

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

<b>Manuscript Number:</b>	GIGA-D-18-00097R1	
<b>Full Title:</b>	Single-cell RNA-seq reveals dynamic transcriptome profiling in human early neural differentiation	
<b>Article Type:</b>	Research	
<b>Funding Information:</b>	Development and Reform Commission of Shenzhen Municipality (DRC-SZ [2016] 884)	Dr. Zhouchun Shang
	Science, Technology and Innovation Commission of Shenzhen Municipality (CXB201108250094A)	Mr. Liang Wu
	China Postdoctoral Science Foundation (2017M622795)	Dr. Dongsheng Chen
<b>Abstract:</b>	<p><b>Background:</b> 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.</p> <p><b>Results:</b> 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 towards commitment to the central nervous system (CNS) lineage at each stage during differentiation, we identified putative novel transcription factors in regulating neural differentiation. In 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 among embryoid body (EB) stage and each cell subpopulation possessed a distinctive spectrum of ligands and receptors associated with neural differentiation which could reflect the identity of each subpopulation.</p> <p><b>Conclusions:</b> 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.</p> <p><b>Key words:</b> single cell RNA-seq, ATAC-seq, neural differentiation, neural rosettes, neural tube, transcription factor, iPSCs</p>	
<b>Corresponding Author:</b>	Xun Xu, Ph.D BGI Shenzhen, Guangdong CHINA	
<b>Corresponding Author Secondary Information:</b>		
<b>Corresponding Author's Institution:</b>	BGI	
<b>Corresponding Author's Secondary Institution:</b>		
<b>First Author:</b>	Zhouchun Shang	
<b>First Author Secondary Information:</b>		
<b>Order of Authors:</b>	Zhouchun Shang	

	Dongsheng Chen
	Quanlei Wang
	Shengpeng Wang
	Qiuting Deng
	Liang Wu
	Chuanyu Liu
	Xiangning Ding
	Shiyu Wang
	Jixing Zhong
	Doudou Zhang
	Xiaodong Cai
	Shida Zhu
	Huanming Yang
	Longqi Liu
	J. Lynn Fink
	Fang Chen
	Xiaoqing Liu
	Zhengliang Gao
	Xun Xu, Ph.D
<b>Order of Authors Secondary Information:</b>	
<b>Response to Reviewers:</b>	<p>Reviewer reports:</p> <p>Reviewer #1: In this study, the authors characterized the single-cell transcriptional profiles as well as chromatin accessibility at several stages during in vitro neuronal differentiation of human iPSCs. Bioinformatics analyses distinguished different subpopulations at each stage and identified transcription factors regulating neural differentiation. Overall, the methods and analyses performed are thoroughly explained, and the experimental results support the author's conclusions.</p> <p>1- The authors captured single cells of human iPSCs (n = 80), embryoid bodies (EBs; n = 81), early and late rosettes (Ros-E, Ros-L; n = 82 and 93, respectively), NPCs (n = 95), and the original somatic fibroblasts (n = 96). In the manuscript, however, the authors do not show any results of the analysis of the fibroblasts (other than the number of expressed genes in Figure S3d). It is stated that bulk ATAC-seq (for chromatin accessibility) was performed on all these stages, but no data for fibroblasts was shown. If the authors have performed single-cell RNA-seq on somatic fibroblasts as stated in the data description (line 114), this data should be presented in the main or supplementary figures.</p> <p>Reply: We agree with the reviewer and have now added heterogeneity study of fibroblasts in the revised version (Additional file 5: Figure S5). Considering this study was mainly focused on the regulation of neural differentiation starting from induced pluripotent stem cells (iPSCs), we thus did not include fibroblasts in the ATAC-seq analyses.</p> <p>2- Figure S2a and Figure S3b/d/e/f are missing y-axis labels. Although it is in the Figure legends, the authors should label the axes directly on the graphs.</p> <p>Reply: We thank the reviewer for pointing out the mistakes, we totally agree and have accordingly corrected them in the revised version (Additional file 3: Figure S3a and Additional file 4: Figure S4b, d, e, f).</p> <p>3- Manuscript lines 239-241 and Figure 1f: The authors state that "some important neural transcription factors exhibited heterogeneous expression within the same cell stage (Figure 1f)", but it is difficult to assess this from Figure 1f. I suggest that the</p>

authors show evidence for this statement with an updated heatmap or separate analysis that focuses only on these stage-specific genes.

Reply: We agree with the reviewer that the Fig. 1f is not clear enough, thus we modified Figure S4h (Additional file 4) to make it more clearly for showing heterogeneous expression of differentially expressed genes within the same cell stage.

4- Could the authors please double-check that Figure 2c displays the expression levels as  $\log_2(\text{RPKM}+1)$ , as stated in the legend? The values seem very low. Could it be the z-score instead of RPKM?

Reply: The expression level shown in Fig. 2c is  $\log_2(\text{RPKM} + 1)$ . All of these genes in Fig. 2c are transcription factors (TFs) coding genes that possess relatively low expression level, which is consistent with the previous publications, e.g., POU5F1, NANOG, SOX2 in Figure 7b; ZIC2, ZIC5 in Figure 7d; KLF4, PRDM14, DPPA2 in Figure 7e (Han et al., 2018); besides, some more published papers with similar gene expression level e.g., Oas2, Lsg20, Lkbke and Tspo in Figure 2 B, F, G, K, L, O, P (Friedman et al., 2018); TH and DCC in Figure 4A (Sousa et al., 2017); IL25 and Tslp in Figure 4c; IL33 in Extended Data Figure 7e; Retnlb, Wars, Pnlipp2 in Extended Data Figure 10d (Haber et al., 2017). Moreover, in our study, several TFs with relatively low expression levels in scRNA-seq data were validated by immunostaining and showed highly enriched at respective cell stage e.g., SOX9 and MAFB at Ros-E stage, SOX9 and PRDM1 at Ros-L stage, and NR2F1 and PRDM1 at NPCs stage (Fig. 4h; Additional file 18, Figure S18).

5- Figure 2d: It is unclear to me why there are two different  $-\log_{10}(\text{P-value})$  graphs for the overlapped GO terms of Ros-E1 and Ros-E2 in Figure 2d (in grey color). Same for the overlap in GO terms of Ros-L1 and Ros-L3 in Figure S5. If the authors perform GO enrichment on the list of genes overlapping between the two stages, there should be one set of GO enrichment results?

Reply: We identified enriched GO terms using up-regulated genes for each subpopulation respectively, and analyzed the relationship between the GO terms in different subpopulations within the same cell stage, so the Venn diagram showed the specific GO terms as well as the overlapped GO terms for the indicated subpopulation. Regarding subpopulation specific GO terms, there is only 1 P-value for each term showing their enrichment in corresponding subpopulation. In contrast, regarding those GO terms shared by Ros-E1 and Ros-E2, they are enriched in both subpopulations with different significances. Therefore, each shared GO term have 2 P-values. To display the P-values of each shared GO term, we present the 2 P-values (one from Ros-E1 and another one from Ros-E2) in Figure 2d.

6- Figure 5 c,d,e: The expression levels in  $\log_2(\text{RPKM}+1)$  for most of the genes indicated are very low (e.g., NMU:  $< 0.05$ ; EPHA7, ACKR3, C5, PTPRZ1, ANGPT2:  $< 0.5$ ). Some of the gene expression levels in other figures (e.g., Figure S4 and Figure S8) are also quite low. Can the authors please explain what  $\log_2(\text{RPKM}+1)$  threshold was used for gene detection and/or the filtering out of non-expressed genes?

Reply: The average expression level of RPKM of 1 was used as a threshold. Ligands and receptors above the threshold were considered as expressed in the corresponding subpopulation in figure 5a and 5b. After checking the expression profiles, we figured out the figure was mislabeled in figure 5c, d, e. Specifically, the y axis should be labeled as  $\log_{10}(\text{RPKM}+1)$  instead of  $\log_2(\text{RPKM}+1)$ . To be consistent with other figures, we have thus regenerated the figure and gene expression level visualized in the form of  $\log_2(\text{RPKM}+1)$ . In terms of the boxplot, we only kept ligands/receptors with average  $\text{RPKM} \geq 1$  in one subpopulation and average  $\text{RPKM} < 1$  in other subpopulations. As a result, WNT5A and EPHB6 from Ros-L1, FZD5 and LPAR4 from Ros-L2, ANGPT2 and PGF from Ros-L3 were visualized in figure 5c, d, e, respectively. Regarding Figure S4 and Figure S8 (correspondingly changed to Figure S6 and Figure S10 in the revised version), as mentioned in the reply to question 4, we did not apply a threshold for TFs as they could potentially function at relatively low expression level.

Minor edits:

- Manuscript lines 196-198: incomplete sentence ("Single cells using Smart-Seq2 method [30], followed by sequencing around 6 million reads per cell.)

Reply: Thanks a lot for pointing out this incomplete sentence, it has been updated into "Single cell RNA-seq libraries were generated using Smart-Seq2 method [30], followed

by sequencing around 6 million reads per cell."

- I suggest that the authors re-order Figure S4 for better readability (Figure S4d first, then Figure S4c and finally Figure S4a,b), since the manuscript discusses results from these figures in this particular order.

Reply: Sorry for the confusion. As suggested, we have re-ordered Figure S4 (changed to Figure S6 in the revised version) and cited corresponding figures in the text.

- Figure 4a,b: For readability, I suggest that the authors provide graph titles for the gene networks displayed, to make clear that one is looking at transcription factors differentially expressed between Ros-E2 and Ros-L3 (in Fig. 4a) and between Ros-L3 and NPC-1 (in Fig. 4b).

Reply: Thanks very much for your kind suggestion. We have added graph titles for the gene networks in Fig. 4a, Fig. 4b and Figure S12 following your advice.

- Line 443: redundant et al. - there is a missing reference here, or a reference has been removed.

Reply: Thanks very much for the kind suggestion. We have made the correction in the text.

Reviewer #2: The present manuscript by Shang et al. performed parallel single cell transcriptome and bulk chromatin landscape profiling during six consecutive stages during neural differentiation of fibroblast-based iPSCs. Based on a comprehensive amount of data and detailed analysis, this well-written and well-visualized paper provides many specific and novel insights into gene expression changes specific to neural differentiation. While I feel that this manuscript is a very interesting read as well as a valuable resource for the field, functional validation of cellular heterogeneity, as well as of the putative novel hub TFs would substantially improve the current study. Below please find my specific points that I feel should be addressed prior to publication.

Major points:

#### 1. Heterogeneity

Definition of the subclasses at the different stages is done based on single-cell RNAseq. Based on the presented data, it is not clear if the observed heterogeneity...

i) ...stems from the fact that the harvested cells at each stage contain lagging and leading cells (slow and fast differentiating cells).

ii) ... represent temporal transcriptional states (e.g. during cell cycle or circadian rhythm).

iii) ... represent distinct cellular subpopulation that occur in parallel, but that have different fates (e.g. dorsal, ventral, neural crest, ...).

While the Monocle analysis appears to suggest iii), the data is not able to convincingly draw a conclusion in this regard. I thus suggest that the authors should mention this as a caveat more prominently. Alternatively, single-cell ATACseq might help to gain insights if the observed subclasses are also distinguishable epigenetically, but the effort seems immense.

Reply: Thanks to the reviewer for the constructive suggestions. The reviewer has raised a very important point that the current scRNA-seq method by its nature only provides a snapshot of the gene expression profile for individual cells. We have added a discussion part regarding the concerns on the heterogeneity study in the manuscript. 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; 1) temporal transcriptional states during transient differentiation process; 2) 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 and which 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 : Figure S7) were located in the reconstructed trajectory prior to the generation of Ros-E populations. In addition, we carried out the cell fate commitment analysis using Branch1, Branch2 and Branch3 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.

## 2. Validation of heterogeneity

For EBs, Ros-E and Ros-L, the authors find strikingly different subclasses of cells. Based on a few selected 'novel' markers, it would be very interesting to see if this heterogeneity can be confirmed by immunostaining with reasonable effort.

Reply: We agree with the reviewer and we have validated several subset-specific markers experimentally (Fig. 4h; Additional file 18: Figure S18). Briefly, Ros-E (SOX9 and MAFB), Ros-L (SOX9 and PRDM1) and NPCs stage (NR2F1 and PRDM1) were validated by immunostaining, as expected, these TFs showed heterogeneous expression level within the same cell stage, moreover, we also validated these TFs in other ESCs and iPSCs e.g., H1\_ESCs, H7\_ESCs, H9\_ESCs and iPS25, the results were in line with that in iPS129.

## 3. Fibroblasts

The study analyzed fibroblasts as well, but in most of the presented the analyses they are not included. It would be interesting to see differences relative to fibroblasts also in the Monocle and ATACseq analysis. Also, why are some fibroblasts clustering with EBs in t-SNE (Fig 2a), and how would a PCA of these data look like?

Reply: We agree with the reviewer and we have added heterogeneity study of fibroblasts (Additional file 5: Figure S5), by applying the same subsets identification method. We identified two subsets of fibroblasts, Fib1 and Fib2, and the results showed significantly higher expression of several important pluripotency- and neural-associated transcription factors e.g., SOX2, LIN28, SOX11, ZIC2, FEZF1 and SIX3 in Fib2 (Additional file 5: Figure S5b). We further analyzed the relationship between fibroblast subsets and EBs and we observed that the majority of cells in Fib2 were clustered together with EB cells (Additional file 5: Figure S5e). The same input as Fig. 2a in PCA was shown as below, which recapitulates the cell stage distribution in Fig. 2a. Together with the molecular features of Fib2 subset (Additional file 5: Figure S5b), we proposed that the Fib2 subset might possess high potential for iPSCs reprogramming and neural conversion. Because the neural differentiation started from iPSCs, and we are focusing on studying the regulation of neural differentiation process, we did not include fibroblasts in the trajectory and ATAC-seq analyses.

## 4. Universal validity / N of genetic background

As that the authors apparently only used one fibroblast line/genetic background for this study, the cogency of the study is limited. While I agree that repeating all experiments with a second line would be very time and money consuming, this caveat leaves the possibility for each finding to be an artifact of this one cell line/genetic background. The authors should try to at least validate some of their key findings/TFs on one another genetic background.

Reply: We understand the concerns raised by the reviewer. To address this, we have performed neural differentiation using ESC with the same protocol and captured bulk transcriptome profiles of the corresponding differentiating cell stages (ESCs, EB, Ros-E, Ros-L and NPCs). The observations in ESCs recapitulated those seen in iPSCs, e.g., 1) PCA analysis; 2) with a high Pearson correlation coefficient between the corresponding cell stage derived from iPSCs and ESCs; 3) validation analysis of subset- specific markers (MAFB, SOX9, PRDM1 and NR2F1) as well as novel neural TF (PRDM1) expression in different genetic cell lines (H1\_ESCs, H7\_ESCs, H9\_ESCs, iPS25 and iPS129) showing consistent with the above heterogeneity study (Fig.4h; Additional file 18: Figure S18).

Regarding the single-cell level, we understand this limitation and are happy to describe

it in the discussion because we agree that repeating the entire study with a second line would indeed be very time and money consuming. However, the results inferred from this single line are largely consistent with previously reported findings from a variety of bulk cell-based studies so we feel that the possibility of our transcriptome profiling results being artefactual is extremely low. Our novel findings comprise new information garnered from high-resolution single-cell sequencing, not from experiments that suggest a complete revision of our fundamental understanding of neural differentiation.

#### 5. Functional validation of new TFs

Based on the identification of 'novel' TFs involved in neural differentiation, it would be interesting if overexpression or knockdown of these factors boost/impair neuronal differentiation of iPSCs.

Reply: We have added validation of subset- specific markers as well as novel markers expression in different genetic cell lines (H1\_ESCs, H7\_ESCs, H9\_ESCs, iPS25 and iPS129). As expected, we observed that MAFB, SOX9, PRDM1, NR2F1 were enriched at respective cell stage across different genetic cell lines, and the immunostaining results were consistent with the heterogeneity study. However, additional experiments to validate our novel findings are more appropriate to follow-up studies that can investigate stage-specific regulatory dynamics in more depth.

#### 6. Chromatin closing

For the ATACseq data, in addition to reporting the % novel peaks for each stage, I would be curious to know the % change in peak diversity between each time, because this would take into account both regions of the genome opening up, and regions of the genome closing up, instead of just opening.

Reply: We agree with the reviewer that only analyzing novel peaks might not reflect all the entire chromatin landscape during differentiation stage transitions. As suggested, we included the analysis of gained and lost peaks at each stage and added annotations, especially on the dynamics of lost peak regions (Additional file 2: Figure S2). Briefly, to reveal the detail of chromatin accessibility dynamics during neural differentiation, we 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: Figure 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: Figure S2b). This observation indicates that distal and promoter regions are more dynamic compared to other genomic regions during neural differentiation process. To gain insight into the potential function of closing (lost) peaks dynamics, we carried out GO enrichment analysis on the genes annotated by lost peaks at each stage. The GO terms analysis showed that "mesoderm morphogenesis", "endoderm development", "gastrulation" and "nodal signalling pathway" were solely enriched at 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 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, e and Additional file 2: Figure S2d, e).

#### 7. Neurons

It would be interesting to see differentiated neurons included in this already very interesting paper!

Reply: Yes, we agree that it would be useful if neurons are included in our study, however, we are afraid that this is probably beyond the scope of this study. This paper mainly focused on early neural differentiation process for following reasons: firstly, early neural development regulatory mechanism remains elusive due to the limited accessibility of human abortive fetuses at such an early stage (week 3 and 4 of human gestation); secondly, another recently published paper has already investigated the differentiation process from neural progenitor cells to neurons (Wang et al., 2017).

	<p>Reference</p> <p>[1] Han X, Chen H, Huang D, Chen H, Fei L, Cheng C, et al. Mapping human pluripotent stem cell differentiation pathways using high throughput single-cell RNA-sequencing. <i>Genome Biol.</i> 2018;19:47.</p> <p>[2] Friedman BA, Srinivasan K, Ayalon G, Meilandt WJ, Lin H, Huntley MA, et al. Diverse Brain Myeloid Expression Profiles Reveal Distinct Microglial Activation States and Aspects of Alzheimer's Disease Not Evident in Mouse Models. <i>Cell Rep.</i> 2018;22:832-847.</p> <p>[3] Sousa AMM, Zhu Y, Raghanti MA, Kitchen RR, Onorati M, Tebbenkamp ATN, et al. Molecular and cellular reorganization of neural circuits in the human lineage. <i>Science.</i> 2017;358:1027-1032.</p> <p>[4] Haber AL, Biton M, Rogel N, Herbst RH, Shekhar K, Smillie C, et al. A single-cell survey of the small intestinal epithelium. <i>Nature.</i> 2017;551:333-339.</p> <p>[5] Wang J, Jenjaroenpun P, Bhinge A, Angarica VE, Del Sol A, Nookaew I, et al. Single-cell gene expression analysis reveals regulators of distinct cell subpopulations among developing human neurons. <i>Genome Res.</i> 2017;27:1783-1794.</p>
<b>Additional Information:</b>	
<b>Question</b>	<b>Response</b>
Are you submitting this manuscript to a special series or article collection?	No
<p><b>Experimental design and statistics</b></p> <p>Full details of the experimental design and statistical methods used should be given in the Methods section, as detailed in our <a href="#">Minimum Standards Reporting Checklist</a>. Information essential to interpreting the data presented should be made available in the figure legends.</p> <p>Have you included all the information requested in your manuscript?</p>	Yes
<p><b>Resources</b></p> <p>A description of all resources used, including antibodies, cell lines, animals and software tools, with enough information to allow them to be uniquely identified, should be included in the Methods section. Authors are strongly encouraged to cite <a href="#">Research Resource Identifiers</a> (RRIDs) for antibodies, model organisms and tools, where possible.</p> <p>Have you included the information requested as detailed in our <a href="#">Minimum Standards Reporting Checklist</a>?</p>	Yes

<p><b>Availability of data and materials</b></p> <p>All datasets and code on which the conclusions of the paper rely must be either included in your submission or deposited in <a href="#">publicly available repositories</a> (where available and ethically appropriate), referencing such data using a unique identifier in the references and in the “Availability of Data and Materials” section of your manuscript.</p> <p>Have you have met the above requirement as detailed in our <a href="#">Minimum Standards Reporting Checklist</a>?</p>	<p>Yes</p>
---	------------



[Click here to view linked References](#)

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

3 Zhouchun Shang<sup>1,2,3,4#</sup>, Dongsheng Chen<sup>2,3#</sup>, Quanlei Wang<sup>2,3,4,6#</sup>, Shengpeng  
4 Wang<sup>2,3</sup>, Qiuting Deng<sup>2,3</sup>, Liang Wu<sup>2,3,5,6</sup>, Chuanyu Liu<sup>2,3,6</sup>, Xiangning Ding<sup>2,3</sup>,  
5 Shiyou Wang<sup>2,3,6</sup>, Jixing Zhong<sup>2,3,6</sup>, Doudou Zhang<sup>7</sup>, Xiaodong Cai<sup>7</sup>, Shida  
6 Zhu<sup>2,3,4</sup>, Huanming Yang<sup>2,8</sup>, Longqi Liu<sup>2,3</sup>, J. Lynn Fink<sup>2,9</sup>, Fang Chen<sup>2,3,10</sup>,  
7 Xiaoqing Liu<sup>1</sup>, Zhengliang Gao<sup>1\*</sup> and Xun Xu<sup>2,3\*</sup>

8 1 Shanghai Tenth People's Hospital, Tongji University School of Medicine,  
9 Shanghai, China

10 2 BGI-Shenzhen, Shenzhen, China

11 3 China National GeneBank, BGI-Shenzhen, Shenzhen, China

12 4 Shenzhen Engineering Laboratory for Innovative Molecular Diagnostics,  
13 BGI-Shenzhen, Shenzhen, China

14 5 Shenzhen Key Laboratory of Neurogenomics, BGI-Shenzhen, Shenzhen,  
15 China

16 6 BGI Education Center, University of Chinese Academy of Sciences,  
17 Shenzhen, China

18 7 Department of Neurosurgery, Shenzhen Second People's Hospital,  
19 Shenzhen University 1st Affiliated Hospital, Shenzhen, Guangdong, China

20 8 James D. Watson Institute of Genome Sciences, Hangzhou, China

21 9 The University of Queensland, Diamantina Institute (UQDI), Brisbane, QLD,  
22 Australia

23 10 Laboratory of Genomics and Molecular Biomedicine, Department of Biology,  
24 University of Copenhagen, DK-2100, Copenhagen, Denmark

25 #These authors contributed equally to this work.

26 \*Correspondence should be addressed to Z.G.  
27 (zhengliang\_gao@tongji.edu.cn) or X.X. (xuxun@genomics.cn).

28 **Abstract:**

29 **Background:** Investigating cell fate decision and subpopulation specification  
30 in the context of the neural lineage is fundamental to understanding

1 31 neurogenesis and neurodegenerative diseases. The differentiation process of  
2 32 neural-tube-like rosettes *in vitro* is representative of neural tube structures,  
3 33 which are composed of radially organized, columnar epithelial cells and give  
4 34 rise to functional neural cells. However, the underlying regulatory network of  
5 35 cell fate commitment during early neural differentiation remains elusive.

6 36 **Results:** In this study, we investigated the genome-wide transcriptome profile  
7 37 of single cells from six consecutive reprogramming and neural differentiation  
8 38 time points and identified cellular subpopulations present at each  
9 39 differentiation stage. Based on the inferred reconstructed trajectory and the  
10 40 characteristics of subpopulations contributing the most towards commitment to  
11 41 the central nervous system (CNS) lineage at each stage during differentiation,  
12 42 we identified putative novel transcription factors in regulating neural  
13 43 differentiation. In addition, we dissected the dynamics of chromatin  
14 44 accessibility at the neural differentiation stages and revealed active  
15 45 *cis*-regulatory elements for transcription factors known to have a key role in  
16 46 neural differentiation as well as for those that we suggest are also involved.  
17 47 Further, communication network analysis demonstrated that cellular  
18 48 interactions most frequently occurred among embryoid body (EB) stage and  
19 49 each cell subpopulation possessed a distinctive spectrum of ligands and  
20 50 receptors associated with neural differentiation which could reflect the identity  
21 51 of each subpopulation.

22 52 **Conclusions:** Our study provides a comprehensive and integrative study of  
23 53 the transcriptomics and epigenetics of human early neural differentiation,  
24 54 which paves the way for a deeper understanding of the regulatory mechanisms  
25 55 driving the differentiation of the neural lineage.

26 56 **Key words:** single cell RNA-seq, ATAC-seq, neural differentiation, neural  
27 57 rosettes, neural tube, transcription factor, iPSCs

## 28 58 **Background**

29 60 The nervous system contains complex molecular circuitry in developmental

1 61 processes. In humans, there is a paucity of data describing early neural  
2 62 development and the corresponding cellular heterogeneity at various stages.  
3  
4 63 To our knowledge, neural tube formation and closure is crucial for embryonic  
5  
6 64 central nervous system (CNS) development and the process of neurulation.  
7  
8 65 Previous studies have reported that neural tube closure is strongly controlled  
9  
10 66 by both genetic and epigenetic factors and is sensitive to environmental  
11  
12 67 influences [1-3]. Perturbations in this delicately balanced and orchestrated  
13  
14 68 process can result in neural tube defects (NTDs) giving rise to birth defects  
15  
16 69 such as spina bifida, anencephaly and encephaloceles. However, the  
17  
18 70 formation and closure of the neural tube *in vivo* during week 3 and 4 of human  
19  
20 71 gestation is a transient event and is therefore difficult to capture. Moreover, the  
21  
22 72 limited accessibility of human abortive fetuses at such an early stage  
23  
24 73 precludes a thorough investigation of human early neural development.  
25  
26

27 74  
28  
29 75 Human pluripotent stem cells (hPSCs), including embryonic stem cells (ESCs)  
30  
31 76 and induced pluripotent stem cells (iPSCs), can be differentiated into all cell  
32  
33 77 types, including neural cells, offering a promising *in vitro* model for tracing  
34  
35 78 early cell lineages and studying the cell fate specification of human neural  
36  
37 79 differentiation [4, 5]. Previous studies have indicated that inhibition of bone  
38  
39 80 morphogenetic protein (BMP) signalling or activation of fibroblast growth factor  
40  
41 81 (FGF) signalling is needed for induction of the neuroectoderm from ESCs [6, 7].  
42  
43 82 A striking feature of differentiating stem cells *in vitro* is that they form neural  
44  
45 83 tube-like rosettes which are composed of radially organized columnar  
46  
47 84 epithelial cells that resemble the process of neurulation. The progenitor cells in  
48  
49 85 rosettes gradually give rise to functional cells (e.g., more restricted progenitors  
50  
51 86 and neuronal precursors, mimicking the process of neurulation and neural tube  
52  
53 87 growth) which represent neural tube structures [8]. These cellular processes  
54  
55 88 suggest that distinct cell fate decisions and lineage commitments occur during  
56  
57 89 rosette formation. However, the corresponding underlying mechanisms of the  
58  
59 90 regulation of cell fate commitment during early neural differentiation remain  
60  
61  
62  
63  
64  
65

1 91 largely unknown.

2 92

3 93 The advance of single cell trans-omics technology has offered incisive tools for  
4 94 revealing heterogeneous cellular contexts and developmental processes [9-11].  
5 95 Single cell RNA-seq (scRNA-seq) has been applied to the study of cellular  
6 96 heterogeneity as well as to the identification of novel subtypes or intermediate  
7 97 cell groups in multiple contexts [12-15], and may help delineate unexpected  
8 98 features of neural developmental biology and facilitate the study of cellular  
9 99 states and neurogenesis processes. In the present study, we used scRNA-seq  
10 100 and ATAC-seq (assay for transposase-accessible chromatin using sequencing)  
11 101 to investigate human early neural differentiation. Our analysis reveals the  
12 102 landscape of the transcriptome and *cis*-regulatory elements during this  
13 103 process and creates an unbiased classification of cell subpopulations during  
14 104 differentiation, providing a comprehensive description of transcriptomic and  
15 105 epigenetic patterns in cell fate decision. The differentiation system of hiPSCs  
16 106 provides access to the very early stage of neural development and may serve  
17 107 as a source of specialized cells for regenerative medicine as well as  
18 108 supporting further investigations of neural tube defects.

