

Spatiotemporal transcriptome atlas of human embryos after gastrulation

<https://doi.org/10.1038/s41586-026-10545-0>

Received: 19 August 2024

Accepted: 15 April 2026

Published online: 27 May 2026

 Check for updates

Jiexue Pan^{1,2,21}, Yuejiao Li^{3,4,5,21}, Zhongliang Lin^{6,7,8,21}, Qing Lan^{3,4,21}, Ying Zhang^{3,9,21}, Huixi Chen^{2,10,21}, Shengwei Sui^{1,21}, Man Zhai^{3,4,21}, Gaochen Zhang^{1,2,21}, Yi Cheng^{1,2}, Yunhui Tang^{1,2}, Qingchen Wang¹, Yue Xu^{1,2}, Guoling Li¹¹, Chunyan Xu^{3,4}, Haoqi Yan^{3,4,12}, Yiting Mao^{1,2,6,7,8}, Xingxia Wang^{3,4}, Hao Li^{3,4}, Yiping Zhu^{3,4,12}, Qinfang Chen¹⁰, Yichun Guan¹³, Nan Meng¹³, Chang Wang^{3,4}, Haiqian Lu¹⁰, Xiangjuan Li¹⁴, Tingting Zheng¹⁴, Xiaoying Yao¹, Tingyu Yang^{3,12}, Xuan Chen^{3,12}, Qiuyu Qin^{3,4}, Bin Jiang^{3,4}, Yuxing Ren^{3,4}, Xinmei Liu^{1,2}, Yuxin Zhang^{3,4}, Minghui Yu^{3,4}, Lifang Wang⁴, Yanrong Wei^{4,12}, Meiqi Luo^{3,4}, Ji Nancuo³, Fuhe Ma^{3,4}, Ziwei Wang³, Zhihua Ou¹⁵, Ying Lei⁴, Xin Jin³, Jianzhong Sheng^{6,7,8}, Congjian Xu¹, Yanting Wu^{1,2}, Chenming Xu^{1,2}, Lijian Zhao^{16,17}, Hongbo Yang¹, Ya Gao^{3,18,19}, Guolian Ding^{1,2,5}, Xun Xu^{3,20} & Hefeng Huang^{1,2,6,7,8,10}

The comprehensive spatiotemporal atlas of gene expression during early human embryonic development is critical for insights into embryogenesis¹, organogenesis² and disease origins^{3,4}. Here, leveraging Stereo-seq technology, we generated spatial transcriptomic profiles across 77 sagittal sections of 13 whole-human embryos ranging from Carnegie stage 12 to 23, integrated with single-nucleus RNA sequencing to elucidate gene expression patterns within defined cellular contexts, revealing the cellular heterogeneity that drives organ-specific differentiation. Our study has established a regulatory profile for the development of 50 organs and 198 substructures, and identified potential tissue-identity regulators. Of note, it uncovered previously uncharacterized gene functions in cardiac and brain development. The atlas not only substantiates and refines the current understanding of human organ development but also highlights key organs susceptible to genetic disorders. Furthermore, we characterized the allelic gene expression within specific organs at different developmental stages. This work presents a comprehensive compilation of genome-wide gene expression profiles for each spatially defined cell population, which can be visualized as a spatial display of the embryonic transcriptional landscape. These results offer the most thorough delineation to data of the spatiotemporal transcriptomic dynamics of human organogenesis.

Human embryo development begins with a fertilized ovum, followed by division and differentiation through the preimplantation, embryonic and fetal stages. After gastrulation, the embryo undergoes early organogenesis that is intricately orchestrated in a spatiotemporal manner. Many childhood and adult disorders originate from the embryonic stages, especially birth defects³ and neurodevelopmental disorders⁴. The periods most susceptible to teratogenesis range from Carnegie stage 12 (CS12) to CS23 (4–8 weeks post-conception) when the embryo

undergoes great expansion of cellular diversity and cell differentiation for organ morphogenesis. Thus, understanding the spatial and temporal expression of genes regulating embryogenesis and organogenesis is crucial in unravelling the aetiology of congenital anomalies and other pathological conditions.

Several groups have attempted to establish the gene-expression landscapes during early human embryogenesis. For instance, Cao et al. used single-cell RNA sequencing (scRNA-seq) to resolve the cellular

¹Obstetrics and Gynecology Hospital, Institute of Reproduction and Development, Fudan University, Shanghai, China. ²Shanghai Key Laboratory of Reproduction and Development, Shanghai, China. ³State Key Laboratory of Genome and Multi-omics Technologies, BGI Research, Shenzhen, China. ⁴State Key Laboratory of Genome and Multi-omics Technologies, BGI Research, Hangzhou, China. ⁵BGI College & Henan Institute of Medical and Pharmaceutical Sciences, Zhengzhou University, Zhengzhou, China. ⁶Key Laboratory of Reproductive Genetics (Ministry of Education), Department of Reproductive Endocrinology, Women's Hospital, Zhejiang University School of Medicine, Hangzhou, China. ⁷Institute of Medical Genetics and Development, Zhejiang University, Hangzhou, China. ⁸Department of Obstetrics and Gynecology, Center for Reproductive Medicine, the Fourth Affiliated Hospital of School of Medicine, and International School of Medicine, International Institutes of Medicine, Zhejiang University School of Medicine, Yiwu, China. ⁹BGI Research, Wuhan, China. ¹⁰The International Peace Maternal and Child Health Hospital, School of Medicine, Shanghai Jiao Tong University, Shanghai, China. ¹¹Zhongshan Institute for Drug Discovery, Shanghai Institute of Materia Medica, Chinese Academy of Sciences, Zhongshan, China. ¹²College of Life Sciences, University of Chinese Academy of Sciences, Beijing, China. ¹³The Third Affiliated Hospital of Zhengzhou University, Zhengzhou, China. ¹⁴Hangzhou Women's Hospital, Hangzhou, China. ¹⁵BGI Research, Beijing, China. ¹⁶BGI Genomics, BGI-Shenzhen, Shenzhen, China. ¹⁷Medical Technology College of Hebei Medical University, Shijiazhuang, China. ¹⁸Shenzhen Engineering Laboratory for Birth Defects Screening, Shenzhen, China. ¹⁹Shanxi Medical University — BGI Collaborative Center for Future Medicine, Shanxi Medical University, Taiyuan, China. ²⁰Guangdong Provincial Key Laboratory of Genome Read and Write, Shenzhen, China. ²¹These authors contributed equally: Jiexue Pan, Yuejiao Li, Zhongliang Lin, Qing Lan, Ying Zhang, Huixi Chen, Shengwei Sui, Man Zhai, Gaochen Zhang. ✉e-mail: zhaolijian@bgi.com; hongboyang@fudan.edu.cn; gaoya@genomics.cn; dingguolian@fudan.edu.cn; xunxu@genomics.cn; huanghefeng@fudan.edu.cn

heterogeneity in 15 organs in fetuses collected from 72 to 129 days of post-conception¹. Similarly, Xu et al. generated a single-cell atlas spanning 18 developmental systems in embryos 4–6 weeks of age². The emergence of spatial transcriptomic has further enabled tissue architecture-resolved gene expression profiling, with applications in gastrulation^{5,6} and fetal organs. Although these studies provide important insights into cellular heterogeneity and tissue architectures during fetal organogenesis, current spatial transcriptomic data lack whole-embryo coverage across critical post-gastrulation stages, impeding comprehensive understanding of the intercellular regulatory networks and pan-embryonic transcriptional coordination during organogenesis.

In this study, we present an integrated spatial and temporal transcriptome atlas of whole-human embryos spanning CS12–23 using the spatial transcriptome and single-nucleus RNA sequencing (snRNA-seq) technology. Our approach reveals the organ-specific differentiation trajectories as well as spatial regulatory programs underlying human embryogenesis. This resource provides a comprehensive transcriptional framework for investigating spatiotemporal dynamics and developmental disorders of human organogenesis.

Human organogenesis spatiotemporal atlas

We generated a whole-embryo spatiotemporal transcriptome atlas by profiling 77 sagittal sections from 13 euploid human embryos (7 male and 6 female individuals) spanning CS12–23 (Fig. 1a). The unsupervised spatially constrained clustering was performed with bin50 (50 × 50 DNA nanoball (DNB) bins, 25 µm) to generate a total of 14,744,703 spots (Supplementary Table 1a and Supplementary Fig. 1a,b). Fifty embryonic organs and/or anatomical regions were annotated using previously reported markers (Fig. 1b and Supplementary Table 2a). Tissue distribution analysis revealed dynamic changes across developmental stages, with early dominance (CS12–15) of the heart, spinal cord and somite, transitioning to the liver, bone, cartilage and skeletal muscle predominance (CS18–23; Fig. 1c, Supplementary Table 1b,c and Supplementary Fig. 1c). Cross-sectional and inter-embryo variability within the same gestational age was minimal (Supplementary Figs. 1d,e and 2a,b), with consistent clustering of the same organs across sections (Supplementary Figs. 1f and 2c). Benchmarking against public CS13 datasets² confirmed transcriptomic concordance and enhanced anatomical annotations (Supplementary Information and Supplementary Fig. 3). The spatial fidelity of structure clustering and annotation was confirmed by in situ hybridization (ISH) of organ-specific markers (*SHH* in the notochord, *STMN2* in the neural tube, *PMEL* in the eye, *LMX1A* in the inner ear, *ALB* in the liver, *MYL7* in the heart, *MYH6* in the atrium and *MYH7* in the ventricle), showing good consistency with the Stereo-seq data (Fig. 1d). Non-coding RNAs, particularly long non-coding RNAs, exhibited a more pronounced increase across development stages than coding RNAs (Extended Data Fig. 1a,b). Functional enrichment analysis of long non-coding RNA-associated genes revealed organ-specific developmental ontologies, indicating their role in organ-specific development processes (Extended Data Fig. 1c).

By integrating cell-type-specific and region-specific biomarkers, we mapped 198 molecularly distinct substructures across 50 organs, resolving their developmental architectures (Fig. 1e, Extended Data Fig. 2a and Supplementary Tables 2b and 3). For instance, in the eye, we delineated melanocytes (*PMEL*), retinal progenitor cells (*HES1*), neurogenic retinal progenitor cells (*ATOH7*), retinal ganglion cells (*GAP43*) and lens cells (*CRYGC*; Extended Data Fig. 2b,c). Differential gene set enrichment analysis aligned with established patterns of ocular development (Extended Data Fig. 2d and Supplementary Table 4). Likewise, in the fetal kidney, we observed compartmentalization into nephron progenitors (*EYA1*), podocytes (*NPHS1*), collecting duct (*CALB1*), distal tubule (*PAPPA2*) and early proximal tubule (*LRP2*), all spatially organized

alongside interstitial populations to establish functional nephron units (Extended Data Fig. 2e–g and Supplementary Table 4).

As a supplement to the spatial transcriptome data, snRNA-seq data of CS19, CS20 and CS23 embryos yielded a total of 607,093 cells, with approximately 2,000 genes per nucleus (Extended Data Fig. 1d and Supplementary Table 5). The merged snRNA dataset resolved 68 cell types through reference-guided annotation (Extended Data Fig. 1e and Supplementary Table 2c). Integrating the snRNA-seq data with corresponding cell-segmented spatial transcriptomics revealed good spatial patterns for different cell types, thereby supporting the subsequent analysis of substructures within individual organs and tissues (Extended Data Fig. 1f). This framework provides superior resolution for investigating human embryogenesis.

Regulatory networks of organogenesis

Our dataset provides a distinctive opportunity for investigating the dynamics of gene expression at the whole-embryo level. Systematic identification of differentially expressed genes (DEGs) revealed a set of genes exhibiting stage-specific patterns in different organs (Supplementary Table 6). Gene Ontology (GO) analysis of these temporally restricted DEGs demonstrated functional stratification along germ layer origins: endodermal derivatives, mesodermal lineages and ectodermal derivatives (Fig. 2a, Extended Data Fig. 3a–c and Supplementary Table 7). Early-stage DEGs predominantly orchestrated fundamental cellular processes, progenitor differentiation and organ morphogenesis. At later stages, the biological functions of DEGs exhibited increasing diversity, including lipid and sterol transport, muscle contraction, otic vesicle development and visual perception (Fig. 2a and Extended Data Fig. 3a–c). Of note, folic acid metabolism and neural tube development and/or closure were prominent as early as CS12–13 (Extended Data Fig. 3c), underscoring the early timeliness of folate supplementation strategy against neural tube defects.

Beyond whole-embryo transcriptional dynamics, organ-specific developmental events were also elucidated. In the liver, genes involved in the erythroid progenitor development (*ID1*)⁷ and energy metabolism (*LDHA*) peaked at CS12–13, whereas genes regulating metabolic homeostasis, such as *APOC3* in sterol transportation⁸ and *TFR3* in iron homeostasis⁹, emerged at later stages (Fig. 2a). This temporal progression was paralleled by spatial reorganization of haematopoiesis: early hepatic dominance gradually transitioned to bone tissue establishment (Extended Data Fig. 3d). In skeletal muscle development, the expression of the α -actin gene, *ACTA1*, was prominent at CS23, contrasting with early-stage dominance of the β -actin gene *ACTB* (Fig. 2a). Meanwhile, the high mobility group (HMG) gene family (*HMG1*, *HMG2* and *HMG3*) showed transient enrichment in the somite and myotome at CS12–17 but silenced thereafter (Fig. 2a). These DEGs and GO analysis collectively identified the orchestrated timeline of organ function and development.

The establishment of cellular identities hinges on orchestrated gene regulatory networks (GRNs), in which a group of transcription factors and cofactors mutually regulate one another, as well as their target genes. Leveraging these DEG and/or GO-defined developmental patterns, the spatial transcriptome data were used to explore the GRNs during organogenesis by identifying organ-specific and/or cell-type-specific regulons with the highest regulon specificity score, revealing hierarchical control architecture governing organogenesis (Supplementary Table 8 and Supplementary Fig. 4). For instance, skeletal muscle specification initiated with MYF5 activity in somite at CS12–13, followed by myotome formation at CS14–17, and later body-wide muscle differentiation (Extended Data Fig. 3e,f), suggesting that the cell fate during skeletal muscle development has been established by CS12–13. In addition, hepatic maturation involved orchestrated KLF1 and MYBL2 activation (CS17–18; Extended Data Fig. 3g,h), which coordinated erythroblast expansion^{10,11}. These processes were also dynamically regulated by multiple functional subunits and regulatory

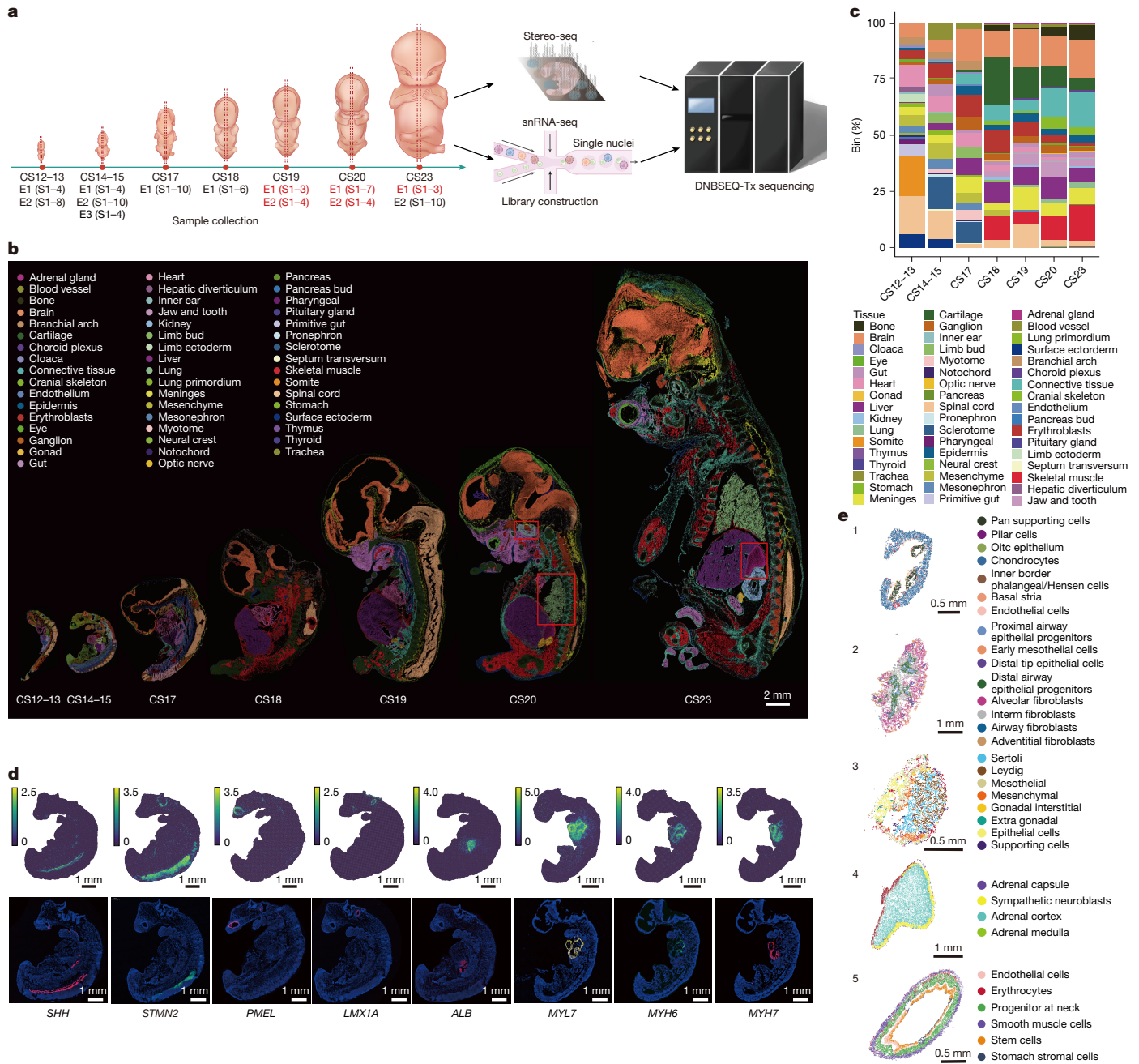


Fig. 1 | Spatiotemporal transcriptomic atlas of human organogenesis.

a, Schematic workflow of this study. Sagittal sections of human embryos across CS12–23 were included for Stereo-seq. Five embryos ranging from CS19 to CS23, highlighted in red, underwent snRNA-seq. Schematic created in BioRender; Ye, L. <https://biorender.com/igs26xe> (2026). E, embryo; S, section. **b**, Unsupervised spatially constrained clustering of human embryo sections across CS12–23. Fifty distinct anatomical tissues were identified and colour coded accordingly. **c**, The percentage of bins annotated for each tissue shows the tissue-type distribution against stages. **d**, Spatial distributions of selected

tissues were visualized in CS14–15 embryos using established tissue-specific markers (*SHH* for the notochord, *STMN2* for the neural tube, *PMEL* for the eye, *LMX1A* for the inner ear, *ALB* for the liver, *MYL7* for the heart, *MYH6* for the atrium and *MYH7* for the ventricle) on Stereo-seq maps (top) and corresponding RNA ISH images using adjacent sections (bottom). The colour bar represents log-normalized expression level. The images are representative of three independent experiments. **e**, Representative substructure mapping was shown for selected organs (the inner ear, lung, gonad, adrenal gland and stomach), and the corresponding sections of specified organs were outlined in red in panel **b**.

networks (Supplementary Fig. 4). Next, the regulons sharing similar spatial patterns at each stage were grouped into modules. Spatiotemporal modularization of co-regulated GRNs revealed increasing functional complexity, with regulon modules expanding from 19 (CS12–13) to 31 (CS23; Fig. 2b and Extended Data Fig. 4a–f). The early-stage modules were enriched in the spinal cord, brain and heart, with the functions of organogenesis, neuron differentiation and other fundamental developmental processes (Extended Data Fig. 4a), whereas later-stage modules

showed expanded tissue and/or organ enrichment governing specialized functions in neurons, hepatocytes and sensory epithelia (Fig. 2b). Integration with snRNA-seq data enabled spatial mapping of regulon modules to specific cell types, including CS23_M4/9 corresponding to cholinergic neurons and GABA neurons in the spinal cord (Fig. 2b and Extended Data Fig. 4g); also, its associated regulons incorporated the evolutionarily conserved HOX family governing the development of central nervous system and pattern specification. This comprehensive

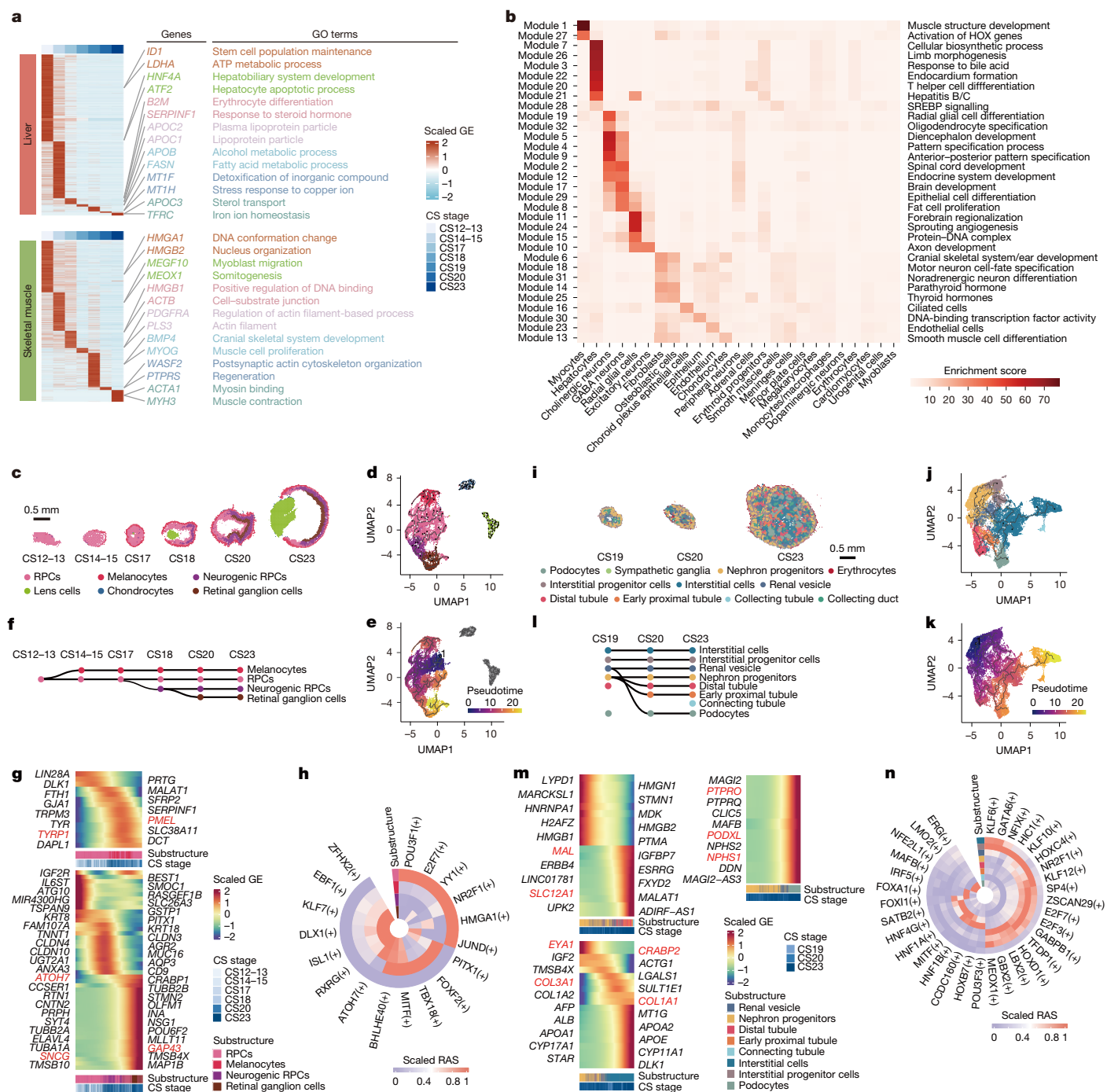


Fig. 2 | Comprehensive identification of gene regulation and substructural lineages in human organs across developmental stages. **a**, DEGs and GO terms of organs derived from the endoderm (liver) and mesoderm (skeletal muscle). GE, gene expression. **b**, Co-occurrence heatmap showing the spatial relationship between regulon modules and cell types at CS23. Each column corresponds to a cell type and each row represents a regulon module. The right side of the heatmap includes GO terms for the transcription factors within each regulon module. The value in each cell reflects an enrichment score, which quantifies the degree to which the regulon module is enriched in the respective cell type. **c**, Spatiotemporal patterning of ocular substructures mapped across developmental stages. RPC, retinal progenitor cell. **d**, **e**, Uniform manifold approximation and projection (UMAP) visualization with pseudotime trajectories (d) and developmental continuum (e) in ocular substructures.

candidate list of regulon modules (Fig. 2b, Extended Data Fig. 4 and Supplementary Table 9) would serve as a rich source for future studies of human organogenesis.

f, Lineage relationship hierarchy of ocular substructures modelled through directed acyclic graph. **g**, Pseudotemporally ordered gene expression waves of melanocytic and neurogenic lineages, with substructural marker genes highlighted in red. RAS, regulon activity score. **h**, Heatmap showing top regulons with differential activity among ocular substructures. **i**, Progressive compartmentalization of renal substructures during morphogenesis. **j**, **k**, Renal developmental trajectories reconstructed through UMAP embedding (j) and pseudotime contouring (k). **l**, Lineage progression network of renal substructural relationships. **m**, Pseudotime-resolved gene activation patterns across renal tri-lineages with annotation markers (in red). **n**, Top regulons with differential activity among renal substructures. The transcription factors marked with (+) in **h** and **n** correspond to regulons consisting of co-regulated target genes.

