

Primate gastrulation and early organogenesis at single-cell resolution


<https://doi.org/10.1038/s41586-022-05526-y>

Received: 9 January 2022

Accepted: 4 November 2022

Published online: 14 December 2022

Open access

 Check for updates

Jinglei Zhai^{1,2,3,10}, Jing Guo^{1,2,3,4,10}, Haifeng Wan^{1,2,3,10}, Luqing Qi^{1,2,3,4,10}, Lizhong Liu^{5,10}, Zhenyu Xiao^{1,2,3}, Long Yan^{1,2,3}, Daniel A. Schmitz⁵, Yanhong Xu^{1,2,3,4}, Dainan Yu^{1,2,3,4}, Xulun Wu^{1,2,3,4}, Wentao Zhao^{1,2,3,4}, Kunyuan Yu^{1,2,3,4}, Xiangxiang Jiang^{1,6,7}, Fan Guo^{1,2,3,4}, Jun Wu^{5,8,9} & Hongmei Wang^{1,2,3,4}

Our understanding of human early development is severely hampered by limited access to embryonic tissues. Due to their close evolutionary relationship with humans, nonhuman primates are often used as surrogates to understand human development but currently suffer from a lack of in vivo datasets, especially from gastrulation to early organogenesis during which the major embryonic cell types are dynamically specified. To fill this gap, we collected six Carnegie stage 8–11 cynomolgus monkey (*Macaca fascicularis*) embryos and performed in-depth transcriptomic analyses of 56,636 single cells. Our analyses show transcriptomic features of major perigastrulation cell types, which help shed light on morphogenetic events including primitive streak development, somitogenesis, gut tube formation, neural tube patterning and neural crest differentiation in primates. In addition, comparative analyses with mouse embryos and human embryoids uncovered conserved and divergent features of perigastrulation development across species—for example, species-specific dependency on Hippo signalling during presomitic mesoderm differentiation—and provide an initial assessment of relevant stem cell models of human early organogenesis. This comprehensive single-cell transcriptome atlas not only fills the knowledge gap in the nonhuman primate research field but also serves as an invaluable resource for understanding human embryogenesis and developmental disorders.

In humans, the developmental periods of gastrulation and early organogenesis largely remain a ‘black box’ due to limited access to research embryos. Recently, single-cell RNA sequencing (scRNA-seq) data from six aborted human embryos (one Carnegie stage¹ (CS) 7 (ref.²) and five CS12–16 (ref.³)) and 15 CS3–7 cynomolgus monkey embryos⁴ became available, providing valuable resources for the study of primate early post-implantation development. Despite these advances, single-cell transcriptomes of CS8–11 human and nonhuman primate (NHP) embryos are still not available, which severely hinders the study of primate perigastrulation development and aetiology underlying several most common forms of congenital malformations. To fill this knowledge gap, we generated a comprehensive single-cell atlas of CS8–11 cynomolgus monkey (herein referred to as monkey) embryos and studied the major molecular and cellular processes during this critical developmental period in primates.

A transcriptome atlas of monkey embryos

We collected six monkey embryos during embryonic day (E) 20–29, which were staged at CS8, CS9 and CS11 (Fig. 1a and Supplementary Table 1). All embryos appeared morphologically normal with expected anatomical

features—for example, primitive streak (PS) and enlarged yolk sac in CS8 and CS9 embryos and forebrain, cardiac structure and somites in CS11 embryos. Samples were dissociated into single cells, with 67,418 sequenced using the 10X Genomics Chromium platform (Supplementary Table 1). After filtering out doublets/multiplets and low-quality cells (fewer than 500 genes detected), a total of 56,636 cells was retained for subsequent analyses with a median of 3,017 genes detected per cell (Extended Data Fig. 1a,b and Supplementary Tables 1 and 2). Based on the expression of known lineage markers and comparison with datasets from mouse embryos at corresponding developmental stages^{5–7}, 38 major clusters were identified (Fig. 1b,c, Extended Data Fig. 1c,d and Supplementary Table 3). The epiblast (EP) and PS cells (cluster nos. 1 and 2) identified from CS11 embryos were greatly under-represented, suggesting that gastrulation was nearing completion at this developmental stage (Fig. 1b, Extended Data Fig. 1c and Supplementary Table 2).

Development landscape of primitive streak

To study the molecular and cellular dynamics during monkey gastrulation and early organogenesis, we used RNA velocity, which predicts

¹State Key Laboratory of Stem Cell and Reproductive Biology, Institute of Zoology, Chinese Academy of Sciences, Beijing, China. ²Institute for Stem Cell and Regeneration, Chinese Academy of Sciences, Beijing, China. ³Beijing Institute for Stem Cell and Regenerative Medicine, Beijing, China. ⁴University of Chinese Academy of Sciences, Beijing, China. ⁵Department of Molecular Biology, University of Texas Southwestern Medical Center, Dallas, TX, USA. ⁶NHC Key Laboratory of Study on Abnormal Gametes and Reproductive Tract, Anhui Medical University, Hefei, China. ⁷Department of Obstetrics and Gynecology, the First Affiliated Hospital of Anhui Medical University, Hefei, China. ⁸Hamon Center for Regenerative Science and Medicine, University of Texas Southwestern Medical Center, Dallas, TX, USA. ⁹Cecil H. and Ida Green Center for Reproductive Biology Sciences, University of Texas Southwestern Medical Center, Dallas, TX, USA. ¹⁰These authors contributed equally: Jinglei Zhai, Jing Guo, Haifeng Wan, Luqing Qi, Lizhong Liu. ✉e-mail: jiangxx@ahmu.edu.cn; guofan@ioz.ac.cn; jun2.wu@utsouthwestern.edu; wanghm@ioz.ac.cn

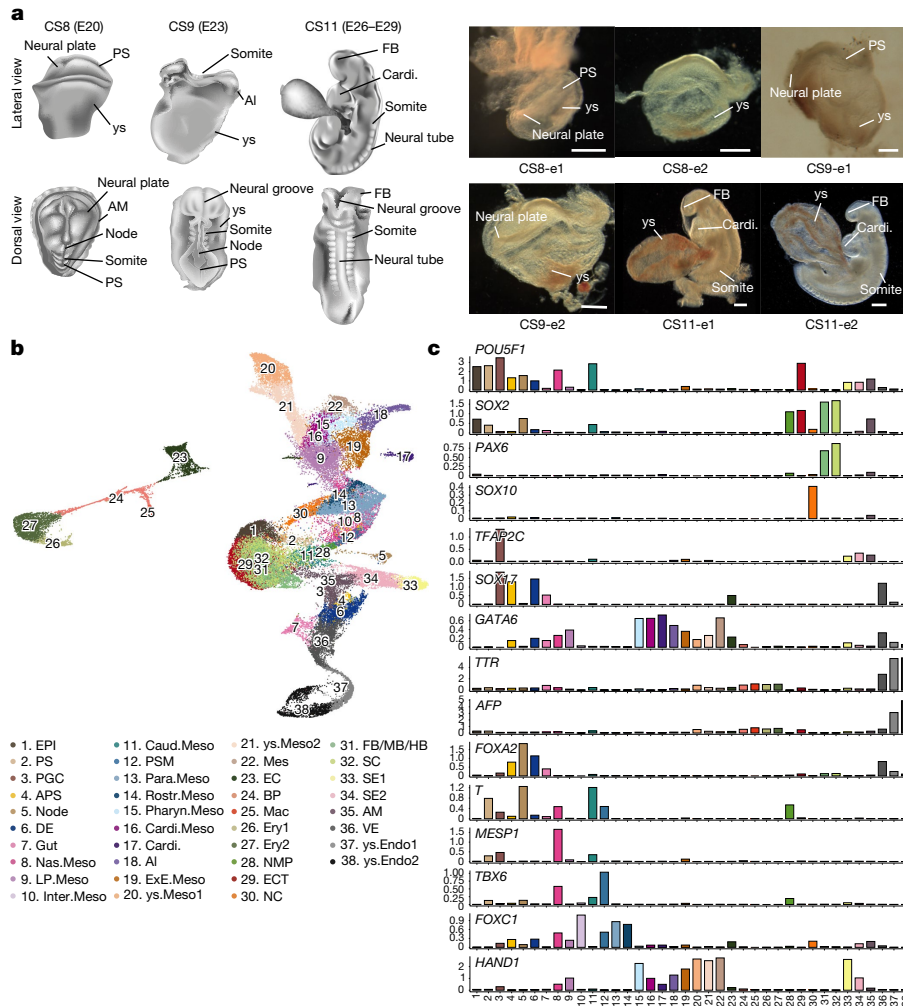


Fig. 1 | Construction of single-cell transcriptome atlas of cynomolgus monkey embryos at CS8–11. **a**, Left, flowchart overview of the dynamic morphogenesis of CS8–11 monkey embryos. ys, yolk sac; Al, allantois; AM, amnion; FB, forebrain; Cardi., cardiac. Right, representative bright-field images of embryos obtained from indicated developmental stages; e, embryo tissue; $n = 2$ for each stage sample. Scale bars, 200 μm . **b**, UMAP plot showing the 38 major cell types. EPI, epiblast; Mes, mesenchyme; EC, endothelial cell; BP, blood progenitor; Mac, macrophage; Ery, erythrocyte; ys.Endo, yolk sac endoderm. **c**, Bar charts showing the over-represented genes in each cell type. *POU5F1* expression is higher in pluripotent cells such as EPI and PGCs;

SOX2 shows higher expression in pluripotent cells, ECT and FB/MB/HB; *PAX6* is essential for neural development, and *SOX10* marks NC. *TFAP2C* is over-represented in PGCs, AM and SE, whereas *SOX17* expression is high in endodermal lineage and PGCs. *GATA6* is overexpressed in mesoderm and endodermal lineages. *TTR* and *AFP* expression levels are high in ys.Endo. *FOXA2* is highly expressed in node and endodermal lineage. *T* is over-represented in PS and mesodermal lineage. *MESP1* marks the Nas.Meso, and *TBX6* is over-represented in PS, Nas.Meso, NMP and PSM. *FOXC1* and *HAND1* show high expression in mesodermal lineage.

differentiation trajectories by leveraging splicing kinetics⁸. We first focused on PS formation-related clusters, which include PS (no. 2), anterior primitive streak (APS, no. 4), definitive endoderm (DE, no. 6), node (no. 5) and nascent mesoderm (Nas.Meso, no. 8) clusters (Fig. 1b). Similar to that in mice⁷, RNA velocity predicted a trifurcating differentiation trajectory of monkey PS/APS towards DE, Nas.Meso and node as gastrulation advances (Fig. 2a, b and Extended Data Fig. 2a). Single-cell regulatory network inference and clustering (SCENIC) and immunofluorescence (IF) analyses showed that several transcription factors (TFs) were enriched in clusters PS (for example, *GATA6* and *PBX2*), APS (for example, *FOXA1* and *HOXD3*), Nas.Meso (for example, *TBX6* and *MEIS1*), DE (for example, *CDX1* and *OTX2*) and node (for example, *TBX* and *HOX*) (Fig. 2c, d), suggesting their roles in different steps of monkey PS formation. Consistent with a study in mice⁹, differentially expressed gene (DEG) and IF analyses provided support that *FOXA2*⁺ cells putatively contributed to DE in monkeys (Extended Data Fig. 1d and Fig. 2d).

Nas.Meso gives rise to most mesoderm cell types. Based on DEG analysis, Nas.Meso derivatives, including intermediate (Inter.Meso),

paraxial (Para.Meso), rostral (Rostr.Meso), pharyngeal (Pharyn.Meso), cardiac (Cardi.Meso), lateral plate (LP.Meso), caudal (Caud.Meso) mesoderm cells and extra-embryonic mesenchymal cells (EXMCs, including allantois (Al), yolk sac (ys.Meso) and extra-embryonic mesoderm cells (ExE.Meso)), were identified (Fig. 2e, Extended Data Figs. 1d and 2b, c and Supplementary Table 3). Cardi.Meso, Inter.Meso, Rostr.Meso, neuromesodermal progenitor (NMP) and presomitic mesoderm (PSM) cells did not manifest as distinct clusters until CS11, when the numbers of cells identified as Nas.Meso and Caud.Meso were greatly reduced (Extended Data Fig. 2c). To gain insight into early mesoderm differentiation, we generated RNA velocity maps of PS and mesoderm cells, which predicted differentiation trajectories of Nas.Meso towards LP.Meso, ExE.Meso and Para.Meso, followed by LP.Meso to ys.Meso1 and Pharyn.Meso (Fig. 2e, f and Extended Data Fig. 2b, d), consistent with studies in chicks and mice¹⁰. It was hypothesized that early primate EXMCs, which putatively originated from hypoblast-derived primary yolk sac before gastrulation, subsequently merged with PS-derived EXMCs to establish the allantoic stalk^{11–14} (Extended Data Fig. 2e).

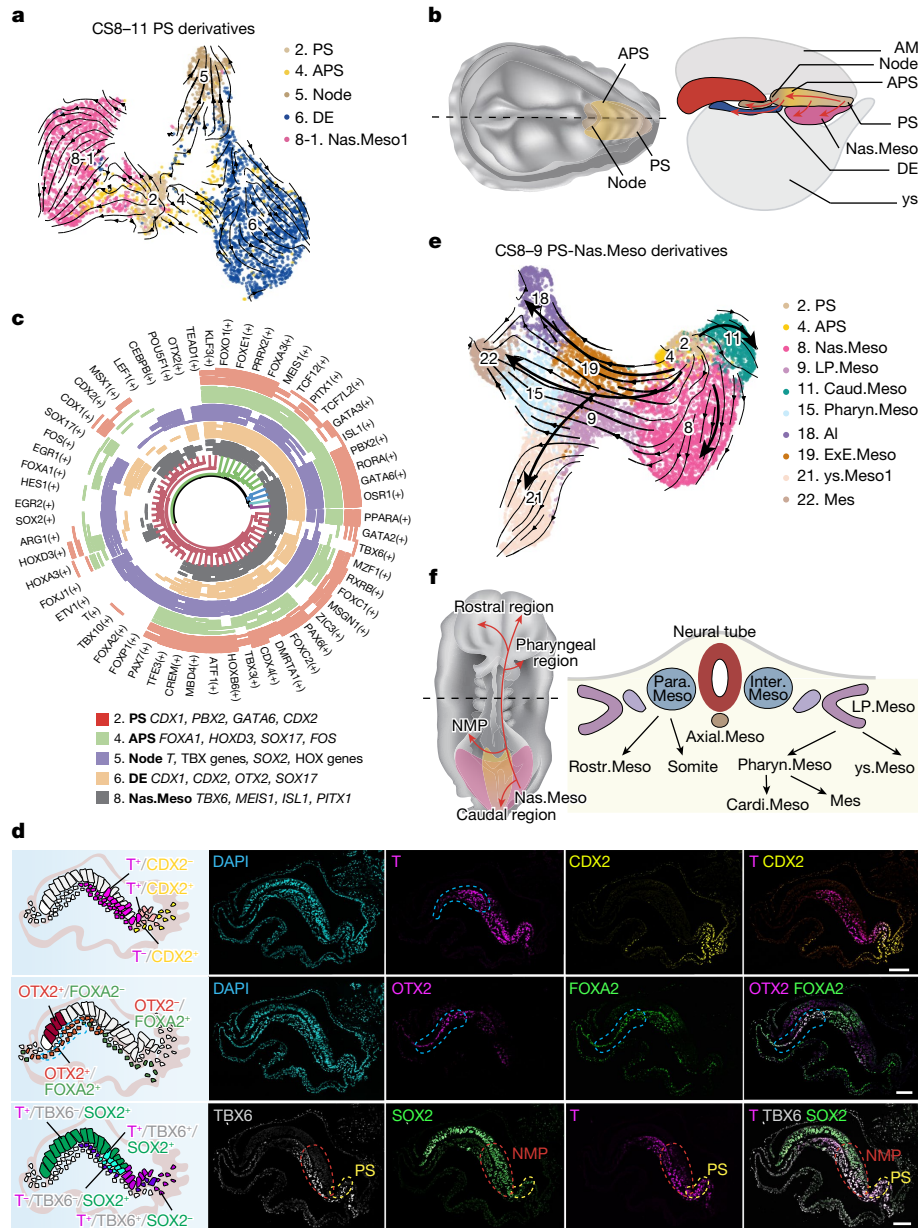


Fig. 2 | Developmental trajectories of PS and its derivatives from monkey embryos. **a**, UMAP showing PS and its derivatives from CS8–11 monkey embryos with trajectories inferred from RNA velocity analysis. Black arrows represent calculated velocity trajectories. **b**, Schematic diagram showing a CS8 embryo. Left, dorsal view; right, lateral view showing potential trajectories from PS to APS, DE, node and Nas.Meso. **c**, Joint hierarchical clustering (HC) showing the circular heatmap with representative TFs upregulated in PS (red), APS (green), node (purple), DE (orange) and Nas.Meso (grey). Genes marked with the same colour by HC show similar expression patterns in different cell types. **d**, Left, diagrams summarizing the distribution of indicated cell types in the region of interest, based on IF images on the right. Right, IF results showing localization of T⁺, CDX2⁺, OTX2⁺, FOXA2⁺, TBX6⁺ and SOX2⁺ cells in a monkey

E22 embryo. T, marker for PS and some mesoderm; CDX2, marker for PS; OTX2, marker for visceral endoderm, definitive endoderm, neural ectoderm and some mesoderm; FOXA2, marker for endoderm and some mesoderm; TBX6, marker for PS, NMP and some mesoderm; SOX2, marker for ECT. DAPI, marker for DNA, here and after. Blue dashed lines indicate potential organizer location; yellow and red dashed lines denote PS and NMP regions, respectively. Scale bars, 100 μm. **e**, RNA velocity overlaid on UMAP coloured by PS and mesodermal fates from CS8–9 embryos. **f**, Left, schematic diagram showing the potential migration and differentiation route of Nas.Meso from the dorsal view of a monkey CS9 embryo. Right, diagram showing mesodermal derivatives from a transverse section.