19 109

## 20 110 **Data description**

21 111 Here, we applied a well-adopted neural induction protocol and generated  
22 112 neural progenitor cells (NPCs) by forming neural rosettes *in vitro* [8, 16]. We  
23 113 analysed several different differentiation stages of cells, including hiPSCs,  
24 114 embryoid body (EB), early rosettes (hereafter termed Ros-E, post-3 days of  
25 115 rosettes formation), late rosettes (hereafter termed Ros-L, post-5 days of  
26 116 rosettes formation), NPCs, and the original somatic fibroblasts (Fib).  
27 117 scRNA-seq was performed at discrete time points (e.g., Fib, iPSCs, EB, Ros-E,  
28 118 Ros-L and NPCs), and we captured 96, 80, 81, 82, 93, and 95 single cells,  
29 119 respectively, for each stage with the purpose of studying differentiation  
30 120 transition events. The quality of sequencing data was evaluated and filtered by

1 121 a quality control (QC) pipeline developed in-house (see Methods for details). In  
2 122 addition, bulk ATAC-seq with two biological replicates was applied to the cell  
3  
4 123 stages iPSCs, EB, Ros-E, Ros-L and NPCs to measure the regulome  
5  
6 124 dynamics during neural differentiation (Fig. 1a).  
7  
8  
9 125

## 10 126 **Analyses**

### 11 127 **Differential transcriptome and regulome dynamics throughout human** 12 128 **early neural differentiation**

13 129 Since the development of human ESCs and iPSCs, the ability to investigate  
14 130 human neurogenesis and neurological diseases via an *in vitro* differentiation  
15 131 model has vastly improved [4, 17]. Subsequently, artificial neural cells have  
16 132 been successfully generated using a variety of protocols by several  
17 133 laboratories [18-23]. Here, we followed a well-adopted neural induction  
18 134 protocol and generated NPCs by forming neural rosettes via inhibition of TGF $\beta$ ,  
19 135 AMPK and BMP signalling pathways and activation of the FGF signalling  
20 136 pathway [8, 16]. We analysed different differentiation stages of the cells  
21 137 including iPSCs, EB, Ros-E, Ros-L, and NPCs as well as the original somatic  
22 138 fibroblasts (Fib). The iPSC aggregates were induced to neuroepithelial cells  
23 139 (NE) and followed by neural tube-like rosettes formation (Fig. 1b). Firstly,  
24 140 pluripotency-associated transcription factors (TFs) (e.g., OCT4, NANOG) were  
25 141 significantly expressed in hiPSCs, suggesting that these cells did exhibit a  
26 142 stem cell phenotype. The subsequent formation of neural rosettes was  
27 143 confirmed by morphology, apical localization of ZO-1, a tight junction protein,  
28 144 and co-localisation of the neuroepithelial marker N-CADHERIN (N-CAD, also  
29 145 known as CDH2) at the junctions. Additional neural markers such as PAX6,  
30 146 NESTIN, SOX2, and SOX1 were also found to be highly enriched in the rosette  
31 147 stage (Fig. 1b).  
32  
33  
34  
35  
36  
37  
38  
39  
40  
41  
42  
43  
44  
45  
46  
47  
48  
49  
50  
51  
52  
53  
54  
55  
56  
57  
58  
59  
60  
61  
62  
63  
64  
65

149 Cell stages are usually determined by a complement of TFs or master  
150 regulators which regulate hundreds of genes associated with various cellular

1 151 functions. To study the genomic features associated with open chromatin  
2 152 regions, we classified ATAC peaks based on the location of the peak centre.  
3  
4 153 More than 16,000 peaks were identified for each cell stage (Additional file 1:  
5  
6 154 Figure S1a) with the majority located in introns and enhancers/promoters,  
7  
8 155 genomic regions that are known to harbour a variety of *cis*-regulatory elements  
9  
10 156 and are subjected to regulation by TFs (Additional file 1: Figure S1b).  
11  
12 157 Furthermore, we observed that ATAC peaks were significantly enriched at  
13  
14 158 regions near transcription start sites (TSS) (Additional file 1: Figure S1c).  
15  
16 159 These observations were reproducible across two replicates with a very high  
17  
18 160 Pearson correlation coefficient ( $\geq 0.954$ ) (Additional file 1: Figure S1d, e).  
19  
20  
21 161

22  
23 162 It is widely reported that chromatin structures undergo widespread  
24  
25 163 reprogramming during cell status transition, with some genomic regions  
26  
27 164 becoming compacted or opened, leading to the switching on or off of a  
28  
29 165 repertoire of genes responsible for cell fate decision [24-29]. We studied the  
30  
31 166 dynamic chromatin landscape by tracing the temporal origins of ATAC peaks at  
32  
33 167 each stage with peaks non-overlapping with existing ones that were annotated  
34  
35 168 as novel peaks. We assumed that those peaks, conserved among  
36  
37 169 differentiation stages, are associated with housekeeping genes while  
38  
39 170 stage-dynamic peaks are likely to represent *cis*-regulatory elements important  
40  
41 171 for cell status transition. As expected, we observed the introduction of roughly  
42  
43 172 10-50% of novel peaks in each stage, accompanied by the disappearance of  
44  
45 173 several pre-existing ATAC peaks. Notably, more novel peaks appeared at the  
46  
47 174 NPCs stage than at other stage (Fig. 1c). GO term analysis of genes residing  
48  
49 175 in novel peaks across the differentiation stages showed enrichment of “axon  
50  
51 176 development”, “positive regulation of nervous system development”, “epithelial  
52  
53 177 tube morphogenesis”, “positive regulation of neurogenesis”, “cell-cell signalling  
54  
55 178 by Wnt”, “forebrain development”, “hindbrain development”, “telencephalon  
56  
57 179 development”, “neural precursor cell proliferation”, and “cell fate commitment”.  
58  
59  
60 180 “Neurotrophin signalling pathway” was also found to be enriched, but was  
61  
62  
63  
64  
65

1 181 specifically associated with NPCs. KEGG enrichment analysis showed that  
2 182 “FoxO signalling pathway”, a pathway which is known to play an important role  
3  
4 183 in NPC proliferation, and “neuroactive ligand–receptor interaction” were  
5  
6 184 enriched in NPCs stage (Fig. 1d, e), suggesting that specific *cis*-regulatory  
7  
8 185 elements regulating neural differentiation are being staged (poised) for stem  
9  
10 186 cell fate specification and conversion.

11  
12 187 To reveal the detail of chromatin accessibility dynamics during neural  
13  
14 188 differentiation, we also analysed the gained or lost peaks at each stage  
15  
16 189 compared with the previously neighbouring one. We observed that the number  
17  
18 190 of gained peaks was with the largest increase at the NPCs stage while the  
19  
20 191 number of lost peaks was relatively high at Ros-E stage (Additional file 2:  
21  
22 192 Figure S2a). Next, we studied the genomic distribution of these dynamic peaks  
23  
24 193 and found that both the gained and lost peaks were located mostly in distal  
25  
26 194 intergenic regions and promoter regions (Additional file 2: Figure S2b). This  
27  
28 195 observation indicates that distal and promoter regions are more dynamic  
29  
30 196 compared to other genomic regions during neural differentiation process.

31  
32 197 To gain insight into the potential function of closing (lost) peaks dynamics, we  
33  
34 198 carried out GO enrichment analysis on the genes associated with lost peaks at  
35  
36 199 each stage. The GO terms analysis showed that “mesoderm morphogenesis”,  
37  
38 200 “endoderm development”, “gastrulation” and “nodal signalling pathway” were  
39  
40 201 solely enriched at EB stage, indicating that upstream, as well as other lineage  
41  
42 202 development, was relatively repressed by closing related *cis*-regulatory  
43  
44 203 regions. Other cell fate conversion terms such as “neural crest cell  
45  
46 204 differentiation”, “osteoclast differentiation”, and “regulation of cartilage  
47  
48 205 development” were enriched at Ros-E stage, together with the annotation  
49  
50 206 results of novel peaks, indicating that the chromatin accessibility prepared for  
51  
52 207 the neural lineage conversion by opening/closing up specific *cis*-regulatory  
53  
54 208 regions which facilitated the neural transition cascades (Fig. 1d, e and  
55  
56 209 Additional file 2: Figure S2d, e).  
57  
58  
59  
60  
61

62 210

211

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

220

221 To more thoroughly investigate the molecular mechanisms governing neural  
222 differentiation we profiled the transcriptomes of 527 single cells. Single cell  
223 RNA-seq libraries were generated using Smart-Seq2 method [30], followed by  
224 sequencing approximately 6 million reads per cell. For subsequent analysis,  
225 we focused on 445 cells that passed the quality control (QC, Methods,  
226 Additional file 4: Figure S4a, b) and ERCC correlation filter (Methods,  
227 Additional file 4: Figure S4c). 7003 to 8560 expressed genes were detected  
228 per cell (Additional file 4: Figure S4d), including TFs that were relatively highly  
229 expressed at the EB and NPCs stages, while, intriguingly, pseudogenes were  
230 relatively highly expressed at the Ros-E and NPCs stages (Additional file 4:  
231 Figure S4e). We also identified a variety of genes: 3524, 3855, 2023, 1804 and  
232 6211 specifically expressed at the iPSCs, EB, Ros-E, Ros-L and NPCs stages,  
233 respectively (Additional file 4: Figure S4f). Many of these stage-specific genes  
234 include some well-known pluripotent genes (*NANOG*, *ID1*, *ID2*, *ZFP42*,  
235 *LIN28A*, *DPPA4*); early neural markers (*SOX2*, *OTX2*, *OTX1*, *PAX6*); and  
236 genes that both regulate neural development and are critical to proliferative  
237 NPCs (*SOX4*, *SIX3*, *CDH2*, *ZIC2*) (Fig. 1f and Additional file 4: Figure S4h).

238

239 Because the neural rosette recapitulates neural tube development *in vitro*, we  
240 paid particular attention to the Ros-E and Ros-L stages. Unsurprisingly, a large



1 241 proportion of up-regulated genes in the Ros-E stage were associated with  
2 242 nervous system development including *TFAP2A*, *CNTN4*, *GLI3*, *DLX5* and  
3 243 *OTX1*) (Fig. 1f). Of particular interest is the gene *GRHL3*. Expression of this  
4 244 gene is associated with neural tube closure in mice [31, 32] and we observed  
5 245 this gene to be highly expressed at Ros-E in human cells, suggesting that its  
6 246 role in neural tube closure may be conserved across mammals or possibly  
7 247 chordates. *TFAP2A* (transcription factor AP-2 alpha) and *TFAP2B*  
8 248 (transcription factor AP-2 beta) have been proposed as master regulators of  
9 249 the neural crest cell and loss of function of transcription factor AP-2 in mice is  
10 250 strongly associated with a cranial neural tube defect phenotype [33]. In our  
11 251 system, *TFAP2B* and *TFAP2A* were relatively highly expressed at both the  
12 252 Ros-E and -L stages, suggesting transcription factor AP-2 may coordinate the  
13 253 specialized distal *cis*-regulatory elements for downstream regulations in  
14 254 human. We also observed expression of *ANLN* (Anillin actin binding protein) at  
15 255 the Ros-L stage, suggesting that neuronal migration and neurite growth might  
16 256 occur by the linking of RhoG to the actin cytoskeleton in neural rosettes [34].  
17 257 Similarly, our data showed that *AURKA* (aurora kinase A) and *AURKB* (aurora  
18 258 kinase B) were both expressed at the Ros-L stage, echoing previous findings  
19 259 that the aPKC–Aurora A–NDEL1 pathway plays an essential role in neurite  
20 260 elongation through modulating microtubule dynamics [35]. Finally, the neuron  
21 261 fate commitment protein, *TGFB2*, the nervous system development regulator,  
22 262 *ZEB2*, and the neural precursor cell proliferation-associated protein, *IFT20*,  
23 263 were enriched at NPCs stage (Fig. 1f).

24 264  
25 265 An unexpected finding was that some of the most important neural TFs  
26 266 exhibited heterogeneous expression within the same cell stage (e.g., *ZIC2*,  
27 267 *OTX2*, *HESX1*, *DLX3*, *LHX5*) (Fig. 1f and Additional file 4: Figure S4h). This  
28 268 inspired us to dissect the subpopulations of cells within each cell stage to  
29 269 better understand the significance of this result.

30 270

1  
2  
3  
4  
5  
6  
7  
8  
9  
10  
11  
12  
13  
14  
15  
16  
17  
18  
19  
20  
21  
22  
23  
24  
25  
26  
27  
28  
29  
30  
31  
32  
33  
34  
35  
36  
37  
38  
39  
40  
41  
42  
43  
44  
45  
46  
47  
48  
49  
50  
51  
52  
53  
54  
55  
56  
57  
58  
59  
60  
61  
62  
63  
64  
65

271 **Heterogeneous cellular subpopulations were identified at each**  
272 **developmental stage**

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

289  
290 **Fibroblasts (Fib) stage**

291 Fibroblasts (Fib) are a very well-adopted original somatic cell resource for  
292 iPSCs reprogramming; many direct conversions from fibroblast to functional  
293 neurons have been reported [37, 38]. Here, we dissected two subpopulations  
294 of human dermal fibroblasts (Fib1 and Fib2) with distinct molecular features,  
295 showing significantly higher expression of several important pluripotency- and  
296 neural-associated transcription factors such as *SOX2*, *LIN28*, *SOX11*, *ZIC2*,  
297 *FEZF1* and *SIX3* in Fib2 (Additional file 5: Figure S5a, b). GO terms identified  
298 by up-regulated genes between the two subsets showed “chromosome  
299 segregation”, “positive regulation of nervous system development”, “stem cell  
300 population maintenance”, “positive regulation of cell cycle”, “neural precursor

1 301 cell proliferation” and “chromatin remodeling” as solely enriched in the Fib2  
2 302 subpopulation (Additional file 5: Figure S5c). KEGG enrichment analysis  
3 303 showed “cell cycle” term was specifically associated with the Fib2 subset  
4 304 (Additional file 5: Figure S5d). Furthermore, we observed that fibroblasts were  
5 305 distributed into two distinct groups called Fib-Group1 and Fib-Group2 based  
6 306 on their location in the Fig. 2a. Of note, the majority of cells in Fib-Group1 and  
7 307 Fib-Groups2 were composed of Fib1 and Fib2, respectively. Moreover, cells  
8 308 from Fib2 subset clustered together with EB cells (Additional file 5: Figure S5e).  
9 309 Together with the molecular features of Fib2 subset (Additional file 5: Figure  
10 310 S5b), we proposed Fib2 subset might possess high potential for iPSCs  
11 311 reprogramming and neural conversion. Thus, based on the differentially  
12 312 expressed genes and CD markers dataset (HUGO Gene Nomenclature  
13 313 Committee, HGNC), we further inferred several cell surface markers of Fib2  
14 314 (e.g., *FGFR2*, *F11R*, *PROM1*, *BST2*, *ITGA6* and *EPCAM*) although these  
15 315 surface markers showed heterogeneously expressed levels within the Fib2  
16 316 subset (Additional file 5: Figure S5f).

17 317

### 18 318 **Embryoid body (EB) stage**

19 319 For the three EB subpopulations (EB1, EB2 and EB3), we identified genes that  
20 320 were up-regulated compared to the iPSCs stage, respectively. These genes  
21 321 were enriched in “fetal brain cortex”, “epithelium” and “brain” terms by DAVID  
22 322 using tissue enrichment analysis (Additional file 6: Figure S6a) which suggests  
23 323 that the biological processes of brain development and neural differentiation  
24 324 initiation are occurring during the iPSCs-to-EB stage transition and these  
25 325 processes are shared by each EB subpopulation. Moreover, most neural TFs  
26 326 and cell-specific markers were expressed commonly among EB  
27 327 subpopulations (e.g., *SOX2*, *ZIC2*, *SOX11*, *SOX4*, *SIX3*) (Additional file 6:  
28 328 Figure S6b) and some of these TFs play a crucial role in neural tube formation.  
29 329 However, some important neural TFs, such as *FOXO1* and *FOXO3*, which play  
30 330 an important role in NPC proliferation and self-renewal [39]; *TULP3*, which

1 331 regulates the SHH signalling pathway and modulates neural tube development  
2 332 [40]; and *POU2F1*, which regulates *NESTIN* gene expression during P19 cell  
3  
4 333 neural differentiation and CNS development [41], showed significantly high  
5  
6 334 expression in the EB3 subpopulation, but low expression in the EB1 and EB2  
7  
8 335 subpopulations (Additional file 6: Figure S6c, d). This suggests that different  
9  
10 336 subpopulations contain specific molecular signatures and different  
11  
12 337 differentiation states or potentials.  
13  
14  
15 338

### 16 339 **Early rosette (Ros-E) stage**

17 340 During the Ros-E stage, which is composed of NE and the cells in the early  
18  
19 341 stage of rosette formation, we observed expression of several master regulator  
20  
21 342 genes associated with neural tube formation and closure including *SOX11*,  
22  
23 343 *ZIC2*, *PAX3*, and *SNAI2* in both Ros-E subgroups (Ros-E1 and Ros-E2).  
24  
25 344 However, genes involved in neural crest specifiers, such as *TWIST1* [42] and  
26  
27 345 *SOX9*, which contribute to the induction and maintenance of neural stem cells  
28  
29 346 and are enriched in neural crest cells [43-45]; and *ETS1*, which regulates  
30  
31 347 neural crest development through mediating BMP signalling [46], were  
32  
33 348 preferentially expressed in the Ros-E1 subpopulation (Fig. 2b, c). The  
34  
35 349 ectoderm marker, *OTX1*, and genes involved in the ventral hindbrain marker  
36  
37 350 (e.g., *IRX3*) were highly expressed in the Ros-E2 subgroup (Fig. 2b, c). GO  
38  
39 351 term annotation analysis showed Ros-E1 and Ros-E2 shared GO terms of “cell  
40  
41 352 cycle G1/S phase transition”, “G1/S transition of mitotic cell cycle”, “epithelial  
42  
43 353 cell proliferation” and “positive regulating of binding” (Fig. 2d) while “negative  
44  
45 354 regulation of neuron differentiation” and “tube morphogenesis” were solely  
46  
47 355 enriched in the Ros-E2 subpopulation (Fig. 2d). KEGG enrichment analysis  
48  
49 356 showed that “base excision repair”, “DNA replication”, “axon guidance”, “cell  
50  
51 357 cycle” and “mismatch repair” were specifically associated with the Ros-E2  
52  
53 358 subset (Fig. 2e). We further performed single-cell differential expression  
54  
55 359 (SCDE) on both Ros-E subpopulations and identified additional differentially  
56  
57 360 expressed genes between the two groups. *SIX3*, *SIX6*, *TFAP2B* and *PBX1*  
58  
59  
60  
61  
62  
63  
64  
65

1 361 were more highly expressed in Ros-E2, whereas *EDN1*, *S100A10* and other  
2 362 genes related to neural crest migration, were highly expressed in Ros-E1 (Fig.  
3  
4 363 2f).  
5  
6 364

### 8 365 **Late rosette (Ros-L) stage**

9  
10 366 At the Ros-L stage the genes *SNAI2*, *OTX2*, *FEZF1*, *ZIC3*, and *HESX1*  
11 367 showed significantly different expression patterns among the three  
12 368 distinguishable subpopulations (Ros-L1, Ros-L2 and Ros-L3) at the Ros-L  
13 369 stage (Additional file 7: Figure S7a, b). Moreover, *SMAD1* and *MYC*, two  
14 370 components in the Wnt signaling pathway which is critical for neural  
15 371 development [47, 48], were specifically enriched in the Ros-L3 subpopulation.  
16 372 Additionally, *JUNB* from the TGF $\beta$  signaling pathway was preferentially  
17 373 expressed in Ros-L3 compared to the other two subpopulations. Interestingly,  
18 374 *HAND1* and *ISL1*, which are mesoderm markers, and *TBX3*, which elicits  
19 375 endodermal determination, were highly expressed in the Ros-L1  
20 376 subpopulation (Additional file 7: Figure S7a, b).  
21  
22  
23  
24  
25  
26  
27  
28  
29  
30  
31  
32

33 377  
34  
35 378 Of 648 GO terms identified by differentially expressed genes among these  
36 379 three subsets, 52 terms were shared by Ros-L1 and Ros-L3, such as “positive  
37 380 regulation of cell motility”, “angiogenesis”, “positive regulation of cellular  
38 381 component movement” and “epithelium migration” (Additional file 7: Figure  
39 382 S7c). A high proportion of cardiac development terms was enriched in Ros-L1,  
40 383 whereas DNA replication- and chromatin remodeling-related terms and  
41 384 pathways were significantly associated with Ros-L2. In addition, cell-substrate  
42 385 adhesion-related terms and cell cycle-related pathways were enriched in  
43 386 Ros-L3 (Additional file 7: Figure S7c, d).  
44  
45  
46  
47  
48  
49  
50  
51  
52  
53

54 387  
55  
56 388 Several subpopulation-specific genes were identified, including *NR2F1*,  
57 389 *ARID3A*, *SIX3*, *OTX2* and *FOXP1* at the NPCs stage (Additional file 8: Figure  
58 390 S8a, b). These observations suggest that significant TF expression patterns  
59  
60  
61  
62  
63  
64  
65

1 391 describe discrepant cell differentiation states or differentiation commitments  
2 392 inside the neural conversion process. Taken together, our results suggest that  
3  
4 393 the subpopulation analyses accurately describe specific gene expression  
5  
6 394 dynamics at each cell stage, which are likely masked in bulk sequencing  
7  
8 395 analyses. Additionally, extrapolating from these observations, we can reason  
9  
10 396 that reconstructing a differentiation trajectory based on the gene expression  
11  
12 397 dynamics of individual subpopulations would allow us to dissect neural  
13  
14 398 differentiation processes that we would otherwise be unable to observe.  
15  
16  
17 399

### 18 19 400 **Tracking a reconstructed trajectory identifies key subpopulations during** 20 21 401 **neural differentiation**

22  
23 402 Based on the subpopulations identified before, we wanted to track the gene  
24  
25 403 expression dynamics of individual subpopulations to parse the neural  
26  
27 404 differentiation processes and dissect the subpopulation with the highest  
28  
29 405 contribution towards commitment to the CNS lineage. First, we reconstructed  
30  
31 406 the differentiation trajectory using 8220 genes with variable expression. This  
32  
33 407 showed that cells in stages from iPSCs to NPCs followed a sequential  
34  
35 408 differentiation process where each stage exhibited a relatively discriminative  
36  
37 409 region with some of the subpopulations overlapping (Fig. 3a). Subsequently,  
38  
39 410 based on the pairwise comparisons of TF expression levels, we inferred the  
40  
41 411 connection of the subpopulations from the iPSCs stage to NPCs stage across  
42  
43 412 the five-stage differentiation process (Fig. 3b). TF expression levels were  
44  
45 413 considered as strong indicators of cell stage and identity [36]. Here, we used  
46  
47 414 the Pearson correlation coefficient to identify more biologically and molecularly  
48  
49 415 similar cell subpopulations and considered them as cells within the same  
50  
51 416 developmental lineage [49]. As a result, iPSCs, EB3, Ros-E2, Ros-L3 and  
52  
53 417 NPC1 were identified as the subpopulations contributing the most to  
54  
55 418 commitment to the CNS lineage (Fig. 3b). These findings were consistent with  
56  
57 419 the specific gene expression pattern in individual subpopulations. For instance,  
58  
59 420 *SOX13*, expressed in the developing nervous system and neural tube [50,51],  
60  
61  
62  
63  
64  
65

1 421 *FOXO1* [39] and *TULP3* [40] were significantly highly expressed in EB3  
2 422 (Additional file 6: Figure S6c, d). *MAFB*, an important TF in hindbrain identity  
3 423 [52], was enriched in Ros-E2 (Fig. 2b, c); and other crucial neural development  
4 424 TFs, especially those involved in CNS development, such as *OTX1*, *DLX3*,  
5 425 *DLX6*, *ZIC3*, *ZIC4*, and *IRX3*, also showed high expression in the Ros-E2  
6 426 subpopulation (Fig. 2b, c). Previously, we assumed that *GRHL3* might be  
7 427 involved in neural tube closure; here, the results showed that *GRHL3* was  
8 428 indeed significantly highly expressed in Ros-L3 (Additional file 7, Figure S7b).  
9 429 Additionally, neural crest regulators (e.g., *ETS1*, *ELK3*, *SOX9*) were enriched  
10 430 in Ros-L3 (Additional file 7, Figure S7b), suggesting that cell fate specification  
11 431 and differential cell status might exist even within subset. Strikingly, Ros-E2  
12 432 and Ros-L3 that were identified in the dominant path to CNS lineage by  
13 433 correlation analysis were shown as a process of sequential conversion in our  
14 434 reconstructed trajectory (Fig. 3a, c). The molecular signature described by  
15 435 these subpopulations was consistent with the analysis that identified the key  
16 436 contributing subpopulations and encouraged us to perform additional cell fate  
17 437 decision analyses.

18 438  
19 439 Of note, there was a clear divarication within the rosette stages (Ros-E and  
20 440 Ros-L) across the differentiation trajectory, indicating cell fate decision might  
21 441 be made at this bifurcation point (Fig. 3c). Here, we focused on the single cells  
22 442 in the rosette stages and called them Branch 1, Branch 2 and Branch 3 based  
23 443 on their location in the developmental trajectory (Fig. 3c). Branch 3 was  
24 444 composed of Ros-E1 (n=27), Ros-L1 (n=15) and small proportion of Ros-E2  
25 445 (n=5) and Ros-L3 (n=9, Fig. 3c). Previously, our observations showed that  
26 446 Ros-E1 was associated with neural crest cells (high expression of *TWIST1*,  
27 447 *SOX9*, *ETS1*, *EDN1* and *S100A10*) and Ros-L1 was likely related to  
28 448 mesoderm and endodermal determination (high expression of *HAND1*, *ISL1*  
29 449 and *TBX3*), and these two subpopulations comprise the majority of cells in  
30 450 Branch 3. Further, we performed a pairwise comparison of gene expression

1 451 across the three branches. The results showed that many neural TFs, such as  
2 452 markers of neural tube formation (*SOX4* and *SOX11*); the NSCs self-renewal  
3 453 and proliferation regulator *FOXO3*; and the NSC markers *NES*, *CDH2* and  
4 454 *FABP7*, were commonly expressed across all three branches, indicating the  
5 455 capacity for neural tube development and NSCs proliferation are a  
6 456 fundamental feature of neural rosettes (Additional file 9: Figure S9a, b).  
7 457 Strikingly, *ZIC2*, a member of the ZIC family of C2H2-type zinc finger proteins,  
8 458 associated with neural tube development [32], showed significantly low  
9 459 expression in Branch 3 (Fig. 3d, e). Some other neural development markers  
10 460 (e.g., *ZIC3*, *HMGB2*, *ID1*, *SIX3*, *SIX6*, *NR6A1*) were significantly lowly  
11 461 expressed in Branch 3 but highly expressed in Branch 1 (Fig. 3d, e, Additional  
12 462 file 9: Figure S9a, c). However, *TFAP2B*, encoding a member of the AP-2  
13 463 family of TFs, and *ELK3*, essential for the progenitor progression to neural  
14 464 crest cell [53], was significantly highly expressed in Branch 3 but lowly  
15 465 expressed in Branch 2. Moreover, *SOX9*, *SNAI2*, *S100A11* and *TFAP2A*,  
16 466 previously shown to be highly expressed in neural crest cells [43,44,45,54],  
17 467 were markedly highly expressed in Branch 3, but not Branch 1 (Fig. 3d, e,  
18 468 Additional file 9: Figure S9a, c). *KLF5* and *IRF6* were significantly highly  
19 469 expressed in Branch 3 as well (Fig. 3d, e). These two TFs have been reported  
20 470 to be involved in phenotypic switching of vascular smooth muscle cells [55]  
21 471 and development of the palate in vertebrates involving cranial neural crest  
22 472 migration [56], respectively. These results indicate that cell fate specification  
23 473 might occur at the bifurcation point and, based on the observations, we  
24 474 speculate that Branch 1-to-Branch 2 has progressed more towards CNS and  
25 475 Branch 3 is probably composed of neural crest cells and other cells comprising  
26 476 this microenvironment.