To further elucidate the differentiation and development lineage of organogenesis across developmental stages, the top 500 pseudotime-variable genes and top 10 substructure-specific regulons in major

organs were analysed to explore inter-substructure regulatory dynamics synchronized with the pseudotime trajectory (Supplementary Tables 10 and 11). For instance, in the ocular system, retinal progenitors diverged into melanocytic and neurogenic lineages (Fig. 2c–f). Pseudotime-dependent gene trajectories and differential regulon analysis revealed finely resolved temporal dynamics of lineage markers and lineage-specific regulons (Fig. 2g,h and Supplementary Tables 10 and 11). We observed strong concordance between transcription factor expression and regulon activity score dynamics across both pseudotime and developmental stages, including MITF in melanocytes¹², ISL1 (ref. 13) and KLF7 (ref. 14) in neurogenic lineages (Extended Data Fig. 5a–d). In the kidney, nephron progenitors exhibited tri-lineage specification: stromal (via interstitial progenitor cells to interstitial cells), nephrogenic (via renal vesicles to distal or proximal tubules) and podocyte (via renal vesicles to podocytes) lineages (Fig. 2i–l). Lineage-specific markers were identified throughout pseudotime progression, whereas key regulators also exhibited coordinated expression–regulon activation along pseudotime (Fig. 2m,n and Extended Data Fig. 5e). These included well-known regulators such as HNF1B in tubular substructures¹⁵, MAFB in podocytes¹⁶ and POU3F3 in the distal tubule¹⁷, with HOXB7 showing synchronized temporal expression–regulon activation in distal tubules, implicating its functional role in lineage commitment (Extended Data Fig. 5f–h and Supplementary Table 11).

Cardiac substructures revealed by GRNs

To elucidate organ-specific GRNs at the substructure level and delineate regulatory patterns during organogenesis, we focused on early heart development, particularly the formation of chamber structures and the cardiac conduction system. Cardiac data were extracted from 58 sections across all stages, with 26 substructures annotated using established markers (Fig. 3a and Supplementary Table 2b). DEG analysis was conducted, followed by GO analysis for each substructure (Supplementary Table 12 and Supplementary Fig. 5a,b). Within the atrium and ventricle, we identified DEGs and highlighted key genes with distinct expression profiles and stage-specific spatial patterns (Fig. 3b and Supplementary Table 13). Of note, *CPNES* and *EPHA4* were identified as potential late-stage atrium genes, whereas *NLGN1* and *FRMD5* as potential late-stage ventricle genes (Supplementary Fig. 5c), which have rarely been indicated in cardiac development before. To further explore chamber-specific regulatory mechanisms, we identified several atrium-specific or ventricle-specific regulators, including the well-known ventricle-specific transcription factors: TBX20 and ESRRG (Fig. 3c). The regulon ETS1 was spatially enriched in the inner layer of the ventricular wall, which encompasses the endocardium and trabecular myocardium (Extended Data Fig. 6a). Its activity and expression patterns colocalized with *NOTCH1*, a critical signalling axis in trabeculation initiation¹⁸. Leveraging published scRNA-seq data from the developing human heart¹⁹ and alignment with our single-cell segmented Stereo-seq data revealed prominent expression of *ETS1* in endothelial cells and endocardial cells, paralleling *NOTCH1* expression (Extended Data Fig. 6b–j). The target genes of ETS1, including *ENG* and *HEG1*, were implicated in heart trabecula morphogenesis (Extended Data Fig. 6k,l). *Ets1* deficiency in endothelial cells is known to cause ventricular noncompaction in mice²⁰, and its cooperative interaction with GATA4 in regulating endocardial development has been documented²¹. Building on these insights, our data underscored the essential function of ETS1 in modulating cardiac trabeculation in humans. In addition to transcriptional regulation, cell–cell communication analysis identified the *NRG3–ERBB2/ERBB4* pathway as a key mediator of endocardium–ventricular cardiomyocyte interaction (Extended Data Fig. 6m–p and Supplementary Information).

To delineate molecular mechanisms governing cardiac pacemaking, we focused on the sinoatrial node (SAN), an intricate structure at

the junction of the superior vena cava and right atrium that regulates heartbeats. At CS17, we precisely localized the SAN in the embryonic heart and identified a SAN-enriched gene module by Hotspot analysis (Fig. 3d,e and Supplementary Table 14). This module contained *SHOX2*, a key pacemaker cell differentiation regulator²², and *VSNL1*, a core GRN component in murine SAN function²³ (Fig. 3f,g). Meanwhile, *KIAA1324L* was identified as a novel SAN-colocalized gene (Fig. 3f). Further GRN analysis highlighted the presence of *SHOX2* and *RORA*, a circadian rhythm regulator²⁴ with previously unspecified roles in cardiac conduction, within the SAN (Fig. 3g). Downstream genes of *SHOX2* and *RORA* were also spatially expressed in the SAN region (Extended Data Fig. 6q,r). ISH confirmed the expression of *KIAA1324L* and *RORA* in the SAN, colocalizing with the pacemaker cell marker *HCN4* (Extended Data Fig. 6s). In vivo CRISPR–Cas9 editing of *SHOX2*, *VSNL1*, *KIAA1324L* and *RORA* orthologues in zebrafish (Fig. 3h) significantly reduced the pacemaker cell count in 2-day post-fertilization embryos (Fig. 3i,j) and decreased the heart rate (Fig. 3k and Supplementary Videos 1–6). These findings collectively underscored the regulatory roles of *KIAA1324L* and *RORA* in the human SAN, providing novel insights into the molecular mechanisms governing cardiac pacemaker cell function.

Central nervous system regionalization

Building on our investigation of substructure-level GRNs in the heart (Fig. 3), we also mapped regulatory networks in the developing brain to uncover the mechanisms driving their formation and specialization. Current understanding of early human neural development remains limited, especially regarding the *in situ* molecular mechanisms governing the formation and differentiation of the circuits and functional brain substructures. To address this gap, we extracted the nervous system data from 65 sections of different stages and annotated brain substructures using reported markers (Fig. 4a, Extended Data Fig. 7a and Supplementary Table 2b). The annotated regions and substructures increased from 5 (CS12–13) to 12 (CS23; Extended Data Fig. 7b), revealing progressive compartmentalization. Transcriptomic similarity among brain substructures during CS12–17 was high but divergence intensified after CS18 (Extended Data Fig. 7c), coinciding with pathway activation for synaptic connections, neural development, perception and signal transduction (Extended Data Fig. 7d and Supplementary Table 15). Cellular diversification during this period featured radial glial cells (RGCs) depletion concurrent with the expansion of inhibitory and excitatory neuronal populations (Fig. 4b, Extended Data Fig. 7e,f, Supplementary Table 2d and Supplementary Fig. 6). Of note, inhibitory neuron markers were detectable as early as CS12–13 in the ganglionic eminence (that is, subpallium), preceding excitatory neuron specification (CS19) in the Pall region of the forebrain, as confirmed by ISH validation (Fig. 4b and Extended Data Fig. 7g). Meanwhile, cross-validation with the external dataset also confirmed the early presence of inhibitory neuron markers and delayed excitatory neuron differentiation²⁵ (Extended Data Fig. 7h–m). It is noteworthy that this observation renewed the timeline of neurogenesis, which has been previously reported that the GABA immunoreactive cells emerged at 6 weeks after conception (CS18)²⁶ and the excitatory neurons appeared between post-conception days 50 and 51 (CS20)²⁶.

We then analysed the expression of transcription factors and regulons in the developing nervous system and identified eight early regulons specifically enriched during CS12–17 and five late regulons enriched during CS18–23 (Fig. 4c and Extended Data Fig. 8a). Of particular interest, HMGA2 exhibited stage-specific regulatory dynamics, transitioning from pan-neural expression with low-regulon activity (CS12–17) to spatially restricted enrichment in the ventricular zone, particularly the pallidum ventricular zone (Pall VZ), accompanied by prominent regulon activity during CS18–23 (Fig. 4d,e and Extended Data Fig. 8b,c). Target prediction and functional annotation revealed

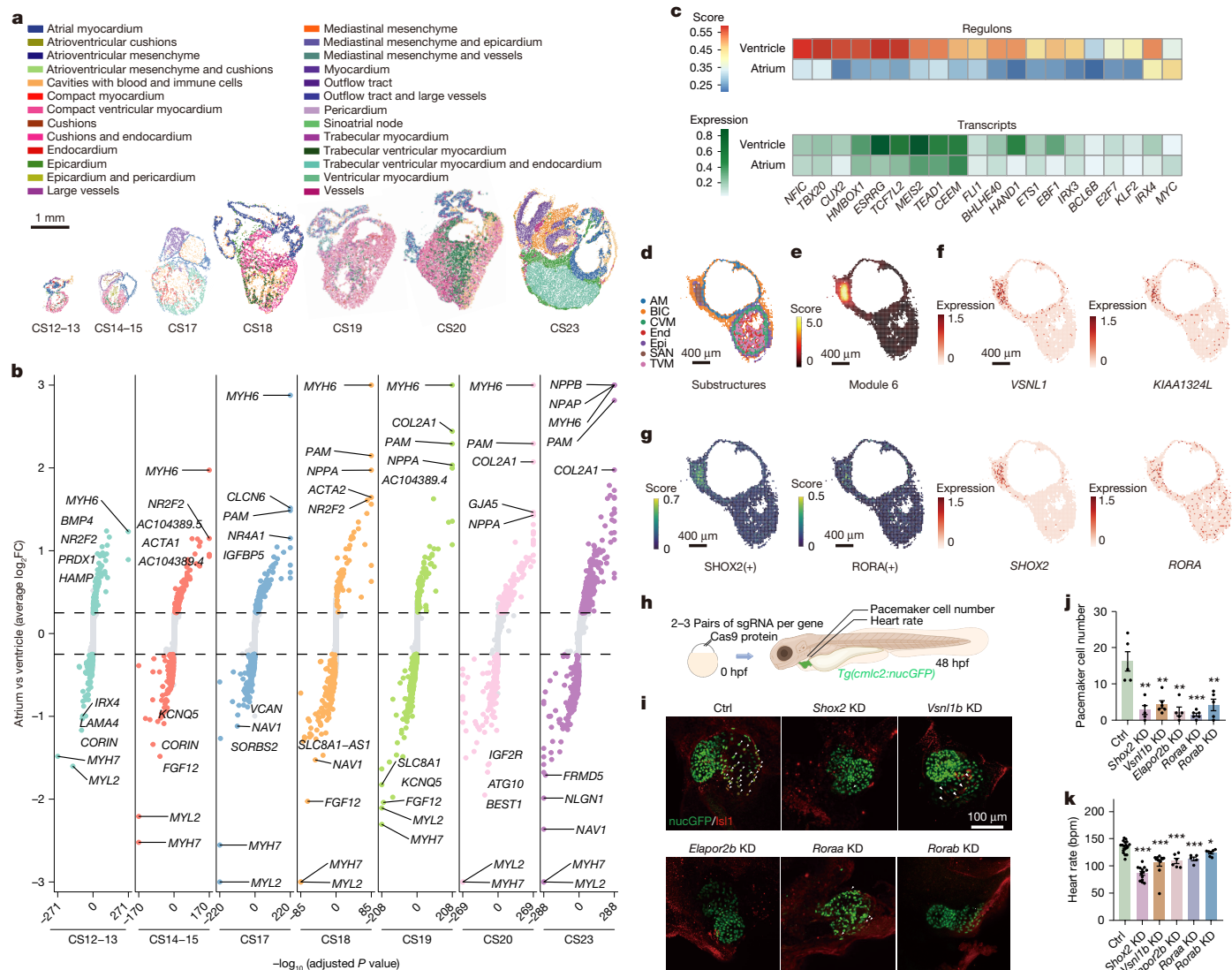


Fig. 3 | Cardiac development revealed by GRNs at the substructure level.

a, Spatial distribution of 26 cardiac substructures across CS12–23. **b**, DEGs in the atrium (top) and ventricle (bottom) from all 58 heart-containing chips. Genes were identified using two-sided Wilcoxon rank-sum test with Bonferroni correction for multiple comparisons (grey: average $\log_2$ fold change ($\log_2FC$) < 0.25 and adjusted $P > 0.01$). The dashed lines indicate the $\log_2FC$ thresholds at 0.25 and -0.25. **c**, Heatmap of atrium-specific or ventricle-specific regulon (pySCENIC) enrichment and their expression patterns. **d**, Spatial visualization of the substructures in the CS17 E1S3 heart. AM, atrial myocardium; BIC, cavity with blood and immune cells; CVM, compact ventricular myocardium; End, endocardium; Epi, epicardium; TVM, trabecular ventricular myocardium. **e**, Spatial visualization of the gene module related to SAN in the CS17 E1S3 heart. **f**, Spatial expression of *VSNL1* and *KIAA1324L* in the CS17 heart. **g**, Spatial visualization of regulon activity (left) and gene expression (right) of *SHOX2* and *RORA* in the CS17 heart. **h**, Schematic of the knockdown experiment in zebrafish. hpf, hours post-fertilization; sgRNA, single guide RNA. Schematic created in BioRender; Ye, L. <https://biorender.com/c5fx4o> (2026). **i**, Immunofluorescence staining of the SAN in *Tg(myl7:H2B-eGFP)* zebrafish at 48 hpf. Pacemaker cell nuclei are *Isl1*⁺ (Isl1⁺; red, denoted by white arrowheads). Ctrl, control; KD, knockdown. **j**, Pacemaker cell counts in control and knockdown zebrafish at 48 hpf. Control ($n = 5$), *Shox2* ($n = 5$, $P = 0.002$), *Vsn1b* ($n = 6$, $P = 0.002$), *Elapor2b* (*KIAA1324L* orthologue, $n = 5$, $P = 0.002$), *Roraa* ($n = 6$, $P = 2.1 \times 10^{-4}$) and *Rorab* ($n = 5$, $P = 0.005$). **k**, Heart rate of control and knockdown zebrafish at 48 hpf. Control ($n = 18$), *Shox2* ($n = 13$, $P = 1.9 \times 10^{-12}$), *Vsn1b* ($n = 11$, $P = 1.9 \times 10^{-5}$), *Elapor2b* ($n = 6$, $P = 5.3 \times 10^{-12}$), *Roraa* ($n = 5$, $P = 3.3 \times 10^{-4}$) and *Rorab* ($n = 7$, $P = 0.020$). Data are mean $\pm$ s.e.m.; P values were determined by unpaired, two-tailed Student's t -test: * $P < 0.05$, ** $P < 0.01$ and *** $P < 0.001$ (j,k).

the predominant involvement of *HMGA2* in Pall development, telencephalon regionalization, neurogenesis and RGC differentiation (Fig. 4f and Supplementary Table 16a,b).

To delineate *HMGA2*-associated cellular characteristics, we integrated snRNA-seq-based annotation of neurocytes and RGCs with spatial cell bin mapping (Fig. 4g and Extended Data Fig. 8d–f). We noted *HMGA2* enrichment in RGCs, particularly in the *PAX6*⁺*FOXG1*⁺ subpopulation within the Pall VZ, and this subpopulation exhibited an expansion at CS23 (Extended Data Fig. 8g,h). Lineage tracing identified a *PAX3*⁺*EFNA5*⁺*HESS*⁺ RGC subpopulation as the developmental

progenitor of neuronal cells, generating three lineages: *PAX6*⁺*FOXG1*⁺ RGC, glioblast and neuroblast (Fig. 4h). The expression of *HMGA2* exhibited dynamic oscillations during *PAX6*⁺*FOXG1*⁺ RGC commitment (Fig. 4i). Perturbational analysis following *HMGA2* knockout revealed accelerated glioblast and/or neuron differentiation at the expense of *PAX6*⁺*FOXG1*⁺ RGC maintenance (Fig. 4j). Given the established role of *HMGA2* in maintaining neural stem cell self-renewal²⁷ and suppressing neuronal differentiation²⁸, these findings show that *HMGA2*-regulated *PAX6*⁺*FOXG1*⁺ RGC lineage commitment constitutes a critical pathway for RGC pool formation.

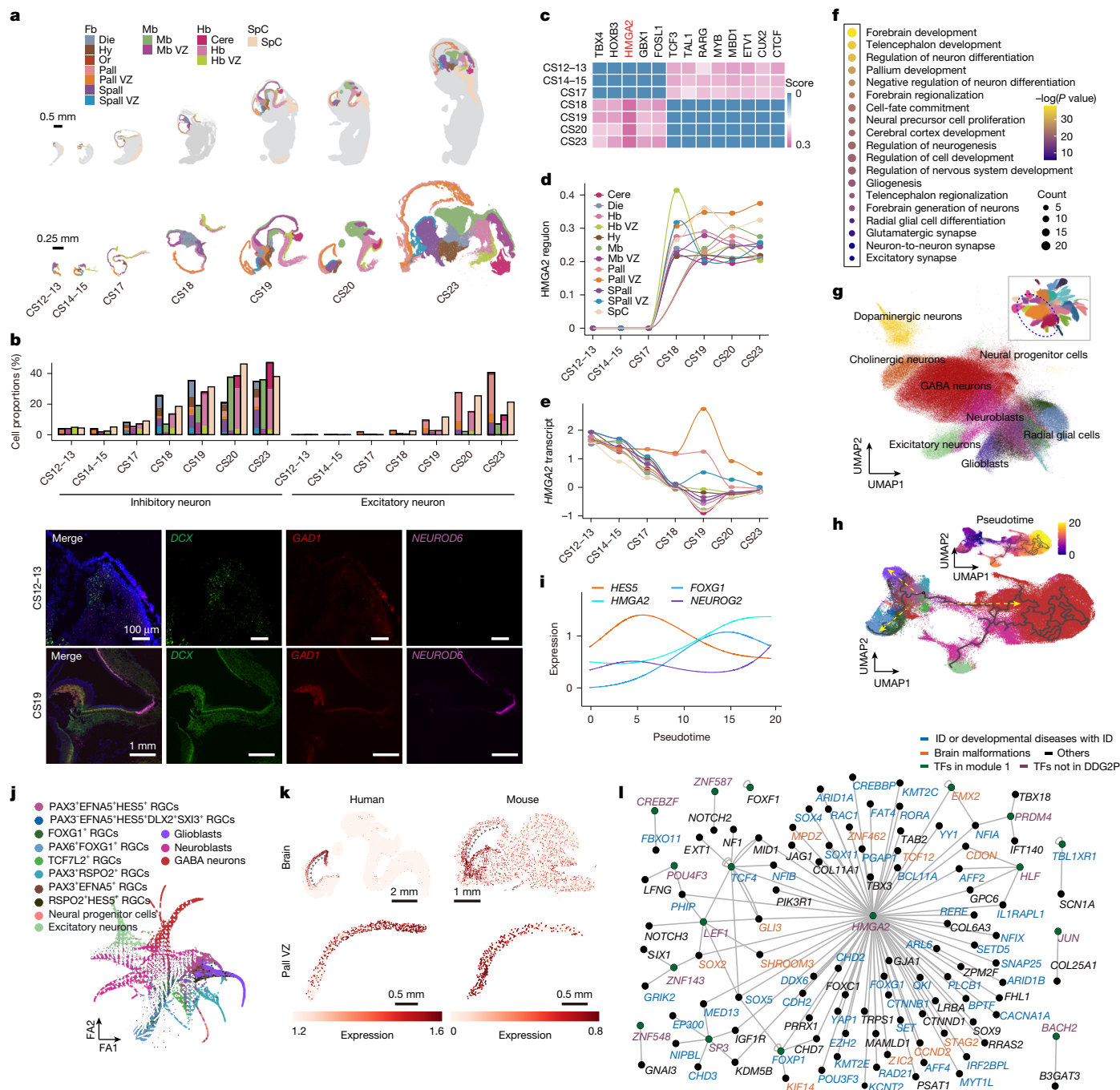


Fig. 4 | Cellular and molecular changes in the regionalization of the human embryonic nervous system. **a**, Spatiotemporal visualization of the colour-coded human embryonic nervous system (top), and the magnified visualization of the brain regions in the top panel (bottom). Cere, cerebellum; Die, diencephalon; Fb, forebrain; Hb, hindbrain; Hy, hypothalamus; Mb, midbrain; Or, optic recess; Spall, subpallium; SpC, spinal cord. **b**, Barplot showing the proportion of excitatory and inhibitory neurons within each region (top; the bar height indicates the cell proportion in the Fb, Mb, Hb and SpC, respectively; and the bar colour denotes the compositional ratio within the finer brain substructures). RNA ISH image (bottom) of *DCX* (immature neurons), *GAD1* (inhibitory neurons) and *NEUROD6* (excitatory neurons) in CS12–13 and CS19 (images are representative of three independent experiments). **c**, Heatmap showing scaled activity of specific regulons across the dynamic stages; the colour bar represents the regulatory activity score. **d, e**, Temporal dynamics

of *HMG2* regulon activity (d) and expression (e) at different substructures (or excluded due to temporal discontinuity). **f**, GO enrichment analysis of *HMG2* target genes. The *P* value was calculated by a one-tailed hypergeometric test to assess the enrichment significance. **g**, UMAP visualizations showing the clustering results of neural cells (CS19–23). **h**, The inferred trajectories for brain parenchymal cell subtypes are shown together with the pseudotime distribution along the trajectory. **i**, Line chart showing the indicated gene expression profiles along pseudotime. **j**, CellOracle simulation of cell-state transition in *HMG2*-knockout perturbation. FA, force atlas. **k**, Spatial visualization of *HMG2* in human (CS20 E1S6) and mouse (E15.5 E1S3) Pall VZ; the bar colour represents the normalized gene expression level. **l**, Gene regulatory networks of the CS18_M1 module. Selected target genes and transcription factors (TFs) reported in the DDG2P are shown. ID, intellectual disability.

To further investigate the spatial features of *HMG2*, we compared its expression patterns between human and mouse embryonic brains. Human *HMG2* showed rostral Pall VZ dominance contrasting with

the caudal enrichment of mouse *Hmg2* (Fig. 4k). To decode this cross-species divergence, we identified the specific gene sets S1 (inverse gradient) and S2 (concordant gradient) relative to *HMG2* expression

in the human dataset (Extended Data Fig. 8i). Integrated analysis of our spatial transcriptomics with published HMGA2 chromatin immunoprecipitation followed by sequencing (ChIP-seq) data²⁹, revealed preferential binding of HMGA2 to genes of the S1 set relative to S2 (Extended Data Fig. 8j). Meanwhile, the ChIP-seq analysis revealed HMGA2 occupancy at gene body regions of cluster 3 genes within the S1, with functional loss confirmed in the *Hmga2*^{-/-} mouse neocortex²⁹ that exhibited transcriptional dependence on HMGA2-mediated regulation (Extended Data Fig. 8k,l). Set 2, a subset of cluster 3, showing an inverse expression pattern with *Hmga2* in the fetal mouse brain, similar to humans, was functionally enriched for neural maturation and development (Extended Data Fig. 8m,n). These findings established HMGA2 as a transcriptional repressor of genes involved in neural maturation, potentially modulating the spatial patterning of RGC pools during embryonic brain development and contributing to species-specific differences in brain zone size between humans and mice.

Previous studies have established a connection between HMGA2 and brain size and cognitive ability³⁰, which indicates its involvement in neurodevelopmental diseases. Leveraging the developmental disorders genotype-to-phenotype database (DDG2P v3.0), we mapped the transcription factors and activated regulons associated with disease-linked genes. Following cluster analysis of regulons exhibiting conserved spatial patterns, we identified a developmentally coordinated regulatory network specifically active in the Pall and Pall VZ at CS18 (module 1). This network comprised 18 regulons targeting 421 genes, including HMGA2 and its downstream genes (Extended Data Fig. 8o and Supplementary Table 17). In particular, 99 of the targets are documented in the DDG2P database, with 51 specifically linked to intellectual disability (Fig. 4l). Subsequent analysis at CS19 revealed *NFIB*, a target gene of the HMGA2 regulon, as a central hub in this network (Extended Data Fig. 8p,q). The established role of *NFIB* in brain development and intellectual disability linkage³¹, combined with the observed HMGA2-to-*NFIB* regulatory cascade, suggests HMGA2 as a key architect of neural trajectories potentially relevant to intellectual disability pathogenesis.

