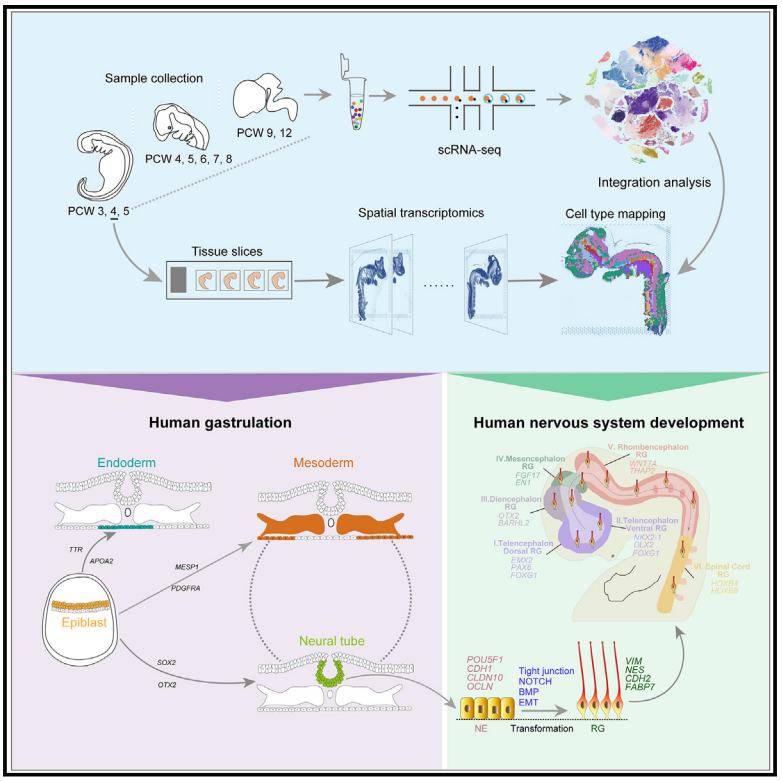
Cell Stem Cell

Resource

# The single-cell and spatial transcriptional landscape of human gastrulation and early brain development

## Graphical abstract




## Highlights

- The spatial patterning of neural tube cells during human gastrulation
- NE to RG transformation is determined by several essential signaling pathways
- RG subtypes exhibit spatial location distribution with distinctive gene expression
- Human-specific features are presented during early nervous system development


## Authors

Bo Zeng, Zeyuan Liu, Yufeng Lu, ..., Qian Wu, Yongjun Wang, Xiaoqun Wang

## Correspondence

oscar.marin@kcl.ac.uk (O.M.), qianwu@bnu.edu.cn (Q.W.), yongjunwang@ncrcnd.org.cn (Y.W.), xiaoqunwang@bnu.edu.cn (X.W.)



## In brief

Zeng et al. illustrate the transcriptional profiles and spatial organization of different cell types during human gastrulation and early nervous system development. They reveal the spatial patterning of neural tube cells, molecular programs driving NE to RG transformation, diversity of RG subtypes, and species differences of early nervous system development.


Zeng et al., 2023, Cell Stem Cell *30*, 851–866
June 1, 2023 
https://doi.org/10.1016/j.stem.2023.04.016

CellPress

Cell Stem Cell

CellPress
OPEN ACCESS

Resource

# The single-cell and spatial transcriptional landscape of human gastrulation and early brain development

Bo Zeng,[1,14,16] Zeyuan Liu,[1,14,16] Yufeng Lu,[2,16] Suijuan Zhong,[1,14,16] Shenyue Qin,[7,8,16] Luwei Huang,[2,9] Yang Zeng,[3] Zixiao Li,[4,5,6] Hao Dong,[2,9] Yingchao Shi,[2,13] Jialei Yang,[4,5] Yalun Dai,[4,5] Qiang Ma,[2,9] Le Sun,[10] Lihong Bian,[11] Dan Han,[12] Youqiao Chen,[1] Xin Qiu,[4,5] Wei Wang,[2] Oscar Marín,[7,8,*] Qian Wu,[1,14,*] Yongjun Wang,[4,5,6,*] and Xiaoqun Wang[1,2,10,13,14,15,17,*]

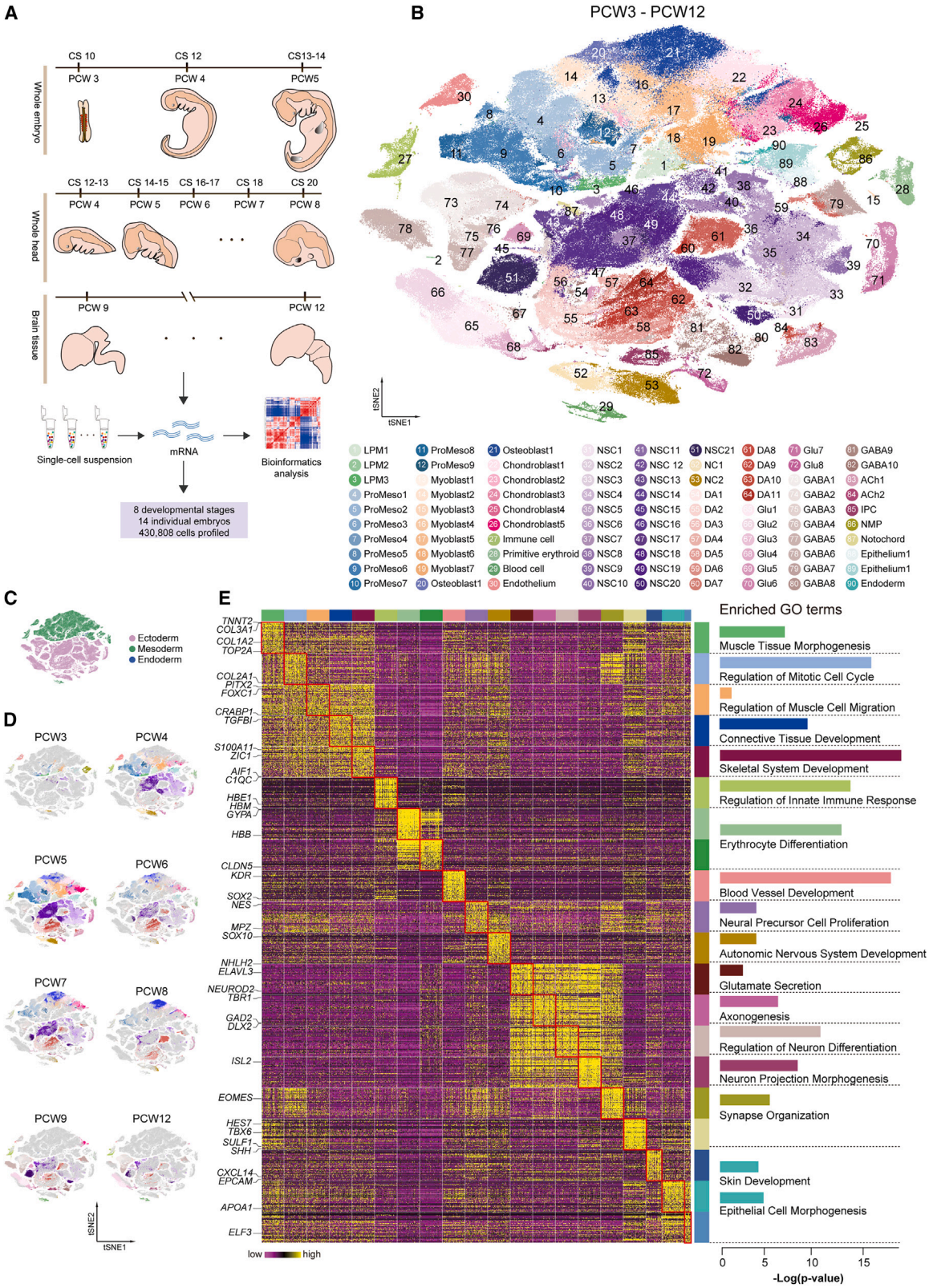
[1]State Key Laboratory of Cognitive Neuroscience and Learning, IDG/McGovern Institute for Brain Research, Beijing Normal University, Beijing 100875, China
[2]State Key Laboratory of Brain and Cognitive Science, Institute of Biophysics, Chinese Academy of Sciences, Beijing 100101, China
[3]State Key Laboratory of Experimental Hematology, Fifth Medical Center of Chinese PLA General Hospital, Beijing 100071, China
[4]China National Clinical Research Center for Neurological Diseases, Beijing 100070, China
[5]Department of Neurology, Beijing Tiantan Hospital, Capital Medical University, Beijing 100070, China
[6]Beijing Advanced Innovation Center for Big Data-Based Precision Medicine, Beihang University & Capital Medical University, Beijing 100069, China
[7]Centre for Developmental Neurobiology, Institute of Psychiatry, Psychology and Neuroscience, King's College London, London SE1 1UL, UK
[8]MRC Centre for Neurodevelopmental Disorders, King's College London, London SE1 1UL, UK
[9]University of Chinese Academy of Sciences, Beijing 100049, China
[10]Beijing Institute of Brain Disorders, Capital Medical University, Beijing 100069, China
[11]Department of Gynecology, Fifth Medical Center of Chinese PLA General Hospital, Beijing 100071, China
[12]Department of Obstetrics & Gynecology, Beijing Tiantan Hospital, Capital Medical University, Beijing 100070, China
[13]Guangdong Institute of Intelligence Science and Technology, Guangdong 519031, China
[14]Changping Laboratory, Beijing 102206, China
[15]New Cornerstone Science Laboratory, Beijing Normal University, Beijing 100875, China
[16]These authors contributed equally
[17]Lead contact
*Correspondence: oscar.marin@kcl.ac.uk (O.M.), qianwu@bnu.edu.cn (Q.W.), yongjunwang@ncrcnd.org.cn (Y.W.), xiaoqunwang@bnu.edu.cn (X.W.)
https://doi.org/10.1016/j.stem.2023.04.016

## SUMMARY

The emergence of the three germ layers and the lineage-specific precursor cells orchestrating organogenesis represent fundamental milestones during early embryonic development. We analyzed the transcriptional profiles of over 400,000 cells from 14 human samples collected from post-conceptional weeks (PCW) 3 to 12 to delineate the dynamic molecular and cellular landscape of early gastrulation and nervous system development. We described the diversification of cell types, the spatial patterning of neural tube cells, and the signaling pathways likely involved in transforming epiblast cells into neuroepithelial cells and then into radial glia. We resolved 24 clusters of radial glial cells along the neural tube and outlined differentiation trajectories for the main classes of neurons. Lastly, we identified conserved and distinctive features across species by comparing early embryonic single-cell transcriptomic profiles between humans and mice. This comprehensive atlas sheds light on the molecular mechanisms underlying gastrulation and early human brain development.

## INTRODUCTION

During early mammalian development, embryos start from a single-cell zygote and progress to a blastocyst, followed by gastrulation. During this period, embryonic pluripotent stem cells diversify into lineage-specific precursors and drive organogenesis. With the utilization of single-cell RNA sequencing (scRNA-seq) technology, recent studies have described the molecular features of gastrulation and organogenesis during early mouse development.[1–3] Under the orchestration of defined gene expression networks, embryonic stem cells progressively lose totipotency, transform into a pluripotent state, and commit to specific lineages during early organogenesis.[1–4]




*(legend on next page)*

Our understanding of human early embryonic development is much more limited. Human organogenesis begins at Carnegie stage (CS) 9, at approximately embryonic day (E) 20. Global transcriptome profiles of human embryos covering early development have been described,[5] followed by two further studies dissecting the cellular and molecular mechanisms underlying human and macaque early gastrulation up to 14 and 20 days post-fertilization, respectively.[6,7] Recent single-cell transcriptomic analyses have begun to illustrate the development of different human tissues and organs.[8–13] However, a comprehensive understanding of the developmental trajectories, spatial gene expression, and essential regulatory networks involved in early human embryo development and organogenesis after gastrulation at single-cell resolution is lacking.

The brain is the most complex organ in the human body. It comprises billions of neurons organized in many interconnected structures, and these regional differences appear as early as the neural tube closure.[14] Neurulation is a complex morphogenetic process requiring precise coordination of cellular events and molecular regulation.[15,16] Many epiblast cells pass through the primitive streak during gastrulation to form the mesoderm and endoderm. The remaining epiblast cells are converted into neuroepithelial (NE) cells that form the developing neural tube.[17] NE cells divide asymmetrically between PCW3 and 4 and undergo a series of molecular and morphological changes that transform them into radial glial (RG) cells, the progenitors underlying neural cell production.[18,19] With the tremendous increase in neuron number, many brain structures substantially expand in humans compared with other mammals,[20] and exploring the diversity of neural stem cells (NSCs) and the regulation of cell fate determination in early human brain development has become critical for understanding how different cell types across brain regions are generated and expanded.

Here, we analyzed 14 samples collected from human embryos and fetuses ranging from PCW3 to 12 to delineate a single-cell landscape of early human gastrulation and neurogenesis. We examined the transcriptional profiles of 430,808 cells and focused on the early dynamics of nervous system development. We resolved the spatial organization of neural tube cells and described the molecular mechanisms through which bipotent neuromesodermal progenitor (NMP) cells contribute to neural and mesodermal tissues. We also identified signaling pathways regulating the epiblast-NE-RG transitions and described the diversification of RG and their spatial localization in the early nervous system. A comparison of early embryonic transcriptomic profiles between humans (PCW3–4) and mice (E8.5–10.5) revealed broadly conserved features between the species, although a human-dominant neural cluster was also identified.

## RESULTS

### A single-cell atlas of early human embryogenesis

To gain a global perspective of cellular diversification during early human embryogenesis, we collected 14 human samples between PCW3 and 12, including four whole embryos (PCW3–5), seven heads (PCW4–8), and three brains (PCW9 and 12), prepared single-cell suspensions, and processed them for scRNA-seq using a droplet-based platform (Figure 1A). The age of the embryonic samples corresponds to CS 10, 12 to 16, 18, and 20 (Figure 1A; Table S1). A total of 430,808 cells were analyzed following quality control processing (Figure S1A). We used Louvain clustering and t-distributed stochastic neighbor embedding (t-SNE) to visualize 90 clusters with cells from the three germ layers and across multiple developmental stages[21,22] with high correlation similarities from samples of the same stages and the same types of tissue (Figures 1B–1D, S1B, and S1C; Table S1). These clusters can be categorized into 20 major groups (Figure S1D) based on the expression of typical markers and gene ontology (GO) analysis of differentially expressed genes (DEGs) (Figures 1E and S1E; Table S1).

We observed that many cell clusters derived from the ectoderm correspond to the nervous system, consistent with the predominant sampling of heads and brains (Figure 1C). These clusters include neural crest (NC) cells, NSCs, intermediate progenitor cells (IPCs), and glutamatergic (Glu), GABAergic (GABA), dopaminergic (DA), and cholinergic (ACh) neurons (Figures 1B and S1D). The expression of *SOX10* and *MPZ* in NC cells (clusters 52 and 53) suggests that these cells may give rise to the peripheral nervous system (PNS) (Figures 1B and S1E). In contrast, the high expression of *POU4F1* in the Glu6 and Glu7 groups (clusters 70 and 71) indicates that these cells might be sensory neurons starting to emerge around PCW4. In addition, ACh1 cells (cluster 83) expressing *ISL2*$^{+}$ as early as PCW4 likely represent spinal cord motor neurons (Figures 1B and S1E).

We also identified several large groups among cells derived from the mesoderm (Figure 1C). For example, myoblasts expressed *MYF5* and *PITX2*, while osteoblasts expressed *RUNX2* (Figures 1B and S1E). In addition, we observed primitive erythroid cells expressing high levels of *GYPA* as early as PCW3, which revealed that the blood system developed very early (Figures 1B–1D and S1E). In contrast to the abundance of ectodermal and mesodermal cells in our dataset, we only captured a few endodermal cells (cluster 90). These cells had a transcriptomic profile similar to epithelial cells derived from the surface ectoderm (clusters 88 and 89). NMPs (cluster 86), a transient cell population that contributes to the development of the spinal cord and the adjacent paraxial mesoderm,[23] were found only at PCW3 (Figures 1B–1D and S1E).

