LETTER

Corrected: Author Correction

https://doi.org/10.1038/s41586-019-1469-8

Molecular architecture of lineage allocation and tissue organization in early mouse embryo

Guangdun Peng[1,2,3,4,11,12]*, Shengbao Suo[5,10,11], Guizhong Cui[1,11], Fang Yu[1,11], Ran Wang[1], Jun Chen[1], Shirui Chen[1], Zhiwen Liu[1], Guoyu Chen[5], Yun Qian[1], Patrick P. L. Tam[6,7], Jing-Dong J. Han[5,8,12]* & Naihe Jing[1,4,9,12]*

During post-implantation development of the mouse embryo, descendants of the inner cell mass in the early epiblast transit from the naive to primed pluripotent state[1]. Concurrently, germ layers are formed and cell lineages are specified, leading to the establishment of the blueprint for embryogenesis. Fate-mapping and lineage-analysis studies have revealed that cells in different regions of the germ layers acquire location-specific cell fates during gastrulation[2–5]. The regionalization of cell fates preceding the formation of the basic body plan—the mechanisms of which are instrumental for understanding embryonic programming and stem-cell-based translational study—is conserved in vertebrate embryos[6–8]. However, a genome-wide molecular annotation of lineage segregation and tissue architecture of the post-implantation embryo has yet to be undertaken. Here we report a spatially resolved transcriptome of cell populations at defined positions in the germ layers during development from pre- to late-gastrulation stages. This spatiotemporal transcriptome provides high-resolution digitized in situ gene-expression profiles, reveals the molecular genealogy of tissue lineages and defines the continuum of pluripotency states in time and space. The transcriptome further identifies the networks of molecular determinants that drive lineage specification and tissue patterning, supports a role of Hippo–Yap signalling in germ-layer development and reveals the contribution of visceral endoderm to the endoderm in the early mouse embryo.

The lineage history and divergence of germ-layer progenitors in early embryos can be annotated by the gene-expression profiles in the temporal (developmental) and spatial (location) dimension. Single-cell RNA-sequencing (RNA-seq) analysis has the potential to reconstruct the developmental trajectories of embryonic cells, but is confounded by the absence of spatial and temporal resolution[9–13]. To achieve a spatial readout of the transcriptome of the epiblast, a genome-wide analysis of the transcriptome in geographically defined epiblast cell populations has been performed by low-input RNA-seq using the Geo-seq protocol at the late mid-streak stage embryo[14]. In the present study, the Geo-seq analysis of spatial transcriptome was extended to a series of post-implantation developmental stages from pre-gastrulation (embryonic day (E)5.5) to late gastrulation (E7.5) on all germ layers in the embryo proper (Fig. 1a and Extended Data Fig. 1a–c). This spatiotemporal transcriptome of discrete cell populations of 5–40 cells provides a comprehensive catalogue of the transcripts in the embryonic germ-layer tissues of unprecedented depth and quality (Extended Data Fig. 1d, e and Supplementary Table 1). An immediate attribute of this spatiotemporal transcriptome is the rich dataset of the dynamic expression profile of coding and noncoding transcripts in defined locations in the germ layers of E5.5 to E7.5 mouse embryos, which could be rendered digitally for visualizing in high-resolution 2D corn plots (Fig. 1b and Extended Data Fig. 1f–i) and a 3D anatomical template[14] (Methods). The identification of region-specific transcripts provides a molecular entry point for functional analysis. For example, the knockout of posterior-enriched long non-coding RNAs (lncRNAs) *310004A29Rik* and *RP23-458G12.2* (Extended Data Fig. 2) led to the dysregulation of mesendoderm-related genes in the differentiating embryonic stem cells, which suggests that these non-coding transcripts have a role in regulating the differentiation of cell lineages derived from the posterior epiblast.

For the analysis of lineage trajectory, we enriched the Geo-seq dataset with the RNA-seq data of E2.5 morula, the E3.5 inner cell mass of the blastocyst and the E4.5 epiblast and primitive endoderm[15] to collate a continuum of transcriptome across the pre- and post-implantation development of the mouse embryos. We used the single-cell regulatory network inference and clustering (SCENIC) pipeline[16] to measure regulon activities in the cell-population samples (Extended Data Fig. 3a and Methods). The regulon-based phylogenetic tree showed that the samples could be divided into six groups that represent the major tissue types of early embryo: pre- and peri-implantation, epiblast, primitive streak, ectoderm, mesoderm and endoderm (Fig. 1c). Of note, the primitive streak cells at E6.5–E7.0 stages were separated from E7.5 primitive streak, which suggests the primitive streak is temporally heterogeneous at early and late gastrulation. The E5.5–E7.5 endoderm samples formed a distinct cluster separate from the epiblast, ectoderm and mesoderm tissues. Additionally, the endoderm cluster was nearer to pre- and peri-implantation embryos than gastrulation-stage embryos. The clustering also indicated that the mesoderm connected with cells of the posterior epiblast that contained the primitive streak at E7.5 stage and the ectoderm emerged from the anterior epiblast (Fig. 1c and Extended Data Fig. 3b). Therefore, the regulon attributes of the spatiotemporal transcriptome furnish the molecular annotation of potential developmental connectivity of cell populations in the early embryos.

Analysis of the regulon activities of the E2.5 to E7.5 tissue samples delineated nine transcription-factor-driven regulon groups associated with the six major lineage clusters (Fig. 2a and Supplementary Table 2). Using biologically driven features inferred by SCENIC, we performed principal component analysis (PCA) and *t*-distributed stochastic neighbour embedding (*t*-SNE) of the tissue samples on the basis of the regulon-activity score. We found that the developmental time

[1]State 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, Shanghai, China. [2]CAS Key Laboratory of Regenerative Biology, Guangdong Provincial Key Laboratory of Stem Cell and Regenerative Medicine, Guangzhou Institutes of Biomedicine and Health, Chinese Academy of Sciences, Guangzhou, China. [3]Guangzhou Regenerative Medicine and Health Guangdong Laboratory (GRMH-GDL), Guangzhou, China. [4]Institute for Stem Cell and Regeneration, Chinese Academy of Sciences, Beijing, China. [5]CAS Key Laboratory of Computational Biology, CAS-MPG Partner Institute for Computational Biology, Shanghai Institute of Nutrition and Health, Chinese Academy of Sciences Center for Excellence in Molecular Cell Science, Collaborative Innovation Center for Genetics and Developmental Biology, Shanghai Institutes for Biological Sciences, Chinese Academy of Sciences, Shanghai, China. [6]Embryology Unit, Children's Medical Research Institute, University of Sydney, Westmead, New South Wales, Australia. [7]School of Medical Sciences, Faculty of Medicine and Health, University of Sydney, Westmead, New South Wales, Australia. [8]Peking-Tsinghua Center for Life Sciences, Academy for Advanced Interdisciplinary Studies, Center for Quantitative Biology (CQB), Peking University, Beijing, China. [9]School of Life Science and Technology, ShanghaiTech University, Shanghai, China. [10]Present address: Department of Biostatistics and Computational Biology, Dana-Farber Cancer Institute and Harvard T.H. Chan School of Public Health, Boston, MA, USA. [11]These authors contributed equally: Guangdun Peng, Shengbao Suo, Guizhong Cui, Fang Yu. [12]These authors jointly supervised this work: Guangdun Peng, Jing-Dong J. Han, Naihe Jing. *e-mail: peng_guangdun@gibh.ac.cn; jackie.han@pku.edu.cn; njing@sibcb.ac.cn

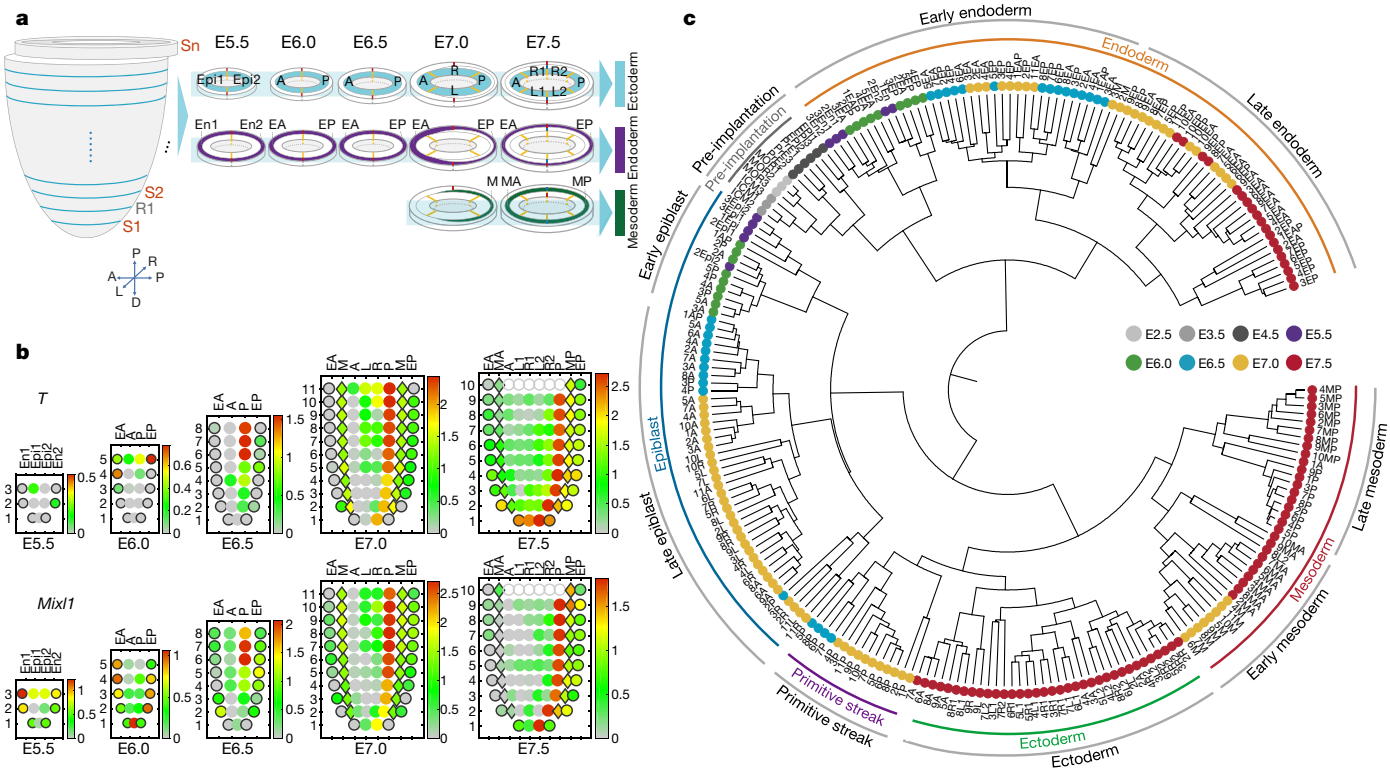


Fig. 1 | Analysis of the spatiotemporal transcriptome of post-implantation mouse embryos by Geo-seq. a, The spatial and temporal coverage of Geo-seq samples of E5.5–E7.5 embryos (Methods). A, anterior; P, posterior; L, left lateral; R, right lateral; L1, anterior left lateral; R1, anterior right lateral; L2, posterior left lateral; R2, posterior right lateral; Epi1 and Epi2, divided epiblast; M, whole mesoderm; MA, anterior mesoderm; MP, posterior mesoderm; En1 and En2, divided endoderm; EA, anterior endoderm; EP, posterior endoderm. S, sample section; R, reference section. Sn: sample section of level *n*. **b**, Corn plots showing the spatial pattern of expression of *T* and *Mixl1*, which marks the length of the primitive streak for developmental staging of gastrulation. Solid circles, epiblast–ectoderm; black outlined rhombus, mesoderm; black outlined circles, endoderm; grey hollow circles, no sample. $n > 2$ for each gene. **c**, The phylogenetic tree shows the classification of embryonic tissues from pre-implantation to gastrulation.

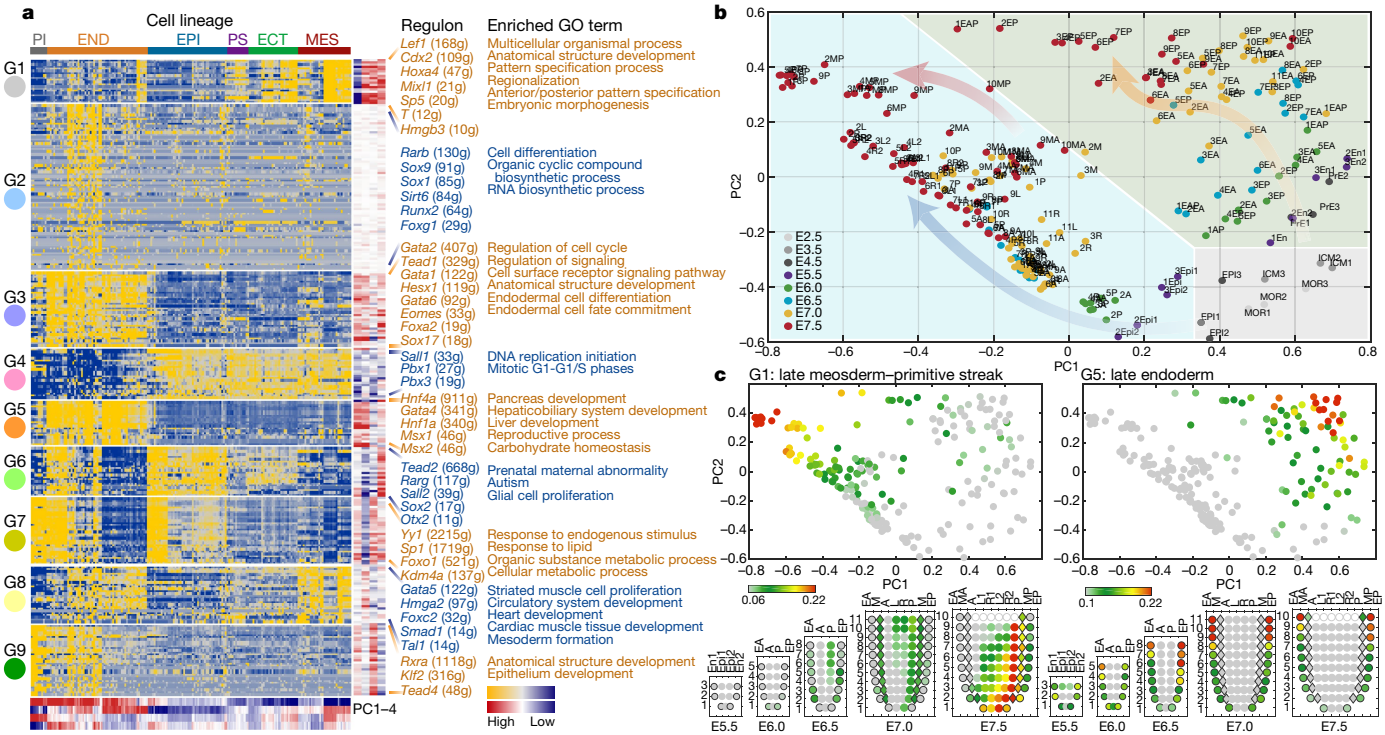


Fig. 2 | Developmental regulons. a, The BIC-SKmeans heat map (on the basis of regulon activity scores; Methods) showing nine regulon groups in tissue samples of E2.5 to E7.5 embryos (ordered on the basis of Fig. 1c) with listing of examples of regulon transcription factors (numbers of predicted target genes by SCENIC in the brackets) for G1–G9 and the enriched Gene Ontology (GO) terms for each regulon group. Right column heat map: contribution of each regulon to the principal component 1–4 (PC1–4). Bottom row heat map: projection scores of each sample on the PC1–4. **b**, PCA plot based on the regulon-activity matrix of tissue samples showing separate developmental trajectories (arrows) across the E2.5 to E7.5 timelines ($n = 226$). **c**, Two-dimensional PCA plots ($n = 226$) and corn plots showing the averaged regulon activities of regulon group G1 and G5 in embryonic tissues. PI, pre-implantation; END, endoderm; PS, primitive streak; EPI, epiblast; ECT, ectoderm; MES, mesoderm.