27 477

## 28 478 **Construction of the TF regulatory network during cell status transition**

29 479 To infer TFs which drive the progression of cell status from one stage to the  
30 480 neighbouring one, we performed SCDE analysis for those cell subpopulations



1 481 committing to CNS lineage, resulting in 58, 123, 98 and 131 TFs differentially  
2 482 expressed among iPSCs vs EB3, EB3 vs Ros-E2, Ros-E2 vs Ros-L3, and  
3  
4 483 Ros-L3 vs NPC1 comparisons (Additional file 10, 11: Figure S10, 11).  
5  
6 484 Interestingly, *PRDM1*, which has been proposed to promote the cell fate  
7  
8 485 specification RB sensory neurons in zebrafish [57], was significantly  
9  
10 486 up-regulated from Ros-E2 to Ros-L3 (Additional file 10: Figure S10). In  
11  
12 487 contrast, several well-characterized TFs were found to be significantly highly  
13  
14 488 expressed in Ros-E2 (mainly resident in Branch 1) and down-regulated during  
15  
16 489 the transition from early to late rosette development: *FOXP1*, cooperating with  
17  
18 490 *Bmi-1* to maintain neural stem cell self-renewal in the forebrain; *MAFB*, the  
19  
20 491 posterior CNS fate identifier and essential for hindbrain choroid plexus  
21  
22 492 development [52, 58]; *DLX3* and *DLX5*, neural plate border specifier genes  
23  
24 493 [58]; and *ID1*, a controller of stem cell proliferation during regenerative  
25  
26 494 neurogenesis in the adult zebrafish telencephalon [59]. These results suggest  
27  
28 495 that the expression patterns of neural-associated TFs undergo dramatic  
29  
30 496 changes during neural differentiation with some TFs activated (*PRDM1*, etc.)  
31  
32 497 and others repressed (*MAFB*, *FOXP1*, *ID1*, etc.) (Additional file 10: Figure  
33  
34 498 S10). Furthermore, it was previously unknown that several of these TFs were  
35  
36 499 involved in neural differentiation so our results have expanded the known  
37  
38 500 biological functions of these molecules.  
39  
40  
41  
42

43 501  
44 502 Among the 131 TFs exhibiting differential expression from Ros-L3 to NPC1, 80  
45  
46 503 TFs were up-regulated while 51 TFs were down-regulated (Additional file 11:  
47  
48 504 Figure S11; Additional file 19: Table S1). Up-regulated TFs included *SNAI2*, a  
49  
50 505 neural crest specifier [58]; *HIF1A*, required for neural stem cell maintenance  
51  
52 506 and vascular stability in the adult mouse [60]; *SIX1*, which drives the neuronal  
53  
54 507 developmental program in the mammalian inner ear [61]; *ETV1*, which  
55  
56 508 orchestrates gene regulation during the terminal maturation program of  
57  
58 509 cerebellar granule cells [62]; and *POU3F3*, which influences neurogenesis of  
59  
60 510 upper-layer cells in the cerebral cortex [63] (Additional file 11: Figure S11). This  
61  
62  
63  
64  
65

1 511 is consistent with our previous observation that the main trajectory has  
2 512 progressed more towards to CNS. Of particular interest is *PRDM1*, whose  
3 513 expression increased from Ros-E2 to Ros-L3 and decreased during the  
4 514 progression from Ros-L3 to NPC1 (Additional file 10, 11: Figure S10, 11),  
5 515 suggesting that it might play multiple specific roles in neural differentiation.  
6  
7  
8  
9

10 516  
11  
12 517 Next, we inferred a regulatory network among those differentially expressed  
13 518 TFs based on known interactions collected in the STRING database [64]. Our  
14 519 results suggested that *SOX2* and *GATA3* were key regulators from iPSCs to  
15 520 EB3 (Additional file 12: Figure S12a); *TP53*, *SOX2*, *RELA*, *SIX3*, *ARNTL*, *ISL1*,  
16 521 *RARA*, *TP63*, *GATA3*, *SNAI2*, and *PAX3* were the key regulators from EB3 to  
17 522 Ros-E2 (Additional file 12: Figure S12b); *MYC*, *SOX2*, *PAX6*, *EGR1*, *PBX1*,  
18 523 *GLI3*, *PAX3*, *SIX3*, *FOXP1*, *OTX2*, *PAX7*, *PPARG*, *SOX9*, *MAFB*, *SIX6* and  
19 524 *ZIC1* were identified as key regulators from Ros-E2 to Ros-L3 (Fig. 4a); and  
20 525 *SOX2*, *AR*, *MYCN*, *LEF1*, *PAX3*, *SNAI2*, *MSX1*, *SOX9*, *NR3C1*, *PARP1*,  
21 526 *RUNX1*, *EBF1*, *HIF1A*, *IRF6*, *IRF1*, *KLF5*, and *LIN28A* were predicted to be  
22 527 key regulators from Ros-L3 to NPC1 (Fig. 4b).  
23  
24  
25  
26  
27  
28  
29  
30  
31  
32  
33  
34  
35  
36

37 528  
38 529 To dissect the *cis*-regulatory elements directing the expression of those  
39 530 regulators, we selected the differentially expressed TFs that showed  
40 531 differential ATAC peaks between neighbouring stages and performed motif  
41 532 scanning on the differential peaks. Focusing on the transition from Ros-E2 to  
42 533 Ros-L3, we found transcription factor binding sites (TFBSs) for TEAD2 and  
43 534 YY1 in a differential ATAC peak downstream of the *PRDM1* gene (Fig. 4c).  
44 535 Multiple motifs for the transcription factor *TFAP2C* were found in a differential  
45 536 peak located in the intron of the *ARID3A* gene, which is a regulator responsible  
46 537 for the transition for Ros-L3 to NPCs (Fig. 4d). Based on the temporal  
47 538 specificity of ATAC peaks and the existence of TF motifs in these regions, we  
48 539 propose that those elements are stage-specific *cis*-regulatory elements  
49 540 regulating the expression of neural regulators in response to their upstream  
50  
51  
52  
53  
54  
55  
56  
57  
58  
59  
60  
61  
62  
63  
64  
65

1 541 regulatory TFs.

2 542

3  
4 543 To infer the putative targets of key regulators, we combined the information  
5  
6 544 from ATAC peaks and motifs for TFs. All peaks containing motifs for a certain  
7  
8 545 TF were annotated as TF-related peaks and genes proximal to the peak were  
9  
10 546 considered as potential targets of that TF. Using these criteria, we predicted  
11  
12 547 thousands of targets for each TF (Additional file 20: Table S2). To dissect the  
13  
14 548 regulatory network of each TF, we conducted GO term and KEGG enrichment  
15  
16 549 analysis for the putative target list of each key regulator. Our results suggested  
17  
18 550 that, from Ros-E2 to Ros-L3, the targets for *PRDM1* were significantly  
19  
20 551 enriched in pathways and GO terms associated with “axon guidance”, “hippo  
21  
22 552 signalling pathway” and “neurotrophin signalling pathway” (Fig. 4e and  
23  
24 553 Additional file 13: Figure S13). From Ros-L3 to NPC1, targets for *HIF1A*,  
25  
26 554 *NR2F1*, *SOX9* and *TFAP2C* were enriched in KEGG pathways associated with  
27  
28 555 “axon guidance” and “hippo signalling pathway” (Additional file 13: Figure S13).  
29  
30 556 We further validated *PRDM1* expression among different genetic background  
31  
32 557 cell lines (H1\_ESCs, H7\_ESCs, H9\_ESCs, iPS25 and iPSC129). The  
33  
34 558 immunostaining showed that *PRDM1* was expressed at Ros-L stage with  
35  
36 559 heterogeneous expression level, though, the scRNA-seq data was not at a  
37  
38 560 high level. Moreover, the results were uniformed across these cell lines (Fig.  
39  
40 561 4g, h).

41 562

### 42 563 **Inferring a cellular communication network among cell subpopulations** 43 44 564 **within specific differentiation stages**

45  
46 565 Cell subpopulations with different functions are proposed to exhibit distinct  
47  
48 566 expression profiles of ligands and receptors which primes cells for  
49  
50 567 cell-type-specific interactions [65]. In this study, the cellular interactions were  
51  
52 568 inferred using public ligand-receptor databases (see Methods). Briefly, 360,  
53  
54 569 182, 261 and 307 ligands/receptors were expressed within EB, Ros-E, Ros-L  
55  
56 570 and NPCs subpopulations respectively, among which 304, 55, 124 and 162

1 571 interactions were identified within subpopulations at each differentiation time  
2 572 point (Fig. 5, Additional file 14-16: Figure S14-16 and Additional file 21: Table  
3 573 S3). The most frequent interactions were observed in the EB stage, implying  
4 574 that cells communicate extensively to coordinate differentiation programs  
5 575 during embryogenesis (Additional file 14: Figure S14). In contrast, much fewer  
6 576 interactions were predicted after the EB stage, suggesting communications  
7 577 decreased dramatically during the progression of lineage commitment. Notably,  
8 578 although comparable number of ligands and receptors were detected at EB  
9 579 (181 receptors and 179 ligands) and NPCs (128 receptors and 179 ligands)  
10 580 stage, only half the interactions (162) were inferred at NPCs stage compared  
11 581 to 304 ligand-receptor interactions at EB stage. (Additional file 14, 16: Figure  
12 582 S14, 16). The interactomes among Ros-L cells, with 31, 32 and 34 receptors  
13 583 from Ros-L1, Ros-L2 and Ros-L3 interacting with ligands from other cell  
14 584 subpopulations were inferred (Fig. 5a). As expected, several interactions  
15 585 involving receptors and ligands previously known to play essential roles during  
16 586 neural development were identified in our study. For example, *WNT5A* and  
17 587 *EPHB6* were enriched in Ros-L1. *FZD5* and *LPAR4* were specifically  
18 588 expressed in Ros-L2. *PGF* and *ANGPT2* were up-regulated in Ros-L3  
19 589 compared to other cell subpopulations (Fig. 5c, d, e). Overall, our study  
20 590 suggests that the specific expression spectrum of ligands and receptors and  
21 591 corresponding interactions can generally reflect the identity of cellular  
22 592 subpopulations.

## 47 594 **Discussion**

48 595 The regulation and molecular programs during embryonic neural development  
49 596 has long been investigated. However, much of this work has been limited to  
50 597 model organisms such as the mouse, zebrafish and *Drosophila* [36,40,56], due  
51 598 to the scarcity of human fetal tissue for research purposes. Our understanding  
52 599 of human early neural development, and particularly neural tube formation and  
53 600 the cell fate commitments of neural precursors in early stages, is still

1 601 incomplete. To circumvent the challenges inherent in these investigations,  
2 602 namely the ability to study these processes *in vivo* in humans, we used hiPSCs  
3  
4 603 and induced differentiation *in vitro* towards a neural cell fate using a  
5  
6 604 well-established model. We characterised both the transcriptional profiles in  
7  
8 605 single cells as well as chromatin accessibility at several critical stages during  
9  
10 606 differentiation to inform this process at unprecedented resolution. This study  
11  
12 607 has unveiled the dynamic transcriptome and regulome underlying the human  
13  
14 608 early neural differentiation and identified functionally-distinct subpopulations  
15  
16 609 within the various stages to have a more precise description of the factors  
17  
18 610 defining the differentiation trajectory. Our analyses hint at the existence of a  
19  
20 611 widespread regulatory network between TFs and their target genes, especially  
21  
22 612 those associated with cellular reprogramming and differentiation. We were  
23  
24 613 also able to construct minimal gene expression profiles based on ligands and  
25  
26 614 receptors in each cell subpopulation which can be used to confidently infer cell  
27  
28 615 identity.  
29  
30

31 616  
32  
33 617 During development *in vivo* the neuroectoderm folds to form the neural tube  
34  
35 618 which is then patterned into regionally specialized subunits composed of  
36  
37 619 progenitor cells. These cells subsequently give rise to regional progenies of  
38  
39 620 neural cells [66]. There is some controversy in this field that formation of the  
40  
41 621 EB would introduce *in vitro* culture variability in regional cells across different  
42  
43 622 batches resulting in a relatively poor model of neural differentiation. The  
44  
45 623 "dual-SMAD inhibition" method (inhibiting the SMAD-dependent TGF $\beta$  and  
46  
47 624 BMP signaling pathways) yielding neural epithelia in "monolayer culture"  
48  
49 625 conditions [18] could alleviate the above concern. However, generation of  
50  
51 626 neural rosette morphology *in vitro* is considered equivalent to neural tube  
52  
53 627 formation, recapitulating neural tube structure, which we believe is a promising  
54  
55 628 research model for early neural differentiation. Neural differentiation of hiPSCs  
56  
57 629 into NPCs starts with initial neural induction by appropriate dosages and  
58  
59 630 gradients of many TFs and morphogenetic factors that are highly expressed in  
60  
61  
62  
63  
64  
65

1 631 the developing brain. In this study, the induction cocktail used in the neural  
2 632 differentiation included SB431542, dorsomorphin, N2, B27, VEGF and bFGF  
3  
4 633 supplemented at specific time points. The self-renewal program in human  
5  
6 634 iPSCs is switched off and differentiation toward NE and NPCs is triggered [8,  
7  
8 635 16]. Previous results have shown that SB431542 enhances neural induction in  
9  
10 636 EB derived from hESCs [65] by inhibiting the Lefty/Activin/TGF $\beta$  pathways and  
11  
12 637 suppresses the mesodermal lineage (Brachyury) induction [18, 42]. Consistent  
13  
14 638 with these previous studies, in our *in vitro* system, treatment with SB431542, in  
15  
16 639 combination with dorsomorphin, results in a dramatic decrease in *NANOG*  
17  
18 640 expression and a concomitant increase in *PAX6* expression (Fig. 1f). In  
19  
20 641 addition, *OTX2*, *ZIC2*, *SOX9*, *HESX1*, *MSX2*, *DLX5*, *SOX4*, *SOX11*, and  
21  
22 642 *SNAI2* were significantly activated during differentiation which demonstrates  
23  
24 643 that the transcriptional program triggering progression towards NPCs was  
25  
26 644 activated (Fig. 1f, Additional file 4: Figure S4h and Additional file 9: Figure  
27  
28 645 S9a-c). Taken together, these results indicate that the induction cocktail  
29  
30 646 effectively achieves efficient neural differentiation.

31  
32  
33 647

34  
35 648 To measure the dynamic changes of *cis*-regulatory elements at each  
36  
37 649 differentiation stage, we performed ATAC-seq and chromatin accessibility  
38  
39 650 analysis on bulk cells. These results showed widespread and comprehensive  
40  
41 651 chromatin structure reprogramming during neural differentiation. In particular,  
42  
43 652 TFBSs for several neural master regulators were enriched in temporally  
44  
45 653 dynamic ATAC peaks, indicating that changes in chromatin accessibility are  
46  
47 654 indeed associated with, and are probably responsive to, the regulation of  
48  
49 655 neural-related TFs. In addition, we also investigated closing (lost) peaks  
50  
51 656 dynamics as well as the functional annotation study, which was in line with the  
52  
53 657 corresponding annotation of novel peaks (Additional file 2, 3: Figure S2, 3). We  
54  
55 658 further identified several enriched TF motifs (e.g., *Pax2* in Ros-L and *FOXO1*  
56  
57 659 in NPCs) (Additional file 17: Figure S17d, e) which are known to play an  
58  
59 660 important role in neural differentiation, consistent with results from previous  
60  
61  
62  
63  
64  
65

1 661 studies [39, 68].  
2 662  
3  
4 663 By integrating single cell-based transcriptome profiling of 391 cells from five  
5  
6 664 differentiation stages, we identified a variety of TFs that were differentially  
7  
8 665 expressed throughout the differentiation process and showed distinct  
9  
10 666 expression profiles among specific cell stages. The TFs *SOX2*, *PAX6*, *OTX2*,  
11  
12 667 *SOX4*, *ZIC2*, *LHX5*, *HESX1*, and *SIX3* were significantly highly expressed at  
13  
14 668 the EB stage (Fig. 1f). It has been reported that members of the  
15  
16 669 grainyhead-like (Grhl) family of TFs, which are well-conserved from *Drosophila*  
17  
18 670 to human, are highly expressed during neurulation in mice and that a  
19  
20 671 *Grhl3*-hypomorphic mutant resulted in NTDs [32, 67]. Remarkably, our results  
21  
22 672 showed that two human Grhl family TFs, *GRHL2* and *GRHL3*, were  
23  
24 673 significantly highly expressed at EB and Ros-E stage, respectively (Fig. 1f and  
25  
26 674 Additional file 4: Figure S4h), and the downstream targets of *GRHL2* (including  
27  
28 675 *E-CADHERIN*, also known as *CDH2*) were highly expressed at the neural  
29  
30 676 rosette stage (Fig. 1b) supporting a role for Grhl TFs in neural tube closure in  
31  
32 677 humans. In addition, previous studies have shown that in the *Drosophila*  
33  
34 678 olfactory system the homeobox gene *distal-less* is required for neuronal  
35  
36 679 differentiation and neurite outgrowth [34]. Our data showed that four homologs  
37  
38 680 of *distal-less* (*DLX3*, *DLX4*, *DLX5*, *DLX6*) were significantly up regulated at the  
39  
40 681 Ros-E stage and were highly expressed in the Ros-E2 subpopulation (Fig. 1f  
41  
42 682 and Fig. 2b) implying that the *distal-less* gene family plays a role in neural  
43  
44 683 differentiation in humans.  
45  
46  
47  
48  
49

50 685 We also applied scRNA-seq to our *in vitro* neural model to dissect the  
51  
52 686 subpopulations present at each differentiation stage (Fig. 2 and Additional file  
53  
54 687 5-8: Figure S5-8). We were then able to reconstruct a differentiation trajectory  
55  
56 688 based on the subpopulations that we identified by variable TF expression  
57  
58 689 within each stage (Fig. 3a). Strikingly, a divarication within the rosette stage  
59  
60 690 across the differentiation trajectory was observed. Comparing Branch 1 to  
61  
62  
63  
64  
65

1 691 Branch 3, Branch 3 possessed the relatively lowly-expressed TFs *LHX5*,  
2 692 *HESX1* and *SIX3* (reported as anterior forebrain markers), as well as other  
3  
4 693 crucial neural TFs (*SOX2*, *HMGB2*, *ZIC2*, *OTX1*, *FEZF1*); and the relatively  
5  
6 694 highly-expressed TFs *TFAP2B*, *SOX9*, *ELK3*, and *SNAI2* (Fig. 3d, e and  
7  
8 695 Additional file 9: Figure S9a, c) which are considered to be neural crest  
9  
10 696 markers [53]. Though *SNAI2* was also expressed at the NPCs stage,  
11  
12 697 combined with other neural crest markers, we proposed that Branch 3 was  
13  
14 698 progressing more towards to neural crest cells (Fig. 3a-c and Additional file 9:  
15  
16 699 Figure S9a-c). Taken together, these observations imply that the main  
17  
18 700 differentiation trajectory (Branch 1 and Branch 2) is heading towards CNS,  
19  
20 701 whereas Branch 3 is progressing towards neural crest cells.  
21  
22

23 702  
24  
25 703 It is important to note that the current scRNA-seq method by its nature only  
26  
27 704 provides a snapshot of the gene expression profile for individual cells. A  
28  
29 705 possible resolution for the above problem is to capture the sample with much  
30  
31 706 more precise time points, which may, to some extent, overcome this limitation.  
32  
33 707 Thus, in spite of the very interesting heterogeneity and cell fate commitment  
34  
35 708 study inferred above, we cannot exclude the following factors that may affect  
36  
37 709 cell subset identification in the above description; 1) temporal transcriptional  
38  
39 710 states during transient differentiation process; 2) differentiation efficiency; and  
40  
41 711 lagging and leading cells remaining in the differentiation process. However, we  
42  
43 712 propose that the subsets dissection analysis facilitates a more precise  
44  
45 713 description of the factors defining the differentiation trajectory. When we  
46  
47 714 constructed the differentiation trajectory using the cells that collected at  
48  
49 715 different time points, the results showed that all subpopulations in stages from  
50  
51 716 iPSCs to NPCs followed a sequential differentiation process where each stage  
52  
53 717 exhibited a relatively discriminative region with some of the subpopulations  
54  
55 718 overlapping (Fig. 3a), indicating that in spite of the above concerns, the  
56  
57 719 trajectory was established by the natural features of the respective subsets  
58  
59 720 and which is also supported by the observations that Ros-L2 possessing many  
60  
61  
62  
63  
64  
65



1 721 early neural differentiation TFs, such as *SOX2*, *OTX2*, *PAX6*, *OTX1*, and *LHX5*,  
2 722 as well as forebrain markers (e.g., *HESX1*) and pluripotency-related TFs  
3  
4 723 (*NANOG*, *SALL4*, *PRDM14*) (Additional file 7: Figure S7) were located in the  
5  
6 724 reconstructed trajectory prior to the generation of Ros-E populations. In  
7  
8 725 addition, we carried out the cell fate commitment analysis using Branch1,  
9  
10 726 Branch2 and Branch3 which were grouped based on the cell locations on the  
11  
12 727 trajectory rather than cell subsets identified by Seurat in order to minimize the  
13  
14 728 above concerns.  
15

16 729  
17  
18 730 Notably, our study reveals the regulatory network of TFs that are differentially  
19  
20 731 expressed among neighbouring cell subpopulations to be likely candidates for  
21  
22 732 promotion of cell fate transition. Based on the topology of this network, we  
23  
24 733 focused on novel regulators (*PRDM1* and *ARID3A*), especially *PRDM1*, which  
25  
26 734 are located on the hub of the network, interacting with both known and novel  
27  
28 735 neural regulators. Although the roles of several TFs have been reported during  
29  
30 736 neural differentiation and brain patterning formation in humans, some TFs have  
31  
32 737 been proposed to play a role in neural fate commitment in non-human species  
33  
34 738 (mouse and zebrafish). However, the interaction partners, *cis*-regulatory  
35  
36 739 elements, and genetic regulatory networks of those TFs are yet to be resolved.  
37  
38 740 Here, we identified the *cis*-regulatory elements for *PRDM1* and *ARID3A* genes  
39  
40 741 and predicted their upstream regulators. Of particular interest, *TFAP2C*'s role  
41  
42 742 in regulating neural development has been widely reported, increasing the  
43  
44 743 confidence of our predictions. In humans, *PRDM1* is reported to promote germ  
45  
46 744 cell fate by suppressing neural effector *SOX2*, but the function of *PRDM1* in  
47  
48 745 neural development is unknown. In zebrafish, *Prdm1a*, the homolog of the  
49  
50 746 *PRDM1* gene, directly activates *foxd3* and *tfap2a* during neural crest  
51  
52 747 specification [57]. Mutation of *prdm1* in zebrafish resulted in severe  
53  
54 748 phenotypes with a decrease in the quantity of neural crest cells and the  
55  
56 749 reduction in the size of structures derived from the neural crest [57]. Similarly,  
57  
58 750 strong expression of *prdm1* was observed in the neural plate border of a basal  
59  
60  
61  
62  
63  
64  
65

1 751 vertebrate lineage, lamprey, implying that the role of *prdm1* in the neural crest  
2 752 formation is likely a conserved, ancestral role [70]. Conversely, *prdm1* is  
3 753 dispensable for neural crest formation in mice, and instead is required for  
4 754 primordial germ cell specification suggesting that the neural crest specification  
5 755 function of *prdm1* in mice has been lost [71]. Overall, previous studies suggest  
6 756 that functions of *prdm1* are quite diverse and need to be investigated in  
7 757 species-, developmental-, and environmental-specific manners. Based on the  
8 758 known interaction between *PRDM1* and *SOX2* in humans, as well as the  
9 759 observation that *PRDM1* expression increased significantly from Ros-E2 to  
10 760 Ros-L3 and was preferentially expressed in Ros-L3 compared to other two  
11 761 subpopulations in the rosette stage (Fig. 4g, h; Additional file 7: Figure S7a, b  
12 762 and Additional file 10: Figure S10), we propose *PRDM1* as a novel neural  
13 763 regulator in early human neural differentiation. Our hypothesis is supported by  
14 764 the GO term and KEGG enrichment analysis of putative targets of *PRDM1*,  
15 765 which are significantly enriched in “axon guidance” and hippo  
16 766 pathway-associated terms (Fig. 4e and Additional file 13: Figure S13a).  
17 767 However, the functions of putative TFs need to be further investigated using  
18 768 experimental methods.

19 769  
20  
21 770 To infer cellular interactions, communication network analysis was applied to  
22 771 the expression profiles of ligands and receptors in stage-specific  
23 772 subpopulations. Two trends were observed in our cellular interaction network  
24 773 analysis: 1) the frequency of cellular interactions peaked at EB stage; and 2)  
25 774 different cell subpopulations showed a certain degree of specificity in their  
26 775 ligand-receptor spectrum. The observation that most interactions were inferred  
27 776 at the EB stage likely reflects the extensive cellular communication during  
28 777 embryogenesis and early neural differentiation (Additional file 14: Figure S14).  
29 778 Regarding the ligand-receptor expression spectra, matched ligand and  
30 779 receptor expression probably underlies the common functions shared by  
31 780 different cell subpopulations within the same stage. In contrast, those specific

1 781 ligands or receptors probably reveal the unique regulatory code of distinct cell  
2 782 subpopulations. For example, *WNT5A*, a crucial regulator of neurogenesis  
3 783 during the development of cerebellum, and *BMP4*, one of the key regulators of  
4 784 dorsal cell identity in the neural tube [72], were highly expressed in Ros-L1  
5 785 compared to other cell subpopulations (Fig. 5c). FZD5 (required for eye and  
6 786 retina development in mouse [73]), and *FGF19* (required for forebrain  
7 787 development in zebrafish [74]) were preferentially expressed in Ros-L2 (Fig.  
8 788 5d and Additional file 22: Table S4). *WNT7A*, involved in several aspects of  
9 789 neurogenesis, including synapse formation and axon guidance [75] and *FGF1*,  
10 790 which maintains the self-renewal and proliferation of NPCs [76], were  
11 791 specifically expressed in Ros-L3 (Additional file 22: Table S4). Pavličev et al.  
12 792 inferred the cell communication network of the maternal-fetal interface and  
13 793 found that ligand-receptor profiles could be a reliable tool for cell type  
14 794 identification [65]. Consistent with their findings, our study suggests that the  
15 795 repertoire of ligands-receptors in neural cell types could probably, to some  
16 796 extent, represent the identity of cell subpopulations.

17 797

18 798 There might be a concern that we only used one genetic background cell line  
19 799 for this study, possibly making the cogency of our findings limited. To address  
20 800 this, we performed ESCs neural differentiation and captured bulk  
21 801 transcriptome profiles of the corresponding differentiation stages (ESCs, EB,  
22 802 Ros-E, Ros-L and NPCs). The observations in ESCs were reproducible in  
23 803 iPSCs with regards to 1) PCA analysis (Additional file 18: Figure S18a); 2) with  
24 804 a high Pearson correlation coefficient between the corresponding cell stage  
25 805 derived from iPSCs and ESCs (Additional file 18: Figure S18b); and 3)  
26 806 validation analysis of subset- specific markers (MAFB, SOX9, PRDM1 and  
27 807 NR2F1). In addition, novel neural TF (PRDM1) expression in different genetic  
28 808 cell lines (H1\_ESCs, H7\_ESCs, H9\_ESCs, iPS25 and iPS129) was consistent  
29 809 with the above heterogeneity study (Additional file 18: Figure S18c, d, e).  
30 810 Together, our findings are supported by different genetic cell lines mitigating

1 811 the concern that our results are limited to the cells forming the basis of this  
2 812 study.

3 813

4 814 Through differential expression analysis, we identified genes specifically  
5 815 expressed at each stage which include both cell status master regulators such  
6 816 as TFs and signalling components, as well as realizators [24] which could  
7 817 directly determine cell growth, cell proliferation, cell morphology and cell-cell  
8 818 interaction. Within each stage, we identified subpopulations with distinct  
9 819 expression signatures, which might represent functional cell clusters or  
10 820 transient cell state given that neural cells have been shown to demonstrate  
11 821 significant heterogeneity as they express different surface proteins, exhibit  
12 822 diversified morphologies and secrete a variety of cytokines. Therefore, it is  
13 823 necessary to explore the heterogeneity of cell subpopulations and study each  
14 824 subpopulation in a case-by-case manner. In summary, our data show  
15 825 conclusively that both transcriptome and regulome dramatically change during  
16 826 neural differentiation, which affects a variety of biological pathways crucial for  
17 827 neural differentiation. We also propose several putative TFs as well as the  
18 828 ligands-receptors interaction spectrum that are important in each  
19 829 differentiation stage which paves the way for a deeper understanding of the  
20 830 cell fate decision and regulatory mechanisms driving the differentiation of the  
21 831 neural lineage.

22 832

## 23 833 **Materials and methods**

### 24 834 **Ethics statement**

25 835 The study was approved by the Institutional Review Boards on Ethics  
26 836 Committee of BGI (Permit No.BGI-IRB 14057). The participant (dermal  
27 837 fibroblast, Fib129) signed informed consent and voluntarily donated the  
28 838 samples for our study.

29 839

1  
2  
3  
4  
5  
6  
7  
8  
9  
10  
11  
12  
13  
14  
15  
16  
17  
18  
19  
20  
21  
22  
23  
24  
25  
26  
27  
28  
29  
30  
31  
32  
33  
34  
35  
36  
37  
38  
39  
40  
41  
42  
43  
44  
45  
46  
47  
48  
49  
50  
51  
52  
53  
54  
55  
56  
57  
58  
59  
60  
61  
62  
63  
64  
65

840 **Cell culture and reprogramming**

841 The human fibroblast cell line was derived from the dermal skin of a healthy  
842 female donor with written informed consent. Briefly, the skin tissue was  
843 washed with DPBS several times, sliced into approximately 1mm or smaller  
844 fragment size, enzymatically dissociated in High Dulbecco's modified Eagle  
845 medium (H-DMEM, Gibco, 11965118) with 100U/ml collagenase type IV  
846 incubating in 37°C overnight, then 0.05% trypsin incubating for 5 min. The  
847 dissociation was terminated by adding 2 ml fibroblast cell culture medium  
848 (H-DMEM +10% FBS + 5ng/ml bFGF+ 2mM Gln) followed by centrifugation at  
849 300g for 5 min. The cells were resuspended with fibroblast cell culture medium,  
850 and cultured at 37°C in a 5% CO<sub>2</sub> incubator. The fibroblast cell culture medium  
851 was changed every 2 days until reaching 80%–90% confluence and cells were  
852 passaged every 3-4 days.

853

854 For reprogramming, non-integrative human iPSCs were generated following a  
855 modified Shinya Yamanaka method [77]. Briefly, 5x10<sup>5</sup> human fibroblast cells  
856 at passage 4 were nucleofected with the program for human dermal fibroblast  
857 NHDF (Lonza, CC-2511) with 2.4ug episomal plasmids, including pCXLE-  
858 hOCT3/4- shp53-F (Addgene, 27077), pCXLE- hSK (Addgene, 27078),  
859 pCXLE- hUL (Addgene, 27080). Transfected cells were cultured in a six-well  
860 plate with culture medium containing H-DMEM supplemented with 10% FBS.  
861 The cells were trypsinized and 1x10<sup>5</sup> cells were seeded onto a 10cm<sup>2</sup> dish  
862 covered with feeder and cultured in a medium containing H-DMEM with 10%  
863 FBS while reaching 80% confluence. After that, the medium was changed to  
864 hiPSCs medium containing DMEM/F12 (Gibco, 11320-033), 20% KSR  
865 (Gibco,10828-028), 2mM L-glutamine (Sigma, G8540), 0.1μM NEAA  
866 (Gibco,11140-050), 0.1μM β-Mercaptoethanol (Gibco, 21985-023) and  
867 10ng/ml human bFGF (Invitrogen, PHG0021). The iPSCs colonies were  
868 picked at around day 25 and maintained in hiPSCs medium.

