

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

Manuscript Number:	GIGA-D-18-00097R2	
Full Title:	Single-cell RNA-seq reveals dynamic transcriptome profiling in human early neural differentiation	
Article Type:	Research	
Funding Information:	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
	National Key R&D Program of China (2017YFA0104100)	Prof. Zhengliang Gao
Abstract:	<p>Background: Investigating cell fate decision and subpopulation specification in the context of the neural lineage is fundamental to understanding neurogenesis and neurodegenerative diseases. The differentiation process of neural-tube-like rosettes in vitro is representative of neural tube structures, which are composed of radially organized, columnar epithelial cells and give rise to functional neural cells. However, the underlying regulatory network of cell fate commitment during early neural differentiation remains elusive.</p> <p>Results: In this study, we investigated the genome-wide transcriptome profile of single cells from six consecutive reprogramming and neural differentiation time points and identified cellular subpopulations present at each differentiation stage. Based on the inferred reconstructed trajectory and the characteristics of subpopulations contributing the most 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>Conclusions: Our study provides a comprehensive and integrative study of the transcriptomics and epigenetics of human early neural differentiation, which paves the way for a deeper understanding of the regulatory mechanisms driving the differentiation of the neural lineage.</p> <p>Key words: single cell RNA-seq, ATAC-seq, neural differentiation, neural rosettes, neural tube, transcription factor, iPSCs</p>	
Corresponding Author:	Xun Xu, Ph.D BGI Shenzhen, Guangdong CHINA	
Corresponding Author Secondary Information:		
Corresponding Author's Institution:	BGI	
Corresponding Author's Secondary Institution:		
First Author:	Zhouchun Shang	
First Author Secondary Information:		

Order of Authors:	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	
Order of Authors Secondary Information:	
Response to Reviewers:	<p>Reviewer reports: Reviewer #1: The authors did great job addressing all the comments provided. I recommend publication.</p> <p>Minor point: The antibodies/dilution used in the revised manuscript should be included in the methods. Reply: We thank the reviewer for pointing out this incomplete description. Accordingly, we have added the related information in "Materials and Methods-Immunofluorescence staining" (page 29, line 823-835).</p> <p>Reviewer #2: The authors addressed all of my comments in a satisfying manner, and I recommend the revised manuscript for publication. I wish the authors luck with this interesting paper and their follow up studies. Reply: We thank the reviewer very much for the insightful and constructive suggestions, which helped us a great deal.</p> <p>Editor: (1) The antibodies/dilution used in the revised manuscript should be included in the methods. Reply: Thanks to the editor, we have added the related information in "Materials and Methods- Immunofluorescence staining" (page 29, line 823-835).</p> <p>(2) Please ensure that the protocols.io citations under "availability of supporting data" are listed in the references and only the reference number cited in the paper. Reply: Thanks to the editor, accordingly, we have modified this in the text (Availability of data and materials; page 35, line 994-997).</p>

Additional Information:	
Question	Response
Are you submitting this manuscript to a special series or article collection?	No
<p>Experimental design and statistics</p> <p>Full details of the experimental design and statistical methods used should be given in the Methods section, as detailed in our Minimum Standards Reporting Checklist. 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>Resources</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 Research Resource Identifiers (RRIDs) for antibodies, model organisms and tools, where possible.</p> <p>Have you included the information requested as detailed in our Minimum Standards Reporting Checklist?</p>	Yes
<p>Availability of data and materials</p> <p>All datasets and code on which the conclusions of the paper rely must be either included in your submission or deposited in publicly available repositories (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>	Yes

Have you have met the above requirement as detailed in our [Minimum Standards Reporting Checklist](#)?

[Click here to view linked References](#)

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

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

8 1 Shanghai Tenth People's Hospital, Tongji University School of Medicine, Shanghai,
9 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 BGI Education Center, University of Chinese Academy of Sciences, Shenzhen,
15 China

16 6 Shenzhen Key Laboratory of Neurogenomics, BGI-Shenzhen, Shenzhen, China

17 7 Department of Neurosurgery, Shenzhen Second People's Hospital, Shenzhen
18 University 1st Affiliated Hospital, Shenzhen, Guangdong, China

19 8 James D. Watson Institute of Genome Sciences, Hangzhou, China

20 9 The University of Queensland, Diamantina Institute (UQDI), Brisbane, QLD,
21 Australia

22 10 Laboratory of Genomics and Molecular Biomedicine, Department of Biology,
23 University of Copenhagen, Copenhagen, Denmark

24 #These authors contributed equally to this work.

25 *Correspondence should be addressed to Z.G. (zhengliang_gao@tongji.edu.cn) or
26 X.X. (xuxun@genomics.cn).

1 30 **Abstract:**

2 31 **Background:** Investigating cell fate decision and subpopulation specification in the
3
4 32 context of the neural lineage is fundamental to understanding neurogenesis and
5
6 33 neurodegenerative diseases. The differentiation process of neural-tube-like rosettes
7
8 34 *in vitro* is representative of neural tube structures, which are composed of radially
9
10 35 organized, columnar epithelial cells and give rise to functional neural cells. However,
11
12 36 the underlying regulatory network of cell fate commitment during early neural
13
14 37 differentiation remains elusive.

15 38 **Results:** In this study, we investigated the genome-wide transcriptome profile of
16
17 39 single cells from six consecutive reprogramming and neural differentiation time points
18
19 40 and identified cellular subpopulations present at each differentiation stage. Based on
20
21 41 the inferred reconstructed trajectory and the characteristics of subpopulations
22
23 42 contributing the most towards commitment to the central nervous system (CNS)
24
25 43 lineage at each stage during differentiation, we identified putative novel transcription
26
27 44 factors in regulating neural differentiation. In addition, we dissected the dynamics of
28
29 45 chromatin accessibility at the neural differentiation stages and revealed active
30
31 46 *cis*-regulatory elements for transcription factors known to have a key role in neural
32
33 47 differentiation as well as for those that we suggest are also involved. Further,
34
35 48 communication network analysis demonstrated that cellular interactions most
36
37 49 frequently occurred among embryoid body (EB) stage and each cell subpopulation
38
39 50 possessed a distinctive spectrum of ligands and receptors associated with neural
40
41 51 differentiation which could reflect the identity of each subpopulation.

42 52 **Conclusions:** Our study provides a comprehensive and integrative study of the
43
44 53 transcriptomics and epigenetics of human early neural differentiation, which paves the
45
46 54 way for a deeper understanding of the regulatory mechanisms driving the
47
48 55 differentiation of the neural lineage.

49 56 **Key words:** single cell RNA-seq, ATAC-seq, neural differentiation, neural rosettes,
50
51 57 neural tube, transcription factor, iPSCs
52
53 58

1 59 **Background**

2 60 The nervous system contains complex molecular circuitry in developmental processes.
3
4 61 In humans, there is a paucity of data describing early neural development and the
5
6 62 corresponding cellular heterogeneity at various stages. To our knowledge, neural tube
7
8 63 formation and closure is crucial for embryonic central nervous system (CNS)
9
10 64 development and the process of neurulation. Previous studies have reported that
11
12 65 neural tube closure is strongly controlled by both genetic and epigenetic factors and is
13
14 66 sensitive to environmental influences [1-3]. Perturbations in this delicately balanced
15
16 67 and orchestrated process can result in neural tube defects (NTDs) giving rise to birth
17
18 68 defects such as spina bifida, anencephaly and encephaloceles. However, the
19
20 69 formation and closure of the neural tube *in vivo* during week 3 and 4 of human
21
22 70 gestation is a transient event and is therefore difficult to capture. Moreover, the limited
23
24 71 accessibility of human abortive fetuses at such an early stage precludes a thorough
25
26 72 investigation of human early neural development.
27
28

29 73

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

1 88 commitment during early neural differentiation remain largely unknown.

2 89

3
4 90 The advance of single cell trans-omics technology has offered incisive tools for
5
6 91 revealing heterogeneous cellular contexts and developmental processes [9-11].
7
8 92 Single cell RNA-seq (scRNA-seq) has been applied to the study of cellular
9
10 93 heterogeneity as well as to the identification of novel subtypes or intermediate cell
11
12 94 groups in multiple contexts [12-15], and may help delineate unexpected features of
13
14 95 neural developmental biology and facilitate the study of cellular states and
15
16 96 neurogenesis processes. In the present study, we used scRNA-seq and ATAC-seq
17
18 97 (assay for transposase-accessible chromatin using sequencing) to investigate human
19
20 98 early neural differentiation. Our analysis reveals the landscape of the transcriptome
21
22 99 and *cis*-regulatory elements during this process and creates an unbiased
23
24 100 classification of cell subpopulations during differentiation, providing a comprehensive
25
26 101 description of transcriptomic and epigenetic patterns in cell fate decision. The
27
28 102 differentiation system of hiPSCs provides access to the very early stage of neural
29
30 103 development and may serve as a source of specialized cells for regenerative
31
32 104 medicine as well as supporting further investigations of neural tube defects.
33
34
35
36

37 105

38 106 **Data description**

39
40 107 Here, we applied a well-adopted neural induction protocol and generated neural
41
42 108 progenitor cells (NPCs) by forming neural rosettes *in vitro* [8, 16]. We analysed
43
44 109 several different differentiation stages of cells, including hiPSCs, embryoid body (EB),
45
46 110 early rosettes (hereafter termed Ros-E, post-3 days of rosette formation), late rosettes
47
48 111 (hereafter termed Ros-L, post-5 days of rosette formation), NPCs, and the original
49
50 112 somatic fibroblasts (Fib). scRNA-seq was performed at discrete time points (e.g., Fib,
51
52 113 iPSCs, EB, Ros-E, Ros-L and NPCs), and we captured 96, 80, 81, 82, 93, and 95
53
54 114 single cells, respectively, for each stage with the purpose of studying differentiation
55
56 115 transition events. We also captured bulk transcriptome profiles of the corresponding
57
58 116 neural differentiation stages derived from iPSCs and ESCs for validation. In addition,
59
60
61
62
63
64
65

1 117 bulk ATAC-seq with two biological replicates was applied to the cell stages iPSCs, EB,
2 118 Ros-E, Ros-L and NPCs to measure the regulome dynamics during neural
3
4 119 differentiation (Fig. 1a). The quality of sequencing data was evaluated and filtered by
5
6 120 a quality control (QC) pipeline developed in-house (see Methods for details).
7
8
9 121

10 122 **Analyses**

11 123 **Differential transcriptome and regulome dynamics throughout human early** 12 124 **neural differentiation**

13 125 Since the development of human ESCs and iPSCs, the ability to investigate human
14 126 neurogenesis and neurological diseases via an *in vitro* differentiation model has vastly
15 127 improved [4, 17]. Subsequently, artificial neural cells have been successfully
16 128 generated using a variety of protocols by several laboratories [18-23]. Here, we
17 129 followed a well-adopted neural induction protocol and generated NPCs by forming
18 130 neural rosettes via inhibition of TGF β , AMPK and BMP signalling pathways and
19 131 activation of the FGF signalling pathway [8, 16]. We analysed different differentiation
20 132 stages of the cells including iPSCs, EB, Ros-E, Ros-L, and NPCs as well as the
21 133 original somatic fibroblasts (Fib). The iPSC aggregates were induced to
22 134 neuroepithelial cells (NE) and followed by neural tube-like rosettes formation (Fig. 1b).
23 135 Firstly, pluripotency-associated transcription factors (TFs) (e.g., OCT4, NANOG) were
24 136 significantly expressed in hiPSCs, suggesting that these cells did exhibit a stem cell
25 137 phenotype. The subsequent formation of neural rosettes was confirmed by
26 138 morphology, apical localization of ZO-1, a tight junction protein, and co-localisation of
27 139 the neuroepithelial marker N-CADHERIN (N-CAD, also known as CDH2) at the
28 140 junctions. Additional neural markers such as PAX6, NESTIN, SOX2, and SOX1 were
29 141 also found to be highly enriched in the Ros-L stage (Fig. 1b).
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

142
143 Cell stages are usually determined by a complement of TFs or master regulators
144 which regulate hundreds of genes associated with various cellular functions. To study
145 the genomic features associated with open chromatin regions, we classified ATAC

1 146 peaks based on the location of the peak centre. More than 16,000 peaks were
2 identified for each cell stage (Additional file 1: Figure S1a) with the majority located in
3 introns and enhancers/promoters, genomic regions that are known to harbour a
4 variety of *cis*-regulatory elements and are subjected to regulation by TFs (Additional
5 file 1: Figure S1b). Furthermore, we observed that ATAC peaks were significantly
6 enriched at regions near transcription start sites (TSS) (Additional file 1: Figure S1c).
7 These observations were reproducible across two replicates with a very high Pearson
8 correlation coefficient (≥ 0.954) (Additional file 1: Figure S1d, e).
9
10
11
12
13
14
15
16
17
18

19 155 It is widely reported that chromatin structures undergo widespread reprogramming
20 during cell status transition, with some genomic regions becoming compacted or
21 opened, leading to the switching on or off of a repertoire of genes responsible for cell
22 fate decision [24-29]. We studied the dynamic chromatin landscape by tracing the
23 temporal origins of ATAC peaks at each stage with peaks non-overlapping with
24 existing ones that were annotated as novel peaks. We assumed that those peaks,
25 conserved among differentiation stages, are associated with housekeeping genes
26 while stage-dynamic peaks are likely to represent *cis*-regulatory elements important
27 for cell status transition. As expected, we observed the introduction of roughly 10-50%
28 of novel peaks in each stage, accompanied by the disappearance of several
29 pre-existing ATAC peaks. Notably, more novel peaks appeared at the NPCs stage
30 than at other stage (Fig. 1c). GO term analysis of genes residing in novel peaks
31 across the differentiation stages showed enrichment of “axon development”, “positive
32 regulation of nervous system development”, “epithelial tube morphogenesis”, “positive
33 regulation of neurogenesis”, “cell-cell signalling by Wnt”, “forebrain development”,
34 “hindbrain development”, “telencephalon development”, “neural precursor cell
35 proliferation”, and “cell fate commitment”. “Neurotrophin signalling pathway” was also
36 found to be enriched, but was specifically associated with NPCs. KEGG enrichment
37 analysis showed that “FoxO signalling pathway”, a pathway which is known to play an
38 important role in NPC proliferation, and “neuroactive ligand–receptor interaction” were
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

1 175 enriched in NPCs stage (Fig. 1d, e), suggesting that specific *cis*-regulatory elements
2 176 regulating neural differentiation are being staged (poised) for stem cell fate
3
4 177 specification and conversion.
5
6

7 178

8
9 179 To reveal the detail of chromatin accessibility dynamics during neural differentiation,
10
11 180 we also analysed the gained or lost peaks at each stage compared with the previously
12
13 181 neighbouring one. We observed that the number of gained peaks was with the largest
14
15 182 increase at the NPCs stage while the number of lost peaks was relatively high at
16
17 183 Ros-E stage (Additional file 2: Figure S2a). Next, we studied the genomic distribution
18
19 184 of these dynamic peaks and found that both the gained and lost peaks were located
20
21 185 mostly in distal intergenic regions and promoter regions (Additional file 2: Figure S2b).
22
23 186 This observation indicates that distal and promoter regions are more dynamic
24
25 187 compared to other genomic regions during neural differentiation process.
26

27 188

28
29 189 To gain insight into the potential function of closing (lost) peaks dynamics, we carried
30
31 190 out GO enrichment analysis on the genes associated with lost peaks at each stage.
32
33 191 The GO terms analysis showed that “mesoderm morphogenesis”, “endoderm
34
35 192 development”, “gastrulation” and “nodal signalling pathway” were solely enriched at
36
37 193 EB stage, indicating that upstream, as well as other lineage development, was
38
39 194 relatively repressed by closing related *cis*-regulatory regions. Other cell fate
40
41 195 conversion terms such as “neural crest cell differentiation”, “osteoclast differentiation”,
42
43 196 and “regulation of cartilage development” were enriched at Ros-E stage, together with
44
45 197 the annotation results of novel peaks, indicating that the chromatin accessibility
46
47 198 prepared for the neural lineage conversion by opening/closing up specific
48
49 199 *cis*-regulatory regions which facilitated the neural transition cascades (Fig. 1d, e and
50
51 200 Additional file 2: Figure S2d, e).
52

53 201

54
55
56 202 Furthermore, we identified stage-specific peaks at iPSCs, EB, Ros-E, Ros-L and
57
58 203 NPCs using motif enrichment analysis (see Methods). Further GO term and KEGG
59
60
61
62
63
64
65

1 204 enrichment analysis showed very similar results with annotation analysis of novel
2 205 peaks in corresponding cell stages (Additional file 3: Figure S3). These findings
3 206 strongly suggest that the novel, gained and lost, as well as stage-specific peaks,
4 207 represent cell status and cell fate transitions that progress neural differentiation and
5 208 that the landscape of *cis*-regulatory element accessibility throughout the differentiation
6 209 process is highly dynamic.
7
8
9
10
11
12
13
14

15 210
16 211 To more thoroughly investigate the molecular mechanisms governing neural
17 212 differentiation we profiled the transcriptomes of 527 single cells. Single cell RNA-seq
18 213 libraries were generated using Smart-Seq2 method [30], followed by sequencing
19 214 approximately 6 million reads per cell. For subsequent analysis, we focused on 445
20 215 cells that passed the quality control (QC, Methods, Additional file 4: Figure S4a, b) and
21 216 ERCC correlation filter (Methods, Additional file 4: Figure S4c). 7003 to 8560
22 217 expressed genes were detected per cell (Additional file 4: Figure S4d), including TFs
23 218 that were relatively highly expressed at the EB and NPCs stages, while, intriguingly,
24 219 pseudogenes were relatively highly expressed at the Ros-E and NPCs stages
25 220 (Additional file 4: Figure S4e). We also identified a variety of genes: 3524, 3855, 2023,
26 221 1804 and 6211 specifically expressed at the iPSCs, EB, Ros-E, Ros-L and NPCs
27 222 stages, respectively (Additional file 4: Figure S4f). Many of these stage-specific genes
28 223 include some well-known pluripotent genes (*NANOG*, *ID1*, *ID2*, *ZFP42*, *LIN28A*,
29 224 *DPPA4*); early neural markers (*SOX2*, *OTX2*, *OTX1*, *PAX6*); and genes that both
30 225 regulate neural development and are critical to proliferative NPCs (*SOX4*, *SIX3*,
31 226 *CDH2*, *ZIC2*) (Fig. 1f and Additional file 4: Figure S4h).
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49

50 227
51 228 Because the neural rosette recapitulates neural tube development *in vitro*, we paid
52 229 particular attention to the Ros-E and Ros-L stages. Unsurprisingly, a large proportion
53 230 of up-regulated genes in the Ros-E stage were associated with nervous system
54 231 development including *TFAP2A*, *CNTN4*, *GLI3*, *DLX5* and *OTX1*) (Fig. 1f). Of
55 232 particular interest is the gene *GRHL3*. Expression of this gene is associated with
56
57
58
59
60
61
62
63
64
65

1 233 neural tube closure in mice [31, 32] and we observed this gene to be highly expressed
2
3 234 at Ros-E in human cells, suggesting that its role in neural tube closure may be
4
5 235 conserved across mammals or possibly chordates. *TFAP2A* (transcription factor AP-2
6
7 236 alpha) and *TFAP2B* (transcription factor AP-2 beta) have been proposed as master
8
9 237 regulators of the neural crest cell and loss of function of transcription factor AP-2 in
10
11 238 mice is strongly associated with a cranial neural tube defect phenotype [33]. In our
12
13 239 system, *TFAP2B* and *TFAP2A* were relatively highly expressed at both the Ros-E and
14
15 240 -L stages, suggesting transcription factor AP-2 may coordinate the specialized distal
16
17 241 *cis*-regulatory elements for downstream regulations in human. We also observed
18
19 242 expression of *ANLN* (Anillin actin binding protein) at the Ros-L stage, suggesting that
20
21 243 neuronal migration and neurite growth might occur by the linking of RhoG to the actin
22
23 244 cytoskeleton in neural rosettes [34]. Similarly, our data showed that *AURKA* (aurora
24
25 245 kinase A) and *AURKB* (aurora kinase B) were both expressed at the Ros-L stage,
26
27 246 echoing previous findings that the aPKC-Aurora A-NDEL1 pathway plays an essential
28
29 247 role in neurite elongation through modulating microtubule dynamics [35]. Finally, the
30
31 248 neuron fate commitment protein, *TGFB2*, the nervous system development regulator,
32
33 249 *ZEB2*, and the neural precursor cell proliferation-associated protein, *IFT20*, were
34
35 250 enriched at NPCs stage (Fig. 1f).

36
37
38 251

39
40 252 An unexpected finding was that some of the most important neural TFs exhibited
41
42 253 heterogeneous expression within the same cell stage (e.g., *ZIC2*, *OTX2*, *HESX1*,
43
44 254 *DLX3*, *LHX5*) (Fig. 1f and Additional file 4: Figure S4h). This inspired us to dissect the
45
46 255 subpopulations of cells within each cell stage to better understand the significance of
47
48 256 this result.

49
50 257

51 52 258 **Heterogeneous cellular subpopulations were identified at each developmental** 53 54 259 **stage**

55
56 260 To evaluate the overall distribution of cells at each of the six stages during
57
58 261 reprogramming and neural differentiation, we first performed an unsupervised

1 262 analysis using all expressed genes (QC, see Methods) as input to t-distributed
2 263 stochastic neighbour embedding (t-SNE) for visualization. This analysis showed
3 264 distinct clusters for each differentiation stage, supporting our observation of
4 265 heterogeneous gene expression during these stages (Fig. 2a). Because previous
5 266 studies have showed that TFs and *cis*-regulatory elements are highly informative in
6 267 reflecting cell identity [36], we used a machine classifier to determine the subsets of
7 268 TFs that best clustered cells into putative cell populations. We were then able to
8 269 identify distinct subpopulations at each cell stage (Fib1, Fib2, EB1, EB2, EB3, Ros-E1,
9 270 Ros-E2, Ros-L1, Ros-L2, Ros-L3, NPC1, NPC2 and NPC3) (Methods, Fig. 2,
10 271 Additional file 5-8: Figure S5-8). As we found no remarkable differential expression of
11 272 pluripotency-associated genes (e.g., *NANOG*, *ID1*, *ID2*, *LIN28A*, *SOX2*, *DPPA4*,
12 273 *ZFP42*, *TRIM28*) at the iPSCs stage (Additional file 4: Figure S4g), we did not include
13 274 iPSCs in the following analyses.

14 275

15 276 **Fibroblasts (Fib) stage**

16 277 Fibroblasts (Fib) are a very well-adopted original somatic cell resource for iPSCs
17 278 reprogramming; many direct conversions from fibroblast to functional neurons have
18 279 been reported [37, 38]. Here, we dissected two subpopulations of human dermal
19 280 fibroblasts (Fib1 and Fib2) with distinct molecular features, showing significantly
20 281 higher expression of several important pluripotency- and neural-associated
21 282 transcription factors such as *SOX2*, *LIN28A*, *SOX11*, *ZIC2*, *FEZF1* and *SIX3* in Fib2
22 283 (Additional file 5: Figure S5a, b). GO terms identified by up-regulated genes between
23 284 the two subsets showed “chromosome segregation”, “positive regulation of nervous
24 285 system development”, “stem cell population maintenance”, “positive regulation of cell
25 286 cycle”, “neural precursor cell proliferation” and “chromatin remodeling” as solely
26 287 enriched in the Fib2 subpopulation (Additional file 5: Figure S5c). KEGG enrichment
27 288 analysis showed “cell cycle” term was specifically associated with the Fib2 subset
28 289 (Additional file 5: Figure S5d). Furthermore, we observed that fibroblasts were
29 290 distributed into two distinct groups called Fib-Group1 and Fib-Group2 based on their

1 291 location in the Fig. 2a. Of note, the majority of cells in Fib-Group1 and Fib-Group2
2 292 were composed of Fib1 and Fib2, respectively. Moreover, cells from Fib2 subset
3 293 clustered together with EB cells (Additional file 5: Figure S5e). Together with the
4 294 molecular features of Fib2 subset (Additional file 5: Figure S5b), we proposed Fib2
5 295 subset might possess high potential for iPSCs reprogramming and neural conversion.
6 296 Thus, based on the differentially expressed genes and CD markers dataset (HUGO
7 297 Gene Nomenclature Committee, HGNC), we further inferred several cell surface
8 298 markers of Fib2 (e.g., *FGFR2*, *F11R*, *PROM1*, *BST2*, *ITGA6* and *EPCAM*) although
9 299 these surface markers showed heterogeneously expressed levels within the Fib2
10 300 subset (Additional file 5: Figure S5f).

11 301

12 302 **Embryoid body (EB) stage**

13 303 For the three EB subpopulations (EB1, EB2 and EB3), we identified genes that were
14 304 up-regulated compared to the iPSCs stage, respectively. These genes were enriched
15 305 in “fetal brain cortex”, “epithelium” and “brain” terms by DAVID using tissue enrichment
16 306 analysis (Additional file 6: Figure S6a) which suggests that the biological processes of
17 307 brain development and neural differentiation initiation are occurring during the
18 308 iPSCs-to-EB stage transition and these processes are shared by each EB
19 309 subpopulation. Moreover, most neural TFs and cell-specific markers were expressed
20 310 commonly among EB subpopulations (e.g., *SOX2*, *ZIC2*, *SOX11*, *SOX4*, *SIX3*)
21 311 (Additional file 6: Figure S6b) and some of these TFs play a crucial role in neural tube
22 312 formation. However, some important neural TFs, such as *FOXO1* and *FOXO3*, which
23 313 play an important role in NPC proliferation and self-renewal [39]; *TULP3*, which
24 314 regulates the SHH signalling pathway and modulates neural tube development [40];
25 315 and *POU2F1*, which regulates *NESTIN* gene expression during P19 cell neural
26 316 differentiation and CNS development [41], showed significantly high expression in the
27 317 EB3 subpopulation, but low expression in the EB1 and EB2 subpopulations
28 318 (Additional file 6: Figure S6c, d). This suggests that different subpopulations contain
29 319 specific molecular signatures and different differentiation states or potentials.