**Figure 1. Survey of human early embryo development at single-cell resolution**

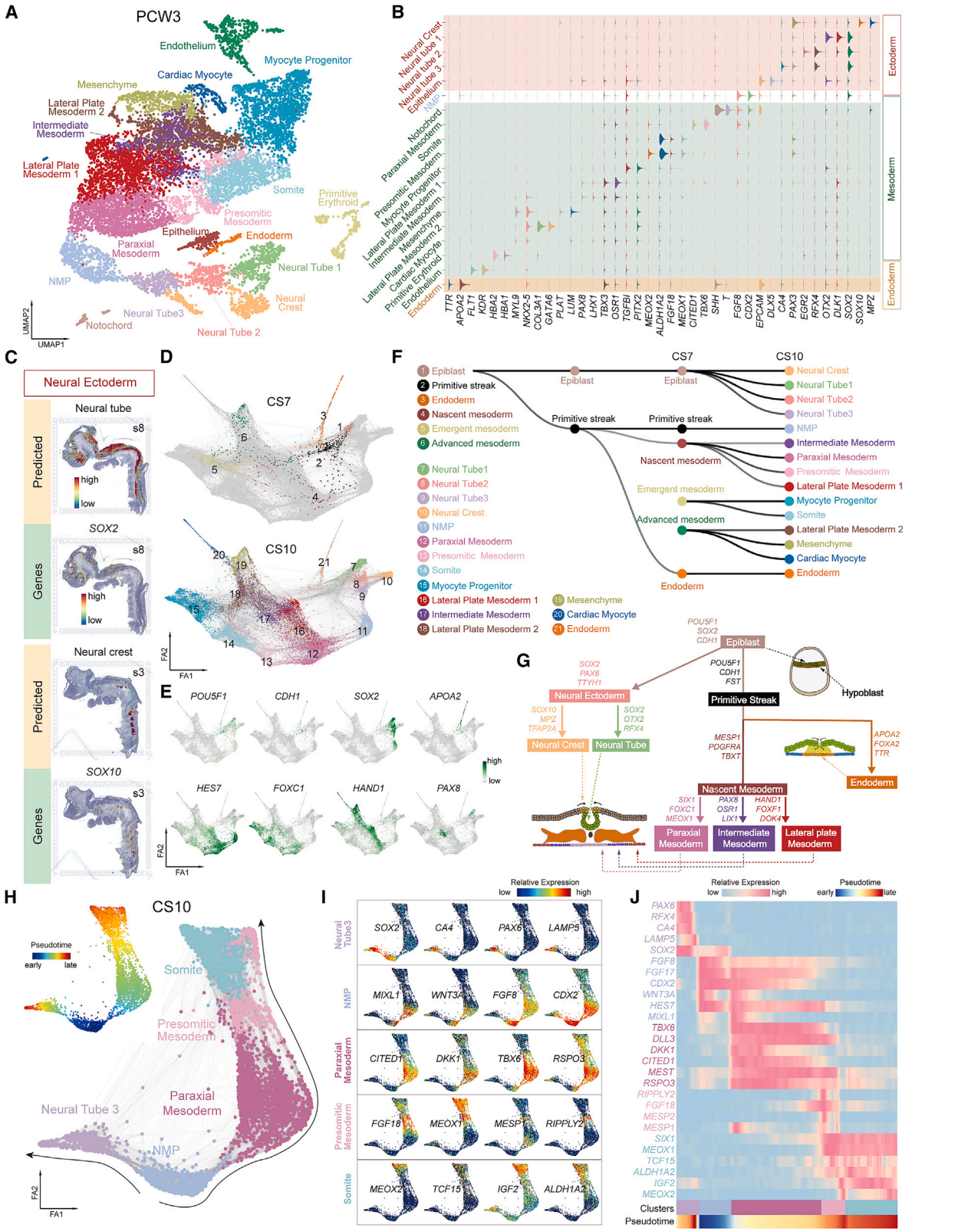
(A) Schematic diagram of sample collection and the scRNA-seq experimental workflow.

(B) t-SNE visualization of all 430,808 human embryo cells color-coded by 90 clusters. LPM, lateral plate mesoderm; ProMeso, proliferating mesoderm; NC, neural crest cells; NSCs, neural stem cells; IPC, intermediate progenitor cells; Glu, glutamatergic neurons; GABA, GABAergic neurons; DA, dopaminergic neurons; ACh, cholinergic neurons; NMPs, neuromesodermal progenitors.

(C) The highlights of the three germ-layer-derived cells on the t-SNE layout.

(D) Cells from each developmental stage are highlighted and color-coded by clusters.

(E) Heatmap showing the mean expression of marker genes of the 20 major groups. Enriched gene ontology (GO) terms for the major gene groups (red boxes) are shown vertically on the right, and the bar plot shows the negative logarithm p values.




*(legend on next page)*

### Cell-type diversity during gastrulation

The third-week post-fertilization is critical for human embryogenesis. During this stage, gastrulation leads to the establishment of the three germ layers, and subsequently, the process of primary neurulation transforms the neural plate into the neural tube.[15,24] To investigate cell diversity during this period, we analyzed a total of 12,323 cells from two embryos of different sex collected during PCW3 (Figure S2A), which corresponds to CS10.[24,25] Unbiased clustering led to the identification of 19 groups of cells across the three germ layers with distinct patterns of gene expression (Figures 2A and 2B; Table S2). Among ectodermal derivatives, three clusters of neural tube cells with distinctive gene expression were identified at this stage, along with NC cells characterized by high expression of *MPZ* and *SOX10* and epithelial cells displaying robust *EPCAM* expression (Figure 2B). Mesodermal cells were segregated into paraxial, intermediate, and lateral plate mesoderm, along with other groups containing more differentiated cells corresponding to the presomitic mesoderm, somites, myocyte progenitors, mesenchyme, endothelial cells, and primitive erythroid progenitors (Figure 2A and 2B). We also identified a cluster of NMPs sharing some features of both ectoderm and mesoderm, as well as a cluster of notochordal cells expressing both *SHH* and *T*. Finally, endodermal cells were distinguished from other germ layers by the specific expression of *TTR* and *APOA2* (Figure 2B).

To map the distribution of different cellular populations in the early human embryo, we performed spatial transcriptomic experiments at PCW4 (Figure S2B). We observed that cluster identities and marker gene expression assigned through the scRNA-seq analysis matched their known anatomical locations (Figures 2C, S2C, and S2D). For example, the predicted neural tube clusters and their marker *SOX2* mapped to the brain and spinal cord, whereas NC cells and their marker *SOX10* showed the highest prediction score adjacent to the somite (Figure 2C). These experiments revealed the precise spatial organization of the different cell groups identified by scRNA-seq during human embryo gastrulation.

Identifying the developmental trajectories followed by different cell types in the human embryo during gastrulation is a significant challenge due to the difficulty of accessing samples at this early age. To circumvent this problem, we integrated our PCW3 (CS10) dataset with a published dataset from a CS7 human embryo, which corresponds to an early stage of gastrulation[26,27] (Figures 2D, 2E, S2E, and S2F). We then used an automated procedure to define cell states and cell-state relationships inferred based on transcriptional similarities.[28] The resulting directed acyclic graph inferred developmental trajectories from the epiblast to the three germ layers (Figures 2F and S2G; Table S2). For instance, the analysis predicted that neuroectoderm cells from CS10 embryos derive from epiblast cells from the CS7 embryo. In contrast, the endoderm and early mesoderm cells found at CS10 (paraxial, intermediate, and lateral plate mesoderm) are pseudo-descendants of the primitive streak induced from the epiblast. We also observed that cells in the emergent mesoderm at CS7 are pseudo-ancestors of the myocyte progenitors and somite cells found at CS10, whereas cells from the advanced mesoderm at CS7 are pseudo-ancestors of the mesenchyme and cardiac myocyte in CS10 embryos. These observations are broadly consistent with our current knowledge of human gastrulation.

We leveraged the inferred developmental trajectories to identify essential genes regulating cell-state transitions during gastrulation. Many of the identified genes are consistent with previously published studies. For example, we observed that the epiblast and the epiblast-induced primitive streak cells express *POU5F1* (also known as *OCT4*) and *CDH1*,[26] whereas cells from the neural ectoderm highly express *SOX2* and *PAX6*. In contrast, the paraxial, intermediate, and lateral plate mesoderm derived from the nascent mesoderm are characterized by the expression of *FOXC1*,[29] *PAX8*,[30] and *HAND1*,[31] respectively (Figures 2E, 2G, and S2F).

We also examined the contributions of NMP cells to the posterior neural tube and somite formation. NMPs are considered bipotent progenitors.[23] They express the neural marker *SOX2* and the mesodermal marker *TBX6* and are positioned between the paraxial mesoderm and neural tube 3 cells in the Uniform Manifold Approximation and Projection (UMAP) (Figures 2A and 2B), highlighting their lineage relationships. To further characterize how NMPs contribute to neural tube and somite formation, we embedded NMPs, neural tube 3 cells, and cells from the paraxial mesoderm, presomitic mesoderm, and somites into a force-directed graph layout (Figure 2H). NMPs were at the start site in the pseudotime trajectory, with cells from the neural tube and somites at either end. This analysis allowed the identification of DEGs between the cell clusters and along the pseudotime trajectories (Figures 2I and 2J; Table S2). We observed that

**Figure 2. Single-cell and spatial transcriptomics map of human gastrulation**

(A) UMAP visualization of the PCW3 cells. Cells were divided into 19 clusters indicated by different colors.

(B) Violin plot showing the marker gene expression profiles of different clusters, grouped by three germ layers, NMP, and notochord. Color shadows depict the three germ layers.

(C) Predicted spatial distributions of the neural tube and neural crest cells and the spatial expression of their corresponding marker genes *SOX2* and *SOX10*.

(D) Integration and the ForceAtlas2 (FA) layout embedding of cells from the CS7 gastrula[26] and CS10 (PCW3) dataset. Corresponding cluster identities are annotated in (F).

(E) Feature plots showing selected marker gene expression of cell types from (D).

(F) Directed acyclic graph showing inferred relationships between the epiblast-derived cell states from CS7 to CS10. The edges between clusters within CS7 and CS10 were also included. All edges with weights above 0.4 are shown in grayscale.

(G) Schematic representation of the gastrulation development tree from epiblasts. Representative marker genes of each cell type are noted next to the arrows.

(H) ForceAtlas2 layout showing the relationship between NMP, neural tube 3, and somite-related cells colored by cluster and pseudotime values. The arrows indicate the predicted developmental trajectories from NMP cells.

(I) Feature plots of genes highly expressed by each cluster in the ForeceAtlas2 layout as in (H).

(J) Heatmap showing the differentially expressed genes along the lineage trajectory clusters (neural tube 3, NMP, paraxial mesoderm, presomitic mesoderm, and somite) and pseudotime of the NMP transformation.

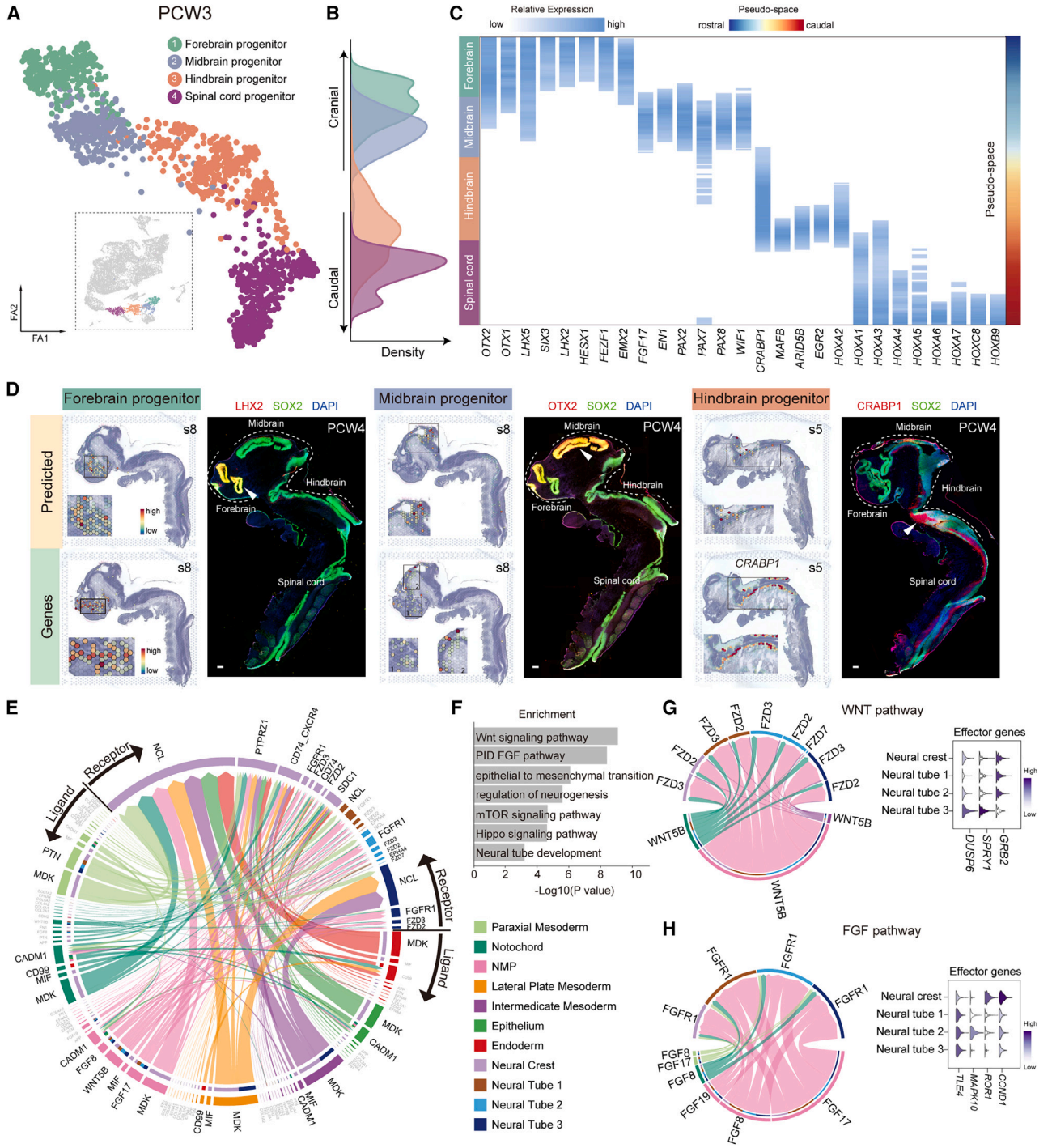



**Figure 3. Spatial organization of the neural tube cells at PCW3**

(A) ForceAtlas2 layout of neural tube cells at PCW3 colored-coded by regionalized clusters. Inset: locations of neural tube clusters in PCW3 UMAP plot from Figure 2A.

(B) The density of cells per cluster along the cranial-caudal axis of the neural tube by the color codes denoted in (A).

(C) Heatmap showing the average expression levels of marker genes along the pseudo-space axis for forebrain, midbrain, hindbrain, and spinal cord progenitor.

(D) Spatial location prediction and transcriptome from sections 8 and 5 and immunostaining reveal the distribution and expression of the forebrain (*LHX2*), midbrain (*OTX2*), and hindbrain progenitor (*CRABP1*) markers. The predicted upper panels indicated predicted regions for each progenitor location. The lower panels show the spatial expression pattern of individual marker genes. Scale bars, 200 μm.