869

1  
2  
3  
4  
5  
6  
7  
8  
9  
10  
11  
12  
13  
14  
15  
16  
17  
18  
19  
20  
21  
22  
23  
24  
25  
26  
27  
28  
29  
30  
31  
32  
33  
34  
35  
36  
37  
38  
39  
40  
41  
42  
43  
44  
45  
46  
47  
48  
49  
50  
51  
52  
53  
54  
55  
56  
57  
58  
59  
60  
61  
62  
63  
64  
65

870 **Neural differentiation**

871 We applied a well-adopted neural differentiation protocol [8,16]. Briefly, human  
872 iPSCs were maintained as described above. To induce neural rosettes,  
873 hiPSCs were mechanically picked and washed with DMEM/F12 twice, and  
874 then cultured for 4 days in suspension with 5 $\mu$ M dorsomorphin (Sigma, P5499)  
875 and 5 $\mu$ M SB431542 (Sigma, S4317) in hiPSCs medium without bFGF for  
876 embryoid bodies (EBs) formation, then the EBs were attached on matrigel (BD,  
877 354277) coated dishes (BD, 354277) and cultured in DMEM/F12 (Gibco,  
878 11320-033) supplemented with 20 ng/ml bFGF, 1 $\times$ N2 (Gibco, 17502-048) and  
879 2 $\mu$ g/ml heparin (Sigma, 1304005) for an additional 3 or 5 days to harvest  
880 rosette-early (Ros-E) and rosette-late (Ros-L) cells, respectively. To collect  
881 neural progenitor cells (NPCs), rosettes structure that appeared in the center  
882 of attached colonies at Ros-L stage were carefully harvested using pulled  
883 glass pipettes and seeded on matrigel-coated dishes and cultured in  
884 DMEM/F12 supplemented with 1 $\times$  N2, 1 $\times$  B27 (Gibco,12587-010), 20 ng/ml  
885 bFGF, 20 ng/ml EGF (Invitrogen, PHG0311) and 2 $\mu$ g/ml heparin  
886 (Sigma,1304005) for additional 7 days, and the medium was changed every 2  
887 days. At day 16, the NPCs reaching approximately 80% confluence were  
888 collected, and all the mass or adherent cell samples were treated with  
889 TrypLE™ Express Enzyme (Gibco, 12604-021) for single cell dissociation and  
890 cryopreservation in gas-phase liquid nitrogen for further sequencing.

891

892 **Immunofluorescence staining**

893 HiPSCs and Ros-L cells were fixed in 4% paraformaldehyde in DPBS for 20  
894 min and permeabilized with 1% Triton X-100 for 20 min at room temperature.  
895 After 60 min blocking with 2% normal goat serum, hiPSCs were incubated with  
896 primary antibodies OCT4 (1: 200, Abcam), NANOG (1: 200, Abcam), and  
897 Ros-L cells were incubated with primary antibodies PAX6 (1: 200, Abcam),  
898 SOX2 (1:200, Abcam), NESTIN (1: 200, Abcam), SOX1 (1: 200, Abcam), Zo-1  
899 (1:100, Abcam) and N-CAD (1: 100, Abcam) overnight at 4 °C, then stained

1 900 with secondary antibodies (goat anti rabbit IgG-Cy3 diluted 1: 300 and goat anti  
2 901 mouse IgG-Cy3 diluted 1: 300) for 60 min at room temperature. DAPI (1: 500)  
3  
4 902 was used as counter-staining for nuclei. The images were captured and  
5  
6 903 analyzed with the Olympus IX73 and Image J.  
7

8 904

### 10 905 **Single cell RNA sequencing**

11  
12 906 Cells at indicated time points were collected for single cell RNA-seq and global  
13  
14 907 transcriptome analysis. TrypLE™ Express Enzyme (Gibco, 12604-021) was  
15  
16 908 applied for single cell dissociation. Single-cell RNA-seq library construction  
17  
18 909 was conducted according to an automated pipeline called microwell full-length  
19  
20 910 mRNA amplification and library construction system (MIRALCS) as described  
21  
22 911 previously [78]. 50bp single-end sequencing was performed using the  
23  
24 912 BGISEQ-500 platform.  
25

26 913

### 27 914 **Assay for transposase-accessible chromatin sequencing (ATAC-seq)**

28  
29 915 We profiled open chromatin accessibility sequencing (ATAC-seq) of neural  
30  
31 916 differentiation process for five stages including iPSCs, EB, Ros-E, Ros-L and  
32  
33 917 NPCs samples. ATAC-seq libraries were prepared using a modified protocol  
34  
35 918 based on previous study [79]. Briefly, 50,000 cells were collected for each  
36  
37 919 sample, washed with pre-cooling PBS and resuspended in 50 µl of ice-cold  
38  
39 920 lysis buffer (10 mM Tris-HCl, pH 7.5, 10 mM NaCl, 3 mM MgCl<sub>2</sub>, 0.1% IGEPAL  
40  
41 921 CA-630). Permeabilized cells were resuspended in 50 µl transposase reaction  
42  
43 922 buffer (1× TAG buffer, 2.0 µl Tn5 transposase enzyme) and incubated for 30 min  
44  
45 923 at 37 °C. PCR amplification and size selection (150–500 bp) were performed  
46  
47 924 using Agincourt AMPure XP (Beckman Coulter) and Bioanalyzer 2100  
48  
49 925 (Agilent). Libraries were pooled at equimolar ratios with barcodes and  
50  
51 926 sequenced on BGISEQ-500 platform.  
52

53 927

### 54 928 **Pre-processing and quality control of single cell RNA-seq**

1 929 The original FASTQ data of the 527 samples were aligned to the rRNA  
2 930 database (downloaded from NCBI) to remove rRNAs and the remaining reads  
3  
4 931 were processed with SOAPnuke (version 1.5.3) [80] to trim adaptors and filter  
5  
6 932 out the low-quality reads. The filtered data were aligned to the reference  
7  
8 933 genome (hg19) using hisat2 (HISAT2 version 2.0.1-beta) [81]. Reads were  
9  
10 934 counted using the R package GenomicAlignments [82] (mode='Union',  
11  
12 935 inter.feature=FALSE), and normalized to RPKM with edgeR [83]. Cells were  
13  
14 936 filtered using following parameters: genome mapping rate more than 70%,  
15  
16 937 fraction of reads mapped to mitochondrial genes less than 20%, mRNA  
17  
18 938 mapping rate more than 80%, ERCC ratio less than 10%, and gene number  
19  
20 939 more than 5000. Further, correlation of ERCC among cells was used to  
21  
22 940 evaluate the quality of each cell (threshold=0.9). At last, 445 single cells  
23  
24 941 remained for further analysis in this project.  
25  
26

27 942

### 28 29 943 **Identification of differentially expressed genes**

30  
31 944 Differential expression of genes in iPSCs (n = 71 cells), EB (n = 57 cells),  
32  
33 945 Ros-E (n = 81 cells), Ros-L (n = 92 cells), and NPCs (n = 90 cells) was  
34  
35 946 determined using SCDE (single cell differential expression analysis) [84] with  
36  
37 947 default parameters except requiring a minimum of 100 genes (parameter  
38  
39 948 min.lib.size = 100 to call scde.error.models function). The Z scores and  
40  
41 949 corrected Z scores (cZ) to adjust for the multiple testing were converted into  
42  
43 950 two-tailed p-values and adjusted to control for FDR using pnorm function in R.  
44  
45 951 The significantly differentially expressed genes were selected based on  
46  
47 952 following criteria: adjusted p-value < 0.01 and fold-change > 2.  
48  
49

50 953

### 51 954 **Constructing trajectory using differentially expressed genes**

52  
53 955 Monocle [85] ordering was conducted for all iPSCs, EB, Ros-E, Ros-L and  
54  
55 956 NPCs cells using the set of variable genes with default parameters except we  
56  
57 957 specified reduction\_method = "DDRTree" in the reduceDimension function. The  
58  
59 958 variable genes were selected using the Seurat R package [86].  
60  
61



1 959

2 **960 Analysis of heterogeneity in each cell stage**

3  
4 961 The heterogeneity of each cell stage was determined using Seurat R package  
5 [86] by the normalized expression level of reported transcription factors  
6 962 [86] by the normalized expression level of reported transcription factors  
7  
8 963 (retrieved from AnimalTFDB 2.0) [89]. Briefly, PCs with a p-value less than  
9 964 0.01 were used for cell clustering with reduction.type="pca" and  
10 resolution="1.0". The FindallMarkers function of Seurat package was used to  
11 965 identify marker genes for each cluster using default parameters.  
12 966  
13  
14  
15  
16

17 967

18 **968 ATAC peak calling**

19  
20  
21 969 We aligned ATAC-seq data to hg19 using Bowtie2 [88] and called peaks using  
22 MACS2 [89]. We established a standard peak set by merging all overlapping  
23 970 peaks. The IDR pipeline [90] was used to identify reproducible peaks between  
24  
25 971 two biological replicates. Only peaks with  $IDR \leq 0.05$  were considered  
26 972 reproducible and retained for downstream analysis. Pearson correlation  
27 973 coefficients of two biological replicates at each stage were calculated.  
28  
29 974 Stage-specific peaks were defined as peaks having no overlap with any peaks  
30 975 in other stages. Novel peaks were defined as peaks non-overlapping with  
31 976 previous stages. In the case of iPSCs, all peaks were annotated as novel  
32 977 peaks.  
33  
34  
35  
36  
37  
38  
39  
40  
41  
42

43 979

44 **980 Targets assignment of ATAC peaks**

45  
46 981 For reproducible peaks, we applied HOMER [91] to assign putative targets for  
47 982 peaks. For stage-specific peaks, ChIPseeker [92] was used for putative target  
48 983 assignment. In both strategies, the putative target of a certain peak is defined  
49 984 as the gene with TSS closest to the peak summit location.  
50  
51  
52  
53

54 985

55 **986 GO term and KEGG enrichment analysis**

56  
57  
58 987 Lists of genes were analysed using DAVID [93,94] and the BH method was  
59 988 used for multiple test correction. GO terms with a FDR less than 0.01 or 0.05  
60  
61  
62  
63  
64  
65

1 989 were considered as significantly enriched. Target genes of stage-specific ATAC  
2 990 peaks were analysed using the R package, clusterProfiler [95], in which an  
3  
4 991 adjusted p-value of 0.05 was used to identify significantly enriched GO and  
5  
6 992 KEGG terms associated with each set of peaks.  
7

8 993

### 10 994 **Regulatory network construction**

11  
12 995 The scRNA-seq profiles among each cell types were compared using SCDE  
13  
14 996 package [84]. TFs significantly differentially expressed, with adjusted p-value  
15  
16 997 threshold of 0.05, among neighboring cell types were submitted to STRING  
17  
18 998 database [64] to infer regulatory networks based on known interaction  
19  
20 999 relationships (supported by data from curated databases, experiments and  
21  
22 1000 text-mining). TFs without any interactions with other proteins were removed  
23  
24 1001 from the network. To select key regulators, we used a threshold of 5 and all  
25  
26 1002 TFs with number of interactions above the threshold were considered as key  
27  
28 1003 regulators.  
29  
30

31 1004

### 33 1005 **Putative targets prediction, GO term and KEGG enrichment analysis**

34  
35 1006 The target prediction and enrichment analyses were performed using the  
36  
37 1007 FIMO [96] and GREAT [97] packages, respectively. Briefly, the peak files in a  
38  
39 1008 certain stage were scanned for the presence or absence of TF motifs, which  
40  
41 1009 were downloaded from the Jasper database [98]. Genes with a TSS closest  
42  
43 1010 to TF motif-containing peaks were considered as putative targets of certain  
44  
45 1011 TFs.  
46  
47

48 1012

### 50 1013 **Construction of cellular communication network**

51  
52 1014 The ligand-receptor interaction relationships were downloaded from the  
53  
54 1015 database, IUPHAR/BPS Guide to PHARMACOLOGY [98], and the Database  
55  
56 1016 of Ligand-Receptor Partners (DLRP) [65, 100]. The average expression level  
57  
58 1017 of TPM of 1 was used as a threshold. Ligands and receptors above the  
59  
60 1018 threshold were considered as expressed in the corresponding cluster.  
61

1 1019 Adjusted  $P$  value of 0.05 was used as a threshold to identify ligands/receptors  
2 1020 specifically expressed in a subpopulation. The R package Circlize [101] was  
3  
4 1021 used to visualize the interactions.  
5

6 1022

### 8 1023 **Motif enrichment analysis**

10 1024 Motifs enriched in each set of ATAC peaks were identified using  
11 findMotifsGenome.pl from HOMER [91] using following parameters: -size  
12 1025 -100,100 -len 4,5,6,7,8,9,10,11,12.  
13  
14 1026  
15

16 1027

### 18 1028 **Additional files**

20 1029 **Additional file 1: Figure S1.** Quality control of ATAC-seq.

22 1030 **Additional file 2: Figure S2.** Dynamics of gained and lost peaks during neural  
23 differentiation.  
24 1031

25 1032 **Additional file 3: Figure S3.** Stage-specific features of cis-regulatory  
26 elements during neural differentiation.  
27 1033

28 1034 **Additional file 4: Figure S4.** Quality control of scRNA-seq.

29 1035 **Additional file 5: Figure S5.** Subgroups identification and key transcriptomic  
30 features within Fib stage.  
31 1036

32 1037 **Additional file 6: Figure S6.** Subgroups identification and key transcriptomic  
33 features within EB stage.  
34 1038

35 1039 **Additional file 7: Figure S7.** Subgroups identification and key transcriptomic  
36 features within Ros-L stage.  
37 1040

38 1041 **Additional file 8: Figure S8.** Subgroups identification and key transcriptomic  
39 features within NPCs stage.  
40 1042

41 1043 **Additional file 9: Figure S9.** Expression pattern of selected transcription  
42 factors (TFs) within rosettes (Ros-E and Ros-L) stage.  
43 1044

44 1045 **Additional file 10: Figure S10.** Differentially expressed transcription factors  
45 (TFs) between Ros-E2 and Ros-L3.  
46 1046

47 1047 **Additional file 11: Figure S11.** Differentially expressed transcription factors  
48 (TFs) between Ros-L3 and NPC1.  
49 1048  
50  
51  
52  
53  
54  
55  
56  
57  
58  
59  
60  
61  
62  
63  
64  
65

1 1049 **Additional file 12: Figure S12.** Key regulators during neural differentiation.

2 1050 **Additional file 13: Figure S13.** GO term and KEGG enrichment analysis of  
3 selected transcription factors (TFs) targets.  
4

5 1051  
6 1052 **Additional file 14: Figure S14.** Putative signaling between expressed  
7 receptors and their ligands in EB subsets.  
8

9 1053  
10 1054 **Additional file 15: Figure S15.** Putative signaling between expressed  
11 receptors and their ligands in Ros-E subsets.  
12

13 1055  
14 1056 **Additional file 16: Figure S16.** Putative signaling between expressed  
15 receptors and their ligands in NPC subsets.  
16

17 1057  
18 1058 **Additional file 17: Figure S17.** Transcription factor motifs enriched in stage  
19 specific peaks.  
20

21 1059  
22 1060 **Additional file 18: Figure S18.** Validation of neural differentiation in different  
23 genetic background cell lines  
24

25 1061  
26 1062 **Additional file 19 : Table S1.** TFs differentially expressed among  
27 neighbouring cell subsets.  
28

29 1063  
30 1064 **Additional file 20: Table S2.** Putative targets of selected regulators.

31 1065 **Additional file 21: Table S3.** Subpopulations interaction networks.

32 1066 **Additional file 22: Table S4.** Differentially expressed receptors and ligands  
33 among Ros-L subpopulations.  
34

35 1067  
36 1068  
37

#### 38 1069 **Availability of data and materials**

39 1070 The detailed protocol of neural differentiation and bioinformatics pipeline was  
40 available in protocol. io (DOI: [dx.doi.org/10.17504/protocols.io.ntrdem6](https://doi.org/10.17504/protocols.io.ntrdem6) and  
41 DOI: [dx.doi.org/10.17504/protocols.io.ntpdemn](https://doi.org/10.17504/protocols.io.ntpdemn)). The sequencing raw data  
42 were in process of submission (Sequence Read Archive submission ID:  
43 SUB4333714).  
44

45 1075  
46

#### 47 1076 **Competing interests**

48 1077 The authors declare that they have no competing interests.  
49

50 1078  
51  
52  
53  
54  
55  
56  
57  
58  
59  
60  
61  
62  
63  
64  
65

1079 **Authors' contributions**

1080 X.X. and Z.G. conceived and designed the project. Z.S., D.C., Q.W., S.W. and  
1081 Q.D. conducted the majority of experiments and data analysis. L.W., X.D., S.W.  
1082 and J.Z. performed computational analyses and prepared figures. C.L.  
1083 participated in validation experiments and assisted with figure preparation for  
1084 revision. D.Z., X.C. and F.C. contributed to sample collection. X.X., Z.G. and  
1085 H.Y. supervised the project. X.L. contributed to the design of the revision, and  
1086 jointly supervised the validation work. Z.S., D.C., Q.W., Z.G. and X.X. prepared  
1087 the manuscript. S.Z., L.L. and J.L.F. contributed to the discussion and revision  
1088 of the manuscript. All authors read and approved the final manuscript.

1090 **Acknowledgements**

1091 We thank Tao Tan for support with antibody, Shiping Liu for bioinformatics help,  
1092 Guibo Li for technical help with preparation of single cell RNA-seq libraries and  
1093 other members of Cell and Developmental Biology Lab for discussions and  
1094 support. This work was supported by Shenzhen Engineering Laboratory for  
1095 Innovative Molecular Diagnostics [grant number DRC-SZ [2016] 884] funded  
1096 by Development and Reform Commission of Shenzhen Municipality; and  
1097 Shenzhen Key Laboratory of Neurogenomics (CXB201108250094A) funded  
1098 by Science, Technology and Innovation Commission of Shenzhen Municipality.  
1099 Dongsheng Chen is supported by China Postdoctoral Science Foundation  
1100 (grant number 2017M622795).

1102 **References**

- 1103 1. Harmacek L, Watkins-Chow DE, Chen J, Jones KL, Pavan WJ, Salbaum  
1104 JM, et al. A unique missense allele of BAF155, a core BAF chromatin  
1105 remodeling complex protein, causes neural tube closure defects in mice.  
1106 Dev Neurobiol. 2014;74:483-97.
- 1107 2. Foster WH, Langenbacher A, Gao C, Chen J, Wang Y. Nuclear  
1108 phosphatase PPM1G in cellular survival and neural development. Dev

1 1109 Dyn. 2013;242:1101-9.

2 1110 3. Wilde JJ, Petersen JR, Niswander L. Genetic, Epigenetic, and

3 1111 Environmental Contributions to Neural Tube Closure. *Annu Rev Genet.*

4 1112 2014;48:583-611.

5 1113 4. Wen Z, Nguyen HN, Guo Z, Lalli MA, Wang X, Su Y, et al. Synaptic

6 1114 dysregulation in a human iPS cell model of mental disorders. *Nature.*

7 1115 2014;515:414-8.

8 1116 5. Tao Y, Zhang SC. Neural Subtype Specification from Human Pluripotent

9 1117 Stem Cells. *Cell Stem Cell.* 2016. p. 573-86.

10 1118 6. Streit A, Berliner AJ, Papanayotou C, Slrulnik A, Stern CD. Initiation of

11 1119 neural induction by FGF signalling before gastrulation. *Nature.*

12 1120 2000;406:74-8.

13 1121 7. Muñoz-Sanjuán I, Brivanlou AH. Neural induction, the default model and

14 1122 embryonic stem cells. *Nat Rev Neurosci.* 2002;3:271-80.

15 1123 8. Zhang S-C, Wernig M, Duncan ID, Brüstle O, Thomson JA. In vitro

16 1124 differentiation of transplantable neural precursors from human embryonic

17 1125 stem cells. *Nat Biotechnol.* 2001;19:1129-33.

18 1126 9. Xu X, Hou Y, Yin X, Bao L, Tang A, Song L, et al. Single-cell exome

19 1127 sequencing reveals single-nucleotide mutation characteristics of a kidney

20 1128 tumor. *Cell.* 2012;148:886-95.

21 1129 10. Hou Y, Song L, Zhu P, Zhang B, Tao Y, Xu X, et al. Single-cell exome

22 1130 sequencing and monoclonal evolution of a JAK2-negative

23 1131 myeloproliferative neoplasm. *Cell.* 2012;148:873-85.

24 1132 11. Johnson MB, Wang PP, Atabay KD, Murphy EA, Doan RN, Hecht JL, et al.

25 1133 Single-cell analysis reveals transcriptional heterogeneity of neural

26 1134 progenitors in human cortex. *Nat Neurosci.* 2015;18:637-46.

27 1135 12. Villani A-C, Satija R, Reynolds G, Sarkizova S, Shekhar K, Fletcher J, et

28 1136 al. Single-cell RNA-seq reveals new types of human blood dendritic cells,

29 1137 monocytes, and progenitors. *Science (80- ).* 2017;356:eaah4573.

30 1138 13. La Manno G, Gyllborg D, Codeluppi S, Nishimura K, Salto C, Zeisel A, et

1 1139 al. Molecular Diversity of Midbrain Development in Mouse, Human, and  
2 1140 Stem Cells. *Cell*. 2016;167:566-580.e19.  
3  
4 1141 14. Buettner F, Natarajan KN, Casale FP, Proserpio V, Scialdone A, Theis FJ,  
5  
6 1142 et al. Computational analysis of cell-to-cell heterogeneity in single-cell  
7  
8 1143 RNA-sequencing data reveals hidden subpopulations of cells. *Nat*  
9  
10 1144 *Biotechnol*. 2015;33:155-60.  
11  
12 1145 15. Li H, Courtois ET, Sengupta D, Tan Y, Chen KH, Goh JJL, et al.  
13  
14 1146 Reference component analysis of single-cell transcriptomes elucidates  
15  
16 1147 cellular heterogeneity in human colorectal tumors. *Nat Genet*.  
17  
18 1148 2017;49:708-18.  
19  
20  
21 1149 16. Kim DS, Lee DR, Kim HS, Yoo JE, Jung SJ, Lim BY, et al. Highly pure and  
22  
23 1150 expandable PSA-NCAM-positive neural precursors from human ESC and  
24  
25 1151 iPSC-derived neural rosettes. *PLoS One*. 2012;7.  
26  
27 1152 17. Ardhanareeswaran K, Mariani J, Coppola G, Abyzov A, Vaccarino FM.  
28  
29 1153 Human induced pluripotent stem cells for modelling neurodevelopmental  
30  
31 1154 disorders. *Nat. Rev. Neurol*. 2017. p. 265-78.  
32  
33 1155 18. Chambers SMSM, Fasano CACA, Papapetrou EP, Tomishima M,  
34  
35 1156 Sadelain M, Studer L. Highly efficient neural conversion of human ES and  
36  
37 1157 iPS cells by dual inhibition of SMAD signaling. *Nat Biotechnol*.  
38  
39 1158 2009;27:275-80.  
40  
41 1159 19. Dolmetsch R, Geschwind DH. The human brain in a dish: The promise of  
42  
43 1160 iPSC-derived neurons. *Cell*. 2011. p. 831-4.  
44  
45 1161 20. Kriks S, Shim JW, Piao J, Ganat YM, Wakeman DR, Xie Z, et al.  
46  
47 1162 Dopamine neurons derived from human ES cells efficiently engraft in  
48  
49 1163 animal models of Parkinson's disease. *Nature*. 2011;480:547-51.  
50  
51  
52 1164 21. Miller JD, Ganat YM, Kishinevsky S, Bowman RL, Liu B, Tu EY, et al.  
53  
54 1165 Human iPSC-based modeling of late-onset disease via progerin-induced  
55  
56 1166 aging. *Cell Stem Cell*. 2013;13:691-705.  
57  
58 1167 22. Maroof AM, Keros S, Tyson JA, Ying S-W, Ganat YM, Merkle FT, et al.  
59  
60 1168 Directed Differentiation and Functional Maturation of Cortical Interneurons

1 1169 from Human Embryonic Stem Cells. *Cell Stem Cell*. 2013;12:559-72.

2 1170 23. Shi Y, Kirwan P, Smith J, Robinson HPC, Livesey FJ. Human cerebral

3 cortex development from pluripotent stem cells to functional excitatory

4 1171 synapses. *Nat Neurosci*. 2012;15:477-86.

5 1172

6 1173 24. Hueber SD, Bezdán D, Henz SR, Blank M, Wu H, Lohmann I.

7 Comparative analysis of Hox downstream genes in *Drosophila*.

8 1174 *Development* 2007;134:381-92.

9 1175

10 1176 25. Chen D, Jiang S, Ma X, Li F. TFBSbank: a platform to dissect the big data

11 of protein-DNA interaction in human and model species. *Nucleic Acids*

12 1177 *Res*. Oxford University Press; 2017;45:D151-7.

13 1178

14 1179 26. Thomas S, Li X-Y, Sabo PJ, Sandstrom R, Thurman RE, Canfield TK, et

15 al. Dynamic reprogramming of chromatin accessibility during *Drosophila*

16 1180 embryo development. *Genome Biol*. BioMed Central; 2011;12:R43.

17 1181

18 1182 27. Buenrostro JD, Wu B, Litzenburger UM, Ruff D, Gonzales ML, Snyder MP,

19 et al. Single-cell chromatin accessibility reveals principles of regulatory

20 1183 variation. *Nature*. Nature Publishing Group; 2015;523:486-90.

21 1184

22 1185 28. Jin F, Li Y, Dixon JR, Selvaraj S, Ye Z, Lee AY, et al. A high-resolution

23 1186 map of the three-dimensional chromatin interactome in human cells.

24 1187 *Nature*. Nature Publishing Group; 2013;503:290-4.

25 1188

26 1189 29. Corces MR, Buenrostro JD, Wu B, Greenside PG, Chan SM, Koenig JL,

27 et al. Lineage-specific and single-cell chromatin accessibility charts

28 1190 human hematopoiesis and leukemia evolution. *Nat Genet*. Nature

29 1191 Publishing Group; 2016;48:1193-203.

30 1192

31 1193 30. Picelli S, Björklund ÅK, Faridani OR, Sagasser S, Winberg G, Sandberg R.

32 Smart-seq2 for sensitive full-length transcriptome profiling in single cells.

33 1194 *Nat Methods*. 2013;10:1096-100.

34 1195

35 1196 31. Kimura-Yoshida C, Mochida K, Ellwanger K, Niehrs C, Matsuo I. Fate

36 Specification of Neural Plate Border by Canonical Wnt Signaling and

37 1197 *Grhl3* is Crucial for Neural Tube Closure. *EBioMedicine*. 2015;2:513-27.

38 1198

39 1198 32. Nikolopoulou E, Galea GL, Rolo A, Greene NDE, Copp AJ. Neural tube



1 1199 closure: cellular, molecular and biomechanical mechanisms.  
2 1200 Development. 2017;144:552-66.  
3  
4 1201 33. Zhang J, Hagopian-Donaldson S, Serbedzija G, Elsemore J,  
5  
6 1202 Plehn-Dujowich D, McMahon AP, et al. Neural tube, skeletal and body  
7  
8 1203 wall defects in mice lacking transcription factor AP-2. Nature.  
9  
10 1204 1996;381:238-41.  
11  
12 1205 34. Hayon Y, Dashevsky O, Shai E, Varon D, Leker RR. Platelet  
13  
14 1206 microparticles promote neural stem cell proliferation, survival and  
15  
16 1207 differentiation. J Mol Neurosci. 2012;47:659-65.  
17  
18 1208 35. Lee HO, Levorse JM, Shin MK. The endothelin receptor-B is required for  
19  
20 1209 the migration of neural crest-derived melanocyte and enteric neuron  
21  
22 1210 precursors. Dev Biol. 2003;259:162-75.  
23  
24 1211 36. Li H, Horns F, Wu B, Xie Q, Li J, Li T, et al. Classifying Drosophila  
25  
26 1212 Olfactory Projection Neuron Subtypes by Single-Cell RNA Sequencing.  
27  
28 1213 Cell. 2017;171:1206-1220.e22.  
29  
30 1214 37. Vierbuchen T, Ostermeier A, Pang ZP, Kokubu Y, Südhof TC, Wernig M.  
31  
32 1215 Direct conversion of fibroblasts to functional neurons by defined factors.  
33  
34 1216 Nature. 2010; 25:1035-1041.  
35  
36 1217 38. Yoo AS, Sun AX, Li L, Shcheglovitov A, Portmann T, Li Y, et al.  
37  
38 1218 MicroRNA-mediated conversion of human fibroblasts to neurons. Nature.  
39  
40 1219 2011; 476:228-31.  
41  
42 1220 39. Kim D-Y, Hwang I, Muller FL, Paik J-H. Functional regulation of FoxO1 in  
43  
44 1221 neural stem cell differentiation. Cell Death Differ. 2015;22:2034-45.  
45  
46 1222 40. Cameron DA, Pennimpede T, Petkovich M. Tulp3 is a critical repressor of  
47  
48 1223 Mouse hedgehog signaling. Dev Dyn. 2009;238:1140-9.  
49  
50 1224 41. Jin Z, Liu L, Bian W, Chen Y, Xu G, Cheng L, et al. Different transcription  
51  
52 1225 factors regulate nestin gene expression during P19 cell neural  
53  
54 1226 differentiation and central nervous system development. J Biol Chem.  
55  
56 1227 2009;284:8160-73.  
57  
58 1228 42. Elkabetz Y, Panagiotakos G, Al Shamy G, Socci ND, Tabar V, Studer L.