Our RNA velocity and IF analyses support the potential contribution of PS (Nas.Meso) towards EXMCs (ExE.Meso, ys. Meso and AI) during CS8–11 in primates¹² (Fig. 2e and Extended Data Fig. 2b,d,f).

Somitogenesis, the process initiated from a subtype of NMP, has been extensively studied in mice¹⁵ but not in primates. To better understand primate somitogenesis, we performed IF analysis and identified two putative NMP populations in a CS8 embryo: SOX2^{low}/T^{high}/TBX6^{high} (contributing to somite) and SOX2^{high}/T^{low}/TBX6^{low} (contributing to spinal

cord (SC)) (Fig. 2d). Based on expression patterns of signalling pathway components and regional markers¹⁶, we identified Rostr.Meso (*PITX2*, *IRX3*) and several somitic cell types in the Para.Meso cluster, which contained somitomere (also called segmentation boundary, *RIPPLY1/2*), early somite (*TCF15*, *FOXC2*, *MEOX1*), sclerotome (*PAX1*, *PAX9*, *NKX3.2*, *SOX9*) and dermomyotome (*PAX7*, *ALX4*, *TFAP2A*) (Extended Data Fig. 2g,h). RNA velocity analysis on NMP, PSM and Para.Meso clusters further revealed their putative lineage relationships (Extended Data Fig. 2g).

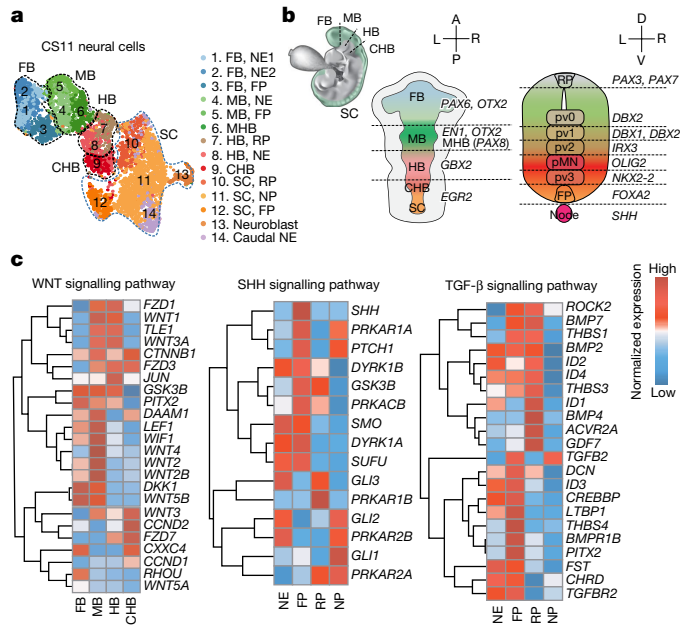


Fig. 3 | Developmental landscape of neural cells from monkey embryos. **a**, UMAP plot showing subtypes of neural cells (including FB/MB/HB and SC cells) from CS11 embryos. **b**, Schematics showing regionalized expression patterns of different brain regions (upper left) along A-P (middle) and D-V axes (right). **c**, Heatmaps showing expression of genes in WNT, SHH and TGF- β signalling pathways in indicated neural cells along the A-P and D-V axes.

To gain insight into gut tube (GT) formation in primates, we focused on DE, visceral endoderm (VE) and gut clusters. Our analyses identified seven subclusters of foregut, midgut and hindgut cells (Extended Data Fig. 2i–k). Foregut contained cells expressing *HHEX* (Foregut1) and *PHLDA2* (Foregut2), Midgut cells were separated into Midgut1 (*MX1*), Midgut2 (*HOXB2, HOXC9*) and Midgut/Hindgut (*HOXA10, CXCL12*), whereas Hindgut cells included subclusters Hindgut1 (*HOXA10*) and Hindgut2 (*CDX2*) (Extended Data Fig. 2j). To help identify the origin(s) of gut cells in primates, we performed RNA velocity analysis which predicted: (1) Foregut1 was solely derived from DE; and (2) Hindgut2 was exclusively contributed by VE whereas Foregut2, Midgut1/2, Midgut/Hindgut and Hindgut1 clusters contained both DE (mostly) and VE cells (Extended Data Fig. 2i,k). These RNA velocity predictions were further validated by transport map, partition-based graph abstraction (PAGA) and pseudotime analyses (Extended Data Fig. 2l), consistent with the DE and VE dual origins of gut cells in mice^{6,17}. VE, through secreted inhibitors of WNT and NODAL pathways, plays an important role in anterior patterning of mouse epiblast (EPI)^{18,19}. We performed CellPhoneDB analysis and identified several conserved ligand–receptor interactions of TGF- β (BMP, NODAL), WNT and FGF pathways between VE and EPI/EPI derivatives (PS, APS, DE, node and Nas.Meso)²⁰. Interestingly, interactions mediated by ligand–receptor pairs of the Notch2 pathway were over-represented between monkey EPI derivatives and VE whereas mouse embryos with perturbed Notch signalling developed normally beyond gastrulation²¹, implying a new role of Notch2 signalling during primate gastrulation. Furthermore, more ligand–receptor interactions were identified between VE and EPI derivatives than between VE and EPI, suggesting dynamic communications between extra-embryonic and embryonic cells during gastrulation (Extended Data Fig. 2m).

Taken together, these analyses identified major cell types during monkey PS development, early mesoderm and endoderm differentiation and shed light on somitogenesis and GT formation in primates.

Developmental landscape of ectoderm

After definitive endoderm and embryonic mesoderm are formed, the remaining epiblast cells become the ectoderm (ECT, cluster no. 29), giving rise to centrally located neural ectoderm (NE, also called neural plate), surface ectoderm (SE, cluster nos. 33 and 34) at the periphery and neural plate border (NPB) between the two, which ventrally delaminates and differentiates into neural crest (NC, cluster no. 30) (Fig. 1b and Extended Data Fig. 3a). The neural plate then thickens, bends and folds to form the neural tube, the precursor of the central nervous system (CNS)²² (Extended Data Fig. 3a).

Specification of NE and SE along the mediolateral axis in zebrafish, chick, mouse embryos and human embryonic stem cell (hESC) derivatives depends on BMP and WNT gradients generated by the Spemann–Mangold organizer—a group of cells that plays a key role in the establishment of dorsal–ventral (D–V) and anterior–posterior (A–P) axes during gastrulation^{22–27}. We found that organizer-related genes²⁷, including *GSC, OTX2, FOXA2, FST, CER1, DKK1, HHEX* and *CHRD*, were highly expressed in some cells from PS, APS, DE and node clusters (Extended Data Fig. 3b,c). IF analyses of *OTX2, FOXA2* and T helped localize putative organizer cells to the anterior region beneath the ectoderm in a CS8 monkey embryo (Fig. 2d), which is consistent with mouse organizer cells in the E7.5 mouse embryo (Extended Data Fig. 3d). In addition, many genes related to TGF- β and WNT pathways were found upregulated in SE but not in NE, suggesting selective activation of these pathways during SE differentiation (Extended Data Fig. 3e). SCENIC analysis further showed elevated expression of TFs including *SOX2, POU3F2, EN2, OTX2* and *NEURUG1* in ECT, and *TFAP2A, TFAP2C, DLX5* and *HOX* family genes in SE (Extended Data Fig. 3f), which may help specify and/or stabilize their lineage identities.

Neural crest is a transient, multipotent and migratory cell population^{22,28}. Notably, the number of cells expressing NC specification genes (for example, *SOX10, SOX9, PAX3, FOXD3* and *SNAI2*) greatly increased in CS11 embryos (Extended Data Fig. 3g,h). We identified eight subpopulations of NC cells from CS11 embryos, which include pre-EMT (*PAX3, ZIC2*), delaminating (*MAFB, MEF2C*), early migratory (*SNAI2, FOXD2*), migratory1/2 (*TWIST1, MCAM*), mesenchymal (*PRRX1*), sensory (*SIX1, EYA2*) and autonomic (*SIOOB, MPZ*) cells (Extended Data Fig. 3g,h). The expression patterns of *HOX* genes helped distinguish cranial and vagal/trunk NC subtypes (Extended Data Fig. 3g,i). RNA velocity analysis further predicted that (1) pre-EMT and delaminating NCs gave rise to migrating progenitors that ultimately contributed to both cranial and vagal/trunk NC subtypes; and (2) cranial, but not vagal/trunk, NC contributed to mesenchymal cells (Extended Data Fig. 3g). These predictions are consistent with mouse studies^{28,29}.

To gain insight into neural tube development in primates, we reanalysed the forebrain/midbrain/hindbrain (FB/MB/HB, cluster no. 31) and SC (cluster no. 32) cells from two CS11 embryos (Fig. 1b). Based on the expression patterns of *OTX2, EN1, EGR2* and *HOXA2*, among others, FB, MB, HB and SC cells along the A–P axis could be annotated (Fig. 3a,b, and Extended Data Fig. 4a,b). In addition, midbrain–hindbrain boundary (MHB), which is derived from MB and characterized by morphological constriction of the neural tube, could also be identified based on the upregulation of *PAX8, FGF8* and *PAX5* (Fig. 3a,b and Extended Data Fig. 4a,b). Next we focused on the WNT pathway and *HOX* family genes, which are known to regulate neural tube patterning along the A–P axis in other vertebrates^{26,30}. Compared with FB, many WNT pathway-related genes were upregulated in MB cells, implying increased WNT activity (Fig. 3c). Many *HOX* genes were found enriched in neural cells from the trunk region, suggesting their roles in A–P patterning of neural tube (Extended Data Fig. 4c). Besides, we identified several specific TFs of FB (for example, *HAND1, HESX1, FOXG1, NFATC4*) and caudal hindbrain (CHB; for example, *HOXA3, MEIS1, WRNIP1, MAFB*) (Extended Data Fig. 4c).

To show the transcriptomic features of monkey neural tube along the D–V axis, we studied roof plate (RP), NE and floor plate (FP) cells

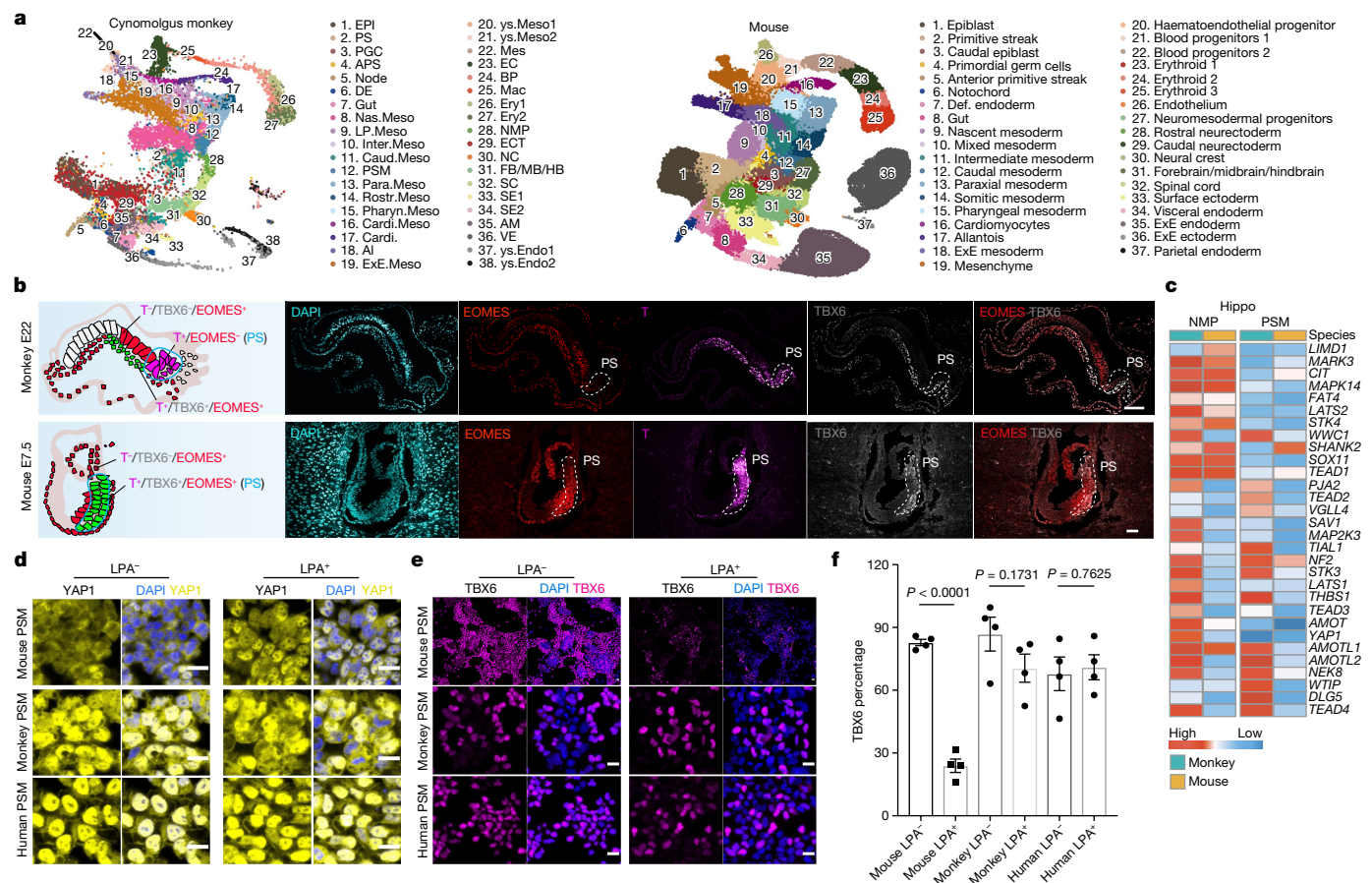


Fig. 4 | Cross-species transcriptomic comparison of early embryonic organogenesis. **a**, UMAPs showing the integrated datasets of CS8–11 monkey embryos (left) and E6.5–8.5 mouse embryos (right). **b**, Left, diagrams summarizing localization of EOMES⁺, T⁺ and TBX6⁺ cells in monkey E22 (top) and mouse E7.5 embryos (bottom). Blue lines circle the PS region. Right, IF results showing differences in EOMES expression of PS in monkey E22 embryo (top, cells coloured magenta, T⁺/EOMES⁺/TBX6^{+/low}) and mouse E7.5 embryo (bottom, cells coloured green, T⁺/EOMES⁺/TBX6⁺). White dashed lines delineate the PS region. Scale bars, 100 μm. **c**, Heatmap showing the different expression patterns of notable molecules on the Hippo signalling pathway between monkeys and mice in terms of NMP and PSM cells. **d**, Different Hippo activities (YAP1) between mouse and primate (monkey and human) PSM-like cells, indicating species-specific requirement for Hippo activation during somitogenesis.

identified by expression patterns of genes including *PAX7*, *PAX6* and *FOXA2* (ref. ³¹) (Fig. 3a,b and Extended Data Fig. 4a,d). Neural progenitor (NP) of ventral interneurons (pv, *DBX1*, *DBX2*, *NKX2-2*) and motor neurons (pMN, *OLIG2*) that regionalized along the D–V axis, and neuroblasts (*TUBB3*, *NEURODG2*), could be annotated in the spinal cord, suggesting that the neural tube was closed here in CS11 embryos (Extended Data Fig. 4d). In mice, D–V patterning of the neural tube is induced by opposing morphogens SHH from notochord (ventral) and TGF-β superfamily from epidermis (dorsal)^{32,33}. Next, we studied gene expression patterns in SHH and TGF-β signalling pathways³³ (Fig. 3c). We found that FP highly expressed *GSK3B*, *PTCH1* and *SHH* whereas RP upregulated the expression of *BMP2*, *BMP4* and *BMP7* (Fig. 3c). From ventral to dorsal, *GLI1* and *GLI2* were upregulated in NP whereas *GLI3* expression was specifically elevated in RP. Notably, in the unclosed part of neural tube, NE simultaneously expressed *GLI2* and *GLI3* but not *GLI1*, which is required for the regulation of pv3 specification^{33,34} (Fig. 3c). CellPhoneDB analysis further indicated a prodigious number of ligand–receptor interactions among different neural cells and their neighbours (Extended Data Fig. 4e).

Inhibition of Hippo activity leads to nuclear accumulation of YAP1 in mouse PSM-like cells. LPA is a Hippo inhibitor. Scale bars, 20 μm. Experiment was repeated independently three times. **e**, Inhibition of Hippo activity impairs differentiation of mouse PSM-like cells but shows only moderate effect on monkey PSM-like cell differentiation. Scale bars, 20 μm. Experiments were repeated independently three times. **f**, Quantification of percentage of TBX6⁺ from total cells; total cell number was calculated by counting the number of nuclei (DAPI). Data shown as mean ± s.d. (*n* = 4 biological replicates; experiments were independently repeated three times and quantified twice, with similar results). *P* values were determined by unpaired two-tailed *t*-test. Mouse LPA⁻ group versus LPA⁺ group, *P* < 0.0001; monkey LPA⁻ group versus LPA⁺ group, *P* = 0.173102; human LPA⁻ group versus LPA⁺ group, *P* = 0.762472.