Molecular basis of developmental diseases

Early embryo development is susceptible to pathogens and congenital diseases. Leveraging our spatial atlas, we systematically mapped the expression dynamics of genes associated with virus entry and developmental disorders across developing organs. By analysing viral receptor expression for pathogens implicated in gestational infections, we illustrated the molecular basis of diseases threatening early life (Fig. 5a, Extended Data Fig. 9 and Supplementary Table 18). Multiple arboviruses, including Zika virus, can disrupt infant brain development; *NCAM1* was enriched in the developing brain and spinal cord during CS18–23 (Extended Data Fig. 9a), potentially facilitating prenatal Zika virus neuroinfection³². Congenital human cytomegalovirus infection is the leading non-genetic cause of sensorineural hearing loss and key contributor to neurological impairment³³, and is associated with neonatal manifestations such as jaundice, hepatomegaly and microcephaly³⁴, with higher vertical transmission rate at the first trimester³³. The observed pathological manifestations in prenatal or neonatal cases align with the spatiotemporal receptor enrichment patterns—including *CD147* and *THY1* within the embryonic brain and cochlear structure as early as CS12–13, and *CD46* and *CD147* complementarily enriched in the embryonic liver (Fig. 5a and Extended Data Fig. 9a–d). For liver infections, *NTCP*, the primary hepatitis B virus receptor³⁵, exhibited low expression in the liver and other organs during CS12–23 (Fig. 5a and Extended Data Fig. 9e). By contrast, a potential hepatitis B virus receptor, *ASGR1* (ref. 36), displayed persistent liver-specific expression during this period (Fig. 5a and Extended Data Fig. 9d), suggesting compensatory viral entry mechanisms. For severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2) susceptibility, the co-expressions of essential entry factors, *ACE2* (ref. 37) and *TMPRSS2*

(ref. 38), were restricted to the fetal gut epithelium at CS23 but rarely co-existed within the same cell (Fig. 5a–d and Extended Data Fig. 9f–h). The limited co-expression of *ACE2* and *TMPRSS2* in other organs and minimal cellular colocalization aligns with clinical reports of low vertical transmission rates³⁹, particularly during CS12–23. By profiling host factors essential for virus attachment and entry, our atlas provides insight into the potential molecular basis of prenatal viral infection.

Next, we analysed the spatial and temporal expression of 1,922 disease-associated genes (Supplementary Table 19) from the DDG2P v3.0, identifying specific enrichment patterns (Fig. 5e, Supplementary Table 20 and Supplementary Fig. 7) that partially mirror the mouse embryonic dataset⁴⁰. Our analysis revealed striking spatiotemporal patterning of disease gene expression. First, stage-specific gene enrichment was identified in specialized structures such as the eye and inner ear, exemplified by ocular-specific expression of the cataract-related genes *CRYBA1* and *BFSP2* (Fig. 5e). This tissue-specific paradigm extended broadly: disease genes showed precise spatial enrichment in their affected organs: adrenal disorder genes in the adrenal tissue, intellectual disability genes in the brain and cardiac defect genes in the heart. Of note, some genes exhibited dual-expression patterns paralleling clinical phenotypes: cataract-associated genes (for example, *SLC33A1*, *GFER* and *FTL*) were co-enriched in ocular and hepatic tissues, whereas cardiomyopathy-related genes (*TTN*, *LDB3* and *LMNA*) showed cardiac skeletal muscle co-expression (Fig. 5e). Conversely, multi-organ disorders displayed broader spatial patterns, consistent with their systemic manifestations (Extended Data Fig. 10a). These results collectively establish that developmental transcriptomes reflect genotype–phenotype relationships across diverse disease contexts. Meanwhile, discrepancies in disease-associated genes between human and mouse embryos were noticed (Supplementary Table 21). For instance, human *ARG1*, critical for the urea cycle in the liver⁴¹ and implicated in growth retardation and neurological issues⁴², showed hepatic dominance from CS14–15, whereas mouse *Arg1* only became liver enriched at embryonic day 16.5 (E16.5; Fig. 5f). Gene enrichment analysis using the upregulated genes in the human liver confirmed the differences and identified the pathways related to arginase 1 deficiency⁴¹, such as the urea cycle and arginine metabolism (Extended Data Fig. 10b and Supplementary Table 22). This temporal mismatch, coupled with urea cycle pathway activation in the human liver, highlights limitations in modelling human argininemia using standard *Arg1*-knockout mice⁴³. Similarly, *CCBE1*, known to regulate lymphatic vascular development and cause Hennekam syndrome⁴⁴, showed high lung-specific expression in human embryos after CS19, whereas the *CCBE1* orthologue in mice was minimally expressed in the lung between E10.5 and E16.5 (Fig. 5f). Enrichment pathways associated with lung development were identified from the upregulated genes in human lungs (Extended Data Fig. 10c and Supplementary Table 22).

In addition, our cross-species temporal transcriptomic analysis identified conserved and divergent expression dynamics of continuously upregulated genes (UTGs) or downregulated genes (DTGs) in the brain, heart and liver (Supplementary Table 23). Although most UTGs or DTGs showed evolutionarily conserved patterns, organ-specific functional signatures emerged: heart UTGs in coagulation and extracellular matrix remodelling and liver UTGs in erythrocyte development, whereas human–mouse-shared DTGs universally mediated DNA replication and non-coding RNA processing, and brain UTGs were enriched in axonogenesis and synaptic signalling (Extended Data Fig. 10d–f and Supplementary Table 24). Of note, genes displaying opposing temporal trajectories between species were identified, including some established genes from the DDG2P (Extended Data Fig. 10f and Supplementary Table 23). Specifically, in the brain, *SUCLG1* and *NHP2* were upregulated during early developmental stages in humans but downregulated in mice, whereas *HNRNPK* and *CAMTA1* expression progressively decreased in humans but increased in mice. These opposing expression patterns suggest species-specific regulatory divergence

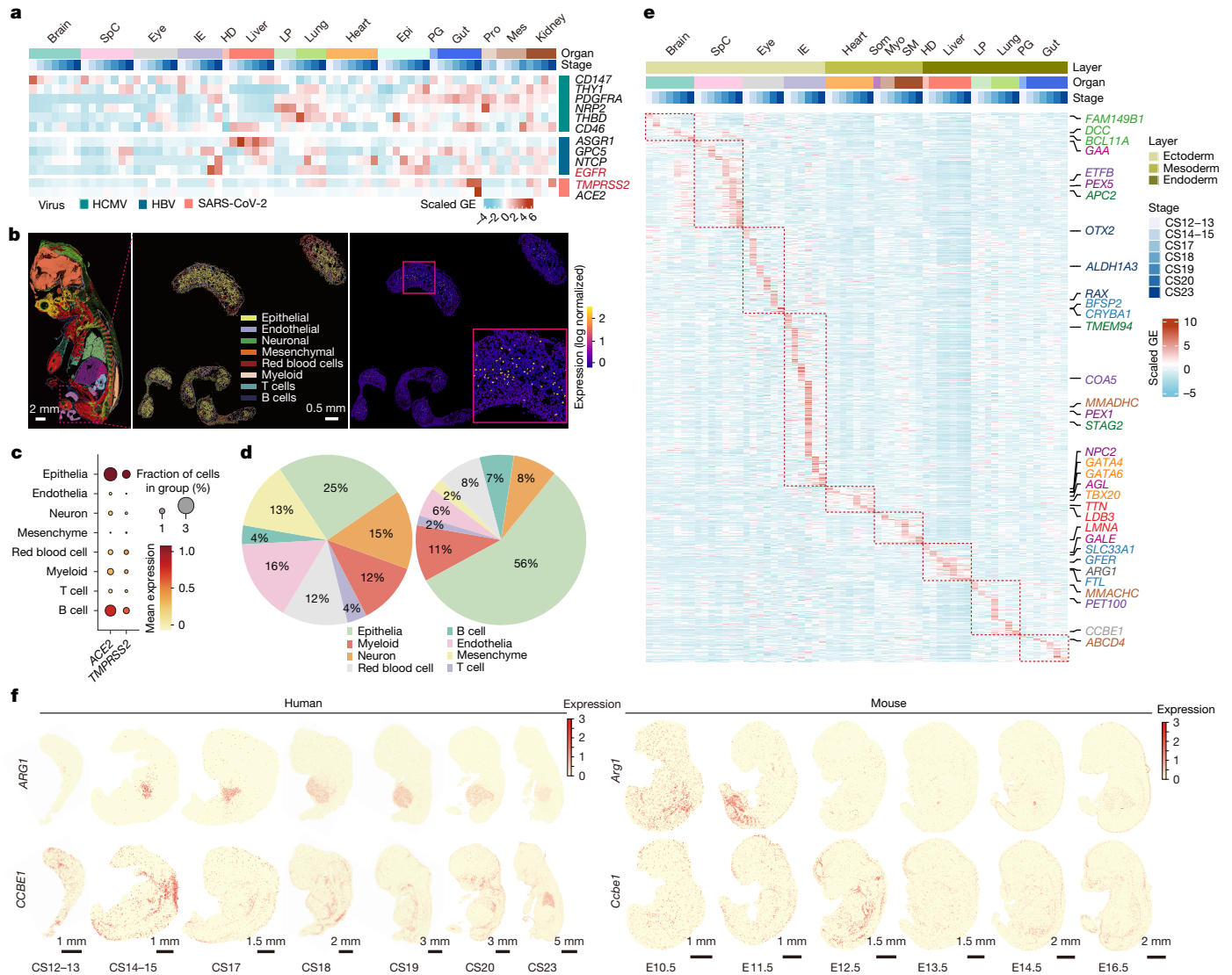


Fig. 5 | Organ enrichment of virus receptors and susceptibility to developmental disorders. **a**, Heatmap showing the normalized gene expression level of receptors in 15 representative anatomical regions from CS12–13 to CS23. The entry receptor of the virus is in black and the important coreceptor is in red. Epi, epidermis; HBV, hepatitis B virus; HCMV, human cytomegalovirus; HD, hepatic diverticulum; IE, inner ear; LP, lung primordium; Mes, mesonephron; PG, primitive gut; Pro, pronephron. **b**, Spatial visualization of cell types and *ACE2* expression based on single-cell segmentation in the gut at CS23. Results are representative of eight sections (from two biologically independent samples) with similar outcomes. **c**, Bubbleplot showing the normalized expression of *ACE2* and *TMPRSS2* in the annotated cell types. **d**, The proportion of cell types in the gut (left) versus cells with *ACE2* expression (right) at CS23. **e**, Heatmap showing the normalized expression of 1,740 genes selected from the DDG2P in the representative anatomical regions across CS12–23. **f**, Spatial visualization of *ARG1* and *CCBE1* expression in the human embryo from CS12–13 to CS23, and *Arg1* and *Ccbe1* expression in the mouse embryo from E10.5 to E16.5.

in neurodevelopmental pathways (Extended Data Fig. 10g). Overall, these results illustrated the differences in genes across species and emphasized the limitations when applying mouse models for certain human disease studies.

Allelic imbalanced expression in embryos

In most circumstances, both alleles of a gene are transcribed; however, certain genes show allelic imbalanced expression (AIE) with imbalanced mRNA abundance between alleles. Selective expression of two alleles of a gene determined by the parent of origin is known

as imprinting, whose dysregulation underlies developmental and neurological disorders. As AIE during human embryogenesis remains largely unknown, we investigated the spatial patterns of AIE using our dataset, identifying 452 genes with reproducible single-nucleotide polymorphisms exhibiting AIE (existed in at least three of five embryos; Extended Data Fig. 11a and Supplementary Table 25a–d). Leveraging haplotype phasing of three embryos (CS14–15E1, CS14–15E3 and CS20E2), we consolidated single-nucleotide polymorphism-level signals into gene-level AIE calls, identifying 104 genes with cross-organ allelic imbalance, including 20 imprinted genes (Extended Data Fig. 11b–d and Supplementary Table 25e). Canonical imprinted genes displayed

spatially resolved expression patterns: paternal dominance was observed in *MAGEL2*, *IGF2* and *MEST*, whereas *MEG3* exhibited maternal dominance, with spatial expression gradients mirroring allelic imbalance (Extended Data Fig. 11a–f). Of note, the pseudogene *RN7SKP255* also demonstrated strong AIE (Extended Data Fig. 11g,i), whereas *SCIPT* and *GFPT2* demonstrated pronounced AIE in the developing brain (Extended Data Fig. 11h,i). We further validated these findings by Sanger sequencing, which confirmed the strong AIE in select genes (Extended Data Fig. 11j).

Discussion

In recent years, dynamic changes occurring during human development have been charted at single-cell resolution^{1,2}, and spatial transcriptomics has elucidated the spatial transcriptional characteristics and complex architectures of certain developing organs^{45–47}. Our study established a spatiotemporal transcriptomic atlas of human embryogenesis from CS12 to CS23, achieving near-single-cell resolution while preserving spatial context. By bridging the gap between gastrulation studies and organ-specific analyses, this resource (<https://db.genomics.cn/hesta>) reveals how transcriptional programs become spatially compartmentalized to orchestrate organ patterning and cellular differentiation. This atlas bridges a substantial gap in developmental biology, as previous spatial studies have focused either on gastrulation (CS7–8)^{5,6} or isolated embryonic stage², leaving the pan-embryonic coordination of organogenesis poorly resolved. Central to this atlas is its unique value in integrating whole-embryo spatial mapping with temporal dynamics across 50 organs and 198 substructures. By reconstructing dynamic transcriptional trajectories and GRNs that orchestrate cellular diversification, we systematically resolved chronological regulation of organogenesis. These findings extend previous single-cell atlases^{19,25,46} by spatially contextualizing developmental checkpoints, offering opportunities to decode developmental coordination during organogenesis.

Leveraging this integrated atlas, we systematically characterized functional hierarchies of cardiac and neural developmental regulators. High-resolution spatial transcriptomic profiling revealed *RORA* and *KIAA1324L* as pivotal regulators within the SAN niche, which not only identified *KIAA1324L* as a previously unrecognized determinant but also crucially recontextualized the ‘known’ gene *RORA*, encoding a nuclear receptor with dual roles in mammalian circadian regulation²⁴ and cerebellum development⁴⁸. These findings may establish a novel regulatory framework for SAN development while opening new directions for cardiac pacemaking. In the brain, our atlas elucidated phylogenetic divergence in neurodevelopmental regulation, offering a more detailed understanding of the neurogenesis timeline than previous single-cell and spatial atlases that focused on specific brain regions. The divergent encephalization patterns between humans (Pall dominant) and mice (olfactory bulb enriched) may be mechanistically explained by species-specific spatial expression of *HMGA2*, which differentially maintains RGC potency through chromatin remodelling^{27,28}. Of note, *HMGA2* has also been implicated in developmental disease, with *HMGA2* and its downstream gene *NFIB* as temporally distinct components of intellectual disability-associated GRNs at CS18 and CS19, respectively. Further studies are required to clarify the precise role of *HMGA2* in maintaining RGC potency and neural diseases.

Beyond development, this atlas also shows its strength in elucidating molecular susceptibilities to pathogens and congenital disorders. Microorganisms implicated in congenital diseases probably use diverse strategies to breach the maternal–fetal interface⁴⁹. The placental barrier, not fully established until the end of the first trimester, renders this period susceptible to vertical transmission of pathogens. Here we identified stage-specific patterns of virus entry factors, delineating the molecular basis and the permissive time window for congenital viral infection. Our results suggested that human embryos express receptors for viruses impacting neural development and liver health as early

as CS12, highlighting the importance of early-pregnancy pathogen surveillance while establishing a molecular framework to study viral teratogenesis. However, the expression of the virus entry factors alone cannot fully explain the tissue or cell tropism, as alternative receptors, glycan-mediated attachment and strain-specific receptor-binding motifs may also have critical roles⁵⁰. Actual infection risk, including maternal immunity and viral load, requires longitudinal cohort studies for definitive. In addition, by systematically comparing the orthologous genes between humans and mice, we identified species-specific discrepancies in disease-associated gene expression, underscoring the limitations of mouse models for studying certain human diseases. Although the developmental importance of the AIE during embryogenesis remains to be elucidated, our data complement existing postnatal literature and provide a foundation for identifying undefined imprinting candidates in human development, which may aid in elucidating the aetiology of developmental disorders in the future.

Although spatial omics technologies represent substantial advances, some challenges persist: low nucleic acid capture efficiency restricts rare-cell detection, persistent two-dimensional imaging obscures three-dimensional morphogenetic dynamics, and multimodal integration requires novel computational pipelines. Future directions should prioritize combinatorial approaches integrating three-dimensional spatial imaging with single-cell multiomics, develop machine learning frameworks to decode regulatory networks from high-dimensional datasets, and leverage cross-species comparisons to distinguish conserved developmental programs from lineage-specific adaptations.

In conclusion, the spatiotemporal transcriptome atlas offered a comprehensive delineation of human embryogenesis after gastrulation. This resource not only refines the existing knowledge on embryonic development but also facilitates the delineation of fine structures and the developmental trajectories of various systems, potentially identifying the cellular and molecular basis of developmental and congenital diseases. Moreover, as developmental biology enters the spatial multiomics era, it serves both as a foundational reference and a catalyst for decoding human-specific life processes – a critical endeavor given the limited predictive power of model organisms in recapitulating human spatial developmental programs.

Online content

Any methods, additional references, Nature Portfolio reporting summaries, source data, extended data, supplementary information, acknowledgements, peer review information; details of author contributions and competing interests; and statements of data and code availability are available at <https://doi.org/10.1038/s41586-026-10545-0>.

1. Cao, J. et al. A human cell atlas of fetal gene expression. *Science* **370**, eaba7721 (2020).
2. Xu, Y. et al. A single-cell transcriptome atlas profiles early organogenesis in human embryos. *Nat. Cell Biol.* **25**, 604–615 (2023).
3. Nishino, T. et al. Single cell multimodal analyses reveal epigenomic and transcriptomic basis for birth defects in maternal diabetes. *Nat. Cardiovasc. Res.* **2**, 1190–1203 (2023).
4. Barnat, M. et al. Huntington's disease alters human neurodevelopment. *Science* **369**, 787–793 (2020).
5. Xiao, Z. et al. 3D reconstruction of a gastrulating human embryo. *Cell* **187**, 2855–2874.e19 (2024).
6. Cui, L. et al. Spatial transcriptomic characterization of a Carnegie stage 7 human embryo. *Nat. Cell Biol.* **27**, 360–369 (2025).
7. Jankovic, V. et al. Id1 restrains myeloid commitment, maintaining the self-renewal capacity of hematopoietic stem cells. *Proc. Natl Acad. Sci. USA* **104**, 1260–1265 (2007).
8. Ballantyne, C. M. et al. Effect of targeting apoC-III with plozasiran on lipoprotein particle size and number in hypertriglyceridemia. *J. Am. Coll. Cardiol.* <https://doi.org/10.1016/j.jacc.2025.03.496> (2025).
9. Fillebeen, C. et al. Transferrin receptor 1 controls systemic iron homeostasis by fine-tuning hepcidin expression to hepatocellular iron load. *Blood* **133**, 344–355 (2019).
10. Siatecka, M. & Bieker, J. J. The multifunctional role of EKLF/KLF1 during erythropoiesis. *Blood* **118**, 2044–2054 (2011).
11. Musa, J., Aynaud, M. M., Mirabeau, O., Delattre, O. & Grunewald, T. G. MYBL2 (B-Myb): a central regulator of cell proliferation, cell survival and differentiation involved in tumorigenesis. *Cell Death Dis.* **8**, e2895 (2017).
12. Ma, X. et al. The transcription factor MITF in RPE function and dysfunction. *Prog. Retin. Eye Res.* **73**, 100766 (2019).

13. Li, R. et al. Isl1 and Pou4f2 form a complex to regulate target genes in developing retinal ganglion cells. *PLoS ONE* **9**, e92105 (2014).
14. Gautam, P. et al. Multi-species single-cell transcriptomic analysis of ocular compartment regulons. *Nat. Commun.* **12**, 5675 (2021).
15. Massa, F. et al. Hepatocyte nuclear factor 1 β controls nephron tubular development. *Development* **140**, 886–896 (2013).
16. Moriguchi, T. et al. MafB is essential for renal development and F4/80 expression in macrophages. *Mol. Cell. Biol.* **26**, 5715–5727 (2006).
17. Cosse-Etchepare, C. et al. Pou3f transcription factor expression during embryonic development highlights distinct pou3f3 and pou3f4 localization in the *Xenopus laevis* kidney. *Int. J. Dev. Biol.* **62**, 325–333 (2018).
18. Del Monte-Nieto, G. et al. Control of cardiac jelly dynamics by NOTCH1 and NRG1 defines the building plan for trabeculation. *Nature* **557**, 439–445 (2018).
19. Farah, E. N. et al. Spatially organized cellular communities form the developing human heart. *Nature* **627**, 854–864 (2024).
20. Wang, L., Lin, L., Qi, H., Chen, J. & Grossfeld, P. Endothelial loss of ETS1 impairs coronary vascular development and leads to ventricular non-compaction. *Circ. Res.* **131**, 371–387 (2022).
21. Zhou, P. et al. GATA4 regulates developing endocardium through interaction with ETS1. *Circ. Res.* **131**, e152–e168 (2022).
22. Espinoza-Lewis, R. A. et al. Shox2 is essential for the differentiation of cardiac pacemaker cells by repressing Nkx2-5. *Dev. Biol.* **327**, 376–385 (2009).
23. Liang, D. et al. Cellular and molecular landscape of mammalian sinoatrial node revealed by single-cell RNA sequencing. *Nat. Commun.* **12**, 287 (2021).
24. Sato, T. K. et al. A functional genomics strategy reveals Rora as a component of the mammalian circadian clock. *Neuron* **43**, 527–537 (2004).
25. Zeng, B. et al. The single-cell and spatial transcriptional landscape of human gastrulation and early brain development. *Cell Stem Cell* **30**, 851–866.e7 (2023).
26. Shi, Y. et al. Mouse and human share conserved transcriptional programs for interneuron development. *Science* **374**, eabj6641 (2021).
27. Nishino, J., Kim, I., Chada, K. & Morrison, S. J. Hmga2 promotes neural stem cell self-renewal in young but not old mice by reducing p16Ink4a and p19Arf expression. *Cell* **135**, 227–239 (2008).
28. Kishi, Y., Fujii, Y., Hirabayashi, Y. & Gotoh, Y. HMGA regulates the global chromatin state and neurogenic potential in neocortical precursor cells. *Nat. Neurosci.* **15**, 1127–1133 (2012).
29. Kuwayama, N. et al. HMGA2 directly mediates chromatin condensation in association with neuronal fate regulation. *Nat. Commun.* **14**, 6420 (2023).
30. Stein, J. L. et al. Identification of common variants associated with human hippocampal and intracranial volumes. *Nat. Genet.* **44**, 552–561 (2012).
31. Schanze, I. et al. NFIB haploinsufficiency is associated with intellectual disability and macrocephaly. *Am. J. Hum. Genet.* **103**, 752–768 (2018).
32. Srivastava, M. et al. Chemical proteomics tracks virus entry and uncovers NCAM1 as Zika virus receptor. *Nat. Commun.* **11**, 3896 (2020).
33. Chatzakis, C. et al. Timing of primary maternal cytomegalovirus infection and rates of vertical transmission and fetal consequences. *Am. J. Obstet. Gynecol.* **223**, 870–883.e11 (2020).
34. Leruez-Ville, M. et al. Consensus recommendation for prenatal, neonatal and postnatal management of congenital cytomegalovirus infection from the European congenital infection initiative (ECCI). *Lancet Reg. Health Eur.* **40**, 100892 (2024).
35. Yan, H. et al. Sodium taurocholate cotransporting polypeptide is a functional receptor for human hepatitis B and D virus. *eLife* **1**, e00049 (2012).
36. Vyas, A. K. et al. Placental expression of asialoglycoprotein receptor associated with hepatitis B virus transmission from mother to child. *Liver Int.* **38**, 2149–2158 (2018).
37. Zhou, P. et al. A pneumonia outbreak associated with a new coronavirus of probable bat origin. *Nature* **579**, 270–273 (2020).
38. Hoffmann, M. et al. SARS-CoV-2 cell entry depends on ACE2 and TMPRSS2 and is blocked by a clinically proven protease inhibitor. *Cell* **181**, 271–280.e8 (2020).
39. Vivanti, A. J. et al. Transplacental transmission of SARS-CoV-2 infection. *Nat. Commun.* **11**, 3572 (2020).
40. Chen, A. et al. Spatiotemporal transcriptomic atlas of mouse organogenesis using DNA nanoball-patterned arrays. *Cell* **185**, 1777–1792.e21 (2022).
41. Sin, Y. Y., Baron, G., Schulze, A. & Funk, C. D. Arginase-1 deficiency. *J. Mol. Med.* **93**, 1287–1296 (2015).
42. Lee, B. H., Jin, H. Y., Kim, G.-H., Choi, J.-H. & Yoo, H.-W. Argininemia presenting with progressive spastic diplegia. *Pediatr. Neurol.* **44**, 218–220 (2011).
43. Iyer, R. K. et al. Mouse model for human arginase deficiency. *Mol. Cell. Biol.* **22**, 4491–4498 (2002).
44. Bos, F. L. et al. CCB1 is essential for mammalian lymphatic vascular development and enhances the lymphangiogenic effect of vascular endothelial growth factor-C in vivo. *Circ. Res.* **109**, 486–491 (2011).
45. Asp, M. et al. A spatiotemporal organ-wide gene expression and cell atlas of the developing human heart. *Cell* **179**, 1647–1660.e19 (2019).
46. Fawcner-Corbett, D. et al. Spatiotemporal analysis of human intestinal development at single-cell resolution. *Cell* **184**, 810–826.e23 (2021).
47. Li, Y. et al. Spatiotemporal transcriptome atlas reveals the regional specification of the developing human brain. *Cell* **186**, 5892–5909.e22 (2023).
48. Guissart, C. et al. Dual molecular effects of dominant RORA mutations cause two variants of syndromic intellectual disability with either autism or cerebellar ataxia. *Am. J. Hum. Genet.* **102**, 744–759 (2018).
49. Megli, C. J. & Coyne, C. B. Infections at the maternal–fetal interface: an overview of pathogenesis and defence. *Nat. Rev. Microbiol.* **20**, 67–82 (2022).
50. Chen, D. et al. Screening of cell-virus, cell-cell, gene-gene crosstalk among animal kingdom at single cell resolution. *Clin. Transl. Med.* **12**, e886 (2022).