*(legend continued on next page)*

genes involved in WNT signaling (*WNT3A*, *DKK1*, *CITED1*, and *RSPO3*) and pattern specification (*FGF8*, *CDX2*, *HES7*, and *TBX6*) are highly expressed by NMPs and the adjacent paraxial mesoderm cells (Figures 2I and 2J). In contrast, genes associated with NOTCH signaling (*SIX1*, *RIPPLY2*, and *MESP2*) and somitogenesis (*MEOX1/2*, *TCF15*, and *ALDH1A2*) are highly expressed in cells from the presomitic mesoderm and somites (Figures 2I and 2J). In addition, we observed that neural lineage markers such as *SOX2* and *LAMP5* are expressed in neural tube 3 cells (Figures 2I and 2J).[32]

## Spatial organization of neural tube cells

We further examined the spatial organization of neural tube cells from the three distinctive neural tube clusters at PCW3 (neural tube 1, 2, and 3 in Figure 2A). We found that neural tube cells were regionalized along the cranial-caudal axis as forebrain, midbrain, hindbrain, and spinal cord progenitors (Figures 3A and 3B) and expressed distinct sets of genes along a pseudo-spatial trajectory (Figure 3C; Table S3).[4] To spatially resolve the expression of marker genes, we mapped the neural tube cells to the spatial transcriptomics data. As expected, the neural tube cells identified by scRNA-seq, and DEGs were spatially segregated (Figures 3D and S3A). For example, we found strong *LHX2* and *FEZF1* expression in the forebrain, while *SIX3* and *EMX2* expression is limited to the ventral and dorsal forebrain, respectively (Figures 3D and S3A). We also observed regionally restricted expression of *OTX2* (forebrain and midbrain) and *FGF17* and *EN1* (midbrain). We noticed that *PAX7* and *FGF8* were strongly expressed in the dorsal and posterior ventral aspects of the midbrain, respectively (Figures 3D and S3A). In contrast, *PAX2* expression in the midbrain was restricted to a narrow stripe at the isthmus (Figure S3B). We found regional *EGR2*, *MAFB*, and *HOXA2* expression in the hindbrain (Figure S3A). In addition, we validated the spatial expression of *CRABP1* in the hindbrain (Figure 3D). We speculate that the lower expression in the anterior hindbrain could be due to the developmental difference between the spatial and single-cell samples used in this analysis. We also observed that the expression of HOX genes (*HOXA1*, *HOXA2*, *HOXA5*, *HOXA7*, and *HOXB9*) correlate well with our current understanding of the anteroposterior patterning of the hindbrain and spinal cord (Figures 3C and S3A).

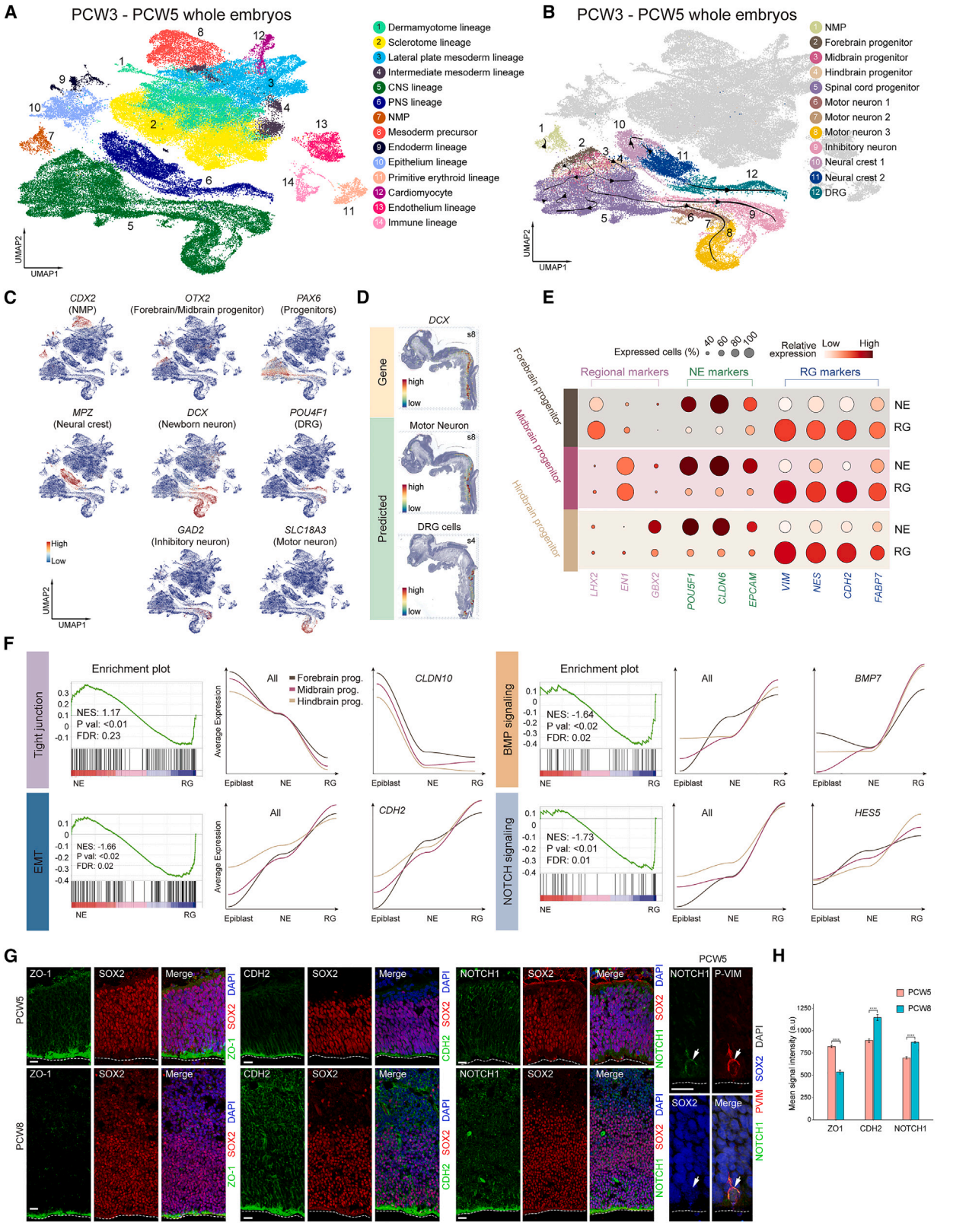
Primary neurulation induction involves not only the neural ectoderm but also the surrounding tissues. The notochord and underlying mesoderm secrete signaling molecules to promote neural ectoderm folding and differentiation.[15,33] To investigate how the mesoderm, notochord, and endoderm influence primary neurulation at PCW3, we performed CellChat analysis[34] and explored intercellular communications among these cells (Figure 3E; Table S3). GO enrichment analysis revealed that genes encoding ligand-receptor pairs in the FGF, WNT, mTOR and Hippo signaling pathways are associated with neural tube development (Figure 3F). For instance, we detected the gene encoding WNT5B in the notochord, NMP and intermediate mesoderm and the genes encoding the receptors FZD2, FZD3, and FZD7 in neural ectoderm cells (Figure 3G). Similarly, we found the genes encoding FGF8, FGF17, and FGF19 in the paraxial mesoderm, notochord, and NMP and the gene encoding FGFR1 in neural ectoderm cells (Figure 3H). Consistent with activated WNT and FGF pathways, we also observed the diverse expression of selected downstream effector genes among the neural tube and NC clusters (Figures 3G and 3H). These observations suggested that non-ectodermal tissues regulate neural ectoderm folding and differentiation through signaling pathways, including WNT and FGF signaling.

## Transitions from epiblasts to neuroepithelial cells to radial glia

We analyzed single-cell transcriptome profiles of whole embryos from PCW3 to 5 to investigate the dynamics of cell fate determination during early neurogenesis. A total of 81,102 cells were categorized into 33 clusters from 11 lineages based on the expression of known markers involved in early organogenesis, including different neural lineages from the ectoderm (Figures 4A and S3C–S3F; Table S4). A clear distinction of the differentiation trajectories between the central nervous system (CNS) and PNS, respectively is apparent from these early stages (Figure 4B). Spinal cord motor neurons and dorsal root ganglion (DRG) cells were abundant at PCW4 and 5 (Figures 4C, S3D, and S3E). The expression of the newborn neuron marker *DCX* was enriched in the posterior CNS and PNS clusters, including those corresponding to spinal cord motor neurons and DRG cells (Figure 4D), in agreement with a posterior-to-anterior gradient of neurogenesis. We observed that CNS progenitors were clustered by regional marker gene expression. For example, *LHX2*, *EN1*, and *EGR2* were expressed by forebrain, midbrain, and hindbrain progenitors, respectively (Figures 4E and S3F).[28,31,35] To better understand the early stages of progenitor development, each brain regional progenitor cluster was further classified into NE and RG cells with marker genes, considering that the NE cells presented tight junction features, while the RG cells showed glial characteristics (Figure 4E).

Although the conversion of epiblast cells into NE and their subsequent transformation into RG are critical developmental milestones,[19,36] the molecular regulation of this process remains unclear. We performed gene set enrichment analysis (GSEA) to identify signaling pathways responsible for the epiblast-NE-RG transformations.[37] We found that genes encoding tight junction proteins, such as *CLDN10*, were highly expressed in epiblast cells, and their expression gradually decreased during the transformation (Figure 4F). Consistently, we observed a substantial reduction in the expression of tight junction protein ZO-1 during the conversion of NE cells into RGs (Figures 4G and 4H). We also found that genes linked to epithelial-mesenchymal transition (EMT), such as the gene encoding N-cadherin (*CDH2*), were up-regulated during this process (Figures 4F–4H). Several other signaling pathways, including BMP and NOTCH, were also implicated in regulating the epiblast-NE-RG transformations.

(E–H) Analysis of cell-cell communications between neural ectoderm and other tissue clusters. The chord diagram shows the ligand-receptor interaction pairs from mesoderm, notochord, and endoderm to neural ectoderm clusters (E). GO enrichment analysis for the ligand-receptor interaction genes from (E) are shown in (F). The right chord diagrams show the significant interaction pairs associated with WNT (G) and FGF (H) signaling pathways, respectively. The right violin plots show the downstream effector genes which differed between the neural tube and neural crest clusters for WNT (G) and FGF (H) pathways, respectively.




*(legend on next page)*

Accordingly, *BMP7* and *HES5* expression increased gradually during this process (Figures 4F–4H). These analyses illuminate the molecular mechanisms regulating the earliest steps in human neurogenesis.

## Diversity of radial glia cells in the CNS

RGs are progenitor cells that give rise to distinct classes of neurons and glial cells in the CNS.[19,36] To identify fundamental molecular mechanisms underlying neurogenesis during early brain development, we extracted 217,778 progenitor cells and neurons from the CNS of all the samples. We used the expression of known markers to classify them as RGs, IPCs, and neurons (Figures S4A–S4C). Since IPCs are primarily described as secondary progenitors in the developing telencephalon, we named progenitor cells with similar neurogenic features but from outside the telencephalon as transient amplifying cells (TACs). We then used velocity analysis to identify the general developmental trajectory directions linking specific populations of RGs, IPCs/TACs, and neurons (Figure S4D). To explore cell-type diversity and heterogeneity, we clustered these cells and identified 24 RG, 5 IPCs/TAC, and 17 neuronal clusters, respectively (Figures 5A, 5B, S4E, and S4F; Table S5).

We integrated the scRNA-seq clusters with those obtained through spatial transcriptomics to reveal their regional organization in the developing CNS[38] (Figures 5C and S4G). We observed that the distinct RG clusters distributed heterogeneously across different brain and spinal cord regions. For example, RG15 and RG21 mapped to the eye and were characterized by the expression of *SIX3* and *SIX6* (Figures 5B and 5C), suggesting a retina progenitor identity.[9] RG3, RG6, RG7, RG9, and RG24 were found in the anterior brain and expressed the telencephalic marker *FOXG1*. RG14 was found in the thalamic region and expressed *BARHL2* and *OTX2*,[39] whereas RG8, RG10, and RG11 mapped to the midbrain and expressed *FGF17* and *EN1*.[40] RG2, RG4, RG12, RG13, RG19, RG20, RG22, and RG23 localized to the hindbrain and spinal cord and expressed *HOX* genes. We also observed spatial segregation along the dorsoventral axis of the CNS, with RG11, RG14, and RG20 expressing dorsal markers such as *ZIC1* and occupying dorsal regions,[41] and RG23 expressing ventral markers such as *SHH* and mapping to ventral regions (Figures 5B and 5C).

We performed correlation analyses to identify putative differentiation trajectories from distinct RG clusters to different classes of neurons. We identified six region-distinctive differentiation lineages based on their predicted spatial location and the expression of region-specific genes (Figures 5C, 5D, and S4H). For example, the first lineage linked the *PAX6*$^+$ RG6, RG7, and RG18 clusters with dorsal telencephalic Glu neurons, which included the *TBR1*$^+$ clusters Glu1, Glu2, and Glu4 and a *RELN*$^+$ and *LHX1*$^+$ cluster that likely corresponds to Cajal-Retzius cells (CRCs).[42] This lineage transverses through the IPC2 and IPC3 clusters, characterized by the expression of *EOMES* (Figures 5D–5H). Among the RG clusters in the first lineage, we noticed that RG6 and RG7 were found as early as PCW4, while RG18 only emerged in PCW9. RG18 and a subcluster of RG6 (RG6-1) were specifically defined by the expression of *HOPX* and *MOXD1*, which are characteristic of outer RG (oRG) (Figures 5B and 5I).[43] We performed velocity analyses to infer the developmental trajectory linking early RG clusters in the dorsal ventricular zone (VZ) (RG6-0 and RG7) and oRG (RG18 and RG6-1) and found a consistent enrichment of this later population from PCW9 onward (Figures 5I–5L). We also found specific gene modules that were mainly enriched in oRG, including genes regulating cell migration (*HOPX*, *FAM107A*, and *PTN*) and several signaling pathways such as WNT (*CTNNB1*, *SOX9*, and *SFRP1*), MAPK (*HSPA1A*, *HSPA1B*, and *HSPA6*), and PI3K-AKT-mTOR (*NGFR, NTRK2, PDGFD*) (Figures 5M, and S4I). We speculate that these signaling pathways may play essential roles in generating oRG cells. In the ventral telencephalon, the second lineage linked the *NKX2-1*$^+$ clusters RG3 and RG9 with two populations of GABA neurons, GABA1, and GABA3, via *ASCL1*-expression IPC1, possibly corresponding to GABA neurogenesis from the medial ganglionic eminence (Figures 5N–5Q). Altogether, our data highlighted the heterogeneity of RG across different regions and their possible developmental lineage in the early developing CNS.

## Distinctive features of human early nervous system development

To investigate similarities and differences across species during early embryonic development, we integrated the single-cell transcriptome profiles of humans (45,354 cells from PCW3 and 4-1) and mice (54,482 cells from Es 8.5, 9.5, and 10.5).[27] We first performed a cross-species comparison of all embryonic cells. We observed a great degree of similarity in cell clusters defined by the expression of marker genes (Figures S5A–5G). We then conducted regulon analyses to compare the critical regulators of early organogenesis in humans and mice (Figures S5H–S5M; Table S6). We identified twelve gene modules likely involved in this process.

**Figure 4. Early embryo neurogenesis from PCW3–5**

(A) UMAP visualization of PCW3–5 whole-embryo cells, colored by developmental lineages. CNS, central nervous system; PNS, peripheral nervous system.

(B) Velocity visualization of the NMP, CNS, and PNS lineages from PCW3 to 5 whole-embryo cells. Individual NMP and neural clusters are color-coded. Cells from other lineages are colored gray. DRG, dorsal root ganglion.

(C) Expression of marker genes of clusters in the early nervous system development between PCW3 and 5.

(D) Spatial transcriptome visualization from PCW4 showing the newborn neuron marker gene *DCX*, together with the predicted spatial locations for motor neuron and DRG cells.

(E) Dot plot of average expression of marker genes for the brain regional NE, and RG cells.