1 320

2 321 **Early rosettes (Ros-E) stage**

3
4 322 During the Ros-E stage, which is composed of NE and the cells in the early stage of
5
6 323 rosette formation, we observed expression of several master regulator genes
7
8 324 associated with neural tube formation and closure including *SOX11*, *ZIC2*, *PAX3*, and
9
10 325 *SNAI2* in both Ros-E subgroups (Ros-E1 and Ros-E2). However, genes involved in
11
12 326 neural crest specifiers, such as *TWIST1* [42] and *SOX9*, which contribute to the
13
14 327 induction and maintenance of neural stem cells and are enriched in neural crest cells
15
16 328 [43-45]; and *ETS1*, which regulates neural crest development through mediating BMP
17
18 329 signalling [46], were preferentially expressed in the Ros-E1 subpopulation (Fig. 2b, c).
19
20 330 The ectoderm marker, *OTX1*, and genes involved in the ventral hindbrain marker (e.g.,
21
22 331 *IRX3*) were highly expressed in the Ros-E2 subgroup (Fig. 2b, c). GO term annotation
23
24 332 analysis showed Ros-E1 and Ros-E2 shared GO terms of “cell cycle G1/S phase
25
26 333 transition”, “G1/S transition of mitotic cell cycle”, “epithelial cell proliferation” and
27
28 334 “positive regulating of binding” (Fig. 2d) while “negative regulation of neuron
29
30 335 differentiation” and “tube morphogenesis” were solely enriched in the Ros-E2
31
32 336 subpopulation (Fig. 2d). KEGG enrichment analysis showed that “base excision
33
34 337 repair”, “DNA replication”, “axon guidance”, “cell cycle” and “mismatch repair” were
35
36 338 specifically associated with the Ros-E2 subset (Fig. 2e). We further performed
37
38 339 single-cell differential expression (SCDE) on both Ros-E subpopulations and identified
39
40 340 additional differentially expressed genes between the two groups. *SIX3*, *SIX6*,
41
42 341 *TFAP2B* and *PBX1* were more highly expressed in Ros-E2, whereas *EDN1*, *S100A10*
43
44 342 and other genes related to neural crest migration, were highly expressed in Ros-E1
45
46 343 (Fig. 2f).
47
48
49

50 344

51
52 345 **Late rosettes (Ros-L) stage**

53
54 346 At the Ros-L stage the genes *SNAI2*, *OTX2*, *FEZF1*, *ZIC3*, and *HESX1* showed
55
56 347 significantly different expression patterns among the three distinguishable
57
58 348 subpopulations (Ros-L1, Ros-L2 and Ros-L3) at the Ros-L stage (Additional file 7:
59
60
61
62
63
64
65

1 349 Figure S7a, b). Moreover, *SMAD1* and *MYC*, two components in the Wnt signaling
2 350 pathway which is critical for neural development [47, 48], were specifically enriched in
3
4 351 the Ros-L3 subpopulation. Additionally, *JUNB* from the TGF β signaling pathway was
5
6 352 preferentially expressed in Ros-L3 compared to the other two subpopulations.
7
8 353 Interestingly, *HAND1* and *ISL1*, which are mesoderm markers, and *TBX3*, which
9
10 354 elicits endodermal determination, were highly expressed in the Ros-L1 subpopulation
11
12 355 (Additional file 7: Figure S7a, b).
13
14
15 356

16
17 357 Of 648 GO terms identified by differentially expressed genes among these three
18
19 358 subsets, 52 terms were shared by Ros-L1 and Ros-L3, such as “positive regulation of
20
21 359 cell motility”, “angiogenesis”, “positive regulation of cellular component movement”
22
23 360 and “epithelium migration” (Additional file 7: Figure S7c). A high proportion of cardiac
24
25 361 development terms was enriched in Ros-L1, whereas DNA replication- and chromatin
26
27 362 remodeling-related terms and pathways were significantly associated with Ros-L2. In
28
29 363 addition, cell-substrate adhesion-related terms and cell cycle-related pathways were
30
31 364 enriched in Ros-L3 (Additional file 7: Figure S7c, d).
32
33 365

34
35 366 Several subpopulation-specific genes were identified, including *NR2F1*, *ARID3A*,
36
37 367 *SIX3*, *OTX2* and *FOXP1* at the NPCs stage (Additional file 8: Figure S8a, b). These
38
39 368 observations suggest that significant TF expression patterns describe discrepant cell
40
41 369 differentiation states or differentiation commitments inside the neural conversion
42
43 370 process. Taken together, our results suggest that the subpopulation analyses
44
45 371 accurately describe specific gene expression dynamics at each cell stage, which are
46
47 372 likely masked in bulk sequencing analyses. Additionally, extrapolating from these
48
49 373 observations, we can reason that reconstructing a differentiation trajectory based on
50
51 374 the gene expression dynamics of individual subpopulations would allow us to dissect
52
53 375 neural differentiation processes that we would otherwise be unable to observe.
54
55 376

56
57
58 377 **Tracking a reconstructed trajectory identifies key subpopulations during neural**
59
60
61
62
63
64
65

1 378 **differentiation**

2 379 Based on the subpopulations identified before, we wanted to track the gene
3 380 expression dynamics of individual subpopulations to parse the neural differentiation
4 381 processes and dissect the subpopulation with the highest contribution towards
5 382 commitment to the CNS lineage. First, we reconstructed the differentiation trajectory
6 383 using 8220 genes with variable expression. This showed that cells in stages from
7 384 iPSCs to NPCs followed a sequential differentiation process where each stage
8 385 exhibited a relatively discriminative region with some of the subpopulations
9 386 overlapping (Fig. 3a). Subsequently, based on the pairwise comparisons of TF
10 387 expression levels, we inferred the connection of the subpopulations from the iPSCs
11 388 stage to NPCs stage across the five-stage differentiation process (Fig. 3b). TF
12 389 expression levels were considered as strong indicators of cell stage and identity [36].
13 390 Here, we used the Pearson correlation coefficient to identify more biologically and
14 391 molecularly similar cell subpopulations and considered them as cells within the same
15 392 developmental lineage [49]. As a result, iPSCs, EB3, Ros-E2, Ros-L3 and NPC1 were
16 393 identified as the subpopulations contributing the most to commitment to the CNS
17 394 lineage (Fig. 3b). These findings were consistent with the specific gene expression
18 395 pattern in individual subpopulations. For instance, *SOX13*, expressed in the
19 396 developing nervous system and neural tube [50,51], *FOXO1* [39] and *TULP3* [40]
20 397 were significantly highly expressed in EB3 (Additional file 6: Figure S6c, d). *MAFB*, an
21 398 important TF in hindbrain identity [52], was enriched in Ros-E2 (Fig. 2b, c); and other
22 399 crucial neural development TFs, especially those involved in CNS development, such
23 400 as *OTX1*, *DLX3*, *DLX6*, *ZIC3*, *ZIC4*, and *IRX3*, also showed high expression in the
24 401 Ros-E2 subpopulation (Fig. 2b, c). Previously, we assumed that *GRHL3* might be
25 402 involved in neural tube closure; here, the results showed that *GRHL3* was indeed
26 403 significantly highly expressed in Ros-L3 (Additional file 7, Figure S7b). Additionally,
27 404 neural crest regulators (e.g., *ETS1*, *ELK3*, *SOX9*) were enriched in Ros-L3 (Additional
28 405 file 7, Figure S7b), suggesting that cell fate specification and differential cell status
29 406 might exist even within subset. Strikingly, Ros-E2 and Ros-L3 that were identified in

1 407 the dominant path to CNS lineage by correlation analysis were shown as a process of
2
3 408 sequential conversion in our reconstructed trajectory (Fig. 3a, c). The molecular
4
5 409 signature described by these subpopulations was consistent with the analysis that
6
7 410 identified the key contributing subpopulations and encouraged us to perform
8
9 411 additional cell fate decision analyses.

10 412
11
12 413 Of note, there was a clear divarication within the rosette stages (Ros-E and Ros-L)
13
14 414 across the differentiation trajectory, indicating cell fate decision might be made at this
15
16 415 bifurcation point (Fig. 3c). Here, we focused on the single cells in the rosette stages
17
18 416 and called them Branch 1, Branch 2 and Branch 3 based on their location in the
19
20 417 developmental trajectory (Fig. 3c). Branch 3 was composed of Ros-E1 (n=27), Ros-L1
21
22 418 (n=15) and small proportion of Ros-E2 (n=5) and Ros-L3 (n=9, Fig. 3c). Previously,
23
24 419 our observations showed that Ros-E1 was associated with neural crest cells (high
25
26 420 expression of *TWIST1*, *SOX9*, *ETS1*, *EDN1* and *S100A10*) and Ros-L1 was likely
27
28 421 related to mesoderm and endodermal determination (high expression of *HAND1*, *ISL1*
29
30 422 and *TBX3*), and these two subpopulations comprise the majority of cells in Branch 3.
31
32 423 Further, we performed a pairwise comparison of gene expression across the three
33
34 424 branches. The results showed that many neural TFs, such as markers of neural tube
35
36 425 formation (*SOX4* and *SOX11*); the NSCs self-renewal and proliferation regulator
37
38 426 *FOXO3*; and the NSC markers *NES*, *CDH2* and *FABP7*, were commonly expressed
39
40 427 across all three branches, indicating the capacity for neural tube development and
41
42 428 NSCs proliferation are a fundamental feature of neural rosettes (Additional file 9:
43
44 429 Figure S9a, b). Strikingly, *ZIC2*, a member of the ZIC family of C2H2-type zinc finger
45
46 430 proteins, associated with neural tube development [32], showed significantly low
47
48 431 expression in Branch 3 (Fig. 3d, e). Some other neural development markers (e.g.,
49
50 432 *ZIC3*, *HMGB2*, *ID1*, *SIX3*, *SIX6*, *NR6A1*) were significantly lowly expressed in Branch
51
52 433 3 but highly expressed in Branch 1 (Fig. 3d, e, Additional file 9: Figure S9a, c).
53
54 434 However, *TFAP2B*, encoding a member of the AP-2 family of TFs, and *ELK3*,
55
56 435 essential for the progenitor progression to neural crest cell [53], was significantly

1 436 highly expressed in Branch 3 but lowly expressed in Branch 2. Moreover, *SOX9*,
2 437 *SNAI2*, *S100A11* and *TFAP2A*, previously shown to be highly expressed in neural
3 438 crest cells [43,44,45,54], were markedly highly expressed in Branch 3, but not Branch
4 439 1 (Fig. 3d, e, Additional file 9: Figure S9a, c). *KLF5* and *IRF6* were significantly highly
5 440 expressed in Branch 3 as well (Fig. 3d, e). These two TFs have been reported to be
6 441 involved in phenotypic switching of vascular smooth muscle cells [55] and
7 442 development of the palate in vertebrates involving cranial neural crest migration [56],
8 443 respectively. These results indicate that cell fate specification might occur at the
9 444 bifurcation point and, based on the observations, we speculate that Branch
10 445 1-to-Branch 2 has progressed more towards CNS and Branch 3 is probably
11 446 composed of neural crest cells and other cells comprising this microenvironment.
12
13
14
15
16
17
18
19
20
21
22
23
24

25 447 26 448 **Construction of the TF regulatory network during cell status transition**

27 449 To infer TFs which drive the progression of cell status from one stage to the
28 450 neighbouring one, we performed SCDE analysis for those cell subpopulations
29 451 committing to CNS lineage, resulting in 58, 123, 98 and 131 TFs differentially
30 452 expressed among iPSCs vs EB3, EB3 vs Ros-E2, Ros-E2 vs Ros-L3, and Ros-L3 vs
31 453 NPC1 comparisons (Additional file 10, 11: Figure S10, 11 and Additional file 19: Table
32 454 S1). Interestingly, *PRDM1*, which has been proposed to promote the cell fate
33 455 specification RB sensory neurons in zebrafish [57], was significantly up-regulated
34 456 from Ros-E2 to Ros-L3 (Additional file 10: Figure S10). In contrast, several
35 457 well-characterized TFs were found to be significantly highly expressed in Ros-E2
36 458 (mainly resident in Branch 1) and down-regulated during the transition from early to
37 459 late rosette development: *FOXP1*, cooperating with *Bmi-1* to maintain neural stem cell
38 460 self-renewal in the forebrain; *MAFB*, the posterior CNS fate identifier and essential for
39 461 hindbrain choroid plexus development [52, 58]; *DLX3* and *DLX5*, neural plate border
40 462 specifier genes [58]; and *ID1*, a controller of stem cell proliferation during regenerative
41 463 neurogenesis in the adult zebrafish telencephalon [59]. These results suggest that the
42 464 expression patterns of neural-associated TFs undergo dramatic changes during
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60
61
62
63
64
65

1 465 neural differentiation with some TFs activated (*PRDM1*, etc.) and others repressed
2
3 466 (*MAFB*, *FOXP1*, *ID1*, etc.) (Additional file 10: Figure S10). Furthermore, it was
4
5 467 previously unknown that several of these TFs were involved in neural differentiation
6
7 468 so our results have expanded the known biological functions of these molecules.
8

9 469
10
11 470 Among the 131 TFs exhibiting differential expression from Ros-L3 to NPC1, 80 TFs
12
13 471 were up-regulated while 51 TFs were down-regulated (Additional file 11: Figure S11;
14
15 472 Additional file 19: Table S1). Up-regulated TFs included *SNAI2*, a neural crest
16
17 473 specifier [58]; *HIF1A*, required for neural stem cell maintenance and vascular stability
18
19 474 in the adult mouse [60]; *SIX1*, which drives the neuronal developmental program in
20
21 475 the mammalian inner ear [61]; *ETV1*, which orchestrates gene regulation during the
22
23 476 terminal maturation program of cerebellar granule cells [62]; and *POU3F3*, which
24
25 477 influences neurogenesis of upper-layer cells in the cerebral cortex [63] (Additional file
26
27 478 11: Figure S11). This is consistent with our previous observation that the main
28
29 479 trajectory has progressed more towards to CNS. Of particular interest is *PRDM1*,
30
31 480 whose expression increased from Ros-E2 to Ros-L3 and decreased during the
32
33 481 progression from Ros-L3 to NPC1 (Fig. 4g and Additional file 10, 11: Figure S10, 11),
34
35 482 suggesting that it might play multiple specific roles in neural differentiation.
36
37

38 483
39
40 484 Next, we inferred a regulatory network among those differentially expressed TFs
41
42 485 based on known interactions collected in the STRING database [64]. Our results
43
44 486 suggested that *SOX2* and *GATA3* were key regulators from iPSCs to EB3 (Additional
45
46 487 file 12: Figure S12a); *TP53*, *SOX2*, *RELA*, *SIX3*, *ARNTL*, *ISL1*, *RARA*, *TP63*, *GATA3*,
47
48 488 *SNAI2*, and *PAX3* were the key regulators from EB3 to Ros-E2 (Additional file 12:
49
50 489 Figure S12b); *MYC*, *SOX2*, *PAX6*, *EGR1*, *PBX1*, *GLI3*, *PAX3*, *SIX3*, *FOXP1*, *OTX2*,
51
52 490 *PAX7*, *PPARG*, *SOX9*, *MAFB*, *SIX6* and *ZIC1* were identified as key regulators from
53
54 491 Ros-E2 to Ros-L3 (Fig. 4a); and *SOX2*, *AR*, *MYCN*, *LEF1*, *PAX3*, *SNAI2*, *MSX1*,
55
56 492 *SOX9*, *NR3C1*, *PARP1*, *RUNX1*, *EBF1*, *HIF1A*, *IRF6*, *IRF1*, *KLF5*, and *LIN28A* were
57
58 493 predicted to be key regulators from Ros-L3 to NPC1 (Fig. 4b).
59
60
61
62
63
64
65

1 494

2 495 To dissect the *cis*-regulatory elements directing the expression of those regulators, we
3
4 496 selected the differentially expressed TFs that showed differential ATAC peaks
5
6 497 between neighbouring stages and performed motif scanning on the differential peaks.
7
8 498 Focusing on the transition from Ros-E2 to Ros-L3, we found transcription factor
9
10 499 binding sites (TFBSs) for TEAD2 and YY1 in a differential ATAC peak downstream of
11
12 500 the *PRDM1* gene (Fig. 4c). Multiple motifs for the transcription factor *TFAP2C* were
13
14 501 found in a differential peak located in the intron of the *ARID3A* gene, which is a
15
16 502 regulator responsible for the transition for Ros-L3 to NPCs (Fig. 4d). Based on the
17
18 503 temporal specificity of ATAC peaks and the existence of TF motifs in these regions, we
19
20 504 propose that those elements are stage-specific *cis*-regulatory elements regulating the
21
22 505 expression of neural regulators in response to their upstream regulatory TFs.
23
24
25
26

27 506

28 507 To infer the putative targets of key regulators, we combined the information from ATAC
29
30 508 peaks and motifs for TFs. All peaks containing motifs for a certain TF were annotated
31
32 509 as TF-related peaks and genes proximal to the peak were considered as potential
33
34 510 targets of that TF. Using these criteria, we predicted thousands of targets for each TF
35
36 511 (Additional file 20: Table S2). To dissect the regulatory network of each TF, we
37
38 512 conducted GO term and KEGG enrichment analysis for the putative target list of each
39
40 513 key regulator. Our results suggested that, from Ros-E2 to Ros-L3, the targets for
41
42 514 *PRDM1* were significantly enriched in pathways and GO terms associated with “axon
43
44 515 guidance”, “hippo signalling pathway” and “neurotrophin signalling pathway” (Fig. 4e
45
46 516 and Additional file 13: Figure S13a). From Ros-L3 to NPC1, targets for *NR2F1*, *SOX9*,
47
48 517 *TFAP2C* were enriched in KEGG pathways associated with “axon guidance” and
49
50 518 “hippo signalling pathway” (Additional file 13: Figure S13b-d). We further validated
51
52 519 *PRDM1* expression among different genetic background cell lines (H1_ESCs,
53
54 520 H7_ESCs, H9_ESCs, iPS25 and iPS129). The immunostaining showed that *PRDM1*
55
56 521 was expressed at Ros-L stage with heterogeneous expression level, though, the
57
58 522 scRNA-seq data was not at a high level. Moreover, the results were uniformed across
59
60
61
62
63
64
65

1 523 these cell lines (Fig. 4g, h).

2 524

3
4
5 525 **Inferring a cellular communication network among cell subpopulations within**
6
7 526 **specific differentiation stages**

8
9 527 Cell subpopulations with different functions are proposed to exhibit distinct expression
10
11 528 profiles of ligands and receptors which primes cells for cell-type-specific interactions
12
13 529 [65]. In this study, the cellular interactions were inferred using public ligand-receptor
14
15 530 databases (see Methods). Briefly, 360, 182, 261 and 307 ligands/receptors were
16
17 531 expressed within EB, Ros-E, Ros-L and NPCs subpopulations respectively, among
18
19 532 which 304, 55, 124 and 162 interactions were identified within subpopulations at each
20
21 533 differentiation time point (Fig. 5, Additional file 14-16: Figure S14-16 and Additional file
22
23 534 21: Table S3). The most frequent interactions were observed in the EB stage, implying
24
25 535 that cells communicate extensively to coordinate differentiation programs during
26
27 536 embryogenesis (Additional file 14: Figure S14). In contrast, much fewer interactions
28
29 537 were predicted after the EB stage, suggesting communications decreased
30
31 538 dramatically during the progression of lineage commitment. Notably, although
32
33 539 comparable number of ligands and receptors were detected at EB (181 receptors and
34
35 540 179 ligands) and NPCs (128 receptors and 179 ligands) stage, only half the
36
37 541 interactions (162) were inferred at NPCs stage compared to 304 ligand-receptor
38
39 542 interactions at EB stage (Additional file 14, 16: Figure S14, 16). The interactomes
40
41 543 among Ros-L cells, with 31, 32 and 34 receptors from Ros-L1, Ros-L2 and Ros-L3
42
43 544 interacting with ligands from other cell subpopulations were inferred (Fig. 5a, b). As
44
45 545 expected, several interactions involving receptors and ligands previously known to
46
47 546 play essential roles during neural development were identified in our study. For
48
49 547 example, *WNT5A* and *EPHB6* were enriched in Ros-L1. *FZD5* and *LPAR4* were
50
51 548 specifically expressed in Ros-L2. *PGF* and *ANGPT2* were up-regulated in Ros-L3
52
53 549 compared to other cell subpopulations (Fig. 5c, d, e). Overall, our study suggests that
54
55 550 the specific expression spectrum of ligands and receptors and corresponding
56
57 551 interactions can generally reflect the identity of cellular subpopulations.
58
59
60
61
62
63
64
65

1 552

2 553 **Discussion**

3 554 The regulation and molecular programs during embryonic neural development has
4
5 555 long been investigated. However, much of this work has been limited to model
6
7 556 organisms such as the mouse, zebrafish and *Drosophila* [36,40,56], due to the
8
9 557 scarcity of human fetal tissue for research purposes. Our understanding of human
10
11 558 early neural development, and particularly neural tube formation and the cell fate
12
13 559 commitments of neural precursors in early stages, is still incomplete. To circumvent
14
15 560 the challenges inherent in these investigations, namely the ability to study these
16
17 561 processes *in vivo* in humans, we used hiPSCs and induced differentiation *in vitro*
18
19 562 towards a neural cell fate using a well-established model. We characterised both the
20
21 563 transcriptional profiles in single cells as well as chromatin accessibility at several
22
23 564 critical stages during differentiation to inform this process at unprecedented resolution.
24
25 565 This study has unveiled the dynamic transcriptome and regulome underlying the
26
27 566 human early neural differentiation and identified functionally-distinct subpopulations
28
29 567 within the various stages to have a more precise description of the factors defining the
30
31 568 differentiation trajectory. Our analyses hint at the existence of a widespread regulatory
32
33 569 network between TFs and their target genes, especially those associated with cellular
34
35 570 reprogramming and differentiation. We were also able to construct minimal gene
36
37 571 expression profiles based on ligands and receptors in each cell subpopulation which
38
39 572 can be used to confidently infer cell identity.
40
41
42
43

44 573

45
46 574 During development *in vivo* the neuroectoderm folds to form the neural tube which is
47
48 575 then patterned into regionally specialized subunits composed of progenitor cells.
49
50 576 These cells subsequently give rise to regional progenies of neural cells [66]. There is
51
52 577 some controversy in this field that formation of the EB would introduce *in vitro* culture
53
54 578 variability in regional cells across different batches resulting in a relatively poor model
55
56 579 of neural differentiation. The "dual-SMAD inhibition" method (inhibiting the
57
58 580 SMAD-dependent TGF β and BMP signaling pathways) yielding neural epithelia in
59
60
61
62
63
64
65

1 581 "monolayer culture" conditions [18] could alleviate the above concern. However,
2 582 generation of neural rosettes morphology *in vitro* is considered equivalent to neural
3 583 tube formation, recapitulating neural tube structure, which we believe is a promising
4 584 research model for early neural differentiation. Neural differentiation of hiPSCs into
5 585 NPCs starts with initial neural induction by appropriate dosages and gradients of
6 586 many TFs and morphogenetic factors that are highly expressed in the developing
7 587 brain. In this study, the induction cocktail used in the neural differentiation included
8 588 SB431542, dorsomorphin, N2, B27, VEGF and bFGF supplemented at specific time
9 589 points. The self-renewal program in human iPSCs is switched off and differentiation
10 590 toward NE and NPCs is triggered [8, 16]. Previous results have shown that SB431542
11 591 enhances neural induction in EB derived from hESCs [67] by inhibiting the
12 592 Lefty/Activin/TGF β pathways and suppresses the mesodermal lineage (Brachyury)
13 593 induction [18, 42]. Consistent with these previous studies, in our *in vitro* system,
14 594 treatment with SB431542, in combination with dorsomorphin, results in a dramatic
15 595 decrease in *NANOG* expression and a concomitant increase in *PAX6* expression (Fig.
16 596 1f and Additional file 4: Figure S4h). In addition, *OTX2*, *ZIC2*, *SOX9*, *HESX1*, *MSX2*,
17 597 *DLX5*, *SOX4*, *SOX11*, and *SNAI2* were significantly activated during differentiation
18 598 which demonstrates that the transcriptional program triggering progression towards
19 599 NPCs was activated (Fig. 1f, Additional file 4: Figure S4h and Additional file 9: Figure
20 600 S9a-c). Taken together, these results indicate that the induction cocktail effectively
21 601 achieves efficient neural differentiation.