1 1229 Human ES cell-derived neural rosettes reveal a functionally distinct early  
2 1230 neural stem cell stage. *Genes Dev.* 2008;22:152-65.  
3  
4 1231 43. Cheung M, Briscoe J. Neural crest development is regulated by the  
5  
6 1232 transcription factor Sox9. *Development.* 2003;130:5681-93.  
7  
8 1233 44. Scott CE, Wynn SL, Sesay A, Cruz C, Cheung M, Gavira MVG, et al.  
9  
10 1234 SOX9 induces and maintains neural stem cells. *Nat Neurosci.*  
11  
12 1235 2010;13:1181-9.  
13  
14 1236 45. Betters E, Liu Y, Kjaeldgaard A, Sundstrom E, Garcia-Castro MI. Analysis  
15  
16 1237 of early human neural crest development. *Dev Biol.* 2010;344:578-92.  
17  
18 1238 46. Wang C, Kam RKT, Shi W, Xia Y, Chen X, Cao Y, et al. The  
19  
20 1239 Proto-oncogene transcription factor Ets1 regulates neural crest  
21  
22 1240 development through histone deacetylase 1 to mediate output of bone  
23  
24 1241 morphogenetic protein signaling. *J Biol Chem.* 2015;290:21925-38.  
25  
26 1242 47. Mulligan KA, Cheyette BNR. Wnt signaling in vertebrate neural  
27  
28 1243 development and function. *J. Neuroimmune Pharmacol.* 2012. p. 774–87.  
29  
30 1244 48. Ille F, Sommer L. Wnt signaling: multiple functions in neural development.  
31  
32 1245 *Cell. Mol. Life Sci.* 2005;62:1100-8.  
33  
34 1246 49. Wang J, Jenjaroenpun P, Bhinge A, Angarica VE, Del Sol A, Nookaew I,  
35  
36 1247 et al. Single-cell gene expression analysis reveals regulators of distinct  
37  
38 1248 cell subpopulations among developing human neurons. *Genome Res.*  
39  
40 1249 2017;27:1783-94.  
41  
42 1250 50. Wang Y, Ristevski S, Harley VR. SOX13 exhibits a distinct spatial and  
43  
44 1251 temporal expression pattern during chondrogenesis, neurogenesis, and  
45  
46 1252 limb development. *J Histochem Cytochem.* 2006;54:1327-33.  
47  
48 1253 51. Ji EH, Kim J. SoxD Transcription Factors: Multifaceted Players of Neural  
49  
50 1254 Development. *Int J stem cells.* Korean Society for Stem Cell Research;  
51  
52 1255 2016;9:3-8.  
53  
54 1256 52. Koshida R, Oishi H, Hamada M, Takei Y, Takahashi S. MafB is required  
55  
56 1257 for development of the hindbrain choroid plexus. *Biochem Biophys Res*  
57  
58 1258 *Commun.* 2017;483:288-93.  
59  
60  
61  
62  
63  
64  
65

1 1259 53. Rogers CD, Phillips JL, Bronner ME. Elk3 is essential for the progression  
2 1260 from progenitor to definitive neural crest cell. *Dev Biol.* 2013;374:255-63.  
3  
4 1261 54. Noisa P, Lund C, Kanduri K, Lund R, La H. Notch signaling regulates the  
5 1262 differentiation of neural crest from human pluripotent stem cells. *J Cell Sci.*  
6 1263 2014;127:2083-94.  
7  
8 1264 55. Zhang J, Zheng B, Zhou P-P, Zhang R-N, He M, Yang Z, et al. Vascular  
9 1265 calcification is coupled with phenotypic conversion of vascular smooth  
10 1266 muscle cells through Klf5-mediated transactivation of the Runx2 promoter.  
11 1267 *Biosci Rep.* 2014;34:e00148.  
12  
13 1268 56. Dougherty M, Kamel G, Grimaldi M, Gfrerer L, Shubinets V, Ethier R, et al.  
14 1269 Distinct requirements for wnt9a and irf6 in extension and integration  
15 1270 mechanisms during zebrafish palate morphogenesis. *Development.*  
16 1271 2013;140:76-81.  
17  
18 1272 57. Hernandez-Lagunas L, Choi IF, Kaji T, Simpson P, Hershey C, Zhou Y, et  
19 1273 al. Zebrafish narrowminded disrupts the transcription factor prdm1 and is  
20 1274 required for neural crest and sensory neuron specification. *Dev Biol.*  
21 1275 2005;278:347-57.  
22  
23 1276 58. Simões-Costa M, Bronner ME. Insights into neural crest development and  
24 1277 evolution from genomic analysis. *Genome Res.* 2013. p. 1069-80.  
25  
26 1278 59. Viales RR, Diotel N, Ferg M, Armant O, Eich J, Alunni A, et al. The  
27 1279 Helix-Loop-Helix Protein Id1 Controls Stem Cell Proliferation During  
28 1280 Regenerative Neurogenesis in the Adult Zebrafish Telencephalon. *Stem*  
29 1281 *Cells.* 2015;33:892-903.  
30  
31 1282 60. Li L, Candelario KM, Thomas K, Wang R, Wright K, Messier A, et al.  
32 1283 Hypoxia Inducible Factor-1 (HIF-1 ) Is Required for Neural Stem Cell  
33 1284 Maintenance and Vascular Stability in the Adult Mouse SVZ. *J Neurosci.*  
34 1285 2014;34:16713-9.  
35  
36 1286 61. Ahmed M, Xu J, Xu P-X. EYA1 and SIX1 drive the neuronal  
37 1287 developmental program in cooperation with the SWI/SNF  
38 1288 chromatin-remodeling complex and SOX2 in the mammalian inner ear.

1 1289 Development. *Company of Biologists*; 2012;139:1965-77.

2 1290 62. Abe H, Okazawa M, Nakanishi S. The Etv1/Er81 transcription factor

3 1291 orchestrates activity-dependent gene regulation in the terminal maturation

4 1292 program of cerebellar granule cells. *Proc Natl Acad Sci U S A. National*

5 1293 *Academy of Sciences*; 2011;108:12497-502.

6 1294 63. Dominguez MH, Ayoub AE, Rakic P. POU-III transcription factors (Brn1,

7 1295 Brn2, and Oct6) influence neurogenesis, molecular identity, and migratory

8 1296 destination of upper-layer cells of the cerebral cortex. *Cereb Cortex.*

9 1297 2013;23:2632-43.

10 1298 64. Szklarczyk D, Morris JH, Cook H, Kuhn M, Wyder S, Simonovic M, et al.

11 1299 The STRING database in 2017: quality-controlled protein-protein

12 1300 association networks, made broadly accessible. *Nucleic Acids Res.*

13 1301 *Oxford University Press*; 2017;45:D362-8.

14 1302 65. Pavličev M, Wagner GP, Chavan AR, Owens K, Maziarz J, Dunn-Fletcher

15 1303 C, et al. Single-cell transcriptomics of the human placenta: Inferring the

16 1304 cell communication network of the maternal-fetal interface. *Genome Res.*

17 1305 2017;27:349-61.

18 1306 66. Stern CD. Initial patterning of the central nervous system: How many

19 1307 organizers? *Nat Rev Neurosci.* 2001;2:92-8.

20 1308 67. Smith JR, Vallier L, Lupo G, Alexander M, Harris WA, Pedersen RA.

21 1309 Inhibition of Activin/Nodal signaling promotes specification of human

22 1310 embryonic stem cells into neuroectoderm. *Dev Biol.* 2008;313:107-17.

23 1311 68. Schwarz M, Alvarez-Bolado G, Dressler G, Urbánek P, Busslinger M,

24 1312 Gruss P. Pax2/5 and Pax6 subdivide the early neural tube into three

25 1313 domains. *Mech Dev.* 1999;82:29-39.

26 1314 69. Brouns MR, de Castro SCP, Terwindt-Rouwenhorst EA, Massa V,

27 1315 Hekking JW, Hirst CS, et al. Over-expression of Grhl2 causes spina bifida

28 1316 in the Axial defects mutant mouse. *Hum Mol Genet.* 2011;20:1536-46.

29 1317 70. Nikitina N, Tong L, Bronner ME. Ancestral network module regulating

30 1318 prdm1 expression in the lamprey neural plate border. *Dev Dyn.*

1 1319 2011;240:2265-71.

2 1320 71. Vincent SD, Dunn NR, Sciammas R, Shapiro-Shalef M, Davis MM,

3 1321 Calame K, et al. The zinc finger transcriptional repressor Blimp1/Prdm1 is

4 1322 dispensable for early axis formation but is required for specification of

5 1323 primordial germ cells in the mouse. *Development*. 2005;132:1315-25.

6 1324 72. Timmer JR, Wang C, Niswander L. BMP signaling patterns the dorsal and

7 1325 intermediate neural tube via regulation of homeobox and helix-loop-helix

8 1326 transcription factors. *Development*. 2002;129:2459-72.

9 1327 73. Burns CJ, Zhang J, Brown EC, Van Bibber AM, Van Es J, Clevers H, et al.

10 1328 Investigation of Frizzled-5 during embryonic neural development in mouse.

11 1329 *Dev Dyn*. 2008;237:1614-26.

12 1330 74. Miyake A, Nakayama Y, Konishi M, Itoh N. Fgf19 regulated by Hh

13 1331 signaling is required for zebrafish forebrain development. *Dev Biol*.

14 1332 2005;288:259-75.

15 1333 75. Qu Q, Sun G, Murai K, Ye P, Li W, Asuelime G, et al. Wnt7a Regulates

16 1334 Multiple Steps of Neurogenesis. *Mol Cell Biol*. 2013;33:2551-9.

17 1335 76. Hsu Y-C, Lee D-C, Chen S-L, Liao W-C, Lin J-W, Chiu W-T, et al.

18 1336 Brain-specific 1B promoter of FGF1 gene facilitates the isolation of neural

19 1337 stem/progenitor cells with self-renewal and multipotent capacities. *Dev*

20 1338 *Dyn*. 2009;238:302-14.

21 1339 77. Okita K, Matsumura Y, Sato Y, Okada A, Morizane A, Okamoto S, et al. A

22 1340 more efficient method to generate integration-free human iPS cells. *Nat*.

23 1341 *Methods*. 2011;8:409-12.

24 1342 78. Wu L, Zhang X, Zhao Z, Wang L, Li B, Li G, et al. Full-length single-cell

25 1343 RNA-seq applied to a viral human cancer: Applications to HPV expression

26 1344 and splicing analysis in HeLa S3 cells. *Gigascience*. 2015;4: 51.

27 1345 79. Buenrostro JD, Giresi PG, Zaba LC, Chang HY, Greenleaf WJ.

28 1346 Transposition of native chromatin for fast and sensitive epigenomic

29 1347 profiling of open chromatin, DNA-binding proteins and nucleosome

30 1348 position. *Nat. Methods*. 2013;10:1213-8.

1 1349 80. Chen Y, Chen Y, Shi C, Huang Z, Zhang Y, Li S, et al. SOAPnuke: a  
2 1350 MapReduce acceleration-supported software for integrated quality control  
3 1351 and preprocessing of high-throughput sequencing data. *Gigascience*  
4 1352 2018;7:1-6.  
5  
6  
7  
8 1353 81. Kim D, Langmead B, Salzberg SL. HISAT: a fast spliced aligner with low  
9 1354 memory requirements. *Nat. Methods*; 2015;12:357-60.  
10  
11 1355 82. Lawrence M, Huber W, Pagès H, Aboyoun P, Carlson M, Gentleman R,  
12 1356 et al. Software for Computing and Annotating Genomic Ranges. Prlic A,  
13 1357 editor. *PLoS Comput. Biol.* 2013;9:e1003118.  
14  
15 1358 83. Robinson MD, McCarthy DJ, Smyth GK. edgeR: a Bioconductor  
16 1359 package for differential expression analysis of digital gene expression  
17 1360 data. *Bioinformatics* 2010;26:139-40.  
18  
19 1361 84. Kharchenko P V, Silberstein L, Scadden DT. Bayesian approach to  
20 1362 single-cell differential expression analysis. *Nat. Methods* 2014;11:740-2.  
21  
22 1363 85. Trapnell C, Cacchiarelli D, Grimsby J, Pokharel P, Li S, Morse M, et al.  
23 1364 The dynamics and regulators of cell fate decisions are revealed by  
24 1365 pseudotemporal ordering of single cells. *Nat. Biotechnol.* 2014;32:381-6.  
25  
26 1366 86. Satija R, Farrell JA, Gennert D, Schier AF, Regev A. Spatial  
27 1367 reconstruction of single-cell gene expression data. *Nat. Biotechnol.*  
28 1368 2015;33:495-502.  
29  
30 1369 87. Zhang H-M, Liu T, Liu C-J, Song S, Zhang X, Liu W, et al. AnimalTFDB  
31 1370 2.0: a resource for expression, prediction and functional study of animal  
32 1371 transcription factors. *Nucleic Acids Res.* 2015;43:D76-81.  
33  
34 1372 88. Langmead B, Salzberg SL. Fast gapped-read alignment with Bowtie 2.  
35 1373 *Nat. Methods.* 2012;9:357-9.  
36  
37 1374 89. Zhang Y, Liu T, Meyer CA, Eeckhoutte J, Johnson DS, Bernstein BE, et  
38 1375 al. Model-based Analysis of ChIP-Seq (MACS). *Genome Biol. BioMed*  
39 1376 *Central*; 2008;9:R137.  
40  
41 1377 90. Li Q, Brown JB, Huang H, Bickel PJ. Measuring reproducibility of  
42 1378 high-throughput experiments. *Ann. Appl. Stat. Institute of Mathematical*

1 1379 Statistics; 2011;5:1752-79.

2 1380 91. Heinz S, Benner C, Spann N, Bertolino E, Lin YC, Laslo P, et al. Simple

3

4 1381 Combinations of Lineage-Determining Transcription Factors Prime

5

6 1382 cis-Regulatory Elements Required for Macrophage and B Cell Identities.

7

8 1383 Mol. Cell 2010;38:576-89.

9

10 1384 92. Yu G, Wang L-G, He Q-Y. ChIPseeker: an R/Bioconductor package for

11

12 1385 ChIP peak annotation, comparison and visualization. Bioinformatics

13

14 1386 2015;31:2382-3.

15

16 1387 93. Huang DW, Sherman BT, Lempicki RA. Systematic and integrative

17

18 1388 analysis of large gene lists using DAVID bioinformatics resources. Nat.

19

20 1389 Protoc. 2009;4:44-57.

21

22 1390 94. Huang DW, Sherman BT, Lempicki RA. Bioinformatics enrichment tools:

23

24 1391 paths toward the comprehensive functional analysis of large gene lists.

25

26 1392 Nucleic Acids Res. 2009;37:1-13.

27

28 1393 95. Yu G, Wang L-G, Han Y, He Q-Y. clusterProfiler: an R Package for

29

30 1394 Comparing Biological Themes Among Gene Clusters. Omi. A J. Integr.

31

32 1395 Biol. 2012;16:284-7.

33

34 1396 96. Grant CE, Bailey TL, Noble WS. FIMO: scanning for occurrences of a

35

36 1397 given motif. Bioinformatics Oxford University Press; 2011;27:1017-8.

37

38 1398 97. McLean CY, Bristor D, Hiller M, Clarke SL, Schaar BT, Lowe CB, et al.

39

40 1399 GREAT improves functional interpretation of cis-regulatory regions. Nat.

41

42 1400 Biotechnol. 2010;28:495-501.

43

44 1401 98. Sandelin A, Alkema W, Engström P, Wasserman WW, Lenhard B.

45

46 1402 JASPAR: an open-access database for eukaryotic transcription factor

47

48 1403 binding profiles. Nucleic Acids Res. Oxford University Press;

49

50 1404 2004;32:D91-4.

51

52 1405 99. Harding SD, Sharman JL, Faccenda E, Southan C, Pawson AJ, Ireland

53

54 1406 S, et al. The IUPHAR/BPS Guide to PHARMACOLOGY in 2018:

55

56 1407 updates and expansion to encompass the new guide to

57

58 1408 IMMUNOPHARMACOLOGY. Nucleic Acids Res. 2018;46:D1091-106.

59

60

61

62

63

64

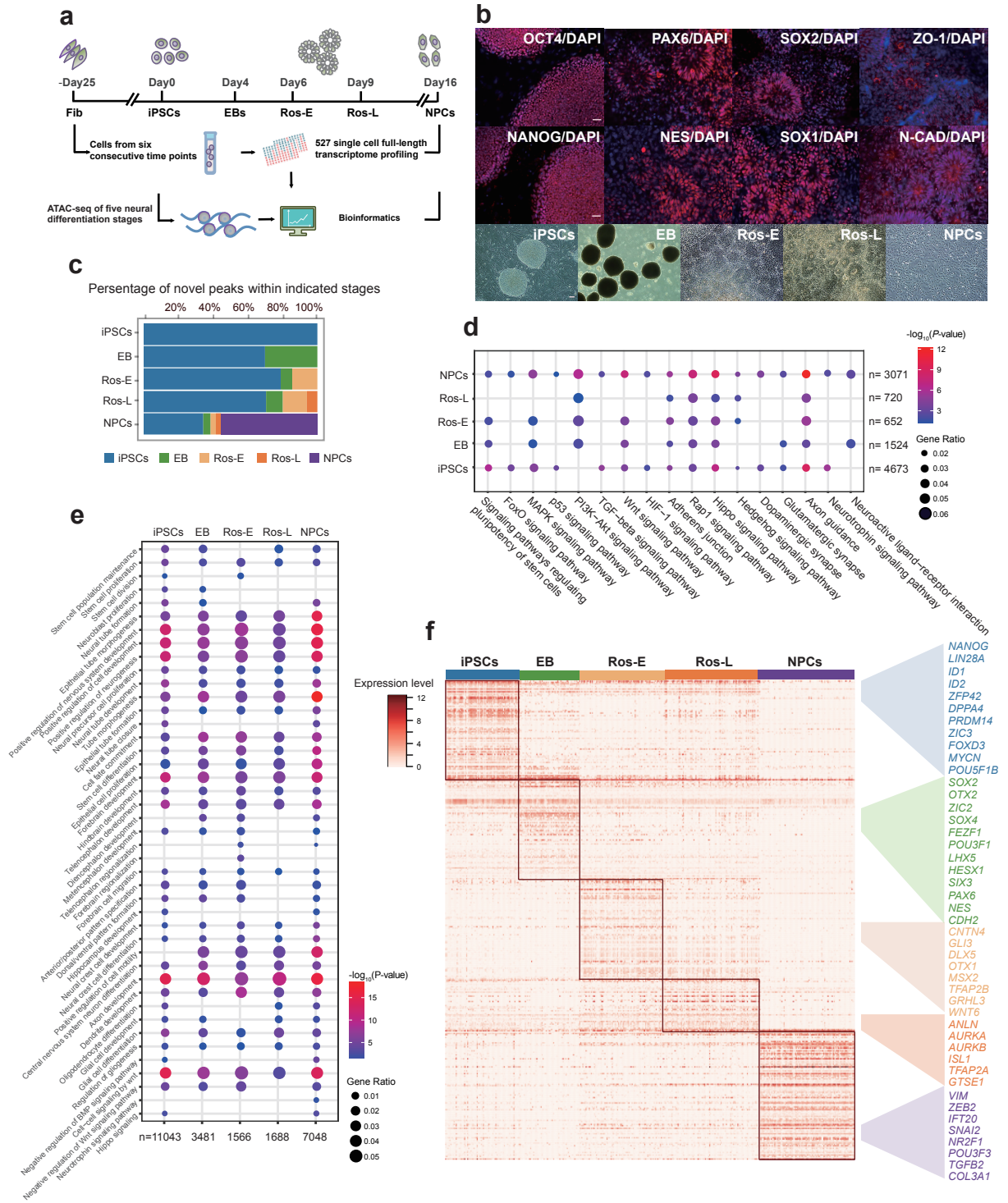
65

1 1409 100.Salwinski L, Miller CS, Smith AJ, Pettit FK, Bowie JU, Eisenberg D. The  
2 Database of Interacting Proteins: 2004 update. Nucleic Acids Res.  
3 1410  
4 2004;32:449D-451.  
5 1411  
6 1412 101.Gu Z, Gu L, Eils R, Schlesner M, Brors B. circlize implements and  
7 enhances circular visualization in R. Bioinformatics 2014;30:2811-2.  
8 1413  
9 1414  
10 1415  
11 1416  
12 1417  
13  
14  
15  
16  
17  
18  
19  
20  
21  
22  
23  
24  
25  
26  
27  
28  
29  
30  
31  
32  
33  
34  
35  
36  
37  
38  
39  
40  
41  
42  
43  
44  
45  
46  
47  
48  
49  
50  
51  
52  
53  
54  
55  
56  
57  
58  
59  
60  
61  
62  
63  
64  
65



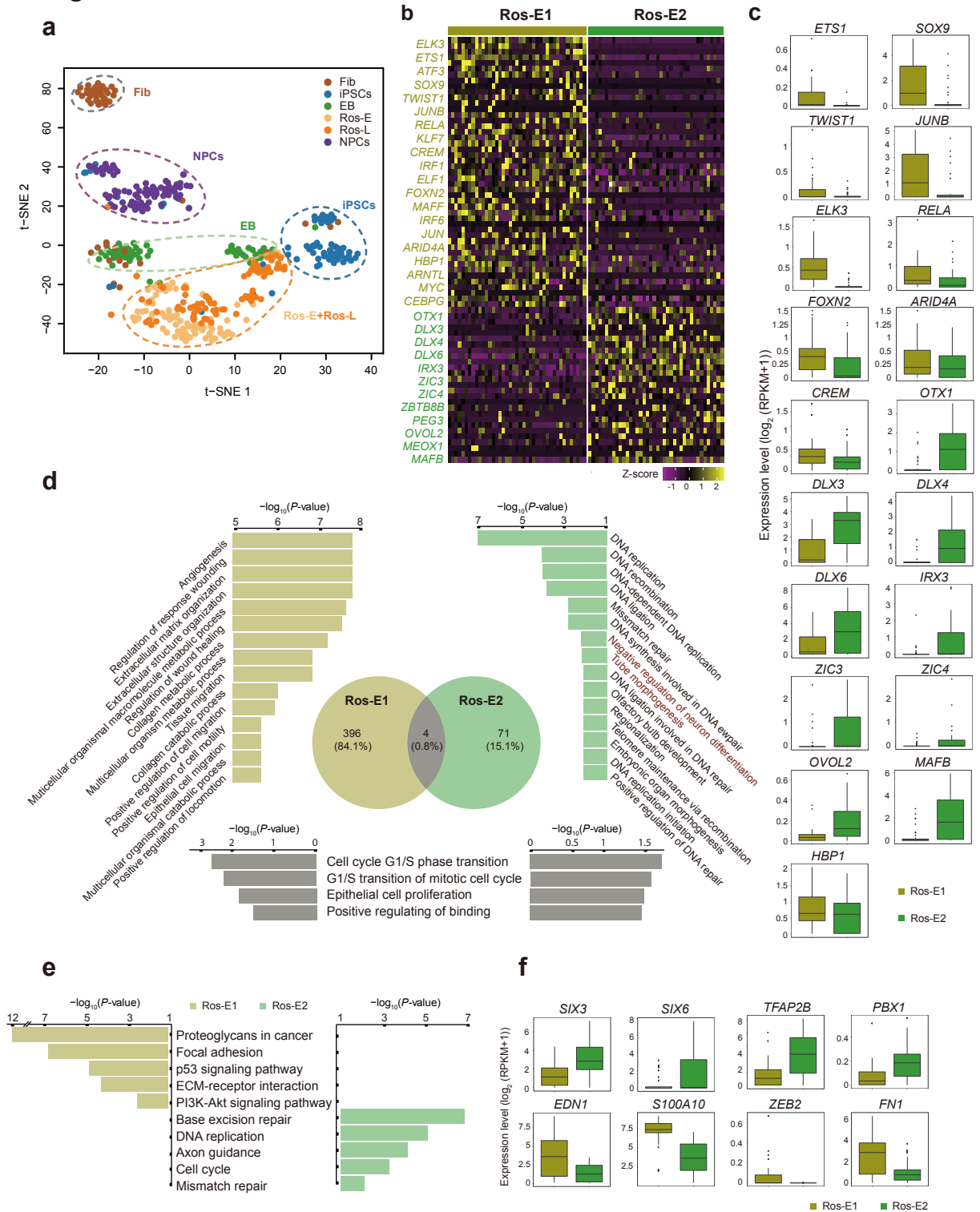


Fig. 1



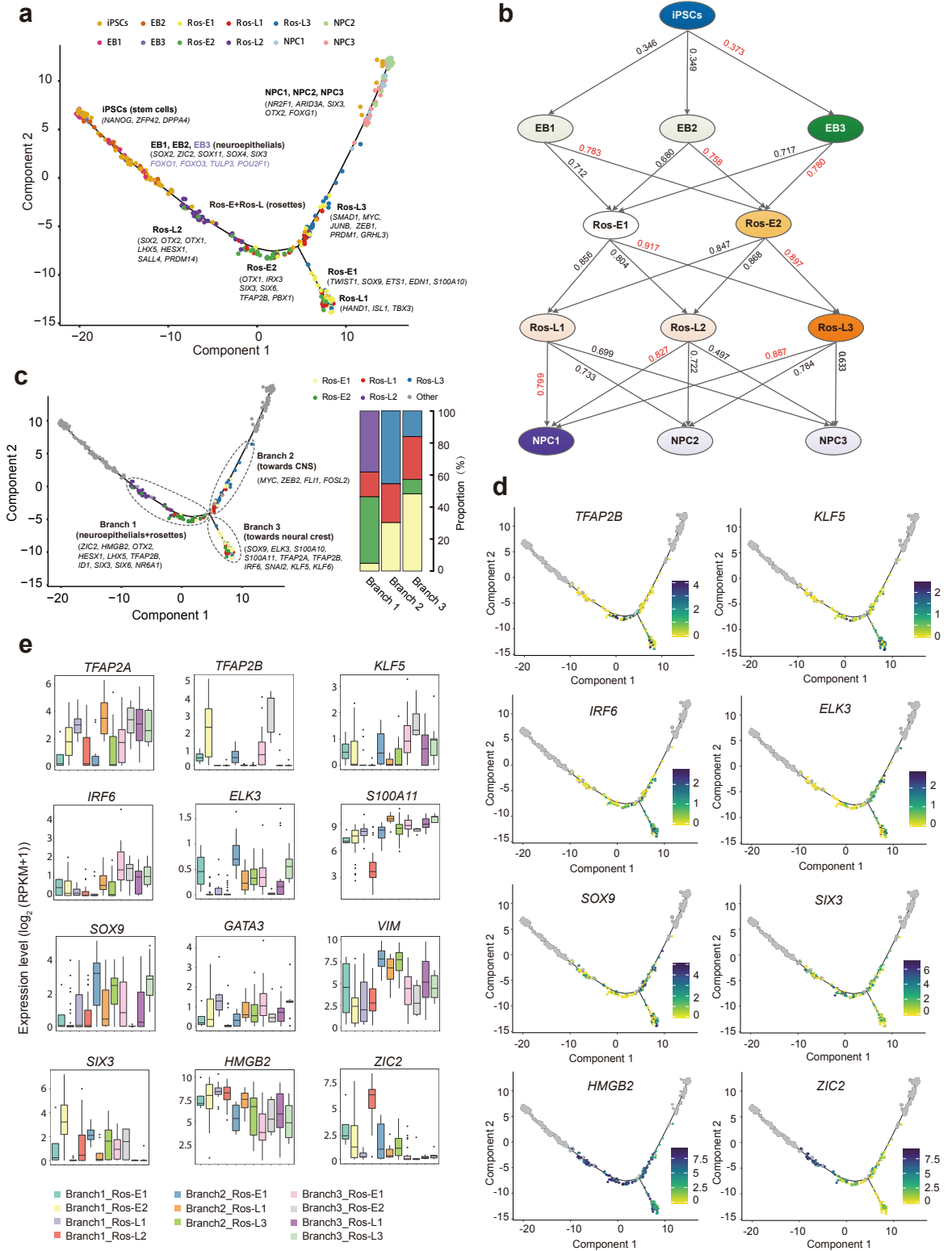
**Fig. 1 Transcriptome and regulome dynamics during human early neural differentiation.** **a** Schematic illustration of experimental strategy. **b** Bright field and immunostaining of well-defined markers for iPSCs including OCT4 and NANOG, and for neural rosettes including PAX6, NES (NESTIN), SOX2,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

**Fig. 2**



**Fig. 2 Cell heterogeneity and identification of subsets within Ros-E stage. a** T-SNE analysis of different cell stages as indicated with different color (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); NPCs (n = 90). Each dot represents an individual cell. **b** Heatmap shows scaled expression [ $\log_2(\text{RPKM}+1)$ ] of discriminative TF sets for each cluster at Ros-E stage,  $P$ -value  $\leq 0.01$ . 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 up-regulated 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 grey). **e** Top 5 differential pathway in Ros-E1 and Ros-E2 respectively by KEGG enrichment analysis. **f** Representative box plots of subpopulation specific genes identified by SCDE (single-cell differential expression), adjusted  $P$ -value  $\leq 0.01$ .