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

Springer Nature or its licensor (e.g. a society or other partner) holds exclusive rights to this article under a publishing agreement with the author(s) or other rightsholder(s); author self-archiving of the accepted manuscript version of this article is solely governed by the terms of such publishing agreement and applicable law.

© The Author(s), under exclusive licence to Springer Nature Limited 2026

Methods

Ethics statement

Ethical approval of human embryo research was provided by the Ethics Committee of Obstetrics and Gynecology Hospital, Fudan University (2021-121) and the Institutional Review Board of BGI-Shenzhen (BGI-IRB 22058). Human embryos and decidua tissues were collected from women who voluntarily terminated the pregnancy at the Affiliated Obstetrics and Gynecology Hospital of Fudan University. No individuals under the 18 years of age were included. The collection and use of human embryos were compliant with the current International Society for Stem Cell Research Guidelines, guidelines for Stem Cell Research and Clinical Translation (2016 and 2021) and the Helsinki declaration (2013). None of the samples used in this study was used in any other experiments. All procedures followed the 'Interim Measures for the Administration of Human Genetic Resources' administered by the Chinese Ministry of Health. Before termination of pregnancy, each participant provided written informed consent for collection for research purposes. Embryo collection was performed only after voluntary informed consent was obtained from each patient undergoing legal pregnancy termination. After the termination of pregnancy, participants were given an additional opportunity to verbally reconfirm their willingness to donate the tissue samples. Discarded embryos were collected and transported immediately to the laboratory for processing. Embryo collection does not involve any invasive procedure or gene editing. There is no risk to the patients. The patients were not compensated for participating in this study.

Embryo collection and preparation for Stereo-seq and snRNA-seq

Sample sizes were not predetermined. One to three embryos were collected at each stage during CS12–23. Once obtained, human embryos were rinsed with PBS to remove surface blood. Embryos were then placed in a cryomold and embedded with optimal cutting temperature compound (4583, Sakura) within 30 min. All reagents were prepared with sterilized water containing diethylpyrocarbonate (B501005-0005, Sangon Biotech), and all instruments were cleaned with sterilized water containing diethylpyrocarbonate and RNAase Zap (AM9780, Invitrogen). Cryosections were performed in sagittal cut with a thickness of 10 μm using a Leica CM1950 cryostat. In each embryo, 1–10 sections with interval distance of 20–1,800 μm (average interval distance of 310 μm) and covered the maximum sagittal section were obtained for Stereo-seq. Extra sagittal sections next to each Stereo-seq section were also obtained for haematoxylin and eosin staining or ISH assays. Each section for Stereo-seq was then gently transferred to a Stereo-seq chip pre-cooled at -20°C . Extra sagittal sections next to Stereo-seq section in five embryos (CS19E1, CS19E2, CS20E1, CS20E2 and CS23E1) were also obtained for snRNA-seq.

Single-nucleus sequencing analysis

Single-nucleus sequencing libraries were prepared with DNBelab C4 kits (MGI) and sequenced on DNBSEQ-Tx sequencer (MGI). Gene count matrices were generated from FASTQ files by an in-house pipeline (BGI-Research). The downstream analyses were performed using Scanpy package (v1.8.2). In brief, in the quality control step, cells with less than 1,000 genes, less than 2,000 counts, more than 10% mitochondrial genes or more than 90 quantiles of total counts per library were filtered out from raw counts by `scanpy.pp.filter_cells`. Genes that were captured in less than three cells were filtered out as well by `scanpy.pp.filter_genes`. A total of 607,093 cells from developing human embryos at stages of CS19, CS20 and CS23 were proceeded to downstream analysis. The filtered raw gene counts were first normalized by `scanpy.pp.normalize_total` with `target_sum=1e4`, and transformed by `scanpy.pp.log1p`. Principal component analysis (PCA) was performed by applying `scanpy.pp.scale` and `scanpy.pp.pca` to log-normalized gene counts. A k -nearest neighbour (kNN)

graph was constructed in the principal component space by `scanpy.pp.neighbor` with $k=10$. Unbiased clusters were further partitioned by applying Leiden algorithm to the kNN graph by `scanpy.pl.leiden` with `resolution=0.8`. Cells were visualized in the umap by `scanpy.tl.umap` and `scanpy.pl.umap`. The DEGs among clusters were identified with Wilcoxon rank-sum by `scanpy.tl.rank_genes_groups`. By integrating the top 100 DEGs within each cluster and canonical marker genes, we established a comprehensive annotation of 68 distinct cell populations (Supplementary Table 2c).

Spatial transcriptome capture and sequencing for Stereo-seq

The spatial transcriptome of human embryo sections was obtained using the Stereo-seq technology, which relies on the DNBs that are photolithographically etched on the chip with 500–750-nm distance to capture tissue RNA⁴⁰. Spatial transcriptomics was captured as previously described^{40,51}. The tissue sections on the Stereo-seq chips were incubated at 37°C for 3–8 min, and fixed in methanol (34860, Sigma) at -20°C for 40 min, followed by the staining of nuclei single-strand DNA using nucleic acid dye (Q10212, Thermo Fisher) and imaging (Ti-7 Nikon Eclipse microscope). The sections were then permeabilized with 0.1% pepsin (P7000, Sigma) in 0.01 M HCl buffer at 37°C for 12 min. The released RNAs were captured by DNBs on the Stereo-seq chip and reverse transcribed overnight using SuperScript II (18064-014, Invitrogen) at 42°C . The chip was then washed with $0.1\times$ SSC buffer (AM9770, Thermo) and digested with Tissue Removal buffer (10 mM Tris-HCl, 25 mM EDTA, 100 mM NaCl and 0.5% SDS) at 37°C for 30 min. cDNAs were treated with exonuclease I (M0293L, NEB) at 37°C for 1 h, amplified with KAPA HiFi Hotstart Ready Mix (KK2602, Roche), purified using $0.6\times$ VAHTSTM DNA Clean Beads, and quantified using Qubit dsDNA HS assay kit (Q32854, Invitrogen). The amplified DNAs were fragmented with in-house Tn5 transposase, amplified with KAPA HiFi Hotstart Ready Mix and purified with AMPure XP Beads. The qualified libraries were then sequenced on a DNBSEQ-Tx sequencer (MGI).

Stereo-seq raw data processing

The Stereo-seq raw data were processed as previously described⁴⁰. The coordinate identity (CID) sequences (1–25 bp) were mapped to the coordinates of the chip, allowing 1 base mismatch. Molecular identity (MID) sequences (26–35 bp) containing N bases or with a quality score below 10 for more than 2 bases were filtered out. cDNA sequences (100 bp) were aligned to the reference genome (GRCh38.p12) by STAR (V2.7.10a)⁵². The expression matrices with coordinates were generated from the above procedures.

Cell segmentation

The nuclei single-stranded DNA images were used to calculate the cell boundaries with the cell segmentation algorithm `spateo`⁵³ (v1.0.2). The greyscale images of nuclei single-stranded DNA were converted to binary images using the calculated Gaussian-weighted threshold. The exact Euclidean distance transformation was performed to segment cell nuclei with overlapped regions. Labels representing different cell nuclei were transferred to pinpoint DNBs corresponding to spatial positions by the watershed algorithm⁵⁴ with parameters of `distance=6`, `max_area=600`.

Unsupervised spatially constrained clustering

Unsupervised spatially constrained clustering was conducted using Scanpy⁵⁵ with the bin50 count matrices and subsequently annotated the partitioned clusters at the organ level based on their highly differential genes. Count matrices were first normalized, and variable genes were identified by Scanpy. For each Stereo-seq chip, the 30-nearest neighbour graph based on gene expressions and the 8-nearest neighbour graph based on spatial coordinates were constructed using Scanpy⁵⁵ and Squidpy (v1.2.0)⁵⁶, respectively. A final unweighted neighbourhood graph was plotted by taking the union of expression and spatial

connectivity, and incorporated into the Leiden algorithm to identify clusters with tuned resolutions.

Identification of spatial developmental substructures

To delineate refined spatial organization patterns, bin50 data from the brain, spinal cord and heart underwent a secondary round of unsupervised spatially constrained clustering at the sectional level, enabling high-resolution identification of organ-scale substructures. For the remaining organs, we implemented a strategic bin aggregation approach: bin50 data from at least two representative sections in each embryo were combined before clustering (Supplementary Table 3), thereby enhancing detection sensitivity for continuous developmental substructures across specimens. Pooled raw counts were integrated to train a scVI model from scvi-tools (v1.2.2) with parameter `early_stopping = True`⁵⁷. This generated batch-corrected latent representations (stored in `X_scvi`) that served as the foundation for constructing combined neighbour graphs and deriving biologically meaningful UMAP embeddings. Subsequent Louvain clustering was applied with a chosen resolution of 1.0. Through systematic annotation using established region-specific and cell-type-defining markers, we ultimately identified 198 distinct spatial substructures (Supplementary Table 3).

Bin-level gene set enrichment analysis

The *irGSEA* (V3.2.5) package in R⁵⁸ was used to perform rank-based gene set enrichment analysis. Pathway activity quantification was performed via the *UCell* algorithm, which computes per-bin50 enrichment scores against GO biological process gene sets. The matrices were used to identify significant substructure-specific biological processes using `rank_genes_groups` in *Scanpy* (adjusted $P < 0.05$; Extended Data Fig. 2d,g and Supplementary Table 4).

Identifying developmental-stage-specific DEGs in key organs

All human embryo chips at each stage were randomly sampled in proportion and then used in the differential expression analysis. All sampled count matrices were normalized by the SCTransform method and integrated by the reciprocal PCA-based pipeline of *Seurat*³⁹. The subset *rds* objects of the same organ from all stages were further merged. *FindMarkers* function in *Seurat* was used to identify the DEGs of each developmental stage in 50 organs (classified into 34 groups) with default parameters, and the normalized RNA assay data from the merged objects were used for the calculations. Among the 34 organ groups, the stomach and pituitary gland were observed in only a single developmental stage, thus DEG analysis was performed for the remaining 32 organ groups. GO enrichment analyses of the upregulated genes (with an average $\log_2FC > 0.25$) at each developmental stage were performed using R package *clusterProfiler* (v4.6.0)⁶⁰. In addition, the expression patterns of DEGs in 11 selected organs (brain, spinal cord, eye, ear, liver, heart, lung, gut, kidney, skeletal muscle and bone) were scaled and visualized using the R package *ComplexHeatmap* (v2.12.0)⁶¹.

GRN analysis

A regulon is the co-expressed target genes of a transcription factor validated by *cis*-regulatory motifs and pruned of indirect targets. GRNs were inferred from Stereo-seq bin matrices according to the *pySCENIC* protocol⁶² using bin50 matrices (CS12–20) or bin200 matrices (CS23). The databases used for *pySCENIC* (v0.11.2) were downloaded from <https://resources.aertslab.org/cistarget/databases/>. Co-expression modules were inferred, pruned and quantified from the count matrix by *GRNBoost2*, *cisTarget* and *AUCell*, respectively. Along with the physical coordinates, regulon activity scores (RASs)⁶³ were visualized in the tissue context across developmental stages using *Scanpy*.

Regulon downstream analysis

Spatially dependent regulon modules were identified by *Hotspot* (v1.1.1)⁶⁴. The RAS matrix from *pySCENIC* was first normalized by a

negative binomial model ('danb'), and the *k*-nearest neighbour graph (10 nearest neighbours) was plotted between bins based on spatial proximity. Autocorrelations were computed for each regulon. The spatially variable regulons were grouped as spatial modules with a minimum of three regulons per module and visualized by a covariance heatmap. Bins with the top five percentile of module scores were considered as the corresponding module locations. Sankey plots representing relationships between GRN modules and annotated organs were visualized by *Plotly* (v5.13.1).

For each stage or chip, regulon specificity scores (RSSs) were calculated to quantify the specificity of the regulons across different organs⁶³. RAS of each regulon was first normalized to a probability distribution. Organs were then represented as a vector of binary labels (target organ: 1; others: 0) and normalized to the probability distribution. The differences between the two normalized probability distributions were calculated by the Jensen–Shannon divergence (JSD) method, with a range from 0 to 1. Finally, JSD scores were converted to the RSSs using the formula: $RSS = 1 - \sqrt{JSD}$. A higher RSS indicates increased enrichment of regulon in the corresponding organ. The regulons in each organ were ranked based on RSSs, and the top five or top ten were selected as organ-identity regulons. Heatmaps of the top five regulons across organs were visualized by *Seaborn* (v0.11.2).

The relationship between GRN modules and annotated organs (for CS12–18) or cell types (for CS19–23) was visualized through heatmap analysis using *Python* (v3.8.13). Organ or cell-type annotation data and module scoring matrices were processed and merged via *pandas* (v1.5.2). To identify highly expressed modules, the top 5% spatial bins for each module were selected based on a 95% quantile threshold, followed by quantification of corresponding organ or cell-type proportions. Frequency counting was implemented using 'collections.Counter' function to establish a module–cell type percentage matrix. Visualization was performed with *Seaborn* (v0.13.2) and *Matplotlib* (v3.6.2).

To infer spatial regulatory network for each organ substructure, the merged raw count matrix was used to infer spatial regulatory network using *Stereopy* (v1.5.1) by setting the method = KNN. The resulting *AUCell* enrichment scores quantify the activity of each regulon per-bin50. The `auc_matrix` was extracted to identify significant differential transcriptional regulators in each substructure using `rank_genes_groups` in *Scanpy* (adjusted $P < 0.05$; Fig. 2h and Supplementary Table 11a).

Constructing developmental trajectories using Monocle3

To construct pseudotime trajectories for organs with continuous spatial transcriptomic profiles, multi-stage transcriptomic matrices for target organ were preprocessed to construct an object in *Monocle3* (v1.3.1)⁶⁵. Initial dimensionality reduction was performed by manually assigning the `X_scvi` representation as PCA embeddings, followed by UMAP recalculation for visualization. By setting the root node at the earliest developmental stage, organ trajectories were automatically computed through the `learn_graph` function, followed by the ordering of cells in pseudotime with the `order_cells` function. The `graph_test` function was used to identify the trajectory-dependent gene with the adjusted *P* value (corrected *P* value) threshold set at less than 0.01. The top 500 trajectory-associated genes, ranked by descending Moran's *I* spatial autocorrelation index derived from pseudotime analysis, are presented in Supplementary Table 10.

To construct single-cell trajectories for neural lineage, brain parenchymal cells were extracted from snRNA-seq data and analysed with *Monocle3*. Following standard preprocessing and Leiden clustering, differentiation trajectories were computed utilizing the `learn_graph` function by setting CS19 as the root stage, followed by the ordering of cells in pseudotime with the `order_cells` function. Subsequently, gene expression profiles along pseudotime were retrieved from the output of the `plot_genes_in_pseudotime` function.

Constructing molecular trajectories using TOME

To visualize the dynamic substructure relationships across developmental stages, we leveraged TOME⁶⁶ to construct the directed acyclic graph for the eye and kidney. The three-dimensional UMAP coordinates, precomputed using Monocle3, served as spatial embeddings for graph topology generation. The graph visualization excluded rare populations comprising less than 1% of the substructure bin composition at any developmental stage. To focus on principal lineage-relevant substructures, we excluded erythrocytic lineages, sympathetic ganglia and collecting duct epithelia from kidney analyses.

Estimation of batch effects and multi-section integration

To estimate the embryo-to-embryo and cross-section variability within the same stage, all chips at each stage were randomly sampled in proportion, and the sampled count matrices were normalized by the SCTransform method and integrated by the reciprocal PCA-based pipeline of Seurat (v4.3.0)⁵⁹. The integrated objects of each stage were further unsupervised clustered and visualized by Seurat.

For the nervous system analysis, the Python package Scanorama (v1.7.4) was used to integrate multiple sections and remove batch effects of multiple datasets⁶⁷. The unique molecular identifiers (UMIs) count matrices were normalized with Scanpy. Integrated data were scaled and underwent PCA for dimension reduction. Scanorama was then used to correct batch effects, and dimensionality reduction was performed with UMAP. Leiden algorithm was used to identify clusters.

Comparison with published human embryonic transcriptome data

To validate the effectiveness and consistency of the data, the CS14–15 E1S4 spatial transcriptome was utilized to integrate with the outer spatial transcriptome data (human S1)² with a similar Carnegie stage by Scanorama (v1.7.4), and it was plotted into UMAP and a spatial expression graph with clusters in common. The average expression for each cluster from the integrated file was extracted to conduct correlation analysis under Pearson method.

Transcriptome dynamics and lncRNA GO enrichment

To delineate transcriptional dynamics of coding and non-coding genes across CS12–23 stages, we performed UMI-based normalization of selected tissue and quantified gene counts (50 or more UMI per gene in each section) in a stage-specific manner. To further analyse the biological function of the long non-coding RNA, the top 150 expressed long non-coding RNAs in each tissue cluster of chips were selected to predict the GO term by Great (<https://great.stanford.edu/great/public/html/>).

GO enrichment analysis in the heart

Twenty-six substructures were identified within the 58 chips that included heart tissue. Each chip was processed in an integrated manner, with batch effects being mitigated through the application of batch balanced KNN (BBKNN; v1.6.0)⁶⁸. Substructures that encompassed over 1,000 bins were selected for analysis of DEGs and genes with adjusted $P < 0.05$ were subsequently subjected to GO enrichment analysis using clusterProfiler (v4.6.0)⁶⁰.

Cell–cell communication analysis

Cell–cell interactions between endocardium (sender cell) and trabecular ventricular cardiomyocytes (receiver cell) were inferred from spatially resolved transcriptomic data (bin50) using the CellChat package (v2.1.2)⁶⁹. A CellChat object was constructed from normalized expression matrices and spatial coordinates, with interactions restricted to intercellular distance (less than 200 μm). The human ligand–receptor database (CellChatDB.human) was subset to secreted signalling pathways. Overexpressed ligand–receptor pairs were identified (truncated mean, trim = 0.1) and communication probabilities were calculated.

Calculation of cell-type percentage in the brain substructure

A total of 65 embryo sections with good morphology and data quality were selected to extract the brain and spinal cord transcriptome data for re-clustering. Functional cell types, including excitatory neurons, inhibitory neurons, RGCs, astrocytes, microglia, oligodendrocytes, oligodendrocyte progenitor cells, neural progenitor cells and glia progenitor cells were annotated using previously reported markers^{70–73} (Supplementary Table 2d). Brain substructures were annotated using previously reported markers^{74–77}. The proportion of a cell type (for example, excitatory neurons) in a brain region (for example, the forebrain) was calculated as the bin50 number of the cell type (for example, excitatory neurons) divided by the total bin50 number of that region (for example, the forebrain).

Hierarchical clustering of gene sets in brain substructures

After performing the bin-level gene set enrichment analysis, we generated matrices of median enrichment scores for each specific brain substructure at various stages and compiled them into a unified score matrix. Hierarchical clustering was then applied to identify the general patterns of activity. The simplifyEnrichment (v1.14.0)⁷⁸ and ComplexHeatmap (v2.20.0)⁶¹ packages were utilized to group functional terms and create visualizations in the form of heatmaps and word clouds. The substructure optic recess was excluded from longitudinal analysis due to incomplete temporal profiling.

Spatial mapping of snRNA-seq data by cell2location

To spatially map the cell types identified from snRNA-seq data, we applied the principled Bayesian model implemented in cell2location (v0.1.3)⁷⁹. This model assigned the reference cell types to their corresponding spatial positions on matched snRNA-seq data and brain cell bins for each developmental stage (CS18 E1S6 and CS19 E2S2 mapping to CS19 snRNA-seq data, CS20 E1S6 mapping to CS20 snRNA-seq, and CS23 E1S2 mapping to CS23 snRNA-seq). After mapping, we obtained a complete map of whole-embryo sections characterized by precise cellular resolution. For more precise spatial mapping within brain regions, we further classified RGCs into eight subpopulations. A similar cell2location process was conducted to spatially map the neuronal cell types to cell bins within the brain regions.

Cell-type deconvolution of public scRNA-seq data by Tangram

Cell segmentation was performed for cell-type mapping analysis in the heart and gut as described above. The scRNA-seq data were then downloaded from the public database and aligned it to the spatial atlas using Tangram (v1.0.4)⁸⁰. The scRNA-seq data of the human embryonic heart was downloaded from the UCSC Cell Browser^{19,81} and the gut data of PCW7.9–8.4 (corresponding to CS22–23) were extracted from the fetal gut cell atlas⁸² for annotation.

Genes associated with virus entry and human developmental disorders

Genes encoding virus receptors and those linked with monogenic disease from the DDG2P database (v3.0)⁸³ were analysed. All chips at each stage were randomly sampled in proportion and then used in this analysis. After filtering out genes with count number less than 1, we obtained 41 virus entry receptors and entry factors, and 1,740 genes associated with developmental diseases. The expression levels in different organs at each developmental stage were aggregated, scaled and further visualized by R package ComplexHeatmap (v2.12.0)⁶¹.

Disease-associated gene set analysis in spatial transcriptomics

Disease-associated gene sets (11 categories including adrenal, cardiac, metabolic and neurodevelopmental disorders) were scored using Seurat's AddModuleScore, applied to Stereo-seq data from 26 human embryonic sections across 22 organ systems. Mean enrichment scores

were z-score normalized per gene set and hierarchically clustered to delineate organ-restricted versus multi-organ disorder patterns.

GRN construction and perturbation simulation

CellOracle (v0.18.0) was utilized to simulate the in silico transcription factor *HMGA2*-knockout effect, following the official tutorial. The workflow comprised four steps: (1) base GRN construction, (2) scRNA-seq data preprocessing and stratified downsampling, (3) GRN model construction, and (4) in silico perturbation analysis. The scRNA dataset from embryonic brain included 147,937 cells (29,480 genes) with pseudotemporal ordering (Monocle3) and annotated cell types.

Base GRN construction. The raw gene expression matrix was analysed using pySCENIC (v0.12.1) with GRNBoost2 to generate co-expression-based regulons, following the official tutorial. Regulons with feature importance scores of less than 3 were filtered out to eliminate weak interactions. The remaining regulons were formatted into a transcription factor–target gene list, as required by CellOracle, to generate the base GRN.

scRNA-seq data stratified downsampling and preprocessing. Dopaminergic and cholinergic neurons were excluded due to lack of pseudotemporal data, retaining 137,518 cells for further study. To reduce computational load, stratified downsampling was performed to retain 30,000 cells. GABA neurons, which constituted 60.2% of the dataset ($n = 82,780$), were disproportionately downsampled owing to their low *HMGA2* expression and limited perturbation response to transcription factor knockout. The downsampling strategy allocated equal numbers of GABA neurons and neuroblasts (the second-largest cell population, $n = 6,564$ after downsampling). The remaining cells ($n = 16,872$) were proportionally sampled from other cell types. The top 3,000 highly variable genes were identified using scanpy.pp.filter_genes_dispersion ($n_top_genes = 3,000$) and subjected to downstream analysis.