(F) Significant Kyoto Encyclopedia of Genes and Genomes (KEGG) signaling pathways and biological processes during the epiblast-NE-RG transformation in each brain region. Dynamic expression patterns of all genes in each pathway and representative genes (*CLDN10*, *CDH2*, *BMP7*, and *HES5*) are presented and colored by spatial regions. Prog, progenitor.

(G) Immunostaining of ZO-1, CDH2, NOTCH1, and SOX2 at PCW5 and 8 validating the change of expression patterns of genes from NE to RG. The right panels show the morphology of RG cells indicated by NOTCH1, P-VIM, and SOX2 staining at PCW5. Scale bars, 20 μm.

(H) The mean signal intensity for ZO-1, CDH2, and NOTCH1 in the apical surface of the ventricular zone at PCW5 and 8. Data are represented as mean ± SEM. Data of 8–11 brain regions from n = 3 brain slices were used for each gene at each time point. Unpaired Student's test was used for statistics. ****$p < 0.0001$.

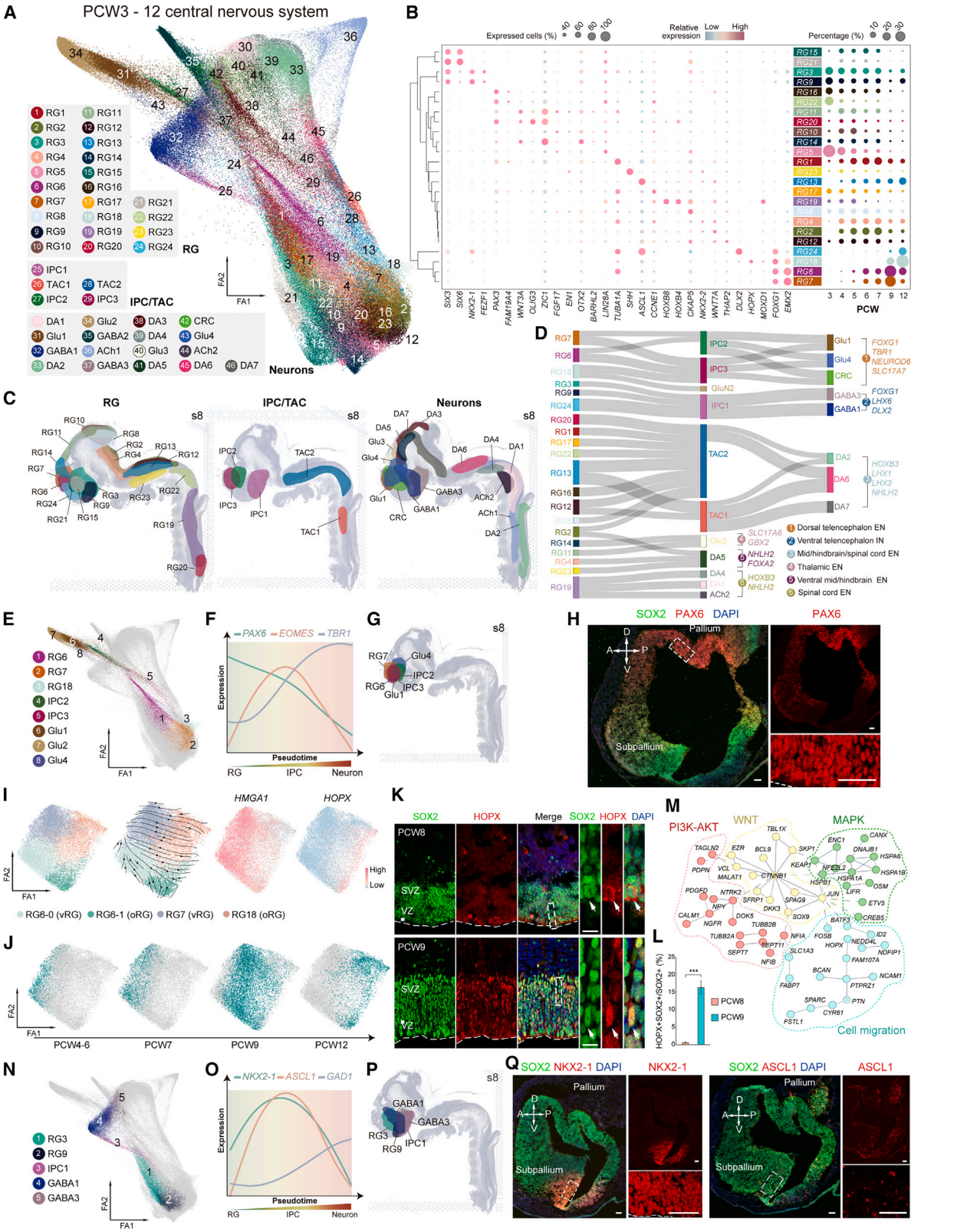



**Figure 5. Early development of the central nervous system from PCW3 to 12**

(A) ForceAtlas2 layout of the 24 RG, 5 IPC/TACs, and 17 neuron subclusters in early human central nervous system development between PCW3 and 12. TAC, transient amplifying cell; CRC, Cajal-Retzius cell.

*(legend continued on next page)*

Among them, transcription factors in module 12 were likely involved in central and PNS development (Figures S5J–S5M).

We next focused on cross-species differences during the early development of the nervous system to compare the molecular and cellular programs between humans and mice. Canonical correlation analysis (CCA) joint clustering assignments indicated that most neural cell types were highly correlated between the two species (Figure 6A). However, we found that cluster 14 was human dominant, whereas cluster 15 (*NKX2-1*$^+$, *DLX6*$^+$) was mouse dominant (Figures 6A, S6A, and S6B). The human-dominant cluster 14 was identified as the cerebellar rhombic lip (RL) by the expression of *LMX1A* and *WLS* (Figures 6B and S6C).[44] The human RL splits into VZ and subventricular zone (SVZ) around PCW10 to sustain prolonged granule cell production. In contrast, the mouse RL does not develop an SVZ, and the VZ thins from E10.5 to E15.5 and disappears by birth.[45] Consistent with the scRNA-seq data analysis, we observed that the human RL contains three layers of progenitor cells at PCW5 (Figure S6D). To investigate how gene expression is regulated in the RL, we analyzed the regulons of the human and mouse neural cells separately. The shared regulons were categorized into ten modules corresponding to specific neural clusters (Figure 6C). *ZNF502*, *TIA1*, *PRDM16*, *CREB3L1*, and *LMX1B* were identified as top regulons for the human-dominant cluster 14 (Figures 6D and S6E), among which we found that *WLS* and *MSX2* highly expressed in cluster 14 were regulated by the regulons of *PRDM16* and *LMX1B*, respectively (Figures 6B and 6E–6G).

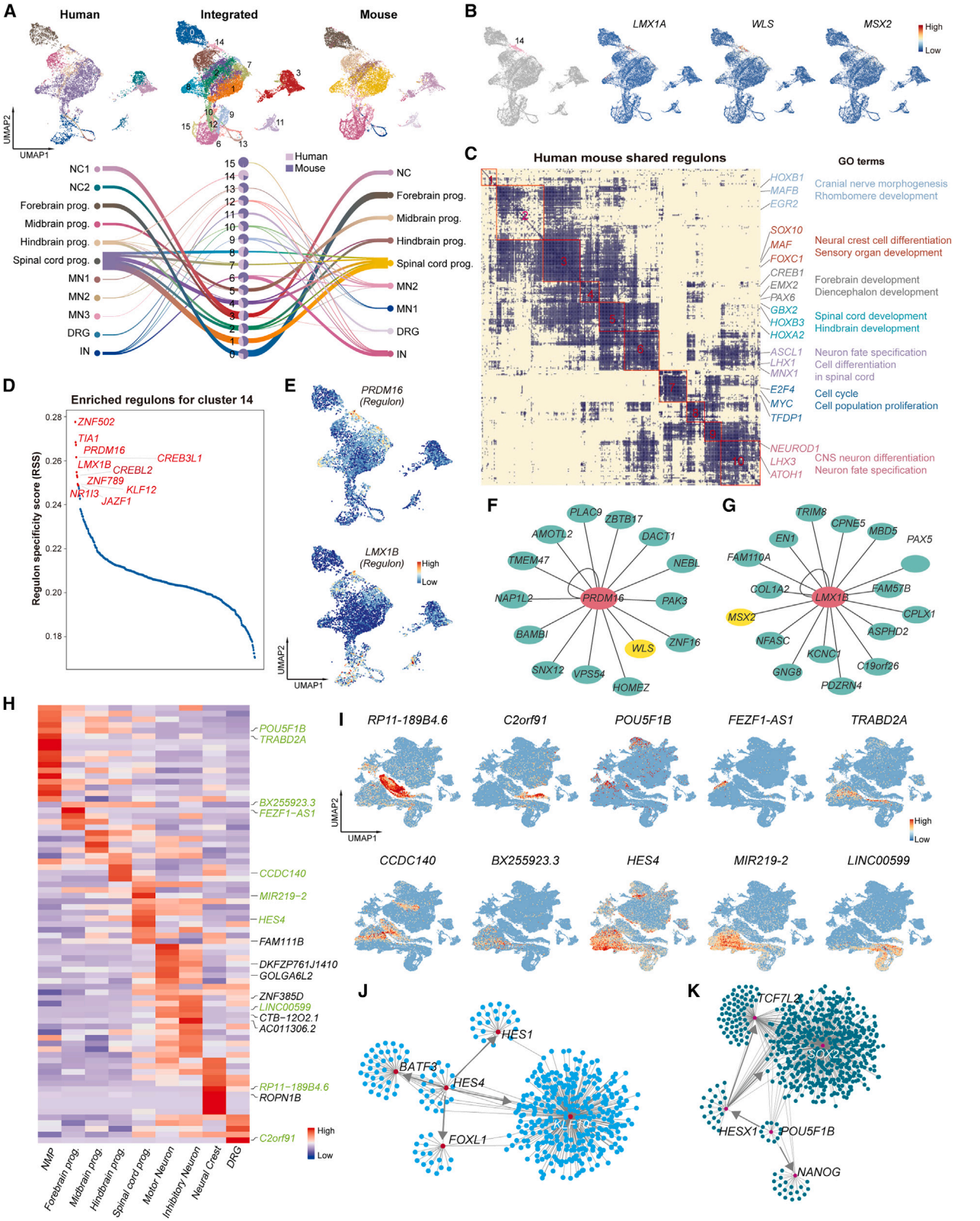
Several primate-specific genes have been reported to play essential roles in human brain development, brain size enlargement, and cortical folding.[46–50] We analyzed the cell-type-specific expression of primate-specific genes[51] and long noncoding RNAs during early organogenesis (PCW3–5) (Figure S6F and S6G), with a particular focus on the nervous system (Figures 6H and 6I). *POU5F1B* was highly expressed in NMPs and neural progenitor cells. *CCDC140* was highly expressed in NC cells and subsets of CNS progenitors. *LINC00599*, a long noncoding RNA, was selectively expressed in neurons from the CNS (Figure 6H and 6I). Next, we conducted regulon analyses to dissect the downstream regulatory network of human-specific genes. *HES4*, which is not present in mice[52] but highly expressed in human progenitors, might orchestrate early neural system development by regulating the transcription of *HES1*, *BATF3*, and other genes in humans (Figures 6H, 6J, and S6H; Table S6). The regulon network suggested that *POU5F1B* regulates the transcription of *NANOG*, which probably maintains stemness, and *SOX2* via *HESX1* for neural lineage specification (Figures 6K and S6I; Table S6). Our results suggest that humans and mice share broadly conserved features linked to cell-type diversity and the transcriptional regulation of early CNS development. However, humans also possess distinct regulons that may underlie brain evolution, some of which may be attributed to primate-specific genes.

## DISCUSSION

Our understanding of the molecular programs regulating human gastrulation is much more limited due to the exceptional difficulty in accessing appropriate samples. Our study fills this gap by analyzing an extraordinary collection of human embryos spanning key gastrulation stages, including early organogenesis and primary neurulation. Identifying gene expression profiles with single-cell resolution recapitulates the transcriptional regulation of human gastrulation, the emergence of cell-type diversity, and the precise spatial patterning of neural tube cells at the gastrulation stage.

Consistent with the characterization of signaling pathways that regulate the transition from cortical NE to RG,[16] we identified the regulation of tight junctions and EMT as essential processes mediating the NE to RG transition, suggesting that progenitor cells progressively lose polarity and cell-cell adhesion activity,[53] and the EMT process promoting the expression switch from E-cadherin (CDH1) to N-cadherin (CDH2) is essential for the maintenance, proliferation, and differentiation of neural progenitors.[54] Eze and colleagues recently described the dynamic expression of genes in the NOTCH signaling pathway. They found that NOTCH1 expression was primarily restricted to the ventricular zone at CS14 and CS16.[16] Similarly, we observed that the NOTCH signaling pathway gene *HES5* was strongly expressed in RG cells and found a steady increase in its expression from CS7 to CS13 in the neural progenitor cells. The study from

(B) The left dot plot shows the relative expression levels of marker genes (color) and the proportion of cells (size) in each RG subcluster from all analyzed RG cells. The dendrogram indicates transcriptome distances between clusters. The right dot plot shows the fraction of RG subcluster cells at each developmental stage.
(C) Predicted spatial distributions of RG, IPC/TAC, and neuron subclusters in the PCW4 spatial transcriptome.
(D) Sankey diagram mapping the cell fate trajectories from RG subclusters to neurons through IPC/TAC cells. The edge width represents the transcriptional correlation between subclusters. Only correlation values larger than 0.3 were shown. EN, excitatory neuron; IN, inhibitory neuron.
(E and N) ForceAtlas2 layouts of the RG-IPC-neuron developmental lineages in the dorsal (E) and ventral (N) telencephalon, respectively.
(F and O) Pseudotemporal expression patterns of markers for RG (dorsal: *PAX6*; ventral: *NKX2-1*), IPC (dorsal: *EOMES*; ventral: *ASCL1*), and neuron (dorsal: *TBR1*; ventral: *GAD1*) along the dorsal (F) and ventral (O) developmental lineages, respectively.
(G and P) Predicted spatial locations of the indicated RG, IPC, and neuronal subclusters from the dorsal (G) and ventral telencephalon lineages (P) mapped to the PCW4 spatial transcriptome.
(H and Q) Immunostaining demonstrating regional PAX6 staining in the pallium (H) and NKX2-1 and ASCL1 staining in the subpallium (Q). Progenitor cells are labeled by SOX2 staining. Ventricular surfaces are denoted by dashed lines in the high-magnification images from white boxes. A, P, D, and V are abbreviations for anterior, posterior, dorsal, and ventral, respectively. Scale bars, 50 μm.
(I and J) Clustering, velocity, and marker genes of dorsal ventricular RG (vRG) marker *HMGA1* and oRG cells marker *HOPX* (I) and cell distributions at different post-conceptional week stages (J).
(K) Immunostaining of HOPX and SOX2 at PCW8 and 9. SOX2$^+$HOPX$^+$ oRG drastically increased in the PCW9 pallium. Scale bars, 10 μm.
(L) Percentage of HOPX+SOX2+ cells among total SOX2+ cells at PCW8 and 9. Data are represented as mean ± SEM. Data of 6–7 brain regions from n = 3 brain slices were used at each time point. Unpaired Student's t test was used for statistics. ***$p < 0.001$.
(M) The protein interaction network between genes highly expressed in oRG cells is constructed from the "string" database; the interacting genes were clustered and formed four functionally enriched groups.




(legend on next page)

Eze et al. covered the developmental stages from CS12 to CS22, indicating that NOTCH signaling peaked around CS16.[16] These two studies together suggested that NOTCH signals might be dynamic during the first trimester for increases in RG generation and maintenance of the progenitor pool.

