalian development. *Nature* **547**, 232–235 (2017).
- Zhang, Y. et al. Dynamic epigenomic landscapes during early lineage specification in mouse embryos. *Nat. Genet.* **50**, 96–105 (2018).
- Zhu, J. et al. Genome-wide chromatin state transitions associated with developmental and environmental cues. *Cell* **152**, 642–654 (2013).
- Rossant, J. & Tam, P. P. L. *Mouse Development: Patterning, Morphogenesis, and Organogenesis*. (San Diego, USA: Gulf Professional Publishing, 2002).
- Edlund, T. & Jessell, T. M. Progression from extrinsic to intrinsic signaling in cell fate specification: a view from the nervous system. *Cell* **96**, 211–224 (1999).
- Bogdanovic, O. et al. Active DNA demethylation at enhancers during the vertebrate phylogenetic period. *Nat. Genet.* **48**, 417–426 (2016).
- Cusanovich, D. A. et al. The cis-regulatory dynamics of embryonic development at single-cell resolution. *Nature* **555**, 538–542 (2018).
- Buenrostro, J. D. et al. Single-cell chromatin accessibility reveals principles of regulatory variation. *Nature* **523**, 486–490 (2015).
- Hainer, S. J., Boskovic, A., McCannell, K. N., Rando, O. J. & Fazio, T. G. Profiling of pluripotency factors in single cells and early embryos. *Cell* **177**, 1319–1329 (2019).
- Behringer, R. R., Gertsenstein, M., Nagy, K. V. & Nagy, A. *Manipulating the mouse embryo: a laboratory manual*. (New York, USA: Cold Spring Harbor Laboratory Press, 2014).

60. Chen, J. et al. Spatial transcriptomic analysis of cryosectioned tissue samples with Geo-seq. *Nat. Protoc.* **12**, 566–580 (2017).
61. Kim, D. et al. TopHat2: accurate alignment of transcriptomes in the presence of insertions, deletions and gene fusions. *Genome Biol.* **14**, R36 (2013).
62. Anders, S., Pyl, P. T. & Huber, W. HTSeq-a Python framework to work with high-throughput sequencing data. *Bioinformatics* **31**, 166–169 (2015).
63. Love, M. I., Huber, W. & Anders, S. Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. *Genome Biol.* **15**, 550 (2014).
64. Leek, J. T., Johnson, W. E., Parker, H. S., Jaffe, A. E. & Storey, J. D. The sva package for removing batch effects and other unwanted variation in high-throughput experiments. *Bioinformatics* **28**, 882–883 (2012).
65. Li, H. & Durbin, R. Fast and accurate short read alignment with Burrows-Wheeler transform. *Bioinformatics* **25**, 1754–1760 (2009).
66. Zhang, Y. et al. Model-based analysis of ChIP-Seq (MACS). *Genome Biol.* **9**, R137 (2008).
67. Landt, S. G. et al. ChIP-seq guidelines and practices of the ENCODE and mod-ENCODE consortia. *Genome Res.* **22**, 1813–1831 (2012).
68. Guo, H. et al. Profiling DNA methylome landscapes of mammalian cells with single-cell reduced-representation bisulfite sequencing. *Nat. Protoc.* **10**, 645–659 (2015).
69. Ernst, J. & Kellis, M. ChromHMM: automating chromatin-state discovery and characterization. *Nat Methods* **9**, 215–216 (2012).
70. McLean, C. Y. et al. GREAT improves functional interpretation of cis-regulatory regions. *Nat. Biotechnol.* **28**, 495–501 (2010).
71. Visel, A., Minovitsky, S., Dubchak, I. & Pennacchio, L. A. VISTA Enhancer Browser - a database of tissue-specific human enhancers. *Nucleic Acids Res.* **35**, D88–D92 (2007).
72. Ran, F. A. et al. Genome engineering using the CRISPR-Cas9 system. *Nat. Protoc.* **8**, 2281 (2013).