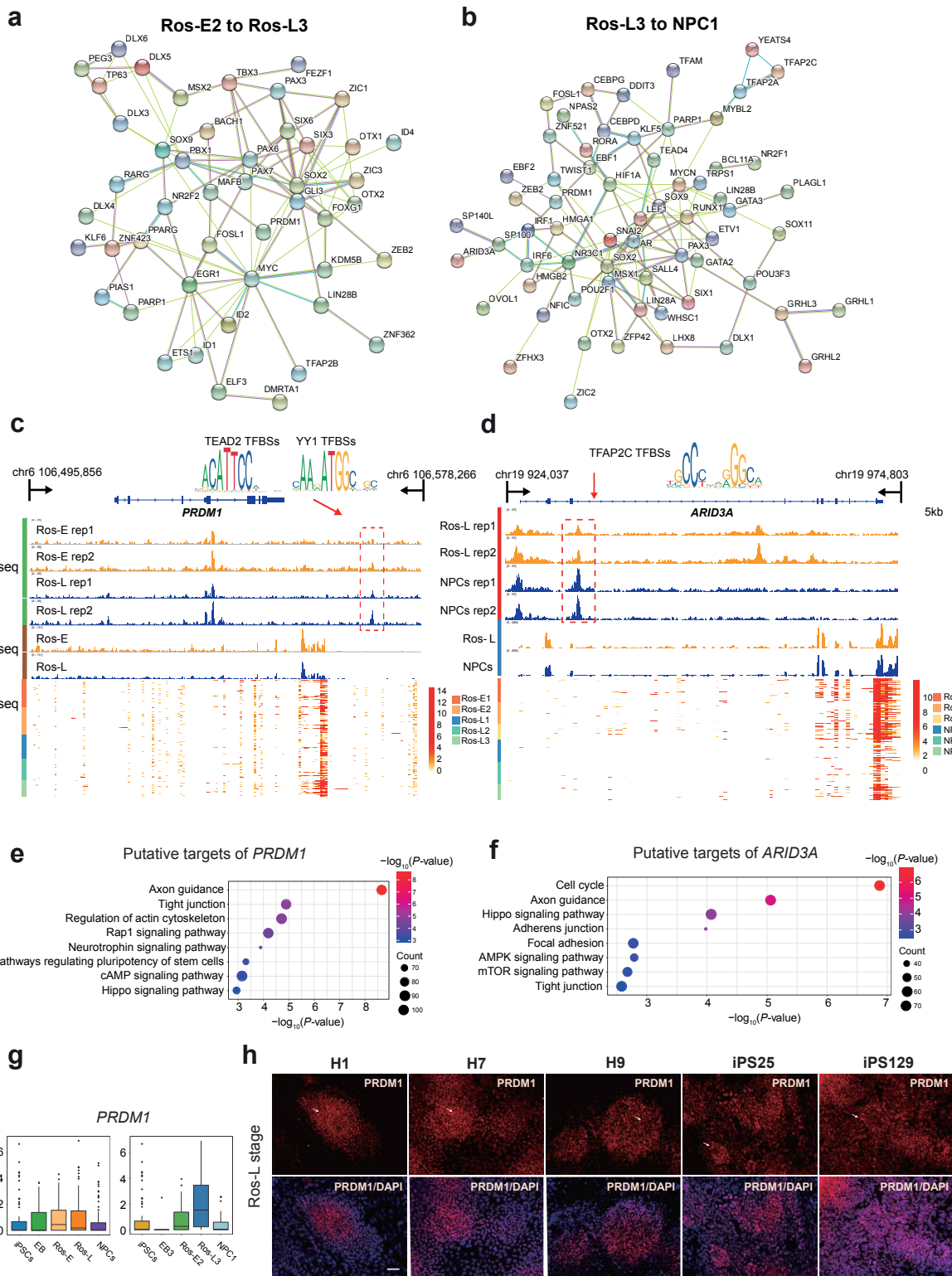
**Fig. 3**



**Fig. 3 Cell fate specification revealed by reconstructed trajectory.**

**a** Differentiation trajectory constructed by 8220 variable genes across different cell stages. Selected marker genes specific to the respective cell stage/ subpopulation are indicated with black/purple color. **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 is indicated on 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  $\leq 0.01$ ). 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.

**Fig. 4**

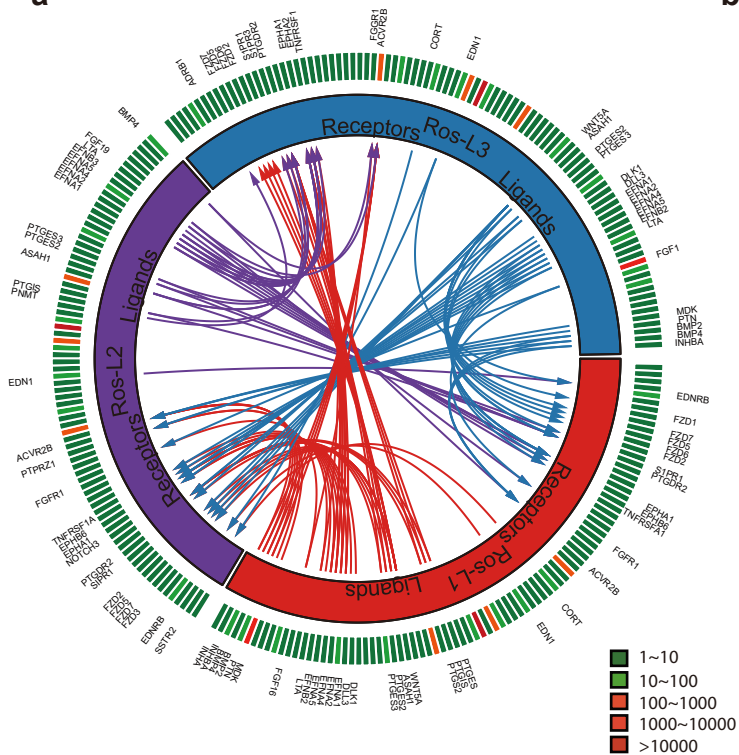


**Fig. 4 Key regulators and corresponding *cis*-regulatory elements during neural differentiation.** **a** Regulatory network of TFs differentially expressed between Ros-E2 and Ros-L3. **b** Regulatory network of differentially expressed TFs between Ros-L3 and NPC1. **c, d** IGV screenshots of ATAC-seq and bulk RNA-seq as well as the corresponding scRNA-seq heatmaps 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  $\mu\text{m}$ .

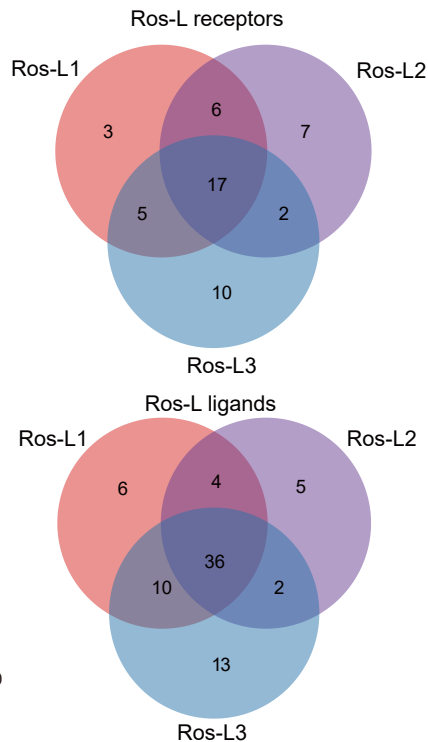


**Fig. 5**

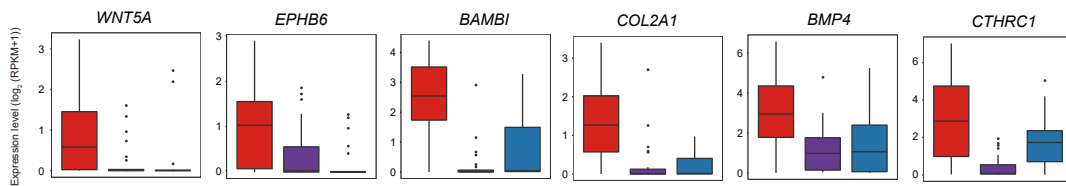
**a**



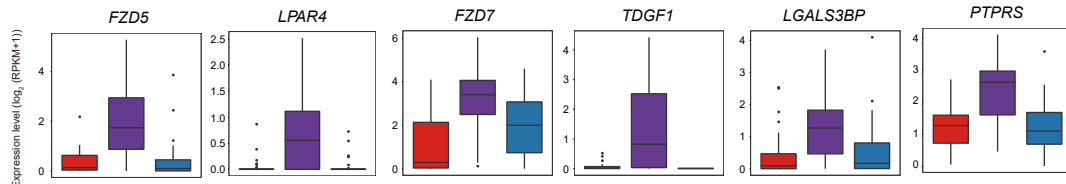
**b**



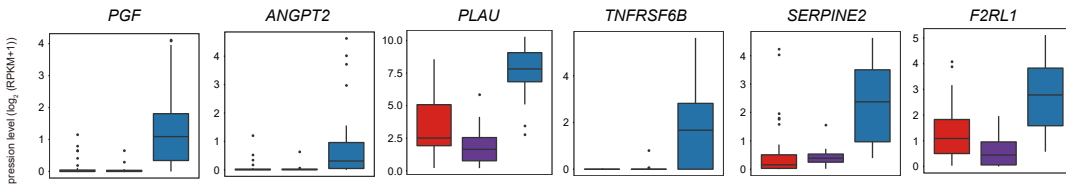
**c**



**d**



**e**

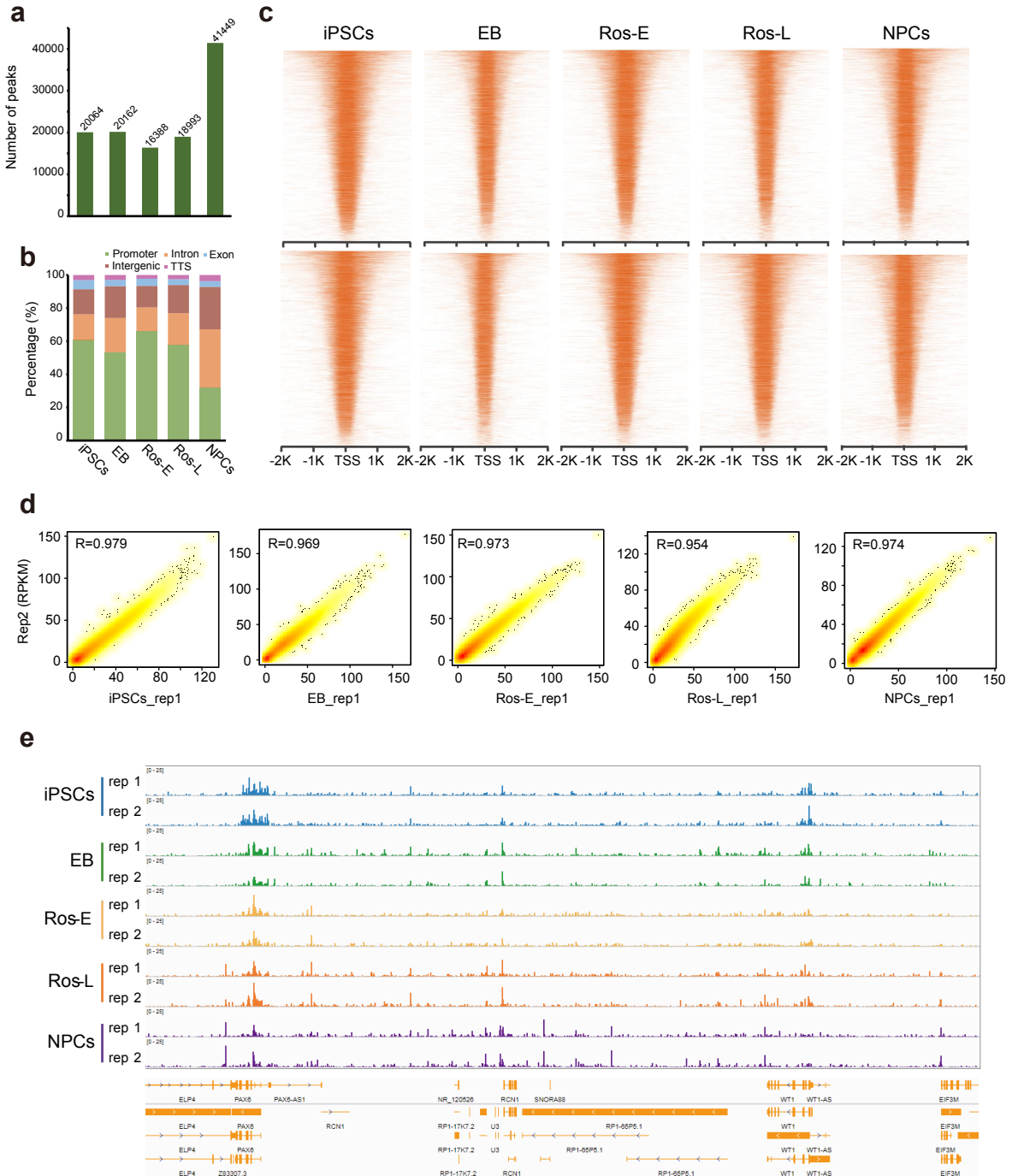


■ Ros-L1 ■ Ros-L2 ■ Ros-L3

**Fig. 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 were shown in red, purple and blue color respectively). The outer layer indicates the expression profiles of ligands and receptors expressed in each cell subset, with low expressed molecules in green color while high expressed ones in red color. 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, d, e** Expression level of receptors/ligands enriched in Ros-L1 (**c**), Ros-L2 (**d**) and Ros-L3 (**e**), respectively.

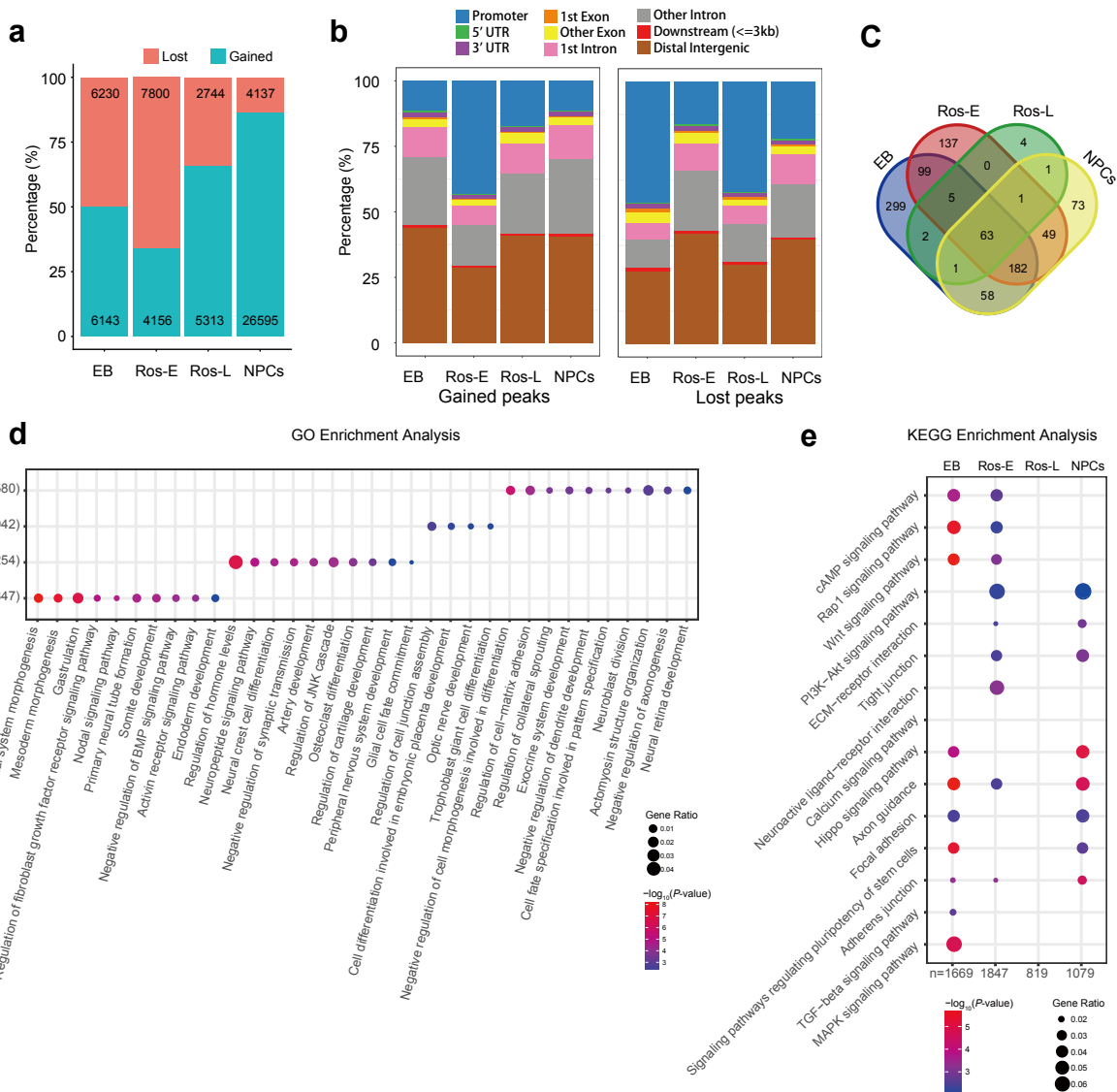


## Additional file 1: Figure S1



**Additional file 1: Figure S1. Quality control of ATAC-seq.** **a** Bar graphs indicate the number of chromatin open regions detected at each cell stage of neural differentiation. **b** Genomic components (distribution) of the peaks in each cell stage during neural differentiation. **c** Heatmaps reporting the chromatin accessibility density within  $\pm 2$  kb of TSSs. **d** Biological replicates of bulk ATAC-seq show high reproducibility. **e** IGV screenshot showing highly correlated ATAC signals in selected region between replicates.

# Additional file 2: Figure S2



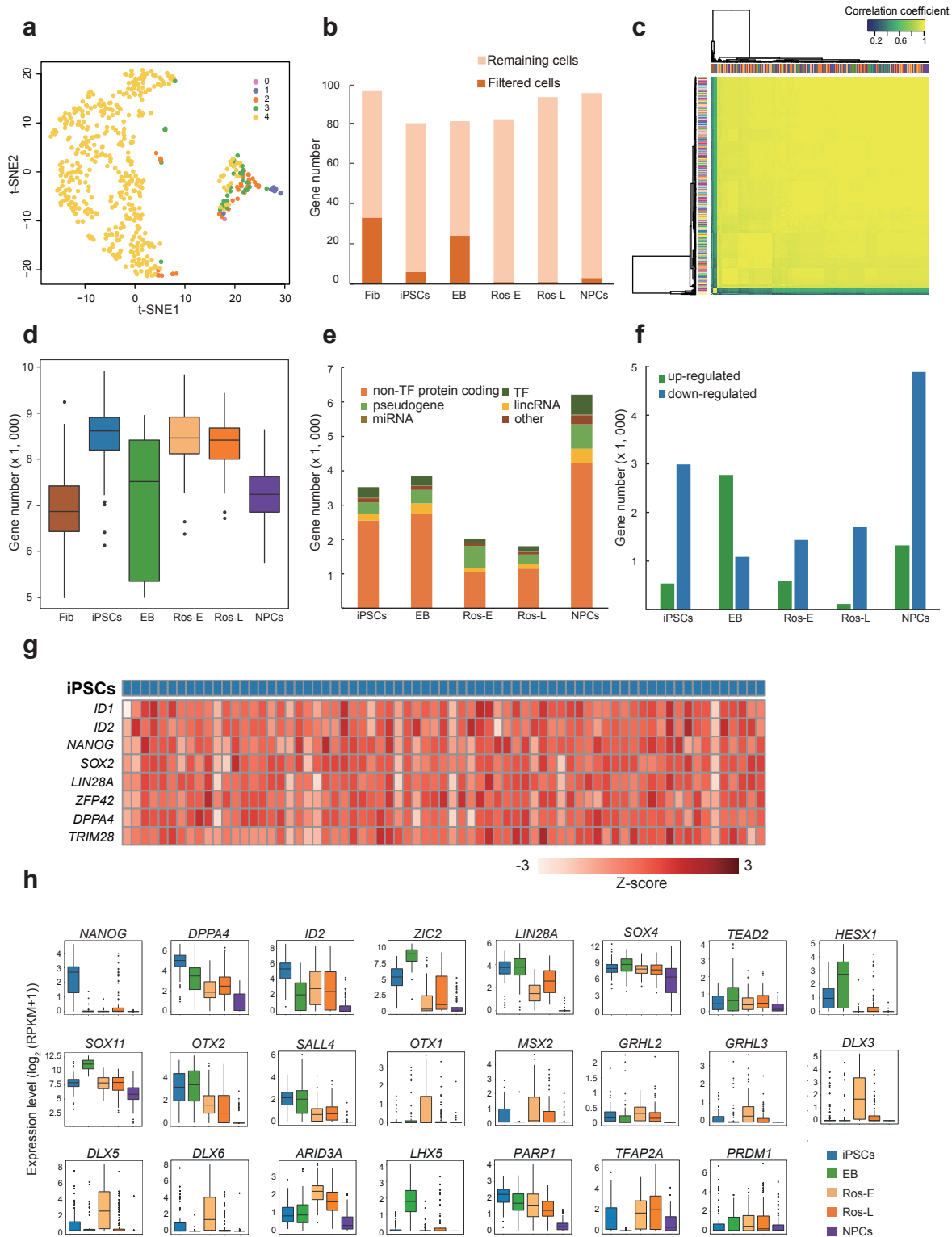
**Additional file 2: Figure S2. Dynamics of gained and lost peaks during neural differentiation.** **a** Bar graph shows the number of gained and lost peaks at each cell stage. **b** Bar graph shows genomic composition of gained and lost peaks at each cell stage respectively. **c** Venn plot of GO enrichment analysis on the genes associated with lost peaks at each stage (adjusted  $P$ -value  $\leq 0.01$ ). **d** Selected GO terms identified by genes associated with lost peaks specific to the respective indicated cell stage (adjusted  $P$ -value  $\leq 0.01$ ). **e** Selected differential pathways identified by genes associated with lost peaks at indicated cell stages (adjusted  $P$ -value  $\leq 0.01$ ).



**Additional file 3: Figure S3. Stage-specific features of *cis*-regulatory elements during neural differentiation.** **a** Bar plot showing the number of stage specific ATAC peaks at iPSCs, EB, Ros-E, Ros-L and NPCs stage (adjusted  $P$ -value  $\leq 0.01$ ). **b** Pie chart shows genomic composition of stage specific peaks respectively. **c, d** GO term and KEGG enrichment analysis of stage specific peaks, respectively (adjusted  $P$ -value  $\leq 0.05$ ).

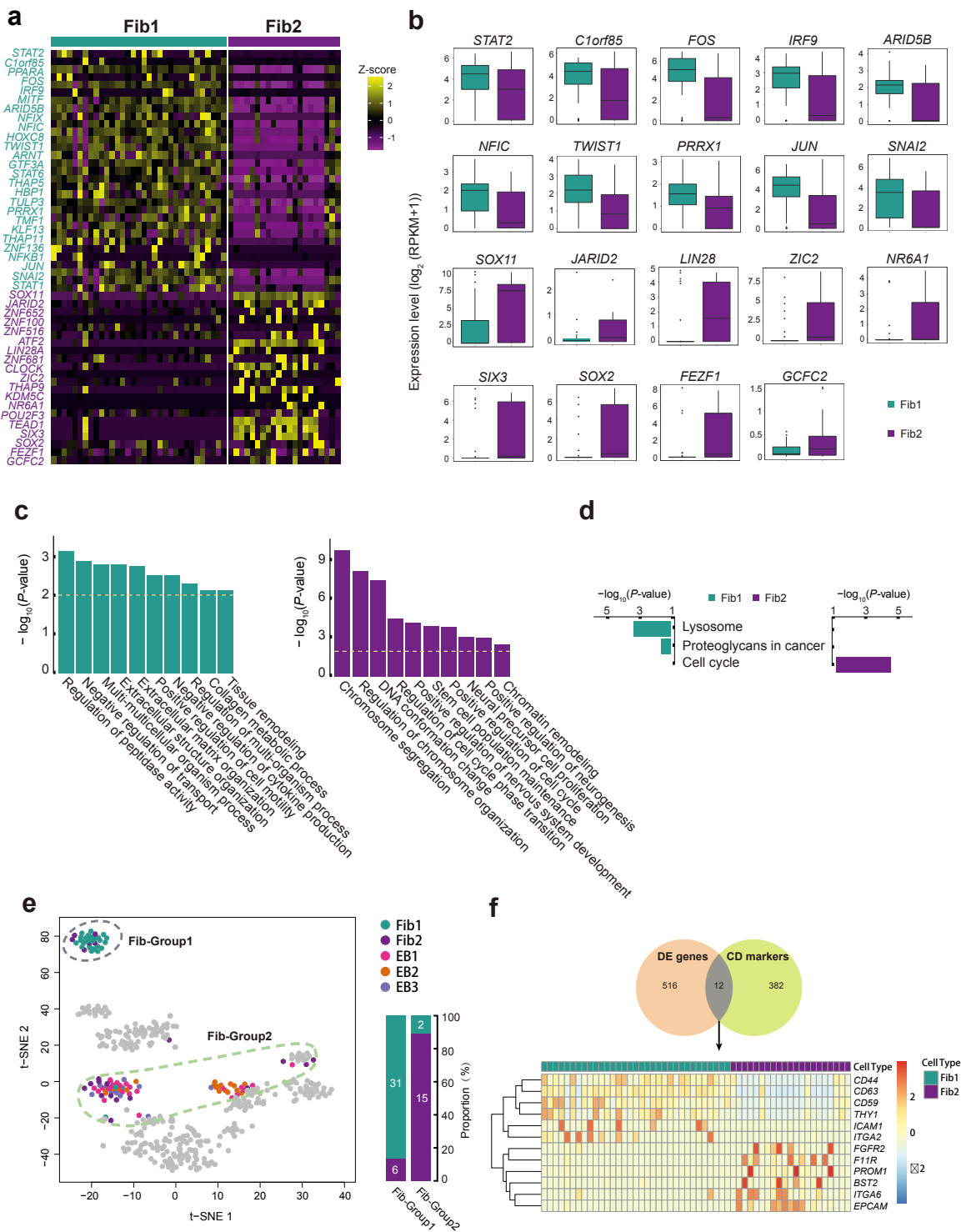


# Additional file 4: Figure S4



**Additional file 4: Figure S4. Quality control of scRNA-seq.** **a** Graph indicates data quality of totally 527 single cells. Color scheme indicates the filter conditions, each dot represents one cell, and yellow dots showing the cells that successfully passed all criteria were used for downstream analysis. **b** Bar plots show the percentage of filtered cells and remaining cells. **c** ERCC correlation analysis of all single cells showing very little batch effects. **d** Box plots report the number of expressed genes for each cell stage after quality control filtering. Each dot represents an outlier gene and each box represents the median and first and third quartiles. **e** Genomic distribution of genes at each cell stage. **f** Summary of up-regulated and down-regulated genes at each cell stage compared to other stages. **g** Expression pattern of pluripotency-associated genes in iPSCs. Color scheme is based on z-score distribution from -3 (light red) to 3 (red). **h** Expression pattern of representative differentially expressed TFs during neural differentiation (adjusted  $P$ -value  $\leq 0.01$ ).

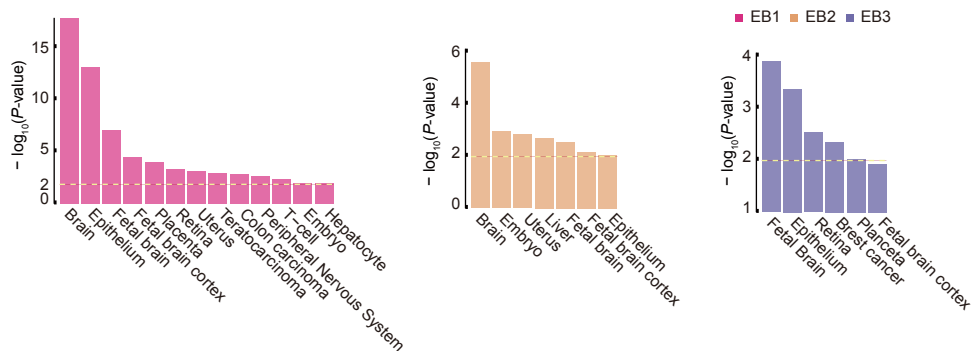
# Additional file 5: Figure S5



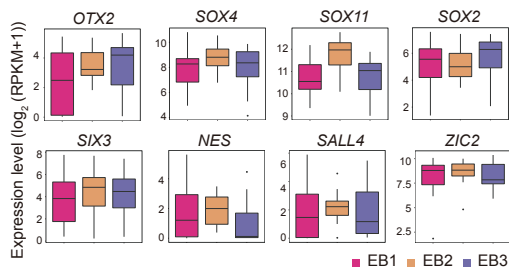
**Additional file 5: Figure S5. Subgroups identification and key transcriptomic features within Fib stage.** **a** Heatmap reports scaled expression [ $\log_2(\text{RPKM}+1)$ ] of discriminative TF sets for each cluster in Fib stage with  $P$ -value cutoff  $\leq 0.01$ . Color scheme is based on z-score distribution from -1 (purple) to 2 (yellow). Gene symbols highlight with color specific to the respective Fib subset. **b** Box plots of selected TFs defined in Figure S5a. **c** Selected GO terms identified by up-regulated genes specific to the respective Fib subpopulation with the color as indicated (Green: GO terms specific to Fib1; purple: GO terms specific to Fib2). **d** KEGG enrichment analysis of all terms in Fib subpopulation, respectively. **e** Fib-Group1 and Fib-Group2 based on their location on the t-SNE are marked by dashed ellipse. The columns represent the components of Fib-Group1 and Fib-Group2, respectively. **f** Comparison of differentially expressed (DE) genes between Fib subpopulation with CD markers dataset (HUGO Gene Nomenclature Committee, HGNC) and the heatmap of differentially expressed CD markers between the two Fib subpopulation.

