



Single-cell RNA-seq reveals dynamic transcriptome profiling in human early neural differentiation

Shang, Zhouchun; Chen, Dongsheng; Wang, Quanlei; Wang, Shengpeng; Deng, Qiuting; Wu, Liang; Liu, Chuanyu; Ding, Xiangning; Wang, Shiyu; Zhong, Jixing; Zhang, Doudou; Cai, Xiaodong; Zhu, Shida; Yang, Huanming; Liu, Longqi; Fink, J Lynn; Chen, Fang; Liu, Xiaoqing; Gao, Zhengliang; Xu, Xun

Published in:
GigaScience

DOI:
[10.1093/gigascience/giy117](https://doi.org/10.1093/gigascience/giy117)

Publication date:
2018

Document version
Publisher's PDF, also known as Version of record

Document license:
[CC BY](https://creativecommons.org/licenses/by/4.0/)


Citation for published version (APA):
Shang, Z., Chen, D., Wang, Q., Wang, S., Deng, Q., Wu, L., ... Xu, X. (2018). Single-cell RNA-seq reveals dynamic transcriptome profiling in human early neural differentiation. *GigaScience*, 7(11), 1-19. [giy117]. <https://doi.org/10.1093/gigascience/giy117>

RESEARCH

Single-cell RNA-seq reveals dynamic transcriptome profiling in human early neural differentiation

Zhouchun Shang^{1,2,3,4,†}, Dongsheng Chen^{2,3,†}, Quanlei Wang^{2,3,4,5,†}, Shengpeng Wang^{2,3}, Qiuting Deng^{2,3}, Liang Wu^{2,3,5,6}, Chuanyu Liu^{2,3,5}, Xiangning Ding^{2,3}, Shiyong Wang^{2,3}, Jixing Zhong^{2,3}, Doudou Zhang⁷, Xiaodong Cai⁷, Shida Zhu^{2,3,4}, Huanming Yang^{2,8}, Longqi Liu^{2,3}, J. Lynn Fink^{2,9,10}, Fang Chen^{2,3,11}, Xiaoqing Liu¹, Zhengliang Gao^{1,*} and Xun Xu^{1,2,3,*}

¹Shanghai Tenth People's Hospital, Tongji University School of Medicine, Shanghai 200072, China, ²BGI-Shenzhen, Shenzhen 518083, China, ³China National GeneBank, BGI-Shenzhen, Shenzhen 518120, China, ⁴Shenzhen Engineering Laboratory for Innovative Molecular Diagnostics, BGI-Shenzhen, Shenzhen 518083, China, ⁵BGI Education Center, University of Chinese Academy of Sciences, Shenzhen 518083, China, ⁶Shenzhen Key Laboratory of Neurogenomics, BGI-Shenzhen, Shenzhen 518083, China, ⁷Department of Neurosurgery, Shenzhen Second People's Hospital, Shenzhen University 1st Affiliated Hospital, Shenzhen 518035, China, ⁸James D. Watson Institute of Genome Sciences, Hangzhou 310058, China, ⁹BGI Australia, L6, CBCRC, 300 Herston Rd, Herston, QLD 4006, Australia, ¹⁰The University of Queensland, Diamantina Institute (UQDI), Brisbane, QLD 4102, Australia and ¹¹Laboratory of Genomics and Molecular Biomedicine, Department of Biology, University of Copenhagen, 2100 Copenhagen, Denmark

*Corresponding authors. Zhengliang Gao, Shanghai Tenth People's Hospital, Tongji University School of Medicine, Shanghai 200072, China; E-mail: zhengliang.gao@tongji.edu.cn; Xun Xu, China National GeneBank, BGI-Shenzhen, Shenzhen 518120, China; E-mail: xuxun@genomics.cn  <http://orcid.org/0000-0002-5338-5173>

[†]These authors contributed equally to this work.

Abstract

Background: Investigating cell fate decision and subpopulation specification in the context of the neural lineage is fundamental to understanding neurogenesis and neurodegenerative diseases. The differentiation process of neural-tube-like rosettes *in vitro* is representative of neural tube structures, which are composed of radially organized, columnar epithelial cells and give rise to functional neural cells. However, the underlying regulatory network of cell fate commitment during early neural differentiation remains elusive. **Results:** In this study, we investigated the genome-wide transcriptome profile of single cells from six consecutive reprogramming and neural differentiation time points and identified cellular subpopulations present at each differentiation stage. Based on the inferred reconstructed trajectory and the characteristics of subpopulations contributing the most toward commitment to the central nervous system lineage at each stage during differentiation, we identified putative novel transcription factors in regulating neural differentiation. In

Received: 17 March 2018; Revised: 27 July 2018; Accepted: 6 September 2018

© The Author(s) 2018. Published by Oxford University Press. This is an Open Access article distributed under the terms of the Creative Commons Attribution License (<http://creativecommons.org/licenses/by/4.0/>), which permits unrestricted reuse, distribution, and reproduction in any medium, provided the original work is properly cited.

addition, we dissected the dynamics of chromatin accessibility at the neural differentiation stages and revealed active cis-regulatory elements for transcription factors known to have a key role in neural differentiation as well as for those that we suggest are also involved. Further, communication network analysis demonstrated that cellular interactions most frequently occurred in the embryoid body stage and that each cell subpopulation possessed a distinctive spectrum of ligands and receptors associated with neural differentiation that could reflect the identity of each subpopulation.

Conclusions: Our study provides a comprehensive and integrative study of the transcriptomics and epigenetics of human early neural differentiation, which paves the way for a deeper understanding of the regulatory mechanisms driving the differentiation of the neural lineage.

Keywords: single-cell RNA-seq; ATAC-seq; neural differentiation; neural rosettes; neural tube; transcription factor; iPSC

Background

The nervous system contains complex molecular circuitry in developmental processes. In humans, there is a paucity of data describing early neural development and the corresponding cellular heterogeneity at various stages. To our knowledge, neural tube formation and closure are crucial for embryonic central nervous system (CNS) development and the process of neurulation. Previous studies have reported that neural tube closure is strongly controlled by both genetic and epigenetic factors and is sensitive to environmental influences [1–3]. Perturbations in this delicately balanced and orchestrated process can result in neural tube defects (NTDs), which give rise to birth defects such as spina bifida, anencephaly, and encephaloceles. However, the formation and closure of the neural tube *in vivo* during weeks 3 and 4 of human gestation are transient events and therefore difficult to capture. Moreover, the limited accessibility of human abortive fetuses at such an early stage precludes a thorough investigation of human early neural development.

Human pluripotent stem cells (hPSCs), including embryonic stem cells (ESCs) and induced pluripotent stem cells (iPSCs), can be differentiated into all cell types, including neural cells, offering a promising *in vitro* model for tracing early cell lineages and studying the cell fate specification of human neural differentiation [4, 5]. Previous studies have indicated that inhibition of bone morphogenetic protein (BMP) signaling or activation of fibroblast growth factor (FGF) signaling is needed for induction of the neuroectoderm from ESCs [6, 7]. A striking feature of differentiating stem cells *in vitro* is that they form neural tube-like rosettes that are composed of radially organized columnar epithelial cells that resemble the process of neurulation. The progenitor cells in rosettes gradually give rise to functional cells (e.g., more restricted progenitors and neuronal precursors, mimicking the process of neurulation and neural tube growth), which represent neural tube structures [8]. These cellular processes suggest that distinct cell fate decisions and lineage commitments occur during rosette formation. However, the corresponding underlying mechanisms of the regulation of cell fate commitment during early neural differentiation remain largely unknown.

The advance of single-cell trans-omics technology has offered incisive tools for revealing heterogeneous cellular contexts and developmental processes [9–11]. Single-cell RNA sequencing (scRNA-seq) has been applied to the study of cellular heterogeneity as well as to the identification of novel subtypes or intermediate cell groups in multiple contexts [12–15] and may help delineate unexpected features of neural developmental biology and facilitate the study of cellular states and neurogenesis processes. In the present study, we used scRNA-seq and assay for transposase-accessible chromatin using sequencing (ATAC-seq) to investigate human early neural differentiation. Our analysis

reveals the landscape of the transcriptome and cis-regulatory elements during this process and creates an unbiased classification of cell subpopulations during differentiation, providing a comprehensive description of transcriptomic and epigenetic patterns in cell fate decisions. The differentiation system of human induced pluripotent stem cells (hiPSCs) provides access to the very early stage of neural development and may serve as a source of specialized cells for regenerative medicine as well as support for further investigations of neural tube defects.

Data Description

Here, we applied a well-adopted neural induction protocol and generated neural progenitor cells (NPCs) by forming neural rosettes *in vitro* [8, 16]. We analyzed several differentiation stages of cells, including hiPSCs, embryoid body (EB), early rosettes (hereafter termed Ros-E, post-3 days of rosette formation), late rosettes (hereafter termed Ros-L, post-5 days of rosette formation), NPCs, and the original somatic fibroblasts (Fib). scRNA-seq was performed at discrete time points (e.g., Fib, iPSCs, EB, Ros-E, Ros-L, and NPCs), and we captured 96, 80, 81, 82, 93, and 95 single cells, respectively, for each stage with the purpose of studying differentiation transition events. We also captured bulk transcriptome profiles of the corresponding neural differentiation stages derived from iPSCs and ESCs for validation. In addition, bulk ATAC-seq with two biological replicates was applied to the cell stages iPSCs, EB, Ros-E, Ros-L, and NPCs to measure the regulome dynamics during neural differentiation (Fig. 1a). The quality of sequencing data was evaluated and filtered by a quality control (QC) pipeline developed in-house (see Methods section for details).

Analyses

Differential transcriptome and regulome dynamics throughout human early neural differentiation

Since the development of human ESCs and iPSCs, the ability to investigate human neurogenesis and neurological diseases via an *in vitro* differentiation model has vastly improved [4, 17]. Subsequently, artificial neural cells have been successfully generated using a variety of protocols by several laboratories [18–23]. Here, we followed a well-adopted neural induction protocol and generated NPCs by forming neural rosettes via inhibition of transforming growth factors β (TGF β), AMP-activated protein kinase, and BMP signaling pathways and activation of the FGF signaling pathway [8, 16]. We analyzed different differentiation stages of the cells including iPSCs, EB, Ros-E, Ros-L, and NPCs as well as the original somatic Fibs. The iPSC aggregates were induced to neuroepithelial (NE) cells and followed by neural tube-like rosettes formation (Fig. 1b). First, pluripotency-

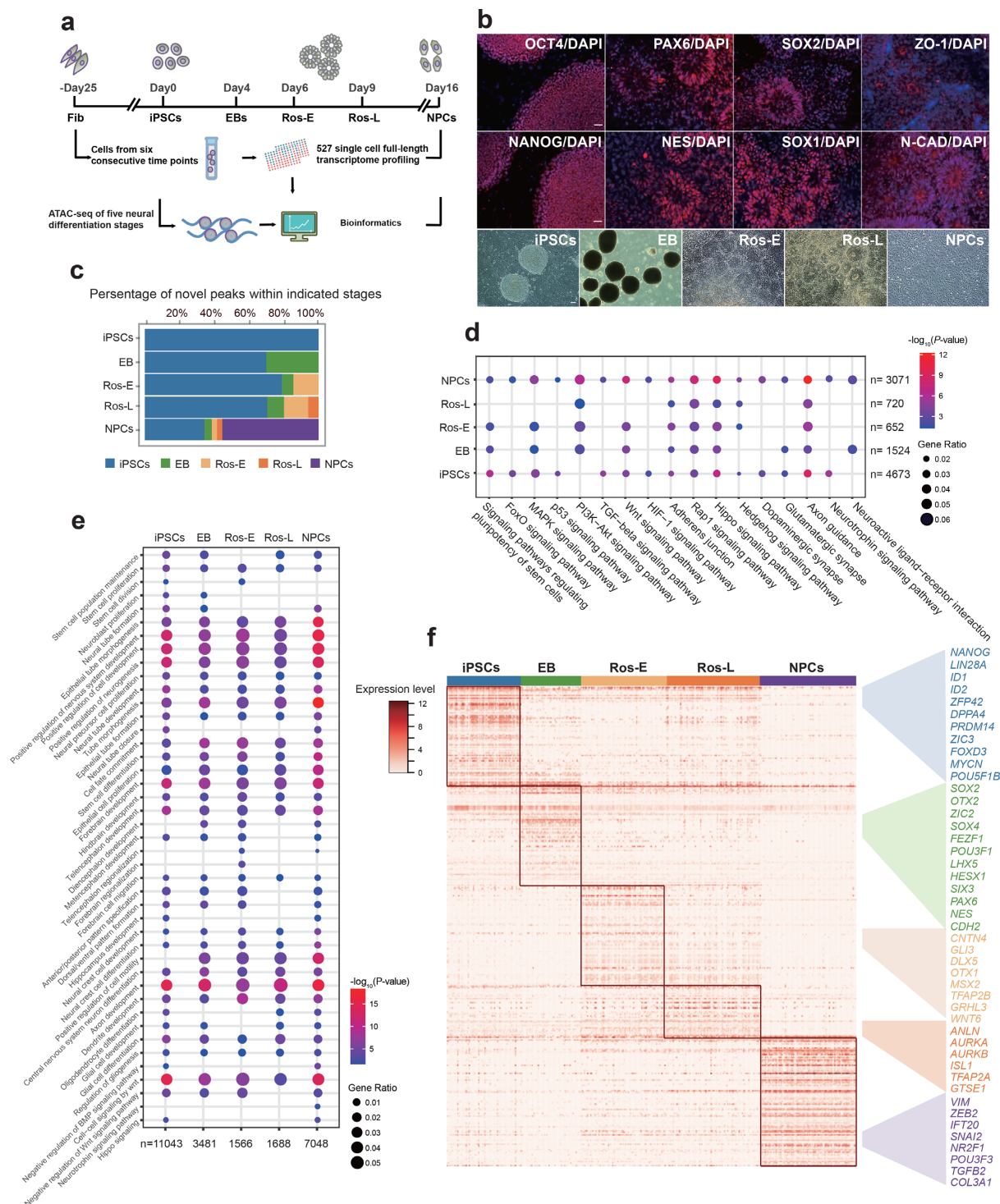


Figure 1: Transcriptome and regulome dynamics during human early neural differentiation. **(a)** Schematic illustration of experimental strategy. **(b)** Bright field and immunostaining of well-defined markers for iPSCs, including OCT4 and NANOG, and for neural rosettes (Ros-L stage), including PAX6, NES (NESTIN), SOX2, SOX1, ZO-1, and N-CAD (N-CADHERIN, also known as CDH2). Scale bar represents 50 μ m. **(c)** Dynamic distribution of novel peaks (active cis-regulatory elements) within indicated cell stages. **(d)** Kyoto Encyclopedia of Genes and Genomes (KEGG) enrichment analysis of novel peaks within each cell stage as indicated respectively. **(e)** Gene Ontology (GO) term annotation of novel peaks within each cell stage as indicated respectively. **(f)** Stage-specific genes highlighted with color specific to the respective neural differentiation cell stage (adjusted P value ≤ 0.01).

associated transcription factors (TFs) (e.g., OCT4, NANOG) were significantly expressed in hiPSCs, suggesting that these cells did exhibit a stem cell phenotype. The subsequent formation of neural rosettes was confirmed by morphology, apical localization

of ZO-1 (also known as TJP1), a tight junction protein, and co-localization of the neuroepithelial marker N-CADHERIN (N-CAD, also known as CDH2) at the junctions. Additional neural mark-

ers such as PAX6, NESTIN, SOX2, and SOX1 were also found to be highly enriched in the Ros-L stage (Fig. 1b).