RGs are very heterogeneous progenitor cells underlying the generation of neuron and glial cell diversity throughout the developing CNS.[18,36] RG are present throughout the developing CNS[55,56] and are very diverse; hence, understanding the relationship between molecular heterogeneity and spatial distribution among RG is a critical question in developmental neurobiology. Previous studies have described 9 clusters of RG throughout the cortex.[16] In contrast, our study identified 24 RG clusters with distinctive transcriptional identities and spatial localization between PCW3 and 12. In addition, we identified unique RG-IPC/TACs-neuron differentiation trajectories, providing a framework for investigating the molecular mechanisms orchestrating early neurodevelopment. oRG derived from RG cells are present in the pallium throughout primate neurogenesis.[11,43] Eze et al. found that the progenitors in the human pallium from PCW6 to 10 were NE, RG, and IPC, but not oRG.[16] However, we found that oRG emerges between PCW9 and 12 in humans, and the WNT and PI3K-AKT-mTOR signaling pathways control this process. WNT signaling has been proposed to play a role in regulating the maturation of RGs into IPCs,[57] which suggests that the RG-derived progenitors (oRG or IPC) may share similar signals.

Our study revealed that the main cell types involved in early embryonic development are similar between humans and mice. In addition, regulon analyses suggested that different species may use shared and unique regulatory mechanisms to guide organogenesis. The evolution of hominids has led to a notable increase in brain size and a tremendous expansion of the cerebral cortex, which is considered the biological base of the expanded cognitive abilities of humans. Despite fundamentally conserved cell types found during early brain development, we identified a subset of RL progenitors (i.e., cluster 14) predominantly present in humans. The SVZ expansion of the cerebellar RL around PCW10 is reported as a human-specific feature. Disruption of RL development is associated with cerebellar hypoplasia and Dandy-Walker malformation.[45] In addition, the human RL SVZ is the origin of group 3 and 4 medulloblastomas (MB) tumors.[58,59]

We investigated the expression of primate-specific genes[51] and found that they exhibit tissue or cell-type preference, and about a third of them are highly expressed in the developing human nervous system, which suggests that the expression of primate-specific genes may contribute to nervous system evolution. Our study offers a valuable large-scale single-cell atlas of early human organogenesis and brain development which may contribute to expanding our current knowledge of early human development and evolution. Understanding the molecular mechanisms of early human embryogenesis could facilitate the development of more robust protocols for the precise differentiation of human embryonic stem cells and induced pluripotent stem cells into desired cell types.

## Limitations of the study

There are several limitations of this study. First, due to the limited number of samples sequenced, we may not be able to fully illustrate the cellular programs during human gastrulation, which may be complemented by further work with large-scale sequencing samples covering the entire gastrulation period. Second, some findings of the TFs, cell-cell interaction proteins, and signaling pathways during the germ layer and neural development were reported only based on transcriptomic data; further work exploring novel molecules and validating these molecules and signaling pathways in humans will be required to understand the distinctiveness of the human embryogenesis. Third, the spatial transcriptomics data lacked single-cell resolution and was performed at a single stage (PCW4). Approaches with enhanced resolution, such as spatial enhanced resolution omics-sequencing (Stero-seq),[60] at multiple developmental time points will provide more refined details to illustrate the spatial relationships among cell types and the emergence of evolutionarily novel cell types. Lastly, we identified a type of RL cell that is predominant in early human embryos and speculated that this cell type is related to the specific expansion of RL SVZ during later human fetal development. However, the limited number of samples sequenced from humans and mice may lead to over- or under-sampling of particular cell types, especially very rare ones. Increasing the number of samples and the range of developmental stages studied and performing further validation experiments should allow addressing this issue in the future.

## STAR★METHODS

Detailed methods are provided in the online version of this paper and include the following:

- KEY RESOURCES TABLE

**Figure 6. Cross-species comparison of nervous system development**

(A) UMAP visualization of the human (PCW3 and 4-1) and mouse (embryonic days 8.5, 9.5, and 10.5) integrated nervous system clusters. The top-left and top-right plots show the color-coded human and mouse clusters separately. River plots compare human and mouse cluster assignments with the joint clusters, and the pie chart indicates their contributions to each cluster. NC, neural crest; MN, motor neuron; IN, inhibitory neuron; DRG, dorsal root ganglia; prog, progenitor.

(B) Representative expression patterns of differentially expressed genes *LMX1A*, *WLS*, and *MSX2* for the human-dominant cluster 14.

(C) Regulon modules based on the correlation matrix of shared regulons between human and mouse nervous system cells. The network shows 332 orthologous transcription factor regulons grouped into 10 major modules (red boxes) with representative transcription factor (TF) regulons and gene ontology (GO) terms highlighted.

(D) The top 10 enriched regulons for the human-dominant cluster 14. The y axis values represent the regulon specificity scores (RSS).

(E) UMAP plots of human regulon activity scores for the PRDM16 and LMX1B regulons.

(F and G) The gene regulatory networks for cluster 14-enriched regulons PRDM16 (F) and LMX1B (G).

(H) Heatmap illustrating the expression levels of primate-specific genes in human CNS and PNS clusters between PCW3 and 5.

(I) Expression of primate-specific genes in PCW3–5 human clusters. Gene expression levels are plotted in the UMAP from Figure 4A.

(J and K) The gene regulatory network of TF HES4 (J) and POU5F1B (K). Arrows indicate the regulatory directions, and each small dot represents the downstream gene of the TF.

## SUPPLEMENTAL INFORMATION

Supplemental information can be found online at https://doi.org/10.1016/j.stem.2023.04.016.

## ACKNOWLEDGMENTS

This work was supported by the National Key R&D Program of China (2019YFA0110100), the National Natural Science Foundation of China (NSFC) (32122037, 32192411, and 81891001), the CAS Project for Young Scientists in Basic Research (YSBR-013), the Science and Technology Innovation 2030—Brain Science and Brain-inspired Intelligence Project of China (2021ZD0200102), the New Cornerstone Science Foundation to X.W., Changping Laboratory, the Wellcome Trust (215556/Z/19/Z to O.M.), and the Collaborative Research Fund of the Chinese Institute for Brain Research, Beijing (2020-NKX-PT-02 and 2020-NKX-PT-03).

## AUTHOR CONTRIBUTIONS

X.W., O.M., Q.W., and Y.W. conceived the project, designed the experiments, and wrote the manuscript. Q.W., Z. Liu, L.H., and Z. Li performed the single-cell RNA-seq and spatial transcriptomic experiment. B.Z., S.Q., Y.Z., and W.W. analyzed the RNA-seq data. Y.L. and Y.C. carried out the cross-species comparison. S.Z. and H.D. analyzed the regulon and network. J.Y., Q.M., L.B., D.H., and L.S. prepared the samples. All authors edited and proofed the manuscript.

## DECLARATION OF INTERESTS

The authors declare no competing interests.

Received: November 16, 2022
Revised: March 14, 2023
Accepted: April 17, 2023
Published: May 15, 2023

## REFERENCES

1. Pijuan-Sala, B., Griffiths, J.A., Guibentif, C., Hiscock, T.W., Jawaid, W., Calero-Nieto, F.J., Mulas, C., Ibarra-Soria, X., Tyser, R.C.V., Ho, D.L.L., et al. (2019). A single-cell molecular map of mouse gastrulation and early organogenesis. Nature *566*, 490–495. https://doi.org/10.1038/s41586-019-0933-9.
2. Cao, J., Spielmann, M., Qiu, X., Huang, X., Ibrahim, D.M., Hill, A.J., Zhang, F., Mundlos, S., Christiansen, L., Steemers, F.J., et al. (2019). The single-cell transcriptional landscape of mammalian organogenesis. Nature *566*, 496–502. https://doi.org/10.1038/s41586-019-0969-x.
3. Argelaguet, R., Clark, S.J., Mohammed, H., Stapel, L.C., Krueger, C., Kapourani, C.A., Imaz-Rosshandler, I., Lohoff, T., Xiang, Y., Hanna, C.W., et al. (2019). Multi-omics profiling of mouse gastrulation at single-cell resolution. Nature *576*, 487–491. https://doi.org/10.1038/s41586-019-1825-8.
4. Nowotschin, S., Setty, M., Kuo, Y.Y., Liu, V., Garg, V., Sharma, R., Simon, C.S., Saiz, N., Gardner, R., Boutet, S.C., et al. (2019). The emergent landscape of the mouse gut endoderm at single-cell resolution. Nature *569*, 361–367. https://doi.org/10.1038/s41586-019-1127-1.
5. Fang, H., Yang, Y., Li, C., Fu, S., Yang, Z., Jin, G., Wang, K., Zhang, J., and Jin, Y. (2010). Transcriptome analysis of early organogenesis in human embryos. Dev. Cell *19*, 174–184. https://doi.org/10.1016/j.devcel.2010.06.014.
6. Ma, H., Zhai, J., Wan, H., Jiang, X., Wang, X., Wang, L., Xiang, Y., He, X., Zhao, Z.A., Zhao, B., et al. (2019). In vitro culture of cynomolgus monkey embryos beyond early gastrulation. Science *366*. https://doi.org/10.1126/science.aax7890.
7. Niu, Y., Sun, N., Li, C., Lei, Y., Huang, Z., Wu, J., Si, C., Dai, X., Liu, C., Wei, J., et al. (2019). Dissecting primate early post-implantation development using long-term in vitro embryo culture. Science *366*. https://doi.org/10.1126/science.aaw5754.
8. Zhong, S., Ding, W., Sun, L., Lu, Y., Dong, H., Fan, X., Liu, Z., Chen, R., Zhang, S., Ma, Q., et al. (2020). Decoding the development of the human hippocampus. Nature *577*, 531–536. https://doi.org/10.1038/s41586-019-1917-5.
9. Lu, Y., Shiau, F., Yi, W., Lu, S., Wu, Q., Pearson, J.D., Kallman, A., Zhong, S., Hoang, T., Zuo, Z., et al. (2020). Single-cell analysis of human retina identifies evolutionarily conserved and species-specific mechanisms controlling development. Dev. Cell *53*, 473–491.e9. https://doi.org/10.1016/j.devcel.2020.04.009.
10. Zhong, S., Zhang, S., Fan, X., Wu, Q., Yan, L., Dong, J., Zhang, H., Li, L., Sun, L., Pan, N., et al. (2018). A single-cell RNA-seq survey of the developmental landscape of the human prefrontal cortex. Nature *555*, 524–528. https://doi.org/10.1038/nature25980.
11. Nowakowski, T.J., Bhaduri, A., Pollen, A.A., Alvarado, B., Mostajo-Radji, M.A., Di Lullo, E., Haeussler, M., Sandoval-Espinosa, C., Liu, S.J., Velmeshev, D., et al. (2017). Spatiotemporal gene expression trajectories reveal developmental hierarchies of the human cortex. Science *358*, 1318–1323. https://doi.org/10.1126/science.aap8809.
12. Bian, Z., Gong, Y., Huang, T., Lee, C.Z.W., Bian, L., Bai, Z., Shi, H., Zeng, Y., Liu, C., He, J., et al. (2020). Deciphering human macrophage development at single-cell resolution. Nature *582*, 571–576. https://doi.org/10.1038/s41586-020-2316-7.
13. Sahara, M., Santoro, F., Sohlmér, J., Zhou, C., Witman, N., Leung, C.Y., Mononen, M., Bylund, K., Gruber, P., and Chien, K.R. (2019). Population and single-cell analysis of human cardiogenesis reveals unique LGR5

ventricular progenitors in embryonic outflow tract. Dev. Cell *48*, 475–490.e7. https://doi.org/10.1016/j.devcel.2019.01.005.