# Additional file 6: Figure S6

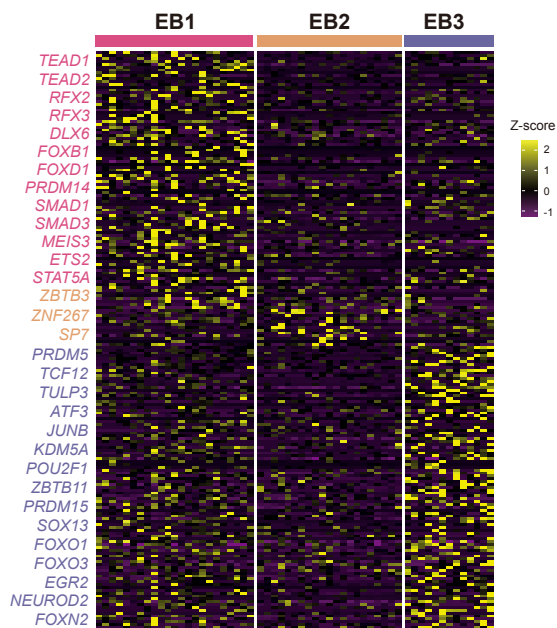
**a**



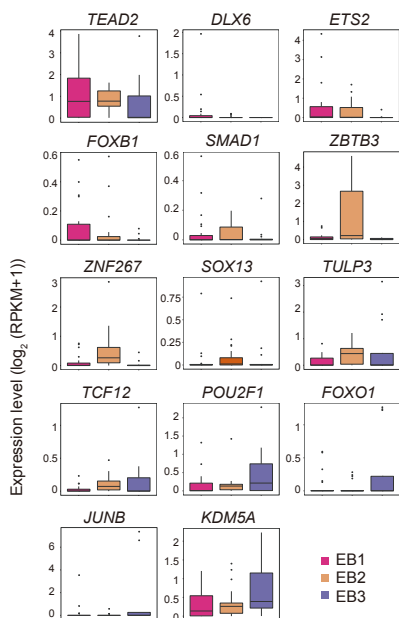
**b**



**c**

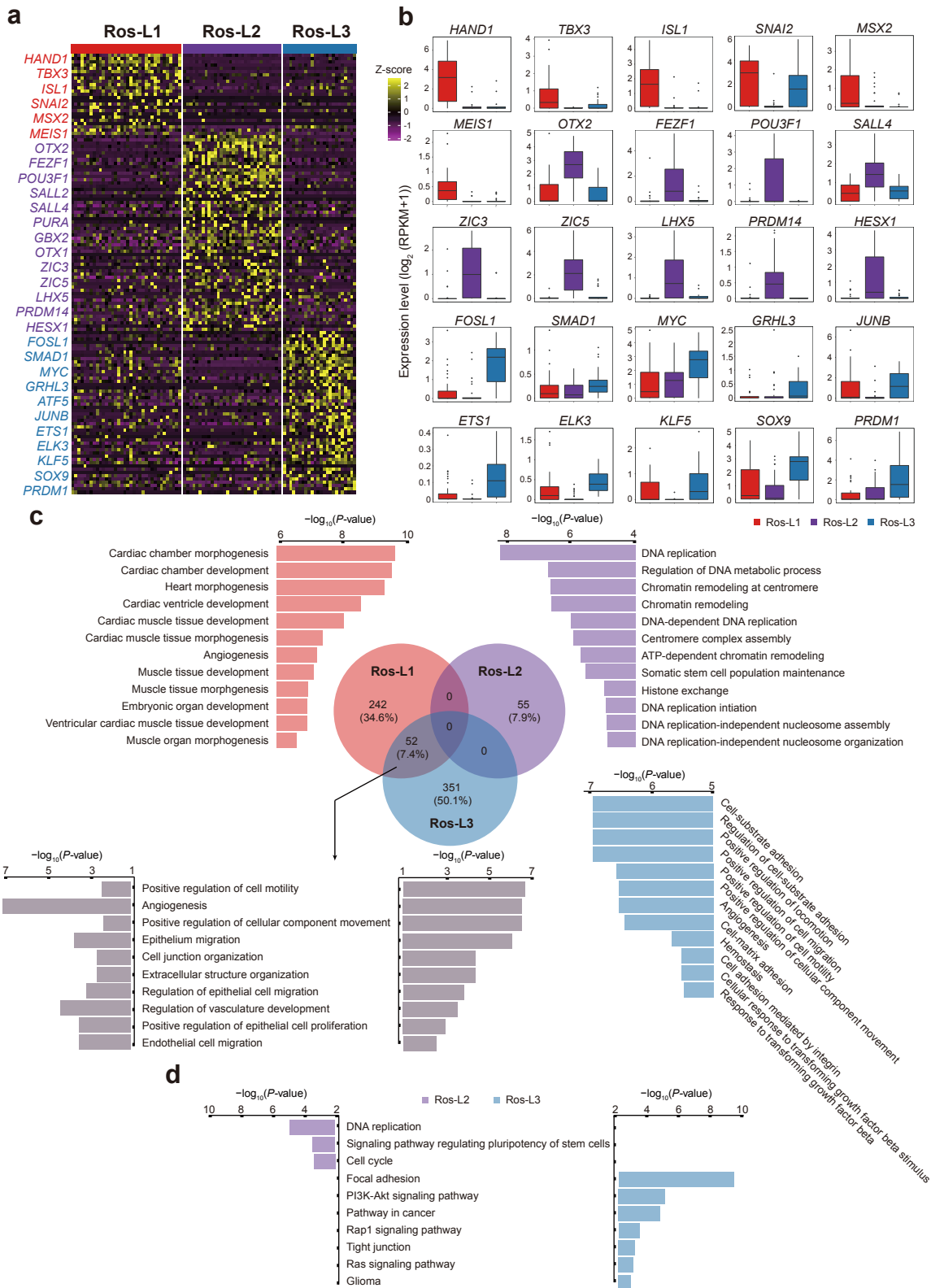


**d**



**Additional file 6: Figure S6. Subgroups identification and key transcriptomic features within EB stage.** **a** David for tissue enrichment analysis of up-regulated genes defined by three EB subgroups compared to iPSCs stage respectively. **b** Box plots of commonly expressed genes across EB subsets. **c** Heatmap reports scaled expression [ $\log_2(\text{RPKM}+1)$ ] of discriminative TF sets for each cluster in EB stage with  $P$ -value cutoff  $\leq 0.01$ . Color scheme is based on z-score distribution from -1 (purple) to 2 (yellow). Gene symbols highlight with color specific to the respective EB subset. **d** Box plot of selected TFs defined in Figure S6a.

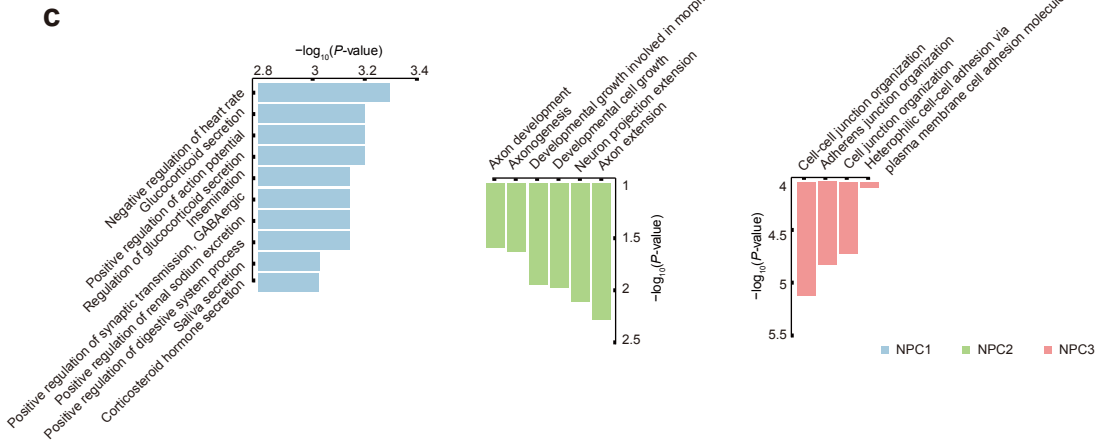
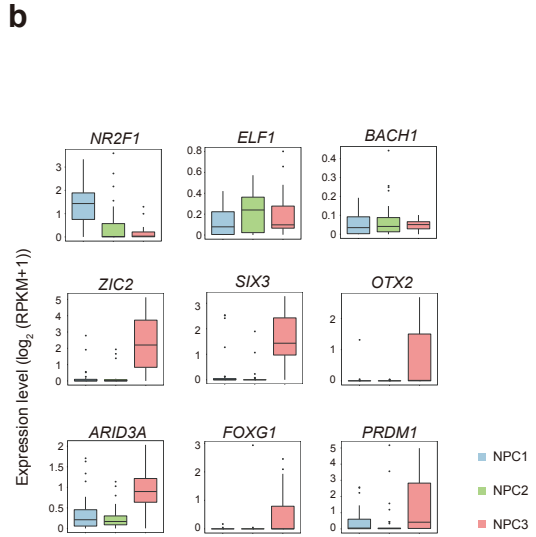
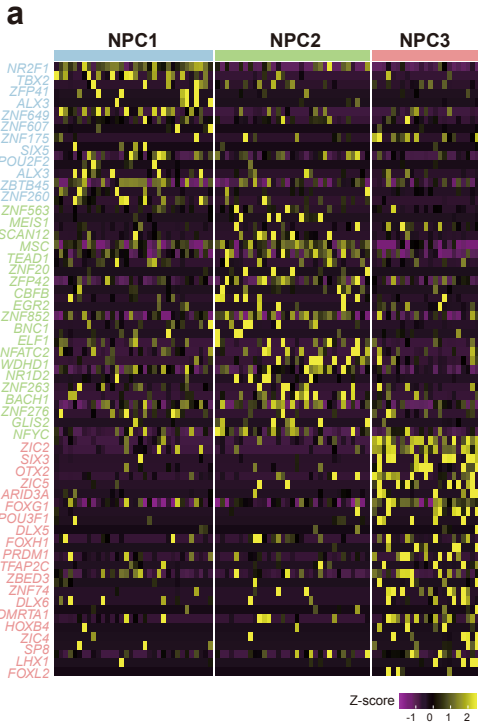
# Additional file 7: Figure S7



**Additional file 7: Figure S7. Subgroups identification and key transcriptomic features within Ros-L stage.** **a** Heatmap reports scaled expression [ $\log_2(\text{RPKM}+1)$ ] of discriminative TF sets for each cluster in Ros-L stage with  $P$ -value cutoff  $\leq 0.01$ . Color scheme is based on z-score distribution from -2 (purple) to 2 (yellow). Gene symbols highlight with color specific to the respective Ros-L subset. **b** Box plots of selected TFs defined in Figure S7a. **c** Top 12 of GO terms identified by up-regulated genes specific to the respective Ros-L subpopulation with the color as indicated (red: GO terms specific to Ros-L1; purple: GO terms specific to Ros-L2; blue: GO terms specific to Ros-L3; gray: selected GO terms shared by Ros-L1 and Ros-L3). **d** KEGG enrichment analysis of Ros-L2 (all terms) and Ros-L3 (selected terms), respectively.

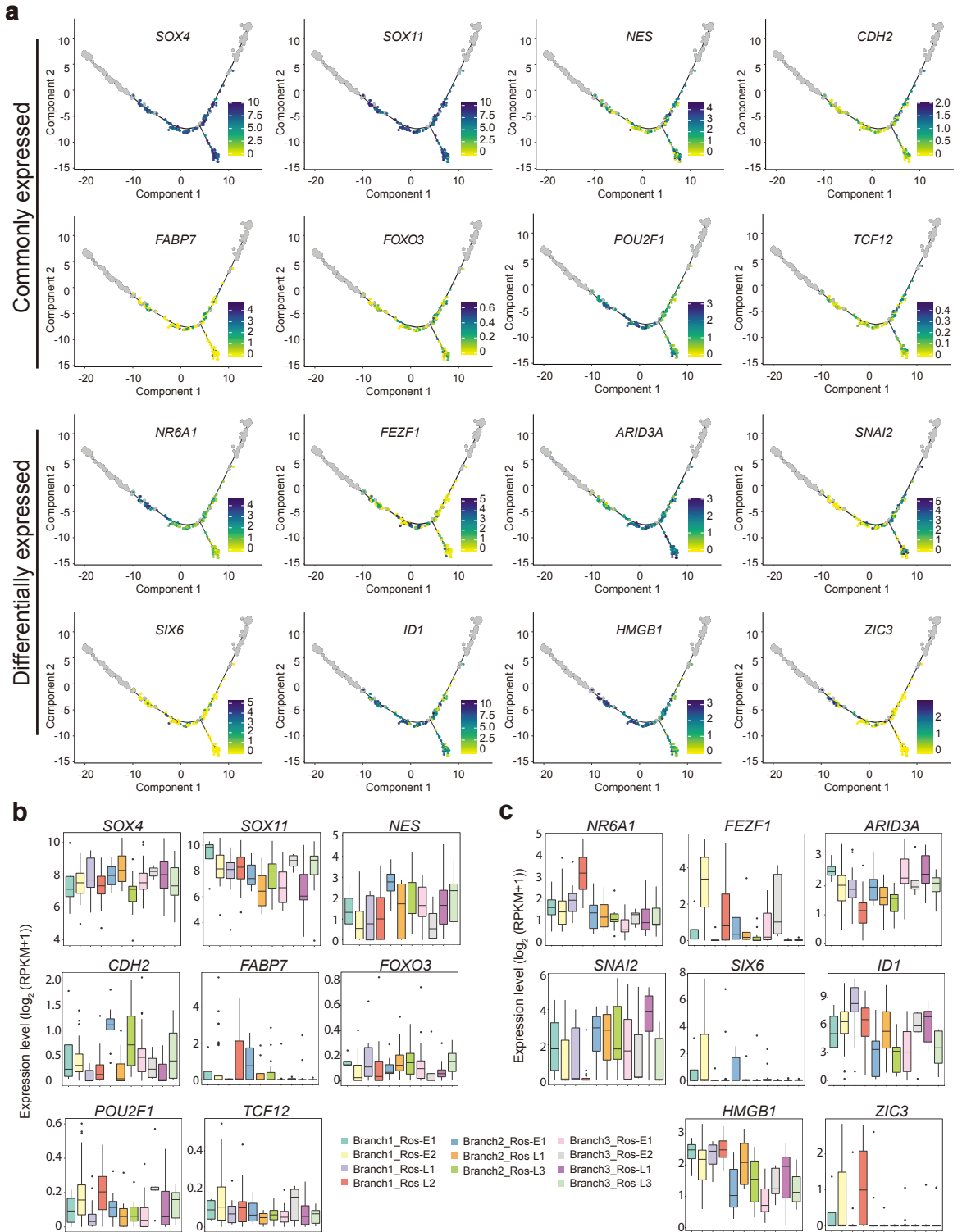


# Additional file 8: Figure S8



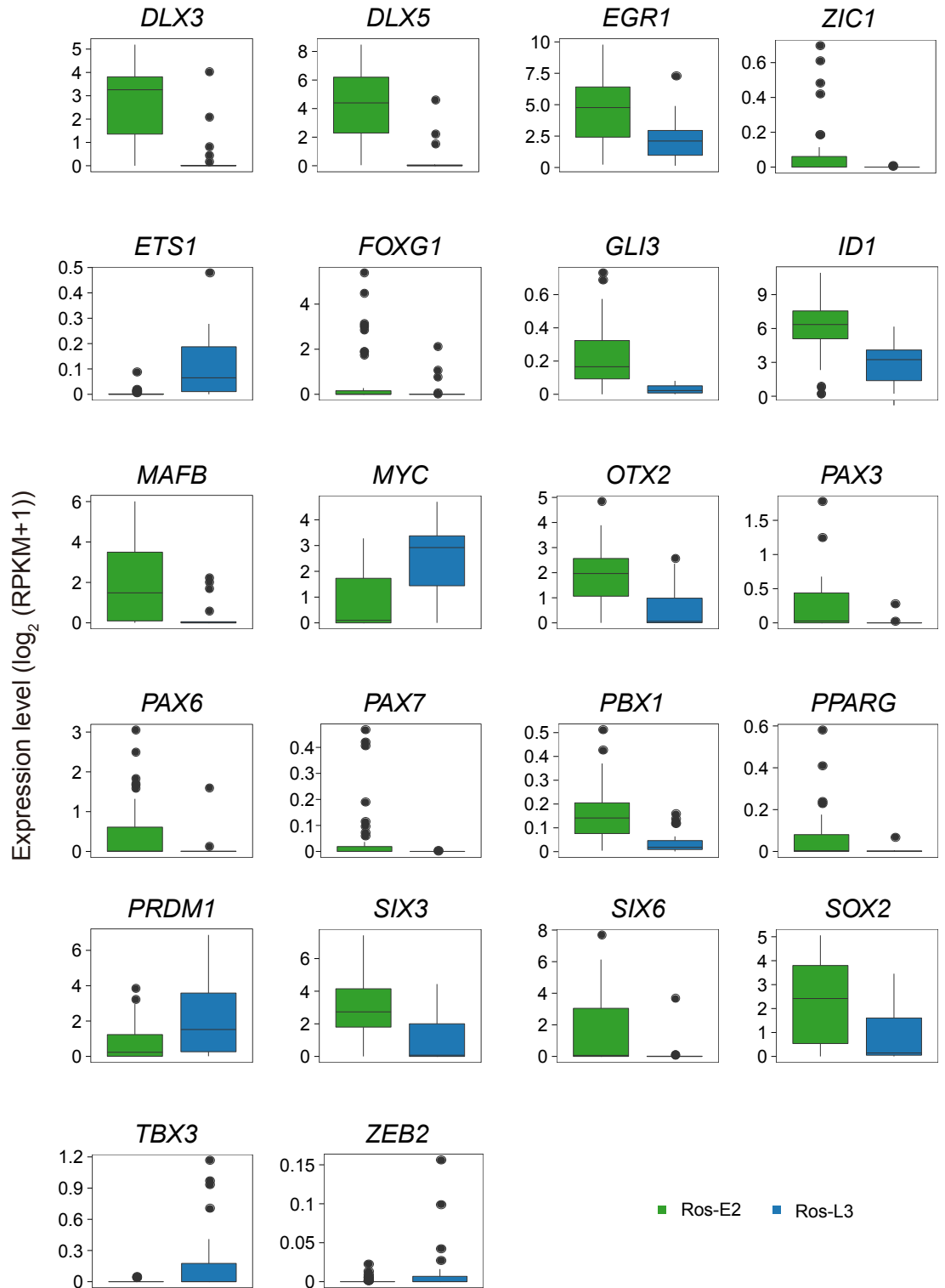
**Additional file 8: Figure S8. Subgroups identification and key transcriptomic features within NPCs stage.** **a** Heatmap reports scaled expression [ $\log_2(\text{RPKM}+1)$ ] of discriminative TF sets for each cluster in NPCs stage with  $P$ -value cutoff  $\leq 0.01$ . Color scheme is based on z-score distribution from -1 (purple) to 2 (yellow). Gene symbols highlight with color specific to the respective NPC subset. **b** Box plot of selected TFs defined in Figure S8a. **c** Top 10 (NPC1) and all (NPC2 and NPC3) of GO terms identified by up-regulated genes specific to the respective Ros-L subpopulation with the color as indicated (blue: GO terms specific to NPC1; green: GO terms specific to NPC2; pink: GO terms specific to NPC3).

# Additional file 9: Figure S9



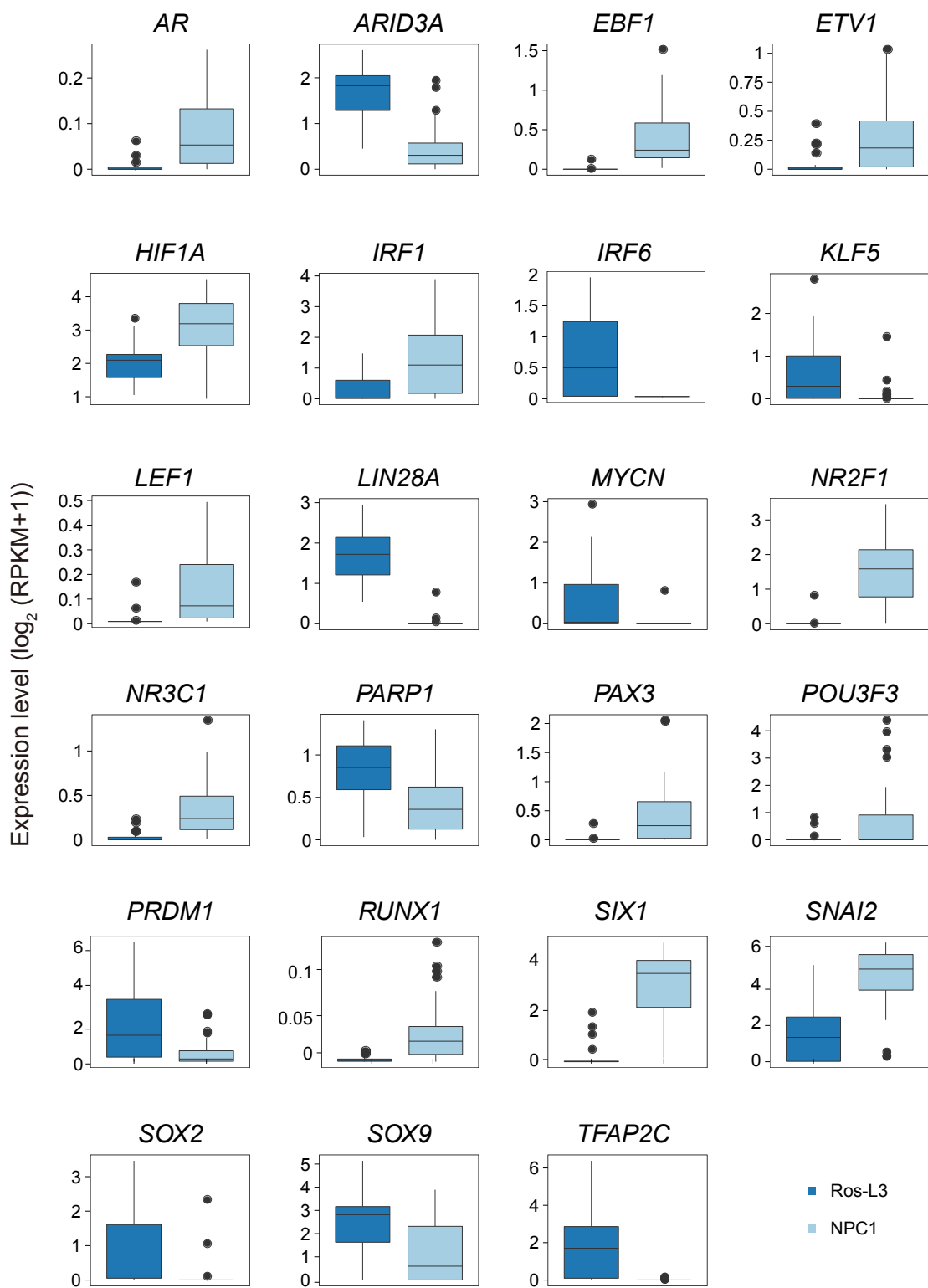
**Additional file 9: Figure S9. Expression pattern of selected transcription factors (TFs) within rosettes (Ros-E and Ros-L) stage. a** Expression enrichment of commonly and differentially expressed TFs along the differentiation trajectory. Color scheme is based on expression [ $\log_2$  (RPKM+1)]. **b, c** Expression pattern of selected TFs with respect to Figure S9a (adjusted  $P$ -value  $\leq 0.01$ ).

# Additional file 10: Figure S10



**Additional file 10: Figure S10. Differentially expressed transcription factors (TFs) between Ros-E2 and Ros-L3.** Ros-E2 and Ros-L3 were shown in green and blue column, respectively (adjusted  $P$ -value  $\leq 0.01$ ).

# Additional file 11: Figure S11



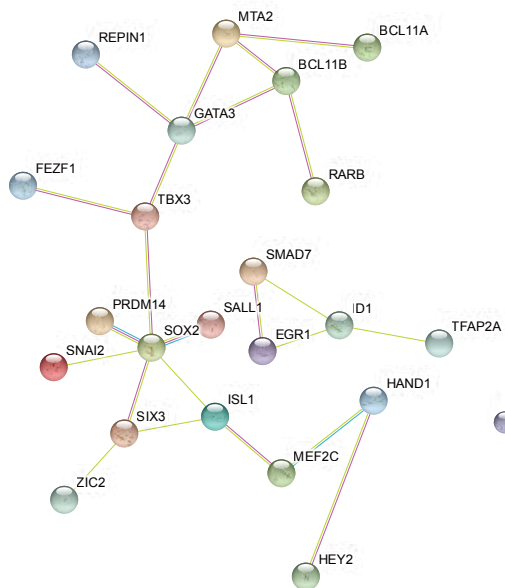
**Additional file 11: Figure S11. Differentially expressed transcription factors (TFs) between Ros-L3 and NPC1.** Ros-L3 and NPC1 were shown in dark blue and light blue column, respectively (adjusted  $P$ -value  $\leq 0.01$ ).



# Additional file 12: Figure S12

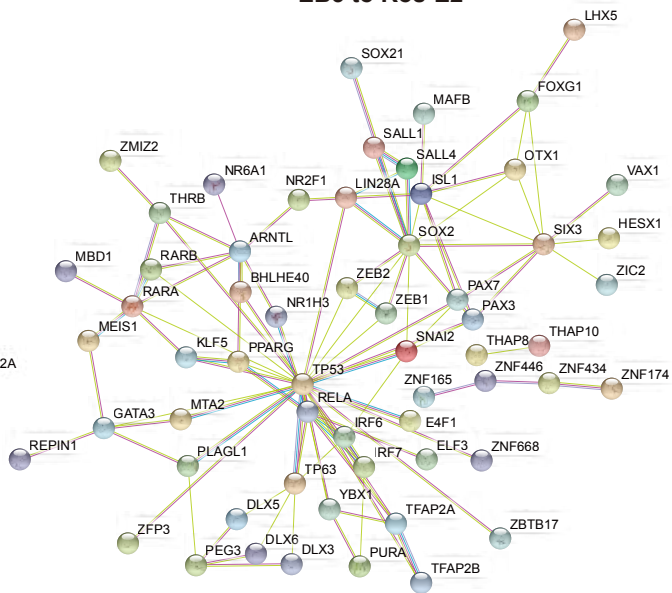
**a**

**iPSCs to EB3**



**b**

**EB3 to Ros-E2**

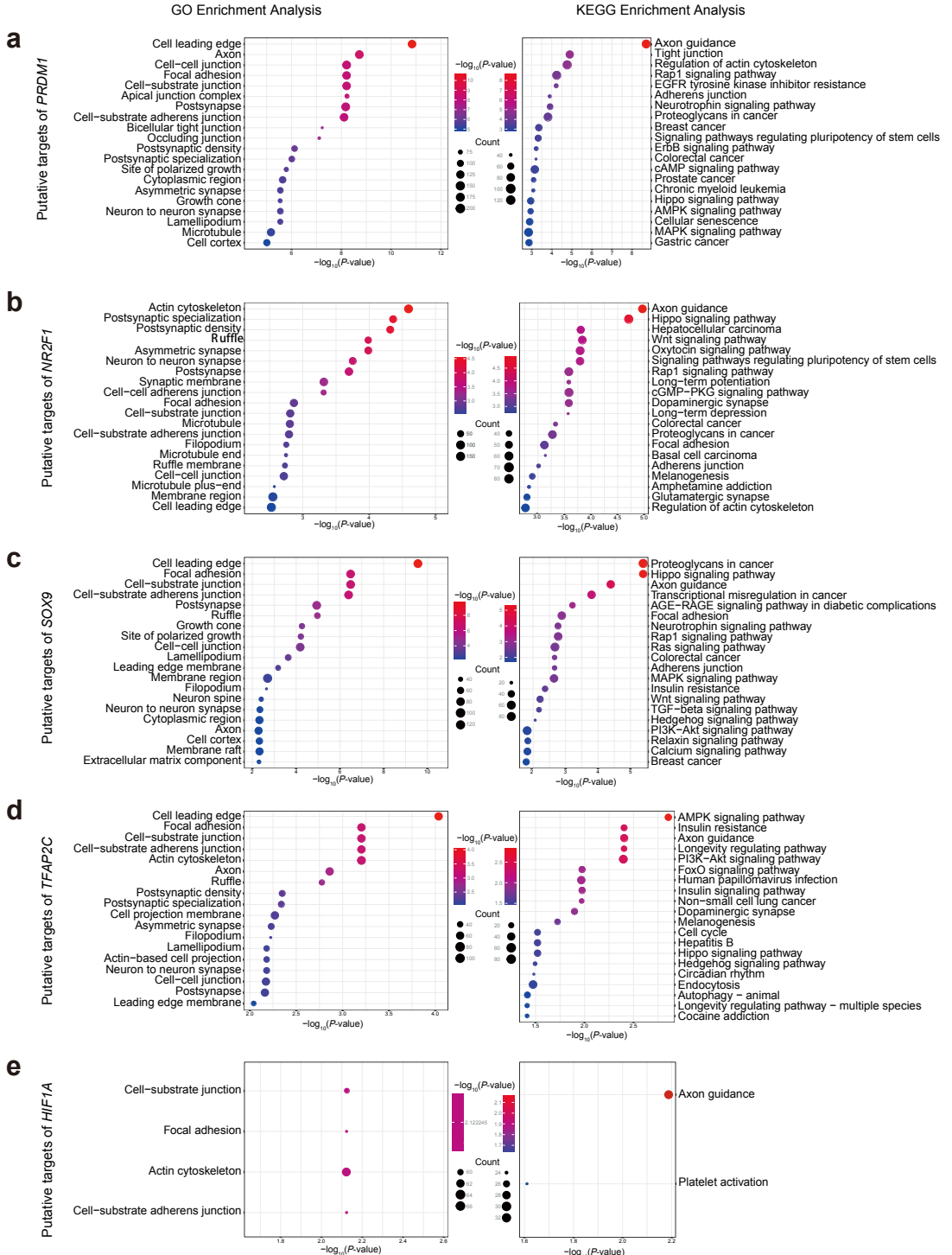


**Additional file 12: Figure S12. Key regulators during neural differentiation. a**

Regulatory network of differentially expressed TFs between iPSCs and EB3. **b**

Regulatory network of differentially expressed TFs between EB3 and Ros-E2.