Cell stages are usually determined by a complement of TFs or master regulators, which regulate hundreds of genes associated with various cellular functions. To study the genomic features associated with open chromatin regions, we classified ATAC peaks based on the location of the peak center. More than 16,000 peaks were identified for each cell stage (Additional file 1: Supplementary Fig. S1a), with the majority located in introns and enhancers/promoters, genomic regions that are known to harbor a variety of cis-regulatory elements and are subjected to regulation by TFs (Additional file 1: Supplementary Fig. S1b). Furthermore, we observed that ATAC peaks were significantly enriched at regions near transcription start site (TSS) (Additional file 1: Supplementary Fig. S1c). These observations were reproducible across two replicates with a very high Pearson correlation coefficient (≥ 0.954) (Additional file 1: Supplementary Fig. S1d, S1e).

It is widely reported that chromatin structures undergo widespread reprogramming during cell status transition, with some genomic regions becoming compacted or opened, leading to the switching on or off of a repertoire of genes responsible for cell fate decision [24–29]. We studied the dynamic chromatin landscape by tracing the temporal origins of ATAC peaks at each stage with peaks nonoverlapping with existing ones that were annotated as novel peaks. We assumed that those peaks, conserved among differentiation stages, are associated with house-keeping genes, while stage-dynamic peaks are likely to represent cis-regulatory elements important for cell status transition. As expected, we observed the introduction of roughly 10%–50% of novel peaks in each stage, accompanied by the disappearance of several pre-existing ATAC peaks. Notably, more novel peaks appeared at the NPCs stage than at any other stage (Fig. 1c). Gene Ontology (GO) term analysis of genes residing in novel peaks across the differentiation stages showed enrichment of “axon development,” “positive regulation of nervous system development,” “epithelial tube morphogenesis,” “positive regulation of neurogenesis,” “cell-cell signaling by Wnt,” “forebrain development,” “hindbrain development,” “telencephalon development,” “neural precursor cell proliferation,” and “cell fate commitment.” “Neurotrophin signaling pathway” was also found to be enriched but was specifically associated with NPCs. Kyoto Encyclopedia of Genes and Genomes (KEGG) enrichment analysis showed that the “FoxO signalling pathway,” a pathway that is known to play an important role in NPC proliferation, and “neuroactive ligand–receptor interaction” were enriched in the NPCs stage (Fig. 1d, 1e), suggesting that specific cis-regulatory elements regulating neural differentiation are being staged (poised) for stem cell fate specification and conversion.

To reveal the detail of chromatin accessibility dynamics during neural differentiation, we also analyzed the gained or lost peaks at each stage compared with the previously neighboring one. We observed that the number of gained peaks was with the largest increase at the NPCs stage, while the number of lost peaks was relatively high at Ros-E stage (Additional file 2: Supplementary Fig. S2a). Next, we studied the genomic distribution of these dynamic peaks and found that both the gained and lost peaks were located mostly in distal intergenic regions and promoter regions (Additional file 2: Supplementary Fig. S2b). This observation indicates that distal and promoter regions are more dynamic compared to other genomic regions during the neural differentiation process.

To gain insight into the potential function of closing (lost) peaks dynamics, we carried out GO enrichment analysis on the

genes associated with lost peaks at each stage. The GO terms analysis showed that “mesoderm morphogenesis,” “endoderm development,” “gastrulation,” and “nodal signaling pathway” were solely enriched at the EB stage, indicating that upstream, as well as other lineage development, was relatively repressed by closing related cis-regulatory regions. Other cell fate conversion terms such as “neural crest cell differentiation,” “osteoclast differentiation,” and “regulation of cartilage development” were enriched at the Ros-E stage, together with the annotation results of novel peaks, indicating that the chromatin accessibility prepared for the neural lineage conversion by opening/closing up specific cis-regulatory regions, which facilitated the neural transition cascades (Fig. 1d, 1e and Additional file 2: Supplementary Fig. S2d, S2e).

Furthermore, we identified stage-specific peaks at iPSCs, EB, Ros-E, Ros-L, and NPCs using motif enrichment analysis (see Methods section). Further GO term and KEGG enrichment analysis showed very similar results with annotation analysis of novel peaks in corresponding cell stages (Additional file 3: Supplementary Fig. S3). These findings strongly suggest that the novel, gained and lost, as well as stage-specific peaks, represent cell status and cell fate transitions that progress neural differentiation and that the landscape of cis-regulatory element accessibility throughout the differentiation process is highly dynamic.

To more thoroughly investigate the molecular mechanisms governing neural differentiation, we profiled the transcriptomes of 527 single cells. scRNA-seq libraries were generated using the Smart-Seq2 method [30], followed by sequencing approximately 6 million reads per cell. For subsequent analysis, we focused on 445 cells that passed the QC (Methods section, Additional file 4: Supplementary Fig. S4a, S4b) and External RNA Controls Consortium (ERCC) correlation filter (Methods section, Additional file 4: Supplementary Fig. S4c). A median number of 7,003 to 8,560 expressed genes were detected per cell (Additional file 4: Supplementary Fig. S4d), including TFs that were relatively highly expressed at the EB and NPCs stages, while, intriguingly, pseudogenes were relatively highly expressed at the Ros-E and NPCs stages (Additional file 4: Supplementary Fig. S4e). We also identified a variety of genes (3,524, 3,855, 2,023, 1,804, and 6,211) specifically expressed at the iPSCs, EB, Ros-E, Ros-L, and NPCs stages, respectively (Additional file 4: Supplementary Fig. S4f). Many of these stage-specific genes include some well-known pluripotent genes (NANOG, ID1, ID2, ZFP42, LIN28A, DPPA4); early neural markers (SOX2, OTX2, OTX1, PAX6); and genes that both regulate neural development and are critical to proliferative NPCs (SOX4, SIX3, CDH2, ZIC2) (Fig. 1f and Additional file 4: Supplementary Fig. S4h).

Because the neural rosette recapitulates neural tube development *in vitro*, we paid particular attention to the Ros-E and Ros-L stages. Unsurprisingly, a large proportion of upregulated genes in the Ros-E stage were associated with nervous system development, including TFAP2A, CNTN4, GLI3, DLX5, and OTX1 (Fig. 1f). Of particular interest is the gene GRHL3. Expression of this gene is associated with neural tube closure in mice [31, 32], and we observed this gene to be highly expressed at Ros-E in human cells, suggesting that its role in neural tube closure may be conserved across mammals or possibly chordates. TFAP2A (transcription factor AP-2 alpha) and TFAP2B (transcription factor AP-2 beta) have been proposed as master regulators of the neural crest cell, and loss of function of transcription factor AP-2 in mice is strongly associated with a cranial neural tube defect phenotype [33]. In our system, TFAP2B and TFAP2A were relatively highly expressed at both the Ros-E and Ros-L stages, suggesting transcription factor AP-2 may coordinate the special-

ized distal cis-regulatory elements for downstream regulations in human. We also observed expression of ANLN (anillin actin binding protein) at the Ros-L stage, suggesting that neuronal migration and neurite growth might occur by the linking of RhoG to the actin cytoskeleton in neural rosettes [34]. Similarly, our data showed that AURKA (aurora kinase A) and AURKB (aurora kinase B) were both expressed at the Ros-L stage, echoing previous findings that the aPKC-Aurora A-NDEL1 pathway plays an essential role in neurite elongation through modulating microtubule dynamics [35]. Finally, the neuron fate commitment protein, TGFB2, the nervous system development regulator, ZEB2, and the neural precursor cell proliferation-associated protein, IFT20, were enriched at the NPCs stage (Fig. 1f).

An unexpected finding was that some of the most important neural TFs exhibited heterogeneous expression within the same cell stage (e.g., ZIC2, OTX2, HESX1, DLX3, LHX5) (Fig. 1f and Additional file 4: Supplementary Fig. S4h). This inspired us to dissect the subpopulations of cells within each cell stage to better understand the significance of this result.

Heterogeneous cellular subpopulations were identified at each differentiation stage

To evaluate the overall distribution of cells at each of the six stages during reprogramming and neural differentiation, we first performed an unsupervised analysis using all expressed genes (QC, see Methods section) as input to t-distributed stochastic neighbor embedding (t-SNE) for visualization. This analysis showed distinct clusters for each differentiation stage, supporting our observation of heterogeneous gene expression during these stages (Fig. 2a). Because previous studies have shown that TFs and cis-regulatory elements are highly informative in reflecting cell identity [36], we used a machine classifier to determine the subsets of TFs that best clustered cells into putative cell populations. We were then able to identify distinct subpopulations at each cell stage (Fib1, Fib2, EB1, EB2, EB3, Ros-E1, Ros-E2, Ros-L1, Ros-L2, Ros-L3, NPC1, NPC2, and NPC3) (Methods section, Fig. 2, Additional file 5–8: Supplementary Figs S5–S8). As we found no remarkable differential expression of pluripotency-associated genes (e.g., NANOG, ID1, ID2, LIN28A, SOX2, DPPA4, ZFP42, TRIM28) at the iPSC stage (Additional file 4: Supplementary Fig. S4g), we did not include iPSCs in the following analyses.

Fib stage

Fibs are a very well-adopted original somatic cell resource for iPSCs reprogramming; many direct conversions from fibroblast to functional neurons have been reported [37, 38]. Here, we dissected two subpopulations of human dermal Fibs (Fib1 and Fib2) with distinct molecular features, showing significantly higher expression of several important pluripotency- and neural-associated TFs such as SOX2, LIN28A, SOX11, ZIC2, FEZF1, and SIX3 in Fib2 (Additional file 5: Supplementary Fig. S5a, S5b). GO terms identified by upregulated genes between the two subsets showed “chromosome segregation,” “positive regulation of nervous system development,” “stem cell population maintenance,” “positive regulation of cell cycle,” “neural precursor cell proliferation,” and “chromatin remodeling” as solely enriched in the Fib2 subpopulation (Additional file 5: Supplementary Fig. S5c). KEGG enrichment analysis showed the term “cell cycle” was specifically associated with the Fib2 subset (Additional file 5: Supplementary Fig. S5d). Furthermore, we observed that Fibs were distributed into two distinct groups called Fib-Group1 and Fib-Group2 based on their location in Fig. 2a. Of note, the major-

ity of cells in Fib-Group1 and Fib-Group2 were composed of Fib1 and Fib2, respectively. Moreover, cells from the Fib2 subset clustered together with EB cells (Additional file 5: Supplementary Fig. S5e). Together with the molecular features of Fib2 subset (Additional file 5: Supplementary Fig. S5b), we proposed the Fib2 subset might possess high potential for iPSCs reprogramming and neural conversion. Thus, based on the differentially expressed genes and CD markers dataset (HUGO Gene Nomenclature Committee), we further inferred several cell surface markers of Fib2 (e.g., FGFR2, F11R, PROM1, BST2, ITGA6, and EPCAM), although these surface markers showed heterogeneously expressed levels within the Fib2 subset (Additional file 5: Supplementary Fig. S5f).

EB stage

For the three EB subpopulations (EB1, EB2, and EB3), we identified genes that were upregulated compared to the iPSCs stage, respectively. These genes were enriched in “fetal brain cortex,” “epithelium,” and “brain” terms by DAVID using tissue enrichment analysis (Additional file 6: Supplementary Fig. S6a), which suggests that the biological processes of brain development and neural differentiation initiation are occurring during the iPSCs-to-EB stage transition and that these processes are shared by each EB subpopulation. Moreover, most neural TFs and cell-specific markers were expressed commonly among EB subpopulations (e.g., SOX2, ZIC2, SOX11, SOX4, SIX3) (Additional file 6: Supplementary Fig. S6b), and some of these TFs play a crucial role in neural tube formation. However, some important neural TFs, such as FOXO1 and FOXO3, play an important role in NPC proliferation and self-renewal [39]. TULP3, which regulates the sonic hedgehog (SHH) signaling pathway and modulates neural tube development [40], and POU2F1, which regulates NESTIN gene expression during P19 cell neural differentiation and CNS development [41], showed significantly high expression in the EB3 subpopulation but low expression in the EB1 and EB2 subpopulations (Additional file 6: Supplementary Fig. S6c, S6d). This suggests that different subpopulations contain specific molecular signatures and different differentiation states or potentials.