GRN model inference. The base GRN and non-log-transformed scRNA-seq data were input into CellOracle to instantiate an Oracle object. GRN computation and network preprocessing were performed as per the tutorial.

In silico perturbation and analysis. The *HMGA2* knockout was simulated by setting its expression to zero. CellOracle calculated cell-state transition probabilities ($n_neighbors = 300$) and visualized the results as vector fields on a two-dimensional grid ($n_grid = 40$, $min_mass = 0.0053$ and $scale_simulation = 0.1$). Pseudotemporal gradients were projected onto the two-dimensional grid ($n_grid = 40$, $min_mass = 0.0053$) using kNN regression. Developmental and perturbation vectors were quantitatively compared via inner product calculations. Positive perturbation scores indicated differentiation promotion, whereas negative scores denoted suppression.

Human–mouse Pall VZ rostrocaudal patterns

The spatial brain domains of human and mouse embryos were systematically digitized into columns along the rostrocaudal axis using the Spateo (v1.1.0)⁵³. For each brain section, the Pall VZ region was processed using the `extract_cluster_contours` function to generate a closed contour line that precisely delineates the boundary of the Pall VZ. Subsequently, leveraging the `st.dd.digitize` function within Spateo, the Pall VZ region was discretized into columns, with the axis defined by the rostral pole (x_min, y_min) and the caudal pole (x_max, y_max) serving as the reference. Normalized anteroposterior axis coordinates were then assigned to each column. These digitized columns were further aggregated into larger bins. Rostrocaudal expression patterns were identified through k -means clustering. ComplexHeatmap

(v2.12.0)⁶¹ packages were utilized to create visualizations in the form of heatmaps.

ChIP-seq analysis

Profile plots were generated using DeepTools to display chromatin signal patterns across genomic regions. Distinct signal patterns were identified through k -means clustering ($k = 4$). Gene body signal intensity (from transcription start site to transcription end site) was calculated in R and visualized using ggplot2.

Cross-species comparisons between human and mouse

For the interspecies analysis, the spatial transcriptome data of mouse embryos from E10.5 to E16.5 were retrieved from the MOSTA database⁴⁰. Mouse embryo sections included E10.5 E1S1, E10.5 E1S3, E10.5 E2S1, E11.5 E1S1, E11.5 E1S2, E11.5 E1S3, E12.5 E1S1, E12.5 E1S2, E12.5 E2S1, E13.5 E1S1, E13.5 E1S3, E13.5 E1S4, E14.5 E1S1, E14.5 E1S4, E14.5 E1S5, E16.5 E1S1, E16.5 E1S5 and E16.5 E12S5. All human embryo chips at each stage were randomly sampled in proportion and then used in this analysis. Correspondence of the developmental stage of human and mouse embryos⁸⁴ is shown in Supplementary Table 21. The count matrices for all sections were processed as follows: (1) the orthologous gene in human and mouse was obtained from Ensembl⁸⁵ and the homologous gene symbols were mapped from mouse to human using the biomaRt (v2.52.0)⁸⁶ and homologene (v1.4.68.19.3.27; <https://github.com/oganm/homologene>), (2) the mouse genes not orthologous to human were filtered, (3) the matrices were filtered based on parameters including the number of counts and features detected, the fraction of mitochondrial genes, and the percentage of ribosomal genes, and (4) the gene expression matrices of all sections from the same developmental stage were normalized and integrated using Seurat⁵⁹. The rds objects of the two species were further integrated at each corresponding developmental stage and subjected to clustering using the standard pipeline of Seurat.

FindMarkers function in Seurat was used to identify the DEGs between human and mouse embryos in the representative anatomical regions (brain, spinal cord, ear, liver, heart, lung, gut, kidney, skeletal muscle, bone and connective tissue) at each developmental stage with default parameters, and the normalized RNA assay data from integrated objects were used for calculations. GO enrichment analysis was performed using R package clusterProfiler (v4.4.1)⁶⁰.

To systematically identify the UTG and DTG for different tissues in human and mice, we aggregated the bin50 Stereo-seq data into pseudobulk RNA expression profiles across developmental stages and performed the Spearman correlation test between the expression vector of each gene and a linearly encoded developmental stage vector. Genes with significant positive correlations ($P < 0.05$ and $r > 0$) were defined UTG, whereas genes with significant negative correlations ($P < 0.05$ and $r < 0$) were defined DTG (Supplementary Table 23). Each gene group was further partitioned into two clusters based on cross-species expression divergence between human and mouse. GO enrichment analysis was then performed for each gene cluster using R package clusterProfiler (v4.4.1; Supplementary Table 24).

Integration of spatial transcriptome and genome for allelic expression

The embryonic samples used in the allelic analysis are listed in Supplementary Table 25a. The duplicate-marked transcriptome BAM file of each chip from Stereo-seq raw data was subjected to the following processes: tissue division (samtools, v1.15.1), sorting (samtools), splitting N cigar reads (SplitNCigarReads, GATK, v4.1.8.1), BQSR with known site from NCBI dbSNP and 1000 G project (BQSR and GATK), variant detection (HaplotypeCaller, GATK) with parameters ‘--ERC GVCF’ and ‘--do-not-run-physical-phasing’. The whole-genome sequencing of the embryo and decidua from CS14–15 E1, CS14–15 E3 and CS20 E2 were performed by DNBseq T1&T5 and then aligned to human genome GRCh38.p12 through SOAPnuke (default: ‘-n 0.1 -q 0.5 -l12 -Q2 -G2 -M2’)

Article

and BWA (built-in software of FPGA heterogeneous computing and multi-stream system (MegaBOLT, v5.0)). The genomic variant detection flow was the same as the transcriptome variant detection. The GVCf files of the embryonic genome, combined with the GVCf files (GenotypeGVCfs, GATK) of the transcriptome of all targeted tissues or maternal genome from the same embryo, transformed into the embryo genome and tissue-specific expression VCF (EaTSE VCF) and the embryo and mother VCF (EaM VCF). These VCF files passed filtering (--genotypegvcfs-stand-call-conf 30, --window 35, --cluster 3, FisherStrand (FS) > 60.0, Quality Depth (QD) < 2.0, Read Depth (DP) < 5) and were annotated in terms of ensGene, cosmic70, exac03 and clinvar_20190305 (GRCh38.p12) provided by annovar (V2020-06-07)⁸⁷.

SNP-level allelic imbalanced expression analysis

The high-quality embryonic heterozygous variant and homozygous expression in certain tissues were extracted into tissue-specific variant imbalanced allelic expression table (Supplementary Table 25d). Five embryos (CS14–15 E1, CS14–15 E2, CS14–15 E3, CS18 E1 and CS20 E2) were used to evaluate the reproducibility of tissue-specific SNP-level imbalanced allelic expression gene (excluding intron region).

Third-generation sequencing and phasing for allelic expression

To improve the phasing result in the difficult-to-map genome regions, third-generation nanopore sequencing for CS14–15 E1 and CS14–15 E3 embryos was performed with PromethION by BGI. Sequencing data were base identified by guppy5 and aligned to GRCh38.p12 with minimap2 (2.24-r1122). The genomic variant detection and phasing were then performed with Clair3 (v0.1-r12)⁸⁸. CS20 E2, additionally sequenced by Pacbio, underwent the same pipeline as CS14–15 embryos. The high-quality (Genotype Quality (GQ) > 10, DP > 4) embryonic heterozygous variant and maternal homozygous variant from the EaM VCF file were selected to generate the preliminary phasing table. With assistance from the phasing blocks of the third-generation phasing VCF, the embryonic heterozygous variants, corresponding to maternal heterozygous variants, were added to the final phasing table.

Phased imbalanced allelic expression analysis

Variants from CS14–15 E1, CS14–15 E3 and CS20 E2 were tagged with haplotype allelic information in reference to the previous final phasing table and consolidated into gene level (Supplementary Table 25e). The confident imbalanced allelic expression genes (the number of variant > 3 or the number of single variant > 300, and imprinted gene from Geneimprint Database) in each selected tissue were chosen to conduct subsequent allelic expression analysis. Variants of each confident imbalanced allelic expression genes were scanned by mpileup of samtools (v1.15.1), and plotted in allelic expression pattern with spatial location and pie plot by matplotlib package (v3.5.2). Furthermore, the allelic expression rate was counted by comparing the number of allelic reads in the same tissue and plotted through ggplot2 (v3.4.0). In addition, the paired Student's *t*-test for the imbalanced allelic expression gene was calculated in the aspect of sections by SciPy (v1.8.1). The allelic expression pattern was further validated by Sanger sequencing of genomic DNA and cDNA from these embryos (the primer sequence is in Supplementary Table 26c).

ISH assay

ISH was performed on 10- μ m cryosections with the RNAScope 2.0 multiplex fluorescent reagent kit assay (323100I, Advanced Cell Diagnostics) following the manufacturer's instructions. In situ probes to human *ALB* (600941-C1), *MYL7* (831731-C1), *PMEL* (410671-C1), *SHH* (600951-C1), *STMN2* (525211-C2), *LMX1A* (540661-C1), *MYH6* (555381-C1), *MYH7* (508201-C2), *ETS1* (533681-C1), *CDH5* (437451-C3), *KIAA1324L* (517661-C1), *RORA* (560761-C2), *HCN4* (495241-C3), *HMGA2* (467111-C1), *DCX* (489551-C2), *GAD1* (404031-C1), *NTCP* (548111-C2), *ASGR1* (424161-C2), *EGFR* (310061-C4), *PDGFRA* (604481-C3), *NRP2*

(424371-C2), *THY1* (430611-C4), *CD46* (430151-C3), *CD147* (472721-C3), *TMPRSS2* (470341-C3), *ACE2* (848151-C2), *EPCAM* (310281-C1) and *NEUROD2* (544301-C3) were used for target detection. Tissue sections were fixed in 10% neutral formalin reagent (G2161, Solarbio), dehydrated through an ethanol series, processed using standard pre-treatment conditions, and followed by incubation with target probe as per the RNAScope 2.0 multiplex fluorescent reagent kit assay protocol. TSA-plus fluorescein, Cy3 and Cy5 fluorophores were used at 1:1,500 dilution for the manual assay. Large slides were imaged by PANNORAMIC MIDI II (3D Histech) and viewed by Slide Converter. High-definition images were taken by a NIKON AXR HD25 laser scanning confocal microscope.

In vivo functional validation of human heart conduction genes

Knockdown of *Roraa*, *Rorab*, *Shox2*, *Vsn11b* and *Elapor2b* was performed in zebrafish embryos by CRISPR–Cas9 editing (Fig. 3h). Zebrafish embryos of unknown sex were randomly assigned to experimental groups. Four to six CRISPR target sites for each gene were designed with CRISPR (<https://crispor.gi.ucsc.edu/>). The oligo pairs for each target are shown in Supplementary Table 26a. The single guide RNAs were generated with the T7 RiboMAX Express Large Scale RNA Production System (P1320, Promega) and purified with the RNeasy Mini Kit (74104, Qiagen). Mixed single guide RNAs for every gene (final concentration of 200 ng μ l⁻¹) and the Cas9 protein (500 ng μ l⁻¹; E365, Novoprotein) were co-injected into one-cell-stage *Tg(myl7:H2B-eGFP)* embryos (2 nl per embryo) using the MPPI-3 Pressure Injector (Smith-Root). The injected embryos were treated with PTU (0.003%; N816213, Macklin) from 24 h post-fertilization and the heart rate was recorded by a fluorescence confocal microscope (BX51WI, Olympus) 2 days post-fertilization. The primer probes for the target genes detection are listed in Supplementary Table 26b. All procedures on live animals have been approved by the Ethics Committee of the Shanghai Medical College, Fudan University. The sample size was not calculated using a statistical power analysis but was chosen based on the previous experience of our laboratory, which has proven sufficient to detect robust effects. Researchers were not blinded to sample identity.

Immunofluorescent staining of pacemaker cells

Two days post-fertilization *Tg(myl7:H2B-eGFP)* transgenic zebrafish embryos were fixed in 1% formaldehyde for 1 h at room temperature, then washed three times in 1 $\times$ PBS. The fixed embryos were blocked in 1 $\times$ PBS containing 10% donkey serum (abs935, Absin), 2 mg ml⁻¹ BSA (A8020, Solarbio) and 0.2% saponin (SAE0073, Sigma) for 1 h at room temperature and then incubated in 200 μ l of Islet-1 (Isl1) primary antibody (GTX128201, GeneTex; 1:500 dilution) and GFP-Tag (7G9) mouse monoclonal antibody (M20004S, Abmart; 1:1,000 dilution) overnight at 4 $^{\circ}$ C. The embryos were then stained with anti-rabbit Alexa 594 (ab150084, Abcam; 1:500 dilution) and anti-mouse Alexa 488 (ab150113, Abcam; 1:500 dilution) secondary antibodies for 2 h in the dark at room temperature. Images were taken by a NIKON AXR HD25 laser scanning confocal microscope and raw data were processed with NIS Elements AR and ImageJ (v2.9.0).

Heart rate analysis

Cardiac rhythm quantification in zebrafish embryos at 2 days post-fertilization was performed using a fluorescence confocal microscopy (BX51WI, Olympus). Original cardiac motion videos were acquired in MP4 format at 40 frames per second and subsequently converted to AVI format using Adobe Premiere Pro 2022 to ensure software compatibility. For quantitative analysis, videos were processed in ImageJ (v2.9.0) with the following protocol: (1) manual selection of the cardiac region of interest, (2) application of the Time Series Analyzer v3 plugin for temporal signal extraction, and (3) peak detection via the Get Average function to derive heart rate measurements. To validate automated analyses, all recordings underwent secondary manual verification

through frame-by-frame beat counting, ensuring measurement accuracy and reliability.

Reporting summary

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

Data availability

The dataset resource for an interactive survey is accessible to the public (<https://db.genomics.cn/hesta>). The data that support the findings of this study have been deposited into the Gene Expression Omnibus (GEO), Genome Sequence Archive in National Genomics Data Center (GSA-Human) and CNGB Sequence Archive (CNSA) of China National GeneBank DataBase (CNGBdb). All processed data generated by Stereo-seq have been deposited to the GEO (<https://www.ncbi.nlm.nih.gov/query/acc.cgi?acc=GSE326326>, accession code GSE326326) and the CNGB Nucleotide Sequence Archive (<https://db.cngb.org/stomics/project/STT0000025/>, accession code STT0000025). All processed data generated by snRNA-seq have been deposited to the GEO (<https://www.ncbi.nlm.nih.gov/query/acc.cgi?acc=GSE325829>, accession code GSE325829) and the CNGB Nucleotide Sequence Archive (<https://db.cngb.org/search/project/CNP0004028/>, accession code CNP0004028). All raw data generated by Stereo-seq have been deposited to the GSA-Human (<https://ngdc.cncb.ac.cn/gsa-human/browse>, accession code HRA017263) and the CNGB Nucleotide Sequence Archive (<https://db.cngb.org/stomics/project/STT0000025/>, accession code STT0000025). All raw data generated by snRNA-seq have been deposited to the GSA-Human (<https://ngdc.cncb.ac.cn/gsa-human/browse>, accession code HRA017541). All raw data generated by whole-genome sequencing have been deposited to the GSA-Human (<https://ngdc.cncb.ac.cn/gsa-human/browse>, accession code HRA004300) and the CNGB Nucleotide Sequence Archive (<https://db.cngb.org/search/project/CNP0004028/>, accession code CNP0004028). All raw data generated by Nanopore and PacBio have been deposited to the GSA-Human (<https://ngdc.cncb.ac.cn/gsa-human/>, accession code HRA004299) and the CNGB Nucleotide Sequence Archive (<https://db.cngb.org/search/project/CNP0004028/>, accession code CNP0004028). The published datasets analysed for this study were retrieved from the GEO repository (GSE155121 and GSE157329), UCSC Cell Browser (<https://cells.ucsc.edu/?ds=hoc>), Gut Cell Survey (<https://www.gutcellatlas.org/>) and STOmics DB (<https://db.cngb.org/stomics/mosta/>). Source data are provided with this paper.

Code availability

The Python and R codes used to process and analyse the Stereo-seq data are available on GitHub (<https://github.com/Yuejiao2025/HESTA>). These scripts orchestrate the use of established bioinformatics tools (for example, cell2location and scvi-tools) and are provided for reproducibility. They are not central to the conceptual conclusions of the paper.

51. Wei, X. et al. Single-cell Stereo-seq reveals induced progenitor cells involved in axolotl brain regeneration. *Science* **377**, eabp9444 (2022).
52. Dobin, A. et al. STAR: ultrafast universal RNA-seq aligner. *Bioinformatics* **29**, 15–21 (2013).
53. Qiu, X. et al. Spatiotemporal modeling of molecular holograms. *Cell* **187**, 7351–7373.e61 (2024).
54. Beucher, S. The watershed transformation applied to image segmentation. *Scan. Microsc.* **1992**, 299–314 (1992).
55. Wolf, F. A., Angerer, P. & Theis, F. J. SCANPY: large-scale single-cell gene expression data analysis. *Genome Biol.* **19**, 15 (2018).
56. Palla, G. et al. Squidpy: a scalable framework for spatial omics analysis. *Nat. Methods* **19**, 171–178 (2022).
57. Lopez, R., Regier, J., Cole, M. B., Jordan, M. I. & Yosef, N. Deep generative modeling for single-cell transcriptomics. *Nat. Methods* **15**, 1053–1058 (2018).
58. Fan, C. et al. irGSEA: the integration of single-cell rank-based gene set enrichment analysis. *Brief. Bioinform.* **25**, bbae243 (2024).

59. Butler, A., Hoffman, P., Smibert, P., Papalexi, E. & Satija, R. Integrating single-cell transcriptomic data across different conditions, technologies, and species. *Nat. Biotechnol.* **36**, 411–420 (2018).
60. Wu, T. et al. clusterProfiler 4.0: a universal enrichment tool for interpreting omics data. *Innovation* **2**, 100141 (2021).
61. Gu, Z. Complex heatmap visualization. *iMeta* **1**, e43 (2022).
62. Aibar, S. et al. SCENIC: single-cell regulatory network inference and clustering. *Nat. Methods* **14**, 1083–1086 (2017).
63. Suo, S. et al. Revealing the critical regulators of cell identity in the mouse cell atlas. *Cell Rep.* **25**, 1436–1445.e3 (2018).
64. DeTomaso, D. & Yosef, N. Hotspot identifies informative gene modules across modalities of single-cell genomics. *Cell Syst.* **12**, 446–456.e9 (2021).
65. Cao, J. et al. The single-cell transcriptional landscape of mammalian organogenesis. *Nature* **566**, 496–502 (2019).
66. Qiu, C. et al. Systematic reconstruction of cellular trajectories across mouse embryogenesis. *Nat. Genet.* **54**, 328–341 (2022).
67. Hie, B., Bryson, B. & Berger, B. Efficient integration of heterogeneous single-cell transcriptomes using Scanorama. *Nat. Biotechnol.* **37**, 685–691 (2019).
68. Polański, K. et al. BBKNN: fast batch alignment of single cell transcriptomes. *Bioinformatics* **36**, 964–965 (2020).
69. Jin, S., Plikus, M. V. & Nie, Q. CellChat for systematic analysis of cell-cell communication from single-cell transcriptomics. *Nat. Protoc.* **20**, 180–219 (2025).
70. Fan, X. et al. Single-cell transcriptome analysis reveals cell lineage specification in temporal-spatial patterns in human cortical development. *Sci. Adv.* **6**, eaaz2978 (2020).
71. Fu, Y. et al. Heterogeneity of glial progenitor cells during the neurogenesis-to-gliogenesis switch in the developing human cerebral cortex. *Cell Rep.* **34**, 108788 (2021).
72. Hatakeyama, J. et al. Hes genes regulate size, shape and histogenesis of the nervous system by control of the timing of neural stem cell differentiation. *Development* **131**, 5539–5550 (2004).
73. Fan, X. et al. Spatial transcriptomic survey of human embryonic cerebral cortex by single-cell RNA-seq analysis. *Cell Res.* **28**, 730–745 (2018).
74. Gui, Y. et al. Single-nuclei chromatin profiling of ventral midbrain reveals cell identity transcription factors and cell-type-specific gene regulatory variation. *Epigenetics Chromatin* **14**, 43 (2021).
75. Luo, Z. et al. Human fetal cerebellar cell atlas informs medulloblastoma origin and oncogenesis. *Nature* **612**, 787–794 (2022).
76. Gonçalves, C. S., Le Boiteux, E., Arnaud, P. & Costa, B. M. HOX gene cluster (de)regulation in brain: from neurodevelopment to malignant glial tumours. *Cell. Mol. Life Sci.* **77**, 3797–3821 (2020).
77. Braun, E. et al. Comprehensive cell atlas of the first-trimester developing human brain. *Science* **382**, eadf1226 (2023).
78. Gu, Z. & Hübschmann, D. simplifyEnrichment: a Bioconductor package for clustering and visualizing functional enrichment results. *Genom. Proteom. Bioinform.* **21**, 190–202 (2023).
79. Kleshchevnikov, V. et al. Cell2location maps fine-grained cell types in spatial transcriptomics. *Nat. Biotechnol.* **40**, 661–671 (2022).
80. Biancalani, T. et al. Deep learning and alignment of spatially resolved single-cell transcriptomes with Tangram. *Nat. Methods* **18**, 1352–1362 (2021).
81. Speir, M. L. et al. UCSC Cell Browser: visualize your single-cell data. *Bioinformatics* **37**, 4578–4580 (2021).
82. Elmentaite, R. et al. Single-cell sequencing of developing human gut reveals transcriptional links to childhood Crohn's disease. *Dev. Cell* **55**, 771–783.e5 (2020).
83. Wright, C. F. et al. Genetic diagnosis of developmental disorders in the DDD study: a scalable analysis of genome-wide research data. *Lancet* **385**, 1305–1314 (2015).
84. Zhai, J., Xiao, Z., Wang, Y. & Wang, H. Human embryonic development: from peri-implantation to gastrulation. *Trends Cell Biol.* **32**, 18–29 (2022).
85. Yates, A. D. et al. Ensembl 2020. *Nucleic Acids Res.* **48**, D682–D688 (2020).
86. Durinck, S., Spellman, P. T., Birney, E. & Huber, W. Mapping identifiers for the integration of genomic datasets with the R/Bioconductor package biomaRt. *Nat. Protoc.* **4**, 1184–1191 (2009).
87. Wang, K., Li, M. & Hakonarson, H. ANNOVAR: functional annotation of genetic variants from high-throughput sequencing data. *Nucleic Acids Res.* **38**, e164 (2010).
88. Zheng, Z. et al. Symphonizing pileup and full-alignment for deep learning-based long-read variant calling. *Nat. Comput. Sci.* **2**, 797–803 (2022).
89. Han, X. et al. Construction of a human cell landscape at single-cell level. *Nature* **581**, 303–309 (2020).

Acknowledgements This work is supported by the National Natural Science Foundation of China (82088102 and 82192873), National Key Research and Development Plan (2024YFA1803000, 2022YFC2703500, 2022YFC2703600, 2022YFC2703803, 2022YFC2703001 and 2022YFC3400400), National Natural Science Foundation of China (82571953, 82171688, 92357306 and 82192864), Shanghai Clinical Research Center for Gynecological Diseases, Shanghai Urogenital System Diseases Research Center and Shanghai Frontiers Science Research Center of Reproduction and Development. We thank all of the participating patients for their kind contributions; L. Wang, B. Jiang, Y. Zhang, J. Jiao and Y. Chen for the experimental assistance; L. Liu, M. A. Esteban, S. Liu, Y. Gu, Y. Li, Y. Bai, L. Han, S. Liao, A. Chen, K. Ma, X. Wang, Q. Tong, Y. Wu, X. Li, S. Cao, L. Miao and R. Zhang for their valuable scientific advice and technical supports; H. Yang for valuable assistance in model organism establishment during manuscript revision; X. Zhai for rigorous oversight of ethical principles; B. Zhu for insightful advice on allele-specific expression analysis; L. Ye for assistance with the graphical work in this paper; Y. Xu; and L. Yu and J. Jiao for providing access to their specialized datasets.