22 602
23 603 To measure the dynamic changes of *cis*-regulatory elements at each differentiation
24 604 stage, we performed ATAC-seq and chromatin accessibility analysis on bulk cells.
25 605 These results showed widespread and comprehensive chromatin structure
26 606 reprogramming during neural differentiation. In particular, TFBSs for several neural
27 607 master regulators were enriched in temporally dynamic ATAC peaks, indicating that
28 608 changes in chromatin accessibility are indeed associated with, and are probably
29 609 responsive to, the regulation of neural-related TFs. In addition, we also investigated

1 610 closing (lost) peaks dynamics as well as the functional annotation study, which was in
2
3 611 line with the corresponding annotation of novel peaks (Fig1d, e and Additional file 2, 3:
4
5 612 Figure S2, 3). We further identified several enriched TF motifs (e.g., *Pax2* in Ros-L
6
7 613 and *FOXO1* in NPCs) (Additional file 17: Figure S17) which are known to play an
8
9 614 important role in neural differentiation, consistent with results from previous studies
10
11 615 [39, 68].
12

13 616
14
15 617 By integrating single cell-based transcriptome profiling of 391 cells from five
16
17 618 differentiation stages, we identified a variety of TFs that were differentially expressed
18
19 619 throughout the differentiation process and showed distinct expression profiles among
20
21 620 specific cell stages. The TFs *SOX2*, *PAX6*, *OTX2*, *SOX4*, *ZIC2*, *LHX5*, *HESX1*, and
22
23 621 *SIX3* were significantly highly expressed at the EB stage (Fig. 1f). It has been reported
24
25 622 that members of the grainyhead-like (*Grhl*) family of TFs, which are well-conserved
26
27 623 from *Drosophila* to human, are highly expressed during neurulation in mice and that a
28
29 624 *Grhl3*-hypomorphic mutant resulted in NTDs [32, 69]. Remarkably, our results showed
30
31 625 that two human *Grhl* family TFs, *GRHL2* and *GRHL3*, were significantly highly
32
33 626 expressed at EB and Ros-E stage, respectively (Fig. 1f and Additional file 4: Figure
34
35 627 S4h), and the downstream targets of *GRHL2* (including *E-CADHERIN*, also known as
36
37 628 *CDH2*) were highly expressed at the neural rosette stage (Fig. 1b) supporting a role
38
39 629 for *Grhl* TFs in neural tube closure in humans. In addition, previous studies have
40
41 630 shown that in the *Drosophila* olfactory system the homeobox gene *distal-less* is
42
43 631 required for neuronal differentiation and neurite outgrowth [34]. Our data showed that
44
45 632 four homologs of *distal-less* (*DLX3*, *DLX4*, *DLX5*, *DLX6*) were significantly up
46
47 633 regulated at the Ros-E stage and were highly expressed in the Ros-E2 subpopulation
48
49 634 (Fig. 1f, Fig. 2b and Additional file 4: Figure S4h) implying that the *distal-less* gene
50
51 635 family plays a role in neural differentiation in humans.
52

53 636
54
55
56 637 We also applied scRNA-seq to our *in vitro* neural model to dissect the subpopulations
57
58 638 present at each differentiation stage (Fig. 2 and Additional file 5-8: Figure S5-8). We
59
60
61
62
63
64
65

1 639 were then able to reconstruct a differentiation trajectory based on the subpopulations
2 640 that we identified by variable TF expression within each stage (Fig. 3a). Strikingly, a
3 641 divarication within the rosette stage across the differentiation trajectory was observed.
4 642 Comparing Branch 1 to Branch 3, Branch 3 possessed the relatively lowly-expressed
5 643 TFs *LHX5*, *HESX1* and *SIX3* (reported as anterior forebrain markers), as well as other
6 644 crucial neural TFs (*SOX2*, *HMGB2*, *ZIC2*, *OTX1*, *FEZF1*); and the relatively
7 645 highly-expressed TFs *TFAP2B*, *SOX9*, *ELK3*, and *SNAI2* (Fig. 3d, e and Additional file
8 646 9: Figure S9a, c) which are considered to be neural crest markers [53]. Though *SNAI2*
9 647 was also expressed at the NPCs stage, combined with other neural crest markers, we
10 648 proposed that Branch 3 was progressing more towards to neural crest cells (Fig. 3a-c
11 649 and Additional file 9: Figure S9a-c). Taken together, these observations imply that the
12 650 main differentiation trajectory (Branch 1 and Branch 2) is heading towards CNS,
13 651 whereas Branch 3 is progressing towards neural crest cells.
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28

29 652
30 653 It is important to note that the current scRNA-seq method by its nature only provides a
31 654 snapshot of the gene expression profile for individual cells. A possible resolution for
32 655 the above problem is to capture the sample with much more precise time points,
33 656 which may, to some extent, overcome this limitation. Thus, in spite of the very
34 657 interesting heterogeneity and cell fate commitment study inferred above, we cannot
35 658 exclude the following factors that may affect cell subset identification in the above
36 659 description; 1) temporal transcriptional states during transient differentiation process;
37 660 2) differentiation efficiency; and lagging and leading cells remaining in the
38 661 differentiation process. However, we propose that the subsets dissection analysis
39 662 facilitates a more precise description of the factors defining the differentiation
40 663 trajectory. When we constructed the differentiation trajectory using the cells that
41 664 collected at different time points, the results showed that all subpopulations in stages
42 665 from iPSCs to NPCs followed a sequential differentiation process where each stage
43 666 exhibited a relatively discriminative region with some of the subpopulations
44 667 overlapping (Fig. 3a), indicating that in spite of the above concerns, the trajectory was
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60
61
62
63
64
65

1 668 established by the natural features of the respective subsets and which is also
2 669 supported by the observations that Ros-L2 possessing many early neural
3 670 differentiation TFs, such as *SOX2*, *OTX2*, *PAX6*, *OTX1*, and *LHX5*, as well as
4 671 forebrain markers (e.g., *HESX1*) and pluripotency-related TFs (*NANOG*, *SALL4*,
5 672 *PRDM14*) (Additional file 7: Figure S7a, b) were located in the reconstructed trajectory
6 673 prior to the generation of Ros-E populations (Fig. 3a, c). In addition, we carried out the
7 674 cell fate commitment analysis using Branch1, Branch2 and Branch3 which were
8 675 grouped based on the cell locations on the trajectory rather than cell subsets identified
9 676 by Seurat in order to minimize the above concerns.

10 677
11 678 Notably, our study reveals the regulatory network of TFs that are differentially
12 679 expressed among neighbouring cell subpopulations to be likely candidates for
13 680 promotion of cell fate transition. Based on the topology of this network, we focused on
14 681 novel regulators (*PRDM1* and *ARID3A*), especially *PRDM1*, which are located on the
15 682 hub of the network, interacting with both known and novel neural regulators. Although
16 683 the roles of several TFs have been reported during neural differentiation and brain
17 684 patterning formation in humans, some TFs have been proposed to play a role in neural
18 685 fate commitment in non-human species (mouse and zebrafish). However, the
19 686 interaction partners, *cis*-regulatory elements, and genetic regulatory networks of those
20 687 TFs are yet to be resolved. Here, we identified the *cis*-regulatory elements for *PRDM1*
21 688 and *ARID3A* genes and predicted their upstream regulators. Of particular interest,
22 689 *TFAP2C*'s role in regulating neural development has been widely reported, increasing
23 690 the confidence of our predictions. In humans, *PRDM1* is reported to promote germ cell
24 691 fate by suppressing neural effector *SOX2*, but the function of *PRDM1* in neural
25 692 development is unknown. In zebrafish, *Prdm1a*, the homolog of the *PRDM1* gene,
26 693 directly activates *foxd3* and *tfap2a* during neural crest specification [57]. Mutation of
27 694 *prdm1* in zebrafish resulted in severe phenotypes with a decrease in the quantity of
28 695 neural crest cells and the reduction in the size of structures derived from the neural
29 696 crest [57]. Similarly, strong expression of *prdm1* was observed in the neural plate

1 697 border of a basal vertebrate lineage, lamprey, implying that the role of *prdm1* in the
2 698 neural crest formation is likely a conserved, ancestral role [70]. Conversely, *prdm1* is
3
4 699 dispensable for neural crest formation in mice, and instead is required for primordial
5
6 700 germ cell specification suggesting that the neural crest specification function of *prdm1*
7
8 701 in mice has been lost [71]. Overall, previous studies suggest that functions of *prdm1*
9
10 702 are quite diverse and need to be investigated in species-, developmental-, and
11
12 703 environmental-specific manners. Based on the known interaction between *PRDM1*
13
14 704 and *SOX2* in humans, as well as the observation that *PRDM1* expression increased
15
16 705 significantly from Ros-E2 to Ros-L3 and was preferentially expressed in Ros-L3
17
18 706 compared to other two subpopulations in the rosette stage (Fig. 4g, h; Additional file 7:
19
20 707 Figure S7a, b and Additional file 10: Figure S10), we propose *PRDM1* as a novel
21
22 708 neural regulator in early human neural differentiation. Our hypothesis is supported by
23
24 709 the GO term and KEGG enrichment analysis of putative targets of *PRDM1*, which are
25
26 710 significantly enriched in “axon guidance” and hippo pathway-associated terms (Fig. 4e
27
28 711 and Additional file 13: Figure S13a). However, the functions of putative TFs need to be
29
30 712 further investigated using experimental methods.
31
32

33 713

34
35 714 To infer cellular interactions, communication network analysis was applied to the
36
37 715 expression profiles of ligands and receptors in stage-specific subpopulations. Two
38
39 716 trends were observed in our cellular interaction network analysis: 1) the frequency of
40
41 717 cellular interactions peaked at EB stage; and 2) different cell subpopulations showed
42
43 718 a certain degree of specificity in their ligand-receptor spectrum. The observation that
44
45 719 most interactions were inferred at the EB stage likely reflects the extensive cellular
46
47 720 communication during embryogenesis and early neural differentiation (Additional file
48
49 721 14: Figure S14). Regarding the ligand-receptor expression spectra, matched ligand
50
51 722 and receptor expression probably underlies the common functions shared by different
52
53 723 cell subpopulations within the same stage. In contrast, those specific ligands or
54
55 724 receptors probably reveal the unique regulatory code of distinct cell subpopulations.
56
57 725 For example, *WNT5A*, a crucial regulator of neurogenesis during the development of
58
59
60
61
62
63
64
65

1 726 cerebellum, and *BMP4*, one of the key regulators of dorsal cell identity in the neural
2 727 tube [72], were highly expressed in Ros-L1 compared to other cell subpopulations (Fig.
3
4 728 5c). *FZD5* (required for eye and retina development in mouse [73]), and *FGF19*
5
6 729 (required for forebrain development in zebrafish [74]) were preferentially expressed in
7
8 730 Ros-L2 (Fig. 5d and Additional file 22: Table S4). *WNT7A*, involved in several aspects
9
10 731 of neurogenesis, including synapse formation and axon guidance [75] and *FGF1*,
11
12 732 which maintains the self-renewal and proliferation of NPCs [76], were specifically
13
14 733 expressed in Ros-L3 (Additional file 22: Table S4). Pavličev et al. inferred the cell
15
16 734 communication network of the maternal-fetal interface and found that ligand-receptor
17
18 735 profiles could be a reliable tool for cell type identification [65]. Consistent with their
19
20 736 findings, our study suggests that the repertoire of ligands-receptors in neural cell
21
22 737 types could probably, to some extent, represent the identity of cell subpopulations.
23
24
25
26

738

27 739 There might be a concern that we only used one genetic background cell line for this
28
29 740 study, possibly making the cogency of our findings limited. To address this, we
30
31 741 performed ESCs neural differentiation and captured bulk transcriptome profiles of the
32
33 742 corresponding differentiation stages (ESCs, EB, Ros-E, Ros-L and NPCs). The
34
35 743 observations in ESCs were reproducible in iPSCs with regards to 1) PCA analysis
36
37 744 (Additional file 18: Figure S18a); 2) with a high Pearson correlation coefficient
38
39 745 between the corresponding cell stage derived from iPSCs and ESCs (Additional file
40
41 746 18: Figure S18b); and 3) validation analysis of subset- specific markers (*MAFB*, *SOX9*,
42
43 747 *PRDM1* and *NR2F1*). In addition, novel neural TF (*PRDM1*) expression in different
44
45 748 genetic cell lines (*H1_ESCs*, *H7_ESCs*, *H9_ESCs*, *iPS25* and *iPS129*) was consistent
46
47 749 with the above heterogeneity study (Fig. 4h and Additional file 18: Figure S18c, d, e).
48
49 750 Together, our findings are supported by different genetic cell lines mitigating the
50
51 751 concern that our results are limited to the cells forming the basis of this study.
52
53
54
55

752

56 753 Through differential expression analysis, we identified genes specifically expressed at
57
58 754 each stage which include both cell status master regulators such as TFs and
59
60
61
62
63
64
65

1 755 signalling components, as well as realizators [24] which could directly determine cell
2 756 growth, cell proliferation, cell morphology and cell-cell interaction. Within each stage,
3
4 757 we identified subpopulations with distinct expression signatures, which might
5
6 758 represent functional cell clusters or transient cell state given that neural cells have
7
8 759 been shown to demonstrate significant heterogeneity as they express different
9
10 760 surface proteins, exhibit diversified morphologies and secrete a variety of cytokines.
11
12 761 Therefore, it is necessary to explore the heterogeneity of cell subpopulations and
13
14 762 study each subpopulation in a case-by-case manner. In summary, our data show
15
16 763 conclusively that both transcriptome and regulome dramatically change during neural
17
18 764 differentiation, which affects a variety of biological pathways crucial for neural
19
20 765 differentiation. We also propose several putative TFs as well as the ligands-receptors
21
22 766 interaction spectrum that are important in each differentiation stage which paves the
23
24 767 way for a deeper understanding of the cell fate decision and regulatory mechanisms
25
26 768 driving the differentiation of the neural lineage.
27
28
29
30
31

32 769

33 770 **Materials and methods**

34 771 **Ethics statement**

35 772 The study was approved by the Institutional Review Boards on Ethics Committee of
36
37 773 BGI (Permit No.BGI-IRB 14057). The participant (dermal fibroblast, Fib129) signed
38
39 774 informed consent and voluntarily donated the samples for our study.
40
41
42
43
44

45 775

46 776 **Cell culture and reprogramming**

47 777 The human fibroblast cell line was derived from the dermal skin of a healthy donor
48
49 778 with written informed consent. Briefly, the skin tissue was washed with DPBS several
50
51 779 times, sliced into approximately 1mm or smaller fragment size, enzymatically
52
53 780 dissociated in High Dulbecco's modified Eagle medium (H-DMEM, Gibco, 11965118)
54
55 781 with 100U/ml collagenase type IV incubating in 37°C overnight, then 0.05% trypsin
56
57 782 incubating for 5 min. The dissociation was terminated by adding 2 ml fibroblast cell
58
59
60
61
62
63
64
65

1 783 culture medium (H-DMEM +10% FBS + 5ng/ml bFGF+ 2mM Gln) followed by
2 784 centrifugation at 300g for 5 min. The cells were resuspended with fibroblast cell
3 785 culture medium, and cultured at 37°C in a 5% CO₂ incubator. The fibroblast cell
4 786 culture medium was changed every 2 days until reaching 80%–90% confluence and
5 787 cells were passaged every 3-4 days.

6 788
7 789 For reprogramming, non-integrative human iPSCs were generated following a
8 790 modified Shinya Yamanaka method [77]. Briefly, 5x10⁵ human fibroblast cells at
9 791 passage 4 were nucleofected with the program for human dermal fibroblast NHDF
10 792 (Lonza, CC-2511) with 2.4ug episomal plasmids, including pCXLE- hOCT3/4-
11 793 shp53-F (Addgene, 27077), pCXLE- hSK (Addgene, 27078), pCXLE- hUL (Addgene,
12 794 27080). Transfected cells were cultured in a six-well plate with culture medium
13 795 containing H-DMEM supplemented with 10% FBS. The cells were trypsinized and
14 796 1x10⁵ cells were seeded onto a 10cm² dish covered with feeder and cultured in a
15 797 medium containing H-DMEM with 10% FBS while reaching 80% confluence. After that,
16 798 the medium was changed to hiPSCs medium containing DMEM/F12 (Gibco,
17 799 11320-033), 20% KSR (Gibco,10828-028), 2mM L-glutamine (Sigma, G8540), 0.1μM
18 800 NEAA (Gibco,11140-050), 0.1μM β-Mercaptoethanol (Gibco, 21985-023) and 10ng/ml
19 801 human bFGF (Invitrogen, PHG0021). The iPSCs colonies were picked at around day
20 802 25 and maintained in hiPSCs medium.

21 803

22 804 **Neural differentiation**

23 805 We applied a well-adopted neural differentiation protocol [8,16]. Briefly, human iPSCs
24 806 were maintained as described above. To induce neural rosettes, hiPSCs were
25 807 mechanically picked and washed with DMEM/F12 twice, and then cultured for 4 days
26 808 in suspension with 5μM dorsomorphin (Sigma, P5499) and 5μM SB431542 (Sigma,
27 809 S4317) in hiPSCs medium without bFGF for embryoid body (EB) formation, then the
28 810 EBs were attached on matrigel (BD, 354277) coated dishes (BD, 354277) and
29 811 cultured in DMEM/F12 (Gibco, 11320-033) supplemented with 20 ng/ml bFGF, 1xN2

1 812 (Gibco, 17502-048) and 2ug/ml heparin (Sigma, 1304005) for an additional 3 or 5
2 813 days to harvest rosette-early (Ros-E) and rosette-late (Ros-L) cells, respectively. To
3 814 collect neural progenitor cells (NPCs), rosettes structure that appeared in the center of
4 815 attached colonies at Ros-L stage were carefully harvested using pulled glass pipettes
5 816 and seeded on matrigel-coated dishes and cultured in DMEM/F12 supplemented with
6 817 1x N2, 1x B27 (Gibco,12587-010), 20 ng/ml bFGF, 20 ng/ml EGF (Invitrogen,
7 818 PHG0311) and 2ug/ml heparin (Sigma,1304005) for additional 7 days, and the
8 819 medium was changed every 2 days. At day 16, the NPCs reaching approximately 80%
9 820 confluence were collected, and all the mass or adherent cell samples were treated
10 821 with TrypLE™ Express Enzyme (Gibco, 12604-021) for single cell dissociation and
11 822 cryopreservation in gas-phase liquid nitrogen for further sequencing.
12
13
14
15
16
17
18
19
20
21
22
23
24

25 824 **Immunofluorescence staining**

26 825 Cells were fixed in 4% paraformaldehyde in DPBS for 20 min and permeabilized with
27 826 1% Triton X-100 for 20 min at room temperature. After 60 min blocking with 2% normal
28 827 goat serum, cells were incubated with primary antibody overnight at 4 °C, washed,
29 828 then stained with secondary antibodies (1:300, goat anti rabbit IgG-Cy3; or 1:300,
30 829 goat anti mouse IgG-Cy3) for 60 min at room temperature and then washed three
31 830 times with PBS. The primary antibodies for respective cells include OCT4 (1:200,
32 831 Abcam), NANOG (1:200, Abcam), PAX6 (1:200, Abcam), SOX2 (1:200, Abcam),
33 832 NESTIN (1:200, Abcam), SOX1 (1: 200, Abcam), ZO-1 (1:100, Abcam), N-CAD (1:
34 833 100, Abcam), MAFB (1:300, Sigma), SOX9 (1:300, EMD Millipore), PRDM1 (1:200,
35 834 Cell Signaling Technology, CST) and NR2F1 (1:300, R&D Systems). DAPI (1:500)
36 835 was used as counter-staining for nuclei. The images were captured and analyzed with
37 836 the Olympus IX73 and Image J.
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53

54 838 **Single cell RNA sequencing**

55 839 Cells at indicated time points were collected for single cell RNA-seq and global
56 840 transcriptome analysis. TrypLE™ Express Enzyme (Gibco, 12604-021) was applied
57
58
59
60
61
62
63
64
65

1 841 for single cell dissociation. Single-cell RNA-seq library construction was conducted
2
3 842 according to an automated pipeline called microwell full-length mRNA amplification
4
5 843 and library construction system (MIRALCS) as described previously [78]. 50bp
6
7 844 single-end sequencing was performed using the BGISEQ-500 platform.
8
9 845

10 846 **Assay for transposase-accessible chromatin sequencing (ATAC-seq)**

11
12 847 We profiled open chromatin accessibility sequencing (ATAC-seq) of neural
13
14 848 differentiation process for five stages including iPSCs, EB, Ros-E, Ros-L and NPCs
15
16 849 samples. ATAC-seq libraries were prepared using a modified protocol based on
17
18 850 previous study [79]. Briefly, 50,000 cells were collected for each sample, washed with
19
20 851 pre-cooling PBS and resuspended in 50 μ l of ice-cold lysis buffer (10 mM Tris-HCl, pH
21
22 852 7.5, 10 mM NaCl, 3 mM MgCl₂, 0.1% IGEPAL CA-630). Permeabilized cells were
23
24 853 resuspended in 50 μ l transposase reaction buffer (1 \times TAG buffer, 2.0 μ l Tn5
25
26 854 transposase enzyme) and incubated for 30 min at 37 $^{\circ}$ C. PCR amplification and size
27
28 855 selection (150–500 bp) were performed using Agincourt AMPure XP (Beckman
29
30 856 Coulter) and Bioanalyzer 2100 (Agilent). Libraries were pooled at equimolar ratios
31
32 857 with barcodes and sequenced on BGISEQ-500 platform.
33
34
35
36 858

37 859 **Pre-processing and quality control of single cell RNA-seq**

38
39 860 The original FASTQ data of the 527 samples were aligned to the rRNA database
40
41 861 (downloaded from NCBI) to remove rRNAs and the remaining reads were processed
42
43 862 with SOAPnuke (version 1.5.3) [80] to trim adaptors and filter out the low-quality reads.
44
45 863 The filtered data were aligned to the reference genome (hg19) using hisat2 (HISAT2
46
47 864 version 2.0.1-beta) [81]. Reads were counted using the R package
48
49 865 GenomicAlignments [82] (mode='Union', inter.feature=FALSE), and normalized to
50
51 866 RPKM with edgeR [83]. Cells were filtered using following parameters: genome
52
53 867 mapping rate more than 70%, fraction of reads mapped to mitochondrial genes less
54
55 868 than 20%, mRNA mapping rate more than 80%, ERCC ratio less than 10%, and gene
56
57 869 number more than 5000. Further, correlation of ERCC among cells was used to
58
59
60
61
62
63
64
65

1 870 evaluate the quality of each cell (threshold=0.9). At last, 445 single cells remained for
2
3 871 further analysis in this project.

4
5 872

6 873 **Identification of differentially expressed genes**

7
8 874 Differential expression of genes in iPSCs (n = 71 cells), EB (n = 57 cells), Ros-E (n =
9 875 81 cells), Ros-L (n = 92 cells), and NPCs (n = 90 cells) was determined using SCDE
10 876 (single cell differential expression analysis) [84] with default parameters except
11 877 requiring a minimum of 100 genes (parameter min.lib.size = 100 to call scde.error.
12 878 models function). The Z scores and corrected Z scores (cZ) to adjust for the multiple
13 879 testing were converted into two-tailed p-values and adjusted to control for FDR using
14 880 pnorm function in R. The significantly differentially expressed genes were selected
15 881 based on following criteria: adjusted p-value < 0.01 and fold-change > 2.

16 882

17 883 **Constructing trajectory using differentially expressed genes**

18 884 Monocle [85] ordering was conducted for all iPSCs, EB, Ros-E, Ros-L and NPCs cells
19 885 using the set of variable genes with default parameters except we specified
20 886 reduction_method = "DDRTree" in the reduceDimension function. The variable genes
21 887 were selected using the Seurat R package [86].

22 888

23 889 **Analysis of heterogeneity in each cell stage**

24 890 The heterogeneity of each cell stage was determined using Seurat R package [86] by
25 891 the normalized expression level of reported transcription factors (retrieved from
26 892 AnimalTFDB 2.0) [87]. Briefly, PCs with a p-value less than 0.01 were used for cell
27 893 clustering with reduction.type="pca" and resolution="1.0". The FindAllMarkers function
28 894 of Seurat package was used to identify marker genes for each cluster using default
29 895 parameters.

30 896

31 897 **ATAC peak calling**

32 898 We aligned ATAC-seq data to hg19 using Bowtie2 [88] and called peaks using MACS2

1 899 [89]. We established a standard peak set by merging all overlapping peaks. The IDR
2 900 pipeline [90] was used to identify reproducible peaks between two biological replicates.
3
4 901 Only peaks with $IDR \leq 0.05$ were considered reproducible and retained for
5
6 902 downstream analysis. Pearson correlation coefficients of two biological replicates at
7
8 903 each stage were calculated. Stage-specific peaks were defined as peaks having no
9
10 904 overlap with any peaks in other stages. Novel peaks were defined as peaks
11
12 905 non-overlapping with previous stages. In the case of iPSCs, all peaks were annotated
13
14 906 as novel peaks.
15
16
17 907

19 908 **Targets assignment of ATAC peaks**

21 909 For reproducible peaks, we applied HOMER [91] to assign putative targets for peaks.
22
23 910 For stage-specific peaks, ChIPseeker [92] was used for putative target assignment. In
24
25 911 both strategies, the putative target of a certain peak is defined as the gene with TSS
26
27 912 closest to the peak summit location.
28

29 913

31 914 **GO term and KEGG enrichment analysis**

33 915 Lists of genes were analysed using DAVID [93,94] and the BH method was used for
34
35 916 multiple test correction. GO terms with a FDR less than 0.01 or 0.05 were considered
36
37 917 as significantly enriched. Target genes of stage-specific ATAC peaks were analysed
38
39 918 using the R package, clusterProfiler [95], in which an adjusted p-value of 0.05 was
40
41 919 used to identify significantly enriched GO and KEGG terms associated with each set
42
43 920 of peaks.
44
45

46 921

48 922 **Regulatory network construction**

50 923 The scRNA-seq profiles among each cell types were compared using SCDE package
51
52 924 [84]. TFs significantly differentially expressed, with adjusted p-value threshold of 0.05,
53
54 925 among neighbouring cell types were submitted to STRING database [64] to infer
55
56 926 regulatory networks based on known interaction relationships (supported by data from
57
58 927 curated databases, experiments and text-mining). TFs without any interactions with
59
60
61
62
63
64
65

1 928 other proteins were removed from the network. To select key regulators, we used a
2
3 929 threshold of 5 and all TFs with number of interactions above the threshold were
4
5 930 considered as key regulators.