Ros-E stage

During the Ros-E stage, which is composed of NE and the cells in the early stage of rosette formation, we observed expression of several master regulator genes associated with neural tube formation and closure, including SOX11, ZIC2, PAX3, and SNAI2, in both Ros-E subgroups (Ros-E1 and Ros-E2). However, genes involved in neural crest specifiers, such as TWIST1 [42] and SOX9, which contribute to the induction and maintenance of neural stem cells and are enriched in neural crest cells [43–45], and ETS1, which regulates neural crest development through mediating BMP signaling [46], were preferentially expressed in the Ros-E1 subpopulation (Fig. 2b, 2c). The ectoderm marker, OTX1, and genes involved in the ventral hindbrain marker (e.g., IRX3) were highly expressed in the Ros-E2 subgroup (Fig. 2b, 2c). GO term annotation analysis showed Ros-E1 and Ros-E2 shared GO terms of “cell cycle G1/S phase transition,” “G1/S transition of mitotic cell cycle,” “epithelial cell proliferation,” and “positive regulating of binding” (Fig. 2d), while “negative regulation of neuron differentiation” and “tube morphogenesis” were solely enriched in the Ros-E2 subpopulation (Fig. 2d). KEGG enrichment analysis showed that “base excision repair,” “DNA replication,” “axon guidance,” “cell cycle,” and “mismatch repair” were specifically associated with the Ros-E2 subset (Fig. 2e). We fur-

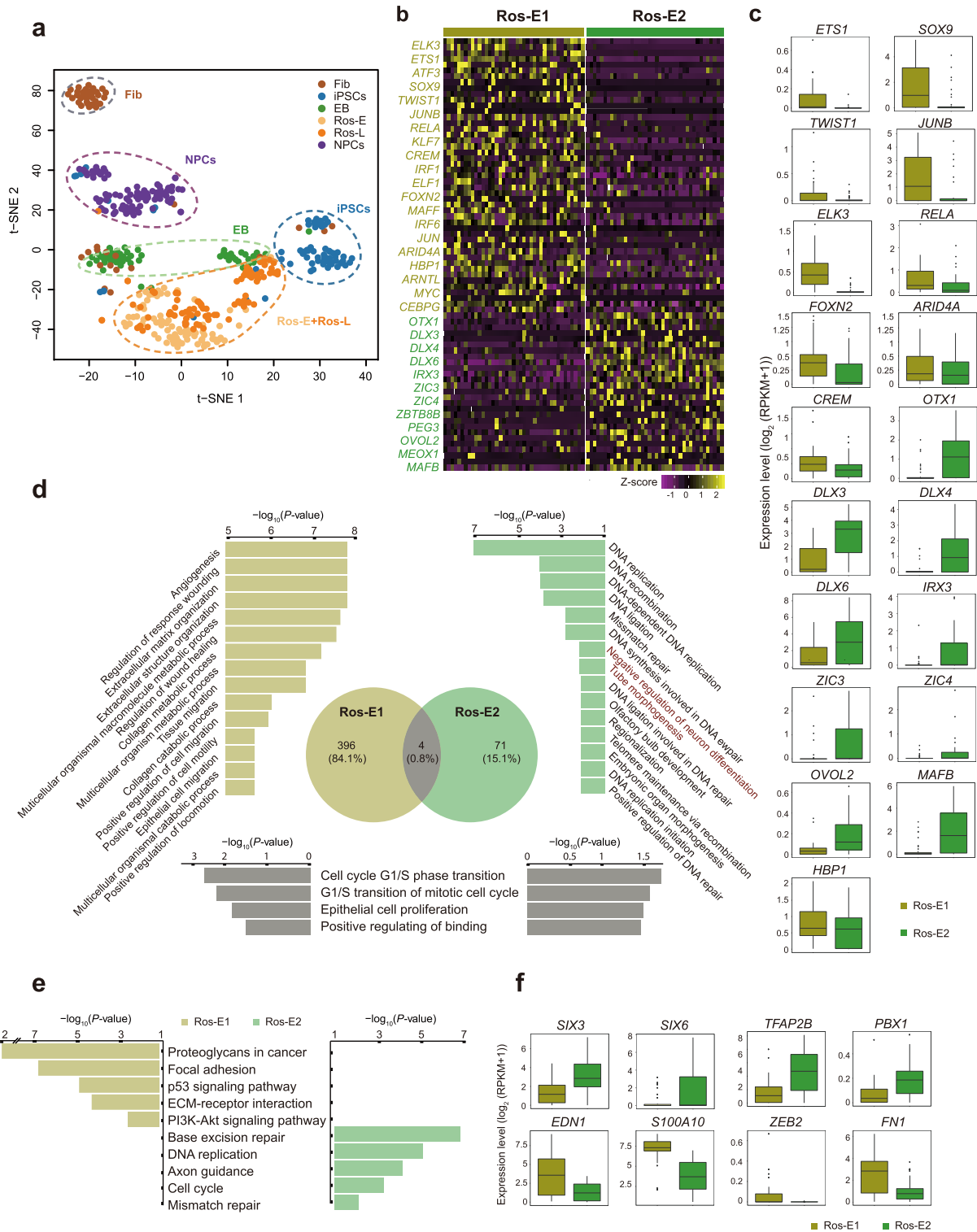


Figure 2: Cell heterogeneity and identification of subsets within Ros-E stage. **(a)** t-SNE analysis of different cell stages as indicated with different colors ($n = 445$). Number of successfully profiled single cells per cell stage: Fib ($n = 54$), iPSCs ($n = 71$), EB ($n = 57$), Ros-E ($n = 81$), Ros-L ($n = 92$), and NPCs ($n = 90$). Each dot represents an individual cell. **(b)** Heat map shows scaled expression [$\log_2(\text{RPKM}+1)$] of discriminative TF sets for each cluster at Ros-E stage, P value ≤ 0.01 . Color scheme is based on z-score distribution from -1 (purple) to 2 (yellow). **(c)** Box plot of discriminative TFs for specific subpopulation at Ros-E stage. **(d)** GO term enrichment of differentially upregulated genes respective to indicated subpopulation (highlighted with color: Ros-E1 is yellow, Ros-E2 is green, overlapped GO terms of Ros-E1 and Ros-E2 are gray). **(e)** Top five differential pathways in Ros-E1 and Ros-E2, respectively, by KEGG enrichment analysis. **(f)** Representative box plots of subpopulation-specific genes identified by SCDE analysis, adjusted P value ≤ 0.01 .

ther performed single-cell differential expression (SCDE) on both Ros-E subpopulations and identified additional differentially expressed genes between the two groups. *SIX3*, *SIX6*, *TFAP2B*, and *PBX1* were more highly expressed in Ros-E2, whereas *EDN1*, *S100A10*, and other genes related to neural crest migration were highly expressed in Ros-E1 (Fig. 2f).

Ros-L stage

At the Ros-L stage, the genes *SNAI2*, *OTX2*, *FEZF1*, *ZIC3*, and *HESX1* showed significantly different expression patterns among the three distinguishable subpopulations (Ros-L1, Ros-L2, and Ros-L3) at the Ros-L stage (Additional file 7: Supplementary Fig. S7a, S7b). Moreover, *SMAD1* and *MYC*, two components in the Wnt signaling pathway, which is critical for neural development [47, 48], were specifically enriched in the Ros-L3 subpopulation. Additionally, *JUNB* from the TGF β signaling pathway was preferentially expressed in Ros-L3 compared to the other two subpopulations. Interestingly, *HAND1* and *ISL1*, which are mesoderm markers, and *TBX3*, which elicits endodermal determination, were highly expressed in the Ros-L1 subpopulation (Additional file 7: Supplementary Fig. S7a, S7b).

Of 648 GO terms identified by differentially expressed genes among these three subsets, 52 terms were shared by Ros-L1 and Ros-L3, such as “positive regulation of cell motility,” “angiogenesis,” “positive regulation of cellular component movement,” and “epithelium migration” (Additional file 7: Supplementary Fig. S7c). A high proportion of cardiac development terms was enriched in Ros-L1, whereas DNA replication- and chromatin remodeling-related terms and pathways were significantly associated with Ros-L2. In addition, cell-substrate adhesion-related terms and cell cycle-related pathways were enriched in Ros-L3 (Additional file 7: Supplementary Fig. S7c, S7d).

Several subpopulation-specific genes were identified, including *NR2F1*, *ARID3A*, *SIX3*, *OTX2*, and *FOXG1* at the NPCs stage (Additional file 8: Supplementary Fig. S8a, S8b). These observations suggest that significant TF expression patterns describe discrepant cell differentiation states or differentiation commitments inside the neural conversion process. Taken together, our results suggest that the subpopulation analyses accurately describe specific gene expression dynamics at each cell stage, which are likely masked in bulk sequencing analyses. Additionally, extrapolating from these observations, we can reason that reconstructing a differentiation trajectory based on the gene expression dynamics of individual subpopulations would allow us to dissect neural differentiation processes that we would otherwise be unable to observe.

Tracking a reconstructed trajectory identifies key subpopulations during neural differentiation

Based on the subpopulations identified before, we wanted to track the gene expression dynamics of individual subpopulations to parse the neural differentiation processes and dissect the subpopulation with the highest contribution toward commitment to the CNS lineage. First, we reconstructed the differentiation trajectory using 8,220 genes with variable expression. This showed that cells in stages from iPSCs to NPCs followed a sequential differentiation process where each stage exhibited a relatively discriminative region with some of the subpopulations overlapping (Fig. 3a). Subsequently, based on the pairwise comparisons of TF expression levels, we inferred the connection of the subpopulations from the iPSCs stage to the NPCs stage across the five-stage differentiation process (Fig. 3b). TF expres-

sion levels were considered as strong indicators of cell state and identity [36]. Here, we used the Pearson correlation coefficient to identify more biologically and molecularly similar cell subpopulations and considered them as cells within the same developmental lineage [49]. As a result, iPSCs, EB3, Ros-E2, Ros-L3, and NPC1 were identified as the subpopulations contributing the most to commitment to the CNS lineage (Fig. 3b). These findings were consistent with the specific gene expression pattern in individual subpopulations. For instance, *SOX13*, expressed in the developing nervous system and neural tube [50, 51], and *FOXO1* [39] and *TULP3* [40] were significantly highly expressed in EB3 (Additional file 6: Supplementary Fig. S6c, d). *MAFB*, an important TF in hindbrain identity [52], was enriched in Ros-E2 (Fig. 2b, 2c); other crucial neural development TFs, especially those involved in CNS development, such as *OTX1*, *DLX3*, *DLX6*, *ZIC3*, *ZIC4*, and *IRX3*, also showed high expression in the Ros-E2 subpopulation (Fig. 2b, c). Previously, we assumed that *GRHL3* might be involved in neural tube closure; here, the results showed that *GRHL3* was indeed significantly highly expressed in Ros-L3 (Additional file 7: Supplementary Fig. S7b). Additionally, neural crest regulators (e.g., *ETS1*, *ELK3*, *SOX9*) were enriched in Ros-L3 (Additional file 7: Supplementary Fig. S7b), suggesting that cell fate specification and differential cell status might exist even within subsets. Strikingly, Ros-E2 and Ros-L3 that were identified in the dominant path to CNS lineage by correlation analysis were shown as a process of sequential conversion in our reconstructed trajectory (Fig. 3a, 3c). The molecular signature described by these subpopulations was consistent with the analysis that identified the key contributing subpopulations and encouraged us to perform additional cell fate decision analyses.

Of note, there was a clear divarication within the rosette stages (Ros-E and Ros-L) across the differentiation trajectory, indicating cell fate decision might be made at this bifurcation point (Fig. 3c). Here, we focused on the single cells in the rosette stages and called them branch 1, branch 2, and branch 3 based on their location in the differentiation trajectory (Fig. 3c). Branch 3 was composed of Ros-E1 ($n = 27$), Ros-L1 ($n = 15$), and a small proportion of Ros-E2 ($n = 5$) and Ros-L3 ($n = 9$; Fig. 3c). Previously, our observations showed that Ros-E1 was associated with neural crest cells (high expression of *TWIST1*, *SOX9*, *ETS1*, *EDN1*, and *S100A10*) and Ros-L1 was likely related to mesoderm and endodermal determination (high expression of *HAND1*, *ISL1*, and *TBX3*); these two subpopulations comprise the majority of cells in branch 3. Further, we performed a pairwise comparison of gene expression across the three branches. The results showed that many neural TFs, such as markers of neural tube formation (*SOX4* and *SOX11*); the neural stem cells (NSCs) self-renewal and proliferation regulator *FOXO3*; and the NSCs markers *NES*, *CDH2*, and *FABP7*, were commonly expressed across all three branches, indicating the capacity for neural tube development and NSCs proliferation are a fundamental feature of neural rosettes (Additional file 9: Supplementary Fig. S9a, S9b). Strikingly, *ZIC2*, a member of the ZIC family of C2H2-type zinc finger proteins, associated with neural tube development [32], showed significantly low expression in branch 3 (Fig. 3d, 3e). Some other neural development markers (e.g., *ZIC3*, *HMG2*, *ID1*, *SIX3*, *SIX6*, *NR6A1*) were significantly lowly expressed in branch 3 but highly expressed in branch 1 (Fig. 3d, 3e, Additional file 9: Supplementary Fig. S9a, S9c). However, *TFAP2B*, encoding a member of the AP-2 family of TFs, and *ELK3*, essential for the progenitor progression to neural crest cell [53], were significantly highly expressed in branch 3 but lowly expressed in branch 2. Moreover, *SOX9*, *SNAI2*, *S100A11*, and *TFAP2A*, previously shown to be highly expressed in neural crest cells [43–45, 54], were markedly highly expressed