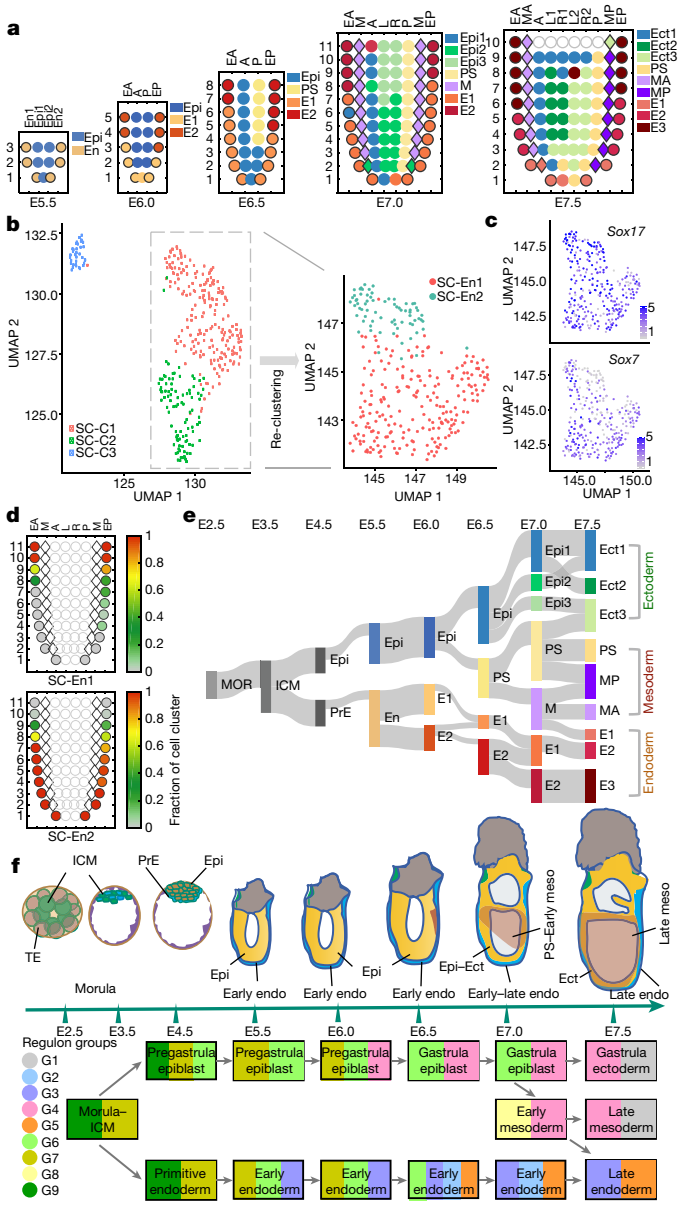


Fig. 3 | The molecular architecture and spatial relationship of the germ-layer tissues. a, The spatial domain of cell populations that are defined by the expression profile of the DEGs in the tissue and germ layers of E5.5–E7.5 embryos. **b**, Clusters of single cells (SC-C1, SC-C2 and SC-C3) from the E7.0 endoderm (left, $n = 345$; right, $n = 295$), reclustered into SC-En1 and SC-En2. **c**, Expression of *Sox17* and *Sox7* in single cells ($n = 295$). **d**, Deconvolution analysis inferred the proportion of two endoderm cell types in Geo-seq endoderm samples. Colour bar indicates the cell-type frequency. **e**, The connection of spatial domains based on the correlation of averaged regulon activities in the tissues of E2.5–E7.5 embryos. **f**, The connectivity of tissues underscored by the continuity of the regulon groups (G1–G9). MOR, morula; ICM, inner cell mass; PrE, primitive endoderm; TE, trophectoderm. Spatial domains: epiblast (Epi), Epi1–3; endoderm (En), E1–3; ectoderm, Ect1–3; mesoderm, M, MA and MP.

and lineage paths were mainly delineated in the first two principal-component space (Fig. 2b and Extended Data Figs. 3c, 4a). The endoderm tissues were aligned with the primitive endoderm and were separated with the epiblast and epiblast derivatives. The mesoderm and ectoderm were in the developmental trajectory of the E4.5 epiblast branch and were closely related, with many regulatory states shared by these two lineages. To visualize the finer segregation of mesoderm and ectoderm, PCA and *t*-SNE analyses were performed on samples that excluded the endoderm. The results showed a clear divergence of mesoderm and ectoderm by E7.5 (Extended Data Fig. 3d). We further analysed the relationship of endoderm, mesoderm and primitive streak. In contrast to the prevailing view that the definitive endoderm originates mainly from the primitive streak, we found that the endoderm of the gastrulating embryo was transcriptionally more similar to primitive or visceral endoderm than primitive streak (Fig. 2b and Extended Data Fig. 3e, f). However, the mesoderm tissues at the distal regions of E7.0 and E7.5 embryo showed regulon connectivity with neighbouring endoderm (for example, G3 and G4 in Extended Data Fig. 3g, h), pointing to a potential location of epiblast-derived endoderm[13,17].

Plotting the distribution of average regulon group activities on both PCA and corn plots revealed population- and germ-layer-related determinants (Fig. 2c and Extended Data 4b). For example, the G1 regulon was enriched in E7.5 mesoderm and posterior epiblast–primitive streak samples, G9 was specifically enriched in cells of pre-implantation embryos and G6—which contains a transcription factor network associated with *Sall2*, *Sox2* and *Otx2*—was enriched in the peri-implantation samples. Notably, the cell-cycle enriched G4 regulon was associated with most post-implantation epiblast tissues except for endoderm, indicating a probable preference in proliferative activity among the germ layers. The PCA display of regulon groups also revealed the association of G3 and G5 with the segregation of early and late endoderm, respectively. Further analysis of the interaction of regulons using connective specificity index[18] identified the prominent regulons (for example, *Sox17*, *Foxa2* and *Gata4*) that have the most connections with other regulons (Extended Data Fig. 4c and Supplementary Table 3). These regulons were populated by many known lineage-specific genes and some regulon members were associated with mutant germ-layer-related phenotypes (Extended Data Fig. 4c, d and Supplementary Table 4), which suggests a contribution to embryonic patterning at gastrulation. To investigate whether the predicted transcription factors of regulons were essential for the development of specific tissue lineages, three transcription factors (*Sp1*, *Hmga2* and *Hmgb3*) in regulon groups G7, G8 and G1, respectively, were knocked out in embryonic stem cells. Ablation of *Sp1* promoted the exit of naive pluripotency to formative pluripotency during embryonic stem cell differentiation (Extended Data Fig. 5a, b). The knockout of two mesoderm-related transcription factors (*Hmga2* and *Hmgb3*) resulted in the aberrant expression of primitive haematopoietic and cardiac genes respectively, indicating that these two transcription factor regulons may play a role in directing mesoderm differentiation (Extended Data Fig. 5c–g and Supplementary Videos 1 and 2). Therefore, the spatiotemporal transcriptome informs the genetic activity regulating lineage segregation and driving germ-layer differentiation in post-implantation development.

The spatiotemporal transcriptome of the post-implantation embryos enables us to scrutinize the architecture and the molecular attributes of the body plan. Spatial domains of differentially expressed genes (DEGs) were identified in embryos at each developmental stage (Supplementary Table 5). Clustering analysis revealed an overall homogenous epiblast in E5.5 and E6.0 embryos (Extended Data Fig. 6a, b). By E6.5, the epiblast was partitioned into anterior and posterior domains, with the posterior domain marked by the expression of primitive streak-related genes such as *T*, *Mixl1* and *Mesp1* (Extended Data Fig. 6c). During gastrulation, the epiblast was further compartmentalized into anterior, posterior and lateral domains (Extended Data Figs. 6e, 7a). Mesoderm was partitioned into anterior and posterior domains by E7.5 (Extended Data Fig. 7a). The spatial domains also pointed to the existence of unique embryonic structures; for example, the node that was revealed by the restricted expression of the G5 DEG group in E7.5 anterior primitive streak and adjacent endoderm tissues (Extended Data Fig. 7b–d). Tracking the expression of signature genes such as *Noto*, *Nodal* and *Foxj1* revealed the genesis of the gastrula organizer[19,20].

In the endoderm, whereas cell populations expressing markers specific to distal visceral endoderm and anterior visceral endoderm were present at E5.5–E6.5 (Extended Data Fig. 1g, h), there was no distinctive regionalization of anterior visceral endoderm and distal visceral endoderm populations (Extended Data Fig. 6a–c). Anterior

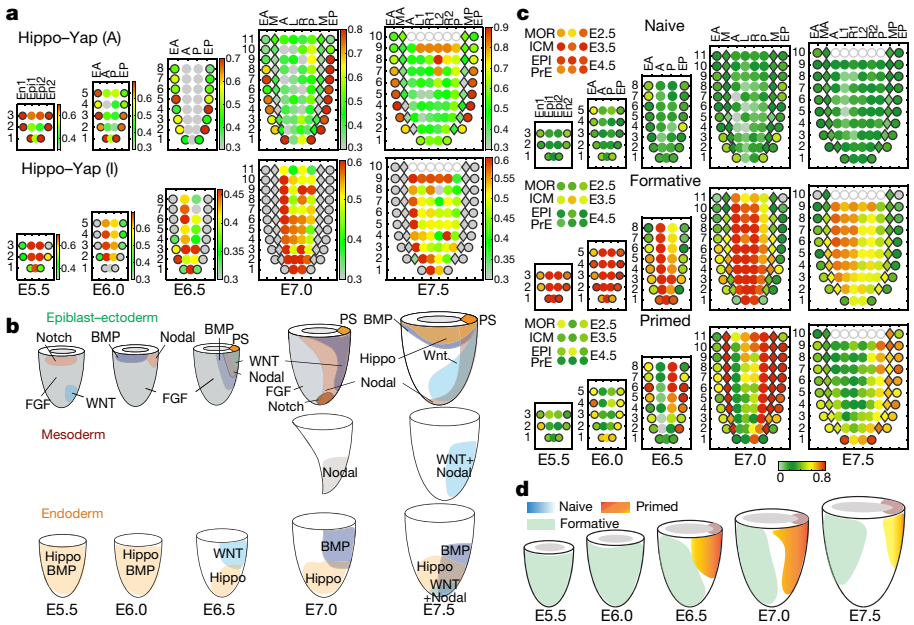


Fig. 4 | Relationship of regionalization of signalling activity and transition of pluripotency state with cell fates in the post-implantation embryos. a, Corn plots showing the activities of the target genes related to the activated and inhibitory states of the Hippo–Yap signalling pathway in the germ layers of E5.5 to E7.5 embryos. **b**, Spatial territory of active bone morphogenetic protein (BMP), Wnt, fibroblast growth factor (FGF), Hippo–Yap, Notch and Nodal signalling in the three germ layers of E5.5 to E7.5 embryos. **c**, Gene-set activity analysis of the expression domains of genes associated with naive, formative and primed pluripotency states in E2.5 to E7.5 stage embryos. **d**, The transition of pluripotency states in the epiblast–ectoderm of E5.5 to E7.5 embryos.

visceral endoderm allocation was revealed only after guilt-by-association co-expression to enrich the regionalized gene set (Extended Data Fig. 6d). An early sign of proximal–distal compartmentalization in endoderm became evident from E6.0 onward (Fig. 3a). To reveal the heterogeneity of endoderm populations in the proximal–distal domains, we assayed 384 single cells from E7.0 endoderm (Extended Data Fig. 7e, f). We found two tightly connected single-cell clusters—with visceral endoderm characteristics but expressing varied levels of *Sox17* and *Sox7*—after removing a minor cluster that expressed primitive streak markers (*T* and *Mixl1*) (Fig. 3b, c and Extended Data Fig. 7g, h). By deconvolution analysis of the endoderm Geo-seq samples using the cell types defined by single-cell RNA-seq, we allocated the single-cell clusters (SC-En1 and SC-En2) to the spatial endoderm domains (E1 and E2) and defined the relative fractions of each cell type. The proximal and distal endoderm domains in E7.0 embryos were populated exclusively by these two single-cell endoderm clusters, respectively, whereas populations with intermediate features were observed at the junction region of the two Geo-seq endoderm domains (Fig. 3a, d).

The general layout of the spatial domains showed substantial inter-embryo consistency (Extended Data Fig. 8 and Supplementary Table 6), which suggests a high synchronicity in embryonic patterning of cell populations in the germ layers during gastrulation. The correlation of regulon-activity scores across the spatial DEG domains of successive developmental stages (Fig. 3e, Methods and Supplementary Table 7) revealed the regulatory network that underpins the molecular architectures along the developmental trajectory from morula towards the nine major germ-layer populations at E7.5 (Fig. 3a, e). Of note, at E7.5, the posterior ectoderm and posterior mesoderm shared a relationship with cells in E7.0 primitive streak. The high activity scores of *T* and *Sox2* in the anterior-distal region of E7.0 primitive streak (Extended Data Fig. 6f), points to the presumptive existence of neural-mesodermal progenitors[21]. Before E7.5, the endoderm was composed mainly of cells connected to the primitive and visceral endoderm (Fig. 3e). However, given the presence of a primitive-streak-like cell type among three endoderm single-cell clusters in the E7.0 endoderm (Extended Data Fig. 7f–h), and as the distal endoderm at E7.5 (E1) was related to the mesoderm (Fig. 3e), the E1-endoderm cells might represent endoderm originating from the primitive streak and egressing through the mesoderm at the late gastrulation stage[13,17,22]. These findings corroborated the lineage tracing study reporting the convergence of cells of extra-embryonic origin to an embryonic endoderm state[23]. Overall, the pattern of regulon switching at the branching points and the sequential transition of regulon usage highlighted the intricate and dynamic molecular control of the specification of tissue fates during germ-layer development (Fig. 3f).

To identify the key signalling activity regulating tissue patterning, we performed enrichment analysis of signalling target genes collated from the regulon data (Methods) for the nine regulon groups. The activating and inhibitory activity of the six signalling pathways was regionalized in the germ layers (Fig. 4a, b and Extended Data Fig. 9a, b). However, only mesoderm- and endoderm-related regulon groups G1, G3, G5 and G8 showed significant enrichment of activated signalling (Extended Data Fig. 9a). The Hippo–Yap signalling pathway, which has a role in trophectoderm segregation[24] but was least characterized in the gastrulating embryo, was enriched exclusively in the endoderm (Fig. 4a, b). The canonical Hippo–Yap pathway factors such as *Tead1*, *Tead4* and *Ctgf* were expressed specifically in the endoderm, although *Yap1* was widely expressed (Extended Data Fig. 9c, d). Blocking the signalling activity in the endoderm explants with the YAP inhibitor verteporfin led to downregulation of visceral-endoderm markers and the early-endoderm-related regulon G3, but no changes in late-endoderm genes (Extended Data Fig. 9e–i). This finding suggested that Hippo signalling may be involved in regulating early-endoderm development.

On the basis of the spatial pattern of the average expression level and activity scores of known germ-layer markers (Extended Data Fig. 10a, b), a series of molecular fate maps was constructed for the epiblast (Extended Data Fig. 10c(i), (ii) and Methods). The maps showed that the anterior epiblast acquired an expanding ectoderm potency at E7.0 and E7.5. The posterior distal epiblast acquired mesoderm–endoderm potency at E7.0 and a predominant mesoderm potency by E7.5. Progenitors of the surface ectoderm were allocated to the anterior proximal epiblast (at E7.0) and then to the proximal-lateral epiblast (at E7.5). These molecular fate maps broadly recapitulated the previously established cell fate map[25] (Extended Data Fig. 10c(iii)).

A continuum of pluripotency is displayed by the embryonic cells during peri-implantation development[1]. Profiling the expression of genes of the pluripotency regulatory network[15] (Supplementary Table 8) revealed that the transcriptional activity of the naive-state pluripotency regulatory network was lost after implantation (Fig. 4c). Transcription of genes associated with formative pluripotency states was detected in E5.5 epiblast and first regionalized in anterior epiblast (E6.5) and retained in the anterior epiblast–ectoderm (E7.0–E7.5) (Fig. 4c). The acquisition of primed pluripotency state was evident in the E6.5 posterior epiblast, coinciding with enrichment of genes associated with epithelial–mesenchymal transition in the posterior epiblast (Extended Data Fig. 10d), which indicates the initiation of germ-layer commitment. Transition to primed pluripotency continued in the posterior epiblast at E7.0 and declined at E7.5 (Fig. 4c). Therefore, the transition to and the exit from pluripotency—the preclude to lineage differentiation—were regionalized and asynchronous among cell populations in the epiblast (Fig. 4d).

The spatiotemporal transcriptome opens avenues for exploring the molecular activity of lineage differentiation and tissue patterning in time and space, unravels the principal regulatory mechanisms driving lineage development[26,27] and enables the delineation of the molecular architecture and developmental trajectory of germ-layer derivatives in unprecedented detail. This dataset will be a valuable resource for guiding future efforts to elucidate the molecular mechanisms of lineage differentiation and morphogenesis of post-implantation embryos, and will enhance the efficiency of directed differentiation of pluripotent stem cells.

Online content

Any methods, additional references, Nature Research reporting summaries, source data, extended data, supplementary information, acknowledgements, peer review information; details of author contributions and competing interests; and statements of data and code availability are available at https://doi.org/10.1038/s41586-019-1469-8.

Received: 17 September 2018; Accepted: 3 July 2019;
Published online 7 August 2019.