6
7 931

9 932 **Putative targets prediction, GO term and KEGG enrichment analysis**

10
11 933 The target prediction and enrichment analyses were performed using the FIMO [96]
12
13 934 and GREAT [97] packages, respectively. Briefly, the peak files in a certain stage were
14
15 935 scanned for the presence or absence of TF motifs, which were downloaded from the
16
17 936 Jasper database [98]. Genes with a TSS closest to TF motif-containing peaks were
18
19 937 considered as putative targets of certain TFs.

20
21 938

22 23 939 **Construction of cellular communication network**

24
25 940 The ligand-receptor interaction relationships were downloaded from the database,
26
27 941 IUPHAR/BPS Guide to PHARMACOLOGY [99], and the Database of
28
29 942 Ligand-Receptor Partners (DLRP) [65, 100]. The average expression level of TPM of
30
31 943 1 was used as a threshold. Ligands and receptors above the threshold were
32
33 944 considered as expressed in the corresponding cluster. Adjusted *P* value of 0.05 was
34
35 945 used as a threshold to identify ligands/receptors specifically expressed in a
36
37 946 subpopulation. The R package Circlize [101] was used to visualize the interactions.

38
39 947

40 41 948 **Motif enrichment analysis**

42
43 949 Motifs enriched in each set of ATAC peaks were identified using findMotifsGenome.pl
44
45 950 from HOMER [91] using following parameters: -size -100,100 -len
46
47 951 4,5,6,7,8,9,10,11,12.

48
49 952

50 51 953 **Additional files**

52
53 954 **Additional file 1: Figure S1.** Quality control of ATAC-seq.

54
55 955 **Additional file 2: Figure S2.** Dynamics of gained and lost peaks during neural
56
57 956 differentiation.

1 957 **Additional file 3: Figure S3.** Stage-specific features of *cis*-regulatory elements
2 during neural differentiation.
3
4 959 **Additional file 4: Figure S4.** Quality control of scRNA-seq.
5
6 960 **Additional file 5: Figure S5.** Subgroups identification and key transcriptomic features
7 within Fib stage.
8
9 961
10
11 962 **Additional file 6: Figure S6.** Subgroups identification and key transcriptomic features
12 within EB stage.
13
14 963
15 964 **Additional file 7: Figure S7.** Subgroups identification and key transcriptomic features
16 within Ros-L stage.
17
18 965
19 966 **Additional file 8: Figure S8.** Subgroups identification and key transcriptomic features
20 within NPCs stage.
21
22 967
23 968 **Additional file 9: Figure S9.** Expression pattern of selected transcription factors (TFs)
24 within rosettes (Ros-E and Ros-L) stage.
25
26 969
27 970 **Additional file 10: Figure S10.** Differentially expressed transcription factors (TFs)
28 between Ros-E2 and Ros-L3.
29
30 971
31 972 **Additional file 11: Figure S11.** Differentially expressed transcription factors (TFs)
32 between Ros-L3 and NPC1.
33
34 973
35 974 **Additional file 12: Figure S12.** Key regulators during neural differentiation.
36
37 975 **Additional file 13: Figure S13.** GO term and KEGG enrichment analysis of selected
38 transcription factors (TFs) targets.
39
40 976
41 977 **Additional file 14: Figure S14.** Putative signaling between expressed receptors and
42 their ligands in EB subsets.
43
44 978
45 979 **Additional file 15: Figure S15.** Putative signaling between expressed receptors and
46 their ligands in Ros-E subsets.
47
48 980
49 981 **Additional file 16: Figure S16.** Putative signaling between expressed receptors and
50 their ligands in NPC subsets.
51
52 982
53 983 **Additional file 17: Figure S17.** Transcription factor motifs enriched in stage specific
54 peaks.
55
56 984
57
58 985 **Additional file 18: Figure S18.** Validation of neural differentiation in different genetic
59
60
61
62
63
64
65

1 986 background cell lines.

2 987 **Additional file 19: Table S1.** Differentially expressed TFs among neighbouring cell
3
4 988 subsets.

5
6 989 **Additional file 20: Table S2.** Putative targets of selected regulators.

7
8 990 **Additional file 21: Table S3.** Ligand-receptor interaction networks among
9
10 991 subpopulations.

11
12 992 **Additional file 22: Table S4.** Differentially expressed receptors and ligands among
13
14 993 Ros-L subpopulations.

15
16 994

17 995 **Availability of supporting data and materials**

18
19 996 The detailed protocol of neural differentiation and bioinformatics pipeline are available
20
21 997 in the protocols.io repository [102-103]. The sequencing raw data are deposited in the
22
23 998 NCBI Sequence Read Archive (SRA) under the accession number SRP155759.
24
25 999 Further supporting data can be found in the GigaScience database, GigaDB [104].
26
27

28
29 1000

30 1001 **Competing interests**

31
32 1002 The authors declare that they have no competing interests.
33

34
35 1003

36 1004 **Authors' contributions**

37
38 1005 Z.S., Z.G. and X.X. conceived and designed the project. Z.S., D.C., Q.W., Shengpeng
39
40 1006 W. and Q.D. conducted the majority of experiments and data analysis. Shengpeng W.,
41
42 1007 L.W., X.D., Shiyou W. and J.Z. performed computational analyses and prepared
43
44 1008 figures. C.L. participated in validation experiments and assisted with figure
45
46 1009 preparation for revision. D.Z., X.C. and F.C. contributed to sample collection. H.Y.,
47
48 1010 X.X., Z.G. and Z.S. supervised the entire study. X.L. contributed to the design of the
49
50 1011 revision and jointly supervised the validation work. Z.S. wrote the manuscript with
51
52 1012 input from D.C., Q.W., L.L., J.L.F., Z.G. and X.X. D.C., L.L., J.L.F., S.Z., F.C., Z.G. and
53
54 1013 X.X. contributed to the discussion and revision of the manuscript. All authors read and
55
56 1014 approved the final manuscript.
57
58
59
60
61
62
63
64
65

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

1015

1016 **Acknowledgements**

1017 We thank Tao Tan for support with antibody, Shiping Liu for bioinformatics help, Guibo
1018 Li for technical help with preparation of single cell RNA-seq libraries and other
1019 members of Cell and Developmental Biology Lab for discussions and support. This
1020 work was supported by National Key R&D Program of China (2017YFA0104100); and
1021 Shenzhen Engineering Laboratory for Innovative Molecular Diagnostics [grant number
1022 DRC-SZ [2016] 884] funded by Development and Reform Commission of Shenzhen
1023 Municipality; and Shenzhen Key Laboratory of Neurogenomics (CXB201108250094A)
1024 funded by Science, Technology and Innovation Commission of Shenzhen Municipality.
1025 Dongsheng Chen is supported by China Postdoctoral Science Foundation (grant
1026 number 2017M622795).

1027

1028 **References**

1029 1. Harmacek L, Watkins-Chow DE, Chen J, Jones KL, Pavan WJ, Salbaum JM, et al.
1030 A unique missense allele of BAF155, a core BAF chromatin remodeling complex
1031 protein, causes neural tube closure defects in mice. *Dev Neurobiol.*
1032 2014;74:483-97.

1033 2. Foster WH, Langenbacher A, Gao C, Chen J, Wang Y. Nuclear phosphatase
1034 PPM1G in cellular survival and neural development. *Dev Dyn.* 2013;242:1101-9.

1035 3. Wilde JJ, Petersen JR, Niswander L. Genetic, Epigenetic, and Environmental
1036 Contributions to Neural Tube Closure. *Annu Rev Genet.* 2014;48:583-611.

1037 4. Wen Z, Nguyen HN, Guo Z, Lalli MA, Wang X, Su Y, et al. Synaptic dysregulation
1038 in a human iPS cell model of mental disorders. *Nature.* 2014;515:414-8.

1039 5. Tao Y, Zhang SC. Neural Subtype Specification from Human Pluripotent Stem
1040 Cells. *Cell Stem Cell.* 2016. p. 573-86.

1041 6. Streit A, Berliner AJ, Papanayotou C, Slrulnik A, Stern CD. Initiation of neural
1042 induction by FGF signalling before gastrulation. *Nature.* 2000;406:74-8.

1043 7. Muñoz-Sanjuán I, Brivanlou AH. Neural induction, the default model and

1 1044 embryonic stem cells. *Nat Rev Neurosci.* 2002;3:271-80.

2

3 1045 8. Zhang SC, Wernig M, Duncan ID, Brüstle O, Thomson JA. In vitro differentiation

4 of transplantable neural precursors from human embryonic stem cells. *Nat*

5 1046 *Biotechnol.* 2001;19:1129-33.

6

7 1047

8

9 1048 9. Xu X, Hou Y, Yin X, Bao L, Tang A, Song L, et al. Single-cell exome sequencing

10 reveals single-nucleotide mutation characteristics of a kidney tumor. *Cell.*

11 1049 2012;148:886-95.

12

13 1050

14

15 1051 10. Hou Y, Song L, Zhu P, Zhang B, Tao Y, Xu X, et al. Single-cell exome

16 sequencing and monoclonal evolution of a JAK2-negative myeloproliferative

17 1052 neoplasm. *Cell.* 2012;148:873-85.

18

19 1053

20

21 1054 11. Johnson MB, Wang PP, Atabay KD, Murphy EA, Doan RN, Hecht JL, et al.

22 Single-cell analysis reveals transcriptional heterogeneity of neural progenitors in

23 1055 human cortex. *Nat Neurosci.* 2015;18:637-46.

24

25 1056

26

27 1057 12. Villani AC, Satija R, Reynolds G, Sarkizova S, Shekhar K, Fletcher J, et al.

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

29 1058 monocytes, and progenitors. *Science.* 2017;356:eaah4573.

30

31 1059

32

33 1060 13. La Manno G, Gyllborg D, Codeluppi S, Nishimura K, Salto C, Zeisel A, et al.

34 Molecular Diversity of Midbrain Development in Mouse, Human, and Stem Cells.

35 1061 *Cell.* 2016;167:566-580.e19.

36

37 1062

38

39 1063 14. Buettner F, Natarajan KN, Casale FP, Proserpio V, Scialdone A, Theis FJ, et al.

40 Computational analysis of cell-to-cell heterogeneity in single-cell

41 1064 RNA-sequencing data reveals hidden subpopulations of cells. *Nat Biotechnol.*

42 2015;33:155-60.

43

44 1065

45

46 1066

47

48 1067 15. Li H, Courtois ET, Sengupta D, Tan Y, Chen KH, Goh JJL, et al. Reference

49 component analysis of single-cell transcriptomes elucidates cellular

50 1068 heterogeneity in human colorectal tumors. *Nat Genet.* 2017;49:708-18.

51

52 1069

53

54 1070 16. Kim DS, Lee DR, Kim HS, Yoo JE, Jung SJ, Lim BY, et al. Highly pure and

55 expandable PSA-NCAM-positive neural precursors from human ESC and

56 1071 iPSC-derived neural rosettes. *PLoS One.* 2012;7.

57

58 1072

59

60

61

62

63

64

65

1 1073 17. Ardhanareeswaran K, Mariani J, Coppola G, Abyzov A, Vaccarino FM. Human
2 1074 induced pluripotent stem cells for modelling neurodevelopmental disorders. *Nat.*
3
4 1075 *Rev. Neurol.* 2017. p. 265-78.

6 1076 18. Chambers SMSM, Fasano CACA, Papapetrou EP, Tomishima M, Sadelain M,
7
8 1077 Studer L. Highly efficient neural conversion of human ES and iPS cells by dual
9
10 1078 inhibition of SMAD signaling. *Nat Biotechnol.* 2009;27:275-80.

12 1079 19. Dolmetsch R, Geschwind DH. The human brain in a dish: The promise of
13
14 1080 iPSC-derived neurons. *Cell.* 2011. p. 831-4.

16 1081 20. Kriks S, Shim JW, Piao J, Ganat YM, Wakeman DR, Xie Z, et al. Dopamine
17
18 1082 neurons derived from human ES cells efficiently engraft in animal models of
19
20 1083 Parkinson's disease. *Nature.* 2011;480:547-51.

22 1084 21. Miller JD, Ganat YM, Kishinevsky S, Bowman RL, Liu B, Tu EY, et al. Human
23
24 1085 iPSC-based modeling of late-onset disease via progerin-induced aging. *Cell*
25
26 1086 *Stem Cell.* 2013;13:691-705.

28 1087 22. Maroof AM, Keros S, Tyson JA, Ying S-W, Ganat YM, Merkle FT, et al. Directed
29
30 1088 Differentiation and Functional Maturation of Cortical Interneurons from Human
31
32 1089 Embryonic Stem Cells. *Cell Stem Cell.* 2013;12:559-72.

34 1090 23. Shi Y, Kirwan P, Smith J, Robinson HPC, Livesey FJ. Human cerebral cortex
35
36 1091 development from pluripotent stem cells to functional excitatory synapses. *Nat*
37
38 1092 *Neurosci.* 2012;15:477-86.

40 1093 24. Hueber SD, Bezdan D, Henz SR, Blank M, Wu H, Lohmann I. Comparative
41
42 1094 analysis of Hox downstream genes in *Drosophila*. *Development*
43
44 1095 2007;134:381-92.

46 1096 25. Chen D, Jiang S, Ma X, Li F. TFBSbank: a platform to dissect the big data of
47
48 1097 protein-DNA interaction in human and model species. *Nucleic Acids Res. Oxford*
49
50 1098 *University Press;* 2017;45:D151-7.

52 1099 26. Thomas S, Li X-Y, Sabo PJ, Sandstrom R, Thurman RE, Canfield TK, et al.
53
54 1100 Dynamic reprogramming of chromatin accessibility during *Drosophila* embryo
55
56 1101 development. *Genome Biol. BioMed Central;* 2011;12:R43.

- 1 1102 27. Buenrostro JD, Wu B, Litzenburger UM, Ruff D, Gonzales ML, Snyder MP, et al.
2
3 1103 Single-cell chromatin accessibility reveals principles of regulatory variation.
4
5 1104 Nature. Nature Publishing Group; 2015;523:486-90.
- 6
7 1105 28. Jin F, Li Y, Dixon JR, Selvaraj S, Ye Z, Lee AY, et al. A high-resolution map of the
8
9 1106 three-dimensional chromatin interactome in human cells. Nature. Nature
10
11 1107 Publishing Group; 2013;503:290-4.
- 12
13 1108 29. Corces MR, Buenrostro JD, Wu B, Greenside PG, Chan SM, Koenig JL, et al.
14
15 1109 Lineage-specific and single-cell chromatin accessibility charts human
16
17 1110 hematopoiesis and leukemia evolution. Nat Genet. Nature Publishing Group;
18
19 1111 2016;48:1193-203.
- 20
21 1112 30. Picelli S, Björklund ÅK, Faridani OR, Sagasser S, Winberg G, Sandberg R.
22
23 1113 Smart-seq2 for sensitive full-length transcriptome profiling in single cells. Nat
24
25 1114 Methods. 2013;10:1096-100.
- 26
27 1115 31. Kimura-Yoshida C, Mochida K, Ellwanger K, Niehrs C, Matsuo I. Fate
28
29 1116 Specification of Neural Plate Border by Canonical Wnt Signaling and Grhl3 is
30
31 1117 Crucial for Neural Tube Closure. EBioMedicine. 2015;2:513-27.
- 32
33 1118 32. Nikolopoulou E, Galea GL, Rolo A, Greene NDE, Copp AJ. Neural tube closure:
34
35 1119 cellular, molecular and biomechanical mechanisms. Development.
36
37 1120 2017;144:552-66.
- 38
39 1121 33. Zhang J, Hagopian-Donaldson S, Serbedzija G, Elsemore J, Plehn-Dujowich D,
40
41 1122 McMahon AP, et al. Neural tube, skeletal and body wall defects in mice lacking
42
43 1123 transcription factor AP-2. Nature. 1996;381:238-41.
- 44
45 1124 34. Hayon Y, Dashevsky O, Shai E, Varon D, Leker RR. Platelet microparticles
46
47 1125 promote neural stem cell proliferation, survival and differentiation. J Mol Neurosci.
48
49 1126 2012;47:659-65.
- 50
51 1127 35. Lee HO, Levorse JM, Shin MK. The endothelin receptor-B is required for the
52
53 1128 migration of neural crest-derived melanocyte and enteric neuron precursors. Dev
54
55 1129 Biol. 2003;259:162-75.
- 56
57
58 1130 36. Li H, Horns F, Wu B, Xie Q, Li J, Li T, et al. Classifying Drosophila Olfactory
59
60
61
62
63
64
65

1 1131 Projection Neuron Subtypes by Single-Cell RNA Sequencing. *Cell*.
2 1132 2017;171:1206-1220.e22.
3
4 1133 37. Vierbuchen T, Ostermeier A, Pang ZP, Kokubu Y, Südhof TC, Wernig M. Direct
5 conversion of fibroblasts to functional neurons by defined factors. *Nature*. 2010;
6 1134 25:1035-1041.
7
8 1135
9
10 1136 38. Yoo AS, Sun AX, Li L, Shcheglovitov A, Portmann T, Li Y, et al.
11 MicroRNA-mediated conversion of human fibroblasts to neurons. *Nature*. 2011;
12 1137 476:228-31.
13
14 1138
15
16 1139 39. Kim D-Y, Hwang I, Muller FL, Paik J-H. Functional regulation of FoxO1 in neural
17 stem cell differentiation. *Cell Death Differ*. 2015;22:2034-45.
18
19 1140
20
21 1141 40. Cameron DA, Pennimpede T, Petkovich M. Tulp3 is a critical repressor of Mouse
22 hedgehog signaling. *Dev Dyn*. 2009;238:1140-9.
23
24 1142
25 1143 41. Jin Z, Liu L, Bian W, Chen Y, Xu G, Cheng L, et al. Different transcription factors
26 regulate nestin gene expression during P19 cell neural differentiation and central
27 nervous system development. *J Biol Chem*. 2009;284:8160-73.
28
29 1144
30
31 1145
32 1146 42. Elkabetz Y, Panagiotakos G, Al Shamy G, Socci ND, Tabar V, Studer L. Human
33 ES cell-derived neural rosettes reveal a functionally distinct early neural stem cell
34 stage. *Genes Dev*. 2008;22:152-65.
35
36 1147
37 1148
38 1149 43. Cheung M, Briscoe J. Neural crest development is regulated by the transcription
39 factor Sox9. *Development*. 2003;130:5681-93.
40
41 1150
42 1151 44. Scott CE, Wynn SL, Sesay A, Cruz C, Cheung M, Gavira MVG, et al. SOX9
43 induces and maintains neural stem cells. *Nat Neurosci*. 2010;13:1181-9.
44
45 1152
46 1153 45. Betters E, Liu Y, Kjaeldgaard A, Sundstrom E, Garcia-Castro MI. Analysis of early
47 human neural crest development. *Dev Biol*. 2010;344:578-92.
48
49 1154
50 1155 46. Wang C, Kam RKT, Shi W, Xia Y, Chen X, Cao Y, et al. The Proto-oncogene
51 transcription factor Ets1 regulates neural crest development through histone
52 deacetylase 1 to mediate output of bone morphogenetic protein signaling. *J Biol*
53
54 1157
55 1158
56 1158 Chem. 2015;290:21925-38.
57
58 1159 47. Mulligan KA, Cheyette BNR. Wnt signaling in vertebrate neural development and
59
60
61
62
63
64
65

1 1160 function. *J. Neuroimmune Pharmacol.* 2012. p. 774–87.

2
3 1161 48. Ille F, Sommer L. Wnt signaling: multiple functions in neural development. *Cell.*
4
5 1162 *Mol. Life Sci.* 2005;62:1100-8.

6
7 1163 49. Wang J, Jenjaroenpun P, Bhinge A, Angarica VE, Del Sol A, Nookaew I, et al.
8
9 1164 Single-cell gene expression analysis reveals regulators of distinct cell
10
11 1165 subpopulations among developing human neurons. *Genome Res.*
12
13 1166 2017;27:1783-94.

14
15 1167 50. Wang Y, Ristevski S, Harley VR. SOX13 exhibits a distinct spatial and temporal
16
17 1168 expression pattern during chondrogenesis, neurogenesis, and limb development.
18
19 1169 *J Histochem Cytochem.* 2006;54:1327-33.

20
21 1170 51. Ji EH, Kim J. SoxD Transcription Factors: Multifaceted Players of Neural
22
23 1171 Development. *Int J stem cells.* Korean Society for Stem Cell Research;
24
25 1172 2016;9:3-8.

26
27 1173 52. Koshida R, Oishi H, Hamada M, Takei Y, Takahashi S. MafB is required for
28
29 1174 development of the hindbrain choroid plexus. *Biochem Biophys Res Commun.*
30
31 1175 2017;483:288-93.

32
33 1176 53. Rogers CD, Phillips JL, Bronner ME. Elk3 is essential for the progression from
34
35 1177 progenitor to definitive neural crest cell. *Dev Biol.* 2013;374:255-63.

36
37 1178 54. Noisa P, Lund C, Kanduri K, Lund R, La H. Notch signaling regulates the
38
39 1179 differentiation of neural crest from human pluripotent stem cells. *J Cell Sci.*
40
41 1180 2014;127:2083-94.

42
43 1181 55. Zhang J, Zheng B, Zhou P-P, Zhang R-N, He M, Yang Z, et al. Vascular
44
45 1182 calcification is coupled with phenotypic conversion of vascular smooth muscle
46
47 1183 cells through Klf5-mediated transactivation of the Runx2 promoter. *Biosci Rep.*
48
49 1184 2014;34:e00148.

50
51 1185 56. Dougherty M, Kamel G, Grimaldi M, Gfrerer L, Shubinets V, Ethier R, et al.
52
53 1186 Distinct requirements for wnt9a and irf6 in extension and integration mechanisms
54
55 1187 during zebrafish palate morphogenesis. *Development.* 2013;140:76-81.

56
57
58 1188 57. Hernandez-Lagunas L, Choi IF, Kaji T, Simpson P, Hershey C, Zhou Y, et al.

- 1 1189 Zebrafish narrowminded disrupts the transcription factor prdm1 and is required
2 for neural crest and sensory neuron specification. *Dev Biol.* 2005;278:347-57.
3 1190
- 4 1191 58. Simões-Costa M, Bronner ME. Insights into neural crest development and
5 evolution from genomic analysis. *Genome Res.* 2013. p. 1069-80.
6 1192
- 7 1193 59. Viales RR, Diotel N, Ferg M, Armant O, Eich J, Alunni A, et al. The
8 Helix-Loop-Helix Protein Id1 Controls Stem Cell Proliferation During
9 Regenerative Neurogenesis in the Adult Zebrafish Telencephalon. *Stem Cells.*
10 1194 2015;33:892-903.
11 1195
- 12 1196
- 13 1197 60. Li L, Candelario KM, Thomas K, Wang R, Wright K, Messier A, et al. Hypoxia
14 Inducible Factor-1 (HIF-1) Is Required for Neural Stem Cell Maintenance and
15 Vascular Stability in the Adult Mouse SVZ. *J Neurosci.* 2014;34:16713-9.
16 1198
- 17 1199
- 18 1200 61. Ahmed M, Xu J, Xu P-X. EYA1 and SIX1 drive the neuronal developmental
19 program in cooperation with the SWI/SNF chromatin-remodeling complex and
20 SOX2 in the mammalian inner ear. *Development.* Company of Biologists;
21 2012;139:1965-77.
22 1201
- 23 1202
- 24 1203
- 25 1204 62. Abe H, Okazawa M, Nakanishi S. The Etv1/Er81 transcription factor orchestrates
26 activity-dependent gene regulation in the terminal maturation program of
27 cerebellar granule cells. *Proc Natl Acad Sci U S A. National Academy of Sciences;*
28 2011;108:12497-502.
29 1205
- 30 1206
- 31 1207
- 32 1208 63. Dominguez MH, Ayoub AE, Rakic P. POU-III transcription factors (Brn1, Brn2,
33 and Oct6) influence neurogenesis, molecular identity, and migratory destination
34 of upper-layer cells of the cerebral cortex. *Cereb Cortex.* 2013;23:2632-43.
35 1209
- 36 1210
- 37 1211 64. Szklarczyk D, Morris JH, Cook H, Kuhn M, Wyder S, Simonovic M, et al. The
38 STRING database in 2017: quality-controlled protein-protein association
39 networks, made broadly accessible. *Nucleic Acids Res. Oxford University Press;*
40 2017;45:D362-8.
41 1212
- 42 1213
- 43 1214
- 44 1215 65. Pavličev M, Wagner GP, Chavan AR, Owens K, Maziarz J, Dunn-Fletcher C, et al.
45 Single-cell transcriptomics of the human placenta: Inferring the cell
46 communication network of the maternal-fetal interface. *Genome Res.*
47 1216
- 48 1217

1 1218 2017;27:349-61.