Taken together, these findings show that NE and SE in monkeys were specified along the mediolateral axis at CS8, followed by NC differentiation at CS11. In addition, as the extension of the body plan took place, spatial organization of CNS divisions was orchestrated along the A–P and D–V axes.

Cross-species comparison

Although mice and monkeys are animal models widely used for understanding human development, single-cell transcriptome comparisons of the earliest steps in organogenesis among mice, monkeys and humans are lacking. To this end, we first annotated the paralogues and one-to-one orthologues from mice⁶, monkeys and humans^{2,3} (Extended Data Fig. 5a,b and Supplementary Tables 4 and 5). By and large, cross-species conserved expression patterns of orthologues were observed during gastrulation and early organogenesis. Notably, *HES4*, known to specify anterior mesoderm within the organizer and to control the proliferation of neural crest and neural cells^{35,36}, was the only paralogue found expressed in humans and monkeys but not in

mice (Extended Data Fig. 5c). Next, we integrated our CS8–11 monkey embryos scRNA-seq dataset with single-cell transcriptomes derived from mouse embryos at corresponding developmental stages (Theiler stage (TS) 9–12). The scmap analysis suggested that cells from CS8–9 and CS11 monkey embryos were mostly comparable to analogous cell types of TS9–12 mouse embryos (Extended Data Fig. 5d–g). Uniform manifold approximation and projection (UMAP) showed well-matched major cell types from both species (Fig. 4a). Based on this integrated dataset, we performed cross-species comparisons and identified many conserved and divergent transcriptomic features of EPI, PS, APS, primordial germ cells (PGCs), ectoderm (NE, FB/MB/HB, SE, SC and NC), mesoderm (node, Nas.Meso, Inter.Meso, Para.Meso, PSM, NMP) and endoderm (VE, DE and Gut) between monkeys and mice (Extended Data Fig. 6 and Supplementary Tables 6–11).

To examine differences in cellular developmental dynamics between mouse and monkey embryos, we studied the expression patterns of selected marker genes during different lineage transitions, including EPI→PS→Nas.Meso→NMP→PSM, VE→DE→Gut and EPI→ECT→FB/MB/HB (Extended Data Fig. 7a). Heatmap and IF analyses showed that genes such as *T*, *EOMES* and *TBX6* exhibited distinct expression patterns between mouse and monkey PS, Nas.Meso, NMP and ectoderm cells (Figs. 2d and 4b and Extended Data Fig. 7a–d). Interestingly, in contrast to mice, many downstream genes of the Hippo signalling pathway were upregulated in monkey NMP and PSM cells (Fig. 4c). To validate and better understand species-specific Hippo signalling activities during PSM differentiation, we took advantage of pluripotent stem cell (PSC)-based in vitro models^{37,38} and recapitulated NMP and PSM differentiation from mouse, monkey and human PSCs (Extended Data Fig. 7e,f). Based on these in vitro models, we confirmed the different expression levels of *MLLT3* and *FOSB* in mouse and monkey PSM-like cells by IF (Extended Data Fig. 7g,h). Consistent with transcriptomic analysis, we found distinct *YAP1* localization between mouse (cytoplasm) and monkey/human (nuclear) PSM-like cells, suggesting lower activities of Hippo kinases in monkey/human PSM (Fig. 4d). In agreement, inhibition of the Hippo pathway by lysophosphatidic acid (LPA) severely impaired mouse but not monkey or human PSM differentiation (Fig. 4e, f).

To gain insight into perigastrulation development in humans, several stem cell embryo models have recently been developed^{39–41}. To date, the fidelity of most models has yet to be evaluated due to limited reference datasets of human embryos. To determine whether our CS8–11 monkey embryos dataset could serve as a surrogate reference, we combined it with datasets from a CS7 (ref. 2) and a CS12 (ref. 3) human embryos. Integrated analysis and annotation of cell types confirmed the high conservation of both humans and monkeys during embryonic development (Extended Data Fig. 8a–f).

Next, we used the monkey reference to evaluate several human stem cell embryo models. By performing integrated analyses, we found that the main cell types from several gastruloids^{42,43}, heart-forming organoids (HFOs)^{44,45}, neuruloids (for example, the MiSTR patterned neural tube that mimics the A–P axis⁴⁶ and the two-dimensional micropatterned, light-induced or chip-based neural tube that mimics the D–V axis^{47–49}) and somitoids⁵⁰ generally overlapped with the CS8–11 monkey embryos reference (Extended Data Figs. 8g,h, 9 and 10 and Supplementary Tables 12 and 13), although notable differences were observed. Using neuruloids and somitoids as proofs of concept, we performed further in-depth comparative analyses of signalling pathways and TFs and found that expression patterns of TGF- β family members *BMP4*, *BMP5* and *BMP7* and WNT ligands in two neural tube models were, by and large, similar to the monkey reference (Extended Data Fig. 8h). However, neither model recapitulated SHH signalling features observed in monkey embryos, suggesting a lack of D–V patterning (Extended Data Fig. 8h). The major cell types during monkey somitogenesis, including NMP, PSM and Para.Meso, were recapitulated in human somitoids (Extended Data Fig. 9a,b). Interestingly,

many genes related to Hippo, WNT, FGF, TGF- β and Notch signalling pathways were found upregulated in monkey embryos when compared with human somitoids (Extended Data Fig. 9c–g). Besides, we identified TFs specifically expressed in monkeys (for example, *EGR1*, *ATF4*, *SRF*, *CHD2*, etc.) and in humans (*POLR2A*) (Extended Data Fig. 9h).

Collectively, these results uncovered conserved and divergent features of embryonic development between monkeys and mice. Moreover, using the CS8–11 monkey embryos dataset as a reference we evaluated several human stem cell embryo models.

Discussion

In this study, through comprehensive scRNA-seq analyses of 56,636 cells, we identified 38 major cell clusters and unveiled the developmental landscapes of all three primary germ layers present during gastrulation and early organogenesis in primates. We then compared the single-cell transcriptomes of CS8–11 monkey embryos with mouse embryos at comparable developmental stages and gained insights into conserved and divergent transcriptomic features across species. In addition, we demonstrated the utility of the CS8–11 monkey embryos dataset as an in vivo reference for authentication of human stem cell embryo models.

The scarcity of CS8–11 human embryos for research, during which primordial organs and the body plan are established, has led to reliance on animal and stem cell embryo models to study this enigmatic period of human embryo development. NHPs, due to their evolutionary resemblance to humans, provide the closest proxy to understanding human embryogenesis. Our comprehensive single-cell transcriptome atlas of a NHP species through CS8 to CS11 not only bridges the knowledge gap in primate embryogenesis but also expands the collection of embryo datasets for comparative developmental biology and benchmarking of embryoid and organoid models.

Online content

Any methods, additional references, Nature Portfolio 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/s41586-022-05526-y>.

1. O’Rahilly, R. & Müller, F. in “Horizons” and a Survey of the Carnegie Collection Section 1, 2–3 (Carnegie Institution of Washington, 1987).
2. Tyser, R. C. V. et al. Single-cell transcriptomic characterization of a gastrulating human embryo. *Nature* **600**, 285–289 (2021).
3. Xu, Y. et al. A single-cell transcriptome atlas of human early embryogenesis. Preprint at *bioRxiv* <https://doi.org/10.1101/2021.11.30.470583> (2021).
4. Nakamura, T. et al. A developmental coordinate of pluripotency among mice, monkeys and humans. *Nature* **537**, 57–62 (2016).
5. Mittenzweig, M. et al. A single-embryo, single-cell time-resolved model for mouse gastrulation. *Cell* **184**, 2825–2842 (2021).
6. Pijuan-Sala, B. et al. A single-cell molecular map of mouse gastrulation and early organogenesis. *Nature* **566**, 490–495 (2019).
7. Chan, M. M. et al. Molecular recording of mammalian embryogenesis. *Nature* **570**, 77–82 (2019).
8. La Manno, G. et al. RNA velocity of single cells. *Nature* **560**, 494–498 (2018).
9. Scheibner, K. et al. Epithelial cell plasticity drives endoderm formation during gastrulation. *Nat. Cell Biol.* **23**, 692–703 (2021).
10. Gilbert, S. F. & Barresi, M. J. F. *Developmental Biology* 12th edn (Sinauer Associates, 2019).
11. Bianchi, D. W., Wilkins-Haug, L. E., Enders, A. C. & Hay, E. D. Origin of extraembryonic mesoderm in experimental animals: relevance to chorionic mosaicism in humans. *Am. J. Med. Genet.* **46**, 542–550 (1993).
12. Ross, C. & Boroviak, T. E. Origin and function of the yolk sac in primate embryogenesis. *Nat. Commun.* **11**, 3760 (2020).
13. Boss, A. L., Chamley, L. W. & James, J. L. Placental formation in early pregnancy: how is the centre of the placenta made? *Hum. Reprod. Update* **24**, 750–760 (2018).
14. Cui, G. et al. Spatial and molecular anatomy of germ layers in the gastrulating cynomolgus monkey embryo. *Cell Rep.* **40**, 11285 (2022).
15. Yoshioka-Kobayashi, K. et al. Coupling delay controls synchronized oscillation in the segmentation clock. *Nature* **580**, 119–123 (2020).
16. Tani, S., Chung, U. I., Ohba, S. & Hojo, H. Understanding paraxial mesoderm development and sclerotome specification for skeletal repair. *Exp. Mol. Med.* **52**, 1166–1177 (2020).

17. Nowotzschin, S. et al. The emergent landscape of the mouse gut endoderm at single-cell resolution. *Nature* **569**, 361–367 (2019).
18. Kimura-Yoshida, C. et al. Canonical Wnt signaling and its antagonist regulate anterior-posterior axis polarization by guiding cell migration in mouse visceral endoderm. *Dev. Cell* **9**, 639–650 (2005).
19. Yamamoto, M. et al. Nodal antagonists regulate formation of the anteroposterior axis of the mouse embryo. *Nature* **428**, 387–392 (2004).
20. Tam, P. P. & Loebel, D. A. Gene function in mouse embryogenesis: get set for gastrulation. *Nat. Rev. Genet.* **8**, 368–381 (2007).
21. Souilhol, C., Cormier, S., Tanigaki, K., Babinet, C. & Cohen-Tannoudji, M. RBP-Jkappa-dependent notch signaling is dispensable for mouse early embryonic development. *Mol. Cell. Biol.* **26**, 4769–4774 (2006).
22. Copp, A. J., Greene, N. D. & Murdoch, J. N. The genetic basis of mammalian neurulation. *Nat. Rev. Genet.* **4**, 784–793 (2003).
23. Spemann, H. & Mangold, H. Induction of embryonic primordia by implantation of organizers from a different species. 1923. *Int. J. Dev. Biol.* **45**, 13–38 (2001).
24. Barth, K. A. et al. Bmp activity establishes a gradient of positional information throughout the entire neural plate. *Development* **126**, 4977–4987 (1999).
25. Patthey, C., Edlund, T. & Gunhaga, L. Wnt-regulated temporal control of BMP exposure directs the choice between neural plate border and epidermal fate. *Development* **136**, 73–83 (2009).
26. Kiecker, C. & Lumsden, A. The role of organizers in patterning the nervous system. *Annu. Rev. Neurosci.* **35**, 347–367 (2012).
27. Martyn, I., Kanno, T. Y., Ruzo, A., Siggia, E. D. & Brivanlou, A. H. Self-organization of a human organizer by combined Wnt and Nodal signalling. *Nature* **558**, 132–135 (2018).
28. Sauka-Spengler, T. & Bronner-Fraser, M. A gene regulatory network orchestrates neural crest formation. *Nat. Rev. Mol. Cell Biol.* **9**, 557–568 (2008).
29. Soldatov, R. et al. Spatiotemporal structure of cell fate decisions in murine neural crest. *Science* **364**, eaas9536 (2019).
30. Nordstrom, U., Jessell, T. M. & Edlund, T. Progressive induction of caudal neural character by graded Wnt signaling. *Nat. Neurosci.* **5**, 525–532 (2002).
31. Briscoe, J., Pierani, A., Jessell, T. M. & Ericson, J. A homeodomain protein code specifies progenitor cell identity and neuronal fate in the ventral neural tube. *Cell* **101**, 435–445 (2000).
32. Fuccillo, M., Joyner, A. L. & Fishell, G. Morphogen to mitogen: the multiple roles of Hedgehog signalling in vertebrate neural development. *Nat. Rev. Neurosci.* **7**, 772–783 (2006).
33. Stamatakis, D., Ulloa, F., Tsoni, S. V., Mynett, A. & Briscoe, J. A gradient of Gli activity mediates graded Sonic Hedgehog signaling in the neural tube. *Genes Dev.* **19**, 626–641 (2005).
34. Lei, Q., Zelman, A. K., Kuang, E., Li, S. & Matisse, M. P. Transduction of graded Hedgehog signaling by a combination of Gli2 and Gli3 activator functions in the developing spinal cord. *Development* **131**, 3593–3604 (2004).
35. Yamaguti, M., Cho, K. W. & Hashimoto, C. Xenopus hairy2b specifies anterior prechordal mesoderm identity within Spemann's organizer. *Dev. Dyn.* **234**, 102–113 (2005).
36. El Yakoubi, W. et al. Hes4 controls proliferative properties of neural stem cells during retinal ontogenesis. *Stem Cells* **30**, 2784–2795 (2012).
37. Diaz-Cuadros, M. et al. In vitro characterization of the human segmentation clock. *Nature* **580**, 113–118 (2020).
38. Matsuda, M. et al. Species-specific segmentation clock periods are due to differential biochemical reaction speeds. *Science* **369**, 1450–1455 (2020).
39. Zheng, Y. et al. Controlled modelling of human epiblast and amnion development using stem cells. *Nature* **573**, 421–425 (2019).
40. Shao, Y. et al. A pluripotent stem cell-based model for post-implantation human amniotic sac development. *Nat. Commun.* **8**, 208 (2017).
41. Hofbauer, P. et al. Cardioids reveal self-organizing principles of human cardiogenesis. *Cell* **184**, 3299–3317 (2021).
42. Minn, K. T. et al. High-resolution transcriptional and morphogenetic profiling of cells from micropatterned human ESC gastruloid cultures. *eLife* **9**, e59445 (2020).
43. Minn, K. T., Dietmann, S., Waye, S. E., Morris, S. A. & Solnica-Krezel, L. Gene expression dynamics underlying cell fate emergence in 2D micropatterned human embryonic stem cell gastruloids. *Stem Cell Rep.* **16**, 1210–1227 (2021).
44. Drakhlis, L., Devadas, S. B. & Zweigerdt, R. Generation of heart-forming organoids from human pluripotent stem cells. *Nat. Protoc.* **16**, 5652–5672 (2021).
45. Drakhlis, L. et al. Human heart-forming organoids recapitulate early heart and foregut development. *Nat. Biotechnol.* **39**, 737–746 (2021).
46. Rifes, P. et al. Modeling neural tube development by differentiation of human embryonic stem cells in a microfluidic WNT gradient. *Nat. Biotechnol.* **38**, 1265–1273 (2020).
47. Haremake, T. et al. Self-organizing neuruloids model developmental aspects of Huntington's disease in the ectodermal compartment. *Nat. Biotechnol.* **37**, 1198–1208 (2019).
48. Karzbrun, E. et al. Human neural tube morphogenesis in vitro by geometric constraints. *Nature* **599**, 268–272 (2021).
49. De Santis, R., Etoc, F., Rosado-Olivieri, E. A. & Brivanlou, A. H. Self-organization of human dorsal-ventral forebrain structures by light induced SHH. *Nat. Commun.* **12**, 6768 (2021).
50. Sanaki-Matsumiya, M. et al. Periodic formation of epithelial somites from human pluripotent stem cells. *Nat. Commun.* **13**, 2325 (2022).

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



Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons license, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons license and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this license, visit <http://creativecommons.org/licenses/by/4.0/>.

© The Author(s) 2022

Methods

Ethical statement

This study was conducted in accordance with the Principles for the Ethical Treatment of Non-Human Primates issued by the Institute of Zoology, Chinese Academy of Sciences (IOZ, CAS), and was approved in advance by the Institutional Animal Care and Use Committee of IOZ, CAS (no. IOZ-EU-20191113 for all monkey experiments, no. IOZ-IACUC-2021-037 for all mouse experiments). Both followed relevant guidelines and regulations. hESC experiments in this study were performed at the UT Southwestern Medical Center and followed the International Society for Stem Cell Research guidelines for Stem Cell Research and Clinical Translation, 2021 (<https://www.isscr.org/policy/guidelines-for-stem-cell-research-and-clinical-translation>). hESC work was reviewed and approved by the UT Southwestern Stem Cell Oversight Committee.

Experiment models and biological sample preparation

Collection of embryonic samples. All *Macaca fascicularis* were of Southeast Asian origin. The animals were maintained at around 25 °C on a 12/12-h light/dark schedule and raised at the Xieerxin Biology Resource with the accreditation of the laboratory animal care facility in Beijing. All animals were given a commercial diet twice per day with tap water ad libitum and were fed vegetables and fruits once daily under careful veterinary supervision. Before the experiment, none of the animals had a clinical or experimental history that would affect physiological ageing or increase susceptibility to diseases.