1. Smith, A. Formative pluripotency: the executive phase in a developmental continuum. *Development* **144**, 365–373 (2017).
2. Lawson, K. A., Meneses, J. J. & Pedersen, R. A. Clonal analysis of epiblast fate during germ layer formation in the mouse embryo. *Development* **113**, 891–911 (1991).
3. Rivera-Pérez, J. A. & Hadjantonakis, A. K. The dynamics of morphogenesis in the early mouse embryo. *Cold Spring Harb. Perspect. Biol.* **7**, a015867 (2014).
4. Tam, P. P. & Loebel, D. A. Gene function in mouse embryogenesis: get set for gastrulation. *Nat. Rev. Genet.* **8**, 368–381 (2007).
5. Arnold, S. J. & Robertson, E. J. Making a commitment: cell lineage allocation and axis patterning in the early mouse embryo. *Nat. Rev. Mol. Cell Biol.* **10**, 91–103 (2009).
6. Irie, N. & Kuratani, S. The developmental hourglass model: a predictor of the basic body plan? *Development* **141**, 4649–4655 (2014).
7. Tam, P. P. & Behringer, R. R. Mouse gastrulation: the formation of a mammalian body plan. *Mech. Dev.* **68**, 3–25 (1997).
8. Tam, P. P. & Quinlan, G. A. Mapping vertebrate embryos. *Curr. Biol.* **6**, 104–106 (1996).
9. Griffiths, J. A., Scialdone, A. & Marioni, J. C. Using single-cell genomics to understand developmental processes and cell fate decisions. *Mol. Syst. Biol.* **14**, e8046 (2018).
10. Kumar, P., Tan, Y. & Cahan, P. Understanding development and stem cells using single cell-based analyses of gene expression. *Development* **144**, 17–32 (2017).
11. Pijuan-Sala, B. et al. A single-cell molecular map of mouse gastrulation and early organogenesis. *Nature* **566**, 490–495 (2019).
12. Cao, J. et al. The single-cell transcriptional landscape of mammalian organogenesis. *Nature* **566**, 496–502 (2019).
13. Nowotschin, S. et al. The emergent landscape of the mouse gut endoderm at single-cell resolution. *Nature* **569**, 361–367 (2019).
14. 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).
15. Boroviak, T. et al. Lineage-specific profiling delineates the emergence and progression of naive pluripotency in mammalian embryogenesis. *Dev. Cell* **35**, 366–382 (2015).
16. Aibar, S. et al. SCENIC: single-cell regulatory network inference and clustering. *Nat. Methods* **14**, 1083–1086 (2017).
17. Viotti, M., Nowotschin, S. & Hadjantonakis, A. K. SOX17 links gut endoderm morphogenesis and germ layer segregation. *Nat. Cell Biol.* **16**, 1146–1156 (2014).
18. Fuxman Bass, J. I. et al. Using networks to measure similarity between genes: association index selection. *Nat. Methods* **10**, 1169–1176 (2013).
19. Robb, L. & Tam, P. P. Gastrula organiser and embryonic patterning in the mouse. *Semin. Cell Dev. Biol.* **15**, 543–554 (2004).
20. Balmer, S., Nowotschin, S. & Hadjantonakis, A. K. Notochord morphogenesis in mice: current understanding & open questions. *Dev. Dyn.* **245**, 547–557 (2016).
21. Henrique, D., Abranches, E., Verrier, L. & Storey, K. G. Neuromesodermal progenitors and the making of the spinal cord. *Development* **142**, 2864–2875 (2015).
22. Kwon, G. S., Viotti, M. & Hadjantonakis, A. K. The endoderm of the mouse embryo arises by dynamic widespread intercalation of embryonic and extraembryonic lineages. *Dev. Cell* **15**, 509–520 (2008).
23. Chan, M. M. et al. Molecular recording of mammalian embryogenesis. *Nature* **570**, 77–82 (2019).
24. Nishioka, N. et al. The Hippo signaling pathway components Lats and Yap pattern Tead4 activity to distinguish mouse trophectoderm from inner cell mass. *Dev. Cell* **16**, 398–410 (2009).
25. Wilson, V. & Beddington, R. S. Cell fate and morphogenetic movement in the late mouse primitive streak. *Mech. Dev.* **55**, 79–89 (1996).
26. Briggs, J. A. et al. The dynamics of gene expression in vertebrate embryogenesis at single-cell resolution. *Science* **360**, eaar5780 (2018).
27. Peng, G., Tam, P. P. L. & Jing, N. Lineage specification of early embryos and embryonic stem cells at the dawn of enabling technologies. *Natl. Sci. Rev.* **4**, 533–542 (2017).

METHODS

Geo-seq of embryonic tissues. Transcriptome analyses of samples of a mixed cell population collected by laser capture microdissection (LCM) of C57BL/6J mouse embryos at the distal-visceral-endoderm (E5.5), anterior-visceral-endoderm (E6.0), early-streak (E6.5), late mid-streak (E7.0) and no-bud (E7.5) stages[28,29] were performed by Geo-seq[14,30]. Whole embryos were embedded in OCT (Leica Microsystems, catalogue no. 020108926) and cryosectioned in the transverse plane serially from the distal region to the proximal region at a thickness of 30 μm (E5.5 embryos), 15 μm (E6.0, E6.5 and E7.0 embryos) or 20 μm (E7.5 embryos). Sections were mounted on polyethylene-terephthalate-coated slides, fixed immediately with ethanol and stained with 1% cresyl violet acetate in 75% ethanol solution (Sigma-Aldrich). All sections from E5.5 embryos were used for LCM (MMI Cellcut Plus system). Alternate sections in the series of E6.0, E6.5, E7.0 and E7.5 embryos were used for LCM (sample sections) and others were used as anatomical templates (reference sections) for the construction of digital 3D display of gene-expression pattern. Samples of 5–40 cells were captured in the distal-to-proximal series of transverse sections. The sampling strategy was modified according to the size of the section. In the epiblast–ectoderm (inner layer), sampling was collected for two sectors (Epi1 and Epi2) of E5.5 embryos and from A (anterior) and P (posterior) sectors of E6.0–E6.5 embryos. Only one sample was collected from the most distal section of E5.5–E7.5 embryos. Lateral sectors (left lateral (L) and right lateral (R)) were captured in E7.0 and E7.5 embryo. For E7.5 embryos, additional samples were collected from the anterior (R1 and L1) and posterior (R2 and L2) regions of the lateral sector. From the endoderm (outer layer) and mesoderm (middle layer), only A and P samples were collected. In sum, 10 (E5.5), 18 (E6.0), 30 (E6.5), 73 (E7.0[14]) and 83 (E7.5) samples were collected for reference embryos (Fig. 1a, Extended Data Fig. 1a–c and Supplementary Table 1). Sequencing was performed on the Illumina HiSeq 2500 or NovaSeq 6000 sequencer (Berry Genomics). The detailed step-by-step procedure for Geo-seq has been deposited in Protocol Exchange[31]. The animal experimentation was performed in compliance with the guidelines of the Animal Ethical Committee of the Shanghai Institute of Biochemistry and Cell Biology, Chinese Academy of Sciences.

Single-cell RNA-seq. Mid-streak (E7.0) C57BL/6J embryos were dissected from the uterus in 10% FBS-DMEM medium, and extra-embryonic tissues were removed. Endoderm germ layer was isolated using proteolytic enzyme digestion (0.25% trypsin in PBS for 3 min) followed by mechanical dissection using tungsten needles[32]. After dissociation using Accutase (Gibco), 384 single cells from endoderm tissues were manually picked under the microscope, subjected to automatic Smart2-seq amplification and library construction with the Bravo robot station[33].

Data analysis. *Pre-processing of RNA-seq data.* Sequencing quality of raw sequence data was evaluated by FASTQC. Tophat2[34] was used to map sequencing reads to the mm10 genome with the following parameters, -g 1 -N 4–read-edit-dist 4–micro-exon-search -G annotation.GTF. Mapping ratio was calculated on the basis of the number of mapped reads and total reads for each sample. All mapped reads were processed by Cufflinks[35] to quantify gene-expression levels, measured by fragment per kilobase per million (FPKM), using default parameters. For each sample, only expressed genes with FPKM >1 in at least two samples across all samples were selected for further analysis. Gene-expression levels for each sample were transformed as $\log_{10}(\mathrm{FPKM}+1)$, and the expression distribution of all samples was computed and plotted to check for sample consistency of individual embryos. We obtained about 11×10^6 mapped reads and about 79% mapping ratio on average for each Geo-seq sample.

Inference of regulons and their activity. The latest SCENIC protocol[16] was used to infer the gene regulatory network (regulon). Instead of using individual gene expression, SCENIC analysis based on rank-based regulatory gene network (regulon) activities inferred from combined co-expression of transcription factor-target modules and enriched *cis*-regulatory motifs enables effective mitigation of potential batch problems and uncovering of stable biological significance compared to analysis based solely on transcriptome[36]. Here we fed the transcriptome profile of all samples from E5.5 to E7.5 embryos (for E7.0 and E7.5 stages, only embryos containing 3, 6 and 9 layers of epiblast together with mesoderm and endoderm samples were used for batch consistency) and data of morula (E2.5), inner cell mass (ICM) of the blastocyst (E3.5) and the primitive endoderm and epiblast of E4.5 embryo (sourced from ArrayExpress submissions E-MTAB-2958 and E-MTAB-2959[15]) to SCENIC to infer the transcription factor regulons. In brief, the regulon inference took three steps: (1) identification of co-expression modules between transcription factors and the potential target genes based on the gene-expression matrix through GENIE3 (v.1.2.1, R package); (2) for each co-expression module, the *cis*-regulatory motif enrichment analysis was performed among all potential target genes by RcisTarget (v.1.0.2), and only the target genes for which the motif of the corresponding transcription factor was enriched were selected as the direct target genes: each transcription factor and its direct target genes were defined as a regulon, and 282 regulons were finally identified; (3) the activity score of each regulon in each sample (226 samples encompassing all epiblast samples and other lineage samples) was computed through AUCell (v.1.2.4) to evaluate whether a regulon gene set was enriched at the top of the ranking for each sample: this strategy ranked the genes in a cell on the basis of their expression level and was therefore independent of the gene-expression units and the normalization procedure. Finally, the regulon-activity matrix for all samples was used for PCA, *t*-SNE, connective specificity index (CSI) network[14] (visualized by Cytoscape v.3.5.1[37]) and clustering analysis (*z*-score normalized across the samples before clustering, implemented by Cluster 3.0[38] and visualized as a heat map by TreeView). To build the phylogenetic tree of all samples based on regulon-activity scores, we first performed *z*-score normalization for each regulon across all samples and then calculated the correlation similarity for any two samples. Finally, the pairwise Euclidean distance calculated from similarity matrix was fed into MEGA[39] to build the UPGMA tree under the Phylogeny menu.

Further, WGCNA[36] and pseudotime analyses[40] based on batch-corrected transcriptome level were included to verify clusters and lineage trees that build on regulons. Both analyses suggested that mesoderm was closely related to primitive streak in development. Of note, early endoderm (visceral endoderm) may have a close relationship to the endoderm in the early mouse embryo.

Identification of DEGs and clustering analysis for each individual embryo. For embryos at E5.5, E6.0 and E6.5, DEGs were identified as follows: (1) calculation of the variance of each expressed gene across all samples and selection of top approximately 5,000 genes as highly variable genes; (2) hierarchical clustering with correlation distance metric based on *z*-score normalized expression of highly variable genes to identify preliminary domains according to distinctly separated dendrogram; and (3) identification of the inter-domain DEGs, on the basis of expression of highly variable genes by pairwise comparisons of preliminary domains using *t*-test[41] ($P < 0.05$) and fold change (FC) (E5.5: FC > 5 or FC < 0.2; E6.0 and E6.5: FC > 10 or FC < 0.1) as the DEGs. For E7.0 and E7.5 samples from different batches, we first used Combat[42] to remove potential batch effects based on expression of all genes ($\log_{10}$-transformed), and then combined top highest and lowest principal component (PC)-loading genes (by using FactoMineR in R) from several selected significant PCs by jackstraw[43] to identify the DEGs (top 500 genes for each of PC1–4). Finally, Kmeans clustering (implemented by Cluster 3.0) was applied to determine the final spatial domains of embryo based on the expression profile of DEGs, and the BIC-SKmeans algorithm[44] was applied to determine the optimal number of gene groups and perform gene clustering analysis based on the *z*-score normalized expression profile of DEGs. The clustering heat map was visualized through TreeView.

Single-cell RNA-seq data processing and deconvolution. Reads from 384 single-cell RNA-seq data (E7.0 stage) were aligned to the mm10 genome by using HISAT (v.2-2.1.0)[45], and gene counts were generated using HTseq (v.0.6.0) with default parameters. We first calculated the percentage of reads mapped to mitochondrial genes for each single cell as a quality-check metric and then filtered cells with mitochondrial percentage larger than 0.2. For each selected cell, all mapped reads were processed by StringTie (v.1.3.3b)[45] to evaluate the gene-expression levels (measured by transcript per million (TPM)) with default parameters. Then, gene-expression values for each cell were transformed as $\log(\mathrm{TPM}+1)$ and cells with fewer than 8,000 detected genes were also discarded. The Seurat package (v.3.0)[46] was used to perform single-cell clustering analysis and UMAP was applied to visualize the results. In brief, we first filtered genes detected in fewer than five cells and then selected highly variable genes to group all cells into three clusters. Then, the expression of several canonical marker genes from different development lineages in mouse were checked, and a cluster that highly expresses primitive streak markers—for example, *T* and *Mixl1*—was removed. Finally, 295 single cells were selected to generate the endodermal cell clusters. Clustering analysis was performed again for above selected endoderm cells to identify potential cell types. To evaluate the proportion of identified cell types in Geo-seq samples, we prepared cell-type labels for each single cell, the single-cell gene-expression matrix and all endoderm Geo-seq samples from E7.0 embryo and applied CIBERSORT[47] to perform cell-type deconvolution analysis with default parameters (without quartile normalization). The proportion of each cell type or cluster were visualized on corn plots.

Functional enrichment analysis and phenotype analysis. We applied g:profiler[48] to perform GO enrichment analysis for each group of regulon transcription factors (Fig. 2a). To systematically investigate the knockout phenotype of regulon transcription factors, we first used modPhEA[49], which incorporates all phenotype ontology information from Mouse Genome Information (MGI), to perform enrichment analysis for regulon transcription factors in each cluster. Then, transcription factors that show knockout phenotype information were selected. We then manually checked the detailed phenotype information of each transcription factor on the MGI website and kept only transcription factors that display strong gastrulation phenotypes. The final selected regulon transcription factors were highlighted and visualized in the regulon network.

The connection of spatial domains based on regulon activity. To build the connection of spatial domains between every two embryos of adjacent developmental stages,

we first averaged the regulon-activity scores for individual spatial domain and then calculated the Pearson correlation coefficient (PCC) of any two domains from two embryos of adjacent stages. After that, the mutual nearest neighbour ($k = 1$) for each domain from one stage against the domains from the adjacent stage was identified and connected based on domain correlation similarity. Finally, we converted the PCC of any two connected spatial domains to 10^{PCC} and visualized the connection map by Sankey plot using Google Charts.

Calculation of specificity score of genes/regulons in spatial domains. The Jensen–Shannon divergence (JSD) algorithm was used to identify spatial-domain-specific genes[50]. Before the application of JSD, a spatial pattern of interest was defined in binary: the value of the corresponding sample was set to 1 if the interested pattern in this sample was expected to be specifically expressed, otherwise the value was set to 0. The specificity score was then calculated on the basis of: (1) normalized both expression value of each gene (P) and predefined value of interested pattern (P) across all samples as $P = V_i/\mathrm{sum}(V)$, in which $i = 1, 2, \ldots n$, and n was the number of samples.

(2) For distribution of each gene P_1 and predefined pattern P_2, the specificity score (S) between P_1 and P_2 was defined as:

$$S = 1 - \sqrt{E\left(\frac{P_1 + P_2}{2}\right) - \frac{E(P_1) + E(P_2)}{2}}$$

$$E(P) = -\sum_{i=1}^{n} p_i \log(p_i)$$

$$P = (p_1, p_2, \ldots, p_n)$$

We used the same method to calculate the specificity score for each regulon in different embryonic spatial domains. To evaluate the significance of regulon specificity in each spatial domain, we permuted the regulon-activity score across all samples 100,000 times, then calculated their specificity scores, and finally a permutation p value was calculated as the number of times that $S_{\mathrm{permutation}} > S_{\mathrm{true}}$ divided by 100,000. For regulon-based analysis in E5.5–E7.5 embryonic embryos, we only used combined significant regulons ($P < 0.01$) that were specifically activated in only E5.5–E7.5 spatial domains.

LncRNA analysis. Annotation for known lncRNAs was based on the GENCODE database (v.M7)[51]. To enhance the consistency of the coordinates of lncRNAs in the annotation file with the corresponding genome, data of another mouse genome from Ensembl database (v.82) were used. All sequencing reads were re-mapped to this genome and the expression level of lncRNAs was quantified by the same methods and parameters as described for coding RNAs. Region specific lncRNAs were identified based on the JSD algorithm described above.

In situ image analysis and comparison. To validate the proximity of Geo-seq data of gene-expression pattern to the whole-mount in situ hybridization (WISH) data of specific genes, WISH image results were digitized using MATLAB-based scripts that were designed to quantify the staining results of WISH. For each WISH specimen, the image data were segmented into layers from up to bottom of epiblast region, with the layer number corresponding to the laser capture sections (8 layers for E6.5 embryo, 11 for E7.0 embryos and 9 for E7.5 embryo, Extended Data Fig. 1h). For each layer, the data were further segmented into subregions (quadrants and sectors). The average coloured pixel density was calculated as the digitized expression level for each subregion. The digital expression value of the gene with its RNA-seq data was compared by: (1) converting the computed digital value and sequenced gene-expression value to integers [1,n] on the basis of the order of their values, in which n is the number of samples in the embryo (for example, converting E6.5 epiblast embryo samples to [1,15]; and (2) calculating Spearman correlation coefficient between these two vectors of the WISH image and RNA-seq data. As the WISH image data could only cover one side of the embryo, only the captured region was analysed in the comparison.