2 1219 66. Stern CD. Initial patterning of the central nervous system: How many organizers?

3 1220 Nat Rev Neurosci. 2001;2:92-8.

4 1221 67. Smith JR, Vallier L, Lupo G, Alexander M, Harris WA, Pedersen RA. Inhibition of

5 1222 Activin/Nodal signaling promotes specification of human embryonic stem cells

6 1223 into neuroectoderm. Dev Biol. 2008;313:107-17.

7 1224 68. Schwarz M, Alvarez-Bolado G, Dressler G, Urbánek P, Busslinger M, Gruss P.

8 1225 Pax2/5 and Pax6 subdivide the early neural tube into three domains. Mech Dev.

9 1226 1999;82:29-39.

10 1227 69. Brouns MR, de Castro SCP, Terwindt-Rouwenhorst EA, Massa V, Hekking JW,

11 1228 Hirst CS, et al. Over-expression of Grhl2 causes spina bifida in the Axial defects

12 1229 mutant mouse. Hum Mol Genet. 2011;20:1536-46.

13 1230 70. Nikitina N, Tong L, Bronner ME. Ancestral network module regulating prdm1

14 1231 expression in the lamprey neural plate border. Dev Dyn. 2011;240:2265-71.

15 1232 71. Vincent SD, Dunn NR, Sciammas R, Shapiro-Shalef M, Davis MM, Calame K, et

16 1233 al. The zinc finger transcriptional repressor Blimp1/Prdm1 is dispensable for early

17 1234 axis formation but is required for specification of primordial germ cells in the

18 1235 mouse. Development. 2005;132:1315-25.

19 1236 72. Timmer JR, Wang C, Niswander L. BMP signaling patterns the dorsal and

20 1237 intermediate neural tube via regulation of homeobox and helix-loop-helix

21 1238 transcription factors. Development. 2002;129:2459-72.

22 1239 73. Burns CJ, Zhang J, Brown EC, Van Bibber AM, Van Es J, Clevers H, et al.

23 1240 Investigation of Frizzled-5 during embryonic neural development in mouse. Dev

24 1241 Dyn. 2008;237:1614-26.

25 1242 74. Miyake A, Nakayama Y, Konishi M, Itoh N. Fgf19 regulated by Hh signaling is

26 1243 required for zebrafish forebrain development. Dev Biol. 2005;288:259-75.

27 1244 75. Qu Q, Sun G, Murai K, Ye P, Li W, Asuelime G, et al. Wnt7a Regulates Multiple

28 1245 Steps of Neurogenesis. Mol Cell Biol. 2013;33:2551-9.

29 1246 76. Hsu Y-C, Lee D-C, Chen S-L, Liao W-C, Lin J-W, Chiu W-T, et al. Brain-specific

1 1247 1B promoter of FGF1 gene facilitates the isolation of neural stem/progenitor cells
2 1248 with self-renewal and multipotent capacities. *Dev Dyn.* 2009;238:302-14.

3 1249 77. Okita K, Matsumura Y, Sato Y, Okada A, Morizane A, Okamoto S, et al. A more
4 1250 efficient method to generate integration-free human iPS cells. *Nat. Methods.*
5 1251 2011;8:409-12.

6 1252 78. Wu L, Zhang X, Zhao Z, Wang L, Li B, Li G, et al. Full-length single-cell RNA-seq
7 1253 applied to a viral human cancer: Applications to HPV expression and splicing
8 1254 analysis in HeLa S3 cells. *Gigascience.* 2015;4: 51.

9 1255 79. Buenrostro JD, Giresi PG, Zaba LC, Chang HY, Greenleaf WJ. Transposition of
10 1256 native chromatin for fast and sensitive epigenomic profiling of open chromatin,
11 1257 DNA-binding proteins and nucleosome position. *Nat. Methods.* 2013;10:1213-8.

12 1258 80. Chen Y, Chen Y, Shi C, Huang Z, Zhang Y, Li S, et al. SOAPnuke: a MapReduce
13 1259 acceleration-supported software for integrated quality control and preprocessing
14 1260 of high-throughput sequencing data. *Gigascience* 2018;7:1-6.

15 1261 81. Kim D, Langmead B, Salzberg SL. HISAT: a fast spliced aligner with low memory
16 1262 requirements. *Nat. Methods*; 2015;12:357-60.

17 1263 82. Lawrence M, Huber W, Pagès H, Aboyoun P, Carlson M, Gentleman R, et al.
18 1264 Software for Computing and Annotating Genomic Ranges. Prlic A, editor. *PLoS*
19 1265 *Comput. Biol.* 2013;9:e1003118.

20 1266 83. Robinson MD, McCarthy DJ, Smyth GK. edgeR: a Bioconductor package for
21 1267 differential expression analysis of digital gene expression data. *Bioinformatics*
22 1268 2010;26:139-40.

23 1269 84. Kharchenko P V, Silberstein L, Scadden DT. Bayesian approach to single-cell
24 1270 differential expression analysis. *Nat. Methods* 2014;11:740-2.

25 1271 85. Trapnell C, Cacchiarelli D, Grimsby J, Pokharel P, Li S, Morse M, et al. The
26 1272 dynamics and regulators of cell fate decisions are revealed by pseudotemporal
27 1273 ordering of single cells. *Nat. Biotechnol.* 2014;32:381-6.

28 1274 86. Satija R, Farrell JA, Gennert D, Schier AF, Regev A. Spatial reconstruction of
29 1275 single-cell gene expression data. *Nat. Biotechnol.* 2015;33:495-502.

1 1276 87. Zhang H-M, Liu T, Liu C-J, Song S, Zhang X, Liu W, et al. AnimalTFDB 2.0: a
2 1277 resource for expression, prediction and functional study of animal transcription
3
4 1278 factors. *Nucleic Acids Res.* 2015;43:D76-81.
5
6 1279 88. Langmead B, Salzberg SL. Fast gapped-read alignment with Bowtie 2. *Nat.*
7
8 1280 *Methods.* 2012;9:357-9.
9
10 1281 89. Zhang Y, Liu T, Meyer CA, Eeckhoute J, Johnson DS, Bernstein BE, et al.
11
12 1282 Model-based Analysis of ChIP-Seq (MACS). *Genome Biol. BioMed Central*;
13
14 1283 2008;9:R137.
15
16 1284 90. Li Q, Brown JB, Huang H, Bickel PJ. Measuring reproducibility of high-throughput
17
18 1285 experiments. *Ann. Appl. Stat. Institute of Mathematical Statistics*;
19
20 1286 2011;5:1752-79.
21
22 1287 91. Heinz S, Benner C, Spann N, Bertolino E, Lin YC, Laslo P, et al. Simple
23
24 1288 Combinations of Lineage-Determining Transcription Factors Prime
25
26 1289 cis-Regulatory Elements Required for Macrophage and B Cell Identities. *Mol. Cell*
27
28 1290 2010;38:576-89.
29
30 1291 92. Yu G, Wang L-G, He Q-Y. ChIPseeker: an R/Bioconductor package for ChIP
31
32 1292 peak annotation, comparison and visualization. *Bioinformatics* 2015;31:2382-3.
33
34 1293 93. Huang DW, Sherman BT, Lempicki RA. Systematic and integrative analysis of
35
36 1294 large gene lists using DAVID bioinformatics resources. *Nat. Protoc.*
37
38 1295 2009;4:44-57.
39
40 1296 94. Huang DW, Sherman BT, Lempicki RA. Bioinformatics enrichment tools: paths
41
42 1297 toward the comprehensive functional analysis of large gene lists. *Nucleic Acids*
43
44 1298 *Res.* 2009;37:1-13.
45
46 1299 95. Yu G, Wang L-G, Han Y, He Q-Y. clusterProfiler: an R Package for Comparing
47
48 1300 Biological Themes Among Gene Clusters. *Omi. A J. Integr. Biol.* 2012;16:284-7.
49
50 1301 96. Grant CE, Bailey TL, Noble WS. FIMO: scanning for occurrences of a given motif.
51
52 1302 *Bioinformatics Oxford University Press*; 2011;27:1017-8.
53
54 1303 97. McLean CY, Bristor D, Hiller M, Clarke SL, Schaar BT, Lowe CB, et al. GREAT
55
56 1304 improves functional interpretation of cis-regulatory regions. *Nat. Biotechnol.*

1 1305 2010;28:495-501.

2 1306 98. Sandelin A, Alkema W, Engström P, Wasserman WW, Lenhard B. JASPAR: an

3 1307 open-access database for eukaryotic transcription factor binding profiles. *Nucleic*

4 1308 *Acids Res. Oxford University Press; 2004;32:D91-4.*

5 1309 99. Harding SD, Sharman JL, Faccenda E, Southan C, Pawson AJ, Ireland S, et al.

6 1310 The IUPHAR/BPS Guide to PHARMACOLOGY in 2018: updates and expansion

7 1311 to encompass the new guide to IMMUNOPHARMACOLOGY. *Nucleic Acids Res.*

8 1312 2018;46:D1091-106.

9 1313 100. Salwinski L, Miller CS, Smith AJ, Pettit FK, Bowie JU, Eisenberg D. The

10 1314 Database of Interacting Proteins: 2004 update. *Nucleic Acids Res.*

11 1315 2004;32:449D-451.

12 1316 101. Gu Z, Gu L, Eils R, Schlesner M, Brors B. circlize implements and enhances

13 1317 circular visualization in R. *Bioinformatics* 2014;30:2811-2.

14 1318 102. Shang Z, Chen D, Wang Q, Wang S. Neural progenitor cells derived from human

15 1319 induced pluripotent stem cells. *protocols.io.* 2018.

16 1320 <https://dx.doi.org/10.17504/protocols.io.ntrdem6>.

17 1321 103. Shang Z, Chen D, Wang Q, Wang S. Bioinformatic pipeline for studying

18 1322 transcriptome and regulome dynamics during neural differentiation. *protocols.io.*

19 1323 2018. <http://dx.doi.org/10.17504/protocols.io.ntpdem6>.

20 1324 104. Shang Z; Chen D; Wang Q; Wang S; Deng Q; Wu L; Liu C; Ding X; Wang S;

21 1325 Zhong J; Zhang D; Cai X; Zhu S; Yang H; Liu L; Fink JL; Chen F; Liu X; Gao Z;

22 1326 Xu X: Supporting data for "Single-cell RNA-seq reveals dynamic transcriptome

23 1327 profiling in human early neural differentiation" *GigaScience Database.* 2018.

24 1328 <http://dx.doi.org/10.5524/100496>

25 1329

26 1330

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

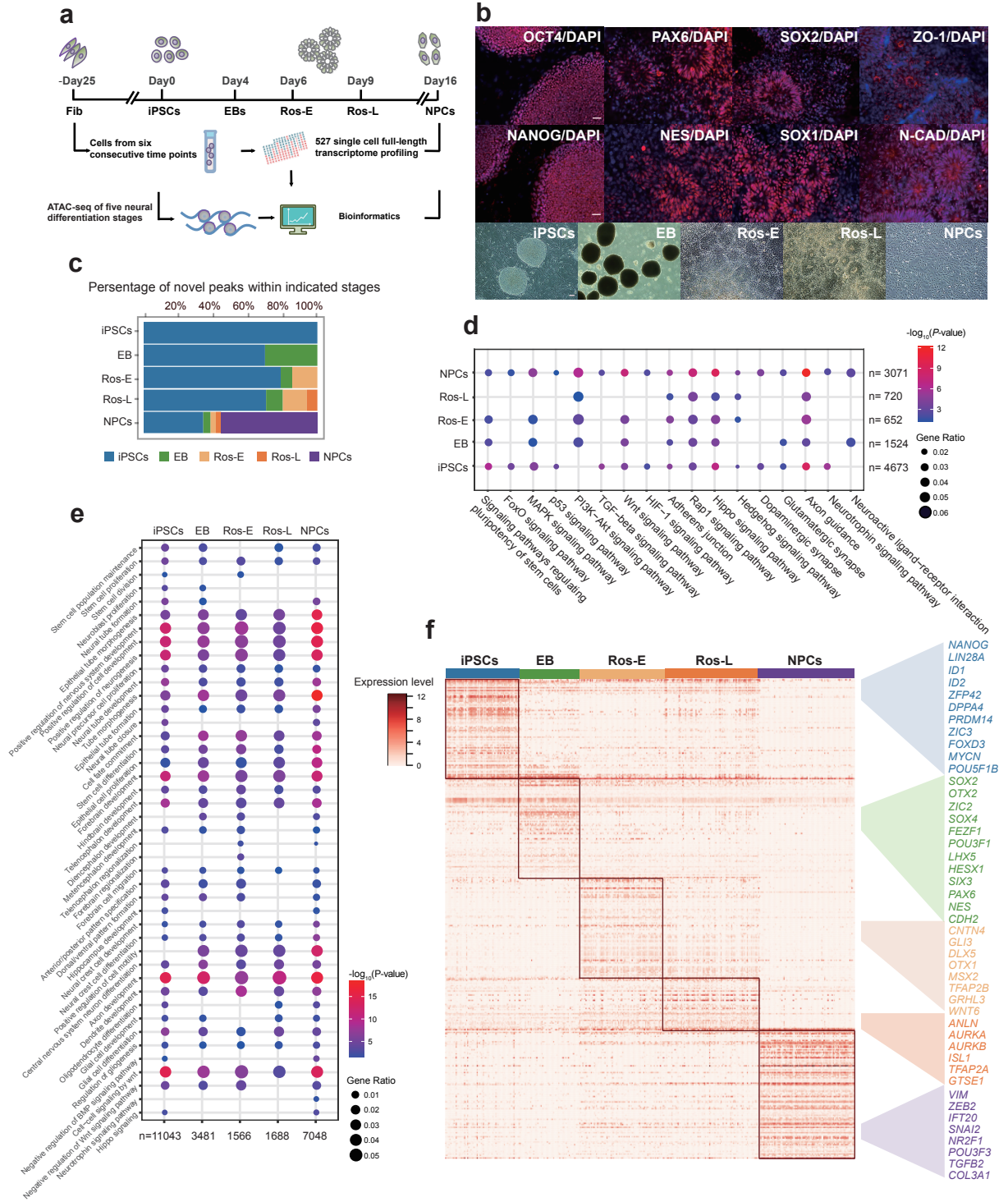


Fig. 1 Transcriptome and regulome dynamics during human early neural differentiation. **a** Schematic illustration of experimental strategy. **b** Bright field and immunostaining of well-defined markers for iPSCs including OCT4 and NANOG, and for neural rosettes (Ros-L stage) including PAX6, NES (NESTIN), SOX2, SOX1, ZO-1 and N-CAD (N-CADHERIN, also known as CDH2). Scale bar represents 50 μ m. **c** Dynamic distribution of novel peaks (active *cis*-regulatory elements) within indicated cell stages. **d** 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 ≤ 0.01).

Fig. 2

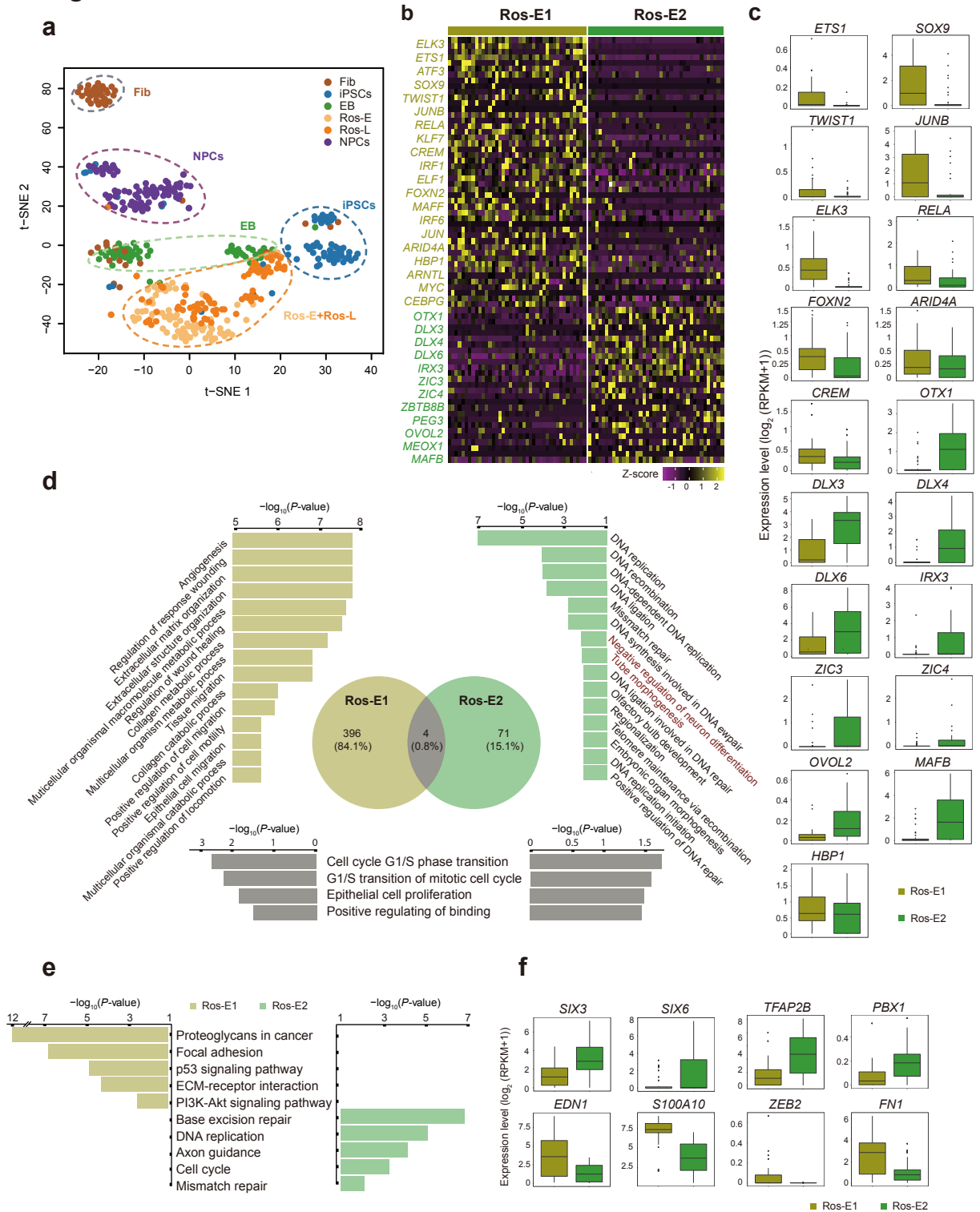


Fig. 2 Cell heterogeneity and identification of subsets within Ros-E stage. a T-SNE analysis of different cell stages as indicated with different color (n = 445). Number of successfully profiled single cells per cell stage: Fib (n = 54); iPSCs (n = 71); EB (n = 57); Ros-E (n = 81); Ros-L (n = 92); NPCs (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 ≤ 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 ≤ 0.01 .

Fig. 3

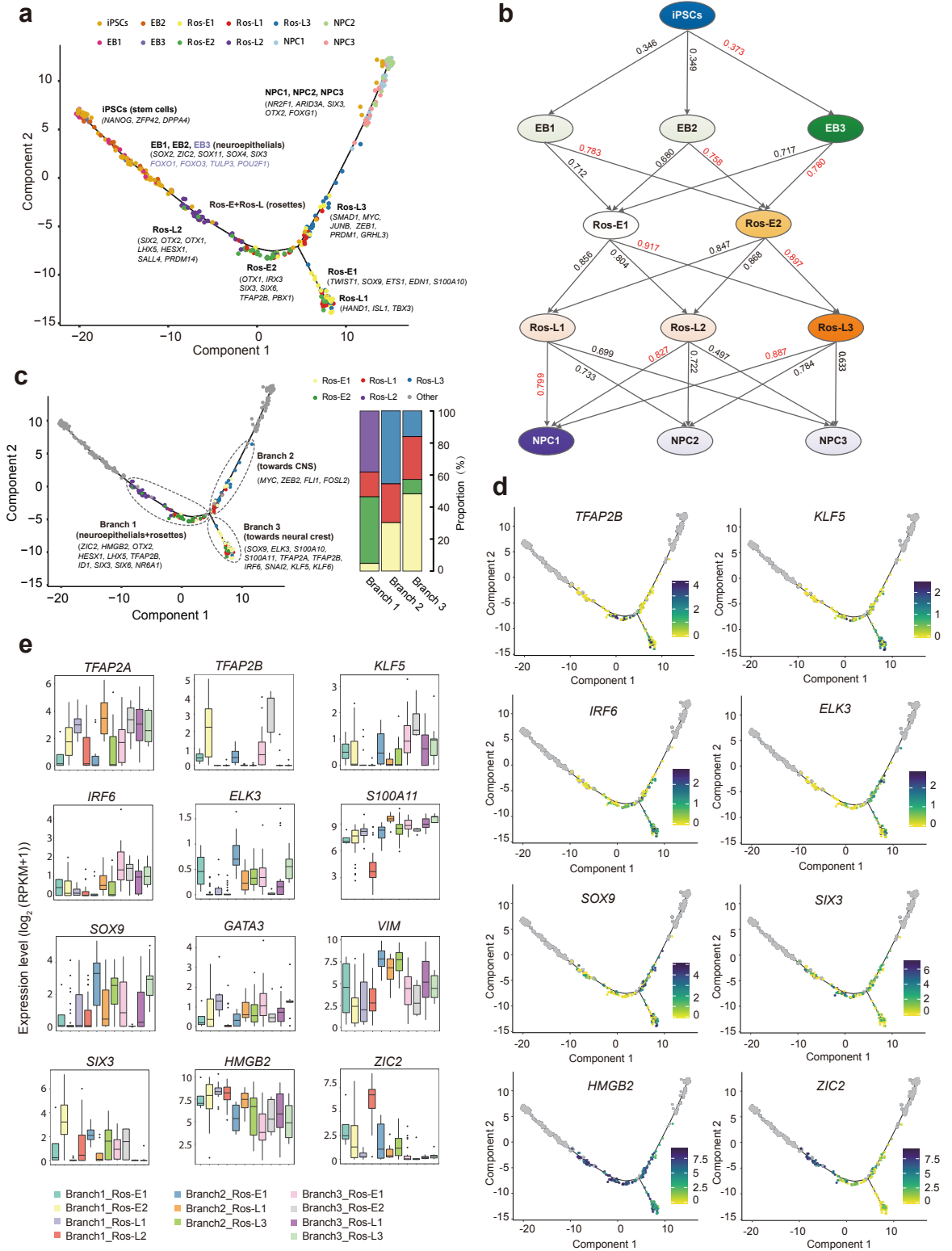


Fig. 3 Cell fate specification revealed by reconstructed trajectory.