Oocyte collection, intracytoplasmic sperm injection, pre-implantation embryo culture and transfer of pre-implantation embryos to foster mothers were performed as described by Yamasaki et al.⁵¹. Briefly, female cynomolgus monkeys around 6–8 years of age were chosen for oocyte collection by superovulation with follicle-stimulating hormone, and an implantable and programmable microfusion device was implanted subcutaneously under ultrasound detection. The day when the collected ova were artificially fertilized by sperm injection was designated as embryonic day 0 (EO). When the embryos developed with blastocoel cavities around E6–7, five or six high-quality embryos were selected and transferred to appropriate recipient female cynomolgus monkeys. The implanted embryos were further monitored by ultrasound scanning from E14 to identify successful pregnancies. Ketamine hydrochloride (0.1–0.2 ml kg⁻¹) was administered by intramuscular injection for the anaesthesia of pregnant monkeys. The implanted uterus was surgically removed at different developmental stages as experimentally designed, from which embryonic tissues could be obtained. The sample size of the study was determined based on the availability of highly regulated primate embryo samples. In compliance with the 3R guidelines we reduced the number of animals used to a minimum, which allowed us to obtain a high-coverage transcriptome for each cell type and confidently perform downstream analyses.

C57BL/6 mice were housed under a 12/12-h light/dark cycle at around 25 °C. Natural mating was established between males and 6–8-week-old females, with 12:00 on the day of vaginal plug insertion considered to be EO.5. Postimplantation embryos were dissected from uteri at E7.5–8.5 for the experiments described below.

Isolation of embryonic cells. Monkey embryonic tissues were transferred to DMEM/F12 (DF12) medium (Gibco, no. 21331020) containing 5% Penicillin-Streptomycin (Gibco) and stored at 4 °C for a short period. After washing in PBS (Gibco), tissues were digested with 0.125% TrypLE (Gibco) and 0.025% DNase (Gibco) in DF12 at 37 °C with stirring for 10 min. The disaggregated cell suspension was passed through 40-µm sterile sieve mesh and washed thoroughly with DF12 containing 10% fetal bovine serum (Invitrogen). Sieved cells were precipitated and collected by centrifugation at 300 g for 5 min. Precipitated cells were resuspended

with 5 ml of red blood cell lysis buffer for 3 min and then diluted with an additional 25 ml of DF12 medium. After removal of red blood cells, cells were recentrifuged and transferred to short-term storage at 4 °C.

Preparation of scRNA-seq library and sequencing. Single-cell libraries were constructed using Single Cell 3 Library & Gel Bead Kit v.3 according to the manufacturer's protocol (10X Genomics)⁵². In short, cell counts were assessed using a haemocytometer (Luna-FL, Logos Biosystems) with cell concentration adjusted to 1,000 µl⁻¹. About 16,000 cells were added to each channel of a 10X loading chip and then around 8,000 were captured. Captured cells were lysed, and the isolated RNA was barcoded through reverse transcription in individual gel bead in the emulsion. cDNA was then amplified to construct the library and the qualities of cDNA and cDNA libraries were assessed using Agilent 2100. Finally, the libraries were sequenced on an Illumina HiSeq X Ten platform (Annoroad Gene Technology).

Single-cell transcriptomic analysis

scRNA-seq data preprocessing. Raw fastq files were processed using Cell Ranger 3.1.0 software with default mapping arguments⁵². Reads were mapped to the *Macaca fascicularis* 5.0 genome. Next, the CellRanger 'aggr' command was used to normalize the sequencing depth of different samples, with mean reads per cell above 30,220 post normalization.

Filtering of cells, integration, dimensionality reduction and clustering. The filtered expression matrix with cell barcodes and gene names was loaded with the 'Read10X' function of the Seurat (v.4.0.0) R package⁵³. First, single cells with the number of detected genes (nFeature_RNA) above 500 and detected transcripts (nCount_RNA) above 1,000 were retained to exclude apoptotic or dead cells. Next, doublet or multiplet cells were determined with Scrublet, according to the recommended multiplet rate reference table from 10X Genomics⁵⁴. Next, Seurat objects of different samples (seven samples, Supplementary Table 1) were created independently, with the expression matrix and metadata containing cell barcodes, and cell multiplet information inferred by Scrublet, followed by merging of these Seurat objects. For monkey genes poorly annotated, gene names annotated by *Macaca fascicularis* 5.0 were further converted to those of human-based genes on the published annotation information to better interpret the data⁴. After exclusion of doublet or multiplet cells, 56,636 embryonic cells remained. Next, we used the dataset integration function of Seurat⁵³ to exclude individual heterogeneities between different monkeys. In brief, after normalization of the Seurat object we selected highly-variably expressed genes by the 'mean.var.plot' method at the FindVariableFeatures step, with 2,117 genes found to have highly variable features. These feature genes of anchor and default 30 dimensions of canonical correlation analysis were used for FindIntegrationAnchors, IntegrateData, RunPCA and so on. A tree number of 50 was set as default when finding integration anchors. Subsequently, the Seurat pipeline was used for dimensionality reduction (UMAP) and unsupervised clustering. In most cases we used the default settings of Seurat during dimensionality reduction and unsupervised clustering. To construct the UMAP plot we selected the number of dimensions mainly according to the 'ElbowPlot' function. For UMAP of 56,636 embryonic cells we used the first 16 principal component analysis dimensions at the RunUMAP procedure; the seed used was 42, minimum distance was 0.3 and n.neighbors was 30 as the default setting of Seurat v.4, except that 'umap.method' was 'umap-learn' and the metric was 'correlation'. For clustering of the 56,636 embryo cells the 'k.parameter' of 20 and 'n.trees parameter' of 50 were the default settings during the neighbour-finding process; the number of dimensions used for neighbour finding was 16, as also used for UMAP construction. A resolution of 0.9 was used at the 'FindClusters' step, as shown in Fig. 1b, different types of single cells grouped well.

Differentially expressed gene (DEG) and Gene Ontology (GO) analyses. We computed the DEGs of each cell cluster with RNA assay using the FindAllMarkers function of the Seurat package⁵³. Heatmaps were plotted based on the top ten highly expressed genes (according to adjusted *P* values and fold change) of each cell cluster. The DEGs of each cell cluster from mouse and monkey were used for GO enrichment and analysed by the clusterProfiler R package⁵⁵. GO terms were enriched by the 'compareCluster' function, and 'ont=BP' was set.

Pseudotime analysis. The 'monocle3' R package⁵⁶ was used to calculate the developmental pseudotime of single cells. The Seurat object was converted to a monocle3 object by the 'as.cell_data_set' command of the SeuratWrappers R package⁵³. The developmental trajectory was then constructed with the 'learn_graph' function of the monocle3 R package. After setting the developmental starting point, the 'order cells' command was used to analyse developmental pseudotimes. Finally, pseudotime trajectory was visualized with the 'plot_cells' function.

RNA velocity analysis. Read annotations for sequenced samples were performed using the 'velocity run 10X' command-line tool with BAM, genome annotation and repeat annotation files⁸. BAM files were generated by the default parameters of Cell Ranger software (10X Genomics)⁵². *Macaca fascicularis* 5.0 genome annotations were used to count molecules while separating them into three categories: spliced, unspliced or ambiguous. Repeat annotation files were downloaded from the UCSC genome browser. We then used the UMAP embedding matrix computed by the Seurat pipeline to construct the velocity map with the scVelo python package⁸. Briefly, the loom file containing three categories of count value was loaded to the R environment by the 'ReadVelocity' function of the SeuratWrappers package when the Seurat pipeline was completed. These data were added to the Seurat object, after which the Seurat object was converted to the 'h5ad file' with the SeuratDisk R package⁵³ and the 'h5ad' file was loaded by the 'scv.read' function of the scVelo python package⁵⁷. After the h5ad file was further filtered and normalized, 'pp.moments', 'tl.velocity' and 'tl.velocity_graph' commands were executed to compute RNA velocities. Finally, using the function 'pl.velocity_embedding_stream', RNA velocity vectors were projected onto to the UMAP produced by the Seurat pipeline.

To address concerns about 10X sequencing depth, PS-mesoderm lineage cells (Extended Data Fig. 2b) were divided into 282 microclusters by the Seurat unsupervised clustering method (resolution, 50) based on transcriptomic similarities. Spliced and unspliced transcripts of each microcluster were further merged (the sum of corresponding transcript count values in all cells of each microcluster was computed separately), then each microcluster was treated as a 'pseudocell'. After microclustering, the new Seurat object was recreated with the merged nCount data and pseudocells were annotated according to the maximum cell population of each microcluster. The total number of detected genes and UMI per cell were increased (nCount (UMI), 1×10^5 unspliced and 4×10^5 spliced; nFeature (genes), 10,000 unspliced and 12,000 spliced), which was helpful in regard to compensating for the depth shortage of 3' sequencing. After the Seurat pipeline, UMAP coordinates were substituted with mean UMAP values of cells in each original microcluster. RNA velocity vectors were then computed with the scVelo python package⁸. The validation of velocity on endoderm lineage based on microclustering was performed using a similar method. In addition, 'Velocity_True' and 'Velocity_False' genes were exported from the 'velocity_genes' of the scVelo object, and 'Conflict' genes were computed based on the methods of Barile et al.⁵⁸.

Pseudotime trajectory analysis of Gut. The Seurat object with scale data of Gut was converted to the h5ad file by the SeuratDisk (v.0.0.0.9013) R package⁵⁹, and the h5ad file was then loaded to the

python environment by the 'sc.read' function of the Scanpy (v.1.8.2) python package⁶⁰. Thereafter, principal components were recomputed with the 'tl.pca' function of Scanpy. The Force-directed graph was constructed with the 14 nearest neighbours with default principal components of the scale data (using the Scanpy 'tl.draw_graph' function), and the layout was generated with the ForceAtlas2 algorithm⁶¹. Graph abstraction was computed with the 'tl.paga' function of Scanpy v.1.8.2. The PAGA plot was drawn with the 'pl.paga_compare' function for improved correlation of cell clusters to the Force-directed graph. The threshold for connection of clusters was set to 0.15, node size scale to 3 and edge width scale to 0.8. Diffusion pseudotime⁶² was computed using the 'tl.dpt' function of Scanpy, with cluster 1 set as the starting point.

TF analysis. The pySCENIC analysis in Docker was carried out following three steps⁶³. The gene expression matrix was converted to loom file by the 'loompy' in python, then the 'pyscenic grn', 'pyscience ctx', and 'pyscience auctell' were used to infer the gene regulatory network, regulon prediction and cellular enrichment (area under the curve, AUC) processes with the corresponding cells. After gene regulatory network was produced by 'pyscenic grn', the regulon specificity scores were computed based on the cell clusters identified by Seurat, and we chose top regulons for each cell cluster following 'pyscience ctx'. The AUC matrix was used to score regulon activity of each cell. The AUC scores identified important regulons in cells by "pyscience auctell". The result was a binary regulon activity matrix (binarized activity matrix) that determined in which cells Regulon is 'on'. The SCENIC AUC heatmap was plotted with binarized activity regulons of each cell cluster by the 'pheatmap' R package with the annotation information in the Seurat object.

Cell-cell communication analysis. Cell annotation information and raw count expression matrix were exported from the Seurat file with suggested scripts using the CellPhoneDB protocol^{64,65}. Cell annotation information and count expression matrix were then used as input for CellPhoneDB statistical analysis with default settings, and this step together with the following plotting step was executed at the Linux command-line interface supported by the protocol. The database of receptor-ligand interactions was generated for human proteins, and the genes of the monkey have been transferred to human genes at the maximum extent to minimize differences in receptor-ligand interactions that might vary between monkeys and humans. Finally, we showed some notable interactions between relevant cell types with the dot-plot function of CellPhoneDB.

Comparison of single-cell transcriptomic dataset among mouse, human and monkey. To project monkey single-cell data onto the mouse UMAP, the mouse single-cell reference dataset was first prepared (Fig. 4a). scRNA-seq data of early mouse embryogenesis⁶ were obtained from EMBL-EBI ArrayExpress under experiment code no. E-MTAB-6967. The count expression matrix and cell annotation files supplied were used to create the Seurat object with the Seurat (v.4.0.0) R package⁵⁹. Using the method of Blanca Pijuan-Sala et al.⁶, 116,312 single-cell transcriptomes remained. The mouse Seurat object (reference dataset) was created with 13,805 monkey/mouse shared genes. Following the RunUMAP procedure (return.model, TRUE), UMAP cell-embedding values were replaced by those supplied in the cell annotation file of no. E-MTAB-6967. Cell clusters in the UMAP plot, as shown in Fig. 4a (right), were also annotated according to the annotation files supplied. The monkey Seurat object (query dataset) was also created, with 13,805 monkey/mouse shared genes. The anchors between mouse and monkey data were found with the FindTransferAnchors function (reference.reduction, 'pca'; dims, 1:50; k.filter, NA), and the function MapQuery (reference.reduction, 'pca'; reduction.model, 'umap') was used to project monkey embryo single-cell data onto the mouse embryo

single-cell data-based UMAP structure. The cell clusters shown in Fig. 1b are shown in the projected UMAP plot in Fig. 4a (left).

To do integration analysis for monkey and human single-cell data, the human CS7 embryo² and various embryoid datasets including gastruloids^{42,43}, HFOs⁴⁵, neuruloids^{46–49} and somitoids⁵⁰ were prepared (Extended Data Figs. 8a–d,g, 9a,b and 10a,d,g). We used the ‘biomaRt’ package to convert genes from cynomolgus monkeys and mice to human homologous genes⁵³. Seurat lists were split by samples or species, and each list was normalized using ‘NormalizeData’ function. Next, 2,000 genes were selected as anchor features. Using the R package ‘Seurat’ with the functions ‘FindIntegrationAnchors’ and ‘IntegrateData’, based on canonical correlation analysis and mutual nearest-neighbours algorithms, we acquired the integration Seurat objects of cynomolgus monkeys with mice then set the default assay as ‘integrated’. UMAPs were calculated using the function of ‘RunUMAP’ with dimensions set as 30. The same methods were performed for integration between natural monkey and human embryos² (or human embryoids).

Developmental staging of monkey and mouse embryos. We selected EPI, rostral neuroectoderm, SE, forebrain/midbrain/hindbrain and NC from the E6.5–8.5 mouse embryo dataset⁶ and compared them with their counterpart cells in CS8–11 cynomolgus monkey embryos using the ‘scmap’ R package^{4,66}. Furthermore, we selected 1,000 genes by setting ‘n_feature=1000’ in the function ‘selectFeatures’ with the parameter threshold set to 0 in ‘scmapCluster’. Default parameters were used for all other steps. We performed the same strategy for the developmental stage comparison of mesoderm and endoderm between cynomolgus monkeys and mice.

Comparison of cellular signalling pathways across species. We downloaded gene information for the WNT, FGF, TGF- β , SHH, Hippo and Notch signalling pathways from the MSigDB database⁶⁷. Gene expression of various cell types was detected in cynomolgus monkey, human and mouse embryos and in human stem cell embryo models. Gene expression was scaled from -1 to 1 from the integrated data, and average expression level was measured by cell type using the ‘Average-Expression’ function in the Seurat R package.

IF analysis on paraffin-fixed embryo sections

Embryonic samples (monkey E22 and mouse E7.5–8.5 embryos) were immediately fixed in 4% paraformaldehyde overnight at 4 °C and subsequently embedded in paraffin. Sections (5 μ m) on slides were dewaxed and rehydrated with xylene and ethanol gradients. Slides were immersed in 0.01 mol l⁻¹ citric acid buffer solution (C₆H₈O₇·H₂O:C₆H₅Na₃O₇·2H₂O, 1:9, pH 6.0) and heated in a microwave oven at 92–98 °C for 15 min for antigen retrieval. After cooling to room temperature, the slides were washed three times with 1 \times PBS (5 min each), incubated with 1% Triton X-100 for 30 min and blocked with 2% bovine serum albumin (BSA) for 30 min at room temperature. Next, the slides were incubated with primary antibodies (Supplementary Table 14), diluted with blocking solution overnight at 4 °C and washed three times with PBST (1 \times PBS with 0.05% Tween-20, 5 min each). The slides were then incubated with secondary antibodies diluted with blocking solution and 1 mg ml⁻¹ DAPI (Invitrogen, no. D3571) for 1 h. Finally, after washing three times with PBST (5 min each) the slides were mounted with anti-fade mounting medium (Gibco). IF images were captured by laser-scanning confocal microscope LSM 880 (Carl Zeiss) and processed with Imaris 9.0.2 software (Bitplane) and Zen 7.0 (Carl Zeiss).