14. Breunig, J.J., Haydar, T.F., and Rakic, P. (2011). Neural stem cells: historical perspective and future prospects. Neuron *70*, 614–625. https://doi.org/10.1016/j.neuron.2011.05.005.
15. Nikolopoulou, E., Galea, G.L., Rolo, A., Greene, N.D., and Copp, A.J. (2017). Neural tube closure: cellular, molecular and biomechanical mechanisms. Development *144*, 552–566. https://doi.org/10.1242/dev.145904.
16. Eze, U.C., Bhaduri, A., Haeussler, M., Nowakowski, T.J., and Kriegstein, A.R. (2021). Single-cell atlas of early human brain development highlights heterogeneity of human neuroepithelial cells and early radial glia. Nat. Neurosci. *24*, 584–594.
17. Solnica-Krezel, L., and Sepich, D.S. (2012). Gastrulation: making and shaping germ layers. Annu. Rev. Cell Dev. Biol. *28*, 687–717. https://doi.org/10.1146/annurev-cellbio-092910-154043.
18. Kriegstein, A., and Alvarez-Buylla, A. (2009). The glial nature of embryonic and adult neural stem cells. Annu. Rev. Neurosci. *32*, 149–184. https://doi.org/10.1146/annurev.neuro.051508.135600.
19. Rakic, P. (2003). Elusive radial glial cells: historical and evolutionary perspective. Glia *43*, 19–32. https://doi.org/10.1002/glia.10244.
20. Herculano-Houzel, S. (2012). The remarkable, yet not extraordinary, human brain as a scaled-up primate brain and its associated cost. Proc. Natl. Acad. Sci. USA *109* (*Suppl 1*), 10661–10668. https://doi.org/10.1073/pnas.1201895109.
21. Van der Maaten, L., and Hinton, G.J. (2008). Visualizing Data Using t-SNE*9*.
22. Blondel, V.D., Guillaume, J.L., Hendrickx, J.M., de Kerchove, C., and Lambiotte, R. (2008). Local leaders in random networks. Phys. Rev. E Stat. Nonlin. Soft Matter Phys. *77*, 036114. https://doi.org/10.1103/PhysRevE.77.036114.
23. Attardi, A., Fulton, T., Florescu, M., Shah, G., Muresan, L., Lenz, M.O., Lancaster, C., Huisken, J., van Oudenaarden, A., and Steventon, B. (2018). Neuromesodermal progenitors are a conserved source of spinal cord with divergent growth dynamics. Development *145*. https://doi.org/10.1242/dev.166728.
24. Müller, F., and O'rahilly, R. (1985). The first appearance of the neural tube and optic primordium in the human embryo at stage 10. Anat. Embryol. (Berl) *172*, 157–169. https://doi.org/10.1007/BF00319598.
25. Hill, M.A. (2007). Early human development. Clin. Obstet. Gynecol. *50*, 2–9. https://doi.org/10.1097/GRF.0b013e31802f119d.
26. Tyser, R.C.V., Mahammadov, E., Nakanoh, S., Vallier, L., Scialdone, A., and Srinivas, S. (2021). Single-cell transcriptomic characterization of a gastrulating human embryo. Nature *600*, 285–289. https://doi.org/10.1038/s41586-021-04158-y.
27. Stuart, T., Butler, A., Hoffman, P., Hafemeister, C., Papalexi, E., Mauck, W.M., 3rd, Hao, Y., Stoeckius, M., Smibert, P., and Satija, R. (2019). Comprehensive integration of single-cell data. Cell *177*, 1888–1902.e21. https://doi.org/10.1016/j.cell.2019.05.031.
28. Qiu, C., Cao, J., Martin, B.K., Li, T., Welsh, I.C., Srivatsan, S., Huang, X., Calderon, D., Noble, W.S., Disteche, C.M., et al. (2022). Systematic reconstruction of cellular trajectories across mouse embryogenesis. Nat. Genet. *54*, 328–341. https://doi.org/10.1038/s41588-022-01018-x.
29. Tani, S., Chung, U.I., Ohba, S., and Hojo, H. (2020). Understanding paraxial mesoderm development and sclerotome specification for skeletal repair. Exp. Mol. Med. *52*, 1166–1177. https://doi.org/10.1038/s12276-020-0482-1.
30. Bouchard, M., Souabni, A., Mandler, M., Neubüser, A., and Busslinger, M. (2002). Nephric lineage specification by Pax2 and Pax8. Genes Dev. *16*, 2958–2970. https://doi.org/10.1101/gad.240102.
31. Lohoff, T., Ghazanfar, S., Missarova, A., Koulena, N., Pierson, N., Griffiths, J.A., Bardot, E.S., Eng, C.L., Tyser, R.C.V., Argelaguet, R., et al. (2022). Integration of spatial and single-cell transcriptomic data elucidates mouse organogenesis. Nat. Biotechnol. *40*, 74–85. https://doi.org/10.1038/s41587-021-01006-2.
32. Koebis, M., Urata, S., Shinoda, Y., Okabe, S., Yamasoba, T., Nakao, K., Aiba, A., and Furuichi, T. (2019). LAMP5 in presynaptic inhibitory terminals in the hindbrain and spinal cord: a role in startle response and auditory processing. Mol. Brain *12*, 20. https://doi.org/10.1186/s13041-019-0437-4.
33. Keller, R., Shih, J., Sater, A.K., and Moreno, C. (1992). Planar induction of convergence and extension of the neural plate by the organizer of Xenopus. Dev. Dyn. *193*, 218–234. https://doi.org/10.1002/aja.1001930303.
34. Jin, S., Guerrero-Juarez, C.F., Zhang, L., Chang, I., Ramos, R., Kuan, C.H., Myung, P., Plikus, M.V., and Nie, Q. (2021). Inference and analysis of cell-cell communication using CellChat. Nat. Commun. *12*, 1088. https://doi.org/10.1038/s41467-021-21246-9.
35. McMahon, A.P., Joyner, A.L., Bradley, A., and McMahon, J.A. (1992). The midbrain-hindbrain phenotype of Wnt-1-/Wnt-1- mice results from stepwise deletion of engrailed-expressing cells by 9.5 days postcoitum. Cell *69*, 581–595. https://doi.org/10.1016/0092-8674(92)90222-x.
36. Kriegstein, A.R., and Götz, M. (2003). Radial glia diversity: a matter of cell fate. Glia *43*, 37–43. https://doi.org/10.1002/glia.10250.
37. Mootha, V.K., Lindgren, C.M., Eriksson, K.F., Subramanian, A., Sihag, S., Lehar, J., Puigserver, P., Carlsson, E., Ridderstråle, M., Laurila, E., et al. (2003). PGC-1α-responsive genes involved in oxidative phosphorylation are coordinately downregulated in human diabetes. Nat. Genet. *34*, 267–273.
38. Biancalani, T., Scalia, G., Buffoni, L., Avasthi, R., Lu, Z., Sanger, A., Tokcan, N., Vanderburg, C.R., Segerstolpe, Å., Zhang, M., et al. (2021). Deep learning and alignment of spatially resolved single-cell transcriptomes with Tangram. Nat. Methods *18*, 1352–1362. https://doi.org/10.1038/s41592-021-01264-7.
39. Zhou, X., Lu, Y., Zhao, F., Dong, J., Ma, W., Zhong, S., Wang, M., Wang, B., Zhao, Y., Shi, Y., et al. (2022). Deciphering the spatial-temporal transcriptional landscape of human hypothalamus development. Cell Stem Cell *29*, 328–343.e5. https://doi.org/10.1016/j.stem.2021.11.009.
40. Nakamura, H., Katahira, T., Matsunaga, E., and Sato, T. (2005). Isthmus organizer for midbrain and hindbrain development. Brain Res. Brain Res. Rev. *49*, 120–126. https://doi.org/10.1016/j.brainresrev.2004.10.005.
41. Aruga, J., Tohmonda, T., Homma, S., and Mikoshiba, K. (2002). Zic1 promotes the expansion of dorsal neural progenitors in spinal cord by inhibiting neuronal differentiation. Dev. Biol. *244*, 329–341. https://doi.org/10.1006/dbio.2002.0598.
42. Loo, L., Simon, J.M., Xing, L., McCoy, E.S., Niehaus, J.K., Guo, J., Anton, E.S., and Zylka, M.J. (2019). Single-cell transcriptomic analysis of mouse neocortical development. Nat. Commun. *10*, 134. https://doi.org/10.1038/s41467-018-08079-9.
43. Pollen, A.A., Nowakowski, T.J., Chen, J., Retallack, H., Sandoval-Espinosa, C., Nicholas, C.R., Shuga, J., Liu, S.J., Oldham, M.C., Diaz, A.J.C., et al. (2015). Molecular identity of human outer radial glia during cortical development. Cell *163*, 55–67.
44. Yeung, J., and Goldowitz, D. (2017). Wls expression in the rhombic lip orchestrates the embryonic development of the mouse cerebellum. Neuroscience *354*, 30–42. https://doi.org/10.1016/j.neuroscience.2017.04.020.
45. Haldipur, P., Aldinger, K.A., Bernardo, S., Deng, M., Timms, A.E., Overman, L.M., Winter, C., Lisgo, S.N., Razavi, F., Silvestri, E., et al. (2019). Spatiotemporal expansion of primary progenitor zones in the developing human cerebellum. Science *366*, 454–460. https://doi.org/10.1126/science.aax7526.
46. Florio, M., Albert, M., Taverna, E., Namba, T., Brandl, H., Lewitus, E., Haffner, C., Sykes, A., Wong, F.K., Peters, J., et al. (2015). Human-specific gene ARHGAP11B promotes basal progenitor amplification and neocortex expansion. Science *347*, 1465–1470. https://doi.org/10.1126/science.aaa1975.
47. Liu, J., Liu, W., Yang, L., Wu, Q., Zhang, H., Fang, A., Li, L., Xu, X., Sun, L., Zhang, J., et al. (2017). The primate-specific gene TMEM14B marks outer radial glia cells and promotes cortical expansion and folding. Cell Stem Cell *21*, 635–649.e8. https://doi.org/10.1016/j.stem.2017.08.013.

48. Namba, T., Dóczi, J., Pinson, A., Xing, L., Kalebic, N., Wilsch-Bräuninger, M., Long, K.R., Vaid, S., Lauer, J., Bogdanova, A., et al. (2020). Human-specific ARHGAP11B acts in mitochondria to expand neocortical progenitors by glutaminolysis. Neuron *105*, 867–881.e9. https://doi.org/10.1016/j.neuron.2019.11.027.

49. Suzuki, I.K., Gacquer, D., Van Heurck, R., Kumar, D., Wojno, M., Bilheu, A., Herpoel, A., Lambert, N., Cheron, J., Polleux, F., et al. (2018). Human-specific NOTCH2NL genes expand cortical neurogenesis through delta/Notch regulation. Cell *173*, 1370–1384.e16. https://doi.org/10.1016/j.cell.2018.03.067.

50. Fiddes, I.T., Lodewijk, G.A., Mooring, M., Bosworth, C.M., Ewing, A.D., Mantalas, G.L., Novak, A.M., van den Bout, A., Bishara, A., Rosenkrantz, J.L., et al. (2018). Human-specific NOTCH2NL genes affect Notch signaling and cortical neurogenesis. Cell *173*, 1356–1369.e22. https://doi.org/10.1016/j.cell.2018.03.051.

51. Shao, Y., Chen, C., Shen, H., He, B.Z., Yu, D., Jiang, S., Zhao, S., Gao, Z., Zhu, Z., Chen, X., et al. (2019). GenTree, an integrated resource for analyzing the evolution and function of primate-specific coding genes. Genome Res. *29*, 682–696. https://doi.org/10.1101/gr.238733.118.

52. Kageyama, R., Ohtsuka, T., and Kobayashi, T. (2007). The Hes gene family: repressors and oscillators that orchestrate embryogenesis. Development *134*, 1243–1251. https://doi.org/10.1242/dev.000786.

53. Bengoa-Vergniory, N., and Kypta, R.M. (2015). Canonical and noncanonical Wnt signaling in neural stem/progenitor cells. Cell. Mol. Life Sci. *72*, 4157–4172. https://doi.org/10.1007/s00018-015-2028-6.

54. Miyamoto, D.T., Zheng, Y., Wittner, B.S., Lee, R.J., Zhu, H., Broderick, K.T., Desai, R., Fox, D.B., Brannigan, B.W., Trautwein, J., et al. (2015). RNA-Seq of single prostate CTCs implicates noncanonical Wnt signaling in antiandrogen resistance. Science *349*, 1351–1356. https://doi.org/10.1126/science.aab0917.

55. Carter, R.A., Bihannic, L., Rosencrance, C., Hadley, J.L., Tong, Y., Phoenix, T.N., Natarajan, S., Easton, J., Northcott, P.A., and Gawad, C. (2018). A single-cell transcriptional atlas of the developing murine cerebellum. Curr. Biol. *28*, 2910–2920.e2. https://doi.org/10.1016/j.cub.2018.07.062.

56. Zhou, X., Zhong, S., Peng, H., Liu, J., Ding, W., Sun, L., Ma, Q., Liu, Z., Chen, R., Wu, Q., et al. (2020). Cellular and molecular properties of neural progenitors in the developing mammalian hypothalamus. Nat. Commun. *11*, 4063. https://doi.org/10.1038/s41467-020-17890-2.

57. Harrison-Uy, S.J., and Pleasure, S.J. (2012). Wnt signaling and forebrain development. Cold Spring Harb. Perspect. Biol. *4*, a008094. https://doi.org/10.1101/cshperspect.a008094.

58. Smith, K.S., Bihannic, L., Gudenas, B.L., Haldipur, P., Tao, R., Gao, Q., Li, Y., Aldinger, K.A., Iskusnykh, I.Y., Chizhikov, V.V., et al. (2022). Unified rhombic lip origins of group 3 and group 4 medulloblastoma. Nature *609*, 1012–1020. https://doi.org/10.1038/s41586-022-05208-9.

59. Hendrikse, L.D., Haldipur, P., Saulnier, O., Millman, J., Sjoboen, A.H., Erickson, A.W., Ong, W., Gordon, V., Coudière-Morrison, L., Mercier, A.L., et al. (2022). Failure of human rhombic lip differentiation underlies medulloblastoma formation. Nature *609*, 1021–1028. https://doi.org/10.1038/s41586-022-05215-w.

60. Chen, A., Liao, S., Cheng, M., Ma, K., Wu, L., Lai, Y., Qiu, X., Yang, J., Xu, J., Hao, S., et al. (2022). Spatiotemporal transcriptomic atlas of mouse organogenesis using DNA nanoball-patterned arrays. Cell *185*, 1777–1792.e21. https://doi.org/10.1016/j.cell.2022.04.003.

61. Lun, A.T., McCarthy, D.J., and Marioni, J.C. (2016). A step-by-step workflow for low-level analysis of single-cell RNA-seq data with Bioconductor.

62. Wolf, F.A., Angerer, P., and Theis. (2018). SCANPY: large-scale single-cell gene expression data analysis F.J.J.G.b. *19*, 1–5.

63. La Manno, G., Soldatov, R., Zeisel, A., Braun, E., Hochgerner, H., Petukhov, V., Lidschreiber, K., Kastriti, M.E., Lönnerberg, P., and Furlan, A.J.N. (2018). RNA velocity of single cells *560*, 494–498.

64. Bergen, V., Lange, M., Peidli, S., Wolf, F.A., and Theis, F.J. (2020). Generalizing RNA velocity to transient cell states through dynamical modeling. Nat. Biotechnol. *38*, 1408–1414. https://doi.org/10.1038/s41587-020-0591-3.

65. Aibar, S., González-Blas, C.B., Moerman, T., Huynh-Thu, V.A., Imrichova, H., Hulselmans, G., Rambow, F., Marine, J.-C., Geurts, P., and Aerts, J. (2017). SCENIC: single-cell regulatory network inference and clustering *14*, 1083–1086.

66. Zhou, Y., Zhou, B., Pache, L., Chang, M., Khodabakhshi, A.H., Tanaseichuk, O., Benner, C., and Chanda, S.K. (2019). Metascape provides a biologist-oriented resource for the analysis of systems-level datasets *10*, 1523.

67. Subramanian, A., Tamayo, P., Mootha, V.K., Mukherjee, S., Ebert, B.L., Gillette, M.A., Paulovich, A., Pomeroy, S.L., Golub, T.R., and Lander, E.S. (2005). Gene set enrichment analysis: a knowledge-based approach for interpreting genome-wide expression profiles *102*, 15545–15550.

68. Shannon, P., Markiel, A., Ozier, O., Baliga, N.S., Wang, J.T., Ramage, D., Amin, N., Schwikowski, B., and Ideker, T. (2003). Cytoscape: a software environment for integrated models of biomolecular interaction networks *13*, 2498–2504.

69. Wolock, S.L., Lopez, R., and Klein, A.M. (2019). Scrublet: computational identification of cell doublets in single-cell transcriptomic data. Cell Syst. *8*, 281–291.e9.

70. Haghverdi, L., Lun, A.T.L., Morgan, M.D., and Marioni, J.C. (2018). Batch effects in single-cell RNA-sequencing data are corrected by matching mutual nearest neighbors. Nat. Biotechnol. *36*, 421–427. https://doi.org/10.1038/nbt.4091.

71. Kanton, S., Boyle, M.J., He, Z., Santel, M., Weigert, A., Sanchís-Calleja, F., Guijarro, P., Sidow, L., Fleck, J.S., Han, D., et al. (2019). Organoid single-cell genomic atlas uncovers human-specific features of brain development. Nature *574*, 418–422. https://doi.org/10.1038/s41586-019-1654-9.

72. Kim, J.B., Greber, B., Araúzo-Bravo, M.J., Meyer, J., Park, K.I., Zaehres, H., and Schöler, H.R. (2009). Direct reprogramming of human neural stem cells by OCT4. Nature *461*. 649-643. https://doi.org/10.1038/nature08436.

73. Aaku-Saraste, E., Hellwig, A., and Huttner, W.B. (1996). Loss of occludin and functional tight junctions, but not ZO-1, during neural tube closure–remodeling of the neuroepithelium prior to neurogenesis. Dev. Biol. *180*, 664–679. https://doi.org/10.1006/dbio.1996.0336.

74. Malatesta, P., Appolloni, I., and Calzolari, F. (2008). Radial glia and neural stem cells. Cell Tissue Res. *331*, 165–178. https://doi.org/10.1007/s00441-007-0481-8.

75. Götz, M., and Huttner, W.B. (2005). The cell biology of neurogenesis. Nat. Rev. Mol. Cell Biol. *6*, 777–788. https://doi.org/10.1038/nrm1739.

76. Castro, D.S., Martynoga, B., Parras, C., Ramesh, V., Pacary, E., Johnston, C., Drechsel, D., Lebel-Potter, M., Garcia, L.G., Hunt, C., et al. (2011). A novel function of the proneural factor Ascl1 in progenitor proliferation identified by genome-wide characterization of its targets. Genes Dev. *25*, 930–945. https://doi.org/10.1101/gad.627811.

77. Shi, Y., Wang, M., Mi, D., Lu, T., Wang, B., Dong, H., Zhong, S., Chen, Y., Sun, L., Zhou, X.J.S., et al. (2021). Mouse and human share conserved transcriptional programs for interneuron development. Science *374*, eabj6641.

78. Macosko, E.Z., Basu, A., Satija, R., Nemesh, J., Shekhar, K., Goldman, M., Tirosh, I., Bialas, A.R., Kamitaki, N., Martersteck, E.M., et al. (2015). Highly parallel genome-wide expression profiling of individual cells using nanoliter droplets. Cell *161*, 1202–1214. https://doi.org/10.1016/j.cell.2015.05.002.

79. Van de Sande, B., Flerin, C., Davie, K., De Waegeneer, M., Hulselmans, G., Aibar, S., Seurinck, R., Saelens, W., Cannoodt, R., Rouchon, Q., et al. (2020). A scalable SCENIC workflow for single-cell gene regulatory network analysis. Nat. Protoc. *15*, 2247–2276. https://doi.org/10.1038/s41596-020-0336-2.