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 ≤ 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

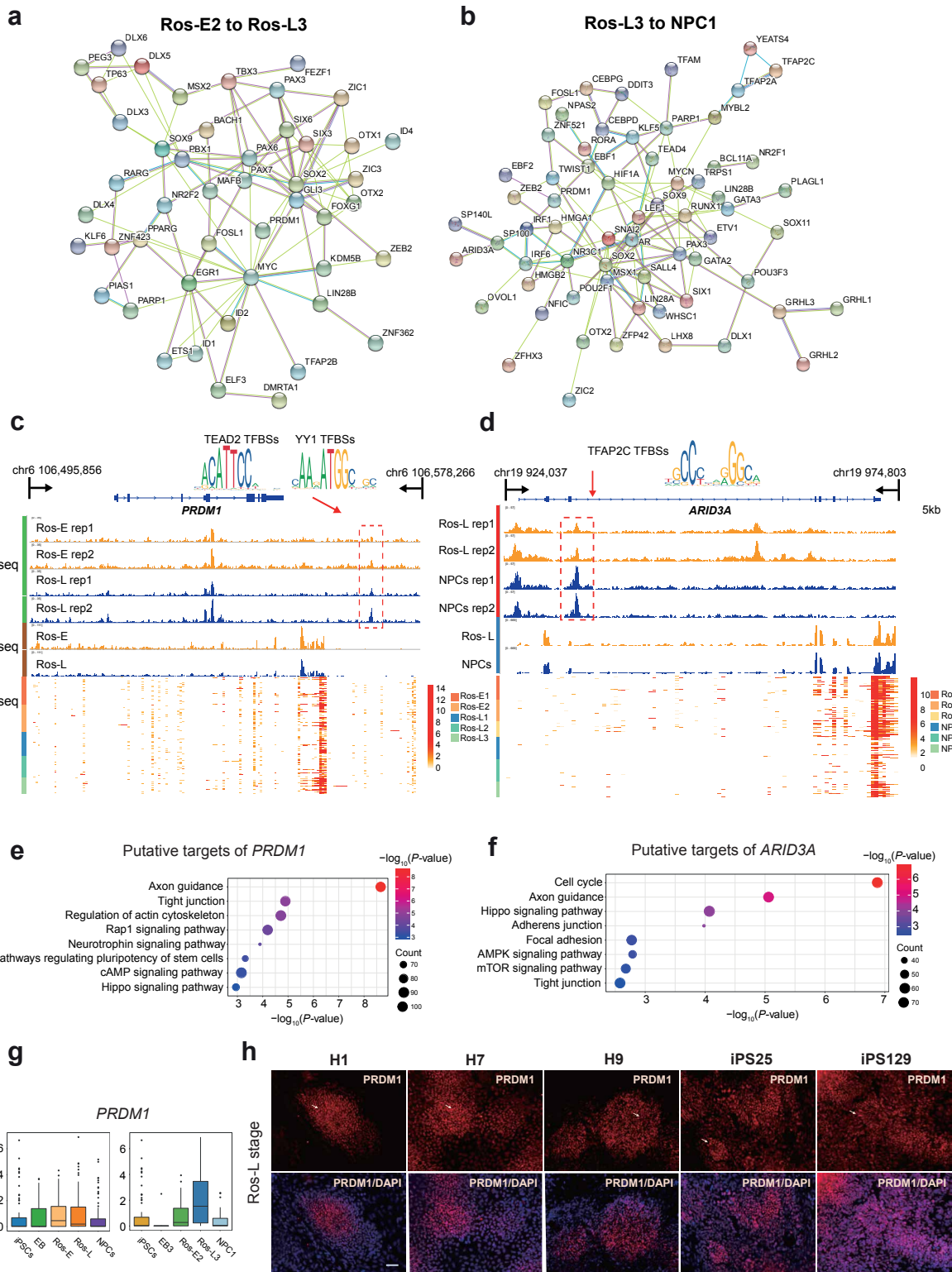
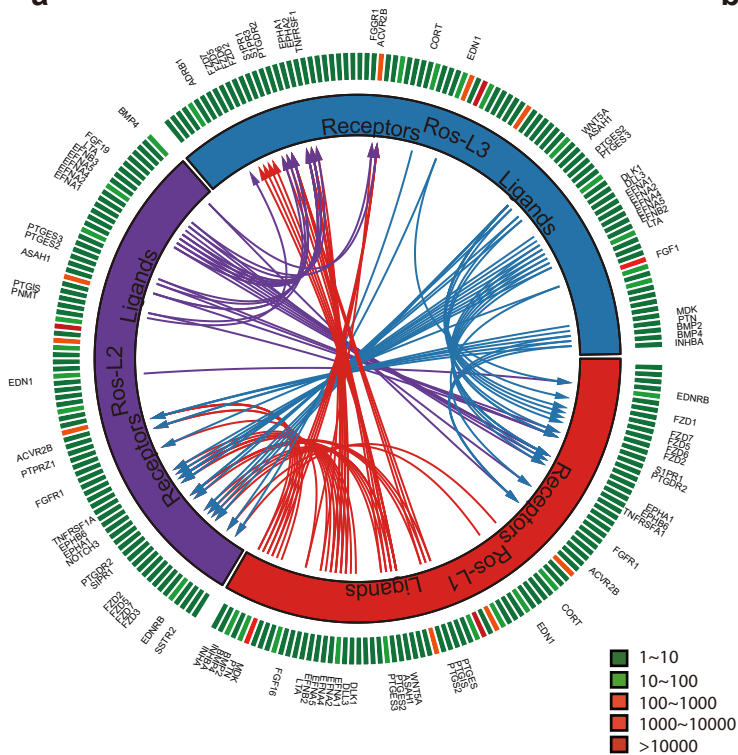


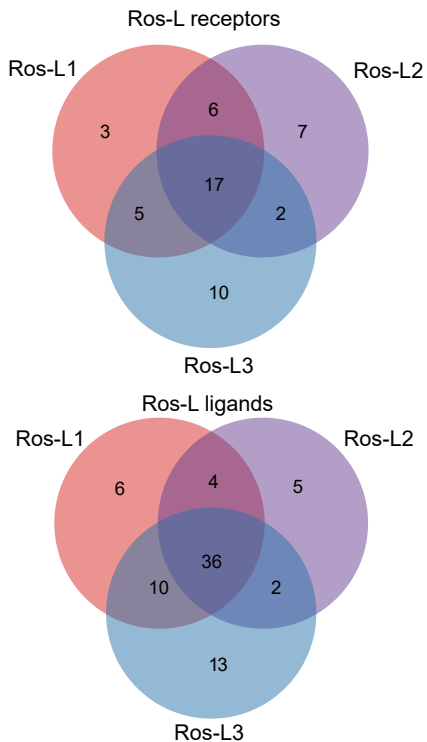
Fig. 4 Key regulators and corresponding *cis*-regulatory elements during neural differentiation. **a** Regulatory network of TFs differentially expressed between Ros-E2 and Ros-L3. **b** Regulatory network of differentially expressed TFs between Ros-L3 and NPC1. **c, d** IGV screenshots of ATAC-seq and 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 μm .

Fig. 5

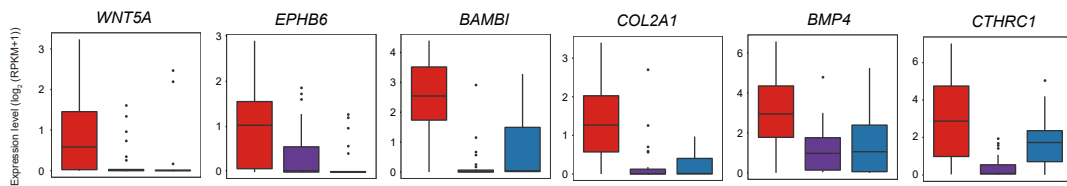
a



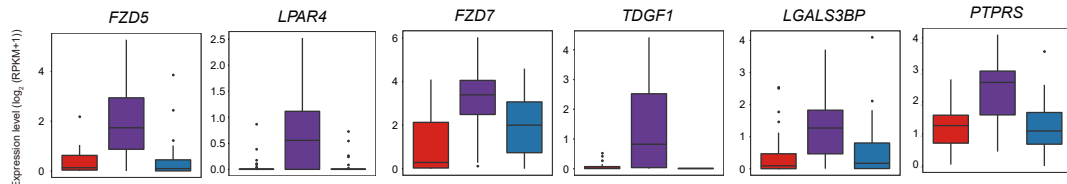
b



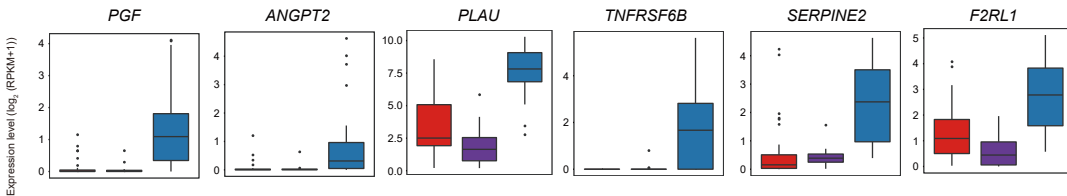
c



d



e

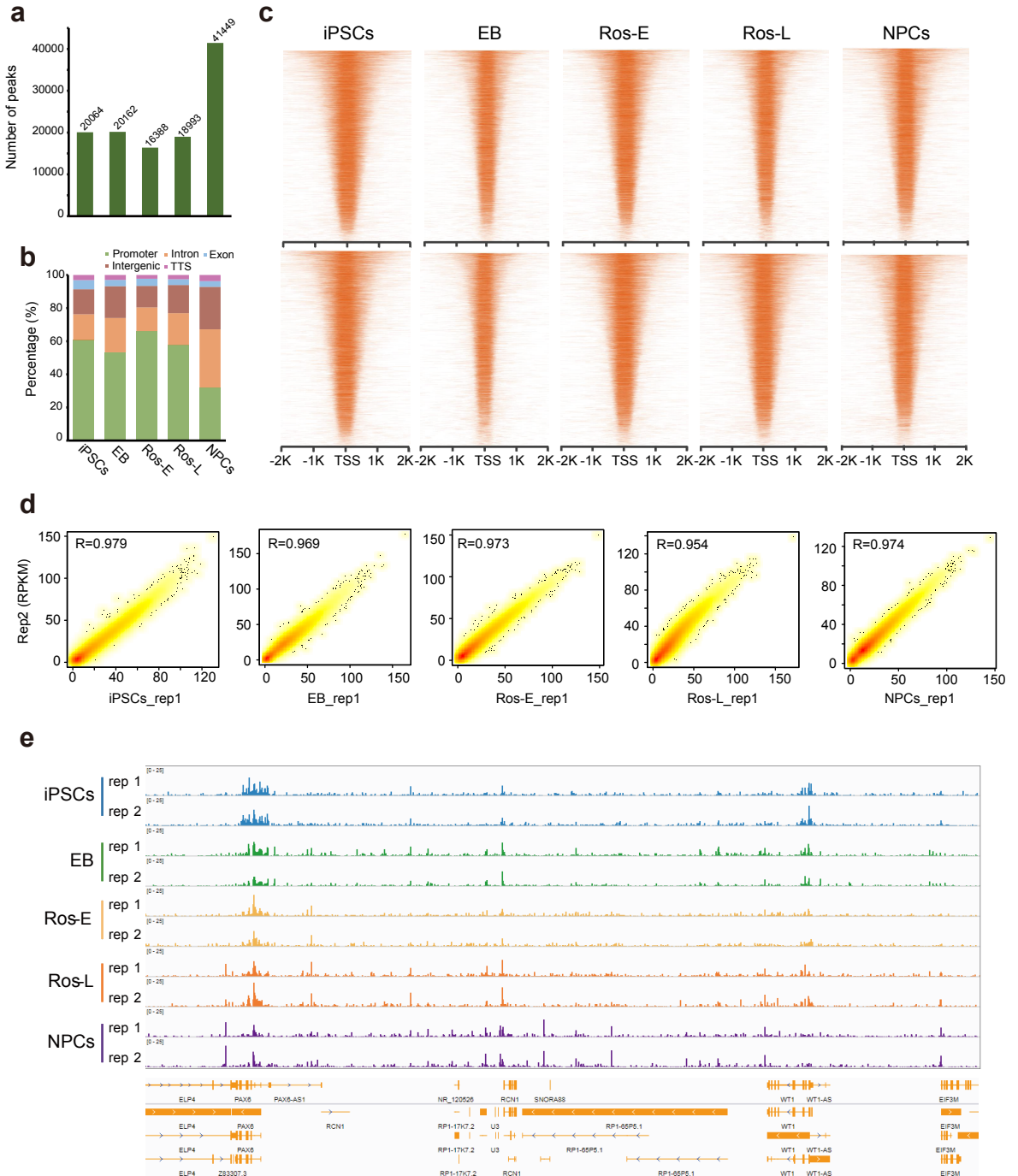


■ Ros-L1 ■ Ros-L2 ■ Ros-L3

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.

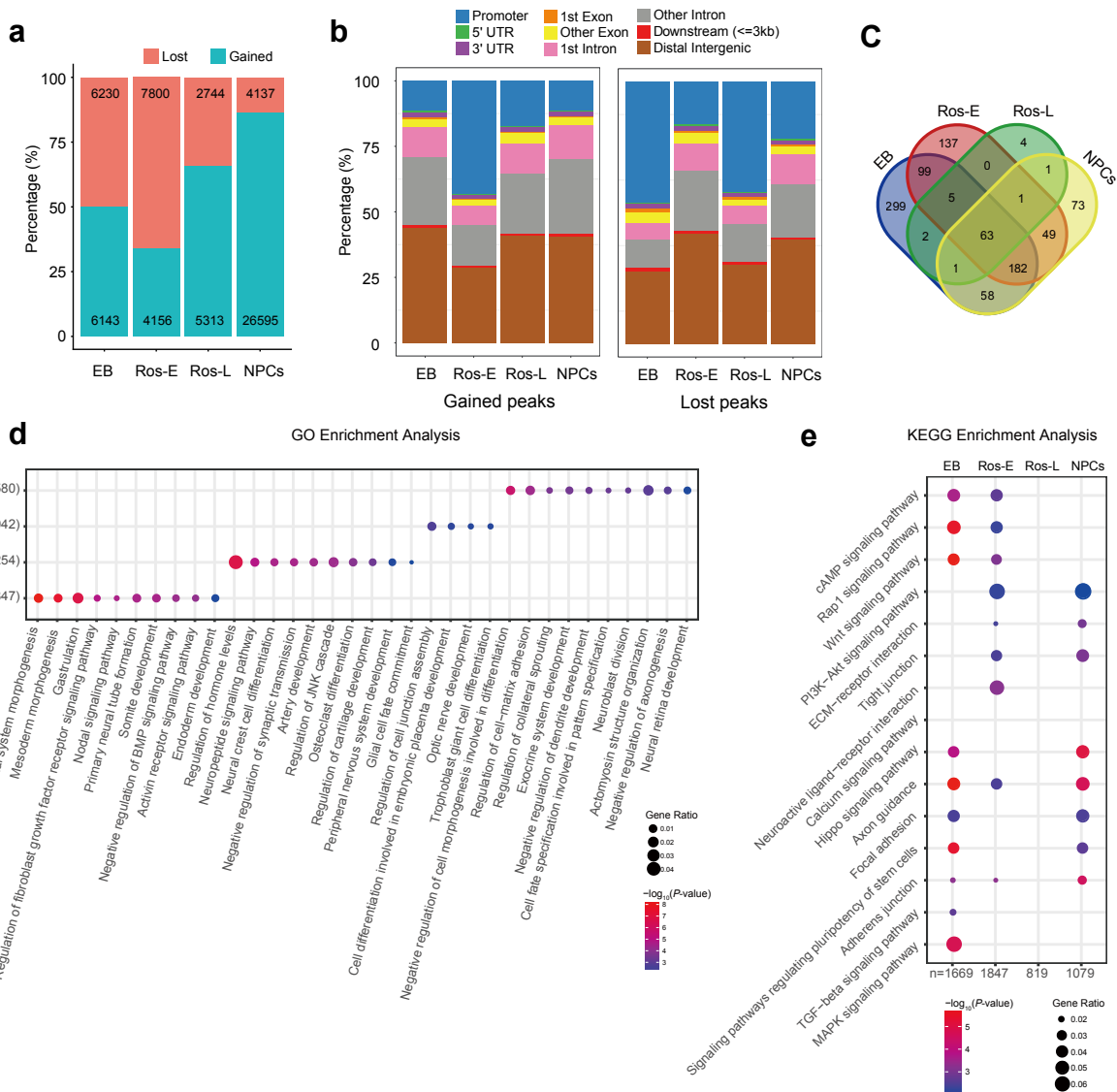


Additional file 1: Figure S1



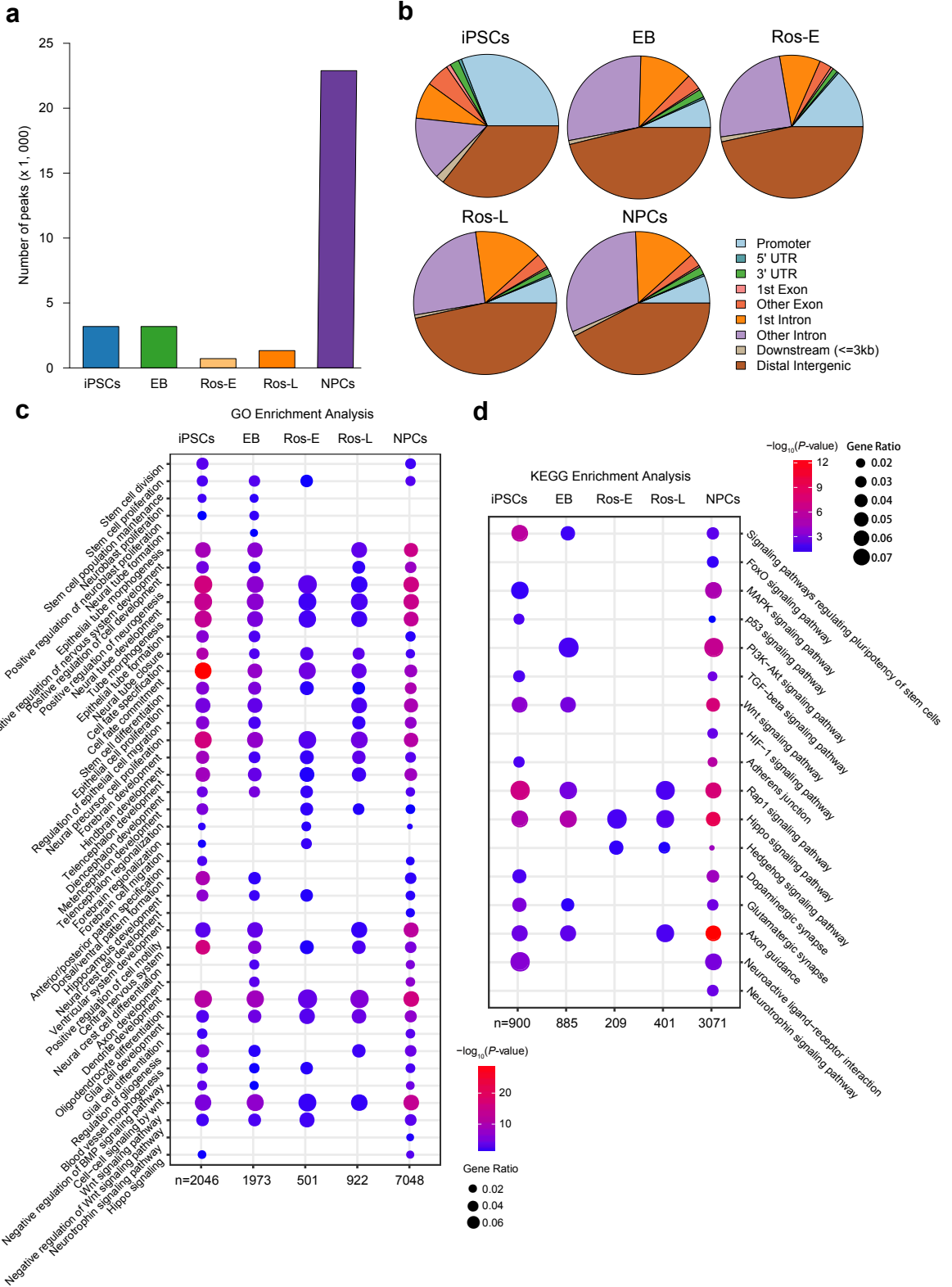
Additional file 1: Figure S1. Quality control of ATAC-seq. **a** Bar graphs indicate the number of chromatin open regions detected at each cell stage of neural differentiation. **b** Genomic components (distribution) of the peaks in each cell stage during neural differentiation. **c** Heatmaps reporting the chromatin accessibility density within ± 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



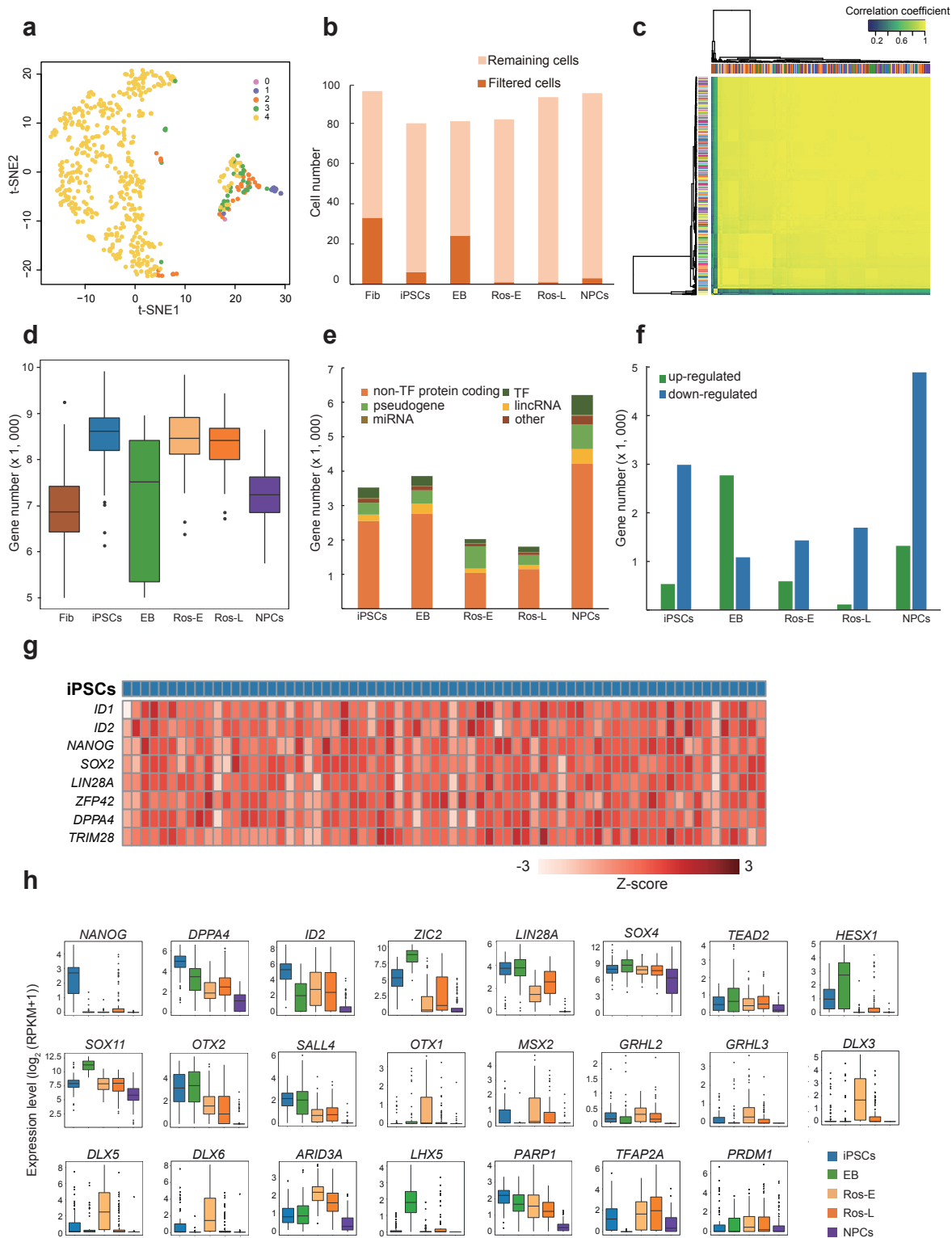
Additional file 2: Figure S2. Dynamics of gained and lost peaks during neural differentiation. **a** Bar graph shows the number of gained and lost peaks at each cell stage. **b** Bar graph shows genomic composition of gained and lost peaks at each cell stage respectively. **c** Venn plot of GO enrichment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

Additional file 3: Figure S3



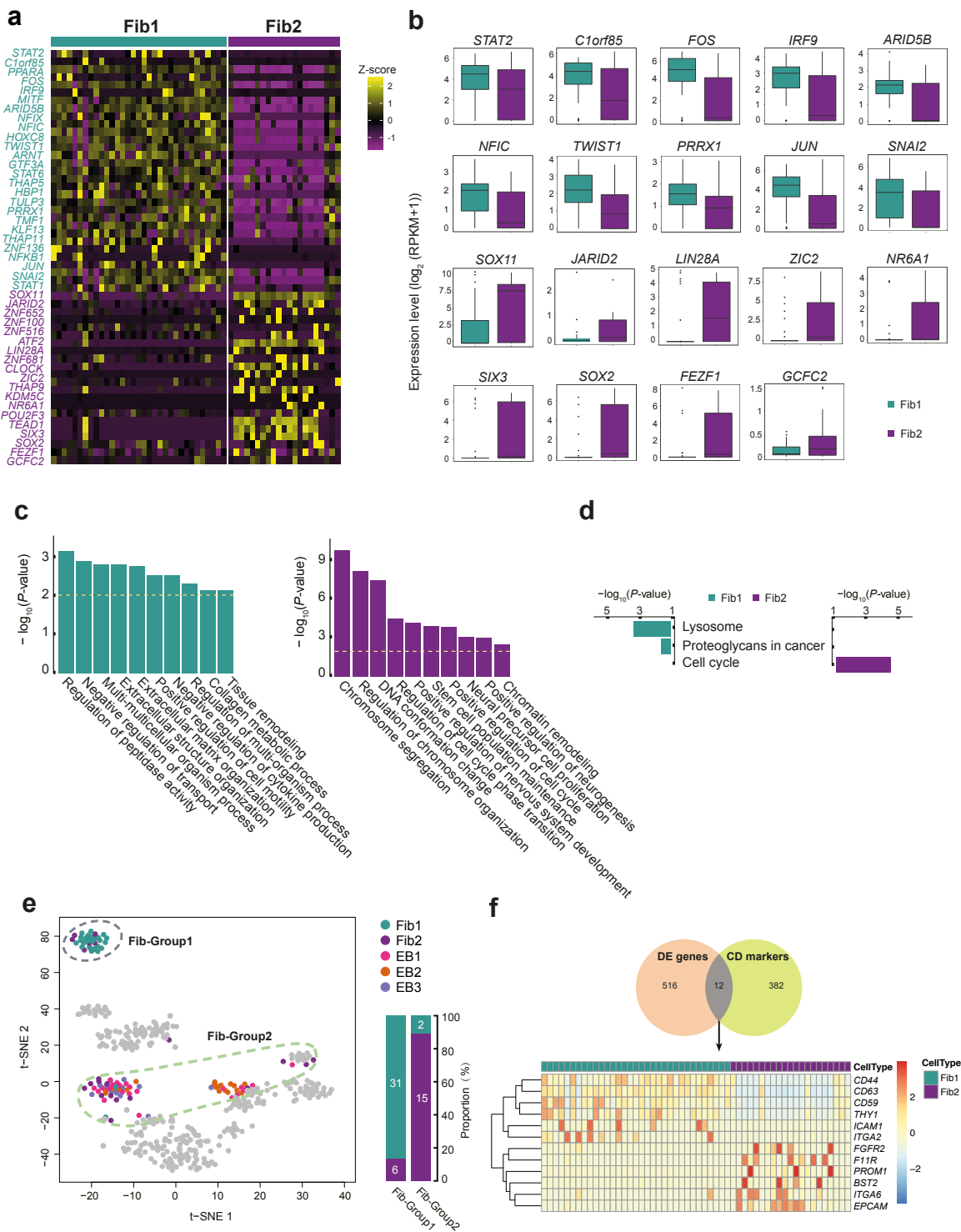
Additional file 3: Figure S3. Stage-specific features of *cis*-regulatory elements during neural differentiation. **a** Bar plot showing the number of stage specific ATAC peaks at iPSCs, EB, Ros-E, Ros-L and NPCs stage (adjusted P -value ≤ 0.01). **b** Pie chart shows genomic composition of stage specific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