Gene-set activity score analysis. To evaluate the enrichment for a gene set in each sample of individual embryos, we applied similar recovery analysis as SCENIC used to calculate the gene-set activity score, which is measured by the area under the recovery curve across the ranking of all genes in each sample. In brief, the x axis of the recovery curve was the ranking of all genes from high to low based on expression level, and the y axis was the number of genes recovered from the tested gene set. Finally, area under the recovery curve was calculated to measure the activity score of this gene set. This activity score analysis was used to evaluate whether any gene set was enriched at the top of the ranking for each sample.

Signalling pathway enrichment and pluripotency activity analysis. In addition to the previously analysed signalling pathways (BMP, FGF, Nodal and Wnt)[14], two other signalling pathways (Notch and Hippo–Yap) were included in this analysis. In brief, potential signalling-target genes of the BMP, FGF, Nodal, Hippo–Yap, Notch and Wnt pathways were identified by comparing control samples with treatment samples using RankProd[52] ($P < 0.01$ for Hippo–Yap and $P < 0.001$ for others) from published perturbation data (Gene Expression Omnibus accesssion numbers GSE48092, GSE41260, GSE17879, GSE69669, GSE15268 and GSE31544). Fisher's exact test followed by Benjamini–Hochberg correction was applied to determine the significance of overlap of the target genes of signalling pathways in different regulon groups. The enrichment for the target gene set of each signalling pathway was evaluated by gene-set activity score analysis.

Naive, formative and primed pluripotency genes were collated from a published embryonic dataset[1,15] (Supplementary Table 8). To determine the spatial enrichment profile of pluripotency-related genes in E5.5–E7.5 embryos, we used the same gene-set activity score analysis to calculate the activity score of naive, formative and primed gene sets in each sample of E5.5–E7.5 embryos.

Analysis based on transcriptome profile. For all combined samples from different stages of embryos, only the expressed genes were selected and then batch effect was removed by ComBat[42]. After that, the signature genes were identified by: (1) defining development-related genes on the basis of the GO database (for example, genes associated with terms containing keywords such as 'differentiation', 'development' and 'morphogenesis'); (2) combining highest and lowest principal-component-loading genes on the basis of their expression across all combined samples (by using FactoMineR in R) from several selected principal components (evaluated using jackstraw[43]; top 300 highest and lowest principal-component-loading genes were symmetrically selected from each principal component (1–6)). Finally, t-SNE analysis[53] was conducted to project the signature expression matrix of all combined samples onto the 2D space for visualizing location of samples in low space.

Gene co-expression analysis. Guilt-by-association analysis[54] was performed to identify co-expressed genes. Specifically, for all samples in one embryo, the expression values of genes were z-score-normalized across all samples and were then converted to a binary matrix with the following strategy: the binary value was set to 1 if the corresponding z-scored value ≥ 0, otherwise the binary value was 0. Finally, the P values for observing a given co-expression of each gene and the query gene were determined by Fisher's exact test based on a 2×2 contingency table.

Web service and 3D reconstruction. The Geo-seq dataset can be accessed at eGastrulation project (http://egastrulation.sibcb.ac.cn). This web service was constructed for visualizing the transcriptome data of all cell populations of E5.5–E7.5 embryos by using django in Python. Currently, four functions can be used to search the spatial gene-expression pattern ('pattern search by gene'), to retrieve guilt-by-association co-expressed genes of query gene ('gene search by gene'), to identify genes that correlate to a customized spatial pattern ('gene search by pattern')[14], and to calculate the enrichment scores of a query gene list as used for signalling pathway enrichment analysis and pluripotency analysis ('gene activity score'). In addition, 3D reconstruction can be performed in pattern search by gene function for the epiblast–ectoderm data from E5.5–E7.5 embryos by segmenting a 2D image of each germ layer into different defined regions, as in the LCM procedure. For each region, the expression level of each gene was presented according to a colour scheme. The Vaa3D utility[55] was applied to stack all the layers to visualize the 3D expression pattern. Image segmentation was performed by a customized MATLAB script[14].

Validation and functional analysis. *Whole-mount or section in situ hybridization.* Probes were synthesized using the Roche DIG RNA labelling kit (11277073910, Roche Applied science) as previous described[14]. Primers for amplifying probe templates are listed in Supplementary Table 9. In brief, embryos were rehydrated through 75%, 50% and 25% methanol at room temperature, washed three times with DEPC-treated PBS and treated with 10 mg ml^{-1} proteinase K (Life Technologies, catalogue no. AM2548) in PBS for 8 min. After post-fixation for 20 min in 4% paraformaldehyde (Sigma; P6148), approximately 1 μg/ml of digoxigenin-labelled RNA probe (Roche) was incubated with the embryo at 68 °C overnight. The embryos were then washed and stained with anti-digoxigenin for imaging.

Mouse embryonic stem cell culture. E14 mouse embryonic stem cells were cultured in DMEM (Life Technologies) supplemented with 15% FBS (VWR; 95042-108), 2 mM glutamine (Gibco–BRL), 1% non-essential amino acids (MEM/NEAA; Hyclone; 16777-186), 2i-LIF (Millipore; ESG1107) and 1% penicillin–streptomycin (Life Technologies; catalogue no. 15140-122) at 37 °C with 5% CO_2. For neural induction, embryonic stem cells were dissociated with 0.05% trypsin and suspended and cultured in DMEM supplemented with 5% knockout serum replacement (KSR; Life Technologies; 10828028), 2 mM glutamine, 1% non-essential amino acids and 1% penicillin–streptomycin for 6 days. For mesodermal differentiation, cells were suspended and cultured in DMEM supplemented with 10% FBS, 2 mM glutamine, 1% non-essential amino acids and 1% penicillin–streptomycin for 6 days.

Ex vivo culture. E7.0 mouse embryos were collected from pregnant mice and the extra-embryonic tissues were removed. The endoderm layer was treated with trypsin (0.25% trypsin in PBS for 3 min), dissected with 30 G1/2 Precision Glided needle (BD; 305106), and dissociated into single cells using Accutase and cultured in N2B27 medium for 8 h with or without 2.5 μM Yap inhibitor verteporfin (TargetMol; T3112).

At the end of culture, cells were collected and lysed, and cDNA was generated for quantitative PCR (qPCR) test or sequencing using the Geo-seq protocol[30].

LncRNA expression analysis. Total RNA was extracted from differentiating knockout embryonic stem cells, using a RNeasy mini-kit (Qiagen). Total RNA (2 mg) was reverse-transcribed into cDNA using FastQuant RT kit (Tiangen; KR106-01). qPCR with reverse transcription (RT–qPCR) was performed with Realplex2 (Eppendorf) using Stormstar SYBR green qPCR maser mix (DBI; DBI-2144). Cycling conditions: initial denaturation at 95 °C for 5 min, followed by 40 cycles of denaturation at 95 °C for 10 s, annealing at 60 °C for 30 s and extension at 72 °C for 30 s. Gene expression was determined using the $\Delta\Delta C_t$ method. Unpaired *t*-test was used to evaluate the significance of differences. Primers for amplifying lncRNAs were listed in Supplementary Table 9.

CRISPR–Cas9-mediated knockout. For lncRNA knockout, upstream and downstream single guide (sg)RNAs of lncRNAs were designed (Supplementary Table 9). For transcription factor knockout, single sgRNA-target coding-sequence regions were designed. The sgRNAs were designed using the CRISPR design website (http://crispr.mit.edu/). Before transfection, sgRNAs were inserted into the pX330-mCherry vector and 5 μg of plasmid was transfected into E14 cells. After 24 h, 10,000 mCherry-positive cells were sorted and seeded into a 10-cm dish. Individual colonies were picked after 4–6 days and expanded in 24-well plates. Genomic DNA was purified and the genomic deletion was validated by PCR or sequencing.

Cardiomyocyte differentiation. Before differentiation, E14 mouse embryonic stem cells were cultured with leukaemia inhibitory factor (LIF) in MEF-coated dishes and passaged two times. Cells were then collected by trypsinization, and 600 cells in 20 μl drops ($n = 50–60$) of embryonic stem cell suspension were placed on the lids of 100-mm bacteriological Petri dishes containing 10 ml PBS without LIF to generate embryoid bodies. After two days, embryoid body cells were transferred to bacterial culture dishes (60 mm) in 5 ml embryonic stem cell medium without LIF. Embryoid bodies were cultured in suspension for 4 days and subsequently reseeded onto gelatin-coated 24-well plates for an additional 6 days to determine the percentage of spontaneously beating embryoid bodies and to perform immunostaining. RT–qPCR was performed at day 6 and day 12.

Immunostaining. For cell immunostaining, the cells were washed once with PBS and fixed with 4% paraformaldehyde at room temperature for 20 min. The fixed cells were permeabilized with 0.1% Triton X-100 for 10 min. The cells were then blocked with PBS containing 10% donkey serum for 1 h at room temperature, subsequently incubated with the appropriate dilution (1:1,000) of primary cTnT antibody (Abcam, ab8295) at 4 °C overnight, then washed with PBS 3 times and incubated with Alexa Fluor 594-conjugated secondary antibody (Invitrogen, A-11032) for 1 h, and then counterstained with DAPI for 10 min and washed with PBS 3 times. Finally, the cells were examined under a fluorescence microscope.

Reporting summary. Further information on research design is available in the Nature Research Reporting Summary linked to this paper.

Data availability

The RNA-seq data generated in this study were deposited in the NCBI Gene Expression Omnibus under accession number GSE120963 and the NODE project under accession number OEP000320. Our resources can be explored at the eGastrulation web portal (http://egastrulation.sibcb.ac.cn). All other data are available from the corresponding authors upon request.

Code availability

Custom codes for generating corn plots for different stages can be obtained from the eGastrulation web portal (http://egastrulation.sibcb.ac.cn).

28. Rivera-Pérez, J. A., Jones, V. & Tam, P. P. Culture of whole mouse embryos at early postimplantation to organogenesis stages: developmental staging and methods. *Methods Enzymol.* **476**, 185–203 (2010).
29. Downs, K. M. & Davies, T. Staging of gastrulating mouse embryos by morphological landmarks in the dissecting microscope. *Development* **118**, 1255–1266 (1993).
30. Chen, J. et al. Spatial transcriptomic analysis of cryosectioned tissue samples with Geo-seq. *Nat. Protoc.* **12**, 566–580 (2017).
31. Cui, G. et al. Spatio-temporal transcriptome construction of early mouse embryo with Geo-seq and Auto-seq. *Protoc. Exch.*, https://doi.org/10.21203/rs.2.10081/v1 (2019).
32. Wells, J. M. & Melton, D. A. Early mouse endoderm is patterned by soluble factors from adjacent germ layers. *Development* **127**, 1563–1572 (2000).
33. Liu, Q. et al. Lung regeneration by multipotent stem cells residing at the bronchioalveolar-duct junction. *Nat. Genet.* **51**, 728–738 (2019).
34. Kim, D. et al. TopHat2: accurate alignment of transcriptomes in the presence of insertions, deletions and gene fusions. *Genome Biol.* **14**, R36 (2013).
35. Trapnell, C. et al. Differential analysis of gene regulation at transcript resolution with RNA-seq. *Nat. Biotechnol.* **31**, 46–53 (2013).
36. Langfelder, P. & Horvath, S. WGCNA: an R package for weighted correlation network analysis. *BMC Bioinformatics* **9**, 559 (2008).
37. Shannon, P. et al. Cytoscape: a software environment for integrated models of biomolecular interaction networks. *Genome Res.* **13**, 2498–2504 (2003).
38. de Hoon, M. J., Imoto, S., Nolan, J. & Miyano, S. Open source clustering software. *Bioinformatics* **20**, 1453–1454 (2004).
39. Kumar, S., Stecher, G. & Tamura, K. MEGA7: Molecular Evolutionary Genetics Analysis version 7.0 for bigger datasets. *Mol. Biol. Evol.* **33**, 1870–1874 (2016).
40. Qiu, X. et al. Reversed graph embedding resolves complex single-cell trajectories. *Nat. Methods* **14**, 979–982 (2017).
41. Klein, C. A. et al. Combined transcriptome and genome analysis of single micrometastatic cells. *Nat. Biotechnol.* **20**, 387–392 (2002).
42. 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).
43. Chung, N. C. & Storey, J. D. Statistical significance of variables driving systematic variation in high-dimensional data. *Bioinformatics* **31**, 545–554 (2015).
44. Zhang, W. et al. Integrating genomic, epigenomic, and transcriptomic features reveals modular signatures underlying poor prognosis in ovarian cancer. *Cell Rep.* **4**, 542–553 (2013).
45. Pertea, M., Kim, D., Pertea, G. M., Leek, J. T. & Salzberg, S. L. Transcript-level expression analysis of RNA-seq experiments with HISAT, StringTie and Ballgown. *Nat. Protocols* **11**, 1650–1667 (2016).
46. Stuart, T. et al. Comprehensive integration of single-cell data. *Cell* **177**, 1888–1902 (2019).
47. Newman, A. M. et al. Robust enumeration of cell subsets from tissue expression profiles. *Nat. Methods* **12**, 453–457 (2015).
48. Reimand, J. et al. g:Profiler-a web server for functional interpretation of gene lists (2016 update). *Nucleic Acids Res.* **44** (W1), W83–W89 (2016).
49. Weng, M. P. & Liao, B. Y. modPhEA: model organism Phenotype Enrichment Analysis of eukaryotic gene sets. *Bioinformatics* **33**, 3505–3507 (2017).
50. Cabili, M. N. et al. Integrative annotation of human large intergenic noncoding RNAs reveals global properties and specific subclasses. *Genes Dev.* **25**, 1915–1927 (2011).
51. Harrow, J. et al. GENCODE: the reference human genome annotation for The ENCODE Project. *Genome Res.* **22**, 1760–1774 (2012).
52. Hong, F. et al. RankProd: a bioconductor package for detecting differentially expressed genes in meta-analysis. *Bioinformatics* **22**, 2825–2827 (2006).
53. van der Maaten, L. & Hinton, G. Visualizing data using *t*-SNE. *J. Mach. Learn. Res.* **9**, 2579–2605 (2008).
54. Walker, M. G., Volkmuth, W., Sprinzak, E., Hodgson, D. & Klingler, T. Prediction of gene function by genome-scale expression analysis: prostate cancer-associated genes. *Genome Res.* **9**, 1198–1203 (1999).
55. Peng, H., Ruan, Z., Long, F., Simpson, J. H. & Myers, E. W. V3D enables real-time 3D visualization and quantitative analysis of large-scale biological image data sets. *Nat. Biotechnol.* **28**, 348–353 (2010).
56. Richardson, L., et al. EMAGE mouse embryo spatial gene expression database: (2014 update). *Nucleic Acids Res.* **42**, D835–D844 (2014).

Acknowledgements The authors thank Q. Zhou for critical discussions, Y. Chen and Y. Xu for technical support, and the chemical biology core facility and animal core facility in SIBCB. This work was supported in part by the 'Strategic Priority Research Program' of the Chinese Academy of Sciences (Grants, XDA16020501 to N.J., XDA16020404 to G.P., and XDA01010303 and XDB19020301 to J.-D.J.H.), National Key Basic Research and Development Program of China (2018YFA0107201 to G.P., 2018YFA0800100, 2018YFA0108000, 2017YFA0102700 and 2015CB964500 to N.J., and 2015CB964803 and 2016YFE0108700 to J.-D.J.H.), National Natural Science Foundation of China (31871456 to G.P., 31661143042, 91519314, 31630043, 31571513 and 31430058 to N.J., and 91329302, 31210103916 and 91749205 to J.-D.J.H.), Science and Technology Planning Project of Guangdong Province, China (2017B030314056), Frontier Research Program of Guangzhou Regenerative Medicine and Health Guangdong Laboratory (2018GZR110105013), Shanghai Natural Science Foundation (18ZR1446200) and Shanghai Municipal Commission for Science and Technology Grants (17411954900). P.P.L.T. is a NHMRC Senior Principal Research Fellow supported by the National Health and Medical Research Council of Australia (Research Fellowship grant 1110751).

Author contributions N.J. and G.P. conceived the study. N.J., G.P. and J.-D.J.H. supervised the project. N.J., P.P.L.T. and G.P. designed the experiments. G.P., G. Cui, J.C., S.C. and Z.L. performed the Geo-seq of embryos. Y.Q. performed animal husbandry. S.S. analysed the sequencing data and built the web portal. F.Y. conducted the embryonic stem cells and ex vivo perturbation experiments. R.W. helped to generate the pipeline for all germ-layer corn plot display. G. Chen helped with data analysis. G.P., P.P.L.T. and S.S. wrote the paper with the help of all other authors.