## STAR★METHODS

### KEY RESOURCES TABLE

| REAGENT or RESOURCE | SOURCE | IDENTIFIER |
|---|---|---|
| Antibodies | | |
| Goat Polyclonal anti-SOX2 | Santa Cruz Biotechnology | Cat# sc-17320; RRID: AB_2286684 |
| Rabbit Polyclonal anti-NOTCH1 | Santa Cruz Biotechnology | Cat# sc-9170, RRID:AB_650334 |
| Mouse Monoclonal anti-ZO-1 | Innovative Research | Cat# 33-9100, RRID:AB_87181 |
| Mouse Monoclonal anti-CDH2 | BD Biosciences | Cat# 610920, RRID:AB_2077527 |
| Mouse Monoclonal anti-P-VIM | MBL International | Cat# D095-3, RRID:AB_592969 |
| Rabbit Polyclonal anti-PAX6 | BioLegend | Cat# 901301, RRID:AB_2565003 |
| Rabbit Polyclonal anti-NKX2-1 | Abcam | Cat# ab86023, RRID:AB_2043125 |
| Mouse anti-PAX3 | DSHB | Cat# Pax3, RRID:AB_2315059) |
| Mouse Monoclonal anti-ASCL1 | BD Biosciences | Cat# 556604, RRID:AB_396479 |
| Rabbit Monoclonal anti-LHX2 | Abcam | Cat# ab184337, RRID: AB_2916270 |
| Goat Polyclonal anti-OTX2 | R and D Systems | Cat# AF1979, RRID: AB_2157172 |
| Mouse Monoclonal anti-CRABP1 | NOVUS | Cat# NB300-539, RRID:AB_2085316 |
| Goat anti-PAX2 | R and D systems | Cat# AF3364, RRID:AB_10889828 |
| Hibernate E medium | Invitrogen | Cat# A1247601 |
| Rabbit Polyclonal anti-LMX1A | Invitrogen | Cat# PA5-34470, RRID:AB_2551822 |
| Collagenase IV | Gibco | Cat# 17104-019 |
| Papain | Sigma | Cat# P4762 |
| DNase I | NEB | Cat# M0303L |
| Triton X-100 | Sigma-Aldrich | Cat# T9284 |
| Phosphate Buffered Saline (PBS) | Gibco | Cat# 10010002 |
| Tissue-Tek O.C.T. Compound | SAKURA | 4583 |
| Mayer's Hematoxylin (Lillie's Modification) Histological Staining Reagent | Dako | S3309 (LOT#11120782) |
| Bluing Buffer | Dako | CS702 (LOT#18021684) |
| Eosin Y-solution, 0.5% aqueous | Sigma | 1098442500 (LOT#HX85174144) |
| Critical Commercial Assays | | |
| 10x Chromium Single Cell 3' Library Kit v2 | 10X Genomics | Cat# PN-120234 |
| 10x Chromium Single Cell 3' Gel Bead Kit v2 | 10X Genomics | Cat# PN-120235 |
| Chromium Single Cell 3' Chip Kit v2 | 10x Genomics | Cat# PN-120236 |
| Chromium i7 Multiplex Kit | 10x Genomics | Cat# PN-120262 |
| 10x Chromium Single Cell 3' GEM, Library & Gel Bead Kit v3 | 10X Genomics | Cat# PN-1000075 |
| Chromium Chip B Single Cell Kit | 10X Genomics | Cat# PN-1000074 |
| Visium Spatial Tissue Optimization Reagent Kit | 10x Gemonics | 1000192 (LOT#154747) |
| Visium Spatial Tissue Optimization Slide Kit | 10x Gemonics | 1000191 (LOT#190827) |
| Visium Spatial Gene Expression Reagent Kit | 10x Gemonics | 1000186 (LOT#154992) |
| Visium Spatial Gene Expression Slide Kit | 10x Gemonics | 1000185 (LOT#155022) |
| Visium Accessory Kit | 10x Gemonics | 1000194 |
| Library Construction Kit | 10x Gemonics | 1000190 (LOT#154997) |
| Dual Index Kit TT Set A | 10x Gemonics | 1000215 (LOT#154902) |
| KAPA SYBR FAST Universal | KAPA Biosystems | KK4601 (LOT#0000092785) |
| Deposited Data | | |
| Human early embryo single-cell RNA-seq data | This paper | GEO: GSE155121 |
| Mouse early embryo single-cell RNA-seq data | This paper | GEO: GSE155121 |
| Human spatial transcriptomics data | This paper | GEO: GSE155121 |

(*Continued on next page*)

*Continued*

| REAGENT or RESOURCE | SOURCE | IDENTIFIER |
|---|---|---|
| Single cell atlas of human gastrulation at CS7 | Richard et al., 2021[26] | E-MTAB-9388 |
| Mouse gastrulation single-cell RNA-seq data | Pijun-Sala et al.[1] | GEO: GSE87038 |
| Software and Algorithms | | |
| CellRanger v.2.0.2 and v.3.0.2 | 10X genomics | https://support.10xgenomics.com/single-cell-gene-expression/software/pipelines/latest/installation |
| Spaceranger v 1.0.0 | 10X genomics | https://support.10xgenomics.com/single-cell-gene-expression/software/downloads/latest |
| R v.3.5.1 and v3.6.3 | R Core Team | https://www.R-project.org/ |
| Seurat v3.1.1 | Stuart et at.[27] | https://satijalab.org/seurat/install.html |
| ggplot2 v3.3.2 | Hadley et al., 2016 | https://cloud.r-project.org/package=ggplot2 |
| scran v1.10.2 | Lun et al., 2016[61] | https://bioconductor.statistik.tu-dortmund.de/packages/3.8/bioc/html/scran.html |
| scanpy v1.4.2 | Wolf et al., 2018[62] | https://github.com/theislab/scanpy |
| cellchat v1.1.0 | Jin et al.[34] | https://github.com/sqjin/CellChat |
| anndata v0.6.20 | N/A | https://anndata.readthedocs.io/en/latest/ |
| python-igraph v0.7.1 | N/A | https://igraph.org/python/ |
| louvain v0.6.1 | N/A | https://github.com/taynaud/python-louvain |
| umap-learn v0.3.8 | N/A | https://github.com/lmcinnes/umap |
| bbknn v1.4.0 | N/A | https://github.com/Teichlab/bbknn |
| pandas v0.25.0 | N/A | github.com/pandas-dev/pandas |
| scipy v1.2.1 | N/A | https://www.scipy.org/ |
| scikit-learn v0.21.3 | N/A | http://scikit-learn.org/ |
| matplotlib v3.1.1 | N/A | github.com/matplotlib/matplotlib |
| numpy v1.17.4 | N/A | https://numpy.org/ |
| holoviews v.14.3 | N/A | https://holoviews.org/ |
| plotly v4.14.3 | N/A | https://plotly.com/python/ |
| rpy2 v3.2.4 | N/A | https://rpy2.github.io/doc/latest/html/index.html |
| anndata2ri v1.0.5 | N/A | https://github.com/theislab/anndata2ri |
| velocyto v 0.17.17 | Gioele et al., 2018[63] | http://velocyto.org/ |
| scVelo v0.2.2 | Vergen et al.[64] | https://scvelo.readthedocs.io/installation.html |
| pySCENIC v0.10.3 | Alibar et al., 2017[65] | https://pyscenic.readthedocs.io/en/latest/installation.html |
| Tangram v1.0.2 | Viancalani et al.[38] | https://github.com/broadinstitute/Tangram |
| metascape | Zhou et al., 2019[66] | https://metascape.org/ |
| GSEA v4.1.0 | Subramanian et al., 2005[67] | http://software.broadinstitute.org/gsea/index.jsp |
| Cytoscape v3.8.2 | Paul et al., 2003[68] | https://cytoscape.org/ |
| ImageJ (1.5.3a) | NIH | https://imagej.nih.gov/ij/ |
| Adobe Photoshop CC2019 | Adobe Systems | http://www.adobe.com |
| Adobe Illustrator CC2019 | Adobe System | http://www.adobe.com |

## RESOURCE AVAILABILITY

### Lead contact

Further information and requests for resources and reagents should be directed to and will be fulfilled by the lead contact, Xiaoqun Wang (xiaoqunwang@bnu.edu.cn).

### Materials availability

This study did not generate new unique reagents.

### Data and code availability

Sequencing data have been deposited at the National Center for Biotechnology Information BioProjects Gene Expression Omnibus (GEO) under accession number GEO: GSE155121. There is no new code generated in this study. To better share this valuable dataset with worldwide researchers, we have built an interactive website that can be explored at: http://wanglaboratory.org:3838/hwb/. Further information and requests for data and code should be directed to and will be fulfilled by the lead contact.

## EXPERIMENTAL MODEL AND STUDY PARTICIPANT DETAILS

### Human sample collection

Human de-identified tissue samples were collected after the donor patients signed an informed consent document that was in strict observance of the legal and institutional ethical regulations. Protocols were approved by the institutional review board (ethics committee) of the Institute of Biophysics, CAS, in compliance with the Interim Measures for the Administration of Human Genetic Resources, administered by the Ministry of Science and Technology of China.

### Animals

Timed pregnant female mice at embryonic days 9.5 and 10.5 were used to collect embryos. All these embryos both included male and female. The protocols used for mouse housing and experiment were approved by the Institutional Animal Care and Use Committee of the Institute of Biophysics, CAS. Food and water were accessible when the mice were housed in the institutional animal care facility with a 12-h light-dark schedule.

## METHOD DETAILS

### Sample dissociation

The age of human embryo and fetal samples was calculated as post-conception weeks. Embryo samples were collected in ice-cold artificial cerebrospinal fluid containing 125.0 mM NaCl, 26.0 mM $NaHCO_3$, 2.5 mM KCl, 2.0mM $CaCl_2$, 1.0 mM $MgCl_2$, 1.25 mM $NaH_2PO_4$ at a pH of 7.4 when oxygenated (95% $O_2$ and 5% $CO_2$). After washing the samples with hibernate E medium (Invitrogen, Cat. A1247601) three times, the samples were digested in 2 mg/ml collagenase IV (Gibco, Cat. 17104-019), 1 mg/ml papain (Sigma, Cat. P4762) and 10 U/μl DNase I (NEB, Cat. M0303L) in hibernate E medium for 20 min. During digestion, samples were vortexed at 300g and 37°C on a thermocycler. Further pipetting was used to digest the tissue fully into single cells. After digestion, the cell suspension was centrifuged at 500g for 5 min to obtain the cell pellet. After removing the digestion medium, the cell pellet was resuspended in 0.04% BSA in PBS and kept on ice. The mouse embryos were dissociated with the same dissociation protocol described above.

### RNA library preparation for high-throughput sequencing

All the libraries were prepared on a 10X GENOMICS platform following the RNA library preparation protocols. Thousands of cells were partitioned into nanoliter-scale Gel Bead-In EMulsions (GEMs) using 10x GemCode Technology, where a common 10x Barcode labeled all the cDNA produced from the same cell. Primers containing an Illumina R1 sequence (read1 sequencing primer), a 16-bp 10x Barcode, a 10-bp randomer, and a poly-dT primer sequence were released and mixed with cell lysate and Master Mix upon dissolution of the single cell 3′ gel bead in a GEM. The GEMs were incubated, and barcoded, full-length cDNA was generated from poly-adenylated mRNA by reverse transcription. Then the GEMs were broken, and the leftover biochemical reagents and primers were removed with silane magnetic beads. Before constructing the library, the cDNA amplicon size was optimized by enzymatic fragmentation and size selection. After that, P5, P7, a sample index, and R2 (read 2 primer sequence) were added to each selected cDNA during end repair and adaptor ligation. P5 and P7 primers were used in Illumina bridge amplification of the cDNA (http://10xgenomics.com). Finally, 150-bp paired-end reads were sequenced from each library using the Illumina HiSeq4000. Concerning the update of the Single Cell 3' Regent Kits version, two libraries were prepared with Single Cell 3' Regent Kits V3, while the others were prepared with Single Cell 3' Regent Kits V2.

### Processing of scRNA-seq data from Chromium system

The sequence data were mapped to the hg19 human genome and mm10 mouse genome to perform quality control and the read counting of Ensemble genes using Cell Ranger (v.2.0.2 and v.3.0.3) (http://10xgenomics.com) with default parameters. The gene-cell sparse matrix was generated for each sample by Cell Ranger software.

### Processing scRNA-seq data for all human samples

All 14 human embryo samples from PCW3 to 12 were integrated and analyzed following the standard pipeline of the Scanpy package (v.1.4.2). In brief, we concatenated the count matrices from all the samples and the merged matrix contains 463,304 cells and 32,738 genes. We only keep good quality cells that meet the following criteria: 1) cells with a number of detected genes greater than 200; 2) cells with between 800 and 6000 genes expressed; 3) cells with mitochondrial gene expression percentages fewer than 20. Genes expressed in less than 20 cells were removed, but we did not set this number too high to avoid restricting rare cell type detection. To detect potential doublets, the Scrublet (v.0.2.1)[69] pipeline was performed on each sample by setting parameters 'n_prin_comps=30',

'expected_doublet_rate=0.06', and 'sim_doublet_rate=20'. A total of 9,847 cells with doublet scores greater than the threshold were identified as doublets and excluded from subsequent analysis. Overall, 430,808 single cells covering 26,434 genes were retained. Next, the filtered gene expression matrix was normalized and log-transformed. Highly variable genes were calculated and used to perform principal component analysis (PCA). Similar to the study from Pijuan-Sala and colleagues,[1] the batch effect correction was conducted on the 50 PCs with function fastMNN[70] from the R Scran package (v.1.10.2), the PCs correction was performed between all the donors/samples, the batch-corrected PCs replaced the non-batch-corrected ones and were used for further analysis such as the nearest-neighbor graphs. Finally, the neighborhood graph computed from the *pp.ngighbors* function was utilized for unsupervised clustering performed by the Louvain algorithm.

### Identification of differentially expressed genes

The differentially expressed genes among the cluster were identified using either the *tl.rank_genes_groups* function in Scanpy (v1.4.2) or the *FindAllMarkers* function in Seurat (v3.1.1). Genes with FDR-corrected p-value < 1e-2 and log fold change > 0.25 were considered significantly high in that cluster. The enriched GO terms of biological processes for the DEGs were identified from the Metascape website (http://metascape.org).

### Systematic reconstruction of the cellular trajectories

The cellular trajectories spanning human gastrulation development from CS7 to CS10 were reconstructed using the TOME (Trajectories Of Mouse Embryogenesis) approach.[28] In brief, to connect each cell state observed at CS10 with its pseudo-ancestors at CS7, we calculated the Euclidean distances between individual cells from CS7 and CS10 using the co-embedding space of the merged dataset, then for cells of each cell state at CS10, the five closest neighbors and the fraction of these neighbors derived from the cell states of CS7 were identified. Edge weights between cell states from CS10 and CS7 were determined by taking the median fractions. Only edge weights greater than 0.4 were retained for the resulting acyclic-directed graph. For clusters within CS7, the edge weights between epiblast and primitive streak, primitive streak and nascent, emergent, advanced mesoderm, and endoderm were calculated, we also calculated the connecting edges between clusters within CS10.

### Pseudo-space definition

In Figure 3C, we used the definition of pseudo-space similar to the study of Nowotschin and colleagues,[4] defining the rostral-caudal pseudo-space ordering of the neural tube cells through the following steps. First, using a diffusion map, we co-embedded the neural tube cells into a low-dimensional space. From the visualization of the first two most informative components of the diffusion map, we can observe a precise alignment of the neural tube cells from the rostral to the caudal ends. Second, to reach a more robust estimate of the rostral-caudal ordering of cells, we computed the multi-scale distances from the most rostral cell in the first diffusion component using ten components. The multi-scale distance between each cell and the most rostral cell was calculated with the following:

$$MSD(c_i, c_0)^2 = \sum_{n=1}^{N} \left(\frac{\gamma}{1-\gamma}\right)^2 \left(c_i^{(n)} - c_0^{(n)}\right)^2$$

Where $c_i$ and $c_0$ represent cell i and the most rostral cell respectively, N represents the number of the dimensions considered, and $\gamma$ is the corresponding eigenvalues from the calculated eigenvectors of the diffusion maps. Lastly, we assigned the multi-scale distances calculated above to each neural tube cell. To further characterize whether these multi-scaled distances represent the real rostral-caudal ordering, we checked the expression pattern of the genes with known spatial rostral-caudal locations along the multi-scaled distances ordering. The gene list used was verified with spatial location expression in this study: *LHX2, FEZF1, SIX3*, and *EMX2* in the rostral forebrain, *FGF17* and *PAX7* in the midbrain, *OTX2* in both the forebrain and midbrain, *CRABP1, EGR2, MAFB*, and *HOXA2* in the hindbrain and *HOXA5, HOXA7*, and *HOXB9* in the caudal spinal cord (Figures 3D and S3A). The gene expression with known location shows remarkable consistency with the multi-scaled distances ordering, therefore, we define the multi-scaled distances as the rostral-caudal pseudo-space.