Validation by stem cell models

Pluripotent stem cell culture. Human embryonic stem cell line H9 (WA09) was obtained from WiCell and authenticated by short tandem repeat profiling. Mouse epiblast stem cells (mEpiSCs) and rhesus macaque ES cells were generated and identified as described in a previous study^{68,69}. Mycoplasma testing for cell lines was negative. hESCs

were maintained in mTeSR Plus medium (STEMCELL Technologies) in a 0.5% Matrigel (BD Biosciences)-coated culture dish at 37 °C and in 5% CO₂. hESCs were dissociated with accutase (STEMCELL Technologies) and split in a 1:1 ratio. A single-cell suspension was seeded into a Matrigel-coated dish in mTeSR Plus medium containing 10 μ M ROCK inhibitor (no. Y27632, Sigma-Aldrich). mEpiSCs were maintained on mouse embryonic fibroblast (MEF) feeder cells in a gelatin-coated dish, in NBFR medium containing DF12 and Neurobasal medium (Invitrogen) mixed in a 1:1 ratio, 0.5 \times N2 supplement (Invitrogen), 0.5 \times B27 supplement (Invitrogen), 2 mM GlutaMax (Gibco), 1 \times nonessential amino acids (NEAA, Gibco), 0.1 mM 2-mercaptoethanol (Sigma-Aldrich), 20 ng ml⁻¹ FGF2 and 2.5 μ M IWR1. mEpiSCs were dissociated with TrypLE (ThermoFisher) and split in a 1:30 ratio. Rhesus macaque ES cells were maintained on mouse embryonic fibroblast feeder cells in a gelatin-coated dish, in NBFR medium supplemented with 5 mg ml⁻¹ BSA (MP Biomedicals). Cells were dissociated with TrypLE and split in a 1:10 ratio. A single-cell suspension was seeded in NBFR (5 mg ml⁻¹ BSA) medium containing 10 μ M ROCK inhibitor.

In vitro differentiation of PSM-like cells. Presomitic mesoderm differentiation was carried out as described in a previous study³⁷. On the day of differentiation (day 0), 100,000–150,000 cells (10,416–15,625 cells cm⁻²) were seeded in a Matrigel-coated, 35 mm dish in pluripotency maintenance medium. Cells were maintained in incubator at 37 °C for about 2 h before changing to differentiation medium; differentiation medium contains DF12 with 1 \times N2 supplement and Neurobasal medium (Invitrogen) with 1 \times B27 supplement in a 1:1 ratio. The medium was also supplemented with 2 mM Glutamax, 0.1 mM nonessential amino acids, 1 mM sodium pyruvate (Gibco), Penicillin-Streptomycin (Gibco), 10 μ M CHIR99021 (Selleckchem) and 0.5 μ M LDN193189 (Selleckchem). Differentiation medium was changed daily. The same protocol was used for differentiation of human, monkey and mouse PSM-like cells. For Hippo inhibition experiments, cells were treated with 0.5 μ M 1-Oleoyl LPA (OCRIS, no. 3854) on day 1.

IF staining and microscopy. For IF staining, 10,416–15,625 cells cm⁻² were initially seeded on Matrigel-coated μ -Slide eight-well chambered coverslips (ibidi, for high-end microscopy). The cells were fixed at the indicated time points (days) in 4% paraformaldehyde for 15 min at room temperature. Fixed cells were washed twice with PBS before permeabilization and blocking with 3% donkey serum in PBST (1 \times PBS with 0.1% Triton X-100) for 1 h at room temperature. Samples were incubated with primary antibody (Supplementary Table 14) and diluted in blocking buffer at room temperature for 2 h or 4 °C overnight followed by 30 min of PBST wash, repeated twice. Secondary antibodies (ThermoFisher) were diluted in blocking buffer in a 1:500 ratio. DAPI staining was performed, together with secondary antibodies, at room temperature for 1 h followed by three PBST washes. Samples were then soaked in PBS before imaging. Fluorescence imaging was performed on either (1) a Nikon CSU-W1 SoRa spinning-disk confocal microscope with objectives \times 20/0.45 numerical aperture (NA), WD 8.9–6.9, air, \times 40/0.6 NA, WD 3.6–2.85, air and \times 100/1.45 NA, oil or (2) a Zeiss LSM 800 laser-scanning confocal microscope with a \times 40/1.3 NA oil objective.

Imaging and statistical analysis. Statistical analyses were repeated at least twice, with consistent results. In the figure captions *n* denotes the number of biological replicates in the same experiment. Raw images were first processed in Fiji⁷⁰ to create maximal intensity projection (MIP) and export look-up table of representative images. For MLLT3 and FOSB data shown in Extended Data Fig. 7g,h, nuclear segmentation was performed in Ilastik⁷¹. MIP images and segmentation masks were processed in MATLAB (R2022a) using custom code, which is available in a public repository. Nuclear localized fluorescence intensity of transcription factors was computed for each cell in a given field, and the value was then normalized to the DAPI intensity of the same

Article

cell. Values of all cells were plotted as mean \pm s.e.m. For data shown in Fig. 4f, total cell and TBX6-positive cell numbers were calculated with Imaris (v.9.9, Oxford Instruments) using the SPOTS function. Total cell number was calculated by counting the number of nuclei (DAPI). The same parameters for computation of the spots were applied to the DAPI and TBX6 channels. Data were shown as mean \pm s.d. *P* values were determined by unpaired *t*-test. GraphPad Prism v.7.0 was used to plot the data shown in Fig. 4f and Extended Data Fig. 7g,h.

Reporting summary

Further information on research design is available in the Nature Portfolio Reporting Summary linked to this article.

Data availability

The dataset of primate gastrulation and early organogenesis generated in the current study is available in the NCBI Gene Expression Omnibus (GEO) under accession no. GSE193007. The dataset of mouse gastrulation and early organogenesis used as reference is available at ArrayExpress under accession no. E-MTAB-6967. The datasets of human CS7 and CS12 embryos used as reference are available at ArrayExpress under accession no. E-MTAB-9388 and at GEO under accession no. GSE157329. The datasets of human neuruloids are available at GEO under accession nos. GSE118682, GSE173492 and GSE163505. The dataset of human somitoids is available at the ArrayExpress database under accession code E-MTAB-11334. The dataset of human gastruloids is available at GEO under accession nos. GSE144897 and GSE169074. The dataset of heart-forming organoids is available at GEO under accession no. GSE150202. The dataset of MiSTR patterned human neuruloids is available at GEO under accession no. GSE135399. Source data are provided with this paper.

Code availability

Custom MATLAB code for image processing is available in a public repository (<https://github.com/ecamacho90/BiologicalImageProcessing>, <https://doi.org/10.5281/zenodo.7367663>). The other codes are available upon request.

51. Yamasaki, J. et al. Vitrification and transfer of cynomolgus monkey (*Macaca fascicularis*) embryos fertilized by intracytoplasmic sperm injection. *Theriogenology* **76**, 33–38 (2011).
52. Zheng, G. X. et al. Massively parallel digital transcriptional profiling of single cells. *Nat. Commun.* **8**, 14049 (2017).
53. Stuart, T. et al. Comprehensive integration of single-cell data. *Cell* **177**, 1888–1902 (2019).
54. Wolock, S. L., Lopez, R. & Klein, A. M. Scrublet: computational identification of cell doublets in single-cell transcriptomic data. *Cell Syst.* **8**, 281–291 (2019).
55. Yu, G., Wang, L. G., Han, Y. & He, Q. Y. clusterProfiler: an R package for comparing biological themes among gene clusters. *OMICS* **16**, 284–287 (2012).
56. Qiu, X. et al. Reversed graph embedding resolves complex single-cell trajectories. *Nat. Methods* **14**, 979–982 (2017).
57. Bergen, V., Lange, M., Peidli, S., Wolf, F. A. & Theis, F. J. Generalizing RNA velocity to transient cell states through dynamical modeling. *Nat. Biotechnol.* **38**, 1408–1414 (2020).

58. Barile, M. et al. Coordinated changes in gene expression kinetics underlie both mouse and human erythroid maturation. *Genome Biol.* **22**, 197 (2021).
59. Hao, Y. et al. Integrated analysis of multimodal single-cell data. *Cell* **184**, 3573–3587 (2021).
60. Wolf, F. A., Angerer, P. & Theis, F. J. SCANPY: large-scale single-cell gene expression data analysis. *Genome Biol.* **19**, 15 (2018).
61. Jacomy, M., Venturini, T., Heymann, S. & Bastian, M. ForceAtlas2, a continuous graph layout algorithm for handy network visualization designed for the Gephi software. *PLoS ONE* **9**, e98679 (2014).
62. Haghverdi, L., Buttner, M., Wolf, F. A., Büttner, F. & Theis, F. J. Diffusion pseudotime robustly reconstructs lineage branching. *Nat. Methods* **13**, 845–848 (2016).
63. Van de Sande, B. et al. A scalable SCENIC workflow for single-cell gene regulatory network analysis. *Nat. Protoc.* **15**, 2247–2276 (2020).
64. Efremova, M., Vento-Tormo, M., Teichmann, S. A. & Vento-Tormo, R. CellPhoneDB: inferring cell-cell communication from combined expression of multi-subunit ligand-receptor complexes. *Nat. Protoc.* **15**, 1484–1506 (2020).
65. Vento-Tormo, R. et al. Single-cell reconstruction of the early maternal-fetal interface in humans. *Nature* **563**, 347–353 (2018).
66. Kiselev, V. Y., Yiu, A. & Hemberg, M. scmap: Projection of single-cell RNA-seq data across data sets. *Nat. Methods* **15**, 359–362 (2018).
67. Subramanian, A. et al. Gene set enrichment analysis: a knowledge-based approach for interpreting genome-wide expression profiles. *Proc. Natl Acad. Sci. USA* **102**, 15545–15550 (2005).
68. Wu, J. et al. Interspecies chimerism with mammalian pluripotent stem cells. *Cell* **168**, 473–486 (2017).
69. Wu, J. et al. An alternative pluripotent state confers interspecies chimaeric competency. *Nature* **521**, 316–321 (2015).
70. Schindelin, J. et al. Fiji: an open-source platform for biological-image analysis. *Nat. Methods* **9**, 676–682 (2012).
71. Kreshuk, A. et al. Automated detection and segmentation of synaptic contacts in nearly isotropic serial electron microscopy images. *PLoS ONE* **6**, e24899 (2011).

Acknowledgements We thank Q. Zhou and W. Li (Chinese Academy of Sciences, CAS) for their invaluable comments and guidance. We thank A. Warmflash (Rice University) for his viewpoints and comments on mammalian gastrulation. We thank Y. Shen for her outstanding support with sample collection and management, and S. Li, X. Zhu and Y. Wang at the CAS imaging platform for their assistance with image analysis. We thank E. Camacho Aguilar for technical support with image processing and the UTSW Harold C. Simmons Cancer Center for their shared resources at the Quantitative Light Microscopy Core. This work was supported by the Strategic Priority Research Program of the Chinese Academy of Sciences (no. XDA16020700), National Key Research and Development Program of China (nos. 2021YFA0805701, 2020YFA0112201 and 2021YFC2700302) and by grants from CAS (nos. XDA16021400 and ZDBS-ZRKJZ-TLCO01) and the National Natural Science Foundation of China (nos. 82192870 and 32100654). J.W. is a New York Stem Cell Foundation–Robertson Investigator and Virginia Murchison Linthicum Scholar in Medical Research and the work is funded by CPRIT (no. RR170076), NIH (no. HD103627-01A1) and Welch (no. 854671).

Author contributions H. Wang, J.W., F.G. and X.J. conceptualized this project and supervised the overall experiments. J.Z., H. Wan, L.Y., Z.X. and D.Y. performed collection of embryonic tissues. J.Z. and X.J. isolated embryonic single cells and established scRNA-seq library construction. J.G. and X.J. performed the original bioinformatics analysis and management. L.L. performed all experiments on stem cell models. J.Z., L.Q. and Y.X. performed IF staining. J.Z., L.Q., L.L., Y.X., W.Z., K.Y. and X.W. performed dataset analysis. H. Wang, J.W., F.G., J.Z., L.Q., L.L. and D.A.S. wrote the manuscript.

Competing interests The authors declare no competing interests.

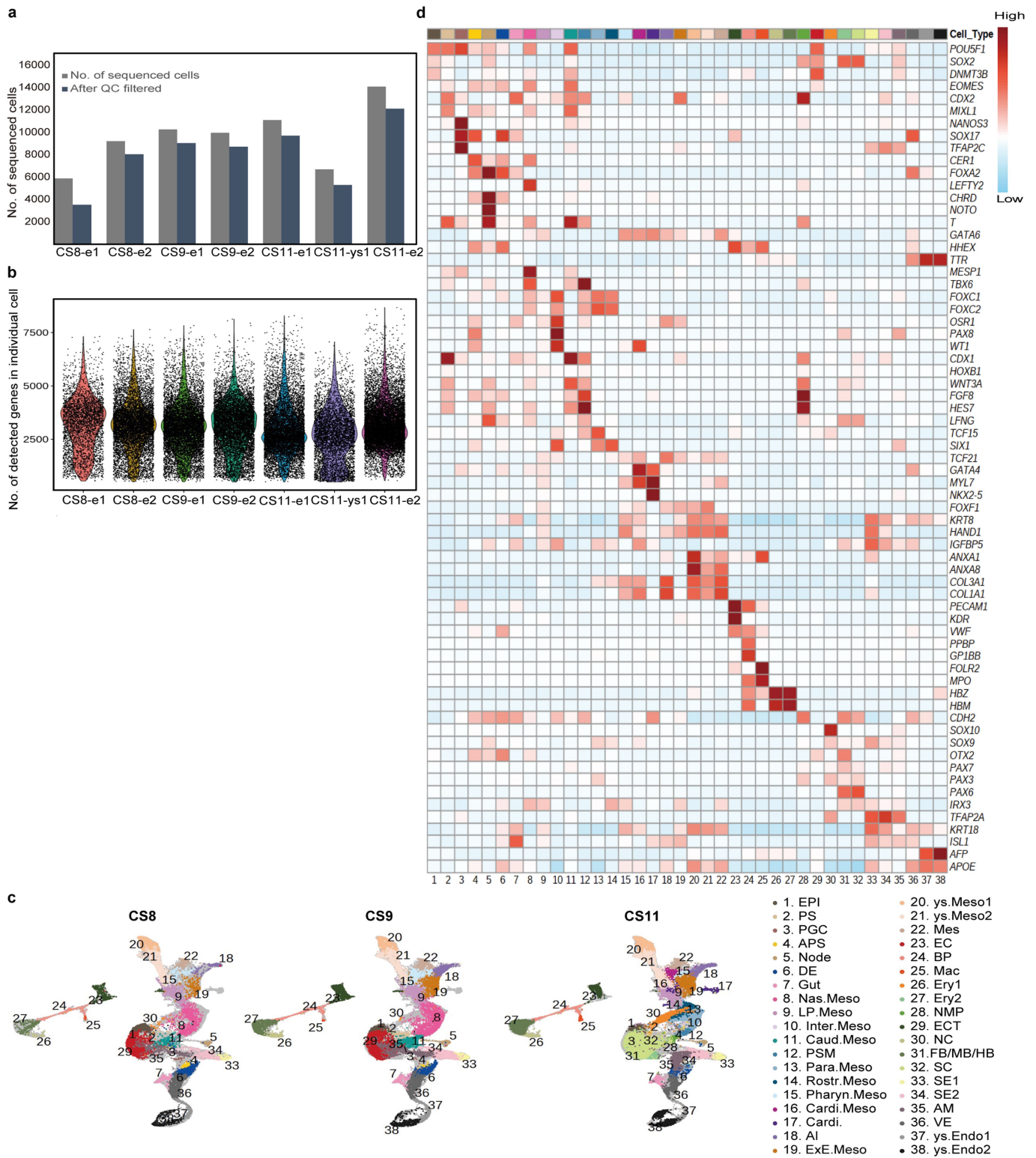
Additional information

Supplementary information The online version contains supplementary material available at <https://doi.org/10.1038/s41586-022-05526-y>.

Correspondence and requests for materials should be addressed to Xiangxiang Jiang, Fan Guo, Jun Wu or Hongmei Wang.