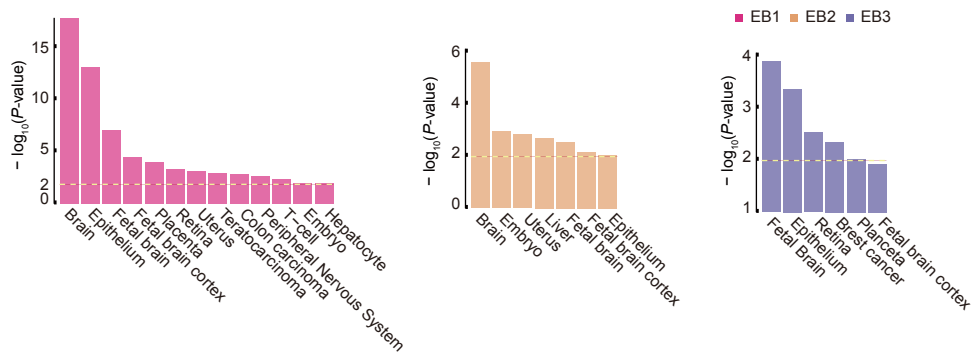
Additional file 5: Figure S5



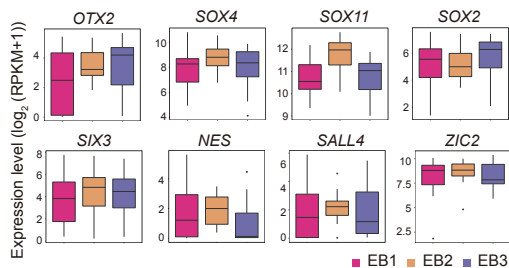
Additional file 5: Figure S5. Subgroups identification and key transcriptomic features within Fib stage. **a** Heatmap reports scaled expression [$\log_2(\text{RPKM}+1)$] of discriminative TF sets for each cluster in Fib stage with P -value cutoff ≤ 0.01 . Color scheme is based on z-score distribution 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

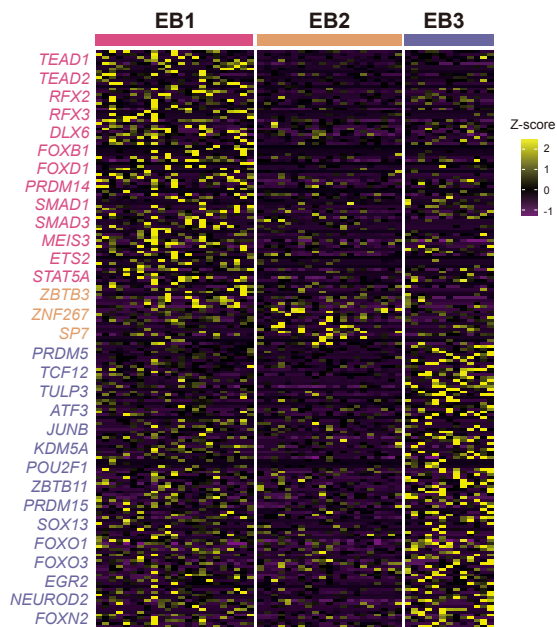
a



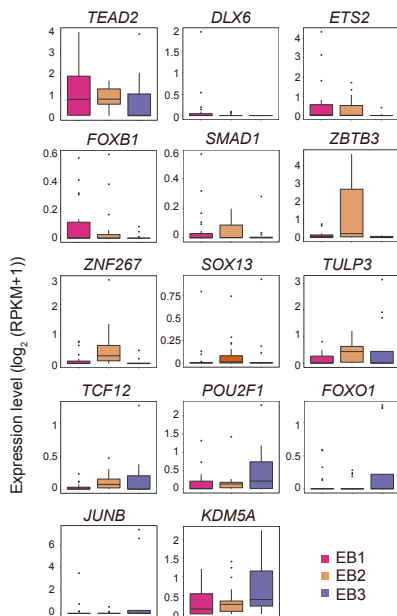
b



c

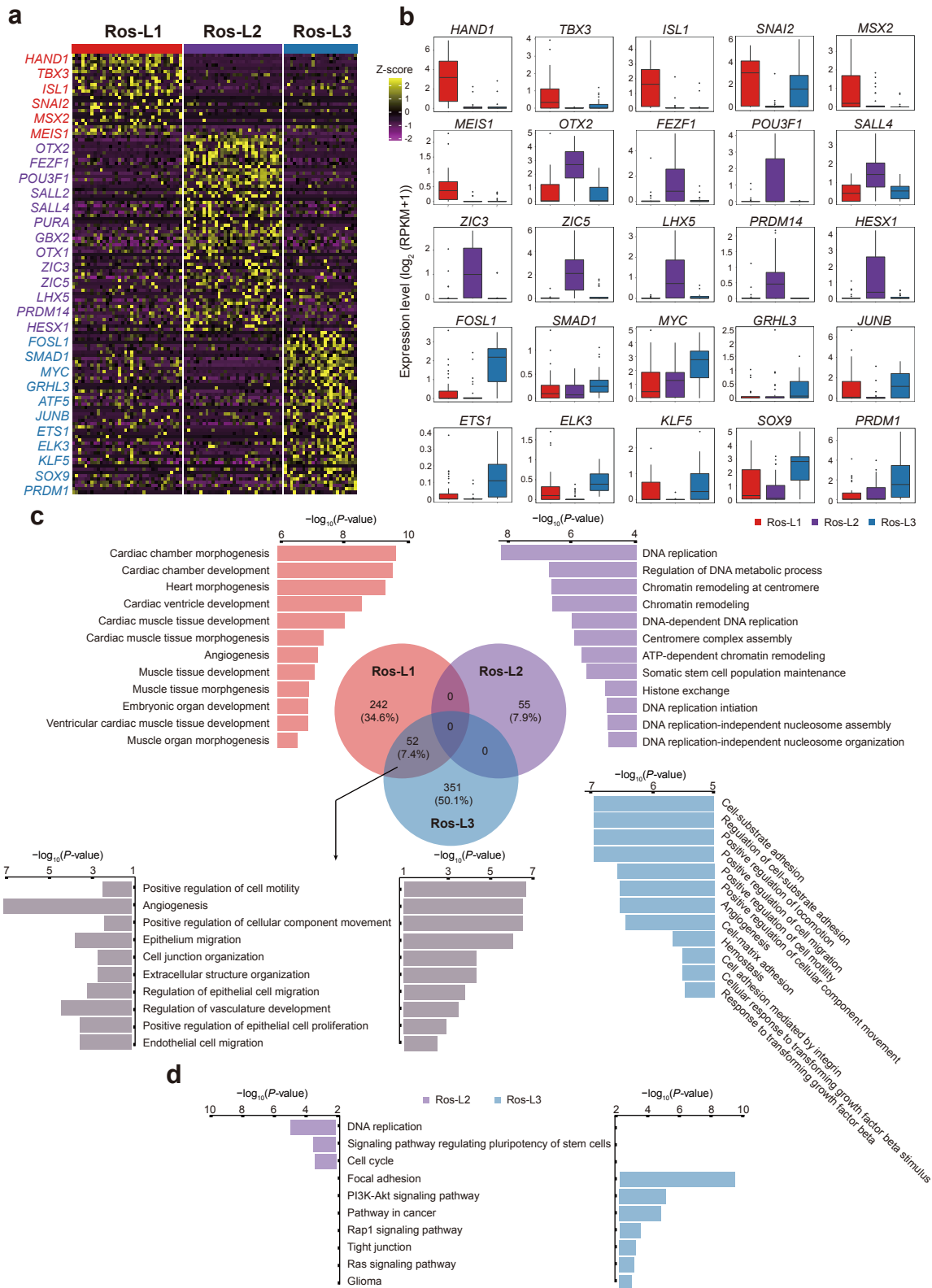


d



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 ≤ 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 S6c.

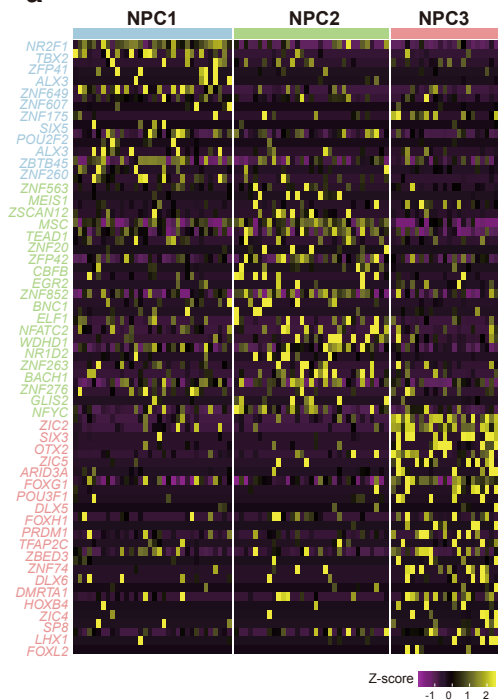
Additional file 7: Figure S7



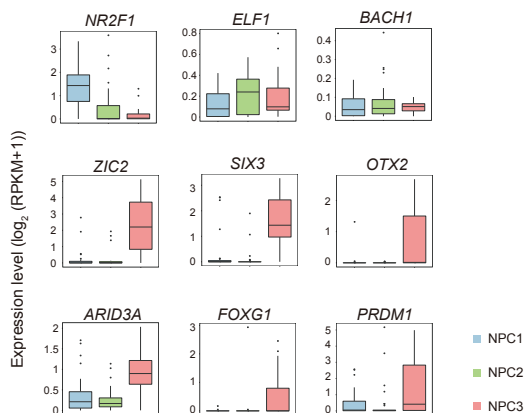
Additional file 7: Figure S7. Subgroups identification and key transcriptomic features within Ros-L stage. **a** Heatmap reports scaled expression [$\log_2(\text{RPKM}+1)$] of discriminative TF sets for each cluster in Ros-L stage with P -value cutoff ≤ 0.01 . Color scheme is based on z-score distribution 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

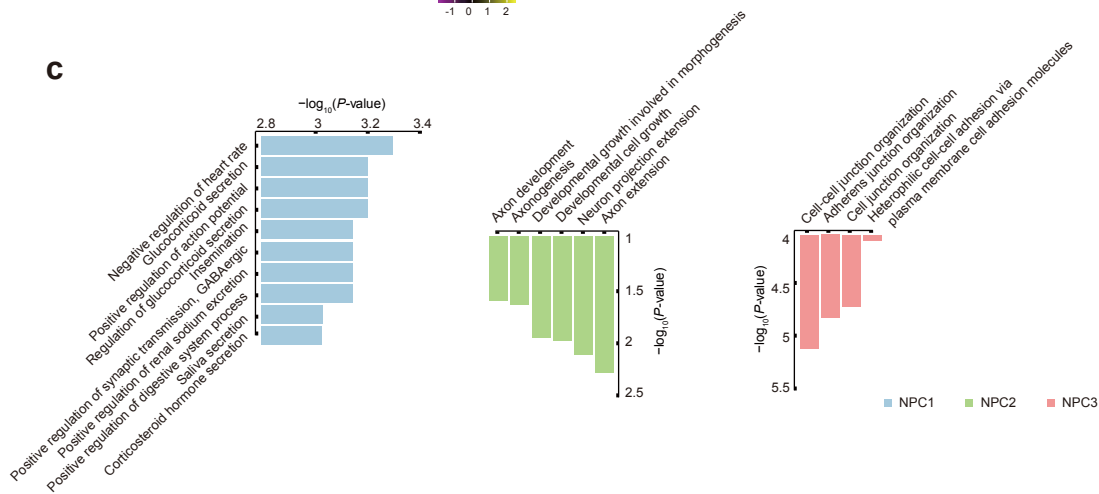
a



b

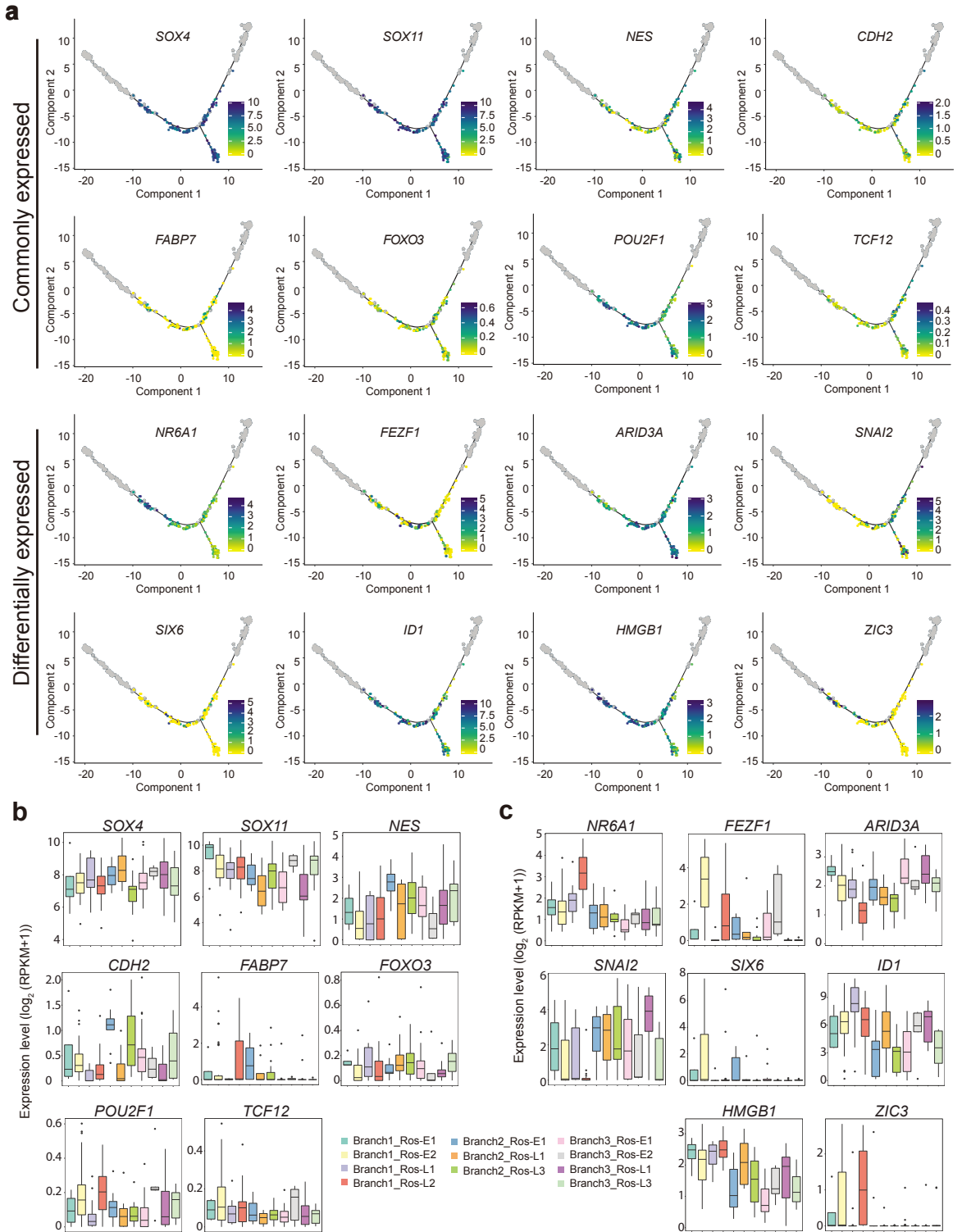


c



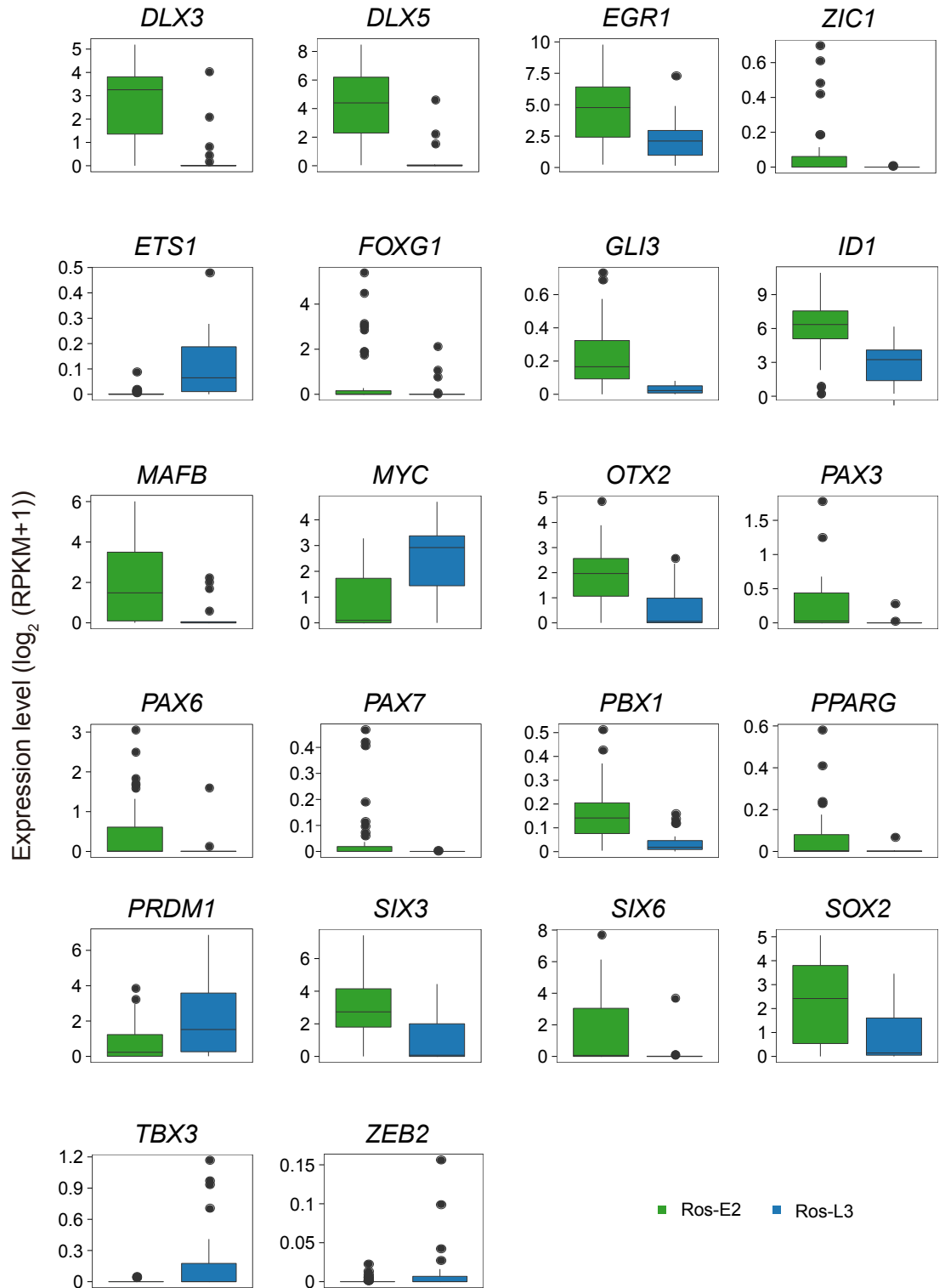
Additional file 8: Figure S8. Subgroups identification and key transcriptomic features within NPCs stage. **a** Heatmap reports scaled expression [$\log_2(\text{RPKM}+1)$] of discriminative TF sets for each cluster in NPCs stage with P -value cutoff ≤ 0.01 . Color scheme is based on z-score distribution 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



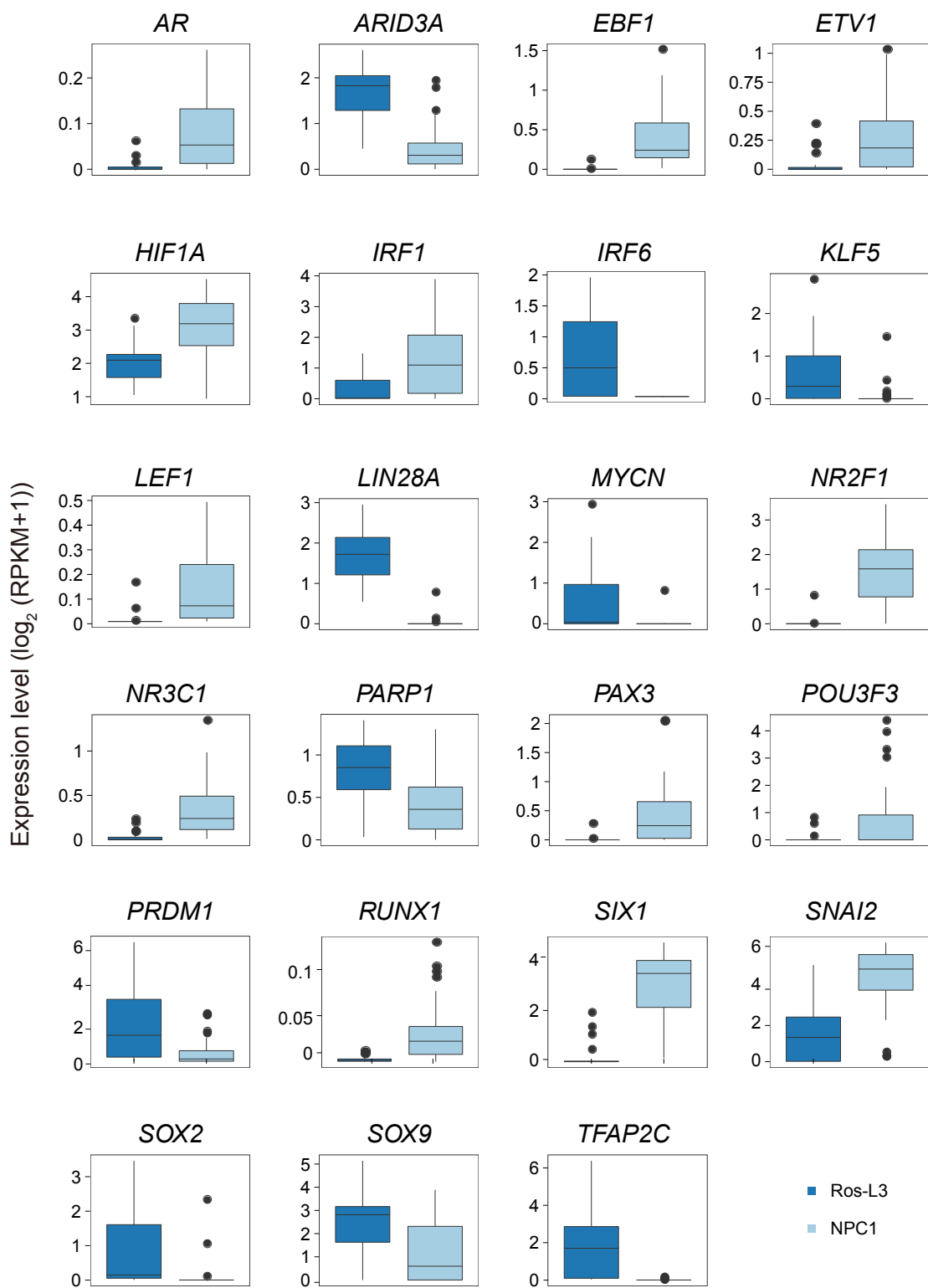
Additional file 9: Figure S9. Expression pattern of selected transcription factors (TFs) within rosettes (Ros-E and Ros-L) stage. a Expression enrichment of commonly and differentially expressed TFs along the differentiation trajectory. Color scheme is based on expression [\log_2 (RPKM+1)]. **b, c**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

Additional file 11: Figure S11

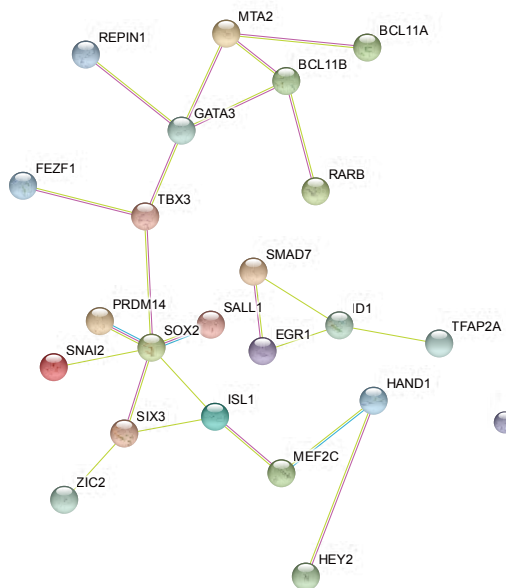


Additional file 11: Figure S11. Differentially expressed transcription factors (TFs) between Ros-L3 and NPC1. Ros-L3 and NPC1 were shown in dark blue and light blue column, respectively (adjusted P -value ≤ 0.01).