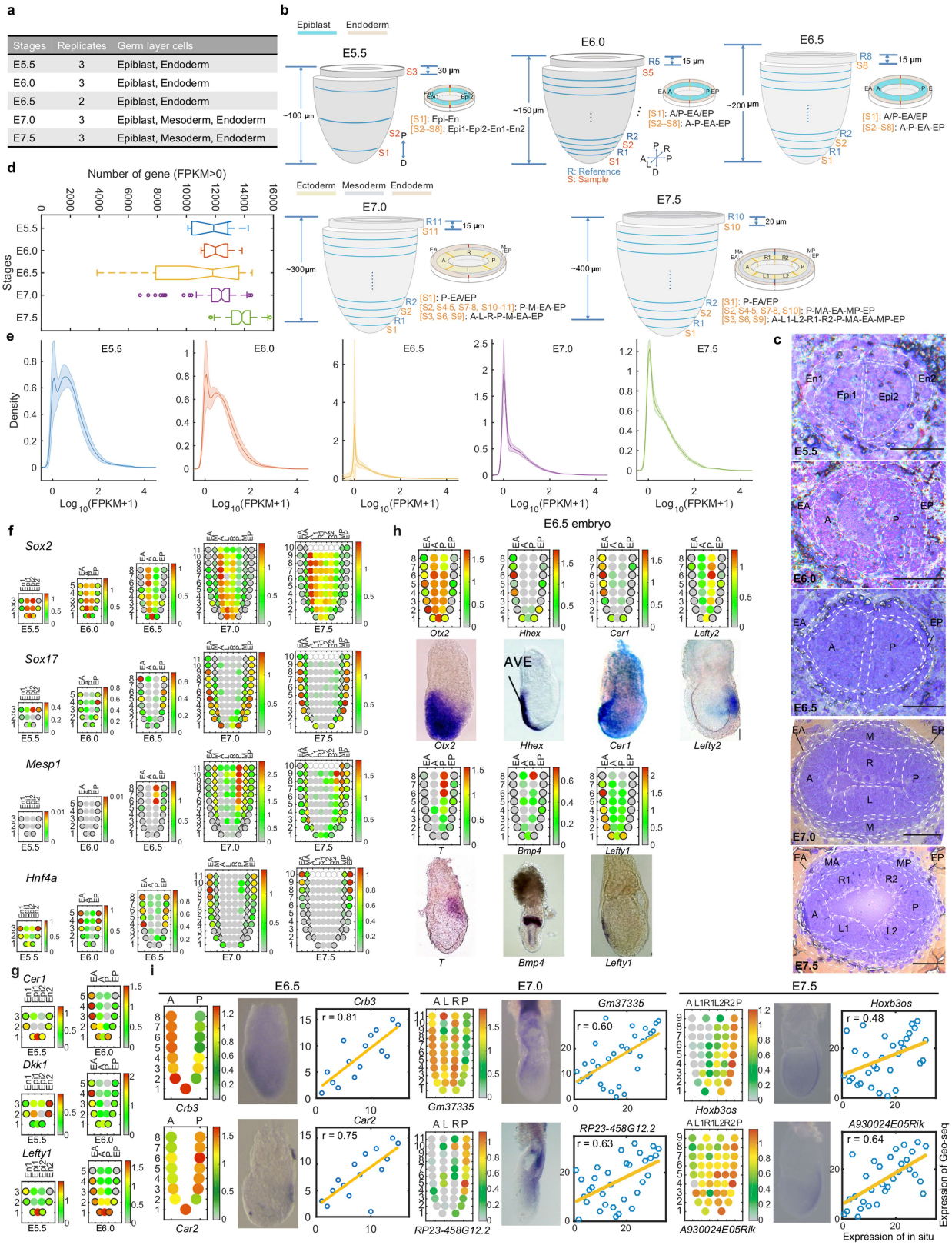
Competing interests The authors declare no competing interests.

Additional information

Supplementary information is available for this paper at https://doi.org/10.1038/s41586-019-1469-8.

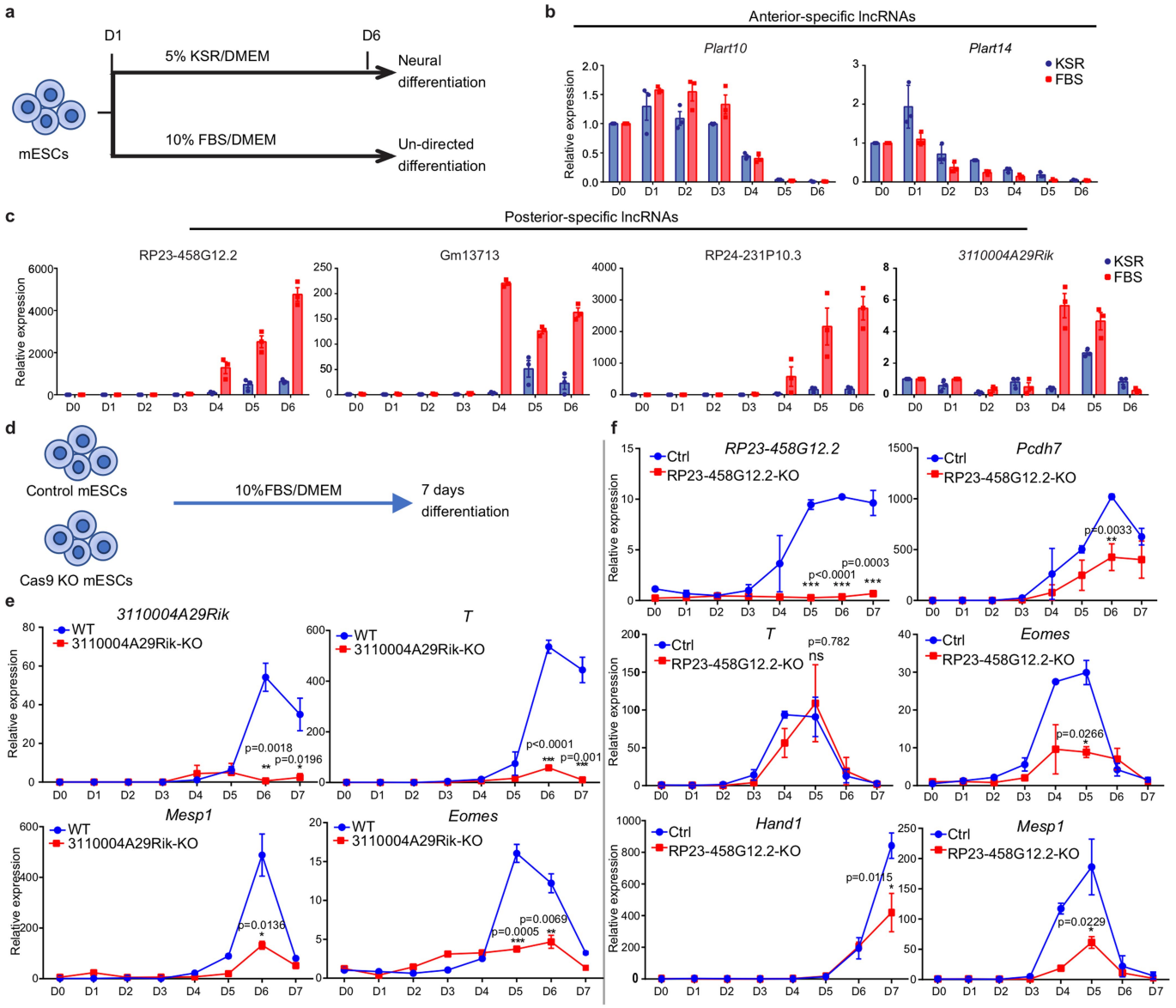
Correspondence and requests for materials should be addressed to G.P., J.-D.J.H. or N.J.

Peer review information *Nature* thanks Thorsten Boroviak, Anna-Katerina Hadjantonakis, Kyle Loh and Valerie Wilson for their contribution to the peer review of this work.

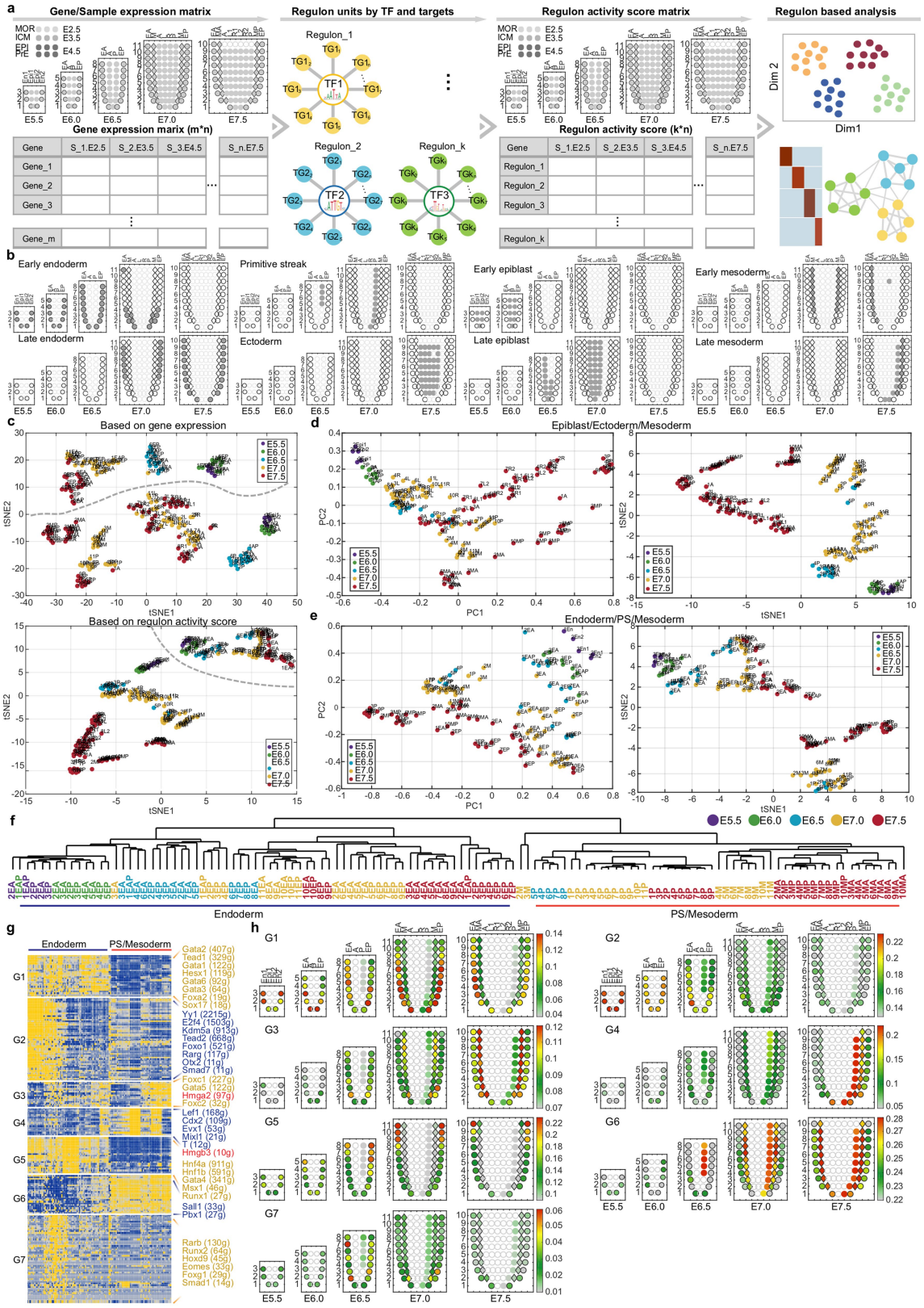


Extended Data Fig. 1 | See next page for caption.

Extended Data Fig. 1 | Geo-seq analysis. a, Samples for Geo-seq: embryonic stages, biological replicates and the germ layers that were sampled for Geo-seq analysis. Stages: E5.5–E6.0, pre-gastrulation stage; E6.5, early streak; E7.0, late mid-streak; E7.5, no-bud stage. See Supplementary Table 1 for complete information. **b**, The strategy of sampling of cell populations in the epiblast–ectoderm and endoderm from E5.5 to E7.5 embryos and mesoderm in the E7.0 to E7.5 embryos (outlined in **c**). Samples were designated by the sequential (ascending) order of serial sections (1 for the most distal section) and the regions in the section (R, reference section; S, sample section; see histology images for orientation). For E5.5 embryo, the anterior–posterior axis has not yet been determined, so samples were denoted as Epi 1/2 and En 1/2. For E5.5–E7.5 embryos: Epi, epiblast; En, endoderm; A, anterior; P, posterior; AP, pooled from anterior and posterior; L, left side (1, anterior; 2, posterior); R, right side (1, anterior; 2, posterior); embryonic axes: anterior–posterior, A↔P; proximal–distal, P↔D; left–right L↔R. Scale bar, 50 μm. **d**, Box plot showing the number of detected genes (FPKM > 0) for samples of E5.5 to E7.5 embryos (E5.5: $n = 10$; E6.0: $n = 18$; E6.5: $n = 30$; E7.0: $n = 73$; E7.5: $n = 83$). The centre line marks the median and box edges represent 25th and 75th percentiles. **e**, Gene-expression density plot of Geo-seq data of samples from E5.5 to E7.5 embryo; the solid line is the mean value and shaded region represents the s.d. **f**, Corn plots showing the expression of representative germ-layer markers (ectoderm: *Sox2*; endoderm: *Sox17*; mesoderm: *Mesp1*; visceral endoderm: *Hnf4a*) at E5.5 to E7.5 stages. Similar results were obtained from embryo replicates. **g**, The expression of anterior visceral endoderm (AVE) and distal visceral endoderm (DVE) markers (*Cer1*, *Dkk1* and *Lefty1*) at E5.5 and E6.0 stages. Similar results were obtained from embryo replicates. **h**, The comparison of digitally rendered gene-expression pattern and WISH results of seven examples of genes in the epiblast and endoderm at E6.5 stage. WISH images were obtained from the EMAP eMouse Atlas Project (www.emouseatlas.org)[56]. Embryos are oriented with posterior to the right and distal to the bottom. **i**, The comparison of the expression of selected coding and noncoding RNAs by Geo-seq and WISH in E6.5 epiblast (*Crb3*, *Car2*), E7.0 (*Gm37335*, *RP23-458G12.2*) and E7.5 ectoderm (*Hoxb3os*, *A930024E05Rik*). $n > 3$ for each gene. Spearman correlation plots of WISH and Geo-seq results are shown on the right. Embryos are oriented with posterior to the right and distal to the bottom.