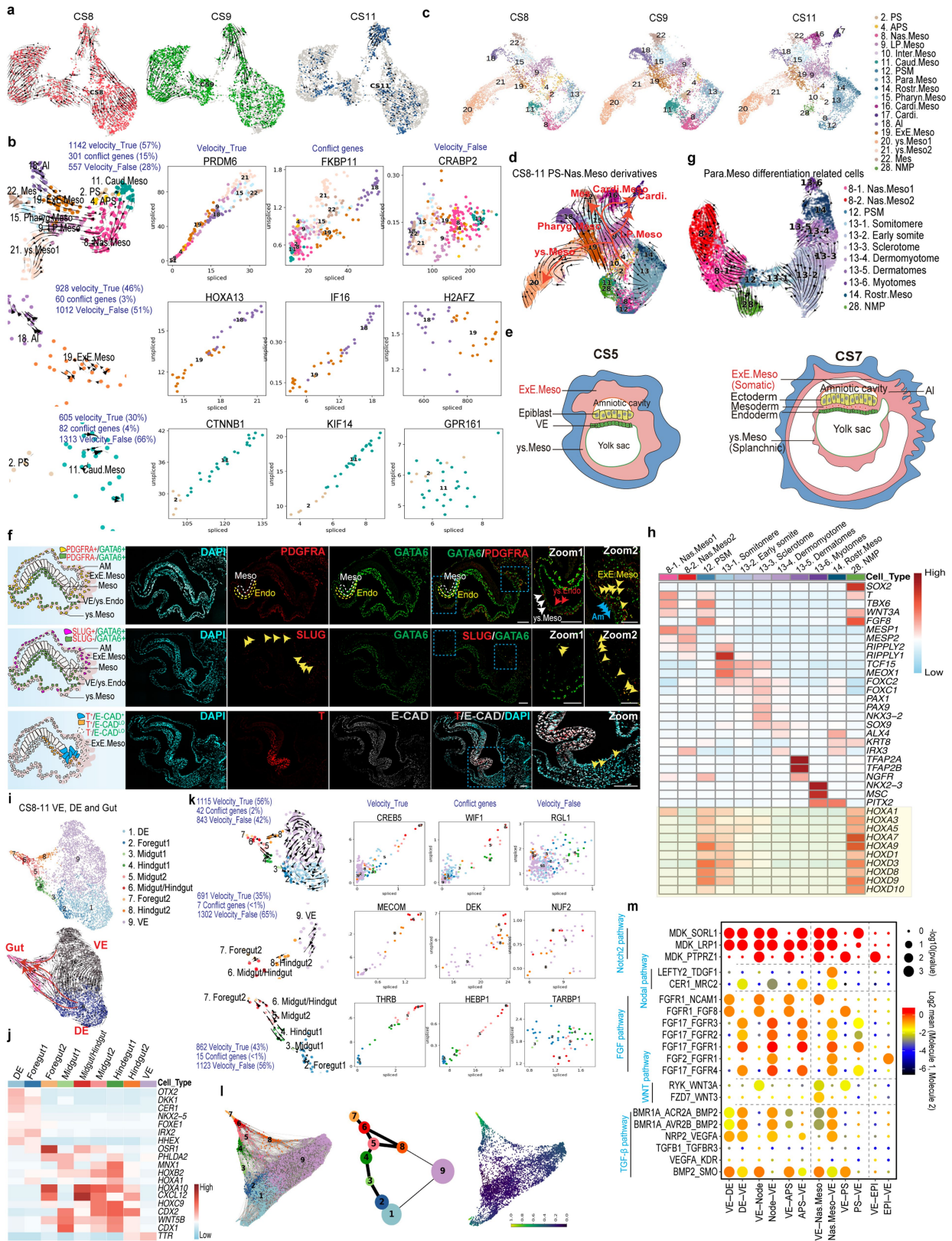
Peer review information Nature thanks Ramiro Alberio and the other, anonymous, reviewer(s) for their contribution to the peer review of this work.

Reprints and permissions information is available at <http://www.nature.com/reprints>.



Extended Data Fig. 1 | Transcriptomic features of the cynomolgus monkey embryos at CS8–11. a, Bar graph showing the number of collected single cells and filtered cells of each sample. **b**, Box plot of the gene numbers detected in each sample. **c**, UMAPs showing the distribution of 38 major cell types in indicated developmental stages (CS8–11). EPI, epiblast; PS, primitive streak; PGC, primordial germ cells; APS, anterior primitive streak; DE, definitive endoderm; Nas.Meso, nascent mesoderm; LP.Meso, lateral plate mesoderm; Inter.Meso, intermediate mesoderm; Caud.Meso, caudal mesoderm; PSM, presomitic mesoderm; Para.Meso, paraxial mesoderm; Rostr.Meso, rostral mesoderm; Pharyn.Meso, pharyngeal mesoderm; Cardi.Meso, cardiac

mesoderm; Cardi., cardiac tissue; AI, allantois; ExE.Meso, extra-embryonic mesoderm; ys.Meso, yolk sac mesoderm; Mes, mesenchyme; EC, endothelial cell; BP, blood progenitor; Mac, macrophage; Ery, erythrocytes; NMP, neuromesodermal progenitor; ECT, ectoderm; NC, neural crest; FB/MB/HB, forebrain/midbrain/hindbrain; SC, spinal cord; SE, surface ectoderm; AM, amnion; VE, visceral endoderm; ys.Endo, yolk sac endoderm. **d**, Heatmap visualizing the representative differentially expressed genes (DEG, here and after) for the 38 cell clusters. Colors from blue to red indicate the low to high expression of different genes in indicated cell types, here and after.

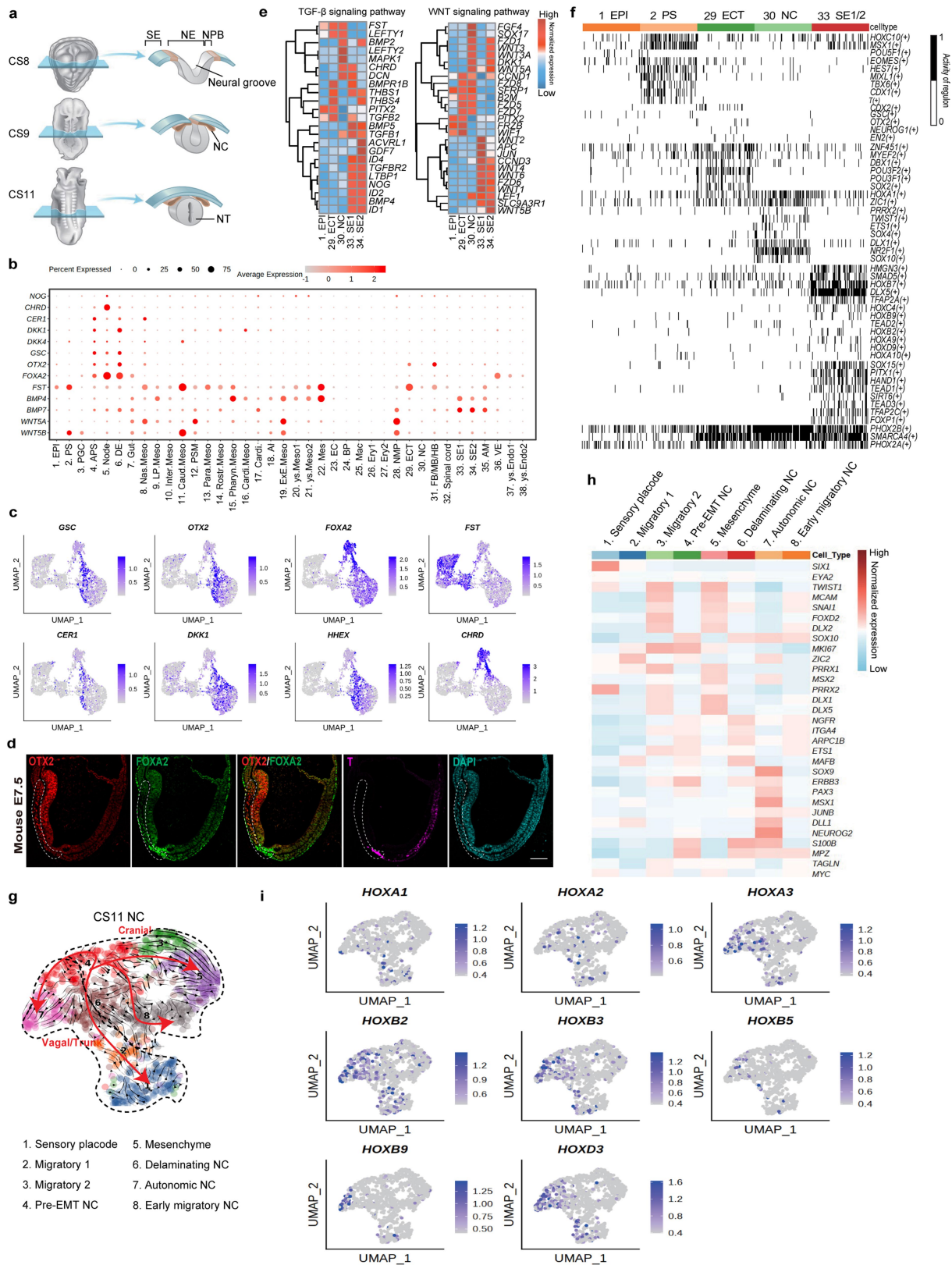


Extended Data Fig. 2 | See next page for caption.

Extended Data Fig. 2 | Transcriptomic characteristics of monkey primitive streak and mesoendoderm.

a, UMAPs overlaid with RNA velocity of PS derivatives from embryos at indicated stages. Black arrows hereafter represent calculated velocity trajectories. **b**, (left) UMAPs overlaid with RNA velocity of PS-mesoderm lineage cells after micro-clustering (see Methods). (right) Phase portraits of representative Velocity_True, Conflict and Velocity_False genes (unspliced versus spliced dependency) that support these velocity analyses. Each point corresponds to a “pseudo” cell, colored by cluster labels, here and after. **c**, UMAPs showing the distribution of PS-mesoderm lineage derivatives at indicated developmental stages. **d**, UMAP overlaid with RNA velocity of PS-mesoderm lineage derivatives from CS8–11 embryos. **e**, Diagrams showing potential originations of extra-embryonic mesenchymal cells (EXMCs, including ys.Meso, ExE.Meso, some mesenchyme, among others) at pre (CS5) or perigastrulation (CS7) stage. The left diagram indicating the EXMCs at pregastrulation are putatively generated from hypoblast derived yolk sac. The right one indicating the putative splanchnic or somatic migrating route of different EXMCs generated from PS/mesoderm at perigastrulation stage. **f**, (left) Diagrams summarizing the distribution of indicated cell types in the region of interest, based on the right IF images. (right) IF results exhibiting the localization of PDGFRA⁺, GATA6⁺, SLUG⁺, T⁺, and E-CAD^{low} cells on a CS8 (E22) monkey embryo. PDGFRA, the marker for mesoderm. GATA6, the marker for EXMCs, endoderm and some mesoderm. T, the marker for primitive streak and some mesoderm. SLUG, the marker for epithelial-mesenchymal transition. E-CAD, the marker for epithelial cells. DAPI, marker for DNA, here and after. Blue dashed frames refer to the zoomed regions; yellow arrowheads indicating the potential ExE.Meso; Blue arrowheads indicating the AM. White arrowheads indicating the ys.Meso (PDGFRA⁺/GATA6⁺) and red arrowheads indicating the

ys.Endo (PDGFRA⁻/GATA6⁺) of the CS8 secondary yolk sac. Scale bars, 100 μm. **g**, UMAP overlaid with RNA velocity of Para.Meso differentiation related cell types from CS8–11 embryos. **h**, Heatmap showing the Para.Meso-related cell types based on DEG analysis. Color bar in heatmap, normalized expression level of different genes in indicated cell types, here and after. **i**, (top) UMAP showing the subtypes of VE, DE and Gut from CS8–11 embryos. (bottom) UMAP overlaid with velocity of VE, DE and Gut cells. **j**, Heatmap showing the representative DEGs of different gut tube subtypes in **i**. **k**, (left) UMAPs overlaid with RNA velocity of endoderm lineage cells after micro-clustering. (right) Phase portraits of representative Velocity_True, Conflict and Velocity_False genes that support the velocity analyses. **l**, (left) FA1 (Force-directed graph) exhibiting expected descendant and ancestor relationship between different cell types. (middle) PAGA analysis on DE, VE and different Gut cells. The boldness of the line indicates the degree of the relationship between clusters. Cell type annotations are color coded. (right) FA1 map showing the diffusion pseudotime analysis of the 9 annotated cell types. The color gradient from dark to bright refers to the level of pseudotime differentiation diffusion from low to high. **m**, Overview of the representative ligand-receptor interactions between VE and embryonic cells including EPI, ECT, PS, APS, Nas.Meso, and DE, analyzed by Cellphone DB. On the horizontal axis, the bottom factors indicating the ligand, and the top factors indicating the receptors. On the vertical axis, the left one indicating the tissue secreting signal-ligands, while the right one indicating the tissue expressing the receptors and receiving the signals. The P value (indicated by circle size, scale on the right) in the analyses indicates the likelihood of cell-type enrichment of each ligand-receptor complex, here and after. The means of the average expression level of interacting molecule 1 in cluster 1 and interacting molecule 2 in cluster 2 are indicated by color.

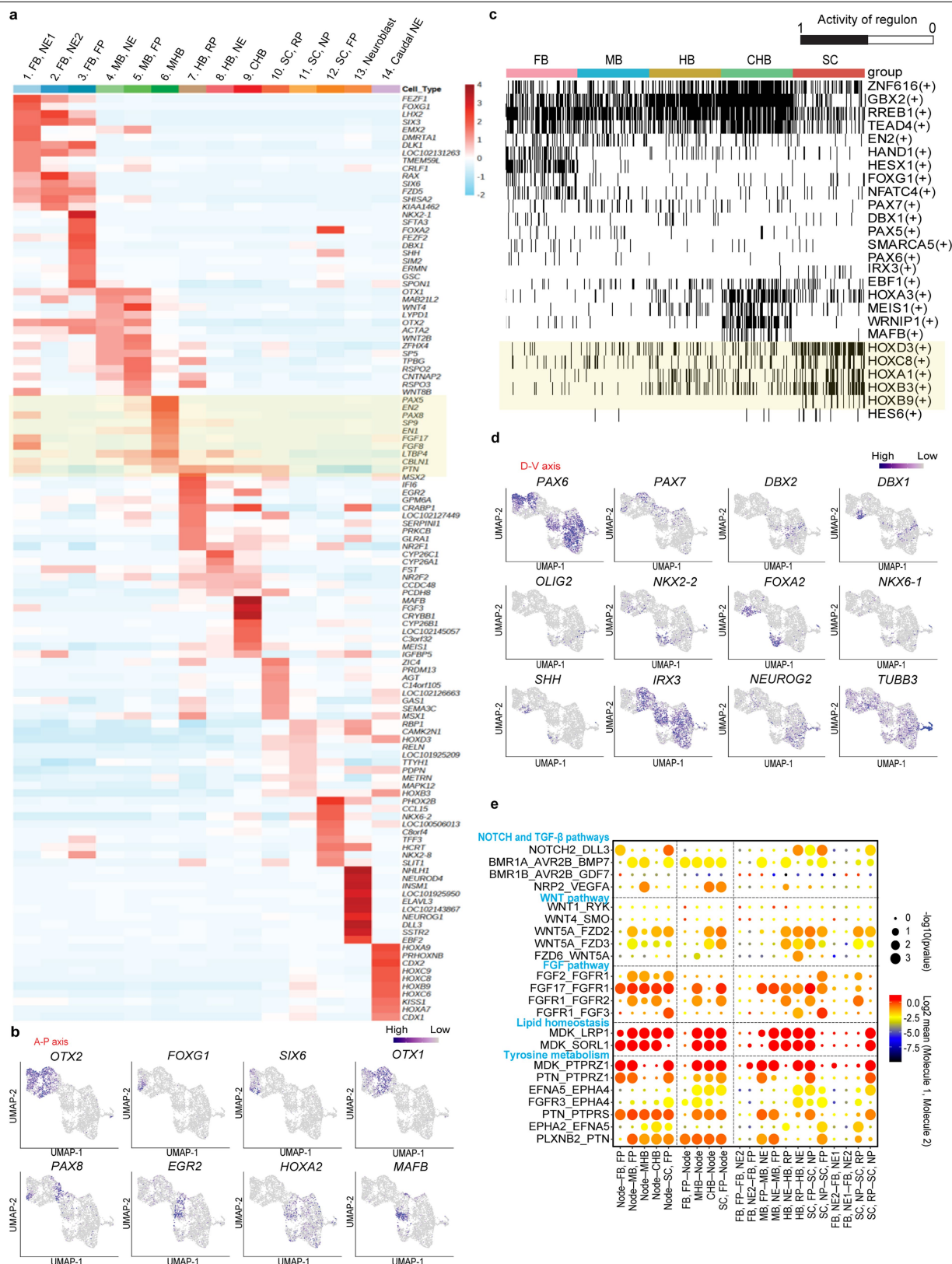


Extended Data Fig. 3 | See next page for caption.

Extended Data Fig. 3 | The single-cell transcriptomic analysis on ectoderm.