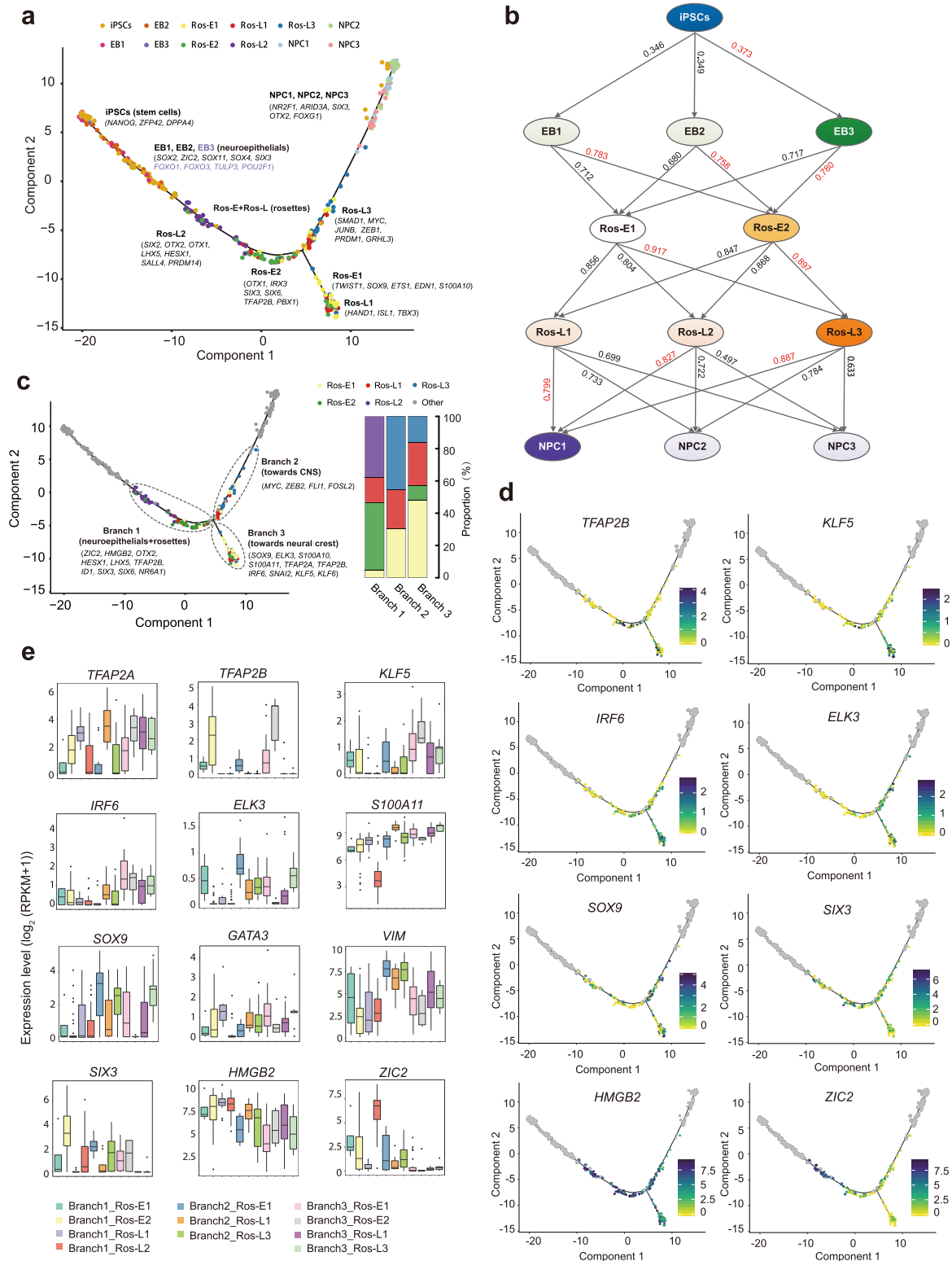


Figure 3: Cell fate specification revealed by reconstructed trajectory. **(a)** Differentiation trajectory constructed with 8,220 variable genes across different cell stages. Selected marker genes specific to the respective cell stage/subpopulation are indicated with black/purple. **(b)** The connection of subpopulations from iPSCs to NPCs stage across the five-differentiation process identified by Pearson correlation coefficient. The Pearson correlation coefficient of the two comparisons in the arrow line, respectively. **(c)** The divarication point within rosette stage (Ros-E and Ros-L) across the differentiation trajectory, branch 1, branch 2, and branch 3 based on their location on the differentiation trajectory are marked by dashed ellipse. Selected discriminative TFs specific to the respective branch are indicated. The columns represent the components of branch 1, branch 2, and branch 3, respectively. **(d)** Expression pattern of selected differentially expressed TFs among the three branches on the reconstructed trajectory (adjusted P value ≤ 0.01). Color scheme is based on expression [$\log_2(\text{RPKM}+1)$]. **(e)** Expression pattern of representative differentially expressed TFs across different components of the three branches.

in branch 3 but not branch 1 (Fig. 3d, 3e, Additional file 9: Supplementary Fig. S9a, S9c). *KLF5* and *IRF6* were significantly highly expressed in branch 3 as well (Fig. 3d, 3e). These two TFs have been reported to be involved in phenotypic switching of vascular smooth muscle cells [55] and development of the palate in vertebrates involving cranial neural crest migration [56], respectively. These results indicate that cell fate specification might occur at the bifurcation point. Based on these observations, we speculate that branch 1-to-branch 2 has progressed more toward CNS and branch 3 is probably composed of neural crest cells and other cells comprising this microenvironment.

Construction of the TF regulatory network during cell status transition

To infer TFs that drive the progression of cell status from one stage to the neighboring one, we performed SCDE analysis for those cell subpopulations committing to CNS lineage, resulting in 58, 123, 98, and 131 TFs differentially expressed among iPSCs vs EB3, EB3 vs Ros-E2, Ros-E2 vs Ros-L3, and Ros-L3 vs NPC1 comparisons (Additional files 10, 11: Supplementary Figs. S10, S11, and Additional file 19: Supplementary Table S1). Interestingly, *PRDM1*, which has been proposed to promote the cell fate specification RB sensory neurons in zebrafish [57], was significantly upregulated from Ros-E2 to Ros-L3 (Additional file 10: Supplementary Fig. S10). In contrast, several well-characterized TFs were found to be significantly highly expressed in Ros-E2 (mainly resident in branch 1) and downregulated during the transition from early to late rosette development: *FOXP1*, cooperating with *Bmi-1* to maintain neural stem cell self-renewal in the forebrain; *MAFB*, the posterior CNS fate identifier and essential for hind-brain choroid plexus development [52, 58]; *DLX3* and *DLX5*, neural plate border specifier genes [58]; and *ID1*, a controller of stem cell proliferation during regenerative neurogenesis in the adult zebrafish telencephalon [59]. These results suggest that the expression patterns of neural-associated TFs undergo dramatic changes during neural differentiation with some TFs activated (e.g., *PRDM1*) and others repressed (e.g., *MAFB*, *FOXP1*, *ID1*) (Additional file 10: Supplementary Fig. S10). Furthermore, it was previously unknown that several of these TFs were involved in neural differentiation, so our results have expanded the known biological functions of these molecules.

Among the 131 TFs exhibiting differential expression from Ros-L3 to NPC1, 80 TFs were upregulated while 51 TFs were downregulated (Additional file 11: Supplementary Fig. S11; Additional file 19: Supplementary Table S1). Upregulated TFs included *SNAI2*, a neural crest specifier [58]; *HIF1A*, required for NSCs maintenance and vascular stability in the adult mouse [60]; *SIX1*, which drives the neuronal developmental program in the mammalian inner ear [61]; *ETV1*, which orchestrates gene regulation during the terminal maturation program of cerebellar granule cells [62]; and *POU3F3*, which influences neurogenesis of upper-layer cells in the cerebral cortex [63] (Additional file 11: Supplementary Fig. S11). This is consistent with our previous observation that the main trajectory has progressed more toward the CNS. Of particular interest is *PRDM1*, whose expression increased from Ros-E2 to Ros-L3 and decreased during the progression from Ros-L3 to NPC1 (Fig. 4g and Additional files 10, 11: Supplementary Figs. S10, S11), suggesting that it might play multiple specific roles in neural differentiation.

Next, we inferred a regulatory network among those differentially expressed TFs based on known interactions collected in the STRING database [64]. Our results suggested that *SOX2* and *GATA3* were key regulators from iPSCs to EB3 (Additional

file 12: Supplementary Fig. S12a); *TP53*, *SOX2*, *RELA*, *SIX3*, *ARNTL*, *ISL1*, *RARA*, *TP63*, *GATA3*, *SNAI2*, and *PAX3* were the key regulators from EB3 to Ros-E2 (Additional file 12: Supplementary Fig. S12b); *MYC*, *SOX2*, *PAX6*, *EGR1*, *PBX1*, *GLI3*, *PAX3*, *SIX3*, *FOXP1*, *OTX2*, *PAX7*, *PPARG*, *SOX9*, *MAFB*, *SIX6*, and *ZIC1* were identified as key regulators from Ros-E2 to Ros-L3 (Fig. 4a); and *SOX2*, *AR*, *MYCN*, *LEF1*, *PAX3*, *SNAI2*, *MSX1*, *SOX9*, *NR3C1*, *PARP1*, *RUNX1*, *EBF1*, *HIF1A*, *IRF6*, *IRF1*, *KLF5*, and *LIN28A* were predicted to be key regulators from Ros-L3 to NPC1 (Fig. 4b).

To dissect the cis-regulatory elements directing the expression of those regulators, we selected the differentially expressed TFs that showed differential ATAC peaks between neighboring stages and performed motif scanning on the differential peaks. Focusing on the transition from Ros-E2 to Ros-L3, we found transcription factor binding sites (TFBSs) for *TEAD2* and *YY1* in a differential ATAC peak downstream of the *PRDM1* gene (Fig. 4c). Multiple motifs for the transcription factor *TFAP2C* were found in a differential peak located in the intron of the *ARID3A* gene, which is a regulator responsible for the transition for Ros-L3 to NPCs (Fig. 4d). Based on the temporal specificity of ATAC peaks and the existence of TF motifs in these regions, we propose that those elements are stage-specific cis-regulatory elements regulating the expression of neural regulators in response to their upstream regulatory TFs.

To infer the putative targets of key regulators, we combined the information from ATAC peaks and motifs for TFs. All peaks containing motifs for a certain TF were annotated as TF-related peaks, and genes proximal to the peak were considered as potential targets of that TF. Using these criteria, we predicted thousands of targets for the inferred TFs (Additional file 20: Supplementary Table S2). To dissect the regulatory network of these TFs, we conducted GO term and KEGG enrichment analysis for the putative target list of selected regulators (e.g., *PRDM1*, *NR2F1*, *SOX9*, and *TFAP2C*). Our results suggested that, from Ros-E2 to Ros-L3, the targets for *PRDM1* were significantly enriched in pathways and GO terms associated with “axon guidance,” “hippo signalling pathway,” and “neurotrophin signalling pathway” (Fig. 4e and Additional file 13: Supplementary Fig. S13a). From Ros-L3 to NPC1, targets for *NR2F1*, *SOX9*, and *TFAP2C* were enriched in KEGG pathways associated with “axon guidance” and “hippo signalling pathway” (Additional file 13: Supplementary Fig. S13b-S13d). We further validated *PRDM1* expression among different genetic background cell lines (H1_ESCs, H7_ESCs, H9_ESCs, iP25, and iP2129). The immunostaining showed that *PRDM1* was expressed at Ros-L stage with heterogeneous expression level, though the scRNA-seq data were not at a high level. Moreover, the results were uniformed across these cell lines (Fig. 4g, 4h).

Inferring a cellular communication network among cell subpopulations within specific differentiation stages

Cell subpopulations with different functions are proposed to exhibit distinct expression profiles of ligands and receptors that prime cells for cell type-specific interactions [65]. In this study, the cellular interactions were inferred using public ligand-receptor databases (see Methods section). Briefly, 360, 182, 261, and 307 ligands/receptors were expressed within EB, Ros-E, Ros-L, and NPCs subpopulations, respectively, among which 304, 55, 124, and 162 interactions were identified within subpopulations at each differentiation time point (Fig. 5, Additional files 14–16: Supplementary Figs. S14-S16 and Additional file 21: Supplementary Table S3). The most frequent interactions were observed in the EB stage, implying that cells communicate extensively to co-

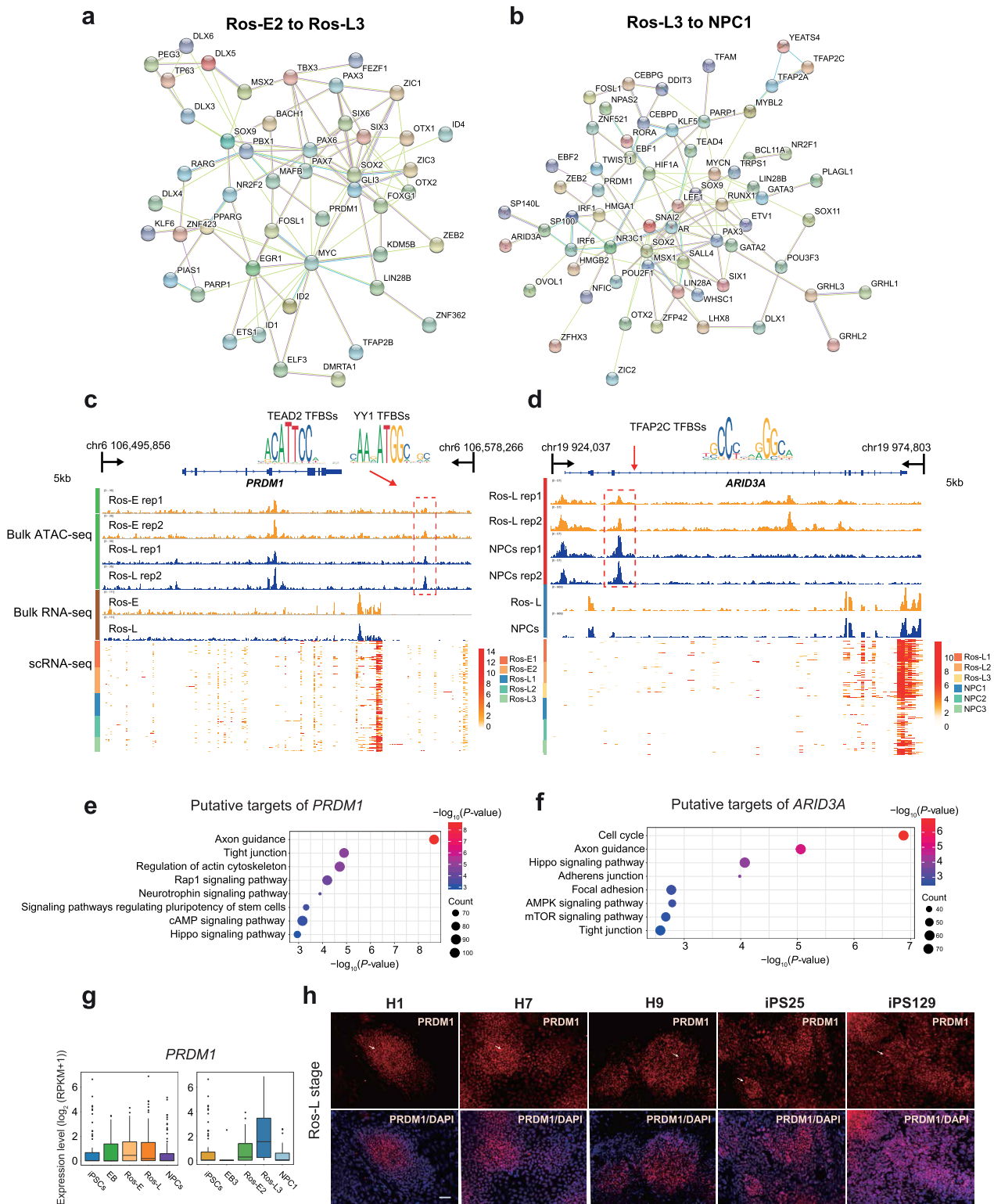


Figure 4: Putative regulators and corresponding cis-regulatory elements during neural differentiation. (a) Regulatory network of TFs differentially expressed between Ros-E2 and Ros-L3. (b) Regulatory network of differentially expressed TFs between Ros-L3 and NPC1. (c, d) IGV screenshots of ATAC-seq and bulk RNA-seq as well as the corresponding scRNA-seq heat maps for putative neural regulator PRDM1 (c) and ARID3A (d). Differential peaks in the dashed boxes possess putative TF motifs outlined in the form of sequence logo. (e, f) KEGG enrichment analysis of putative target genes under the regulation of PRDM1 (e) and ARID3A (f). (g) Expression pattern of PRDM1 at indicated cell stages (left) and subsets (right) during neural differentiation. (h) Immunostaining of PRDM1 at Ros-L stage across different genetic background cell lines (H1.ESCs, H7.ESCs, H9.ESCs, iPS25, and iPS129). Scale bar represents 50 μm .