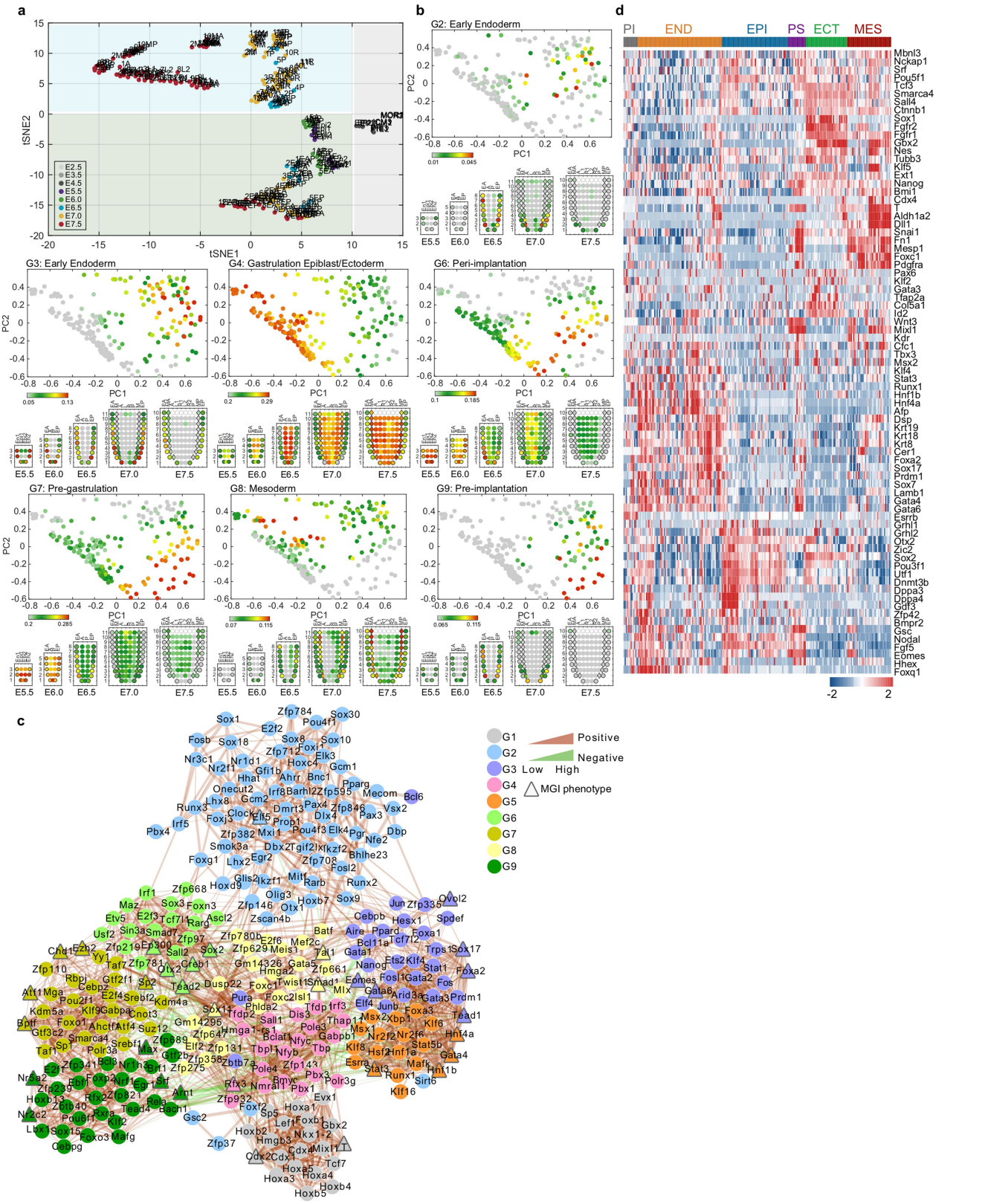


Extended Data Fig. 2 | Functional analysis of lncRNA. a, Experimental design for tracking the expression of lncRNAs during in vitro differentiation of mouse embryonic stem cells (mESCs). **b**, **c**, Expression of anterior-specific (**b**) and posterior-specific lncRNAs (**c**), assayed by real-time PCR, during in vitro differentiation (days 0–6) of mouse embryonic stem cells ($n = 3$ for KSR and FBS). **d**, Experimental design for assessing the effect of the loss of lncRNA function on the differentiation of mouse embryonic stem cells in vitro. **e**, Ablation of lncRNA *3110004A29Rik* led to downregulation of mesoderm genes: *T*, *Eomes* and *Mesp1* at days 5 and 6 of culture ($n = 3$). **f**, Ablation of lncRNA *RP23-458G12.2* disrupted the expression of mesendoderm genes: *Pcdh7*, *Eomes*, *Mesp1* and *Hand1* at later time points of the seven-day period of differentiation in vitro ($n = 3$). Gene expression was assayed by real-time PCR. Data are mean ± s.e.m. Significant difference by two-sided *t*-test; $^{*}P < 0.05$, $^{**}P < 0.01$ and $^{***}P < 0.001$.

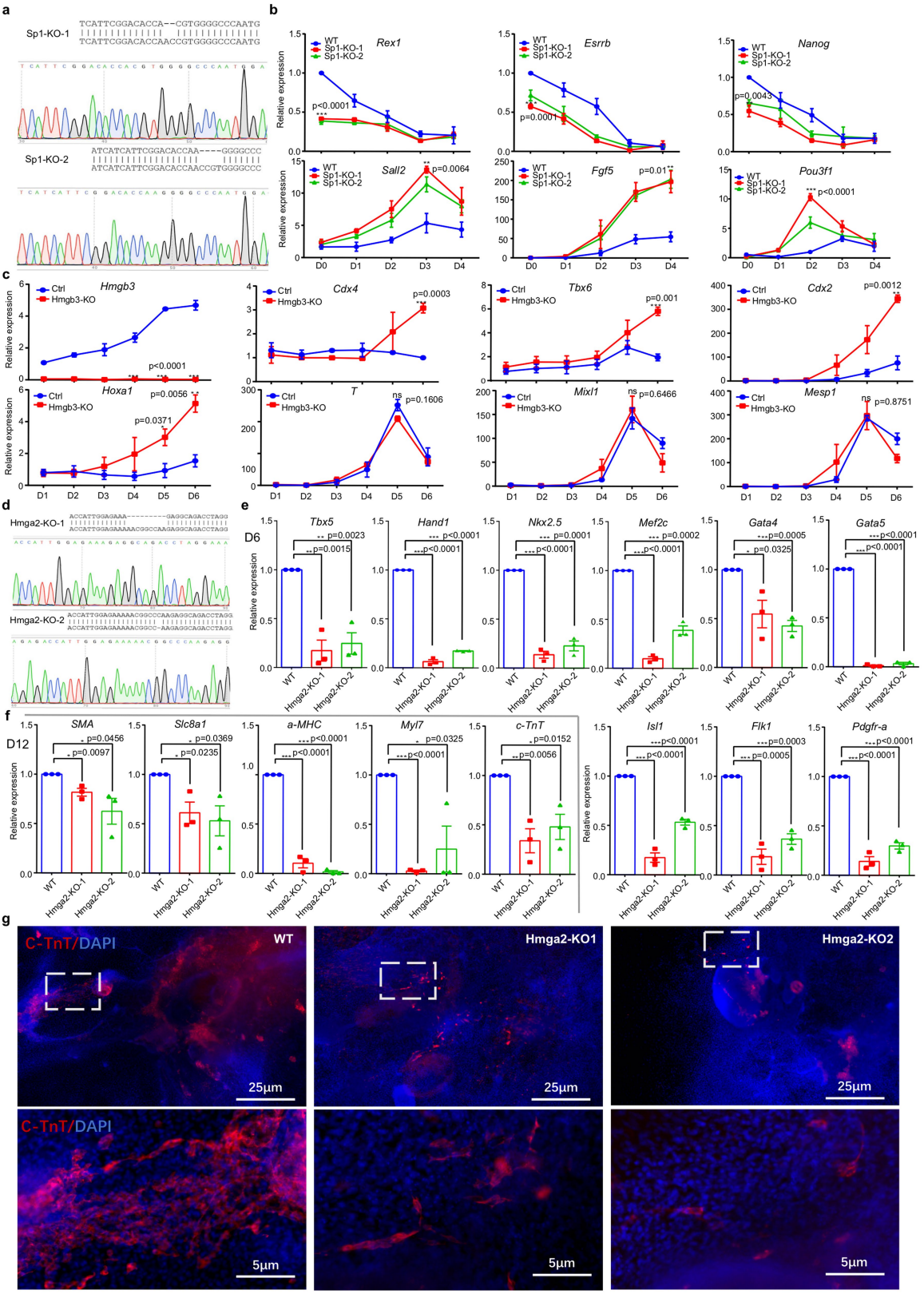


Extended Data Fig. 3 | See next page for caption.

Extended Data Fig. 3 | The segregation of germ-layer tissues by the regulon and transcriptome. a, The workflow of regulon analysis in the SCENIC pipeline. Tissue samples from embryos at E2.5 to E7.5 were collated for mining the co-expression genes, *cis*-regulatory motif analysis and building of regulon units comprising of transcription factors (TF) and targets (TG), and the regulon-activity matrix for each sample were used in subsequent bioinformatic analysis (Methods). Dim, dimension. Transcriptome data of blastomeres of the morula, inner cell mass of the blastocyst and the primitive endoderm and epiblast of E4.5 embryo were sourced from ArrayExpress accession numbers E-MTAB-2958 and E-MTAB-2959. **b**, Corn plots show the spatial domain of the cell populations of each branch of the phylogenetic tree at E5.5–E7.5. **c**, *t*-SNE plots of tissue samples from E5.5 to E7.5 based on gene expression and the regulon-activity matrix ($n = 214$). **d**, **e**, PCA and *t*-SNE analysis for regulon-activity scores in the epiblast–ectoderm and mesoderm ($n = 145$) (**d**) and the primitive streak, mesoderm and endoderm of E5.5 to E7.5 embryos ($n = 122$) (**e**). **f**, Hierarchical clustering analysis for the posterior epiblast/primitive streak, the mesoderm and the endoderm for embryonic samples from E5.5 to E7.5 stages. **g**, The regulon-activity heat map for primitive streak, mesoderm and endoderm of E5.5 to E7.5 embryo. Sample in each column was ordered based on **f**, and examples of regulon transcription factors were listed. Regulon transcription factors coloured red were experimentally tested in this study. **h**, Corn plots showing the pattern of averaged activities of regulon groups in the endoderm, mesoderm and primitive streak samples.

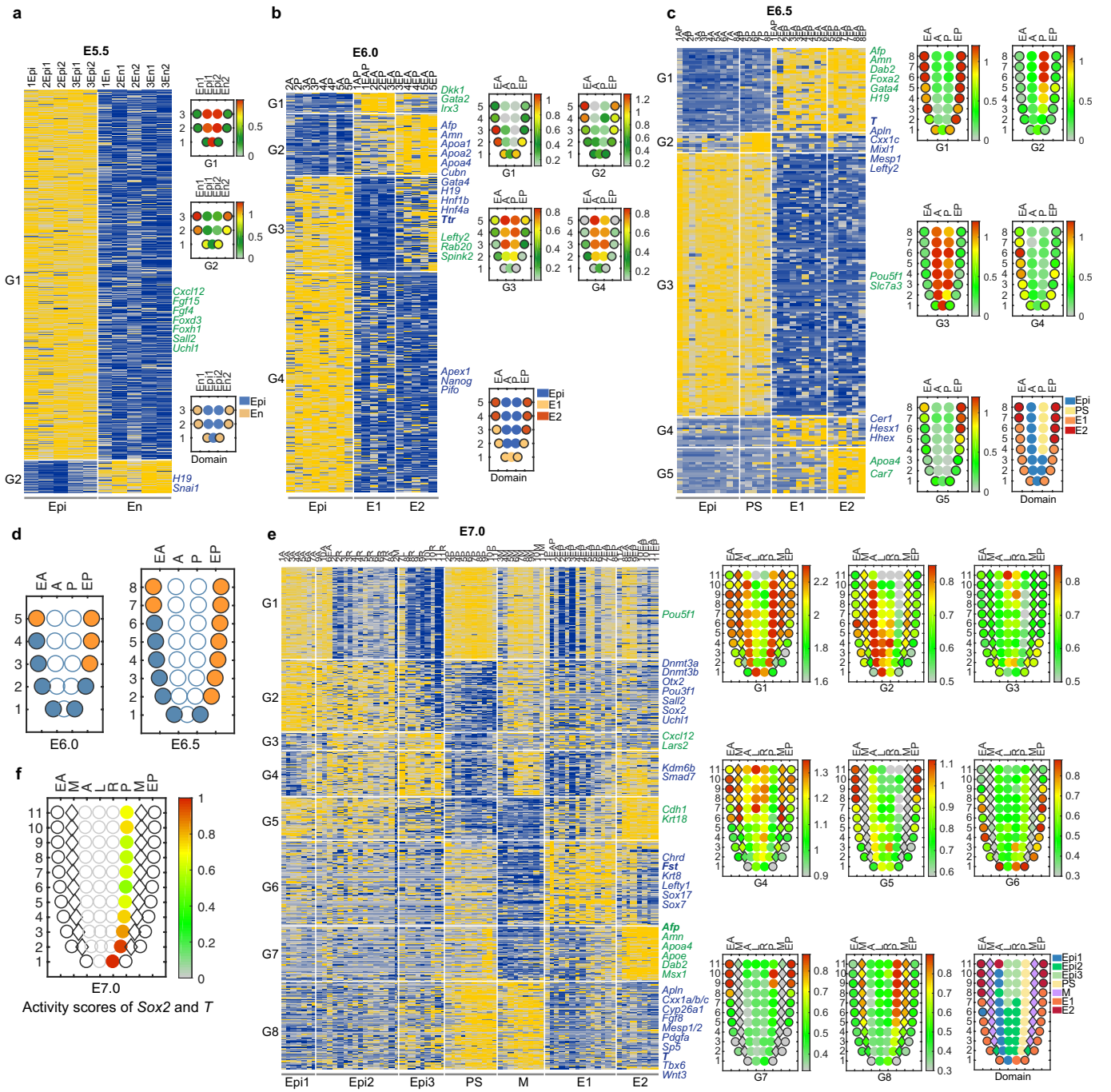


Extended Data Fig. 4 | The regulon sample gene sets. a, *t*-SNE distribution of tissue samples of E2.5 to E7.5 embryos based on regulon-activity scores ($n = 226$). **b**, Two-dimensional PCA plots ($n = 226$) and corn plots showing the averaged regulon activities of G2, G3, G4, G6, G7, G8 and G9 regulon groups. **c**, The CSI co-expression network analysis (edges were filtered by CSI >0.85) for the nine regulon groups. Edges: positive interaction (brown), negative interaction (green). Triangle nodes denote transcription factors with knockout gastrulation phenotype in MGI database. A wider edge signifies higher correlation. **d**, The gene-expression profile of lineage markers in the different regulon sample groups.

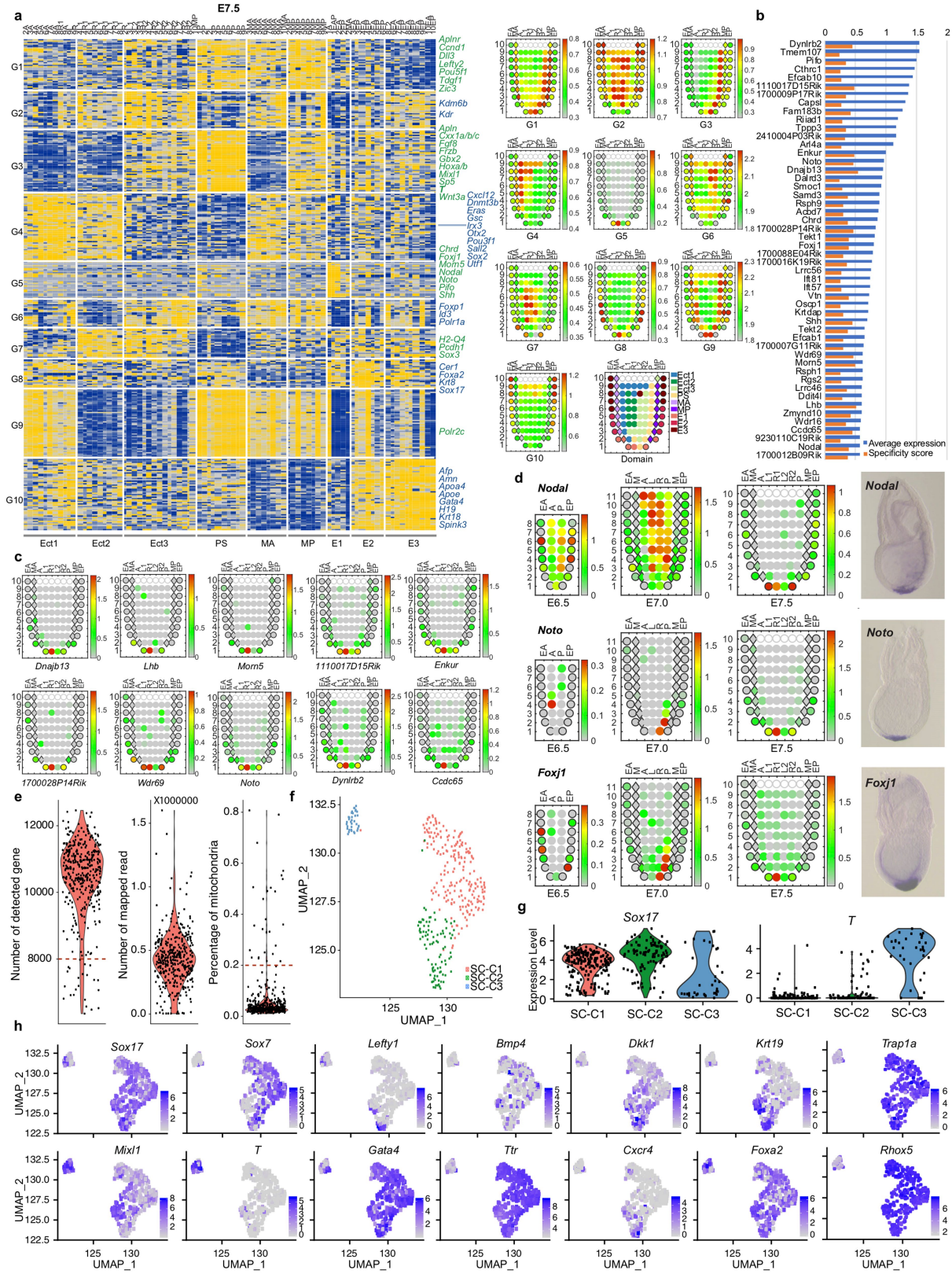


Extended Data Fig. 5 | See next page for caption.