Author contributions H.H. conceived the idea. H.H., X.X., G.D., Y. Gao, J.P., H. Yang and L.Z. made the study design. J.P., Y. Li, Y. Gao, G.D., H. Yang, X.X. and H.H. supervised the work. J.P. and G.D. designed the sample collection protocol. H.C., J.P., Z.L., Y.C., Y.T., X.Y., G.Z. and Y.X. collected the samples with the help of G.D., Q.C., Y. Guan, N.M., H. Lu, X. Li, T.Z. and Congjian

Article

Xu, Y. Gao, Y. Li, Z.L. and X.J. designed the spatial transcriptome experiments. Y. Li, Z.L., Q.L., J.P., M.Z., S.S., G.Z., H.C., Ying Zhang, Z.W., Z.O., F.M., Q.Q., B.J., Y.R., L.W., Y. Wei, J.N., Chunyan Xu, H. Yan, X.W., H. Li, Y. Zhu, C.W., T.Y., X.C., Yuxin Zhang, M.Y. and M.L. performed the bioinformatics analysis, statistical analysis and result visualization with the help of Y. Gao, J.P., G.D., X.J., H. Yang, Z.L., G.Z., Y.X., Y. Wei, Y.C., Y. Lei, J.S., H.H., X. Liu, Y. Wu and Chenming Xu. J.P., G.Z., Y.M. and G.D. conducted the ISH experiments. J.P., G.Z., Z.L. and G.L. performed the model organism validation experiments. Q.W. and H. Yang performed the zebrafish validation experiments. J.P., Y. Li, Z.L., Q.L., H.C., M.Z., G.Z. and S.S. wrote the first draft of the manuscript. J.P., Y. Gao, G.D., H.H., H. Yang, X.X., Z.L. and Y. Li revised and finalized the manuscript.

Competing interests None of the results in this study has been used for patent application. X.X. is the key inventor of DNBSEQ and Stereo-seq technologies covered in patents and patents applications, including ‘Array and method for detecting spatial information of nucleic acids’ (applicants: BGI Research and MGI Tech; inventors: A. Chen, X.X., J. Yang, L. Liu, O. Wang, Y. Li, S. Liao, G. Tang, Y. Jiang, C. Xu, M. Ni, W. Zhang, R. Drmanac and S. Drmanac; application numbers: CN202080036185.5 and RU2021135125), ‘Automatic control system of biochip and automatic control method thereof’ (applicant: BGI Research; inventors: Y. Hong, H. Kuang, Q. Yu, Q. Li, X. Sun, F. He, J. Xu, M. Shen, A. Chen, Y. Li, W. Zhang and X.X.; application number: PCT/CN2022/123075), ‘Biochip automatic control method and automatic control system thereof’ (applicant: BGI Research; inventors: F. He, X. Lei, A. Chen, Z. Qiu, Y. Hong, Z. Wan, M. Shen, A. Chen, Y. Li, W. Zhang and X.X.; application number: PCT/CN2022/123191), ‘Biochip and use and preparation method thereof’ (applicant: BGI Research; inventors: Y. Hong, Q. Li, A. Chen, Y. Li, W. Zhang and X.X.; application number: PCT/CN2022/123027),

‘Biochip container and methods for preparation and use thereof’ (applicant: BGI Research; inventors: Q. Yu, Y. Hong, Q. Li, A. Chen, Y. Li, W. Zhang and X.X.; application number: PCT/CN2022/127996), ‘Freezing and embedding device’ (applicants: BGI Research and BGI; inventors: J. Yi, L. Cui, S. Liao, A. Chen, W. Zhang and X.X.; application number: CN202223054698.4), ‘Methods, apparatus and applications for preparing candidate sequencing probe sets’ (applicant: MGI Tech; inventors: X.X., H. Jiang, C. Geng, G. Fan, E. Liang and Z. Zhu; application number: CN201610075006.4), and ‘Modified nucleoside or nucleotide’ (applicant: BGI Research; inventors: X.X., B. Teng, W. Zhang, A. Chen, H. Li, S. Yan, S. Zhuo, L. Shen, Y. Zhang, N. Gao, J. Zhao and S. Liao; application number: PCT/CN2021/125262). X.X. and L.Z. are employees of BGI Genomics, and hold stocks in BGI. Y. Li, Q.L., M.Z., Ying Zhang, Z.W., H. Yan, F.M., Q.Q., Y.R., B.J., L.W., J.N., Y. Wei, M.L., Chunyan Xu, X.W., H. Li, Y. Zhu, C.W., T.Y., X.C., Yuxin Zhang, M.Y., Z.O., Y. Lei, X.J. and Y. Gao are employees of BGI Research, and hold no intellectual property and commercial interests in DNBSEQ and Stereo-seq. J.P., Z.L., H.C., S.S., G.Z., Y.C., Y.T., Q.W., Y.X., G.L., Y.M., Q.C., Y. Guan, N.M., H. Lu, X. Li, T.Z., X.Y., X. Liu, J.S., Congjian Xu, Y. Wu, Chenming Xu, H. Yang, G.D. and H.H. declare no competing interests.

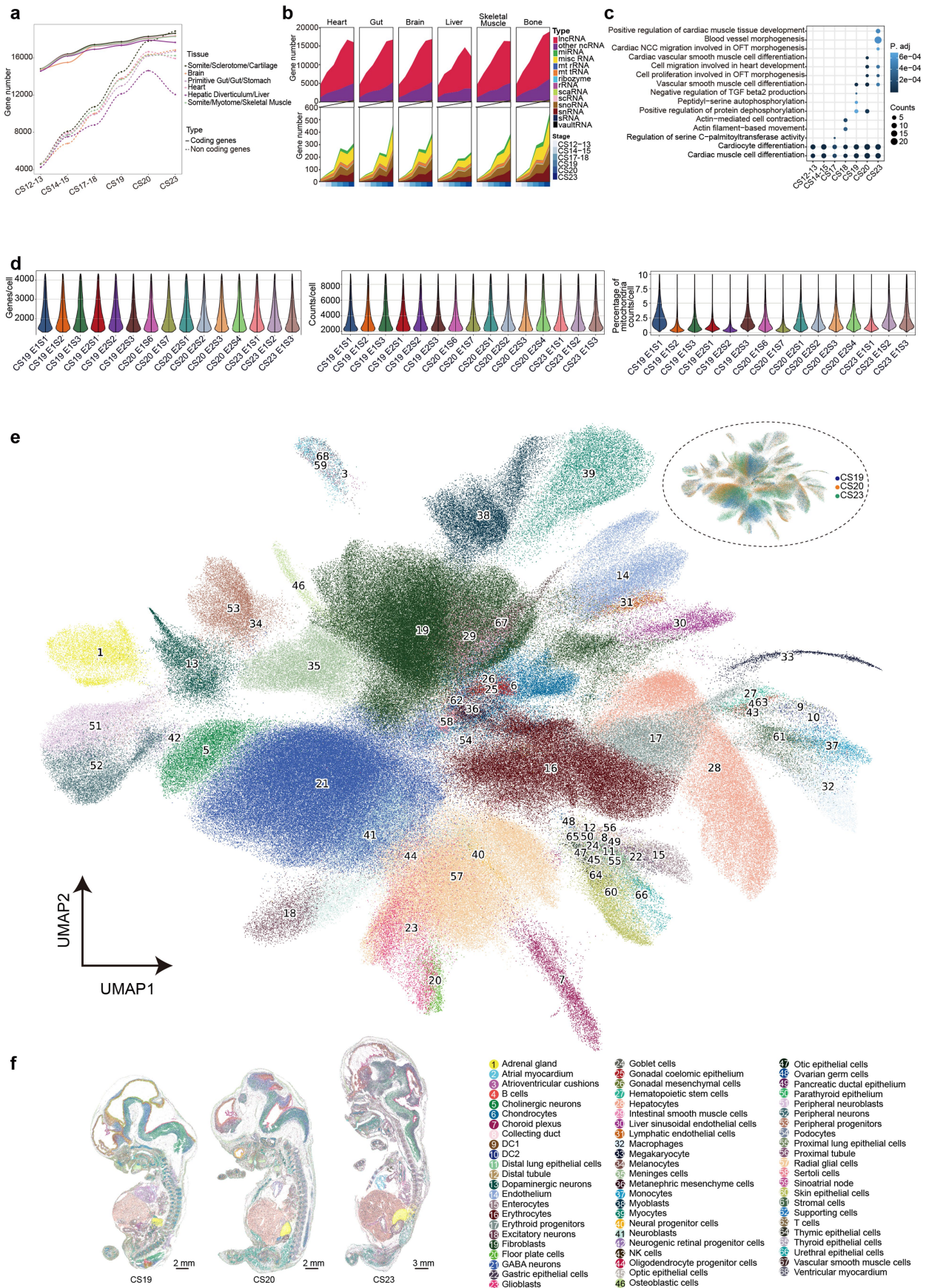
Additional information

Supplementary information The online version contains supplementary material available at <https://doi.org/10.1038/s41586-026-10545-0>.

Correspondence and requests for materials should be addressed to Lijian Zhao, Hongbo Yang, Ya Gao, Guolian Ding, Xun Xu or Hefeng Huang.

Peer review information *Nature* thanks Konstantin Khodosevich, Malte Spielmann and the other, anonymous, reviewer(s) for their contribution to the peer review of this work.

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



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

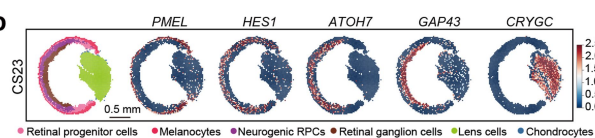
Extended Data Fig. 1 | Dynamics of coding and non-coding RNA across CS12-23, and, annotation of snRNA-seq data. **a**, Temporal progression of coding and non-coding transcripts in the main organs across CS12-23. **b**, The distribution and abundance of non-coding RNA subtypes in the main organs across CS12-23. lncRNA, long noncoding RNA; miRNA, microRNA; miscRNA, miscellaneous RNA; mt rRNA, mitochondrial ribosomal RNA; mt tRNA, mitochondrial transfer RNA; rRNA, ribosomal RNA; scaRNA, small cajal body-specific RNA; scRNA, small cytoplasmic RNA; snoRNA, small nucleolar RNA; snRNA, small nuclear RNA; sRNA, small non-coding RNA; vaultRNA, vault RNA.

c, The GO terms of lncRNA predicted target genes in heart across CS12-23. NCC, neural crest cell; OFT, outflow tract; TGF, transforming growth factor. **d**, Violin plot showing the count number of genes/ the count number of UMIs/ percentage of mitochondria in each cell. **e,f**, Annotation of snRNA-seq data. **e**, UMAP of all cells colored by cell type for the annotation of snRNA-seq data. The inset dashed circle shows the same UMAP colored by developmental stage. **f**, Mapping of snRNA data to corresponding adjacent spatial transcriptomic data using Cell2location. Colors and numbers in **e** and **f** correspond to the 68 listed major cell cluster annotations.

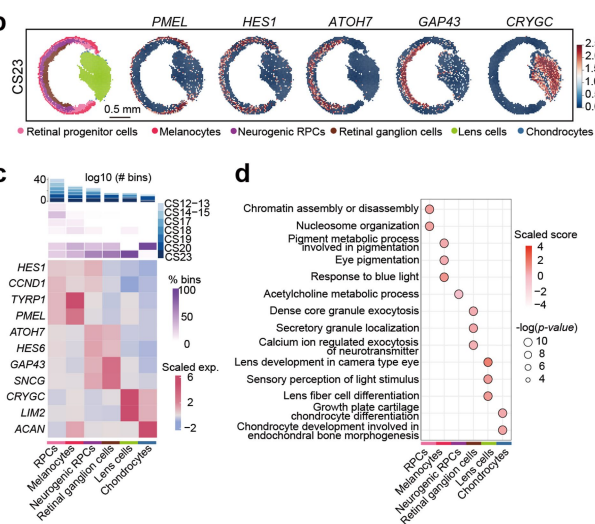
a



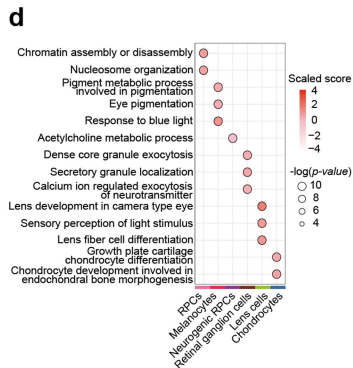
b



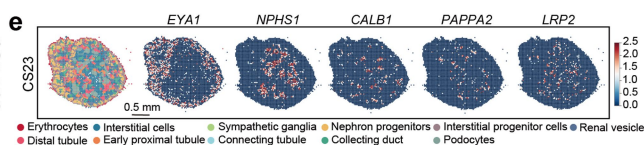
c



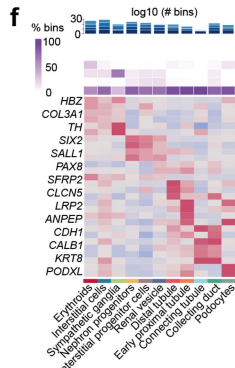
d



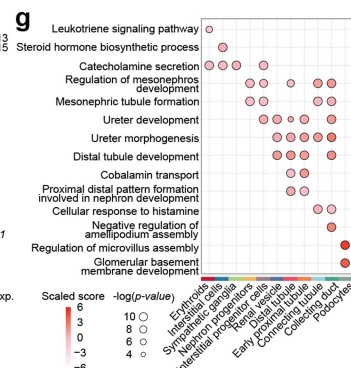
e



f



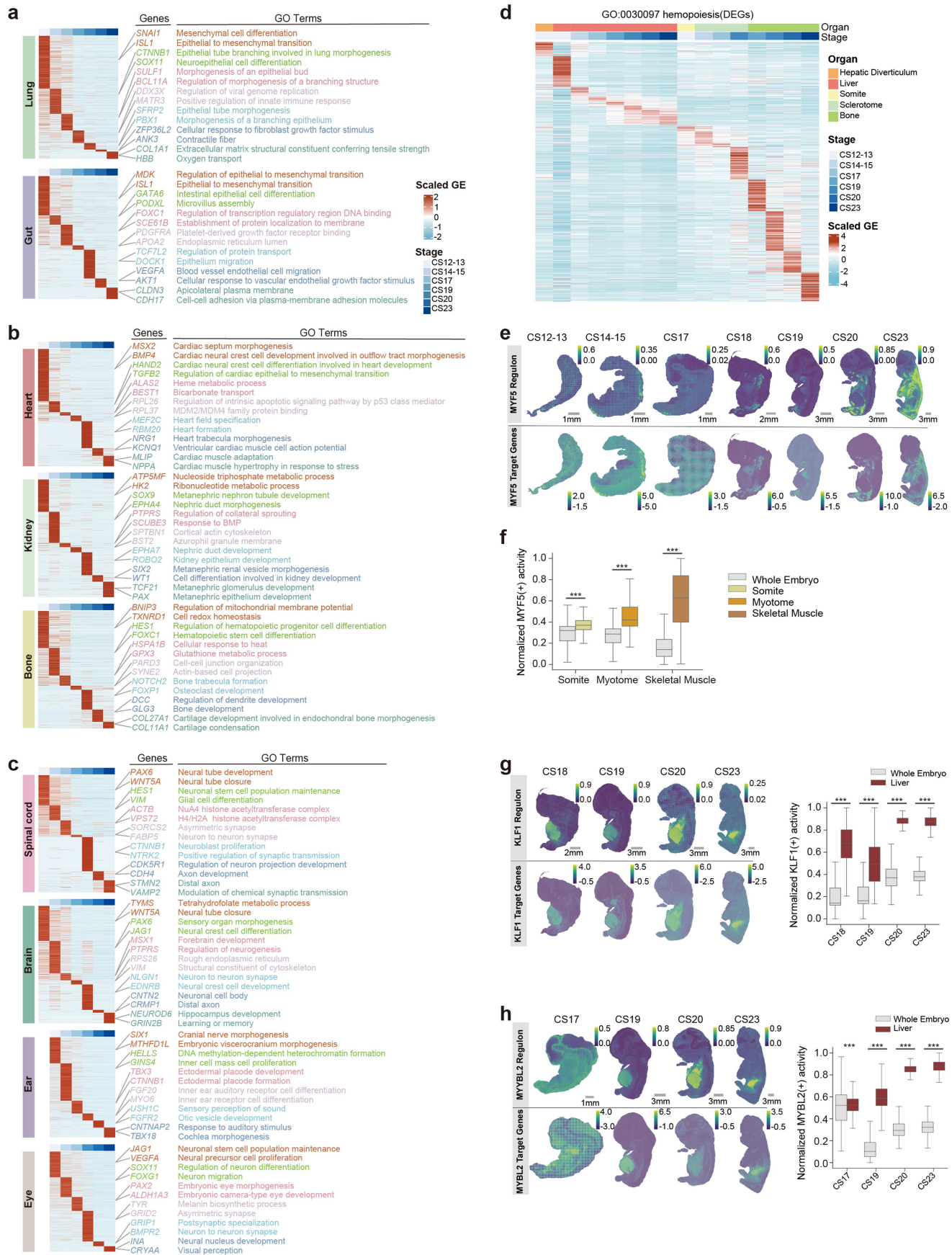
g



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

Extended Data Fig. 2 | Spatiotemporal mapping of substructure during human embryonic organogenesis. **a**, Schematic diagram of substructures in representative sections across CS12-23. 198 distinct anatomical substructures across 50 organ systems were identified and color-coded accordingly. **b**, Spatial localization of substructure-enriched marker genes in the eye. **c**, Heatmap showing scaled expression profile of ocular marker genes. The upper stacked barplot and middle annotation heatmap each displays the log-transformed bin counts and bin proportion of ocular substructures across developmental stages. RPCs, retinal progenitor cells. **d**, Dotplot showing enriched pathways among ocular substructures. **e**, Spatial distribution substructure-enriched

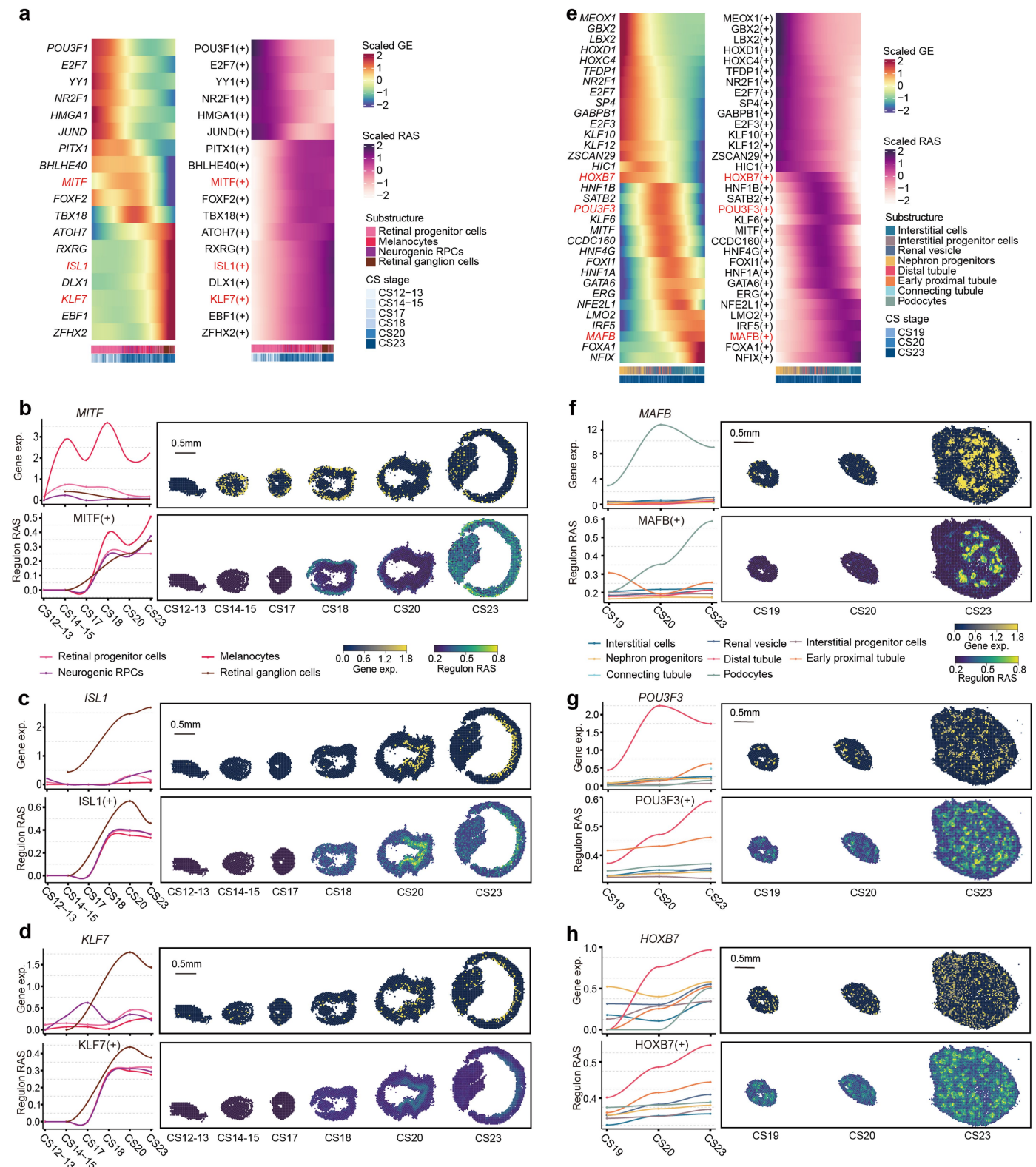
genes in the kidney. **f**, Heatmap showing scaled expression profile of renal marker genes. The upper stacked barplot and middle annotation heatmap each displays the log-transformed bin counts and bin proportion of renal substructures across developmental stages. **g**, Dotplot showing enriched pathways among renal substructures. For panel **d** and panel **g**, differential pathway activity was assessed by the two-tailed Wilcoxon rank-sum test on UCell scores, with *p*-values adjusted for multiple comparisons via the Benjamini-Hochberg procedure. Dot color represents the scaled UCell activity score and dot size represents the statistical significance -log10 (adjusted *p*-value).



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

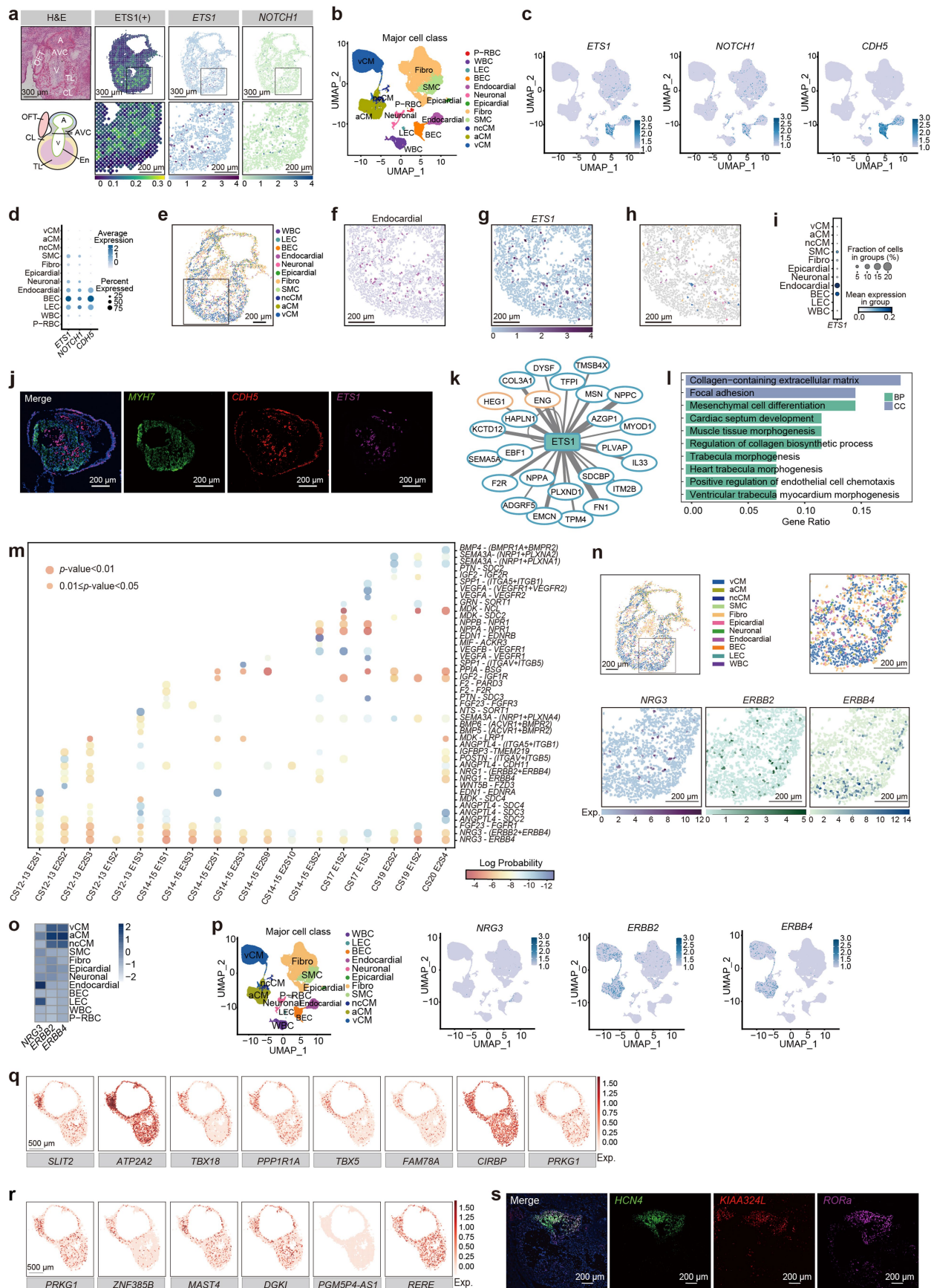
Extended Data Fig. 3 | DEGs, GO and regulon analysis. a-c, GO terms of organs derived from endoderm (a), mesoderm (b) and ectoderm (c). **d,** Heatmap showing the representative DEGs for hemopoiesis in liver and somite/sclerotome/bone. **e,** Spatial patterns of MYF5 regulon activity and target gene expressions from CS12 to CS23. **f,** Boxplot showing normalized MYF5 regulon activity among somite, myotome and skeletal muscle. n = 71,074 bins (whole embryo) and 8,621 bins (somite) examined in two biologically independent samples for somite, n = 200,205 bins (whole embryo) and 8,677 bins (myotome) examined in two biologically independent samples for myotome, n = 763,470 bins (whole embryo) and 38,679 bins (skeletal muscle) examined in four biologically independent samples for skeletal muscle. $p < 1.0 \times 10^{-100}$. **g,** Spatial patterns of KLF1 regulon activity and target gene expressions from CS18 to CS23. Boxplot showing normalized KLF1 regulon activity enriched in liver. n = 132,277 bins (whole embryo) and 11,540 bins (liver) examined over one biologically independent sample for CS18, n = 263,202 bins (whole embryo) and 16,945 bins (liver) examined in one biologically independent sample for

CS19, n = 32,858 bins (whole embryo) and 2,343 bins (liver) examined in one biologically independent sample for CS20, n = 87,350 bins (whole embryo) and 3,530 bins (liver) examined in one biologically independent sample for CS23. $p < 1.0 \times 10^{-100}$. **h,** Spatial patterns of MYBL2 regulon activity and target genes expression at CS17, CS19, CS20 and CS23. Boxplot showing normalized MYBL2 regulon activity enriched in liver. n = 66,586 bins (whole embryo) and 2,800 bins (liver) examined in one biologically independent sample for CS17, n = 510,985 bins (whole embryo) and 30,785 bins (liver) examined in one biologically independent sample for CS19, n = 16,977 bins (whole embryo) and 1,219 bins (liver) examined in one biologically independent sample for CS20, n = 87,350 bins (whole embryo) and 3,530 bins (liver) examined in one biologically independent sample for CS23. $p = 7.8 \times 10^{-7}$ for CS17, $p < 1.0 \times 10^{-100}$ for CS19, CS20 and CS23. For **f-h**, center line denotes median value, whisker lengths are within 1.5 interquartile range of the 25th and 75th percentile box boundaries, and outlier points are not shown. Statistical significance was assessed using the two-tailed Mann-Whitney U test. *** $p < 0.001$.