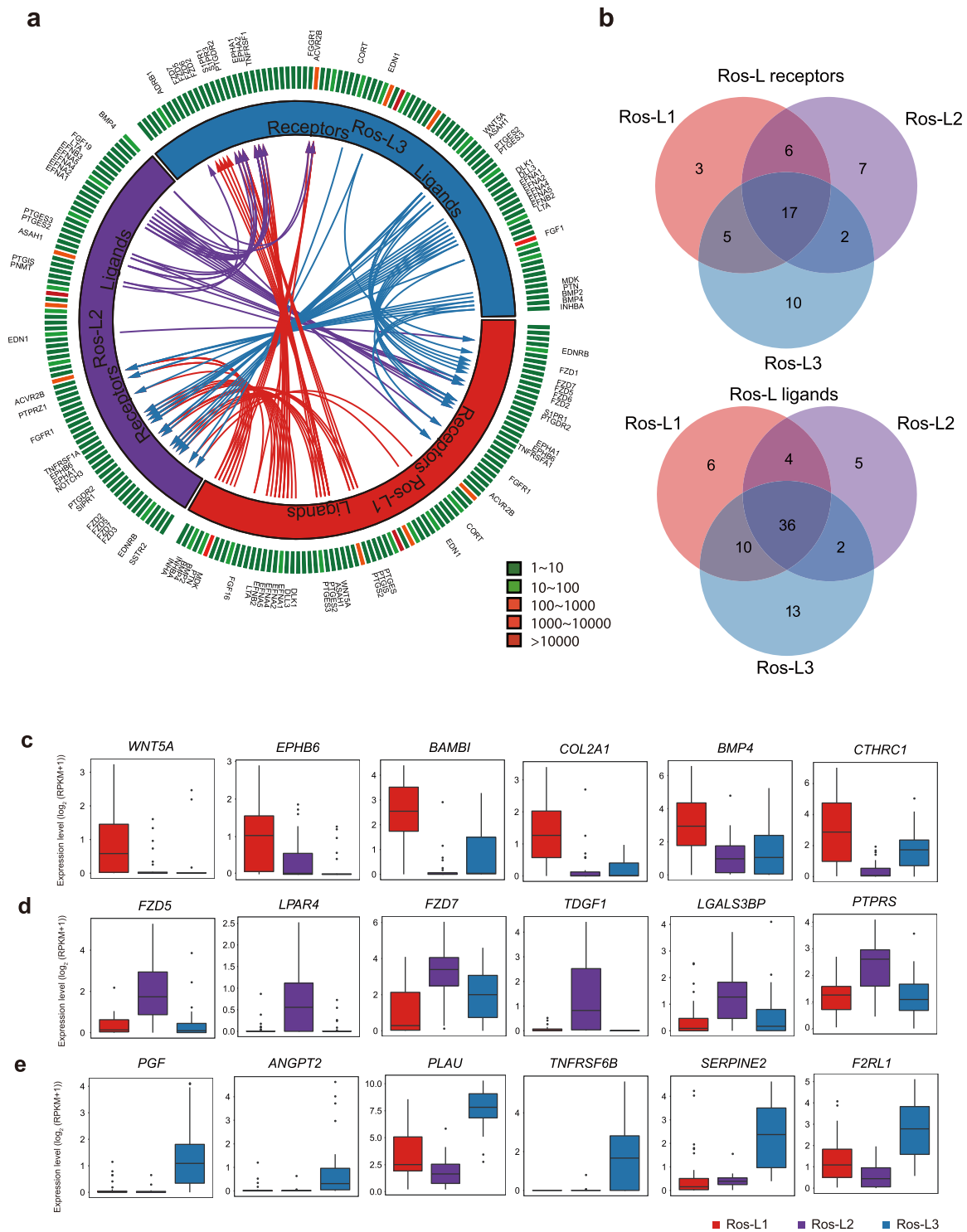


Figure 5: Putative receptor-ligand interactions in Ros-L subsets. **(a)** Putative signaling between expressed receptors and their ligands in Ros-L subsets. The inner layer compartments represent different cell subpopulations (Ros-L1, Ros-L2, and Ros-L3 shown in red, purple, and blue, respectively). The outer layer indicates the expression profiles of ligands and receptors expressed in each cell subset, with low expressed molecules in green and high expressed molecules in red. Arrows indicate putative interactions between ligands and receptors among cell subsets. **(b)** Venn plot showing the overlapping of ligands and receptors among cellular subpopulations. **(c-e)** Expression level of receptors/ligands enriched in Ros-L1 **(c)**, Ros-L2 **(d)**, and Ros-L3 **(e)**, respectively.

ordinate differentiation programs during embryogenesis (Additional file 14: Supplementary Fig. S14). In contrast, much fewer interactions were predicted after the EB stage, suggesting communications decreased dramatically during the progression of

lineage commitment. Notably, although comparable numbers of ligands and receptors were detected at the EB (181 receptors and 179 ligands) and NPCs (128 receptors and 179 ligands) stage, only half the interactions (162) were inferred at the NPCs stage com-

pared to 304 ligand-receptor interactions at the EB stage (Additional files 14, 16: Supplementary Figs. S14, S16). The interactomes among Ros-L cells, with 31, 32, and 34 receptors from Ros-L1, Ros-L2, and Ros-L3 interacting with ligands from other cell subpopulations, were inferred (Fig. 5a, 5b). As expected, several interactions involving receptors and ligands previously known to play essential roles during neural development were identified in our study. For example, *WNT5A* and *EPHB6* were enriched in Ros-L1. *FZD5* and *LPAR4* were specifically expressed in Ros-L2. *PGF* and *ANGPT2* were upregulated in Ros-L3 compared to other cell subpopulations (Fig. 5c, 5d, 5e). Overall, our study suggests that the specific expression spectrum of ligands and receptors and corresponding interactions can generally reflect the identity of cellular subpopulations.

Discussion

The regulation and molecular programs during embryonic neural development have long been investigated. However, much of this work has been limited to model organisms such as the mouse, zebrafish, and *Drosophila* [36, 40, 56] due to the scarcity of human fetal tissue for research purposes. Our understanding of human early neural development and, particularly, neural tube formation and the cell fate commitments of neural precursors in early stages is still incomplete. To circumvent the challenges inherent in these investigations, namely, the ability to study these processes *in vivo* in humans, we used hiPSCs and induced differentiation *in vitro* toward a neural cell fate using a well-established model. We characterized both the transcriptional profiles in single cells as well as chromatin accessibility at several critical stages during differentiation to inform this process at unprecedented resolution. This study has unveiled the dynamic transcriptome and regulome underlying early human neural differentiation and identified functionally distinct subpopulations within the various stages to have a more precise description of the factors defining the differentiation trajectory. Our analyses hint at the existence of a widespread regulatory network between TFs and their target genes, especially those associated with cellular reprogramming and differentiation. We were also able to construct minimal gene expression profiles based on ligands and receptors in each cell subpopulation that can be used to confidently infer cell identity.

During development *in vivo*, the neuroectoderm folds to form the neural tube that is then patterned into regionally specialized subunits composed of progenitor cells. These cells subsequently give rise to regional progenies of neural cells [66]. There is some controversy in this field that formation of the EB would introduce *in vitro* culture variability in regional cells across different batches, resulting in a relatively poor model of neural differentiation. The “dual-SMAD inhibition” method (inhibiting the SMAD-dependent TGF β and BMP signaling pathways) yielding neural epithelia in “monolayer culture” conditions [18] could alleviate the above concern. However, generation of neural rosettes morphology *in vitro* is considered equivalent to neural tube formation, recapitulating neural tube structure, which we believe is a promising research model for early neural differentiation. Neural differentiation of hiPSCs into NPCs starts with initial neural induction by appropriate dosages and gradients of many TFs and morphogenetic factors that are highly expressed in the developing brain. In this study, the induction cocktail used in the neural differentiation included SB431542, dorsomorphin, N2, B27, VEGF, and bFGF supplemented at specific time points. The self-renewal program in hiPSCs is switched off and differen-

tiation toward NE and NPCs is triggered [8, 16]. Previous results have shown that SB431542 enhances neural induction in EB derived from hESCs [67] by inhibiting the Lefty/Activin/TGF β pathways and suppressing the mesodermal lineage (Brachyury) induction [18, 42]. Consistent with these previous studies, in our *in vitro* system, treatment with SB431542, in combination with dorsomorphin, results in a dramatic decrease in NANOG expression and a concomitant increase in PAX6 expression (Fig. 1f and Additional file 4: Supplementary Fig. S4h). In addition, *OTX2*, *ZIC2*, *SOX9*, *HESX1*, *MSX2*, *DLX5*, *SOX4*, *SOX11*, and *SNAI2* were significantly activated during differentiation, which demonstrates that the transcriptional program triggering progression toward NPCs was activated (Fig. 1f, 3d, 3e, Additional file 4: Supplementary Fig. S4h and Additional file 9: Supplementary Fig. S9a-S9c). Taken together, these results indicate that the induction cocktail effectively achieves efficient neural differentiation.

To measure the dynamic changes of cis-regulatory elements at each differentiation stage, we performed ATAC-seq and chromatin accessibility analysis on bulk cells. These results showed widespread and comprehensive chromatin structure reprogramming during neural differentiation. In particular, TFBSs for several neural master regulators were enriched in temporally dynamic ATAC peaks, indicating that changes in chromatin accessibility are indeed associated with, and are probably responsive to, the regulation of neural-related TFs. In addition, we also investigated closing (lost) peak dynamics as well as the functional annotation study, which was in line with the corresponding annotation of novel peaks (Fig. 1d, 1e and Additional files 2, 3: Supplementary Figs. S2, S3). We further identified several enriched TF motifs (e.g., *Pax2* in Ros-L and *FOXO1* in NPCs) (Additional file 17: Supplementary Fig. S17), which are known to play an important role in neural differentiation, consistent with results from previous studies [39, 68].

By integrating single cell-based transcriptome profiling of 391 cells from five differentiation stages, we identified a variety of TFs that were differentially expressed throughout the differentiation process and showed distinct expression profiles among specific cell stages. The TFs *SOX2*, *PAX6*, *OTX2*, *SOX4*, *ZIC2*, *LHX5*, *HESX1*, and *SIX3* were significantly highly expressed at the EB stage (Fig. 1f). It has been reported that members of the grainyhead-like (*Grhl*) family of TFs, which are well conserved from *Drosophila* to human, are highly expressed during neurulation in mice and that a *Grhl3*-hypomorphic mutant resulted in NTDs [32, 69]. Remarkably, our results showed that two human *Grhl* family TFs, *GRHL2* and *GRHL3*, were significantly highly expressed at EB and Ros-E stages, respectively (Fig. 1f and Additional file 4: Supplementary Fig. S4h), and the downstream targets of *GRHL2* (including *E-CADHERIN*, also known as *CDH2*) were highly expressed at the neural rosette stage (Fig. 1b), supporting a role for *Grhl* TFs in neural tube closure in humans. In addition, previous studies have shown that in the *Drosophila* olfactory system, the homeobox gene *distal-less* is required for neuronal differentiation and neurite outgrowth [34]. Our data showed that four homologs of *distal-less* (*DLX3*, *DLX4*, *DLX5*, *DLX6*) were significantly upregulated at the Ros-E stage and were highly expressed in the Ros-E2 subpopulation (Figs. 1f and 2b and Additional file 4: Supplementary Fig. S4h), implying that the *distal-less* gene family plays a role in neural differentiation in humans.

We also applied scRNA-seq to our *in vitro* neural model to dissect the subpopulations present at each differentiation stage (Fig. 2 and Additional files 5–8: Supplementary Figs. S5–S8). We were then able to reconstruct a differentiation trajectory based on the subpopulations that we identified by variable TF expression within each stage (Fig. 3a). Strikingly, a divarication within

the rosette stage across the differentiation trajectory was observed. Comparing branch 1 to branch 3, branch 3 possessed the relatively lowly expressed TFs *LHX5*, *HESX1*, and *SIX3* (reported as anterior forebrain markers), as well as other crucial neural TFs (*SOX2*, *HMGB2*, *ZIC2*, *OTX1*, *FEZF1*) and the relatively highly expressed TFs *TFAP2B*, *SOX9*, *ELK3*, and *SNAI2* (Fig. 3d, 3e and Additional file 9: Supplementary Fig. S9a, S9c), which are considered to be neural crest markers [53]. Though *SNAI2* was also expressed at the NPCs stage, combined with other neural crest markers, we proposed that branch 3 was progressing more toward neural crest cells (Fig. 3a-3c and Additional file 9: Supplementary Fig. S9a-S9c). Taken together, these observations imply that the main differentiation trajectory (branch 1 and branch 2) is heading toward CNS, whereas branch 3 is progressing towards neural crest cells.

It is important to note that the current scRNA-seq method, by its nature, only provides a snapshot of the gene expression profile for individual cells. A possible resolution for the above problem is to capture the sample with much more precise time points, which may, to some extent, overcome this limitation. Thus, in spite of the very interesting heterogeneity and cell fate commitment study inferred above, we cannot exclude the following factors that may affect cell subset identification in the above description: temporal transcriptional states during transient differentiation process, differentiation efficiency, and lagging and leading cells remaining in the differentiation process. However, we propose that the subsets dissection analysis facilitates a more precise description of the factors defining the differentiation trajectory. When we constructed the differentiation trajectory using the cells that collected at different time points, the results showed that all subpopulations in stages from iPSCs to NPCs followed a sequential differentiation process where each stage exhibited a relatively discriminative region with some of the subpopulations overlapping (Fig. 3a), indicating that in spite of the above concerns, the trajectory was established by the natural features of the respective subsets. This is also supported by the observations that Ros-L2 possessing many early neural differentiation TFs, such as *SOX2*, *OTX2*, *PAX6*, *OTX1*, and *LHX5*, as well as forebrain markers (e.g., *HESX1*) and pluripotency-related TFs (*NANOG*, *SALL4*, *PRDM14*) (Additional file 7: Supplementary Fig. S7a, S7b) were located in the reconstructed trajectory prior to the generation of Ros-E populations (Fig. 3a, 3c). In addition, we carried out the cell fate commitment analysis using branch 1, branch 2, and branch 3, which were grouped based on the cell locations on the trajectory rather than cell subsets identified by Seurat in order to minimize the above concerns.