### Cell-cell communication analysis

To study the cell-cell communication between neural and non-neural ectoderm cells at PCW3, CellChat (v1.1.0) was applied to infer the ligand-receptor pairs between cell types.[34] In brief, gene expression data of cells and assigned cell types were used as input for CellChat. First, overexpressed ligands or receptors in one cell group were identified, and then gene expression data were projected onto the protein-protein interaction network. The biologically significant cell-cell communication pathways were identified with a probability value and performed a permutation test. The used human database contains 1,939 ligand-receptor (LR) interactions, including 'Secreted Signaling', 'ECM-Receptor', and 'Cell-Cell Contact' interactions. To obtain strong signaling pathways, we set the parameter of 'trim' to be 0.4 for the *ComputeCommunProb* function with the argument 'type=truncatedMean' to calculate the communication probabilities between the interesting cell types.

### Classification of NE and RG cells

For the NE and RG classification, we first reanalyzed and subclustered the progenitor cells for each region separately, then separated the NE and RG cell for the brain progenitors (forebrain, midbrain, and hindbrain) using the NE markers (POU5F1/OCT4, EPCAM, and

CLDN6) and RG markers (CDH2, NES, VIM, and FABP7). The NE cells in the brain also named neural stem cells are present in the neural plate or neural tube at the early embryonic stage, suggesting higher stemness or multipotent capacity (POU5F1/OCT4) than RG cells.[71,72] In addition, the epithelium-related protein expression such as tight junction (EPCAM, Claudins) was downregulated during the NE to RG transformation, in contrast, the upregulation of N-cadherin (CDH2).[16,73,74] In comparison to NE cells, RG cells upregulated the glial-specific genes such as the intermediate filament vimentin (VIM),[74] nestin RC2 (NES),[16,39,73] and the brain-lipid-binding protein BLBP (FABP7).[39,75]

### Identification of signaling pathways regulating early neural stem cell transformation

To identify critical genes or signaling pathways regulating neural stem cell differentiation during the epiblast-NE-RG transformations, the gene set enrichment analysis (GSEA v4.1.0) was performed to identify gene sets with statistically significant differences. Then the gene sets of KEGG pathways and gene ontology (Biological Process) were tested in the Molecular Signature Database (MSigDB). Gene sets enriched in several signaling pathways were identified, including the tight junction, EMT, BMP, and NOTCH signaling pathways. For each of the CNS regions of the forebrain, midbrain, and hindbrain, the average gene expression was smoothed across cell states during epiblast-NE-RG transformations for the gene sets from each signaling pathway, as well for the single representative gene.

### RG diversity analysis and cell fate lineages determination

To analyze the human early central nervous system development, we firstly selected all the neural-related clusters 31 to 85 from the whole dataset in Figure 1B, then dropped the neural crest clusters 52 and 53, and sensory neuron clusters 70 and 71. Subsequently, we reanalyzed the retained cells from raw data by running the analysis pipeline in Scanpy, including normalization, cell cycle regression, dimension reduction, and batch correction. Finally, the cells were projected onto a low-dimensional embedding using the force-directed graph layout. Except for the *EOMES*$^{+}$ IPC cells in Figure 1, here we further noticed some cells positioned between neural progenitors and neurons as well as the high expression of *ASCL1*, which was reported as a proneural gene expressed in neural progenitor cells and promote cell cycle exit and neuronal differentiation.[76,77] Thus, we defined the three main cell types of RG, IPC/TACs and Neurons, which highly express (*SOX2, VIM*), (*EOMES* or *ASCL1*), and (*DCX, STMN2*), respectively. Next, the Louvain sub-clustering was performed for RG, IPC/TACs and neurons, and we obtained 24 RG, 5 IPC/TACs, and 17 Neuron subtypes. According to the differential express genes, three of the IPC/TACs subclusters were defined as IPC (the subcluster expresses *ASCL1* and *NKX2-1* was named IPC1 (sub-pallium IPC),[77] the other two subclusters express *EOMES* and the pallium marker *EMX1* were named IPC2-3 (pallium IPC)). The other two non-telencephalic subclusters lacking the *FOXG1* expression were named TACs. The correlations of gene expression between the RG, IPC/TACs, and neuron subclusters were calculated with the Pearson correlation function in R. Then the Sankey diagram was constructed using the *sankeyNetwork* function in the R package networkD3, and only correlation values greater than 0.3 were retained, and the lines were not shown between RG and neuron pairs if the IPC/TACs cluster connected them robustly. We have constructed the directed differentiation trajectory using both the velocity and the correlation analysis, this is because of the advantage and limitations of these methods. The velocity analysis can show us the general main cluster differentiation direction. However, this method presents a resolution limitation when too many clusters were analyzed simultaneously, especially in our dataset with more than 200, 000 cells and 46 clusters. In contrast, the correlation analysis exhibits the advantage of high resolution which can separate the differentiation paths from numerous clusters, although this method has the limitation of unknown differentiation direction.

### Gene module enrichment analysis

For the gene module enrichment analysis presented in Figure 5M, we first calculated the differential express genes for the oRG cells over the vRG cells. Second, we input the DEGs into the STRING database (v.11.5) (https://cn.string-db.org/) and construct the protein interaction network, and the vertices in the network represent the proteins, and after removing those proteins without interaction, we performed the K-means clustering of the network proteins according to the protein interacting scores, and we obtained 4 functional similar gene modules, then we performed the biological enrichment analysis for each gene module.

### Cell cycle regression

To minimize the effect of cell cycle heterogeneity, we first downloaded the cell cycle gene list,[78] then the cell cycle scores for every single cell were calculated with the *score_genes_cell_cycle* function in Scanpy, and the computed *S_score* and *G2M_score* for each cell were stored in the metadata. Finally, before the PCA calculation and batch effect correction steps, the data variation caused by the cell cycle was regressed with the *regress_out* function in Scanpy.

### RNA velocity analysis

We ran the velocyto v0.17.17 using the *run10x* mode to process the aligned bam files outputting from Cell Ranger, the count matrix made of spliced and unspliced read counts was created in the output loom file. Next, the merged spliced/unspliced counts object was further merged with the transcriptional object using the *scv.utils.merge* function in the scVelo v0.2.2 package.[64] To process the data, we run the *scv.pp.filter_and_normalize* function. For the genes used for velocity calculation, we used the default parameter to calculate the top 2,000 highly variable genes that passed a minimum threshold of 20 expressed counts. After running the *scv.pp.comments* function, the nearest model graph (with 30 neighbors) was calculated based on Euclidean distances in the 30 principal

components on spliced counts. The stochastic model was selected for velocity estimation by running the *scv.tl.velocity.graph*. As a result, transition probabilities were estimated to form a velocity graph. Finally, we embedded the resulting velocities on the low dimensional space using the *velocity_embedding_stream* function.

### Spatial transcriptomics

Fresh human PCW4 tissue was embedded in Optimal Cutting Temperature (SAKURA) and frozen in a dry ice-ethanol mixture. The frozen tissue block was cut into 16 μM sections in a cryostat (Lecai CM3050 S) and mounted on Tissue Optimization Slides and Gene Expression Slides (10x Genomics).

Permeabilization time was optimized following the manufacturer's instructions (10x Genomics, Visium Spatial Tissue Optimization, CG000238 Rev D), and 10 minutes of permeabilization time was chosen for the Visium Spatial Gene Expression workflow. H&E-stained tissue sections were prepared following the manufacturer's instructions (10x Genomics, Methanol Fixation, H&E Staining & Imaging for Visium Spatial Protocols, CG000160 Rev B). H&E images and Tissue optimization images were taken using an Olympus FV3000 imaging system with an Olympus DP80 CCD camera and an Olympus 10X/0.40 objective.

Tissue slices were then processed for gene expression following the manufacturer's instructions (10x Genomics, Visium Spatial Gene Expression Reagent Kits, CG000239 Rev D).

In brief, H&E-stained tissue sections were permeabilized for 10 minutes, followed by reverse transcription, second-strand synthesis, and denaturation. qPCR experiment was processed using KAPA SYBR FAST kit (KAPA Biosystems) and QuantStudio 6 Flex system (ThermoFisher). The cDNA amplification cycle number was determined by ~25% of the peak fluorescence value. The final libraries were processed on the Illumina HiSeq Xten system for sequencing 150 bp pair-end reads.

Space Ranger (version 1.0.0) software from 10X Genomics was used to perform process, alignment, and barcode/UMI counting against the human hg38 reference genome for each spot on the Visum spatial transcriptomic array. The raw UMI count matrix, images, spot-image coordinates, and scale factors were imported to Seurat (v3.1.1). In brief, the *SCTransform* function was performed to normalize expression values for total UMI count per cell, following the dimensionality reduction and clustering with the *RunPCA*, *FindNeighbors*, *FindClusters*, and *RunUMAP* functions. The *SpatialDimPlot* was applied to visualize the spatial gene expression from the 10x Visum assay overlaid on the tissue image. To predict the spatial locations of the PCW3 clusters, we first normalized the scRNA-seq data using the *SCTransfrom* function, then the cluster prediction probability on each spot was calculated using factor analysis via *FindTransferAnchors* and *TranserData* functions.[27] Therefore, we get the cluster compositions for each spatial spot and visualize the cluster compositions in the spot with a pie graph.

### Mapping the single-cell and spatial transcriptomics data with Tangram

Tangram was applied to map the single-cell gene expression data from the neural system onto spatial transcriptional data to resolve cell types in the space.[38] Firstly, a subset of the differentially expressed genes across cell types for the single-cell dataset was calculated and treated as the training genes. The average gene expression on the training gene of the mapped cells vs spatial data was calculated, and the optimal spatial alignment for scRNA-seq profiles was performed with the *map_cells_to_space* function. The mapping mode was set to 'mode=clusters' and 'num_epochs=500', and run Tangram with CPU by setting 'devise=cpu'. The *project_cell_annotation* function was performed to transform the single-cell cluster annotations to space, and the *plot_cell_annotation* function with 'perc=0.02' was applied to visualize the annotation in space. The cluster prediction scores for the spatial dataset from Tangram were also imported into Seurat for visualization with the function of *SpatialDimPlot*.

### Cross-species transcriptome comparison

To investigate the developmental similarities and differences between humans (PCW3, PCW4-1) and mice (E8.5, E9.5 and E10.5) from the aspect of single-cell transcriptome profiles, we performed the canonical correlation analysis (CCA) algorithm to integrate human and mouse whole embryo datasets with function *IntegrateData* and obtained the new integrated expression matrix, the mouse E8.5 dataset was from the published mouse embryo single-cell RNA-seq dataset.[1] Cell clustering and dimensionality reduction of the integrated matrix were performed with the functions *FindClusters* and *RunUMAP*, respectively. We used the R package river plot to compare the cluster assignments and visualize the cell type assignments for human and mouse datasets with CCA joint clusters. The integration and analysis for the human and mouse neural system cells were computed similarly.

### Cross-species regulatory network comparison

Human and mouse gene regulatory network analysis was performed using the pyscenic CLI pipeline with default parameters, respectively (https://github.com/aertslab/pySCENIC).[79] To determine the 'on/off' state of each regulon in each cell, we used binarize function from pyscenic with default parameters. We obtained a total of 542 human and 615 mouse regulons for whole-body data, among which we defined 386 orthologous transcription factor (TF) regulons. We obtained a total of 539 human and 572 mouse regulons for neural tube cells, among which we defined 332 orthologous TF regulons. We used Cytoscape to visualize our regulon network for our interested TFs, with weights being normalized by z-score transformation for comparability between human and mouse node size. Based on orthologous TF regulons, we obtained a merged AUC score matrix from the human and mouse AUC score matrix. Then the Pearson Correlation Coefficient (PCC) of the AUC score was calculated for each pair of orthologous TF regulons. The Connection Specificity Index (CSI) was used to quantify interaction-profile similarity. The interaction profile describes the relationship between a pair of genes (regulons). As the CSI formula calculation shows below, for a fixed pair of regulons *a* and *b*, the

corresponding CSI was defined as the fraction of regulons whose PCC of AUC scores with a and b were lower than the PCC between *a* and *b* themselves. The CSI score means the fraction of other regulons that have an interaction profile similarity with *a* and *b* that is lower than the interaction profile similarity between *a* and *b*, and a high CSI score represents the high interaction profile similarity between *a* and *b*.

$$CSI_{ab} = 1 - \frac{PCC_a > PCC_{ab} + PCC_b > PCC_{ab}}{(n-1)/2}$$

From the analysis, we identified 10 major modules for neural system cells data and 12 major modules for whole-body data using 'ward.D' clustering method based on the CSI matrix. GO was conducted by Metascape with default parameters.

### Immunohistochemistry

Human brain samples were fixed in 4% paraformaldehyde (in PBS, pH=7.4) for 24 hours. After the fixation, the samples were dehydrated in 30% sucrose (in PBS, pH=7.4) at 4°C, embedded in O.C.T Compound (Thermo Scientific), and frozen at -80°C. The tissues were sectioned into 20 μm thin slices with Leica CM3050S. After antigen retrieval, slices were pretreated with PBST (0.1% Triton X-100 in PBS, pH=7.4) and incubated in a blocking solution (5% donkey serum and 0.1% Triton X-100 in PBS, pH=7.4) for 1 hour. Then the slices were treated with primary antibodies at 4°C for 12 hours. All the antibodies used were listed as follows: Goat anti-SOX2 (1:200, Santa Cruz Biotechnology, sc-17320), Rabbit anti-LHX2 (1:200, Abcam, ab184337), Goat anti-OTX2 (1:200, R and D Systems, AF1979), Mouse anti-CRABP1 (1:300, NOVUS, NB300-539), Goat anti-PAX2 (1:200, R&D, AF3364), Rabbit anti-NOTCH1 (1:200, Santa Cruz Biotechnology, sc-9170), Mouse anti-ZO-1 (1:500, Innovative Research, 33-9100), Mouse anti-CDH2 (1:500, BD Biosciences, 610920), Mouse anti-P-VIM (1:300, MBL International, D095-3), Rabbit anti-PAX6 (1:200, BioLegend, 901301), Rabbit anti-NKX2-1 (1:200, Abcam, ab86023), Mouse anti-ASCL1 (1:200, BD Biosciences, 556604), Rabbit anti-LMX1A (1:200, Invitrogen, PA5-34470). Slices were then washed three times by PBST at room temperature for ten minutes. After that, the slices were incubated with secondary antibodies for 1 hour, which were donkey anti-goat 488 (1:300), donkey anti-rabbit 594 (1:300), and donkey anti-mouse 647 (1:300). DAPI was added to the secondary incubation buffer to stain the nuclei. When the incubation was finished, slices were mounted with a Mounting medium (Sigma Aldrich) and covered by a cover glass. Immunohistochemistry images were captured with Olympus FV1000 confocal microscope. The acquired image data were analyzed with ImageJ (v1.5.3a) and Adobe Photoshop CC2019.

## QUANTIFICATION AND STATISTICAL ANALYSIS

All the data were represented as mean ± SEM. The quantification graphs were made using the ggplot2 package in R language. The number of samples used for each analysis can be found in the figure legends. The two-tailed unpaired Student's t-test was performed for the quantification statistics and the p-values below 0.01 were considered statistically significant.