Extended Data Fig. 5 | Spatiotemporal dynamics of representative gene-regulon networks in ocular and renal substructures. a, Pseudotemporal profiles of gene expression (left) and RAS (right) of TFs in ocular system. RPCs, retinal progenitor cells. RAS, regulon activity score. **b-d,** For ocular substructures, left panels depict the temporal profiles of gene expression

and RAS of each TF along developmental stages. Right panels show spatial pattern of corresponding transcriptomic expression and regulon activity. **e,** Pseudotemporal patterns of gene expression (left) and RAS (right) of TFs in renal system. **f-h,** Coordinated gene-regulon activation with spatiotemporal organization were shown for renal substructures.

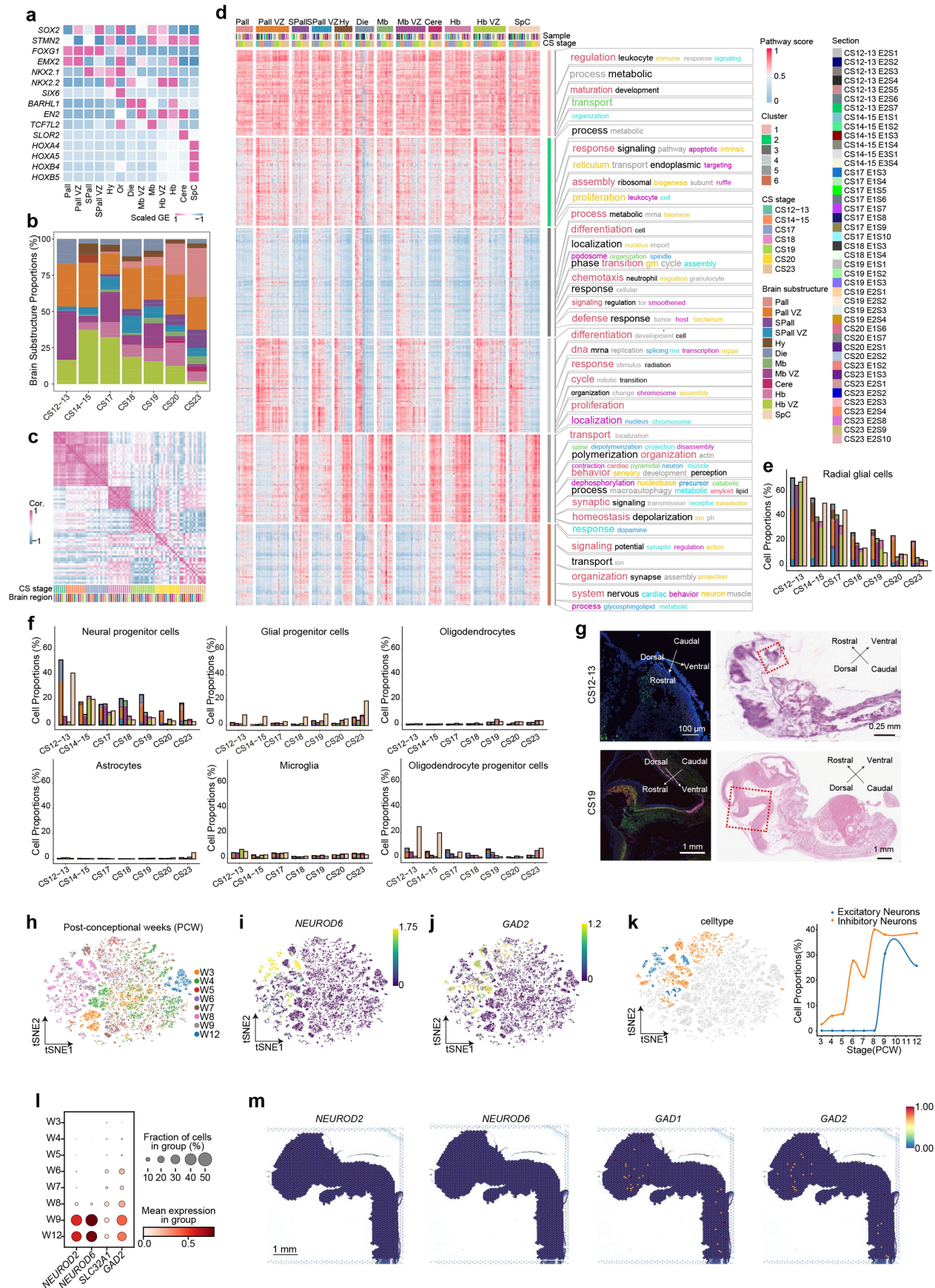


Extended Data Fig. 6 | See next page for caption.

Article

Extended Data Fig. 6 | GRNs governing cardiac development at the substructure level, and, signaling interactions between endocardium and trabecular ventricular cardiomyocytes. **a**, H&E staining and schematic diagram of trabecular and compact layers of CS14-15 E3S3 heart. The bin50 spatial distribution of the *ETSI* regulon activity, and, cell-segmented spatial distribution of *ETSI* and *NOTCH1* expression. CL, compact layer; TL, trabecular layer; A, atrium; V, ventricle; OFT, outflow tract; En, endocardium; AVC, atrioventricular canal. Results are representative of 7 sections (from four biologically independent samples) with similar outcomes. **b**, UMAP visualization of major cell types in the developing human heart between 9 and 16 post-conception weeks (PCW), based on a public single-cell RNA sequencing (scRNA-seq) dataset. P-RBC, platelet-red blood cells; WBC, white blood cells; LEC, lymphatic endothelial cells; BEC, blood endothelial cells; Fibro, fibroblasts; SMC, smooth muscle cells; ncCM, non-chambered cardiomyocytes; aCM, atrial cardiomyocytes; vCM, ventricular cardiomyocytes. **c**, UMAP visualization of the expression profiles of *ETSI*, *NOTCH1*, and *CDH5* (pan-endothelial marker). **d**, Dotplot illustrating the expression profiles of *ETSI*, *NOTCH1*, and *CDH5* using the scRNA-seq dataset. **e**, Stereo-seq data of the CS14-15 E3S3 heart, with segmented cells color-coded by the cell-type annotation of the 9 PCW cardiac scRNA-seq profile, mapped to their spatial positions using Tangram. **f,g**, Spatial distribution of endocardial cells (**f**) and *ETSI* (**g**) within the squared regions highlighted in (**e**). **h**, Spatial localization of *ETSI*⁺ cells, color-coded by cell type. **i**, Dotplot depicting the expression profiles of *ETSI* using cell-segmented Stereo-seq data. **j**, RNA-scope analysis showing the expression of *ETSI*, *MYH7*

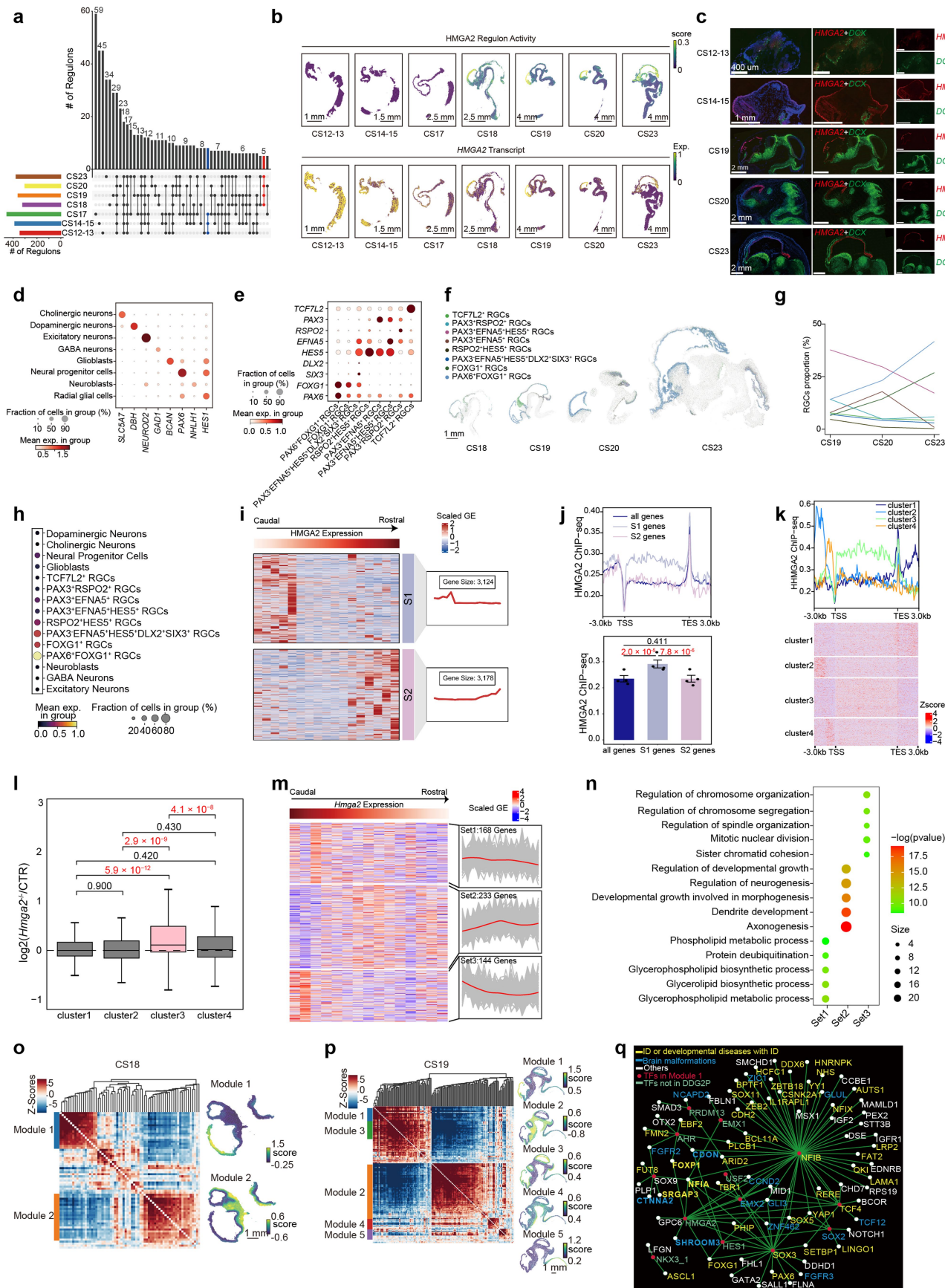
(ventricular cardiomyocyte marker), and *CDH5* in the human embryonic heart. The images are representative of 3 independent experiments. **k**, Regulatory network of *ETSI*, in which genes highlighted with orange circles represent genes for trabecular morphogenesis. The line width represents the weight of motif enrichment. **l**, Barplot exhibiting the representative GO enrichment terms of genes in (**k**). **m**, Dotplot illustrating specific signaling interactions between endocardium and trabecular ventricular cardiomyocytes across Carnegie stages (CS) 12–20. **n**, Stereo-seq data of the CS14-15 E3S3 heart, with segmented cells color-coded by the cell-type annotation of the 9 post-conception weeks (PCW) cardiac scRNA-seq profile¹⁹ mapped to their spatial positions using Tangram. The spatial distribution of *NRG3*, *ERBB2*, and *ERBB4* is shown within the squared regions in the left panel. Results are representative of 7 sections (from three biologically independent samples) with similar outcomes. **o**, Heatmap depicting the cell type-specific expression of *NRG3*, *ERBB2*, and *ERBB4* in the developing human heart between 9 and 16 PCW, based on a public dataset. P-RBC, platelet-red blood cells; WBC, white blood cells; LEC, lymphatic endothelial cells; BEC, blood endothelial cells; Fibro, fibroblasts; SMC, smooth muscle cells; ncCM, non-chambered cardiomyocytes; aCM, atrial cardiomyocytes; vCM, ventricular cardiomyocytes. **p**, UMAP visualization of major cell types, and, *NRG3*, *ERBB2* and *ERBB4* derived from this dataset. **q,r**, Spatial visualization for the downstream genes of *SHOX2* (**q**) and *RORA* (**r**) in CS17 E1S3 heart. **s**, RNA-scope analysis showing the expression of *KIAA1324L*, *RORA* and the pacemaker cell marker *HCN4*. The images are representative of 3 independent experiments.



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

Extended Data Fig. 7 | Regionalization of the embryonic brain facilitates the characterization of timelines for functional specialization, and, cellular proportions of nerve-associated cells in the embryonic central nervous system. **a**, Heatmap showing the expression of marker genes across the substructures of human embryonic nervous system, and the color bar represents scaled gene expression level. **b**, The proportion of bins assigned to brain substructures at each Carnegie stage. **c**, Heatmap showing the Spearman correlation of transcriptional profiles among brain substructures at each Carnegie stage. **d**, Heatmap illustrating the activities of pathways across brain substructures, accompanied by enriched terms within each cluster. Colors in **b** correspond to the substructure identity in **d**. **e**, Barplots showing the cell proportion of radial glial cells in each brain region. **f**, Barplots showing the cell proportion of neural progenitor cells, glial progenitor cells, oligodendrocytes, astrocytes, microglia and oligodendrocyte progenitor cells in each brain region. For panel **e** and **f**, the bar height indicates cell proportion in the Fb, Mb, Hb, and SpC, respectively, and, bar color denotes the compositional ratio within the

finer brain substructures. **g**, Schematic representation of immunofluorescence localization patterns from Fig. 4b. RNA ISH image of *DCX* (green, immature neurons), *GADI* (red, inhibitory neurons) and *NEUROD6* (purple, excitatory neurons) in CS12-13 and CS19 (same in Fig. 4, left), along with their specific locations and orientations within the embryo. **h-m**, Cross validation of the temporal sequence of neurogenesis with data from Zeng *et al.* 2023 Cell Stem Cell²⁵. **h-j**, tSNE of all cells colored by stage (**h**), expression of excitatory neuron maker (**i**) and inhibitory neuron maker (**j**). The scale bar (**i,j**) represents normalized gene expression level. **k**, tSNE of all cells colored by cell types (left) and temporal dynamics of excitatory and inhibitory neuron ratios across different stages (right). **l**, Bubbleplot showing the proportion of cells with marker gene expression at different developmental stages. W, weeks post-conception. **m**, Spatial visualization of excitatory (*NEUROD2*, *NEUROD6*) and inhibitory (*GADI*, *GAD2*) neurons in Zeng *et al.*'s spatial transcriptomics (CS12 S5), and the scale bar represents normalized gene expression level.

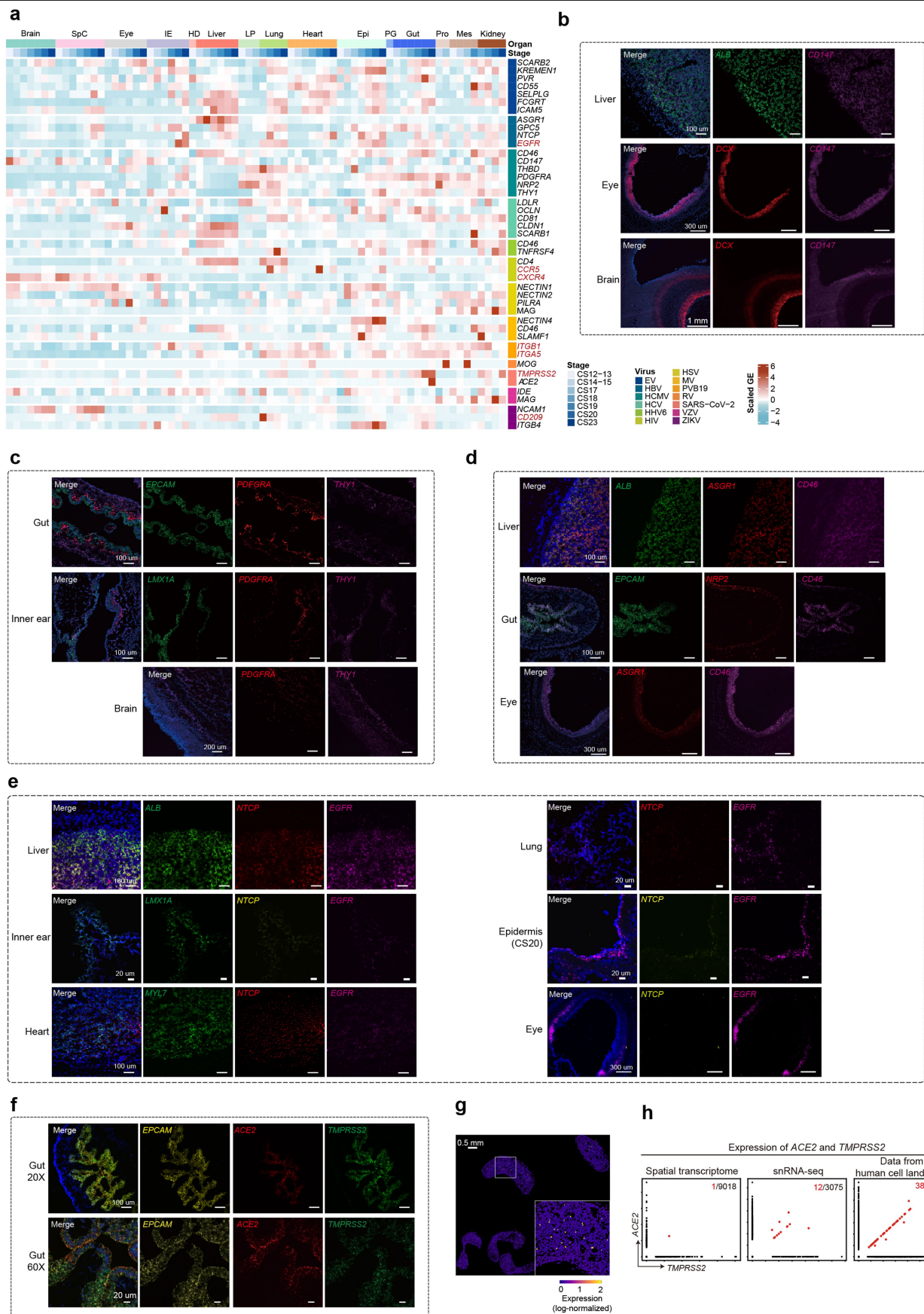


Extended Data Fig. 8 | See next page for caption.

Article

Extended Data Fig. 8 | Spatiotemporal characterization of the molecular and cellular patterns in embryonic nervous system, and, differences in HMGA2 rostrocaudal axis patterning within the Pall VZ between humans and mice. **a**, Upset plot showing intersections of regulons at central nervous system at each stage. **b**, Spatial visualization of the HMGA2 regulon activity (top) and expression (bottom) in the brain and spinal cord, and the color bar represents regulatory activity score (top) and scaled gene expression level (bottom). **c**, RNA ISH image of *HMGA2* and *DCX* in the developing brain. The images are representative of 3 independent experiments. **d, e**, Bubbleplot showing the expression of cell marker genes for different neural cells (**d**) and RGCs subpopulations (**e**). **f**, Spatial mapping of different RGCs subpopulations. **g**, The shift in proportions of RGCs subpopulations over different developmental stages. **h**, Expression of *HMGA2* in different neural cells. **i**, Heatmap shows the normalized expression of gene sets S1 (exhibiting an inverse expression gradient) and S2 (a concordant gradient) relative to *HMGA2* expression in human Pall VZ, and the color bar represents scaled gene expression level. **j**, Profile plot shows the averaged HMGA2 ChIP-seq signals of gene sets S1 and S2 (top). Barplot shows the averaged HMGA2 ChIP-seq signals of gene sets S1 and S2 (bottom). Data are presented as mean $\pm$ SEM ($n = 4$ biologically independent experiments). Initially, ANOVA was conducted to assess overall differences, followed by Tukey's multiple comparison test for multiple

comparisons to determine specific pairwise differences between the gene sets. **k**, HMGA2 binding patterns were grouped into four clusters using k-means clustering ($k = 4$). Profile plot shows the averaged HMGA2 ChIP-seq signals of 4 clusters (top). Heatmap shows the HMGA2 ChIP-seq signals ranging from 3,000 bp upstream of the TSS to 3,000 bp downstream of the TES (bottom). **l**, Differences in gene expression for each cluster between neocortical NPCs of *Hmga2*^{-/-} mice and those of control mice. Data are presented as box plots, with the boxes representing the median and upper and lower quartiles and the whiskers indicating the range. Statistical significance was assessed using Kruskal-Wallis test followed by Dunn's post hoc test with Bonferroni correction for multiple comparisons. **m**, The expression of the rostrocaudal axis expression of the gene set cluster3 in mouse (E15.5 E1S3) Pall VZ were grouped into 3 Sets using k-means clustering ($k = 3$). Data are presented as heatmap, and the color bar represents scaled gene expression level. **n**, GO enrichment analysis of the 3 Sets. The p -value was calculated by the one-tailed hypergeometric test to assess the statistical significance of the enrichment. **o**, Left panel, heatmap showing the regulon modules grouped by Hotspot at sections CS18 E1S1. Right panel, spatial visualization of modules in the brain. **p**, Left panel, heatmap showing the regulon modules grouped by Hotspot at sections CS19 E1S1. Right panel, spatial visualization of modules in the brain. **q**, Gene regulatory networks of module CS19_M1. Selected target genes and TFs in the DDG2P were shown.

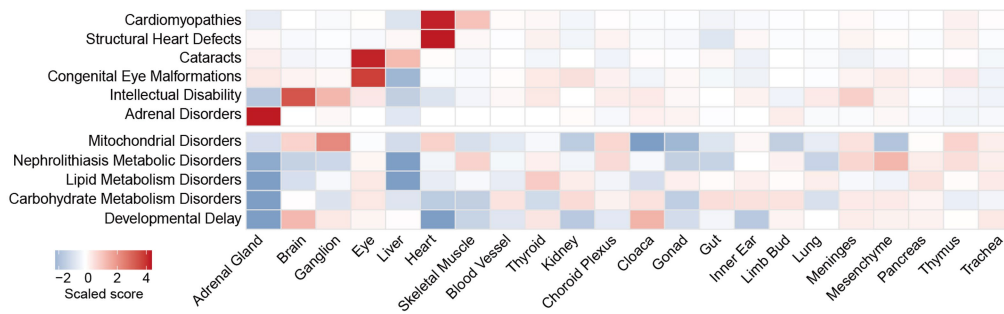


Extended Data Fig. 9 | See next page for caption.