Notably, our study reveals the regulatory network of TFs that are differentially expressed among neighboring cell subpopulations to be likely candidates for promotion of cell fate transition. Based on the topology of this network, we focused on novel regulators (*PRDM1* and *ARID3A*), especially *PRDM1*, which are located on the hub of the network, interacting with both known and novel neural regulators. Although the roles of several TFs have been reported during neural differentiation and brain patterning formation in humans, some TFs have been proposed to play a role in neural fate commitment in non-human species (mouse and zebrafish). However, the interaction partners, cis-regulatory elements, and genetic regulatory networks of those TFs are yet to be resolved. Here, we identified the cis-regulatory elements for *PRDM1* and *ARID3A* genes and predicted their upstream regulators. Of particular interest, *TFAP2C*'s role in regulating neural development has been widely reported, increasing the confidence of our predictions. In humans, *PRDM1* is reported

to promote germ cell fate by suppressing neural effector *SOX2*, but the function of *PRDM1* in neural development is unknown. In zebrafish, *Prdm1a*, the homolog of the *PRDM1* gene, directly activates *foxd3* and *tfap2a* during neural crest specification [57]. Mutation of *prdm1* in zebrafish resulted in severe phenotypes with a decrease in the quantity of neural crest cells and the reduction in the size of structures derived from the neural crest [57]. Similarly, strong expression of *prdm1* was observed in the neural plate border of a basal vertebrate lineage, lamprey, implying that the role of *prdm1* in neural crest formation is likely a conserved, ancestral role [70]. Conversely, *prdm1* is dispensable for neural crest formation in mice and, instead, is required for primordial germ cell specification, suggesting that the neural crest specification function of *prdm1* in mice has been lost [71]. Overall, previous studies suggest that functions of *prdm1* are quite diverse and need to be investigated in species-, developmental-, and environmental-specific manners. Based on the known interaction between *PRDM1* and *SOX2* in humans, as well as the observation that *PRDM1* expression increased significantly from Ros-E2 to Ros-L3 and was preferentially expressed in Ros-L3 compared to the other two subpopulations in the rosette stage (Fig. 4g, 4h; Additional file 7: Supplementary Fig. S7a, S7b; and Additional file 10: Supplementary Fig. S10), we propose *PRDM1* as a novel neural regulator in early human neural differentiation. Our hypothesis is supported by the GO term and KEGG enrichment analysis of putative targets of *PRDM1*, which are significantly enriched in "axon guidance" and hippo pathway-associated terms (Fig. 4e and Additional file 13: Supplementary Fig. S13a). However, the functions of putative TFs need to be further investigated using experimental methods.

To infer cellular interactions, communication network analysis was applied to the expression profiles of ligands and receptors in stage-specific subpopulations. Two trends were observed in our cellular interaction network analysis: the frequency of cellular interactions peaked at the EB stage and different cell subpopulations showed a certain degree of specificity in their ligand-receptor spectrum. The observation that most interactions were inferred at the EB stage likely reflects the extensive cellular communication during embryogenesis and early neural differentiation (Additional file 14: Supplementary Fig. S14). Regarding the ligand-receptor expression spectra, matched ligand and receptor expression probably underlies the common functions shared by different cell subpopulations within the same stage. In contrast, those specific ligands or receptors probably reveal the unique regulatory code of distinct cell subpopulations. For example, *WNT5A*, a crucial regulator of neurogenesis during the development of cerebellum, and *BMP4*, one of the key regulators of dorsal cell identity in the neural tube [72], were highly expressed in Ros-L1 compared to other cell subpopulations (Fig. 5c). *FZD5* (required for eye and retina development in mouse [73]), and *FGF19* (required for forebrain development in zebrafish [74]) were preferentially expressed in Ros-L2 (Fig. 5d and Additional file 22: Supplementary Table S4). *WNT7A*, involved in several aspects of neurogenesis, including synapse formation and axon guidance [75], and *FGF1*, which maintains the self-renewal and proliferation of NPCs [76], were specifically expressed in Ros-L3 (Additional file 22: Supplementary Table S4). Pavličev et al. inferred the cell communication network of the maternal-fetal interface and found that ligand-receptor profiles could be a reliable tool for cell type identification [65]. Consistent with their findings, our study suggests that the repertoire of ligands-receptors in neural cell types could probably, to some extent, represent the identity of cell subpopulations.

There might be a concern that we only used one genetic background cell line for this study, possibly making the cogency of our findings limited. To address this, we performed ESCs neural differentiation and captured bulk transcriptome profiles of the corresponding differentiation stages (ESCs, EB, Ros-E, Ros-L, and NPCs). The observations in ESCs were reproducible in iPSCs with regards to PCA analysis (Additional file 18: Supplementary Fig. S18a); with a high Pearson correlation coefficient between the corresponding cell stage derived from iPSCs and ESCs (Additional file 18: Supplementary Fig. S18b); and validation analysis of subset-specific markers (MAFB, SOX9, PRDM1, and NR2F1). In addition, novel neural TF (PRDM1) expression in different genetic cell lines (H1.ESCs, H7.ESCs, H9.ESCs, iP525, and iP5129) was consistent with the above heterogeneity study (Fig. 4h and Additional file 18: Supplementary Fig. S18c, S18d, S18e). Together, our findings are supported by different genetic cell lines mitigating the concern that our results are limited to the cells forming the basis of this study.

Through differential expression analysis, we identified genes specifically expressed at each stage, which include both cell status master regulators such as TFs and signaling components, as well as realizators [24], which could directly determine cell growth, cell proliferation, cell morphology, and cell-cell interaction. Within each stage, we identified subpopulations with distinct expression signatures, which might represent functional cell clusters or transient cell state given that neural cells have been shown to demonstrate significant heterogeneity as they express different surface proteins, exhibit diversified morphologies, and secrete a variety of cytokines. Therefore, it is necessary to explore the heterogeneity of cell subpopulations and study each subpopulation in a case-by-case manner. In summary, our data show conclusively that both transcriptome and regulome dramatically change during neural differentiation, which affects a variety of biological pathways crucial for neural differentiation. We also propose several putative TFs as well as the ligands-receptors interaction spectrum that are important in each differentiation stage. This paves the way for a deeper understanding of the cell fate decision and regulatory mechanisms driving the differentiation of the neural lineage.

Materials and Methods

Ethics statement

The study was approved by the institutional review boards on the ethics committee of BGI (permit BGI-IRB 14057). The participants (dermal fibroblast, Fib129 and Fib 25) signed informed consent and voluntarily donated the samples for our study.

Cell culture and reprogramming

The human fibroblast cell line was derived from the dermal skin of a healthy donor, with written informed consent. Briefly, the skin tissue was washed several times with Dulbecco's phosphate-buffered saline (DPBS), sliced into approximately 1 mm or smaller fragment size, enzymatically dissociated in high Dulbecco's modified Eagle medium (H-DMEM, Gibco, 11965118) with 100 U/mL collagenase type IV incubating in 37°C overnight and then 0.05% trypsin incubating for 5 minutes. The dissociation was terminated by adding 2 mL of fibroblast cell culture medium (H-DMEM + 10% fetal bovine serum [FBS] + 5 ng/mL basic fibroblast growth factor [bFGF] + 2 mM glutamine [Gln]) followed by centrifugation at 300 g for 5 minutes. The cells were resuspended with fibroblast cell culture medium and cultured at

37°C in a 5% CO₂ incubator. The fibroblast cell culture medium was changed every 2 days until reaching 80%–90% confluence, and cells were passaged every 3–4 days.

For reprogramming, non-integrative human iPSCs were generated following a modified Shinya Yamanaka method [77]. Briefly, 5×10^5 human fibroblast cells at passage 4 were nucleofected with the program for human dermal fibroblast NHDF (Lonza, CC-2511) with 2.4 μ g episomal plasmids, including pCXLE-hOCT3/4-shp53-F (Addgene, 27077), pCXLE-hSK (Addgene, 27078), and pCXLE-hUL (Addgene, 27080). Transfected cells were cultured in a six-well plate with culture medium containing H-DMEM supplemented with 10% FBS. The cells were trypsinized, and 1×10^5 cells were seeded onto a 10 cm² dish covered with feeder and cultured in a medium containing H-DMEM with 10% FBS while reaching 80% confluence. After that, the medium was changed to hiPSCs medium containing DMEM/F12 (Gibco, 11320-033), 20% knockout serum replacement (KSR) (Gibco, 10828-028), 2 mM L-glutamine (Sigma, G8540), 0.1 μ M non-essential amino acid (NEAA) (Gibco, 11140-050), 0.1 μ M β -mercaptoethanol (Gibco, 21985-023), and 10 ng/mL human bFGF (Invitrogen, PHG0021). The iPSCs colonies were picked at around day 25 and maintained in hiPSCs medium.

Neural differentiation

We applied a well-adopted neural differentiation protocol [8, 16]. Briefly, human iPSCs were maintained as described above. To induce neural rosettes, hiPSCs were mechanically picked and washed with DMEM/F12 twice and then cultured for 4 days in suspension with 5 μ M dorsomorphin (Sigma, P5499) and 5 μ M SB431542 (Sigma, S4317) in hiPSCs medium without bFGF for EB formation. Then, the EBs were attached on matrigel (BD, 354277) coated dishes (BD, 354277) and cultured in DMEM/F12 (Gibco, 11320-033) supplemented with 20 ng/mL bFGF, $1 \times N2$ (Gibco, 17502-048) and 2 μ g/mL heparin (Sigma, 1304005) for an additional 3 or 5 days to harvest Ros-E and Ros-L cells, respectively. To collect NPCs, rosettes structure that appeared in the center of attached colonies at the Ros-L stage were carefully harvested using pulled glass pipettes and seeded on matrigel-coated dishes and cultured in DMEM/F12 supplemented with $1 \times N2$, $1 \times B27$ (Gibco, 12587-010), 20 ng/mL bFGF, 20 ng/mL epidermal growth factor (EGF) (Invitrogen, PHG0311), and 2 μ g/mL heparin for an additional 7 days; the medium was changed every 2 days. At day 16, the NPCs reaching approximately 80% confluence were collected, and all the mass or adherent cell samples were treated with TrypLE Express Enzyme (Gibco, 12604-021) for single-cell dissociation and cryopreservation in gas-phase liquid nitrogen for further sequencing.

Immunofluorescence staining

Cells were fixed in 4% paraformaldehyde in DPBS for 20 minutes and permeabilized with 1% Triton X-100 for 20 minutes at room temperature. After 60 minutes blocking with 2% normal goat serum, cells were incubated with primary antibody overnight at 4°C, washed, and stained with secondary antibodies (1:300, goat anti-rabbit IgG-Cy3; or 1:300, goat anti-mouse IgG-Cy3) for 60 minutes at room temperature and then washed three times with phosphate-buffered saline (PBS). The primary antibodies for respective cells include OCT4 (1:200, Abcam), NANOG (1:200, Abcam), PAX6 (1:200, Abcam), SOX2 (1:200, Abcam), NESTIN (1:200, Abcam), SOX1 (1:200, Abcam), ZO-1 (1:100, Abcam), N-CAD (1:100, Abcam), MAFB (1:300, Sigma), SOX9 (1:300, EMD Millipore), PRDM1 (1:200, Cell Signaling Technology), and NR2F1 (1:300,

R&D Systems). DAPI (4', 6-diamidino-2-phenylindole) (1:500) was used as counter-staining for nuclei. The images were captured and analyzed with the Olympus IX73 and Image J.

Single-cell RNA sequencing

Cells at indicated time points were collected for single-cell RNA-seq and global transcriptome analysis. TrypLE Express Enzyme (Gibco, 12604-021) was applied for single-cell dissociation. Single-cell RNA-seq library construction was conducted according to an automated pipeline called microwell full-length mRNA amplification and library construction system (MIRALCS) as described previously [78]. The 50-bp single-end sequencing was performed using the BGISEQ-500 platform.

ATAC-seq

We profiled chromatin accessibility of the neural differentiation process for five stages, including iPSCs, EB, Ros-E, Ros-L, and NPCs samples. ATAC-seq libraries were prepared using a modified protocol based on a previous study [79]. Briefly, 50,000 cells were collected for each sample, washed with pre-cooling PBS, and resuspended in 50 μ L of ice-cold lysis buffer (10 mM Tris-HCl, pH 7.5, 10 mM NaCl, 3 mM MgCl₂, 0.1% IGEPAL CA-630). Permeabilized cells were resuspended in 50 μ L transposase reaction buffer (1 \times TAG buffer, 2.0 μ L Tn5 transposase enzyme) and incubated for 30 minutes at 37°C. Polymerase chain reaction amplification and size selection (150-500 bp) were performed using Agencourt AMPure XP (Beckman Coulter) and Bioanalyzer 2100 (Agilent). Libraries were pooled at equimolar ratios with barcodes and sequenced on the BGISEQ-500 platform.

Pre-processing and quality control of single-cell RNA-seq

The original FASTQ data of the 527 samples were aligned to the rRNA database (downloaded from the National Center for Biotechnology Information [NCBI]) to remove rRNAs; the remaining reads were processed with SOAPnuke (version 1.5.3) [80] to trim adaptors and filter out the low-quality reads. The filtered data were aligned to the reference genome (hg19) using hisat2 (HISAT2 version 2.0.1-beta) [81]. Reads were counted using the R package GenomicAlignments [82] (mode = "Union," inter.feature = FALSE) and normalized to reads per kilobase million (RPKM) with edgeR [83]. Cells were filtered using the following parameters: genome mapping rate was more than 70%, fraction of reads mapped to mitochondrial genes was less than 20%, mRNA mapping rate was more than 80%, ERCC ratio was less than 10%, and gene number was more than 5,000. Further, the correlation of ERCC among cells was used to evaluate the quality of each cell (threshold = 0.9). At last, 445 single cells remained for further analysis in this study.

Identification of differentially expressed genes

Differential expression of genes in iPSCs (n = 71 cells), EB (n = 57 cells), Ros-E (n = 81 cells), Ros-L (n = 92 cells), and NPCs (n = 90 cells) was determined using SCDE analysis [84] with default parameters except for requiring a minimum of 100 genes (parameter min.lib.size = 100 to call scde.error.models function). The Z scores and corrected Z scores (cZ) to adjust for the multiple testing were converted into 2-tailed P values and adjusted to control for false discovery rate (FDR) using the pnorm function in R. The significantly differentially expressed genes were

selected based on the following criteria: adjusted P value < 0.01 and fold change > 2.