Extended Data Fig. 5 | Functional analysis of *Sp1* of the G7 (pre-gastrulation), *Hmgb3* of the G1 (late mesoderm/primitive streak) regulon group and *Hmga2* of G8 (mesoderm) regulon group on differentiation of the mouse embryonic stem cells. a, DNA sequence of two *Sp1*-knockout cell lines. **b**, Loss of *Sp1* function led to downregulation of the naive marker genes (*Rex1*, *Esrrb* and *Nanog*) and upregulation of formative marker genes (*Sall2*, *Fgf5* and *Pou3f1*) ($n = 3$). **c**, Loss of *Hmgb3* function led to upregulation of *Cdx2*, *Cdx4*, *Hoxa1* and *Tbx6* genes related to precursors of the trunk mesoderm and no changes to markers of mesendoerm progenitors. Gene expression was assayed by qPCR. Data are mean ± s.e.m. ($n = 3$). Significant difference by two-sided *t*-test; $^{*}P < 0.05$, $^{**}P < 0.01$ and $^{***}P < 0.001$. **d**, DNA sequence and sequencing peak map of two *Hmga2*-knockout cell lines. **e**, Knockout of *Hmga2* led to down regulation of cardiac progenitor marker genes at differentiation day 6 ($n = 3$). **f**, Knockout of *Hmga2* led to down regulation of cardiomyocyte marker genes at differentiation day 12 ($n = 3$). **g**, Immunostaining of embryoid body at day 12 shows lower expression of cardiomyocyte marker c-TnT in *Hmga2*-knockout cells (scale bars: top, 25 μm; bottom, 5 μm) ($n = 3$). Data are mean ± s.e.m. Significant difference by two-sided *t*-test; $^{*}P < 0.05$, $^{**}P < 0.01$ and $^{***}P < 0.001$.



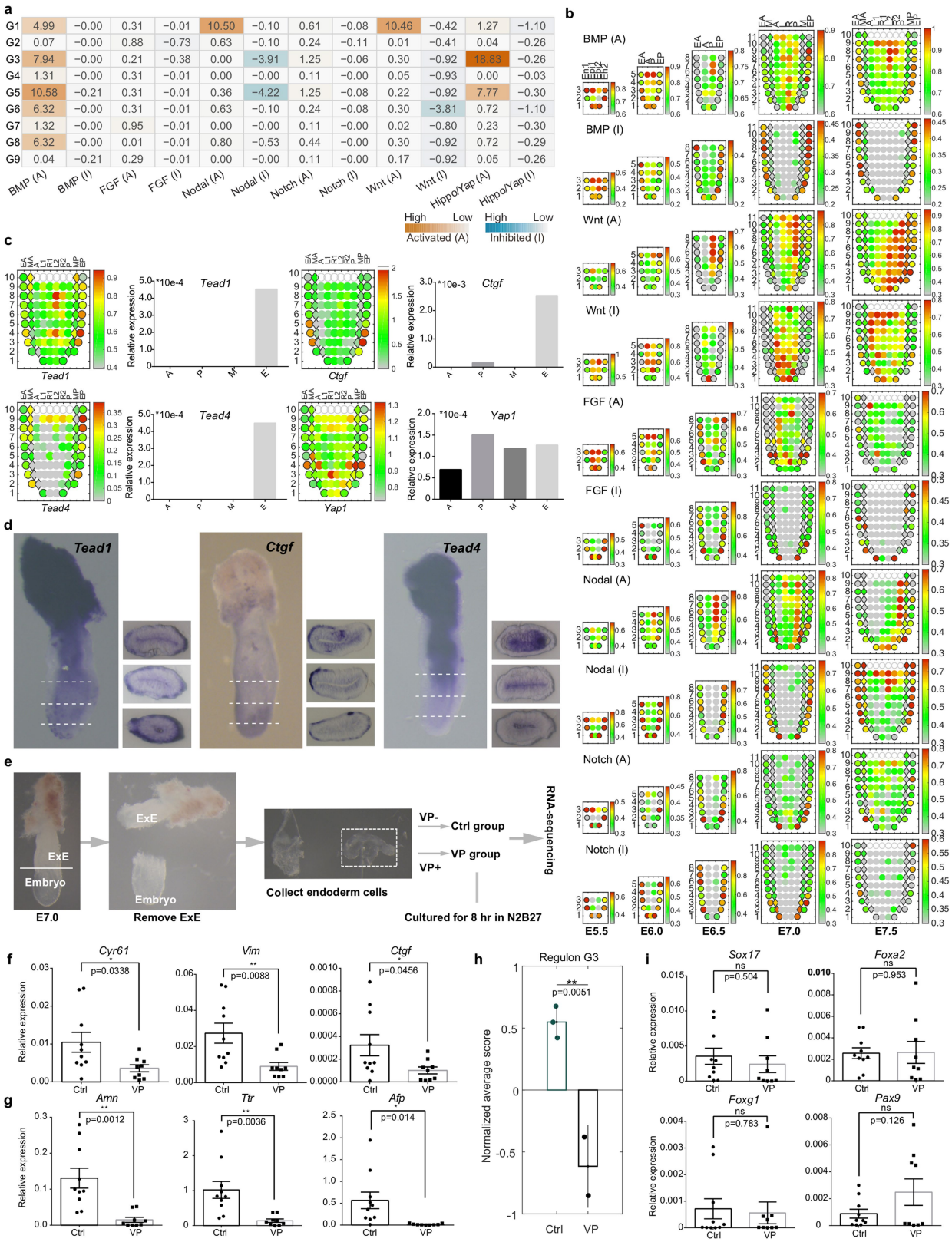
Extended Data Fig. 6 | The spatial domains in the germ layers of E5.5–E7.0 embryos delineated by the DEG groups. a–c, Heat maps and corn plots of DEG groups in E5.5 (**a**), E6.0 (**b**) and E6.5 (**c**) embryos. **d**, The transcriptional endoderm domains at E6.0 and E6.5 identified by analysis based on guilt-by-association co-expressed genes of *Dkk1*, *Cer1* and *Lefty1*. **e**, Heat map and corn plots of DEG groups in E7.0 embryos. Representative genes of each gene group are listed. Corn plots show the spatial domains and the average expression level of the DEG groups. **f**, The activity score of *Sox2* and *T* in the E7.0 primitive streak region.



Extended Data Fig. 7 | See next page for caption.

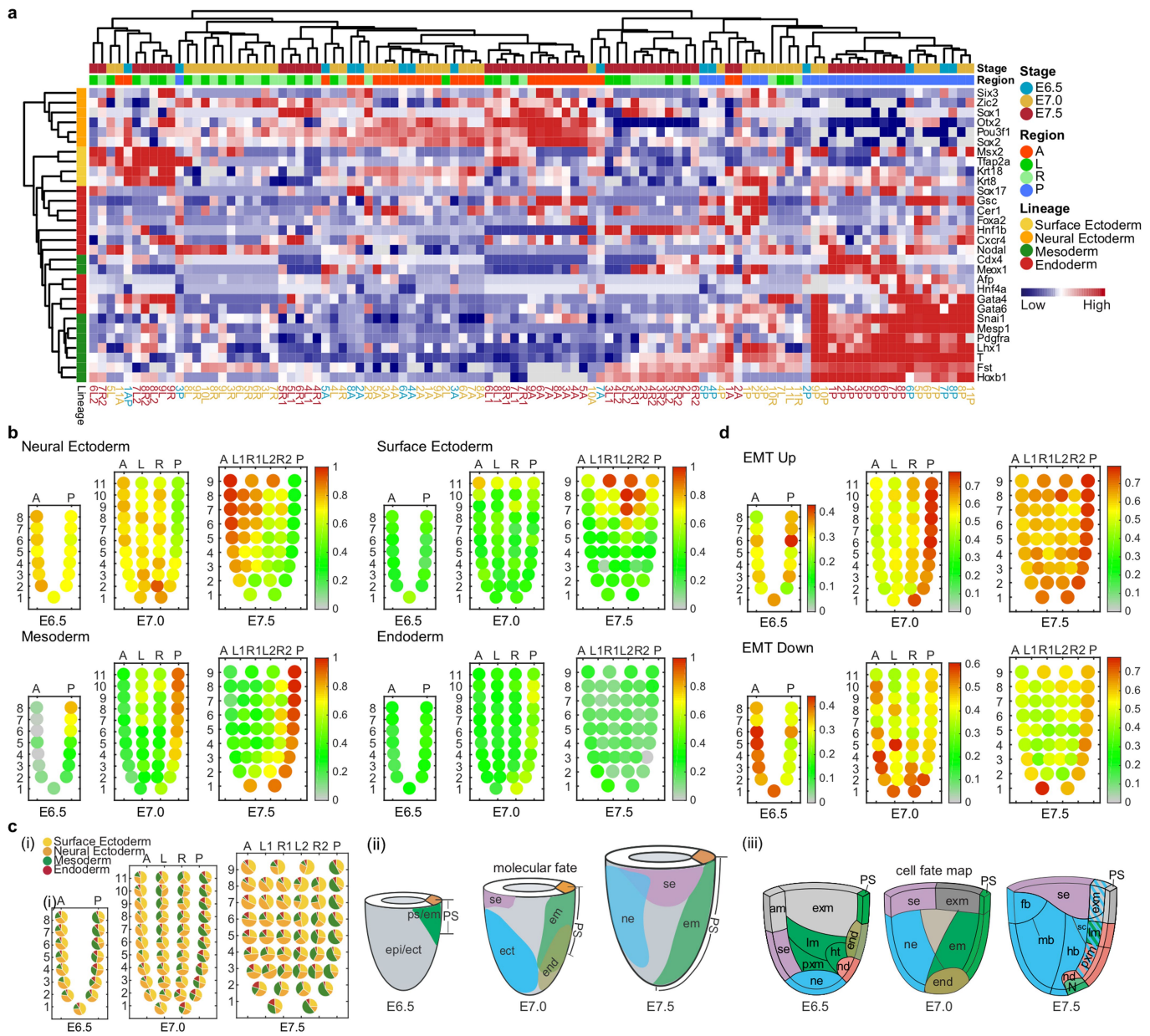
Extended Data Fig. 7 | The tissue spatial domains in the germ layers of E7.5 embryo and endoderm single-cell proportion in the endoderm. **a**, Heat map and corn plots of DEG groups at the E7.5 stage. Representative genes of each DEG group are listed. Corn plots show the spatial domains and the average expression level of the gene group. **b**, Genes that are specifically expressed in the node region (DEG G5) of E7.5 embryos with high expression values and the corresponding JSD specificity score (in the distal-endoderm domain). **c**, Top ten (with the highest JSD scores) examples of node-specific expression pattern shown in corn plots. **d**, The dynamic expression of three E7.5 G5 genes during gastrulation shown in corn plots and compared with WISH results at E7.5 stage. Embryos are oriented with posterior to the right and distal to the bottom ($n > 3$ for each gene). **e**, Gene-filter criteria of single-cell RNA-seq data ($n = 384$ cells; top and bottom of each violin plot represent maxima and minima). Cells with number of detected genes less than 8,000 were discarded (left, red dotted line) and cells with the percentage of reads mapped to mitochondrial genes less than 0.2 were reserved for further analysis (right, red dotted line). **f**, UMAP plots of 345 single cells (that passed the quality check) in three clusters, indicted by different colours. **g**, Violin plots of gene-expression level of *Sox17* and *T* across three clusters (SC-C1, $n = 213$ cells; SC-C2, $n = 96$ cells; SC-C3, $n = 36$ cells; top and bottom of each violin plot represent maximum and minimum, respectively). **h**, Feature plots of represented lineage markers ($n = 345$ cells). The colour gradient indicates levels of expression.

Extended Data Fig. 8 | The inter-embryo correlation of spatial domains. The Spearman correlation coefficient of spatial domains (from Extended Data Figs. 6, 7) between embryonic replicates at E5.5 (**a**), E6.0 (**b**), E6.5 (**c**), E7.0 (**d**) and E7.5 (**e**) stages based on regulon-activity scores. For E7.0 replicate 2 and E7.5 replicate 2, only epiblast domains were assessed.



Extended Data Fig. 9 | See next page for caption.

Extended Data Fig. 9 | Signalling activity in the post-implantation embryo. a, Enrichment for target and response genes of signalling pathways in the nine regulon gene groups (from Fig. 2a). Brown shows activating (A) and blue shows inhibitory (I) signalling activity. The significance of $-\log_{10}$(FDR) value in each cell was calculated by one-sided Fisher's exact test followed by Benjamini–Hochberg correction. **b**, Corn plots showing the activity scores of the target and response genes related to the activated and inhibitory activity of BMP, Wnt, FGF, Nodal and Notch pathways in all germ-layer samples of E5.5 to E7.5 embryos. Similar results were obtained from embryo replicates. **c**, The expression of four Hippo–Yap signalling pathway components shown in corn plots of E7.5 embryo and in the dissected germ-layer tissues assayed by qPCR ($n = 3$). Dissected tissues: A, anterior epiblast, P, posterior epiblast, M, mesoderm; E, endoderm. **d**, WISH of E7.0 embryo (left) and visualization of hybridization signals in the germ layers (right) of Hippo pathway components *Tead1* and *Tead4*, and *Ctgf*, the *Yap1* downstream target gene ($n > 3$ for each gene). Dotted lines indicate plane of sectioning. **e**, The experimental design of ex vivo culture and treatment by inhibitor. E7.0 endoderm layer was explanted, treated with or without Yap inhibitor (verteporfin, VP) in N2B27 for 8 h and prepared for cDNA amplification and Geo-seq analysis. ExE, extra-embryonic tissues. **f**, **g**, The expression of Hippo–Yap pathway factors (*Ctgf*, $P = 4.56 \times 10^{-2}$; *Cyr61*, $P = 3.38 \times 10^{-2}$ and *Vim*, $P = 8.8 \times 10^{-3}$) (**f**) and visceral endoderm markers (*Afp*, $P = 1.4 \times 10^{-2}$; *Amn*, $P = 1.2 \times 10^{-3}$ and *Ttr*, $P = 3.6 \times 10^{-3}$) (**g**) following verteporfin treatment of endoderm explants. Data are mean ± s.e.m. Control, $n = 10$; verteporfin, $n = 9$. Significant difference by two-sided *t*-test; $^{*}P < 0.05$ and $^{**}P < 0.01$. **h**, Changes in the G3 regulon-activity scores following Yap inhibitor treatment in endoderm explants (regulon-activity scores were *z*-score-normalized across all verteporfin and control samples). Data are mean ± s.d. One-tailed *t*-test; $P = 5.1 \times 10^{-3}$; control, $n = 3$; verteporfin, $n = 2$. **i**, No effect on definitive endoderm genes (*Sox17*, *Foxa2*, *Foxg1* and *Pax9*) after 8 h Yap inhibition of endoderm tissues ex vivo. Data are mean ± s.e.m. wild type, $n = 10$; verteporfin, $n = 9$.



Extended Data Fig. 10 | The molecular fates and correlation with experimental cell fate maps. a, Heat map showing the expression pattern of germ-layer marker genes. Top coloured bars indicate the developmental stage (stages: cyan, E6.5; yellow, E7.0; dark red, E7.5) and the location of the cell population in epiblast–ectoderm (region: red, anterior; blue, posterior; green, left; light green, right). The lineage fates (left bar) were assigned on the basis of the expression of markers of neuroectoderm (orange), surface ectoderm (yellow), mesoderm (green) and endoderm (red). **b**, Corn plots showing the activity scores of different lineage markers in embryos at different stages. **c**, (i) Corn plots showing the relative level of expression of germ-layer-related genes (as pie charts in the corn plots) in different cell populations of the epiblast–ectoderm in E6.5–E7.5 embryos; (ii) the molecular fate maps of major germ-layer derivatives in E6.5–E7.5 epiblast–ectoderm constructed based on the expression pattern of tissue markers; and (iii) the prospective fate maps (shown for the right lateral half of the epiblast–ectoderm) constructed from the data of the fate-mapping and lineage tracking experiments. Abbreviations: ne, neural ectoderm; fb, forebrain; mb, midbrain; hb, hindbrain; sc, spinal cord; se, surface ectoderm; am, amnion ectoderm; em, embryonic mesoderm; exm, extra-embryonic mesoderm; ht, heart; lm, lateral mesoderm; meso, mesoderm; pxm, paraxial mesoderm; end, endoderm; N, node; nd, notochord. **d**, Average expression of genes associated with epithelial–mesenchymal transition (EMT) showing enrichment of genes associated with enhanced EMT in the posterior epiblast (EMT up) and reduced EMT in the anterior epiblast (EMT down) in E6.5 to E7.5 embryos.

nature research

Corresponding author(s): Dr. Naihe Jing

Reporting Summary

Nature Research wishes to improve the reproducibility of the work that we publish. This form provides structure for consistency and transparency in reporting. For further information on Nature Research policies, see Authors & Referees and the Editorial Policy Checklist.

Statistical parameters

When statistical analyses are reported, confirm that the following items are present in the relevant location (e.g. figure legend, table legend, main text, or Methods section).

n/a	Confirmed
☒	☐ The exact sample size (*n*) for each experimental group/condition, given as a discrete number and unit of measurement
☐	☒ An indication of whether measurements were taken from distinct samples or whether the same sample was measured repeatedly
☐	☒ The statistical test(s) used AND whether they are one- or two-sided *Only common tests should be described solely by name; describe more complex techniques in the Methods section.*
☒	☐ A description of all covariates tested
☐	☒ A description of any assumptions or corrections, such as tests of normality and adjustment for multiple comparisons
☐	☒ A full description of the statistics including central tendency (e.g. means) or other basic estimates (e.g. regression coefficient) AND variation (e.g. standard deviation) or associated estimates of uncertainty (e.g. confidence intervals)
☐	☒ For null hypothesis testing, the test statistic (e.g. *F*, *t*, *r*) with confidence intervals, effect sizes, degrees of freedom and *P* value noted *Give P values as exact values whenever suitable.*
☒	☐ For Bayesian analysis, information on the choice of priors and Markov chain Monte Carlo settings
☒	☐ For hierarchical and complex designs, identification of the appropriate level for tests and full reporting of outcomes
☐	☒ Estimates of effect sizes (e.g. Cohen's *d*, Pearson's *r*), indicating how they were calculated
☐	☒ Clearly defined error bars *State explicitly what error bars represent (e.g. SD, SE, CI)*

Our web collection on statistics for biologists may be useful.