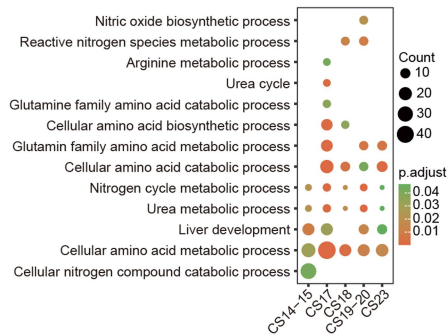
Extended Data Fig. 9 | Organ enrichment of genes related to virus entry. **a**, Heatmap showing the normalized gene expression level of receptors in 15 representative anatomic regions from CS12-13 to CS23. The entry receptor of the virus in black and the important coreceptor in red. SpC, spinal cord; IE, inner ear; HD, hepatic diverticulum; LP, lung primordium; Epi, epidermis; PG, primitive gut; Pro, pronephron; Mes, mesonephron. **b**, RNA ISH images of *CD147* in embryonic organs (CS23). **c**, RNA ISH images of *PDGFRA* and *THY1* in embryonic organs (CS23). **d**, RNA ISH images of *CD46*, *ASGR1* and *NRP2* in embryonic organs (CS23). **e**, RNA ISH images of *NTCP* and *EGFR* in embryonic

organs. The epidermis image was obtained from sample at Carnegie Stage 20 (CS20); all other images were from CS23 embryos. **f**, RNA ISH images of *ACE2* and *TMPRSS2* in embryonic gut (CS23). The images (**b-f**) are representative of 3 independent experiments. **g**, Spatial visualization of *TMPRSS2* expression based on single-cell segmentation in gut at CS23. Results are representative of 8 sections (from two biologically independent samples) with similar outcomes. **h**, Scatterplot showing the co-expression pattern (red spots) of *ACE2* and *TMPRSS2* within the same cell in the gut epithelia based on Stereo-seq, snRNA data and human cell landscape⁸⁹.

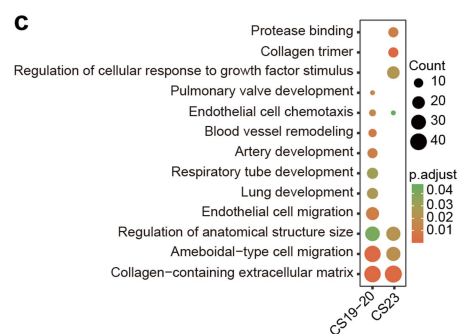
a



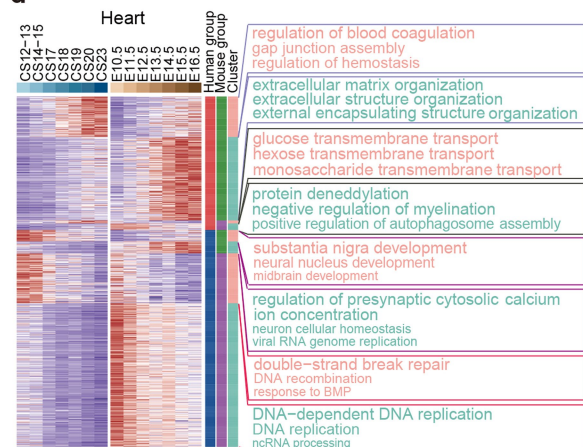
b



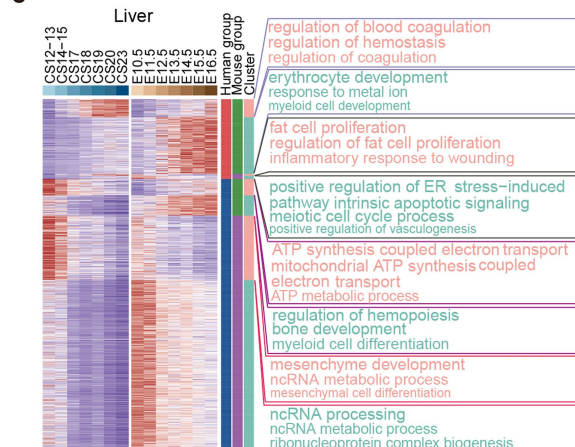
c



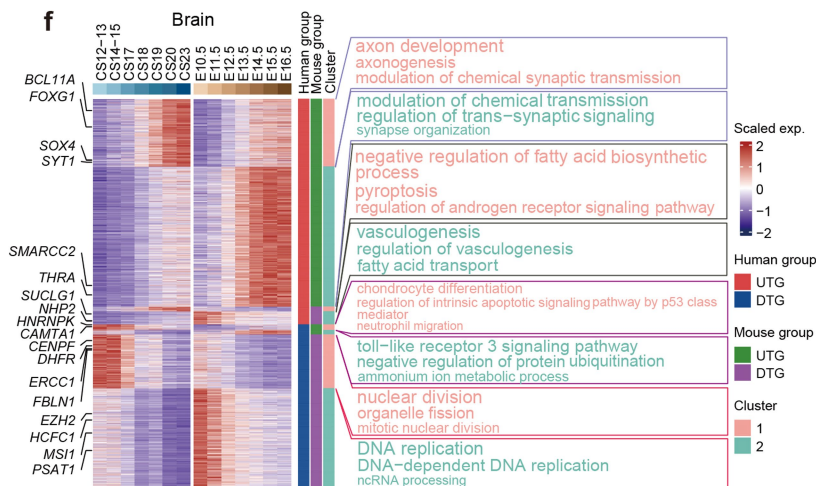
d



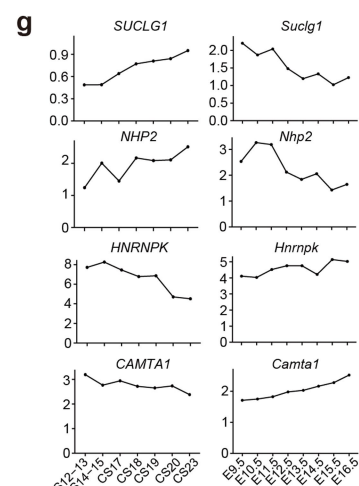
e



f



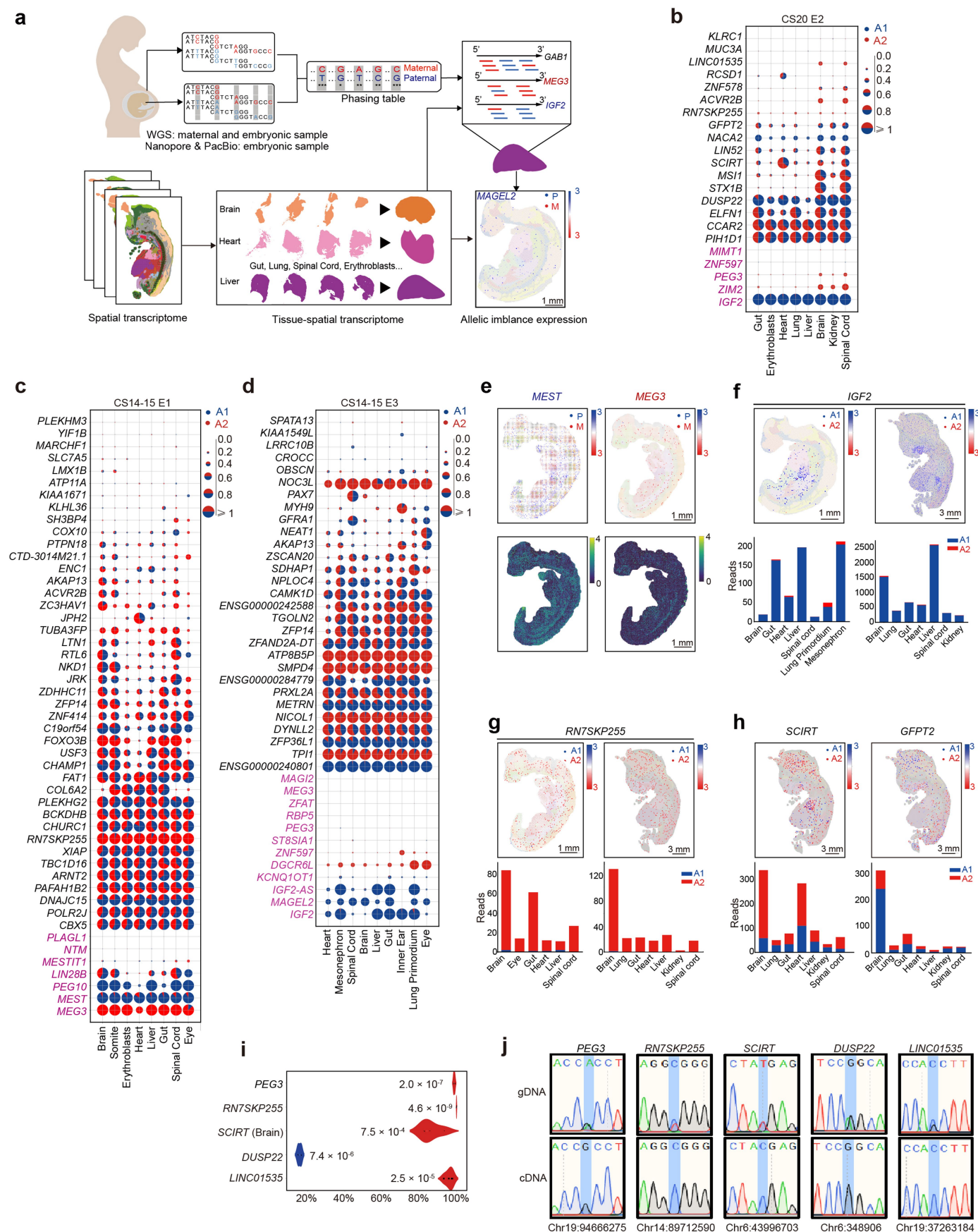
g



Extended Data Fig. 10 | See next page for caption.

Extended Data Fig. 10 | Enrichment of genes and pathways associated with development and cross-species comparisons of temporal profiles between human and mouse. **a**, The enrichment score of the causative genes implicated in developmental disorders. **b**, Bubbleplot showing the GO enrichment pathways of up-regulated genes in human liver compared with that of mouse from CS14-15 to CS23. **c**, Bubbleplot showing the GO enrichment pathways of up-regulated genes in human lung compared with that of mouse at CS19-20 and CS23. For panel **b,c**, the GO enrichment analysis was assessed by the one-sided

GO over-representation test with Benjamini-Hochberg multiple comparison correction. **d-f**, Comparative heatmap of temporal expression profiles for UTGs (up-regulated genes) and DTGs (down-regulated genes) across human and mouse in the heart (**d**), liver (**e**) and brain (**f**). Selected genes from DDG2P are marked on the left. Top enriched GO terms are displayed on the right, with color-coded bars and text corresponding to functional gene clusters. **g**, Dynamic expression profiles of four representative genes (*SUCLG1*, *NHP2*, *HNRNPK*, *CAMTA1*) in human brain and their orthologous genes in mouse brain.



Extended Data Fig. 11 | See next page for caption.

Extended Data Fig. 11 | The existence of imbalanced allelic expression pattern. **a**, The allelic expression analysis flowchart of spatial transcriptome from multiple chips. **b-d**, The imbalanced allelic expressions of 104 phased genes in tissues of CS20 E2 (**b**), CS14-15 E1 (**c**) and CS14-15 E3 (**d**) embryos. The known imprinted genes were highlighted in purple at the bottom of the chart. The pooled proportions of multiple sections are presented, with the size of each pie chart indicating the expressed bin rate within the corresponding tissue. **e**, The spatial allelic expression pattern (top) and common expression pattern (bottom) of *MEST* and *MEG3*. **f**, The spatial expression pattern of imprinted gene *IGF2* in CS14-15 E3S4 and CS20 E2S2. **g**, The spatial imbalanced allelic expression pattern of *RN7SKP255* in CS14-15 E1S4, accompanied by the

validation in CS20 E2S2. **h**, Brain-specific imbalanced allelic expression of *SCIRT* and *GFPT2* in CS20 E2S2. Representative sections (top) and pooled proportions for multiple sections; A1, blue; A2, red; color bar represents UMI count (**f-h**). **i**, The allelic expression of *PEG3*, *RN7SKP255*, *SCRIPT* (in brain), *DUSP22* and *LINCO1535* in the spatial transcriptome data from 4 sections of CS20 E2. X-axis represents the proportion of the A2 allele. Red indicates a higher proportion of the A2 allele, while blue indicates a higher proportion of the A1 allele. Each dot corresponds to a chip. Statistical analysis was based on the comparisons of the allele proportion differences (A1 vs. A2) for each gene. Two-tailed *p* values by paired Student's t-test. **j**, Sanger sequencing of genomic DNA and cDNA of allelic imbalanced expression gene from CS20 E2.

Corresponding author(s): Hefeng Huang 
 Last updated by author(s): Feb 9, 2026

Reporting Summary

Nature Portfolio wishes to improve the reproducibility of the work that we publish. This form provides structure for consistency and transparency in reporting. For further information on Nature Portfolio policies, see our [Editorial Policies](#) and the [Editorial Policy Checklist](#).

Statistics

For all statistical analyses, confirm that the following items are present in the figure legend, table legend, main text, or Methods section.

n/a Confirmed

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

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

Software and code

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

Data collection

Single-nucleus sequencing libraries were prepared with DNBelab c4 kits (MGI, China) and sequenced on DNBSEQ-Tx sequencer (MGI, China). The gene count matrices were generated from FASTQ files by an in-house pipeline (scRNA-seq_C4_analysis_pipeline, BGI-Research, China). The Stereo-seq libraries were sequenced on a DNBSEQ-Tx sequencer (MGI, China). Raw data was aligned to the reference genome (GRCh38.p12) by STAR V2.7.10a. The third-generation sequencing was performed with PromethION by BGI and Pacbio HiFi platform. Raw data was base-identified by guppy5 and aligned to GRCh38.p12 with minimap2 (2.24-r1122). The WGS was performed on the DNBseq T1&T5 sequencing platform. Raw data was aligned to human genome GRCh38.p12 through SOAPnuke (Default: "-n 0.1 -q 0.5 -l 12 -Q 2 -G 2 -M 2") and BWA (Built-in software of FPGA heterogeneous computing and multi-stream system (MegaBOLT, v 5.0)).

Data analysis

Unsupervised spatially constrained clustering was conducted using Scanpy (V1.8.2) and Squidpy (V1.2.0). Gene regulatory networks were inferred from Stereo-seq bin matrices according to the pySCENIC (V0.11.2) protocol. Spatially dependent regulon modules were identified by Hotspot (V1.1.1). The nuclei ssDNA images were used to calculate the cell boundaries with the cell segmentation algorithm spatelo (V1.0.2). Other packages involved in are Seurat (V4.3.0), SCTransform, Monocle3 (V1.3.1), clusterProfiler (V4.4.1, V4.6.0), ComplexHeatmap (V2.12.0, V2.20.0), IRGSEA (V3.2.5), simplifyEnrichment (V1.14.0), CellChat (V2.1.2), ggplot2 (V3.4.0), TOME, spatelo (V1.1.0), Stereopy (V1.5.1), Plotly (V5.13.1), Seaborn (V0.11.2, V0.13.2), SciPy (V1.8.1), pandas (V1.5.2), matplotlib (V3.5.2, V3.6.2), BBKNN (V1.6.0), Scanorama (V1.7.4), scvi-tools (V1.2.2), Tangram (V1.0.4), cell2location (V0.1.3), pySCENIC (V0.12.1), CellOracle (V0.18.0), samtools (V1.15.1), GATK (V4.1.8.1), Clair3 (V0.1-r12), annovar (V2020-06-07), DeepTools, ImageJ (V2.9.0). Databases involved in are blomaRt (V2.52.0), homologue (V1.4.68.19.3.27), DDG2P database (V3.0), GeneImprint.

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

Data

Policy information about availability of data

All manuscripts must include a [data availability statement](#). This statement should provide the following information, where applicable:

- Accession codes, unique identifiers, or web links for publicly available datasets
- A description of any restrictions on data availability
- For clinical datasets or third party data, please ensure that the statement adheres to our [policy](#)

The dataset resource for an interactive survey is accessible to the public (<https://db.genomics.cn/hesta>). The data that support the findings of this study have been deposited into OMIX and Genome Sequence Archive in National Genomics Data Center (GSA-Human). All processed data generated by Stereo-seq have been deposited to OMIX (<https://ngdc.cncb.ac.cn/omix>, accession code: OMIX003254). All processed data generated by snRNA-seq have been deposited to OMIX (<https://ngdc.cncb.ac.cn/omix>, accession code: OMIX003254). All raw data generated by WGS have been deposited to GSA-Human (<https://ngdc.cncb.ac.cn/gsa-human>, accession code: HRA004300). All raw data generated by Nanopore and PacBio have been deposited to GSA-Human (<https://ngdc.cncb.ac.cn/gsa-human>, accession code: HRA004299). The published datasets analysed for this study were retrieved from the GEO repository (GSE155121, GSE157329), UCSC Cell Browser (<https://cells.ucsc.edu/?ds=hoc>), <https://www.gutcellatlas.org/> and <https://db.cngb.org/stomics/mosta/>.

Research involving human participants, their data, or biological material

Policy information about studies with [human participants or human data](#). See also policy information about [sex, gender \(identity/presentation\), and sexual orientation](#) and [race, ethnicity and racism](#).

Reporting on sex and gender	The embryo samples were all derived from women who have voluntarily terminated their pregnancies, and the genders of the embryo samples include both male and female.
Reporting on race, ethnicity, or other socially relevant groupings	The embryo samples were all derived from Chinese Han population.
Population characteristics	The embryo samples were all derived from Chinese Han population, aged from 21 to 35. No prior history of systemic diseases, and all subjects were women who underwent elective termination of pregnancy.
Recruitment	Human embryos and decidua tissues were collected from women who voluntarily terminated the pregnancy at the Affiliated Obstetrics and Gynecology Hospital of Fudan University. No individuals under the age of 18 were included. No individuals had serious diseases or medical history. The collection and use of human embryos were compliant with the current International Society for Stem Cell Research (ISSCR) Guidelines, guidelines for Stem Cell Research and Clinical Translation (2016, 2021), and the Helsinki declaration (2013). None of the samples used in this study were used in any other experiments. All procedures followed the 'Interim Measures for the Administration of Human Genetic Resources' administered by the Chinese Ministry of Health. Before the abortion procedure, each participant provided written informed consent for collection for research purposes. Embryo collection was performed only after voluntary informed consent was obtained from each patient undergoing legal pregnancy termination. After the termination of pregnancy, participants were given an additional opportunity to verbally reconfirm their willingness to donate the tissue samples. Discarded embryos were collected and transported immediately to the laboratory for processing. Embryo collection does not involve any invasive procedure or gene editing. There is no risk to the patients. The patients were not compensated for participating in this study. The sample is restricted to women seeking elective abortions at a single hospital, which may not represent broader populations (e.g., varying socioeconomic backgrounds, geographic regions, or reasons for pregnancy termination). This limits generalizability to other contexts. This bias could compromise external validity (by limiting applicability beyond the study population). For example, associations observed in this sample might not hold for women with different demographic profiles or in regions with distinct healthcare practices.
Ethics oversight	Ethical approval of human embryo research was provided by the Ethics Committee of Obstetrics and Gynecology Hospital, Fudan University (2021-121), and the Institutional Review Board of BGI-Shenzhen (BGI-IRB 22058). Human embryos and decidua tissues were collected from women who voluntarily terminated the pregnancy at the Affiliated Obstetrics and Gynecology Hospital of Fudan University. The collection and use of human embryos were compliant with the current International Society for Stem Cell Research (ISSCR) Guidelines, guidelines for Stem Cell Research and Clinical Translation (2016, 2021), and the Helsinki declaration (2013). None of the samples used in this study were used in any other experiments. All procedures followed the 'Interim Measures for the Administration of Human Genetic Resources' administered by the Chinese Ministry of Health.

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

Field-specific reporting

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

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

For a reference copy of the document with all sections, see nature.com/documents/nr-reporting-summary-flat.pdf

Life sciences study design

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

Sample size	Sample sizes were not predetermined, but our sample sizes are larger than those reported in previous publications (Cell 2024, PMID: 38657603; Nat Cell Biol. 2025, PMID: 39794460) A total of 77 sagittal sections from 13 euploid human embryos (7 males and 6 females) dating from CS12 to CS23 were sequenced to obtain a spatial-temporal transcriptome atlas of organogenesis. One to three embryos were collected at each stage during CS12-23.
Data exclusions	A total of 90 sagittal sections from 16 euploid human embryos (9 males, 7 females) were sequenced to obtain spatial-temporal transcriptome atlas of organogenesis. We went through every section and removed 13 low-quality sections from the total 90 sections, leaving 77 sections from 13 euploid human embryos (7 males and 6 females) for the final analysis. We also performed data QC by removing the gene clusters of cavities which represent the tissue-hollow/blank areas of the sections. For snRNA-seq, cells with less than 1k genes, less than 2k counts, more than 10% mitochondrial genes, or more than 90 quantiles of total counts per library were filtered out from raw counts by scanpy.pp.filter_cells. Genes that were captured in less than 3 cells were filtered out as well by scanpy.pp.filter_genes. A total of 607,093 cells from developing human embryos at stages of CS19, CS20, and CS23 were proceeded to downstream analysis.
Replication	For the spatial atlas, all attempts at replication were successful. Specifically, we performed 3 biological replicates at CS14-15 and 2 biological replicates at CS12-13, CS19, CS20, and CS23. Only for CS17 and CS18 did we use a single embryo due to the rarity of early-stage human embryos and considerations of data quality. All integrative analyses and key findings were verified using at least two biological replicates or multiple technical replicates (different sections from the same embryos). We also collected extra slides of cryosections from the embryos (CS19, CS20, CS23) near the locations where the Stereo-seq data have been obtained to generate our snRNA-seq data. We used both the public scRNA-seq data and our snRNA-seq data to validate the findings that from the spatial transcription data. For ISH assay and animal studies, the number of replicates for each experiment is mentioned in the figure legend. All attempts at replication were successful.
Randomization	Sample allocation into experimental groups was not randomized. Instead, samples (human embryos) were objectively categorized into groups based on their developmental stage, as determined by Carnegie staging (CS)—a standardized, morphology-based classification system for embryonic development. Groups corresponded to distinct developmental periods (e.g., CS12, CS14, CS19), reflecting natural progression rather than experimental intervention.
Blinding	Not applicable since no specific grouping.

Reporting for specific materials, systems and methods

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

Materials & experimental systems		Methods	
n/a	Involved in the study	n/a	Involved in the study
☐	☒ Antibodies	☒	☐ ChIP-seq
☒	☐ Eukaryotic cell lines	☒	☐ Flow cytometry
☒	☐ Palaeontology and archaeology	☒	☐ MRI-based neuroimaging
☐	☒ Animals and other organisms		
☒	☐ Clinical data		
☒	☐ Dual use research of concern		
☒	☐ Plants		

Antibodies

Antibodies used	Primary antibody anti-Islet-1 (Isl1) (Genetex, GTX128201, 1:500 dilution); anti-GFP-Tag (7G9) mouse mAb (Abmart, M200045, 1:1000 dilution). Secondary antibodies Anti-rabbit Alexa 594 (Abcam, ab150084, 1:500 dilution); anti-mouse Alexa 488 (Abcam, ab150113, 1:500 dilution).
Validation	All commercial antibodies have been tested for specificity by their respective suppliers. Antibody validations and validation criteria are available on the following websites: Anti-Islet-1 (Isl1) — https://www.genetex.cn/Product/Detail/Is11-antibody/GTX128201 Anti-GFP-Tag(7G9) Mouse mAb(Abmart, M200045)— https://www.ab-mart.com.cn/page.aspx?node=%2059%20&id=%20971 Anti-rabbit Alexa 594— https://www.abcam.cn/products/secondary-antibodies/goat-rabbit-igg-hl-alexa-fluor-594-preadsorbed-ab150084.html

Animals and other research organisms

Policy information about studies involving animals: ARRIVE guidelines recommended for reporting animal research, and Sex and Gender in Research

Laboratory animals	Tg(myl7:H2B-EGFP) transgenic zebrafish fish was used to incross to obtain 1-5 dpf embryos with eGFP labeled cardiomyocytes.
Wild animals	This study didn't involve wild animals.
Reporting on sex	The genders of the zebrafish embryo samples were unknown.
Field-collected samples	The study didn't involve any samples from the field.
Ethics oversight	All procedures on live animals were approved by the the Ethics Committee of the Shanghai Medical College, Fudan University.

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

Plants

Seed stocks	not applicable
Novel plant genotypes	not applicable
Authentication	not applicable