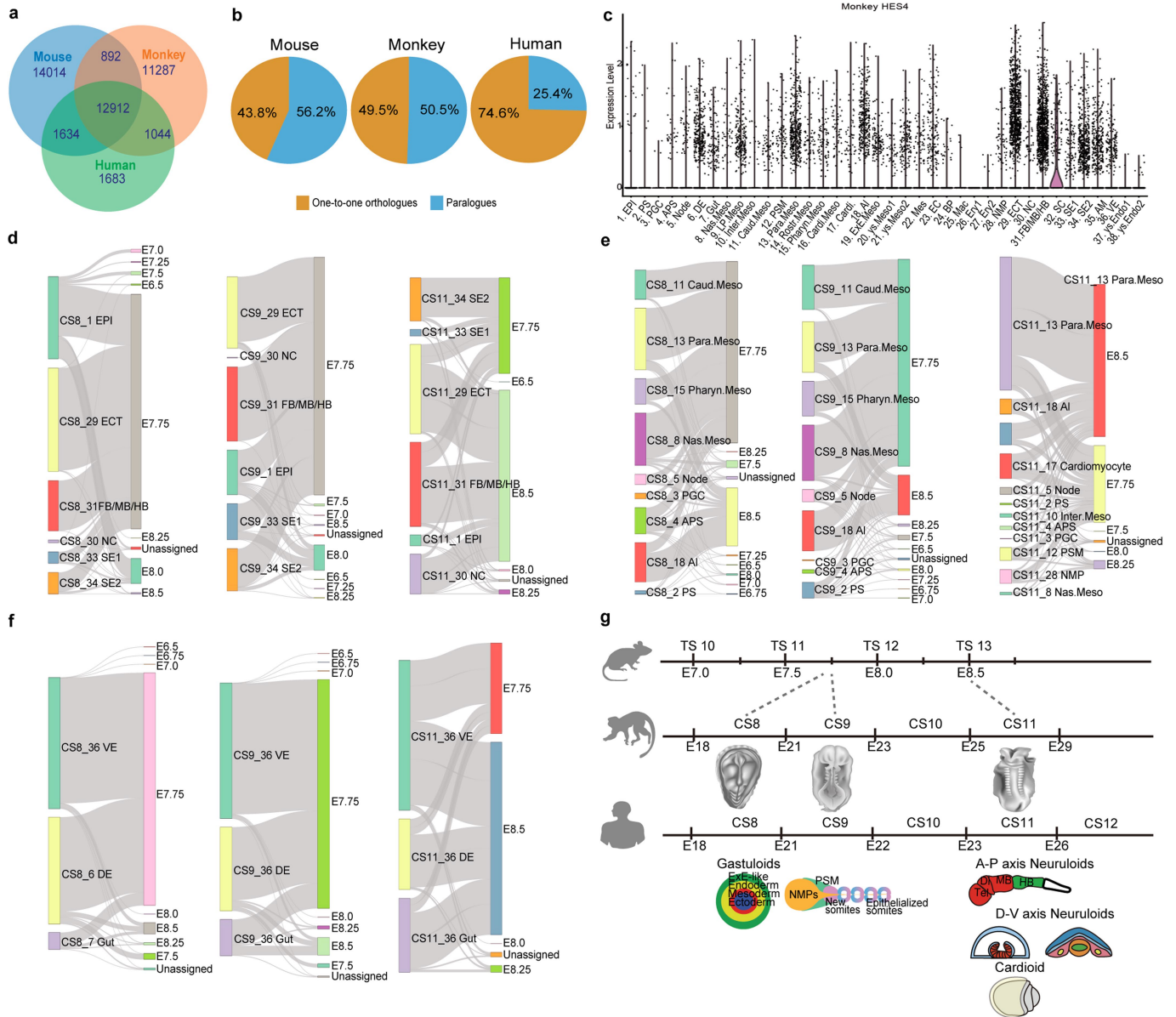
a, Diagram summarizing the formation of neural tube and the differentiation of neural and non-neural ectoderm in the CS8–11 primate embryos. NPB, neural plate border; NT, neural tube. **b**, Dot plot map showing the expression levels of specific marker genes of the potential organizer. The plot sizes from 0–75 indicate the percentage of cells expressed the indicated genes. Colors from grey to red indicate the average expression level of different genes in indicated cell types. **c**, UMAPs showing the expression of organizer marker genes in PS, node, DE, Nas.Meso, which are particularly analyzed in Fig. 2a. Color bars in UMAPs, normalized expression level of indicated genes from high (purple) to low (grey), here and after. **d**, Representative IF results exhibiting the localization of organizer ($OTX2^+/FOXA2^+/T^{+Low}$) in E7.5 mouse embryos. White dashed lines

highlight the potential region of organizer cells. DAPI, marker for DNA. Scale bar, 100 μm . $n = 3$ biologically independent samples. **e**, Heatmaps showing the expression of genes in TGF- β and WNT signalling pathways in relative cell types. Color bar in heatmap, the normalized expression level of different genes in indicated cell types, here and after. **f**, Heatmap showing the representative transcription factors (TFs) in the differentiation of ECT. Color bar, activity of regulons, indicating the active (black, 1) or non-active (white, 0) of different transcriptome factors. **g**, UMAP overlaid with the velocity analysis on different NC subtypes. EMT, epithelial-mesenchymal transition. **h**, Heatmap showing the representative DEGs in various NC subtypes. **i**, UMAPs showing the expression patterns of HOX family genes in NC.



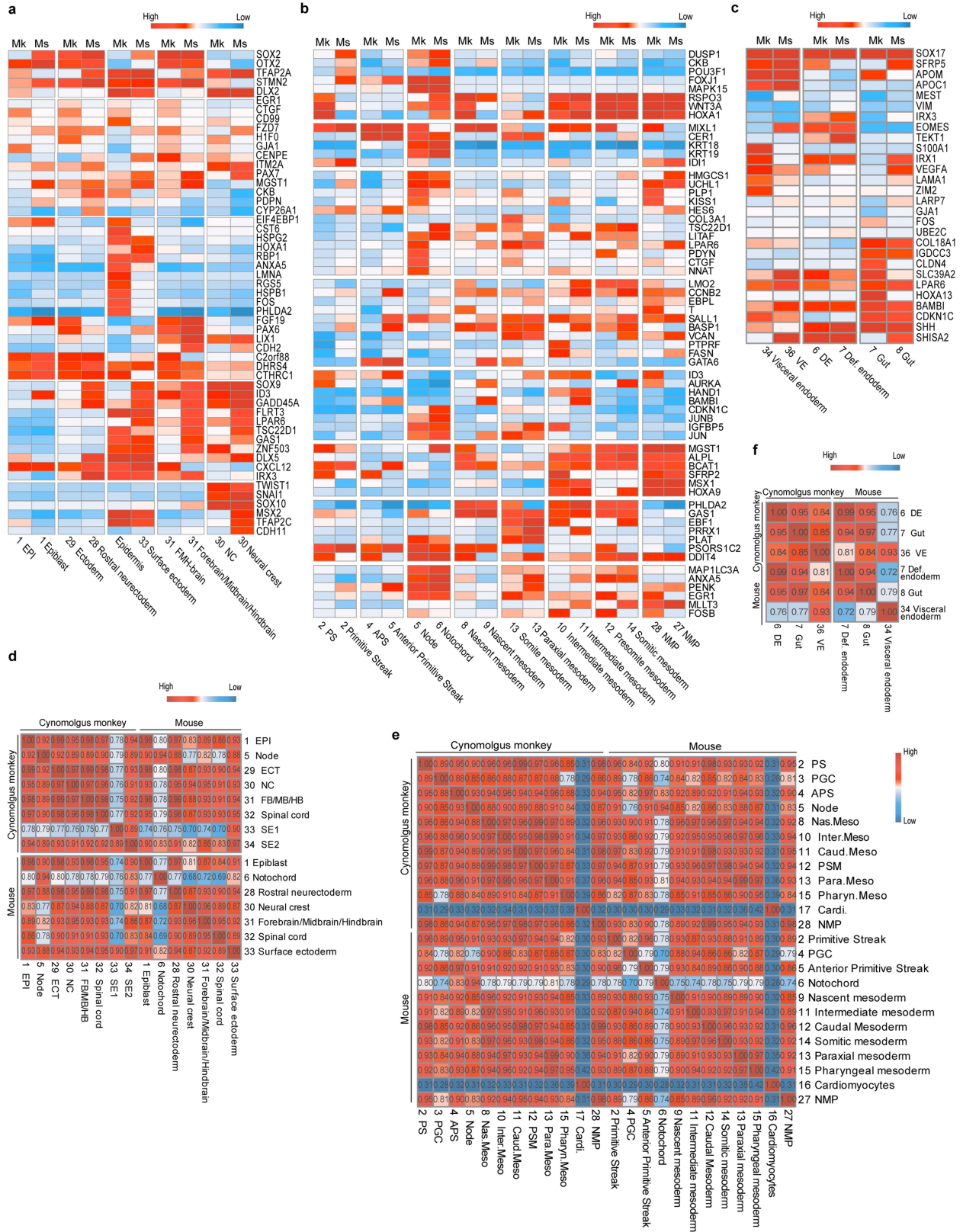
Extended Data Fig. 4 | The transcriptomic features of different neural cells.
a, Heatmap showing the representative DEGs in various neural cells from CS11 embryos. MHB, midbrain and hindbrain boundary; CHB, caudal hindbrain; NE, neural ectoderm (in unclosed neural tube); NP, neural progenitor (in closed neural tube); FP, floor plate; RP, roof plate. Color bar, normalized expression level of different genes in indicated cell types. **b**, UMAPs showing the expression pattern of marker genes along the A-P axis. Color bar in UMAPs, normalized

expression level of indicated genes from high (purple) to low (grey), here and after. **c**, Heatmap showing the representative TFs in different brain regions and spinal cord (SC) along the A-P axis. Color bar, activity of regulons, indicating the active (black, 1) or non-active (white, 0) of different TFs. **d**, UMAPs showing the expression pattern of marker genes along the D-V axis of closed or unclosed neural tube. **e**, Overview of the representative ligand-receptor interactions between node and different neural cells analyzed by Cellphone DB.



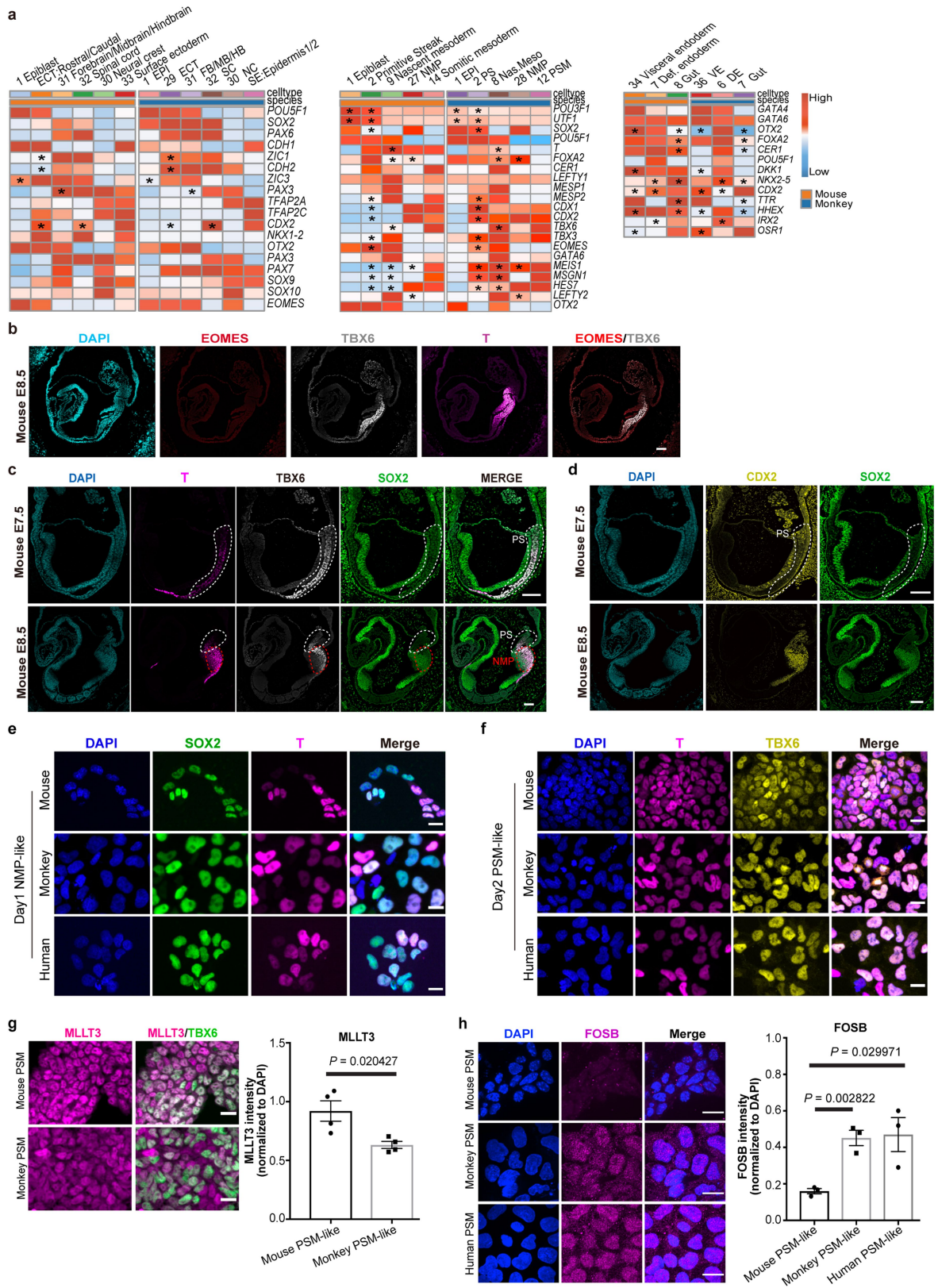
Extended Data Fig. 5 | The cross-species comparison on orthologues and developmental stages. **a**, The number of orthologues and paralogues between mouse, monkey and human. In total 26135 genes from macaque embryonic dataset (orange), 29452 genes from mouse embryonic dataset (blue), and 17273 genes from human embryonic dataset (green) (CS7 and CS12) are included in this analysis. **b**, The percentage of one-to-one orthologues in

each species. **c**, The expression of HES4 in different cell types in monkey embryos. **d-f**, Curve graphs showing the comparable developmental stage between monkey and mouse of different cell types from ectoderm (**d**), mesoderm and PGC (**e**), and endoderm (**f**). **g**, Schematic summarizing the cross-species transcriptomic and developmental stage comparison between mouse and monkey natural embryos and human embryos.



Extended Data Fig. 6 | The transcriptomic comparison of three germ layers between monkey and mouse. a-c. Heatmaps showing the DEGs indicated monkey (Mk) and mouse (Ms) cell types related with ectoderm (a), mesoderm (b) and endoderm (c) development. Color bars, normalized expression level of different genes from high (red) to low (blue) in indicated cell types. **d-f.** Heatmaps of the correlation coefficients among the indicated cells including those reported

by mouse dataset based on the cynomolgus monkey ontogenic gene levels, performed by Pearson Correlation Coefficient. Gene expression values of the integrated dataset were calculated by “cor()” in R, here and after. No other statistical analysis. Colors from blue to red indicate the low to high levels of Pearson Correlation Coefficient between each cell types, here and after.



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

Article

Extended Data Fig. 7 | The differences in gene expression pattern between monkey and mouse embryos. **a**, Heatmaps showing the dynamic expression of marker genes in the differentiation of different lineages, including EPI-PS-Nas, Meso-NMP-PSM (left), EPI-ECT-neural (middle), and VE-DE-Gut (right) cells. Black asterisks indicating genes (i.e., *FOXA2*, *TBX6*, *EOMES*, and *OTX2*) that show distinct expression patterns between mouse and monkey. Colors from blue to red indicate the low to high expression level of different genes in indicated cell types. **b**, IF results exhibiting the localization of $EOMES^{low}/TBX6^+/T^+$ cells on E8.5 mouse embryos. DAPI, marker for DNA, here and after. Scale bar, 100 μm . **c**, IF results exhibiting the localization of $T^+/TBX6^+/SOX2^+$ NMP on E7.5-E8.5 mouse embryos. White dashed lines highlight the PS region ($SOX2^+/T^+/TBX6^+$), here and after. Red dashed lines highlight the potential NMP region ($SOX2^+/T^+/TBX6^+$). Scale bars, 100 μm . **d**, IF results exhibiting the localization of $CDX2^{low}/SOX2^+$ cells on E7.5-E8.5 mouse embryos. Scale bars, 100 μm . All the IF experiments (**b-d**) were independently repeated in 3 samples. **e-f**, IF results

showing the lineage-specific markers ($SOX2$ and T for NMP, T and $TBX6$ for PSM) of NMP-like and PSM-like cells differentiated from pluripotent stem cells. Scale bars, 20 μm . Experiments were independently repeated 3 times. **g**, (left) IF validation of *MLLT3* captured by scRNA-seq in mouse and monkey PSM-like cells. Scale bars, 20 μm . (right) Bar graph indicating the quantification of *MLLT3* immunofluorescent intensity, which is normalized to DAPI. Data are shown as mean \pm s.e.m. ($n = 4$, biological replicates, experiments were independently repeated twice with similar results). Mouse PSM versus monkey PSM, the two-tailed P value = 0.020427. **h**, (left) IF validation of FOSB captured by scRNA-seq in mouse and monkey PSM-like cells. Scale bars, 20 μm . (right) Quantification of FOSB immunofluorescent intensity, which is normalized to DAPI. Data are shown as mean \pm s.e.m. ($n = 3$, biological replicates, experiments were independently repeated twice with similar results). The mouse PSM versus monkey PSM, two-tailed P value = 0.002822; the mouse PSM versus human PSM, two-tailed P value = 0.029971.

Article

Extended Data Fig. 8 | The transcriptomic comparison between monkey, human embryos, and human neuruloids. **a**, UMAP showing the integrated single-cell transcriptome datasets of our monkey CS8-11 embryos and human CS7 and CS12 embryos. **b**, UMAP indicating the 38 cell types from monkey embryonic cells. **c**, UMAP showing the 18 cell types of human CS7 embryo². **d**, UMAP showing the 19 major cell types of human CS12 embryo³. LPM, lateral plate mesoderm; IM, intermediate mesoderm. **e-f**, Heatmaps of the correlation coefficients among the indicated cells including those reported by human dataset based on the cynomolgus monkey ontogenic gene levels, performed by

Pearson Correlation Coefficient. No other statistical analysis. **g**, Comparison of CS11 monkey neural cells and cells of different human neuruloids with dorsal-ventral patterning⁴⁷⁻⁴⁹. The top row showing the annotated cell types from different datasets, and the bottom row highlighting the comparable cell types. **h**, Heatmap showing the expression of key genes of the SHH, TGF- β , and WNT signalling pathways enriched in monkey neural cells and human neuruloid cells. Colors from blue to red indicate the low to high expression level of different genes in indicated cell types.