Constructing trajectory using differentially expressed genes

Monocle [85] ordering was conducted for all iPSCs, EB, Ros-E, Ros-L, and NPCs cells using the set of variable genes with default parameters, except we specified reduction.method = "DDRTree" in the reduceDimension function. The variable genes were selected using the Seurat R package [86].

Analysis of heterogeneity in each cell stage

The heterogeneity of each cell stage was determined using the Seurat R package [86] by the normalized expression level of reported transcription factors (retrieved from AnimalTFDB 2.0) [87]. Briefly, principle components (PCs) with a P value less than 0.01 were used for cell clustering with reduction.type = "pca" and resolution = "1.0." The FindAllMarkers function of the Seurat package was used to identify marker genes for each cluster using default parameters.

ATAC peak calling

We aligned ATAC-seq data to hg19 using Bowtie2 [88] and called peaks using MACS2 [89]. We established a standard peak set by merging all overlapping peaks. The IDR (irreproducible discovery rate) pipeline [90] was used to identify reproducible peaks between two biological replicates. Only peaks with IDR \leq 0.05 were considered reproducible and retained for downstream analysis. Pearson correlation coefficients of two biological replicates at each stage were calculated. Stage-specific peaks were defined as peaks having no overlap with any peaks in other stages. Novel peaks were defined as peaks nonoverlapping with previous stages. In the case of iPSCs, all peaks were annotated as novel peaks.

Targets assignment of ATAC peaks

For reproducible peaks, we applied HOMER [91] to assign putative targets for peaks. For stage-specific peaks, ChIPseeker [92] was used for putative target assignment. In both strategies, the putative target of a certain peak is defined as the gene with TSS closest to the peak summit location.

GO term and KEGG enrichment analysis

Lists of genes were analyzed using DAVID [93, 94], and the BH method was used for multiple test correction. GO terms with an FDR less than 0.01 or 0.05 were considered as significantly enriched. Target genes of stage-specific ATAC peaks were analyzed using the R package, clusterProfiler [95], in which an adjusted P value of 0.05 was used to identify significantly enriched GO and KEGG terms associated with each set of peaks.

Regulatory network construction

The scRNA-seq profiles among each cell type were compared using the SCDE package [84]. TFs significantly differentially expressed, with adjusted P value threshold of 0.05, among neighboring cell types were submitted to the STRING database [64] to infer regulatory networks based on known interaction relationships (supported by data from curated databases, experiments, and text-mining). TFs without any interactions with other pro-

teins were removed from the network. To select key regulators, we used a threshold of 5; all TFs with number of interactions above the threshold were considered as key regulators.

Putative targets prediction, GO term, and KEGG enrichment analysis

The target prediction and enrichment analyses were performed using the FIMO [96] and GREAT [97] packages, respectively. Briefly, the peak files in a certain stage were scanned for the presence or absence of TF motifs, which were downloaded from the Jasper database [98]. Genes with a TSS closest to TF motif-containing peaks were considered as putative targets of certain TFs.

Construction of cellular communication network

The ligand-receptor interaction relationships were downloaded from the database IUPHAR/BPS Guide to PHARMACOLOGY [99] and the Database of Ligand-Receptor Partners [65, 100]. The average expression level of transcript per million (TPM) of 1 was used as a threshold. Ligands and receptors above the threshold were considered as expressed in the corresponding cluster. An adjusted *P* value of 0.05 was used as a threshold to identify ligands/receptors specifically expressed in a subpopulation. The R package Circlize [101] was used to visualize the interactions.

Motif enrichment analysis

Motifs enriched in each set of ATAC peaks were identified using findMotifsGenome.pl from HOMER [91] using the following parameters: -size -100 100 -len 4,5,6,7,8,9,10,11,12.

Availability of supporting data

The detailed protocol of neural differentiation and bioinformatics pipeline are available in the protocols.io repository [102–103]. The sequencing raw data are deposited in the NCBI Sequence Read Archive under accession number SRP155759. Further supporting data can be found in the GigaScience database, GigaDB [104].

Additional files

Additional file 1: Figure S1. Quality control of ATAC-seq.
Additional file 2: Figure S2. Dynamics of gained and lost peaks during neural differentiation.
Additional file 3: Figure S3. Stage-specific features of cis-regulatory elements during neural differentiation.
Additional file 4: Figure S4. Quality control of scRNA-seq.
Additional file 5: Figure S5. Subgroups identification and key transcriptomic features within Fib stage.
Additional file 6: Figure S6. Subgroups identification and key transcriptomic features within EB stage.
Additional file 7: Figure S7. Subgroups identification and key transcriptomic features within Ros-L stage.
Additional file 8: Figure S8. Subgroups identification and key transcriptomic features within NPCs stage.
Additional file 9: Figure S9. Expression pattern of selected TFs within rosettes (Ros-E and Ros-L) stage.
Additional file 10: Figure S10. Differentially expressed TFs between Ros-E2 and Ros-L3.
Additional file 11: Figure S11. Differentially expressed TFs between Ros-L3 and NPC1.

Additional file 12: Figure S12. Key regulators during neural differentiation.

Additional file 13: Figure S13. GO term and KEGG enrichment analysis of selected TFs targets.

Additional file 14: Figure S14. Putative signaling between expressed receptors and their ligands in EB subsets.

Additional file 15: Figure S15. Putative signaling between expressed receptors and their ligands in Ros-E subsets.

Additional file 16: Figure S16. Putative signaling between expressed receptors and their ligands in NPC subsets.

Additional file 17: Figure S17. Transcription factor motifs enriched in stage specific peaks.

Additional file 18: Figure S18. Validation of neural differentiation in different genetic background cell lines.

Additional file 19: Table S1. Differentially expressed TFs among neighbouring cell subsets.

Additional file 20: Table S2. Putative targets of selected regulators.

Additional file 21: Table S3. Ligand-receptor interaction networks among subpopulations.

Additional file 22: Table S4. Differentially expressed receptors and ligands among Ros-L subpopulations.

Abbreviations

ATAC-seq: assay for transposase-accessible chromatin using sequencing; bFGF: basic fibroblast growth factor; BMP: bone morphogenetic protein; CNS: central nervous system; DPBS: Dulbecco's phosphate-buffered saline; EB: embryoid body; ERCC: External RNA Controls Consortium; ESCs: embryonic stem cells; FBS: fetal bovine serum; FDR: false discovery rate; FGF: fibroblast growth factor; Fib: fibroblast; Gln: glutamine; GO: Gene Ontology; HGNC: HUGO Gene Nomenclature Committee; hPSCs: human pluripotent stem cells; hiPSCs: human induced pluripotent stem cells; iPSCs: induced pluripotent stem cells; KEGG: Kyoto Encyclopedia of Genes and Genomes; NCBI: National Center for Biotechnology Information; NE: neuroepithelial; NPCs: neural progenitor cells; NTDs: neural tube defects; PBS: phosphate-buffered saline; QC: quality control; Ros-E: early rosette; Ros-L: late rosette; RPKM: reads per kilobase million; SCDE: single-cell differential expression; scRNA-seq: single-cell RNA sequencing; SHH: sonic hedgehog; t-SNE: t-distributed stochastic neighbor embedding; TGF β : transforming growth factors β ; TF: transcription factor; TFBS: transcription factor binding site; TPM: transcript per million; TSS: transcription start site.

Competing interests

The authors declare that they have no competing interests.

Funding

This work was supported by National Key R&D Program of China (2017YFA0104100) and Shenzhen Engineering Laboratory for Innovative Molecular Diagnostics (grant DRC-SZ [2016] 884) funded by Development and Reform Commission of Shenzhen Municipality; and Shenzhen Key Laboratory of Neurogenomics (CXB201108250094A) funded by Science, Technology and Innovation Commission of Shenzhen Municipality. D.C. is supported by China Postdoctoral Science Foundation (grant 2017M622795).

Author contributions

Z.S., Z.G., and X.X. conceived and designed the project. Z.S., D.C., Q.W., Shengpeng W., and Q.D. conducted the majority of experiments and data analysis. Shengpeng W., L.W., X.D., Shiyong W., and J.Z. performed computational analyses and prepared figures. C.L. participated in validation experiments and assisted with figure preparation for revision. D.Z., X.C., and F.C. contributed to sample collection. H.Y., X.X., Z.G., and Z.S. supervised the entire study. X.L. contributed to the design of the revision and jointly supervised the validation work. Z.S. wrote the manuscript with input from D.C., Q.W., L.L., J.L.F., Z.G., and X.X. D.C., L.L., J.L.F., S.Z., F.C., Z.G., and X.X. contributed to the discussion and revision of the manuscript. All authors read and approved the final manuscript.

Acknowledgements

We thank Tao Tan for support with antibody, Shiping Liu for bioinformatics help, Guibo Li for technical help with preparation of single-cell RNA-seq libraries, and other members of Cell and Developmental Biology Lab for discussions and support.

References

- Harmacek L, Watkins-Chow DE, Chen J, et al. A unique missense allele of BAF155, a core BAF chromatin remodeling complex protein, causes neural tube closure defects in mice. *Dev Neurobiol* 2014;**74**:483–97.
- Foster WH, Langenbacher A, Gao C, et al. Nuclear phosphatase PPM1G in cellular survival and neural development. *Dev Dyn* 2013;**242**:1101–9.
- Wilde JJ, Petersen JR, Niswander L. Genetic, epigenetic, and environmental contributions to neural tube closure. *Annu Rev Genet* 2014;**48**:583–611.
- Wen Z, Nguyen HN, Guo Z, et al. Synaptic dysregulation in a human iPSC model of mental disorders. *Nature* 2014;**515**:414–8.
- Tao Y, Zhang SC. Neural subtype specification from human pluripotent stem cells. *Cell Stem Cell* 2016;**19**:573–86.
- Streit A, Berliner AJ, Papanayotou C, et al. Initiation of neural induction by FGF signalling before gastrulation. *Nature* 2000;**406**:74–8.
- Muñoz-Sanjuán I, Brivanlou AH. Neural induction, the default model and embryonic stem cells. *Nat Rev Neurosci* 2002;**3**:271–80.
- Zhang SC, Wernig M, Duncan ID, et al. In vitro differentiation of transplantable neural precursors from human embryonic stem cells. *Nat Biotechnol* 2001;**19**:1129–33.
- Xu X, Hou Y, Yin X, et al. Single-cell exome sequencing reveals single-nucleotide mutation characteristics of a kidney tumor. *Cell* 2012;**148**:886–95.
- Hou Y, Song L, Zhu P, et al. Single-cell exome sequencing and monoclonal evolution of a JAK2-negative myeloproliferative neoplasm. *Cell* 2012;**148**:873–85.
- Johnson MB, Wang PP, Atabay KD, et al. Single-cell analysis reveals transcriptional heterogeneity of neural progenitors in human cortex. *Nat Neurosci* 2015;**18**:637–46.
- Villani AC, Satija R, Reynolds G, et al. Single-cell RNA-seq reveals new types of human blood dendritic cells, monocytes, and progenitors. *Science* 2017;**356**:eaah4573.
- La Manno G, Gyllborg D, Codeluppi S, et al. Molecular diversity of midbrain development in mouse, human, and stem cells. *Cell* 2016;**167**:566–80. e19.
- Buettner F, Natarajan KN, Casale FP, et al. Computational analysis of cell-to-cell heterogeneity in single-cell RNA-sequencing data reveals hidden subpopulations of cells. *Nat Biotechnol* 2015;**33**:155–60.
- Li H, Courtois ET, Sengupta D, et al. Reference component analysis of single-cell transcriptomes elucidates cellular heterogeneity in human colorectal tumors. *Nat Genet* 2017;**49**:708–18.
- Kim DS, Lee DR, Kim HS, et al. Highly pure and expandable PSA-NCAM-positive neural precursors from human ESC and iPSC-derived neural rosettes. *PLoS One* 2012;**7**:e39715.
- Ardhanareeswaran K, Mariani J, Coppola G, et al. Human induced pluripotent stem cells for modelling neurodevelopmental disorders. *Nat Rev Neurol* 2017;**13**:265–78.
- Chambers SMSM, Fasano CACA, Papapetrou EP, et al. Highly efficient neural conversion of human ES and iPSC cells by dual inhibition of SMAD signaling. *Nat Biotechnol* 2009;**27**:275–80.
- Dolmetsch R, Geschwind DH. The human brain in a dish: the promise of iPSC-derived neurons. *Cell* 2011;**145**:831–4.
- Kriks S, Shim JW, Piao J, et al. Dopamine neurons derived from human ES cells efficiently engraft in animal models of Parkinson's disease. *Nature* 2011;**480**:547–51.
- Miller JD, Ganat YM, Kishinevsky S, et al. Human iPSC-based modeling of late-onset disease via progerin-induced aging. *Cell Stem Cell* 2013;**13**:691–705.
- Maroof AM, Keros S, Tyson JA, et al. Directed differentiation and functional maturation of cortical interneurons from human embryonic stem cells. *Cell Stem Cell* 2013;**12**:559–72.
- Shi Y, Kirwan P, Smith J, et al. Human cerebral cortex development from pluripotent stem cells to functional excitatory synapses. *Nat Neurosci* 2012;**15**:477–86.
- Hueber SD, Bezdan D, Henz SR, et al. Comparative analysis of Hox downstream genes in Drosophila. *Development* 2007;**134**:381–92.
- Chen D, Jiang S, Ma X, et al. TFBSbank: a platform to dissect the big data of protein-DNA interaction in human and model species. *Nucleic Acids Res* 2017;**45**:D151–7.
- Thomas S, Li X-Y, Sabo PJ, et al. Dynamic reprogramming of chromatin accessibility during Drosophila embryo development. *Genome Biol BioMed Central* 2011;**12**:R43.
- Buenrostro JD, Wu B, Litzenburger UM, et al. Single-cell chromatin accessibility reveals principles of regulatory variation. *Nature* 2015;**523**:486–90.
- Jin F, Li Y, Dixon JR, et al. A high-resolution map of the three-dimensional chromatin interactome in human cells. *Nature* 2013;**503**:290–4.
- Corces MR, Buenrostro JD, Wu B, et al. Lineage-specific and single-cell chromatin accessibility charts human hematopoiesis and leukemia evolution. *Nat Genet* 2016;**48**:1193–203.
- Picelli S, Björklund ÅK, Faridani OR, et al. Smart-seq2 for sensitive full-length transcriptome profiling in single cells. *Nat Methods* 2013;**10**:1096–100.
- Kimura-Yoshida C, Mochida K, Ellwanger K, et al. Fate specification of neural plate border by canonical Wnt signaling and Grhl3 is crucial for neural tube closure. *EBioMedicine* 2015;**2**:513–27.
- Nikolopoulou E, Galea GL, Rolo A, et al. Neural tube closure: cellular, molecular and biomechanical mechanisms. *Development* 2017;**144**:552–66.
- Zhang J, Hagopian-Donaldson S, Seedzija G, et al. Neural tube, skeletal and body wall defects in mice lacking tran-