Additional file 12: Figure S12

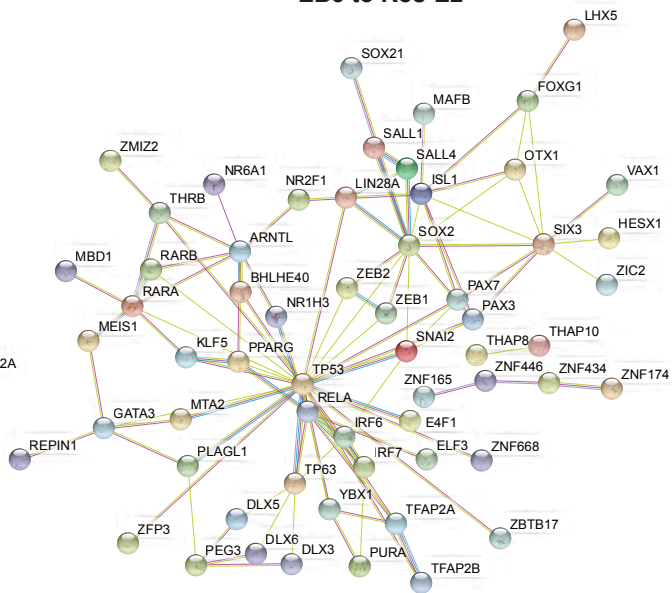
a

iPSCs to EB3



b

EB3 to Ros-E2

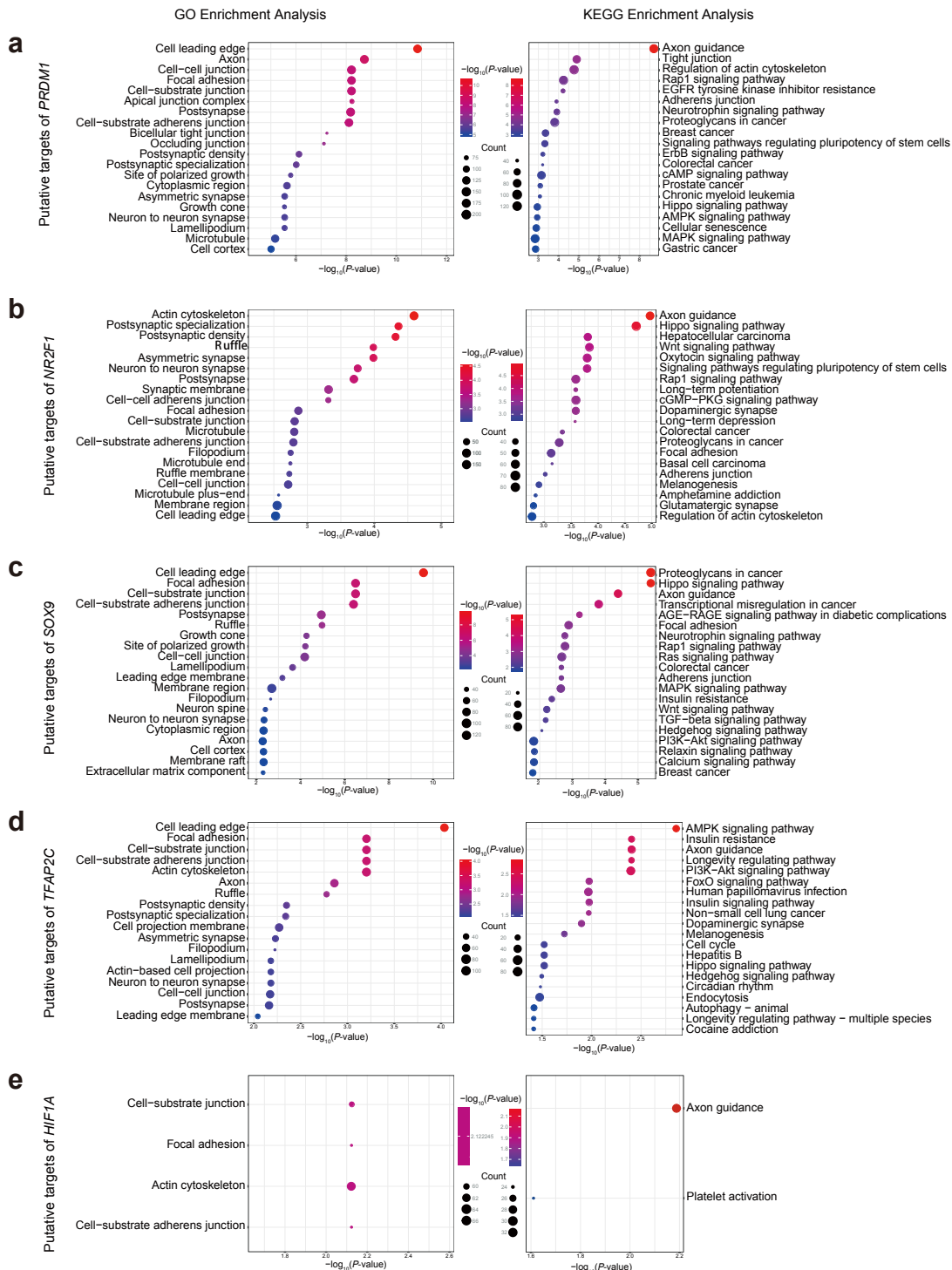


Additional file 12: Figure S12. Key regulators during neural differentiation. a

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

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

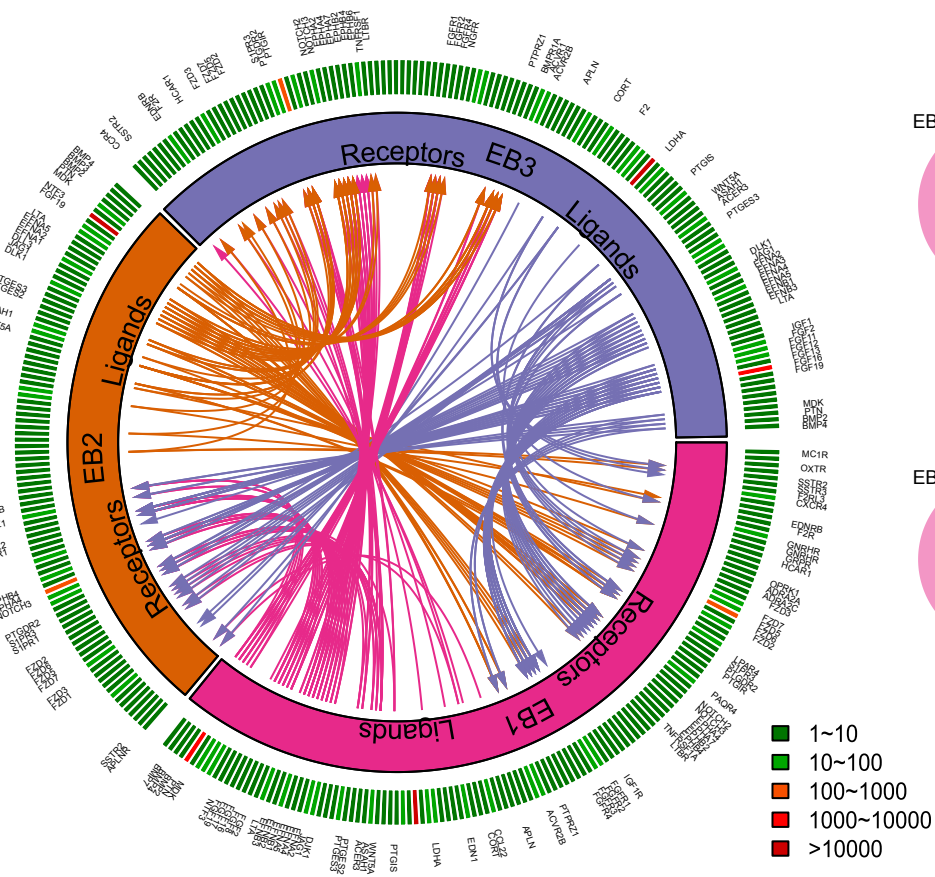
Additional file 13: Figure S13



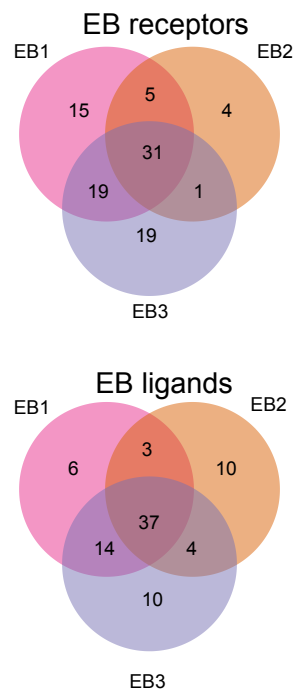
Additional file 13: Figure S13. GO term and KEGG enrichment analysis of selected transcription factors (TFs) targets. GO term and KEGG enrichment analysis for putative targets of *PRDM1* (a), *NR2F1* (b), *SOX9* (c), *TFAP2C* (d) and *HIF1A* (e), adjusted *P*-value ≤ 0.05 .

Additional file 14: Figure S14

a

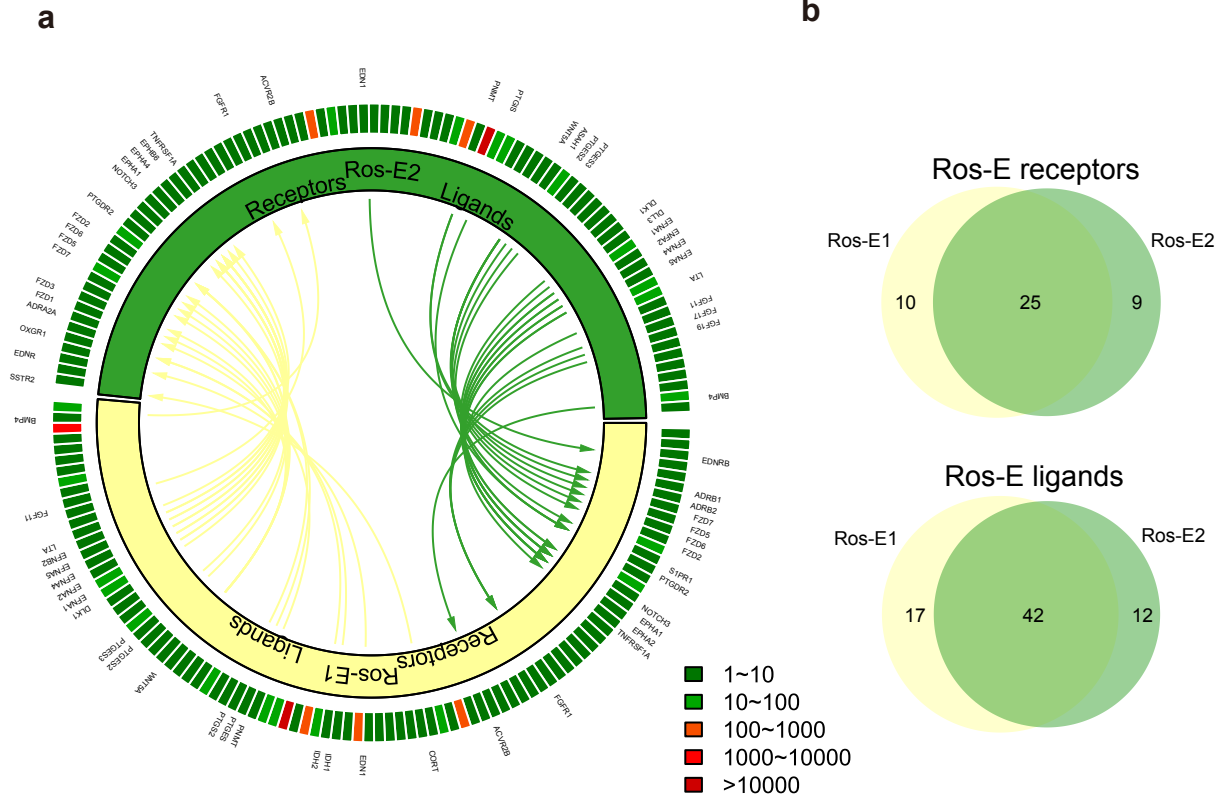


b



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 EB subpopulations.

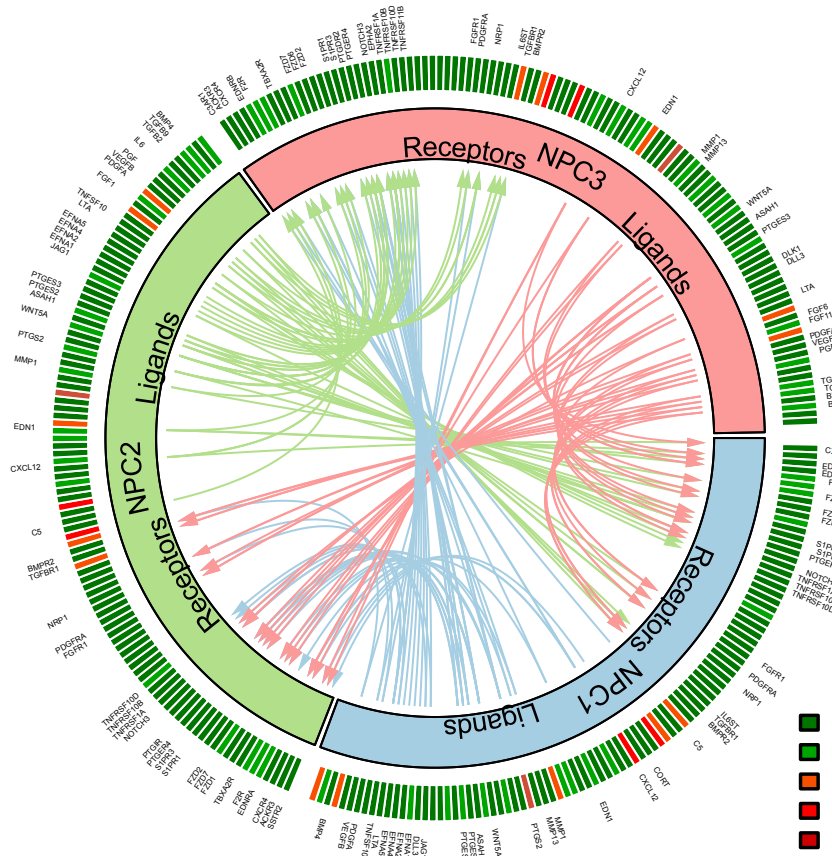
Additional file 15: Figure S15



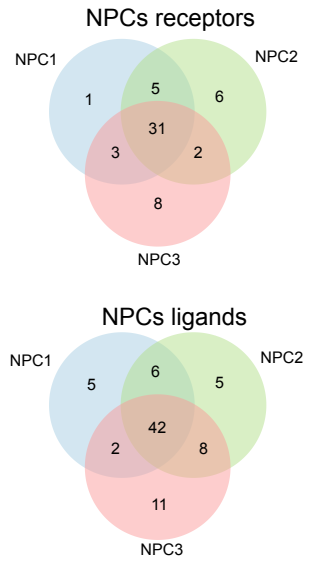
Additional file 15: Figure S15. Putative signaling between expressed receptors and their ligands in Ros-E subsets. **a** The inner layer compartments represent different cell subpopulations (Ros-E1 and Ros-E2). The outer layer indicates the expression profiles of ligands and receptors expressed in each 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 between Ros-E 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 NPC 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

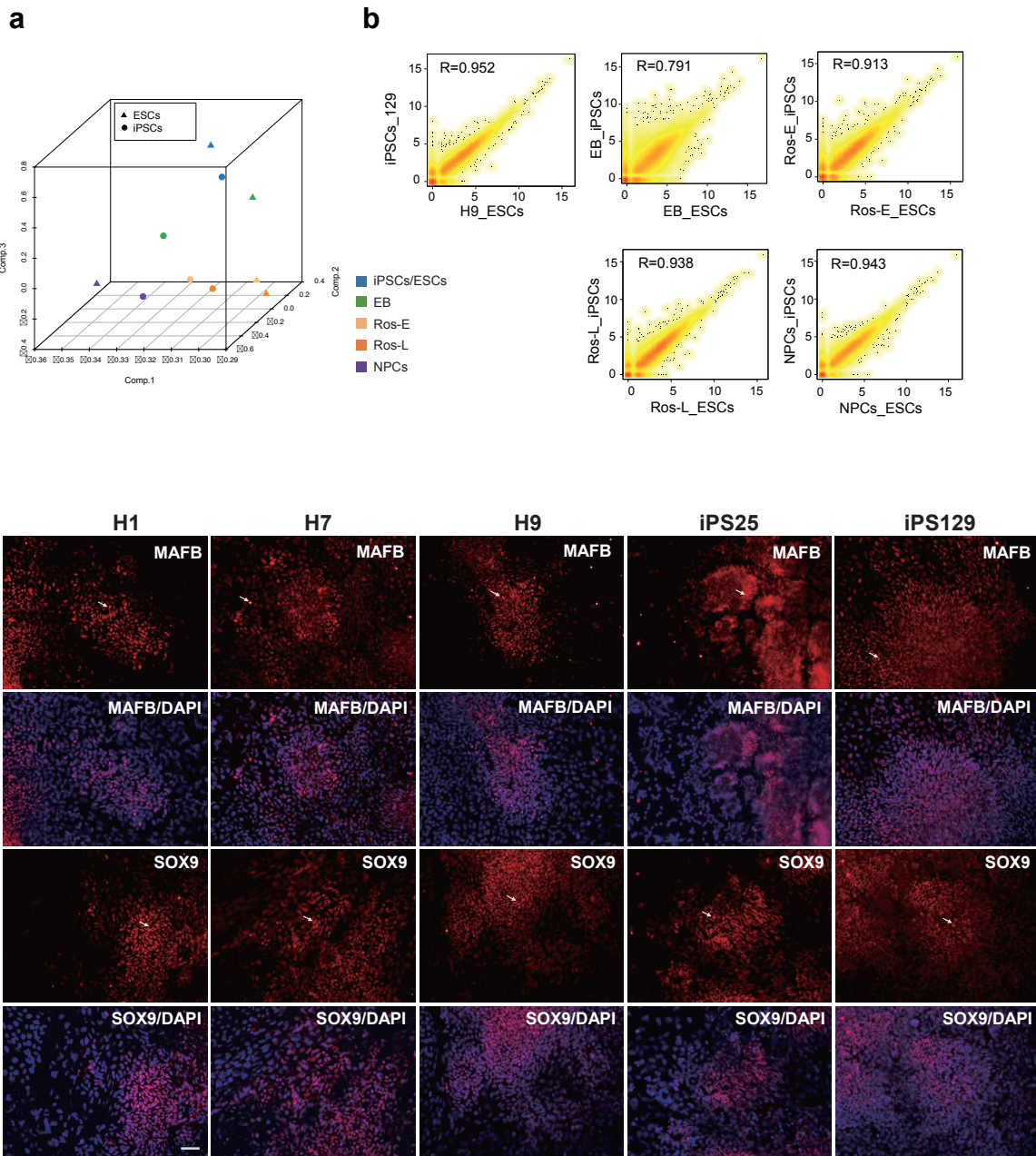
Motif	P-value	Best Match/Details
	1e-156	TEAD4/MA0809.1/Jaspar(0.970)
	1e-61	FOSL1/MA0477.1/Jaspar(0.944)
	1e-45	AP-2alpha(AP2)/Hela-AP2alpha-ChIP-Seq(GSE31477)/Homer(0.967)
	1e-38	Tcf3(HMG)/mES-Tcf3-ChIP-Seq(GSE11724)/Homer(0.891)
	1e-27	Pax2/MA0067.1/Jaspar(0.780)
	1e-26	Nur77(NR)/K562-NR4A1-ChIP-Seq(GSE31363)/Homer(0.750)
	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

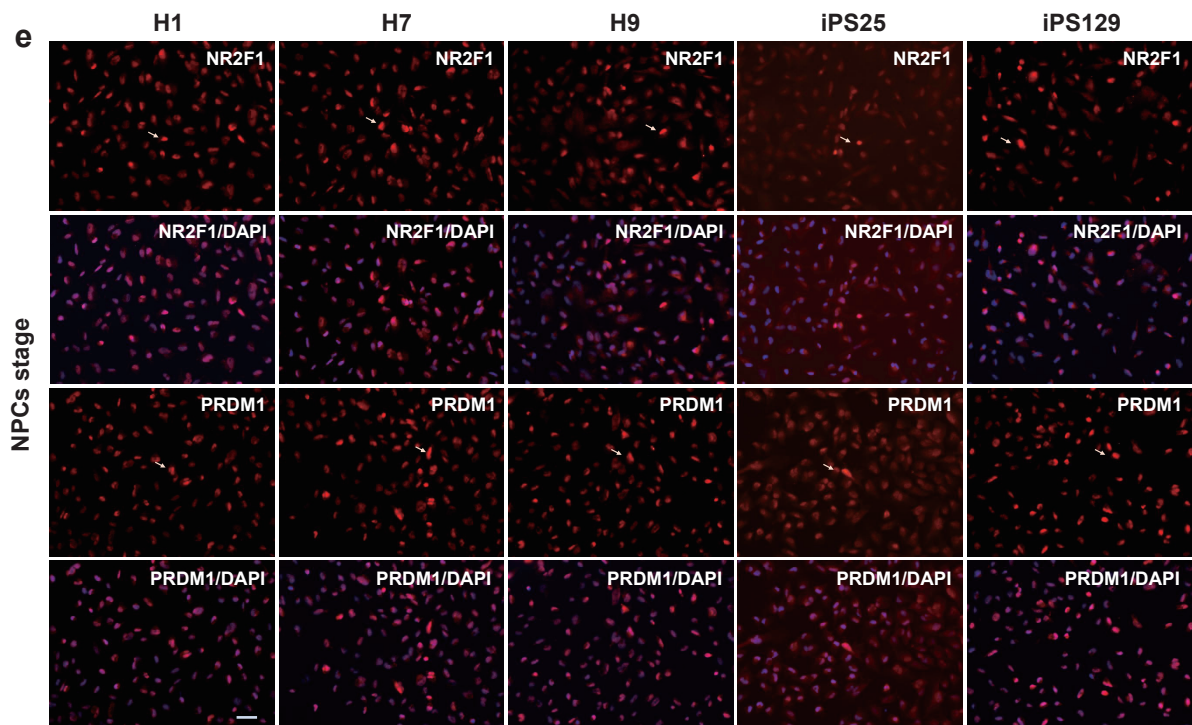
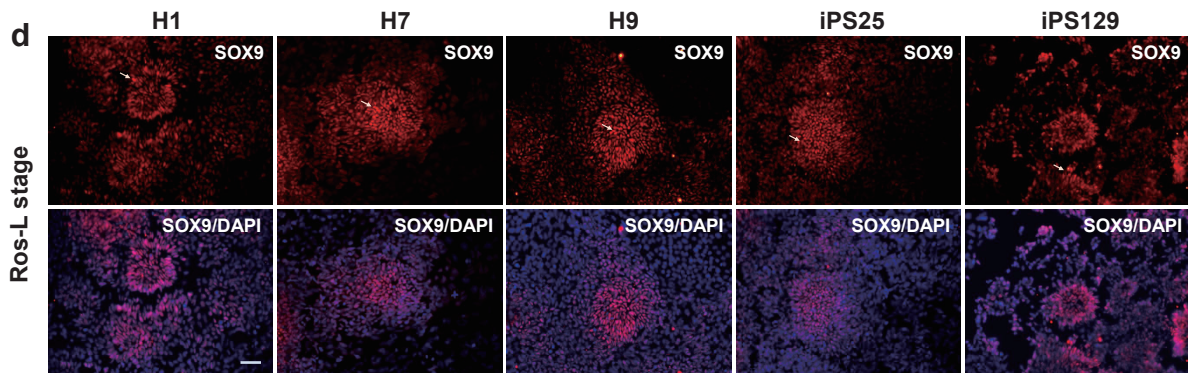
Motif	P-value	Best Match/Details
	1e-7054	AP-1(bZIP)/ThioMac-PU.1-ChIP-Seq(GSE21512)/Homer(0.990)
	1e-1065	CTCF(Zf)/CD4+-CTCF-ChIP-Seq(Barski_et_al.)/Homer(0.920)
	1e-928	Atf1/MA0604.1/Jaspar(0.898)
	1e-776	RUNX(Runt)/HPC7-Runx1-ChIP-Seq(GSE22178)/Homer(0.989)
	1e-524	Ascl1(bHLH)/NeuralTubes-Ascl1-ChIP-Seq(GSE55840)/Homer(0.961)
	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 (top 15) 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 μm .

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

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

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

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



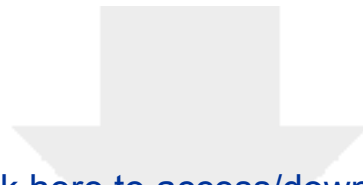
Click here to access/download
Supplementary Material
Additional file 19_Table S1.xlsx



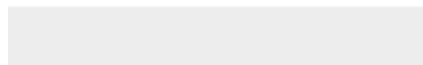


Click here to access/download
Supplementary Material
Additional file 20_Table S2.xlsx





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



Differentially expressed receptors and ligands among Ros-L subpopulations.



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