Extended Data Fig. 10 | The transcriptomic comparison between monkey embryonic cells and human embryoid (gastruloids, neuruloids with dorsal-ventral patterning, and heart-forming organoid) cells. **a**, UMAPs showing the integrated single-cell transcriptome datasets with human gastruloids^{42,43} projected on monkey map. The top row showing the annotated cell types from different datasets, and the bottom row highlighting the comparable cell types, here and after. H1 hESC, human embryonic stem cell H1; hPGCLC, human primordial germ cell like cells; ExE-like, extra-embryonic mesenchymal-like cells. **b**, Heatmap of the correlation coefficients among the indicated cells including those reported by human dataset based on the cynomolgus monkey ontogenic gene levels, performed by Pearson Correlation Coefficient. No other statistical analysis. **c**, Heatmap showing the expression of specific marker genes in each annotated cell types and key genes of the TGF- β and WNT signalling pathways enriched in monkey embryonic cells and human gastruloid cells. Color bar, the normalized expression of different genes in indicated cell types, here and after. **d**, UMAPs showing comparison of

CS11 monkey neural cells and MiSTR⁴⁶ patterned human neuruloid cells. Cells of human neuruloid are compared with cells of different brain regions in monkey embryos. DiEn, diencephalon; VFB, visceral forebrain; VMB, visceral midbrain. **e**, Heatmap of the correlation coefficients among the indicated cells including those reported by human dataset based on the cynomolgus monkey gene levels, performed by Pearson Correlation Coefficient. No other statistical analysis. **f**, Heatmap showing the expression of key genes of the WNT signalling pathway enriched in monkey neural cells and human neuruloid cells. **g**, UMAPs showing comparison of monkey cardiogenesis-related cells and human heart-forming organoid (HFO)⁴⁵. AFE, anterior foregut endoderm. **h**, Heatmap of the correlation coefficients among the indicated cells including those reported by human dataset based on the cynomolgus monkey ontogenic gene levels. **i**, Heatmap showing the expression of key genes of the RA (retinoic acid), WNT and Insulin activity signalling pathways, respectively enriched in monkey cardiogenesis-related cells and human heart-forming organoid cells.

Reporting Summary

Nature Portfolio 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 Portfolio policies, see our [Editorial Policies](#) and the [Editorial Policy Checklist](#).

Statistics

For all statistical analyses, confirm that the following items are present in the 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
- A statement on 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 statistical parameters 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

Our web collection on [statistics for biologists](#) contains articles on many of the points above.

Software and code

Policy information about [availability of computer code](#)

Data collection Single-cell libraries were constructed using the Single cell 3' Library & Gel Bead Kit v3 of 10X Genomics, and were sequenced on an Illumina HiSeq X Ten platform.

Data analysis We used existing published sequence analysis packages and methods, as detailed in the Methods, including "CellRanger3.1.0", "Seurat (v4.0.0)", "Scrublet(0.2.1)", "clusterProfiler(v4.0.5)", "monocle3 (v1.0.0)", "SeuratDisk (v0.0.0.9013)", "Scanpy(v1.8.2)", "SeuratWrappers(0.3.0)", "scVelo(0.2.4)", "pySCENIC(v0.10.0)", "CellPhoneDB(2.1.7)", "biomaRt (v2.46.3)", "scmap(v1.20.0)", Matlab (R2019a), and GraphPad Prism (7.0). Custom MATLAB code for image processing is available in a public repository (<https://github.com/ecamacho90/BiologicalImageProcessing>, <https://doi.org/10.5281/zenodo.7367663>). The other codes are available upon request.

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 and reviewers. We strongly encourage code deposition in a community repository (e.g. GitHub). See the Nature Portfolio [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 description of any restrictions on data availability
- For clinical datasets or third party data, please ensure that the statement adheres to our [policy](#)

Raw data and processed data were uploaded to the NCBI Gene Expression Omnibus (GEO) database with the accession number GSE193007.

Datasets used as references include:
 mouse E6.5-E8.6 data: E-MTAB-6967;
 human CS7 data: E-MTAB-9388;
 human CS12 embryo data: GSE157329;
 human gastruloids: GSE144897 and GSE169074;
 human MiSRT neuruloid data: GSE135399;
 human 2D micropatterned neuruloid: GSE118682;
 light induced D-V patterned neuruloid: GSE163505;
 chip-based neuruloid: GSE173492;
 heart forming organoid: GSE150202;
 somitoid: E-MTAB-11334.

Field-specific reporting

Please select the one below that is 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/documents/nr-reporting-summary-flat.pdf](https://www.nature.com/documents/nr-reporting-summary-flat.pdf)

Life sciences study design

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

Sample size	The sample size of this study was determined based on availability of highly regulated primate embryo samples. In compliance with the 3R guidelines, we reduced the number of used animals to minimum and obtained pregnant females uteri at E20 (n=2), E22 (n=1), E23 (n=2), E26 (n=1), and E29 (n=1), which allowed us to obtain high coverage transcriptome for each cell type, and perform confident downstream analyses.
Data exclusions	First, single cells with a number of detected genes (nFeature_RNA) above 500 and detected transcripts (nCount_RNA) above 1000 were retained to exclude the apoptotic or dead cells. Then, the doublet or multiplet cells were figured out with the Scrublet, according to the recommended multiplet rate reference table from 10X Genomics (Wolock et al., 2019). Next, the Seurat objects of different samples were created independently with the expression matrix and metadata containing cell barcodes, cell status, and assignment information identified by Souporecell and cell multiplet information inferred by Scrublet, then these Seurat3 objects were merged.
Replication	Sequenced samples from two independent embryos of the same stages showed similar gene expression patterns. Since developmental stage of embryo in utero is uncontrollable, though we collected monkey embryos by calculating the day post fertilization and combining with b-ultrasound, CS10 embryo was not successfully collected, so not performed. The IF experiments on mouse embryos were repeated in three independent biological samples. The stem cells experiments were independently repeated at least three times. All attempts of experiment replication were successful.
Randomization	Samples were not allocated into randomized groups. Randomization was not relevant to the study. All embryo samples were analyzed individually.
Blinding	Blinding of the investigators was not possible due to study design and was not relevant to the study. It was not possible to blind the experiments during neither embryo collection nor single cell collection. We performed lineage assignment in an unbiased way, in detail we assigned samples to lineages based on their gene expression profile and then validated our findings by their localization within the embryo.

Reporting for specific materials, systems and methods

We require information from authors about some types of materials, experimental systems and methods used in many studies. Here, indicate whether each material, system or method listed is relevant to your study. If you are not sure if a list item applies to your research, read the appropriate section before selecting a response.

Materials & experimental systems

n/a	Involved in the study
<input type="checkbox"/>	<input checked="" type="checkbox"/> Antibodies
<input type="checkbox"/>	<input checked="" type="checkbox"/> Eukaryotic cell lines
<input checked="" type="checkbox"/>	<input type="checkbox"/> Palaeontology and archaeology
<input type="checkbox"/>	<input checked="" type="checkbox"/> Animals and other organisms
<input checked="" type="checkbox"/>	<input type="checkbox"/> Human research participants
<input checked="" type="checkbox"/>	<input type="checkbox"/> Clinical data
<input checked="" type="checkbox"/>	<input type="checkbox"/> Dual use research of concern

Methods

n/a	Involved in the study
<input checked="" type="checkbox"/>	<input type="checkbox"/> ChIP-seq
<input checked="" type="checkbox"/>	<input type="checkbox"/> Flow cytometry
<input checked="" type="checkbox"/>	<input type="checkbox"/> MRI-based neuroimaging

Antibodies used

Primary antibodies used for IF staining of embryos and embryonic stem cells:

OCT4 (mouse, 1:100, Santa Cruz Biotechnology, sc5279, clone C-10)
 SOX2 (mouse, 1:100, Santa Cruz Biotechnology, sc365823, clone E-4)
 E-Cadherin (goat, 1:200, R&D Systems, AF748)
 OTX2 (goat, 1:100, R&D Systems, AF1979)
 Slug (rabbit, 1:100, Cell Signaling Technology, 9585S, clone C19G7)
 FOXA2 (rabbit, 1:200, Abcam, ab108422, clone EPR4466)
 T (rabbit, 1:100, Cell Signaling Technology, 81694, clone D2Z3J)
 AF9, MLLT3 (rabbit, 1:200, Abcam, ab154492)
 Brachyury, T (goat, 1:200, R&D Systems, AF2085)
 FOSB (goat, 1:200, R&D Systems, AF2214)
 TBX6 (goat, 1:200, R&D Systems, AF4744)
 YAP1 (mouse, 1:200, Abnova, H00010413-M01, clone 2F12)
 EOMES (rabbit, 1:100, Abcam, ab23345)
 CDX2 (mouse, 1:100, BioGenex, MU392A-5UC, clone CDX2-88)
 GATA6 (goat, 1:100, R&D Systems, AF1700)
 PDGF Receptor α (rabbit, 1:1, Cell Signaling Technology, 3174T, clone D1E1E)

Secondary antibodies used for IF staining of embryos and embryonic stem cells:

Alexa Fluor 488 Donkey anti-mouse (1:200, Thermo Fisher, A21202)
 Alexa Fluor 647 Goat anti-mouse (1:200, Thermo Fisher, A31571)
 Alexa Fluor 568 Donkey anti-rabbit (1:200, Thermo Fisher, A10042)
 Alexa Fluor 647 Donkey anti-rabbit (1:200, Thermo Fisher, A31573)
 Alexa Fluor 488 Donkey anti-goat (1:200, Thermo Fisher, A11055)
 Alexa Fluor 488 Donkey anti-rat (1:200, Thermo Fisher, A21208)
 Alexa Fluor 568 Donkey anti-sheep (1:200, Thermo Fisher, A21099)

Validation

All the above primary antibodies were validated in this work, the subcellular localization of the analyzed proteins has been previously reported. We used this to validate the specificity of the antibody in this study.

Primary antibodies used for IF staining of embryos and embryonic stem cells:

SOX2 (mouse, 1:100, Santa Cruz Biotechnology, sc365823, clone E-4): Correctly stained on a E22 monkey embryo, embryonic day (E)7.5-E8.5 mouse embryos for ectoderm as expected, and previously reported (PeerJ (2019), doi: 10.7717/peerj.5840 in porcine blastocyst).

SOX2 (rabbit, 1:200, Cell Signaling Technology, 5024S, clone D6D9): Correctly stained on human, monkey, and mouse ESC-derived neuroesodermal progenitor (NMP)-like cells for in vitro differentiation as expected, and previously reported (PLoS ONE (2014), doi: 10.1371/journal.pone.0106694 in human fetal normal brain sections and glioblastoma patient derived cell lines).

E-Cadherin (goat, 1:200, R&D Systems, AF748): Correctly stained a E22 monkey embryo for epithelial cells as expected, and previously reported (Developmental Cell (2017), doi: 10.1016/j.devcel.2017.05.004 in mouse ESCs).

OTX2 (goat, 1:100, R&D Systems, AF1979): Correctly stained a E22 monkey embryo and E7.5 mouse embryo for visceral endoderm, definitive endoderm, neural ectoderm and some mesoderm, as expected, and previously reported (Development (2018), doi: 10.1242/dev.167833 in monkey and human late blastocysts).

SLUG (rabbit, 1:100, Cell Signaling Technology, 9585S, clone C19G7): Correctly stained a E22 monkey embryo for epithelial-mesenchymal transition as expected, and previously reported (Scientific Reports (2018), doi: 10.1038/s41598-018-30939-z in human ESCs).

FOXA2 (rabbit, 1:200, Abcam, ab108422, clone EPR4466): Correctly stained a E22 monkey embryo and E7.5 mouse embryo for endoderm and some mesoderm, as expected, and previously reported (BioRxiv (2022), doi: 10.1101/2022.03.07.483315 in E7.5 mouse embryos).

T (rabbit, 1:200, Cell Signaling Technology, 81694, clone D2Z3J): Correctly stained a E22 monkey embryo and E7.5 mouse embryo, and E8.5 mouse embryos for primitive streak (PS) and some mesoderm, as expected, and previously reported (BioRxiv (2022), doi: 10.1101/2022.03.07.483315 in E7.5 mouse embryos).

Correctly stained on human, monkey, and mouse ESC-derived NMP-like cells and presomitic mesoderm (PSM)-like cells for in vitro differentiation as expected, and previously reported (BioRxiv Preprint (2022), doi: 10.1101/2022.03.07.483315 in E7.5 mouse embryos).

Brachyury (T) (goat, 1:300, R&D Systems, AF2085-SP): Correctly stained on human, monkey, and mouse ESC-derived NMP-like cells and presomitic mesoderm (PSM)-like cells for in vitro differentiation as expected, and previously reported (eLife (2018), doi: 10.7554/eLife.38279 in human ESCs).

AF9, MLLT3 (rabbit, 1:200, Abcam, ab154492): Correctly stained on monkey and mouse ESC-derived PSM-like cells for in vitro differentiation as expected, and previously reported (Transcription (2017), doi: 10.1080/21541264.2017.1295831 in HEK-293T cells).

FOSB (goat, 1:200, R&D Systems, AF2214): Correctly stained on human, monkey and mouse PSM-like cells for in vitro differentiation as expected, and previously reported (Molecular Systems Biology (2021), doi: 10.15252/msb.202010125 in U937 cells for western blot).

TBX6 (goat, 1:200, R&D Systems, AF4744): Correctly stained a E22 monkey embryo and E7.5 mouse embryo, and E8.5 mouse embryos for PS, NMP, and some mesoderm, as expected, and previously reported (Development (2018), doi: 10.1242/dev.164319 in mouse embryos between E9.5-10.5).

Correctly stained human, monkey and mouse ESC-derived PSM-like cells for evaluation of PSM-like cell differentiation as expected, and previously reported (Cell Reports (2019), doi: 10.1016/j.celrep.2019.07.090 in human ESCs-differentiated PSM cells).

YAP1 (rabbit, 1:200, Abcam, ab52771, clone EP1674Y): Correctly stained human, monkey and mouse ESC-derived PSM-like cells for Hippo-YAP activities, as expected, and previously reported (Oncology Reports (2019), doi: 10.3892/or.2019.7065 in human glioblastoma cells U-372 MG cells).

EOMES (rabbit, 1:100, Abcam, ab23345): Correctly stained a E22 monkey embryo, a E7.5 mouse embryo, and E8.5 mouse embryos for PS as expected, and previously reported (PLoS Biology (2020), doi: 10.1371/journal.pbio.3000705 in organoid models consisting of human ESCs and induced pluripotent stem cells).

CDX2 (mouse, 1:100, BioGenex, MU392A-5UC, clone CDX2-88): Correctly stained a E22 monkey embryo, E7.5-E8.5 mouse embryos for PS as expected, and previously reported (Developmental Cell (2017), doi: 10.1016/j.devcel.2017.05.004 in E3.5 mouse blastocysts).

GATA6 (goat, 1:100, R&D Systems, AF1700): Correctly stained a E22 monkey embryo for extraembryonic mesenchymal cells, endoderm and some mesoderm, as expected, and previously reported (Development (2018), doi: 10.1242/dev.167833 in monkey and human late blastocysts).

PDGF Receptor α (rabbit, 1:1, Cell Signaling Technology, 3174T, clone D1E1E): Correctly stained a E22 monkey embryo for mesoderm as expected, and previously reported (Disease Models & Mechanisms (2013), doi: 10.1242/dmm.013748 in E13.5 mouse embryo head skin sections).

Eukaryotic cell lines

Policy information about [cell lines](#)

Cell line source(s)	Human embryonic stem cell line H9 (WA09) was obtained from WiCell and authenticated by short tandem repeat (STR) profiling. Mouse epiblast stem cells (EpiSCs), rhesus macaque ES cells (ESCs) were generated and described in previous study (Wu, J. et al., Cell, 2017; Wu, J. et al., Nature, 2015). HES7 promoter-luciferase reporter line was generated by following a protocol described in the previous study (Matsuda, M. et al., Nature, 2020).
Authentication	Human embryonic stem cells (H9) were obtained from WiCell and authenticated by short tandem repeat (STR) profiling. Monkey and mouse cell lines were validated by IF staining, western blots using specific antibodies and genomic PCR and qRT-PCR using species-specific primers.
Mycoplasma contamination	All cell lines are negative for mycoplasma contamination.
Commonly misidentified lines (See ICLAC register)	No commonly misidentified cell lines were used in this study.

Animals and other organisms

Policy information about [studies involving animals](#); [ARRIVE guidelines](#) recommended for reporting animal research

Laboratory animals	Male and female common cynomolgus monkeys (<i>Macaca fascicularis</i>) with an age range of 6-8 years old were used in this study. Male and female C57BL/6 mouse (<i>Mus musculus</i>) around 8 weeks old were used in this study.
Wild animals	This study did not use any wild animals.
Field-collected samples	This study did not use any field-collected samples.
Ethics oversight	This study was conducted in accordance with the "Principles for the Ethical Treatment of Non-Human Primates" issued by Institute of Zoology, Chinese Academy of Sciences (IOZ, CAS), and was approved in advance by the Institutional Animal Care and Use Committee of the IOZ, CAS (Appl.No: IOZ-EU-20191113 for all monkey experiments, Appl.No: IOZ-IACUC-2021-037 for all mouse experiments). Both followed relevant guidelines and regulations.

Note that full information on the approval of the study protocol must also be provided in the manuscript.