- scription factor AP-2. *Nature* 1996;**381**:238–41.
34. Hayon Y, Dashevsky O, Shai E, et al. Platelet microparticles promote neural stem cell proliferation, survival and differentiation. *J Mol Neurosci* 2012;**47**:659–65.
 35. Lee HO, LeVorse JM, Shin MK. The endothelin receptor-B is required for the migration of neural crest-derived melanocyte and enteric neuron precursors. *Dev Biol* 2003;**259**:162–75.
 36. Li H, Horns F, Wu B, et al. Classifying *Drosophila* olfactory projection neuron subtypes by single-cell RNA sequencing. *Cell* 2017;**171**:1206–20. e22.
 37. Vierbuchen T, Ostermeier A, Pang ZP, et al. Direct conversion of fibroblasts to functional neurons by defined factors. *Nature* 2010;**25**:1035–41.
 38. Yoo AS, Sun AX, Li L, et al. MicroRNA-mediated conversion of human fibroblasts to neurons. *Nature* 2011;**476**:228–31.
 39. Kim D-Y, Hwang I, Muller FL, et al. Functional regulation of FoxO1 in neural stem cell differentiation. *Cell Death Differ* 2015;**22**:2034–45.
 40. Cameron DA, Pennimpede T, Petkovich M. Tulp3 is a critical repressor of mouse hedgehog signaling. *Dev Dyn* 2009;**238**:1140–9.
 41. Jin Z, Liu L, Bian W, et al. Different transcription factors regulate nestin gene expression during P19 cell neural differentiation and central nervous system development. *J Biol Chem* 2009;**284**:8160–73.
 42. Elkabetz Y, Panagiotakos G, Al Shamy G, et al. Human ES cell-derived neural rosettes reveal a functionally distinct early neural stem cell stage. *Genes Dev* 2008;**22**:152–65.
 43. Cheung M, Briscoe J. Neural crest development is regulated by the transcription factor Sox9. *Development* 2003;**130**:5681–93.
 44. Scott CE, Wynn SL, Sesay A, et al. SOX9 induces and maintains neural stem cells. *Nat Neurosci* 2010;**13**:1181–9.
 45. Betters E, Liu Y, Kjaeldgaard A, et al. Analysis of early human neural crest development. *Dev Biol* 2010;**344**:578–92.
 46. Wang C, Kam RKT, Shi W, et al. The proto-oncogene transcription factor Ets1 regulates neural crest development through histone deacetylase 1 to mediate output of bone morphogenetic protein signaling. *J Biol Chem* 2015;**290**:21925–38.
 47. Mulligan KA, Cheyette BNR. Wnt signaling in vertebrate neural development and function. *J Neuroimmune Pharmacol* 2012;**7**:774–87.
 48. Ille F, Sommer L. Wnt signaling: multiple functions in neural development. *Cell Mol Life Sci* 2005;**62**:1100–8.
 49. Wang J, Jenjaroenpun P, Bhinge A, et al. Single-cell gene expression analysis reveals regulators of distinct cell subpopulations among developing human neurons. *Genome Res* 2017;**27**:1783–94.
 50. Wang Y, Ristevski S, Harley VR. SOX13 exhibits a distinct spatial and temporal expression pattern during chondrogenesis, neurogenesis, and limb development. *J Histochem Cytochem* 2006;**54**:1327–33.
 51. Ji EH, Kim J. SoxD transcription factors: multifaceted players of neural development. *Int J Stem Cells Korean Society for Stem Cell Research* 2016;**9**:3–8.
 52. Koshida R, Oishi H, Hamada M, et al. MafB is required for development of the hindbrain choroid plexus. *Biochem Biophys Res Commun* 2017;**483**:288–93.
 53. Rogers CD, Phillips JL, Bronner ME. Elk3 is essential for the progression from progenitor to definitive neural crest cell. *Dev Biol* 2013;**374**:255–63.
 54. Noisa P, Lund C, Kanduri K, et al. Notch signaling regulates the differentiation of neural crest from human pluripotent stem cells. *J Cell Sci* 2014;**127**:2083–94.
 55. Zhang J, Zheng B, Zhou P-P, et al. Vascular calcification is coupled with phenotypic conversion of vascular smooth muscle cells through Klf5-mediated transactivation of the Runx2 promoter. *Biosci Rep* 2014;**34**:e00148.
 56. Dougherty M, Kamel G, Grimaldi M, et al. Distinct requirements for wnt9a and irf6 in extension and integration mechanisms during zebrafish palate morphogenesis. *Development* 2013;**140**:76–81.
 57. Hernandez-Lagunas L, Choi IF, Kaji T, et al. Zebrafish narrowminded disrupts the transcription factor prdm1 and is required for neural crest and sensory neuron specification. *Dev Biol* 2005;**278**:347–57.
 58. Simões-Costa M, Bronner ME. Insights into neural crest development and evolution from genomic analysis. *Genome Res* 2013;**23**:1069–80.
 59. Viales RR, Diotel N, Ferg M, et al. The helix-loop-helix protein Id1 controls stem cell proliferation during regenerative neurogenesis in the adult zebrafish telencephalon. *Stem Cells* 2015;**33**:892–903.
 60. Li L, Candelario KM, Thomas K, et al. Hypoxia inducible factor-1 (HIF-1) is required for neural stem cell maintenance and vascular stability in the adult mouse SVZ. *J Neurosci* 2014;**34**:16713–9.
 61. Ahmed M, Xu J, Xu P-X. EYA1 and SIX1 drive the neuronal developmental program in cooperation with the SWI/SNF chromatin-remodeling complex and SOX2 in the mammalian inner ear. *Development Company of Biologists* 2012;**139**:1965–77.
 62. Abe H, Okazawa M, Nakanishi S. The Etv1/Er81 transcription factor orchestrates activity-dependent gene regulation in the terminal maturation program of cerebellar granule cells. *Proc Natl Acad Sci U S A* 2011;**108**:12497–502.
 63. Dominguez MH, Ayoub AE, Rakic P. POU-III transcription factors (Brn1, Brn2, and Oct6) influence neurogenesis, molecular identity, and migratory destination of upper-layer cells of the cerebral cortex. *Cereb Cortex* 2013;**23**:2632–43.
 64. Szklarczyk D, Morris JH, Cook H, et al. The STRING database in 2017: quality-controlled protein-protein association networks, made broadly accessible. *Nucleic Acids Res* 2017;**45**:D362–8.
 65. Pavličev M, Wagner GP, Chavan AR, et al. Single-cell transcriptomics of the human placenta: inferring the cell communication network of the maternal-fetal interface. *Genome Res* 2017;**27**:349–61.
 66. Stern CD. Initial patterning of the central nervous system: how many organizers? *Nat Rev Neurosci* 2001;**2**:92–8.
 67. Smith JR, Vallier L, Lupo G, et al. Inhibition of activin/nodal signaling promotes specification of human embryonic stem cells into neuroectoderm. *Dev Biol* 2008;**313**:107–17.
 68. Schwarz M, Alvarez-Bolado G, Dressler G, et al. Pax2/5 and Pax6 subdivide the early neural tube into three domains. *Mech Dev* 1999;**82**:29–39.
 69. Brouns MR, de Castro SCP, Terwindt-Rouwenhorst EA, et al. Over-expression of Grhl2 causes spina bifida in the axial defects mutant mouse. *Hum Mol Genet* 2011;**20**:1536–46.
 70. Nikitina N, Tong L, Bronner ME. Ancestral network module regulating prdm1 expression in the lamprey neural plate border. *Dev Dyn* 2011;**240**:2265–71.
 71. Vincent SD, Dunn NR, Sciammas R, et al. The zinc finger transcriptional repressor Blimp1/Prdm1 is dispensable for early axis formation but is required for specifica-

- tion of primordial germ cells in the mouse. *Development* 2005;132:1315–25.
72. Timmer JR, Wang C, Niswander L. BMP signaling patterns the dorsal and intermediate neural tube via regulation of homeobox and helix-loop-helix transcription factors. *Development* 2002;129:2459–72.
 73. Burns CJ, Zhang J, Brown EC, et al. Investigation of Frizzled-5 during embryonic neural development in mouse. *Dev Dyn* 2008;237:1614–26.
 74. Miyake A, Nakayama Y, Konishi M, et al. Fgf19 regulated by Hh signaling is required for zebrafish forebrain development. *Dev Biol* 2005;288:259–75.
 75. Qu Q, Sun G, Murai K, et al. Wnt7a regulates multiple steps of neurogenesis. *Mol Cell Biol* 2013;33:2551–9.
 76. Hsu Y-C, Lee D-C, Chen S-L, et al. Brain-specific 1B promoter of FGF1 gene facilitates the isolation of neural stem/progenitor cells with self-renewal and multipotent capacities. *Dev Dyn* 2009;238:302–14.
 77. Okita K, Matsumura Y, Sato Y, et al. A more efficient method to generate integration-free human iPS cells. *Nat Methods* 2011;8:409–12.
 78. Wu L, Zhang X, Zhao Z, et al. Full-length single-cell RNA-seq applied to a viral human cancer: applications to HPV expression and splicing analysis in HeLa S3 cells. *GigaScience* 2015;4:51.
 79. Buenrostro JD, Giresi PG, Zaba LC, et al. Transposition of native chromatin for fast and sensitive epigenomic profiling of open chromatin, DNA-binding proteins and nucleosome position. *Nat Methods* 2013;10:1213–8.
 80. Chen Y, Chen Y, Shi C, et al. SOAPnuke: a MapReduce acceleration-supported software for integrated quality control and preprocessing of high-throughput sequencing data. *GigaScience* 2018;7:1–6.
 81. Kim D, Langmead B, Salzberg SL. HISAT: a fast spliced aligner with low memory requirements. *Nat Methods* 2015;12:357–60.
 82. Lawrence M, Huber W, Pagès H, et al. Software for computing and annotating genomic ranges. Pric A, editor. *PLoS Comput Biol* 2013;9:e1003118.
 83. Robinson MD, McCarthy DJ, Smyth GK. edgeR: a bioconductor package for differential expression analysis of digital gene expression data. *Bioinformatics* 2010;26:139–40.
 84. Kharchenko P V, Silberstein L, Scadden DT. Bayesian approach to single-cell differential expression analysis. *Nat Methods* 2014;11:740–2.
 85. Trapnell C, Cacchiarelli D, Grimsby J, et al. The dynamics and regulators of cell fate decisions are revealed by pseudotemporal ordering of single cells. *Nat Biotechnol* 2014;32:381–6.
 86. Satija R, Farrell JA, Gennert D, et al. Spatial reconstruction of single-cell gene expression data. *Nat Biotechnol* 2015;33:495–502.
 87. Zhang H-M, Liu T, Liu C-J, et al. AnimalTFDB 2.0: a resource for expression, prediction and functional study of animal transcription factors. *Nucleic Acids Res* 2015;43:D76–81.
 88. Langmead B, Salzberg SL. Fast gapped-read alignment with Bowtie 2. *Nat Methods* 2012;9:357–9.
 89. Zhang Y, Liu T, Meyer CA, et al. Model-based analysis of ChIP-Seq (MACS). *Genome Biol BioMed Central* 2008;9:R137.
 90. Li Q, Brown JB, Huang H, et al. Measuring reproducibility of high-throughput experiments. *Ann Appl Stat* 2011;5:1752–79.
 91. Heinz S, Benner C, Spann N, et al. Simple combinations of lineage-determining transcription factors prime cis-regulatory elements required for macrophage and B cell identities. *Mol Cell* 2010;38:576–89.
 92. Yu G, Wang L-G, He Q-Y. ChIPseeker: an R/Bioconductor package for ChIP peak annotation, comparison and visualization. *Bioinformatics* 2015;31:2382–3.
 93. Huang DW, Sherman BT, Lempicki RA. Systematic and integrative analysis of large gene lists using DAVID bioinformatics resources. *Nat Protoc* 2009;4:44–57.
 94. Huang DW, Sherman BT, Lempicki RA. Bioinformatics enrichment tools: paths toward the comprehensive functional analysis of large gene lists. *Nucleic Acids Res* 2009;37:1–13.
 95. Yu G, Wang L-G, Han Y, et al. clusterProfiler: an R package for comparing biological themes among gene clusters. *Omi A J Integr Biol* 2012;16:284–7.
 96. Grant CE, Bailey TL, Noble WS. FIMO: scanning for occurrences of a given motif. *Bioinformatics* 2011;27:1017–8.
 97. McLean CY, Bristor D, Hiller M, et al. GREAT improves functional interpretation of cis-regulatory regions. *Nat Biotechnol* 2010;28:495–501.
 98. Sandelin A, Alkema W, Engström P, et al. JASPAR: an open-access database for eukaryotic transcription factor binding profiles. *Nucleic Acids Res* 2004;32:D91–4.
 99. Harding SD, Sharman JL, Faccenda E et al. The IUPHAR/BPS Guide to Pharmacology in 2018: updates and expansion to encompass the new guide to Immunopharmacology. *Nucleic Acids Res* 2018;46:D1091–106.
 100. Salwinski L, Miller CS, Smith AJ, et al. The Database of Interacting Proteins: 2004 update. *Nucleic Acids Res* 2004;32:449D–451.
 101. Gu Z, Gu L, Eils R, et al. Circlize implements and enhances circular visualization in R. *Bioinformatics* 2014;30:2811–2.
 102. Shang Z, Chen D, Wang Q, et al. Neural Progenitor Cells Derived from Human Induced Pluripotent Stem Cells. [protocols.io](https://dx.doi.org/10.17504/protocols.io.ntrdem6). 2018. <https://dx.doi.org/10.17504/protocols.io.ntrdem6>.
 103. Shang Z, Chen D, Wang Q, et al. Bioinformatic Pipeline for Studying Transcriptome and Regulator Dynamics During Neural Differentiation. [protocols.io](https://dx.doi.org/10.17504/protocols.io.ntpdem6). 2018. [http://dx.doi.org/10.17504/protocols.io.ntpdem6](https://dx.doi.org/10.17504/protocols.io.ntpdem6).
 104. Shang Z, Chen D, Wang Q, et al. Supporting data for “Single-cell RNA-seq reveals dynamic transcriptome profiling in human early neural differentiation.” *GigaScience Database* 2018. <http://dx.doi.org/10.5524/100496>.