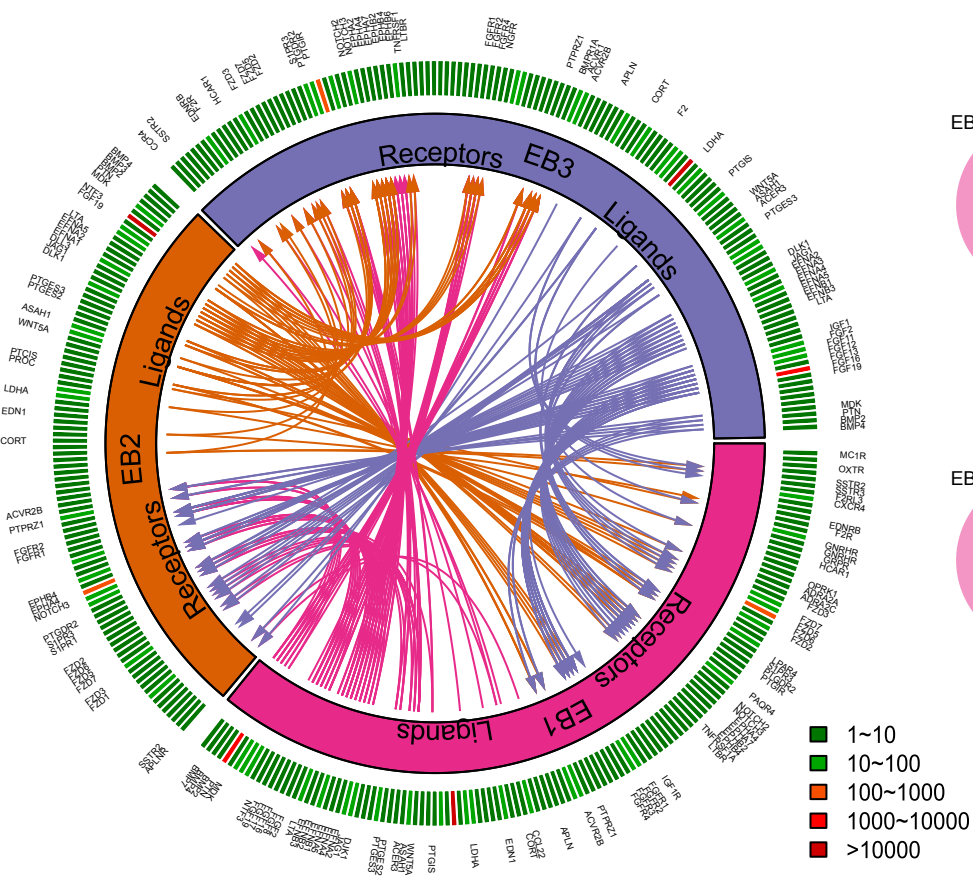
# Additional file 13: Figure S13



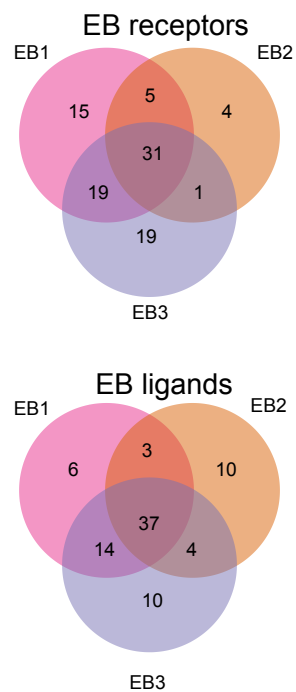
**Additional file 13: Figure S13. GO term and KEGG enrichment analysis of selected transcription factors (TFs) targets.** GO term and KEGG enrichment analysis for putative targets of *PRDM1* (a), *NR2F1* (b), *SOX9* (c), *TFAP2C* (d) and *HIF1A* (e), adjusted *P*-value  $\leq 0.05$ .

# Additional file 14: Figure S14

**a**



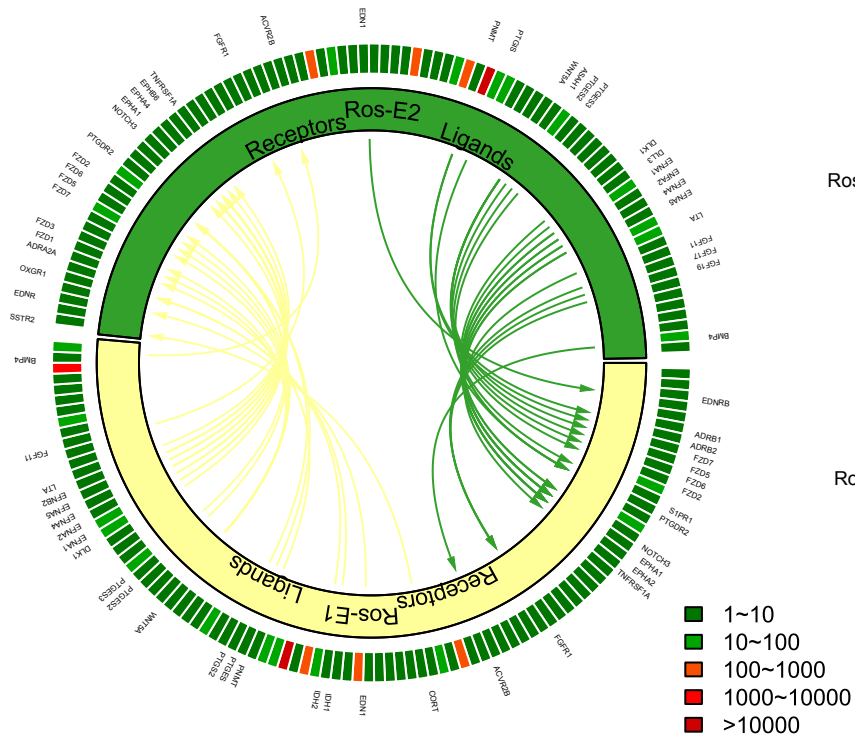
**b**



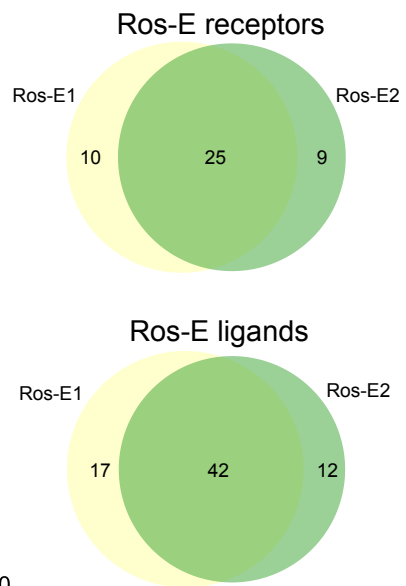
**Additional file 14: Figure S14. Putative signaling between expressed receptors and their ligands in EB subsets.** **a** The inner layer compartments represent different cell subpopulations (EB1, EB2 and EB3). The outer layer indicates the expression profiles of ligands and receptors expressed in each cell subset, with low expressed molecular in green color while high expressed ones in red color. Arrows indicate putative interactions between ligands and receptors among cell subsets. **b** Venn plot showing the overlapping of ligands and receptors among cellular subpopulations.

# Additional file 15: Figure S15

**a**



**b**

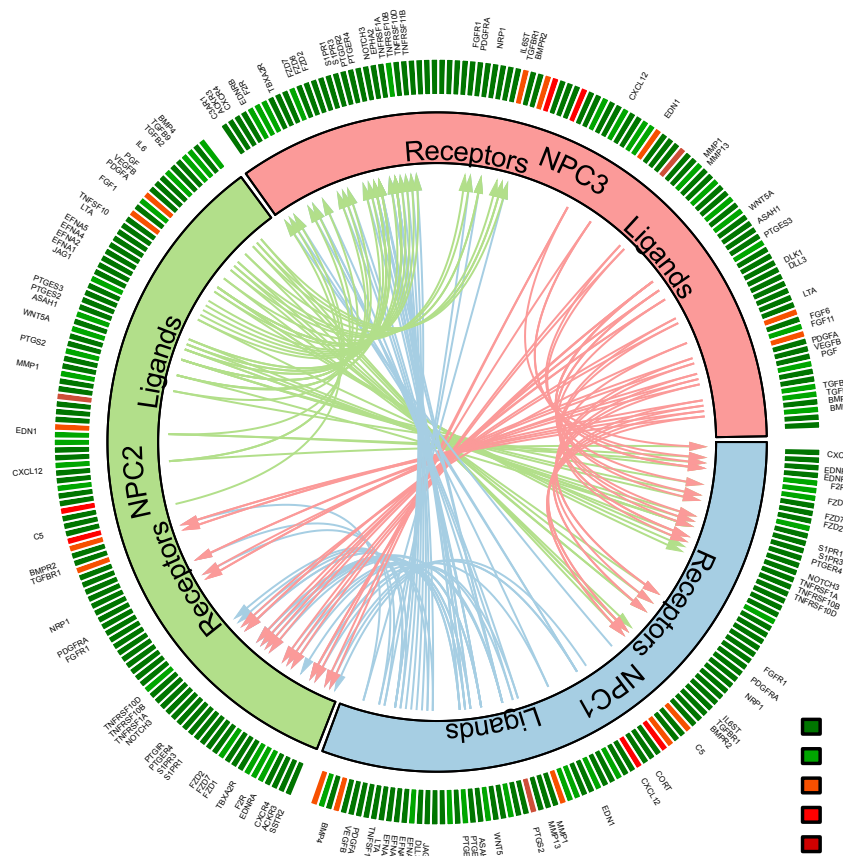


**Additional file 15: Figure S15. Putative signaling between expressed receptors and their ligands in Ros-E subsets.** **a** The inner layer compartments represent different cell subpopulations (Ros-E1 and Ros-E2). The outer layer indicates the expression profiles of ligands and receptors expressed in each cell subset, with low expressed molecular in green color while high expressed ones in red color. Arrows indicate putative interactions between ligands and receptors among cell subsets. **b** Venn plot showing the overlapping of ligands and receptors among cellular subpopulations.

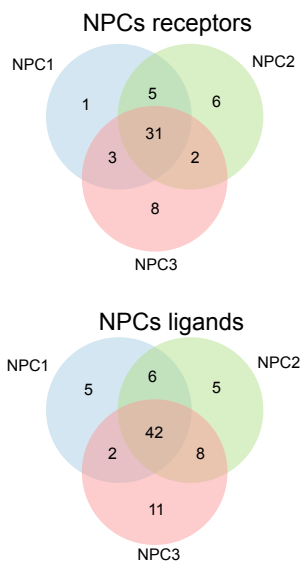


# Additional file 16: Figure S16

**a**


















**b**


















**Additional file 16: Figure S16. Putative signaling between expressed receptors and their ligands in NPC subsets.** **a** The inner layer compartments represent different cell subpopulations (NPC1, NPC2 and NPC3). The outer layer indicates the expression profiles of ligands and receptors expressed in each cell subset, with low expressed molecular in green color while high expressed ones in red color. Arrows indicate putative interactions between ligands and receptors among cell subsets. **b** Venn plot showing the overlapping of ligands and receptors among cellular subpopulations.

## Additional file 17: Figure S17
















### a. iPSCs stage

Motif	P-value	Best Match/Details
	1e-279	Pou5f1::Sox2/MA0142.1/Jaspar(0.924)
	1e-73	BORIS(Zf)/K562-CT CFL-ChIP-Seq(GSE32465)/Homer(0.899)
	1e-57	Sox3(HMG)/NPC-Sox3-ChIP-Seq(GSE33059)/Homer(0.956)
	1e-47	Otd6(POU,Homeobox)/NPC-Otd6-ChIP-Seq(GSE35496)/Homer(0.693)
	1e-43	FOXB1/MA0845.1/Jaspar(0.759)
	1e-41	PH0098.1_Lhx8/Jaspar(0.881)
	1e-37	BORIS(Zf)/K562-CT CFL-ChIP-Seq(GSE32465)/Homer(0.753)
	1e-33	POL003.1_GC-box/Jaspar(0.820)
	1e-28	LIN54/MA0619.1/Jaspar(0.778)
	1e-27	Unknown-ESC-element/mES-Nanog-ChIP-Seq(GSE11724)/Homer(0.797)
	1e-25	PB0097.1_Zfp281_1/Jaspar(0.919)
	1e-25	NFYB/MA0502.1/Jaspar(0.736)
	1e-24	MF0006.1_bZIP_cEBP-like_subclass/Jaspar(0.595)
	1e-23	MZF1/MA0056.1/Jaspar(0.610)
	1e-22	YY2/MA0748.1/Jaspar(0.660)
















## b. EB stage

Motif	P-value	Best Match/Details
	1e-298	BORIS(Zf)/K562-CTCF-L-ChIP-Seq(GSE32465)/Homer(0.932)
	1e-74	AP-2alpha(AP2)/Hela-AP2alpha-ChIP-Seq(GSE31477)/Homer(0.877)
	1e-58	Sox3(HMG)/NPC-Sox3-ChIP-Seq(GSE33059)/Homer(0.962)
	1e-57	MEOX1/MA0661.1/Jaspar(0.902)
	1e-53	PB0099.1_Zfp691_1/Jaspar(0.622)
	1e-51	GRHL1/MA0647.1/Jaspar(0.859)
	1e-50	POL010.1_DCE_S_III/Jaspar(0.698)
	1e-44	PRDM9(Zf)/Testis-DMC1-ChIP-Seq(GSE35498)/Homer(0.685)
	1e-44	Rhox11/MA0629.1/Jaspar(0.782)
	1e-42	TEAD(TEA)/Fibroblast-PU.1-ChIP-Seq(Unpublished)/Homer(0.810)
	1e-41	POU2F2/MA0507.1/Jaspar(0.671)
	1e-39	Nr5a2(NR)/mES-Nr5a2-ChIP-Seq(GSE19019)/Homer(0.624)
	1e-36	Six1(Homeobox)/Myoblast-Six1-ChIP-Chip(GSE20150)/Homer(0.836)
	1e-34	Myb/MA0100.2/Jaspar(0.769)
	1e-33	PH0137.1_Pitx1/Jaspar(0.653)













### c. Ros-E stage

Motif	P-value	Best Match/Details
	1e-30	TEAD4/MA0809.1/Jaspar(0.941)
	1e-22	CTCF(Zf)/CD4+-CTCF-ChIP-Seq(Barski_et_al.)/Homer(0.833)
	1e-19	Lhx2(Homeobox)/HFSC-Lhx2-ChIP-Seq(GSE48068)/Homer(0.951)
	1e-18	TFCP2/MA0145.3/Jaspar(0.933)
	1e-17	Rfx1(HTH)/NPC-H3K4me1-ChIP-Seq(GSE16256)/Homer(0.937)
	1e-16	Sox15(HMG)/CPA-Sox15-ChIP-Seq(GSE62909)/Homer(0.882)
	1e-16	Nrf2(bZIP)/Lymphoblast-Nrf2-ChIP-Seq(GSE37589)/Homer(0.772)
	1e-15	Pknox1(Homeobox)/ES-Prep1-ChIP-Seq(GSE63282)/Homer(0.853)
	1e-15	Ets1-distal(ETS)/CD4+-PolII-ChIP-Seq(Barski_et_al.)/Homer(0.676)
	1e-15	RUNX1(Runt)/Jurkat-RUNX1-ChIP-Seq(GSE29180)/Homer(0.691)
	1e-14	ZFX(Zf)/mES-Zfx-ChIP-Seq(GSE11431)/Homer(0.790)
	1e-14	CRX(Homeobox)/Retina-Crx-ChIP-Seq(GSE20012)/Homer(0.688)
	1e-13	MEIS1/MA0498.2/Jaspar(0.906)
	1e-13	FoxL2(Forkhead)/Ovary-FoxL2-ChIP-Seq(GSE60858)/Homer(0.610)
	1e-13	AR-halbsite(NR)/LNCaP-AR-ChIP-Seq(GSE27824)/Homer(0.627)

#### d. Ros-L stage

Motif	P-value	Best Match/Details
	1e-156	TEAD4/MA0809.1/Jaspar(0.970)
	1e-61	FOSL1/MA0477.1/Jaspar(0.944)
	1e-45	AP-2alpha(AP2)/Hela-AP2alpha-ChIP-Seq(GSE31477)/Homer(0.967)
	1e-38	Tcf3(HMG)/mES-Tcf3-ChIP-Seq(GSE11724)/Homer(0.891)
	1e-27	Pax2/MA0067.1/Jaspar(0.780)
	1e-26	Nur77(NR)/K562-NR4A1-ChIP-Seq(GSE31363)/Homer(0.750)
	1e-23	Six1(Homeobox)/Myoblast-Six1-ChIP-Chip(GSE20150)/Homer(0.903)
	1e-21	MYB(HTH)/ERMYB-Myb-ChIPSeq(GSE22095)/Homer(0.749)
	1e-20	Dux/MA0611.1/Jaspar(0.880)
	1e-18	Rhox11/MA0629.1/Jaspar(0.851)
	1e-18	TFCP2/MA0145.3/Jaspar(0.756)
	1e-17	ZFX(Zf)/mES-Zfx-ChIP-Seq(GSE11431)/Homer(0.814)
	1e-17	FOXO3/MA0157.2/Jaspar(0.748)
	1e-17	KLF5(Zf)/LoVo-KLF5-ChIP-Seq(GSE49402)/Homer(0.804)
	1e-16	Bach1::Mafk/MA0591.1/Jaspar(0.632)

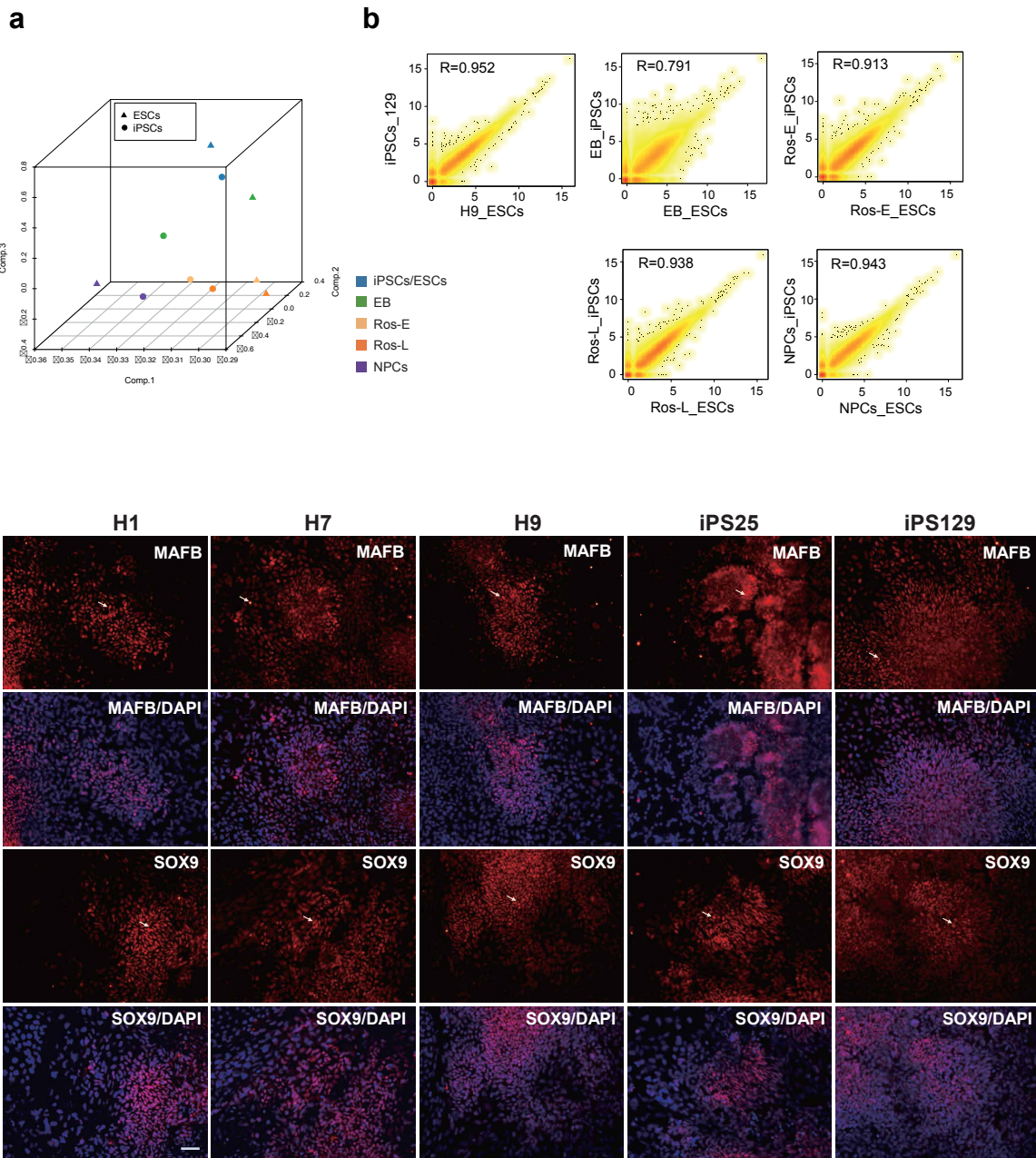
### e. NPCs stage

Motif	P-value	Best Match/Details
	1e-7054	AP-1(bZIP)/ThioMac-PU.1-ChIP-Seq(GSE21512)/Homer(0.990)
	1e-1065	CTCF(Zf)/CD4+-CTCF-ChIP-Seq(Barski_et_al.)/Homer(0.920)
	1e-928	Atf1/MA0604.1/Jaspar(0.898)
	1e-776	RUNX(Runt)/HPC7-Runx1-ChIP-Seq(GSE22178)/Homer(0.989)
	1e-524	Ascl1(bHLH)/NeuralTubes-Ascl1-ChIP-Seq(GSE55840)/Homer(0.961)
	1e-475	NFATC1/MA0624.1/Jaspar(0.870)
	1e-259	Ets1-distal(ETS)/CD4+-PolII-ChIP-Seq(Barski_et_al.)/Homer(0.935)
	1e-251	TFAP2A(var.2)/MA0810.1/Jaspar(0.780)
	1e-241	Smad4(MAD)/ESC-SMAD4-ChIP-Seq(GSE29422)/Homer(0.652)
	1e-226	TEAD(TEA)/Fibroblast-PU.1-ChIP-Seq(Unpublished)/Homer(0.754)
	1e-202	POL010.1_DCE_S_III/Jaspar(0.646)
	1e-189	c-Jun-CRE(bZIP)/K562-cJun-ChIP-Seq(GSE31477)/Homer(0.958)
	1e-186	POL010.1_DCE_S_III/Jaspar(0.711)
	1e-170	Foxo1(Forkhead)/RAW-Foxo1-ChIP-Seq(Fan_et_al.)/Homer(0.767)
	1e-168	POL010.1_DCE_S_III/Jaspar(0.795)

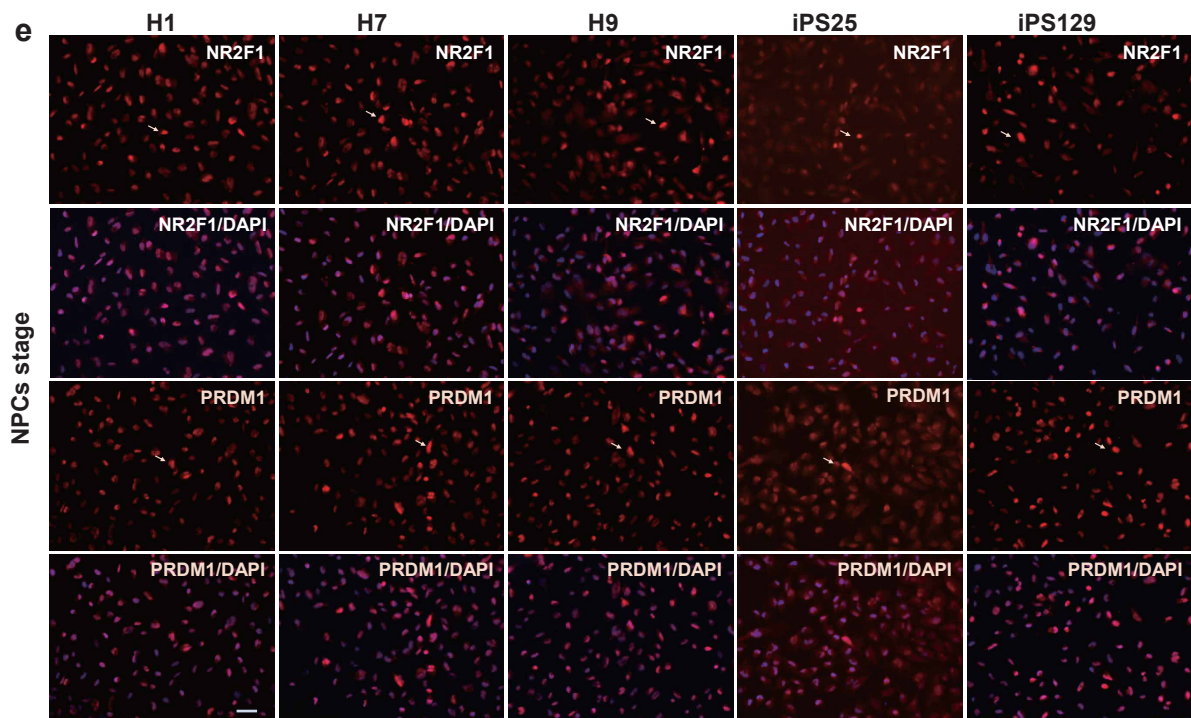
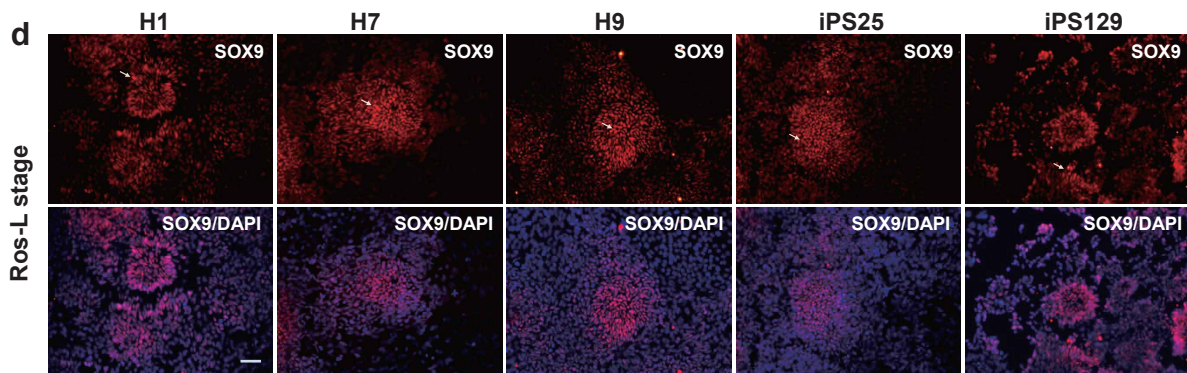
**Additional file 17: Figure S17. Transcription factor motifs enriched in stage specific peaks.** Motifs enriched in stage specific ATAC peaks were listed in tables containing the following information: motif, *P*-value and best match/details for iPSCs (a), EB (b), Ros-E (c), Ros-L (d) and NPCs stage (e), respectively.



# Additional file 18: Figure S18



# Additional file 18: Figure S18



**Additional file 18: Figure S18. Validation of neural differentiation in different genetic background cell lines.** **a** 3D PCA plot of the indicated cell stage derived from ESCs or iPSCs designated by colors and symbols. **b** The Pearson correlation coefficient between the corresponding cell stage derived from iPSCs and ESCs. **c, d, e** Immunostaining of MAFB and SOX9 at Ros-E stage (**c**), SOX9 at Ros-L stage (**d**), NR2F1 and PRDM1 at NPCs stage (**e**) across different genetic background cell lines (H1\_ESCs, H7\_ESCs, H9\_ESCs, iPS25 and iPS129). Scale bar represents 50  $\mu\text{m}$ .

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

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

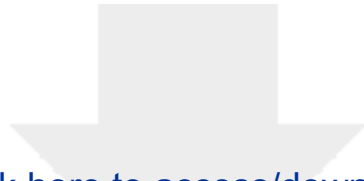
**Additional file 21: Table S3. Subpopulations interaction networks.**

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

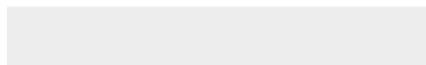
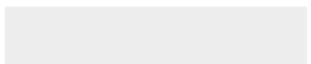


Click here to access/download  
**Supplementary Material**  
Additional file 19\_Table S1.xlsx



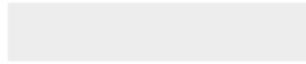


Click here to access/download  
**Supplementary Material**  
Additional file 20\_Table S2.xlsx





Click here to access/download  
**Supplementary Material**  
Additional file 21\_Table S3.pdf





Click here to access/download  
**Supplementary Material**  
Additional file 22\_Table S4.xls