Software and code

Policy information about availability of computer code

Data collection	Cells were sequenced using HiSeq2500 and NovaSeq 6000 and corresponding Illumina commercial software.
Data analysis	Software/Tool used-Matlab R2017a/b, R v3.5.1, Python v2.7.13, Cytoscape v3.5.1, FastQC v0.11.2, Tophat2 v2.1.1, Cufflinks v2.1.1, HITSAT v2-2.1.0, HTseq v0.6.0, StringTie v1.3.3b, MEGA v7 (https://www.megasoftware.net/), Cluster 3.0 (http://bonsai.hgc.jp/~mdehoon/software/cluster/software.htm), TreeView (http://jtreeview.sourceforge.net), Vaa3D (http://home.penglab.com/proj/vaa3d/Vaa3D/About_Vaa3D.html). R package used: GENIE3 v1.2.1, RcisTarget v1.0.2, AUCell v1.2.4, SCENIC v1.0.1, ComBat (from sva v3.30.0), FactoMineR v1.41, Seurat v3.0, RankProd v2.20.0, pheatmap v1.0.10, WGCNA v1.66, monocle v2.10.1, Python package used: django v1.11.10, Matlab toolbox used: Dimensionality Reduction for t-SNE (https://lvdmaaten.github.io/drtoolbox//). Codes based on literatures: JSD algorithm (based on Cabili et al. 2011), Jackstraw (based on Chung et al. 2015). Web service used: CIBERSORT (https://cibersort.stanford.edu), g:profiler (https://biit.cs.ut.ee/gprofiler/gost), modPhEA (http://evol.nhri.org.tw/phenome2/), Google Charts (https://developers.google.com/chart/interactive/docs/gallery/sankey).

For manuscripts utilizing custom algorithms or software that are central to the research but not yet described in published literature, software must be made available to editors/reviewers upon request. We strongly encourage code deposition in a community repository (e.g. GitHub). See the Nature Research guidelines for submitting code & software for further information.

Data

Policy information about availability of data

All manuscripts must include a data availability statement. This statement should provide the following information, where applicable:

- Accession codes, unique identifiers, or web links for publicly available datasets
- A list of figures that have associated raw data
- A description of any restrictions on data availability

Raw sequencing data is available on NCBI Gene Expression Omnibus under accession number GEO, GSE110284, and NODE project (http://www.biosino.org/node, accession OEP000320). Our resource can be explored at http://egastrulation.sibcb.ac.cn. All other data are available from the corresponding authors on reasonable request.

Field-specific reporting

Please select the best fit for your research. If you are not sure, read the appropriate sections before making your selection.

☒ Life sciences ☐ Behavioural & social sciences ☐ Ecological, evolutionary & environmental sciences

For a reference copy of the document with all sections, see nature.com/authors/policies/ReportingSummary-flat.pdf

Life sciences study design

All studies must disclose on these points even when the disclosure is negative.

Sample size	Atlas sample sizes were chosen to maximise the number of recovered cells from each experiment and to obtain whole embryo part with all germ layers in mouse embryos at their respective stages. The sampling strategy was modified in accordance to the size of the section. For example, In the epiblast/ectoderm, sampling was collected for two sectors A (anterior) and P (posterior) sectors of E6.0-E6.5 embryos and lateral sectors (left lateral and right lateral) were captured in E7.0 and E7.5 embryo. At least three embryo replicates at each developmental stages were studied and the correlation between these embryonic was assessed. High correlation was found. Therefore, one of the embryos was chosen as the representative embryo for data presentation.
Data exclusions	No data were excluded. All replicates were included and only the representative data set were shown.
Replication	All replications were consistent. The Geo-seq on developmental stage embryos were repeated at least three times and the dataset were compared and verified for data consistency. For example, the spearman correlation coefficient of spatial domains were more than 0.9.
Randomization	The Geo-seq was conducted by three different people independently.
Blinding	The samples were collected at different time by different people without knowing the outcome of sequencing data.

Behavioural & social sciences study design

All studies must disclose on these points even when the disclosure is negative.

Study description	*Briefly describe the study type including whether data are quantitative, qualitative, or mixed-methods (e.g. qualitative cross-sectional, quantitative experimental, mixed-methods case study).*
Research sample	*State the research sample (e.g. Harvard university undergraduates, villagers in rural India) and provide relevant demographic information (e.g. age, sex) and indicate whether the sample is representative. Provide a rationale for the study sample chosen. For studies involving existing datasets, please describe the dataset and source.*
Sampling strategy	*Describe the sampling procedure (e.g. random, snowball, stratified, convenience). Describe the statistical methods that were used to predetermine sample size OR if no sample-size calculation was performed, describe how sample sizes were chosen and provide a rationale for why these sample sizes are sufficient. For qualitative data, please indicate whether data saturation was considered, and what criteria were used to decide that no further sampling was needed.*
Data collection	*Provide details about the data collection procedure, including the instruments or devices used to record the data (e.g. pen and paper, computer, eye tracker, video or audio equipment) whether anyone was present besides the participant(s) and the researcher, and whether the researcher was blind to experimental condition and/or the study hypothesis during data collection.*
Timing	*Indicate the start and stop dates of data collection. If there is a gap between collection periods, state the dates for each sample cohort.*
Data exclusions	*If no data were excluded from the analyses, state so OR if data were excluded, provide the exact number of exclusions and the rationale behind them, indicating whether exclusion criteria were pre-established.*

Non-participation	*State how many participants dropped out/declined participation and the reason(s) given OR provide response rate OR state that no participants dropped out/declined participation.*
Randomization	*If participants were not allocated into experimental groups, state so OR describe how participants were allocated to groups, and if allocation was not random, describe how covariates were controlled.*

Ecological, evolutionary & environmental sciences study design

All studies must disclose on these points even when the disclosure is negative.

Study description	*Briefly describe the study. For quantitative data include treatment factors and interactions, design structure (e.g. factorial, nested, hierarchical), nature and number of experimental units and replicates.*
Research sample	*Describe the research sample (e.g. a group of tagged Passer domesticus, all Stenocereus thurberi within Organ Pipe Cactus National Monument), and provide a rationale for the sample choice. When relevant, describe the organism taxa, source, sex, age range and any manipulations. State what population the sample is meant to represent when applicable. For studies involving existing datasets, describe the data and its source.*
Sampling strategy	*Note the sampling procedure. Describe the statistical methods that were used to predetermine sample size OR if no sample-size calculation was performed, describe how sample sizes were chosen and provide a rationale for why these sample sizes are sufficient.*
Data collection	*Describe the data collection procedure, including who recorded the data and how.*
Timing and spatial scale	*Indicate the start and stop dates of data collection, noting the frequency and periodicity of sampling and providing a rationale for these choices. If there is a gap between collection periods, state the dates for each sample cohort. Specify the spatial scale from which the data are taken*
Data exclusions	*If no data were excluded from the analyses, state so OR if data were excluded, describe the exclusions and the rationale behind them, indicating whether exclusion criteria were pre-established.*
Reproducibility	*Describe the measures taken to verify the reproducibility of experimental findings. For each experiment, note whether any attempts to repeat the experiment failed OR state that all attempts to repeat the experiment were successful.*
Randomization	*Describe how samples/organisms/participants were allocated into groups. If allocation was not random, describe how covariates were controlled. If this is not relevant to your study, explain why.*
Blinding	*Describe the extent of blinding used during data acquisition and analysis. If blinding was not possible, describe why OR explain why blinding was not relevant to your study.*

Did the study involve field work? ☐ Yes ☐ No

Field work, collection and transport

Field conditions	*Describe the study conditions for field work, providing relevant parameters (e.g. temperature, rainfall).*
Location	*State the location of the sampling or experiment, providing relevant parameters (e.g. latitude and longitude, elevation, water depth).*
Access and import/export	*Describe the efforts you have made to access habitats and to collect and import/export your samples in a responsible manner and in compliance with local, national and international laws, noting any permits that were obtained (give the name of the issuing authority, the date of issue, and any identifying information).*
Disturbance	*Describe any disturbance caused by the study and how it was minimized.*

Reporting for specific materials, systems and methods

Materials & experimental systems

n/a	Involved in the study
☒	☐ Unique biological materials
☐	☒ Antibodies
☐	☒ Eukaryotic cell lines
☒	☐ Palaeontology
☐	☒ Animals and other organisms
☒	☐ Human research participants

Methods

n/a	Involved in the study
☒	☐ ChIP-seq
☒	☐ Flow cytometry
☒	☐ MRI-based neuroimaging

Unique biological materials

Policy information about availability of materials

Obtaining unique materials	No unique materials.

Antibodies

Antibodies used	Primary antibody: Mouse Cardiac Troponin T antibody [1C11], abcam, monoclonal, catalog#ab829, Lot number #GR3237888-6, dilution #1:1000. Secondary antibody: Goat anti-Mouse Alexa Fluor-594, Invitrogen, polyclonal, Catalog#A-11032, Lot number#1922849, dilution #1:400.
Validation	Primary antibody tested applications: Flow Cyt, IHC-Fr, WB, ELISA, IHC-FoFr, IP, IHC-P, Sandwich ELISA, ICC/IF. Species reactivity:Reacts with Mouse, Rat, Dog, Human; Antibody notes: This antibody detects Troponin T in human cardiac muscle. No cross-reaction with skeletal troponin T, cTnI and TnC. Revelent refereenc: Ren L et al. A Disintegrin and Metalloprotease-22 Attenuates Hypertrophic Remodeling in Mice Through Inhibition of the Protein Kinase B Signaling Pathway. J Am Heart Assoc 7:N/A (2018). Read more (PubMed: 29358191)

Eukaryotic cell lines

Policy information about cell lines

Cell line source(s)	E14 cell line was a gift from Jinsong Li's Lab, which was obtained from CellBank of SIBCB.
Authentication	E14 is maintained in the CellBank of SIBCB and STR analysis was routinely conducted.
Mycoplasma contamination	Cell lines were tested for mycoplasma contamination monthly. Result shows negative for mycoplasma contamination.
Commonly misidentified lines (See ICLAC register)	No commonly misidentified cell lines were used.

Palaeontology

Specimen provenance	N.A.
Specimen deposition	N.A.
Dating methods	N.A.

☐ Tick this box to confirm that the raw and calibrated dates are available in the paper or in Supplementary Information.

Animals and other organisms

Policy information about studies involving animals; ARRIVE guidelines recommended for reporting animal research

Laboratory animals	For Geo-seq analysis, mus musculus C57BL/6 males and females at E5.5, E6.0, E6.5, E7.0 and E7.5. For mouse embryos for WISH or in vitro culture, ICR mouse embryo at E6.5, E7.0 and E7.5 was used and no sex preference.
Wild animals	The study did not involve wild animals
Field-collected samples	The study did not involve field-collected samples

Human research participants

Policy information about studies involving human research participants

Population characteristics	No human research participants were used.
Recruitment	N.A.

ChIP-seq

Data deposition

☐ Confirm that both raw and final processed data have been deposited in a public database such as GEO.

☐ Confirm that you have deposited or provided access to graph files (e.g. BED files) for the called peaks.

Data access links *May remain private before publication.*	*For "Initial submission" or "Revised version" documents, provide reviewer access links. For your "Final submission" document, provide a link to the deposited data.*
Files in database submission	*Provide a list of all files available in the database submission.*
Genome browser session (e.g. UCSC)	*Provide a link to an anonymized genome browser session for "Initial submission" and "Revised version" documents only, to enable peer review. Write "no longer applicable" for "Final submission" documents.*

Methodology

Replicates	*Describe the experimental replicates, specifying number, type and replicate agreement.*
Sequencing depth	*Describe the sequencing depth for each experiment, providing the total number of reads, uniquely mapped reads, length of reads and whether they were paired- or single-end.*
Antibodies	*Describe the antibodies used for the ChIP-seq experiments; as applicable, provide supplier name, catalog number, clone name, and lot number.*
Peak calling parameters	*Specify the command line program and parameters used for read mapping and peak calling, including the ChIP, control and index files used.*
Data quality	*Describe the methods used to ensure data quality in full detail, including how many peaks are at FDR 5% and above 5-fold enrichment.*
Software	*Describe the software used to collect and analyze the ChIP-seq data. For custom code that has been deposited into a community repository, provide accession details.*

Flow Cytometry

Plots

Confirm that:

☐ The axis labels state the marker and fluorochrome used (e.g. CD4-FITC).

☐ The axis scales are clearly visible. Include numbers along axes only for bottom left plot of group (a 'group' is an analysis of identical markers).

☐ All plots are contour plots with outliers or pseudocolor plots.

☐ A numerical value for number of cells or percentage (with statistics) is provided.

Methodology

Sample preparation	*Describe the sample preparation, detailing the biological source of the cells and any tissue processing steps used.*
Instrument	*Identify the instrument used for data collection, specifying make and model number.*
Software	*Describe the software used to collect and analyze the flow cytometry data. For custom code that has been deposited into a community repository, provide accession details.*
Cell population abundance	*Describe the abundance of the relevant cell populations within post-sort fractions, providing details on the purity of the samples and how it was determined.*
Gating strategy	*Describe the gating strategy used for all relevant experiments, specifying the preliminary FSC/SSC gates of the starting cell population, indicating where boundaries between "positive" and "negative" staining cell populations are defined.*

☐ Tick this box to confirm that a figure exemplifying the gating strategy is provided in the Supplementary Information.

Magnetic resonance imaging

Experimental design

Design type	*Indicate task or resting state; event-related or block design.*
Design specifications	*Specify the number of blocks, trials or experimental units per session and/or subject, and specify the length of each trial or block (if trials are blocked) and interval between trials.*
Behavioral performance measures	*State number and/or type of variables recorded (e.g. correct button press, response time) and what statistics were used to establish that the subjects were performing the task as expected (e.g. mean, range, and/or standard deviation across subjects).*

Acquisition

Imaging type(s)	*Specify: functional, structural, diffusion, perfusion.*
Field strength	*Specify in Tesla*
Sequence & imaging parameters	*Specify the pulse sequence type (gradient echo, spin echo, etc.), imaging type (EPI, spiral, etc.), field of view, matrix size, slice thickness, orientation and TE/TR/flip angle.*
Area of acquisition	*State whether a whole brain scan was used OR define the area of acquisition, describing how the region was determined.*

Diffusion MRI ☐ Used ☐ Not used

Preprocessing

Preprocessing software	*Provide detail on software version and revision number and on specific parameters (model/functions, brain extraction, segmentation, smoothing kernel size, etc.).*
Normalization	*If data were normalized/standardized, describe the approach(es): specify linear or non-linear and define image types used for transformation OR indicate that data were not normalized and explain rationale for lack of normalization.*
Normalization template	*Describe the template used for normalization/transformation, specifying subject space or group standardized space (e.g. original Talairach, MNI305, ICBM152) OR indicate that the data were not normalized.*
Noise and artifact removal	*Describe your procedure(s) for artifact and structured noise removal, specifying motion parameters, tissue signals and physiological signals (heart rate, respiration).*
Volume censoring	*Define your software and/or method and criteria for volume censoring, and state the extent of such censoring.*

Statistical modeling & inference

Model type and settings	*Specify type (mass univariate, multivariate, RSA, predictive, etc.) and describe essential details of the model at the first and second levels (e.g. fixed, random or mixed effects; drift or auto-correlation).*
Effect(s) tested	*Define precise effect in terms of the task or stimulus conditions instead of psychological concepts and indicate whether ANOVA or factorial designs were used.*

Specify type of analysis: ☐ Whole brain ☐ ROI-based ☐ Both

Statistic type for inference (See Eklund et al. 2016)	*Specify voxel-wise or cluster-wise and report all relevant parameters for cluster-wise methods.*
Correction	*Describe the type of correction and how it is obtained for multiple comparisons (e.g. FWE, FDR, permutation or Monte Carlo).*

Models & analysis

n/a	Involved in the study
☐	☐ Functional and/or effective connectivity
☐	☐ Graph analysis
☐	☐ Multivariate modeling or predictive analysis

Functional and/or effective connectivity	*Report the measures of dependence used and the model details (e.g. Pearson correlation, partial correlation, mutual information).*
Graph analysis	*Report the dependent variable and connectivity measure, specifying weighted graph or binarized graph, subject- or group-level, and the global and/or node summaries used (e.g. clustering coefficient, efficiency, etc.).*

Multivariate modeling and predictive analysis

Specify independent variables, features extraction and dimension reduction, model, training and evaluation metrics.