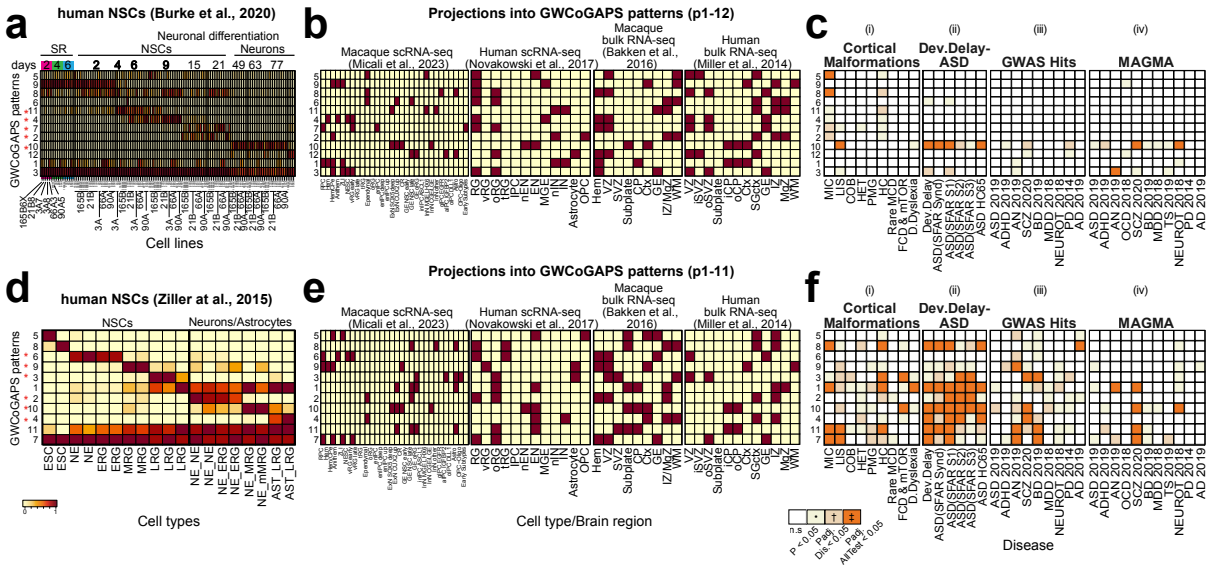
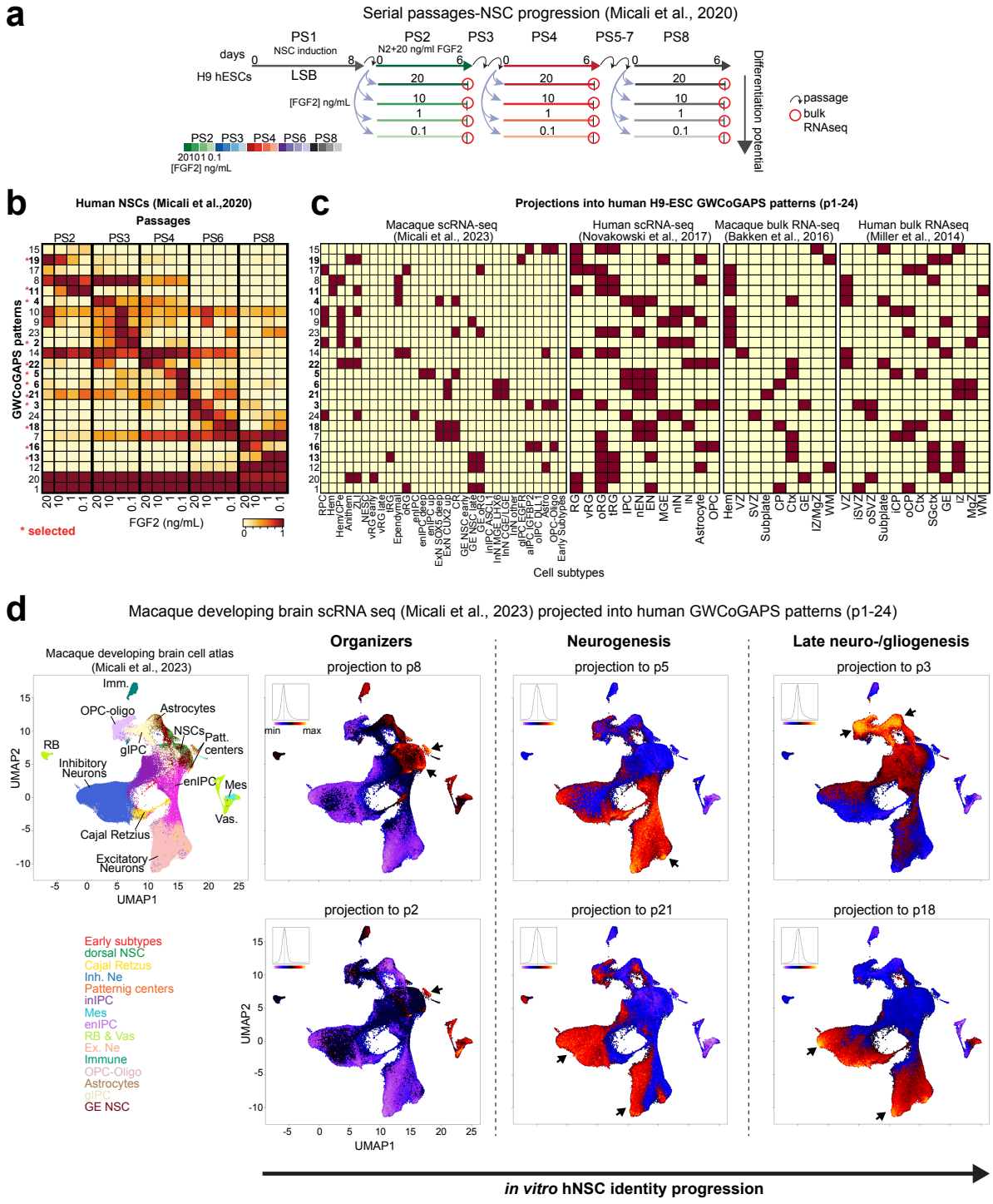
SUPPLEMENTARY FIGURES



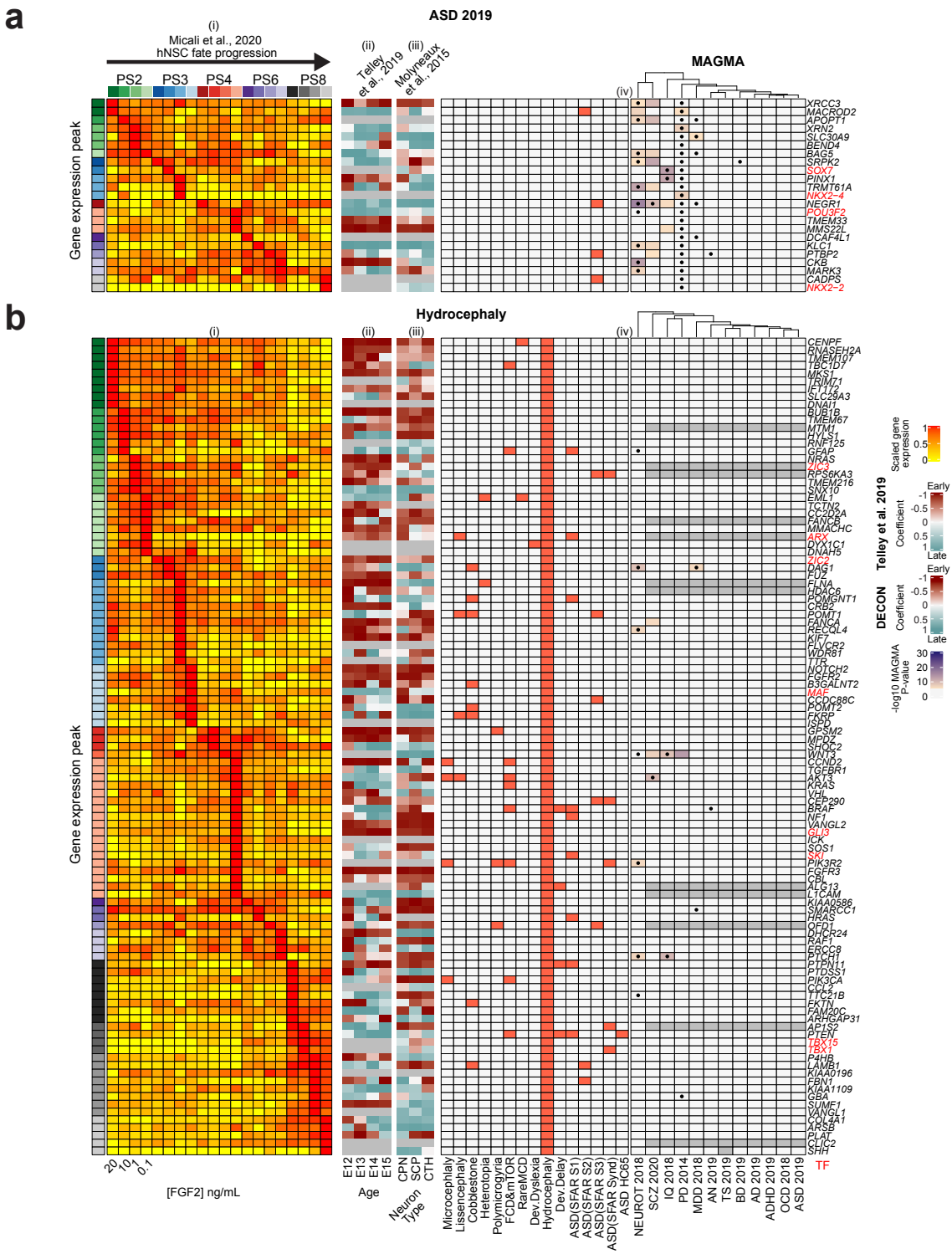
Supplementary Fig. 1. Related to Fig. 1. Cortical disorder-associated risk genes expressed in human neurons in vitro. a and d) GWCoGAPS analysis defined transcriptomic patterns across (a) neuronal differentiation of human NSC-derived from multiple iPSC lines [44] (p1-12), and (d) across sequentially passaged human NSCs derived from H9 ESCs [51] (p1-11). Column labels indicate steps of differentiation from self-renewal (SR) of the iPSCs to NSCs and terminally differentiated neurons (a), and from NSCs to terminally differentiated neurons and astrocytes (d). (*) indicates the patterns defining the strongest changes across time and distinguishing discrete cortical cells in (b and e). **b and e**) The telencephalic identity of the in vitro cells at different phases of differentiation was confirmed by projection of scRNA-seq data from the developing macaque [37] and human [48] telencephalon, and bulk RNA-seq data from microdissected developing human [49] and macaque [50] cortex into the GWCoGAPS patterns from (a) and (d). Thresholding within each in vivo dataset was used to show tissues and cells with highest expression of each GWCoGAPS pattern, i.e. dark cells indicate high levels of in vitro transcriptomic patterns in the in vivo data. D: dorsal; V: ventral; ESC: embryonic stem cells; NE: neuroepithelial stem cells; ERG: early radial glia; MRG: midradial glia; LRG: late radial glia; NE-NE: neurons derived from NE; NE-ERG: neurons derived from ERG; NE-MRG: neurons derived from MRG; NE-LRG: neurons derived from LRG; AST-LRG: astrocytes derived from LRG. RG: radial glia; oRG: outer RG; vRG: ventricular RG; tRG:truncated RG; IPC: intermediate precursor cell; nEN: new excitatory neurons; EN:

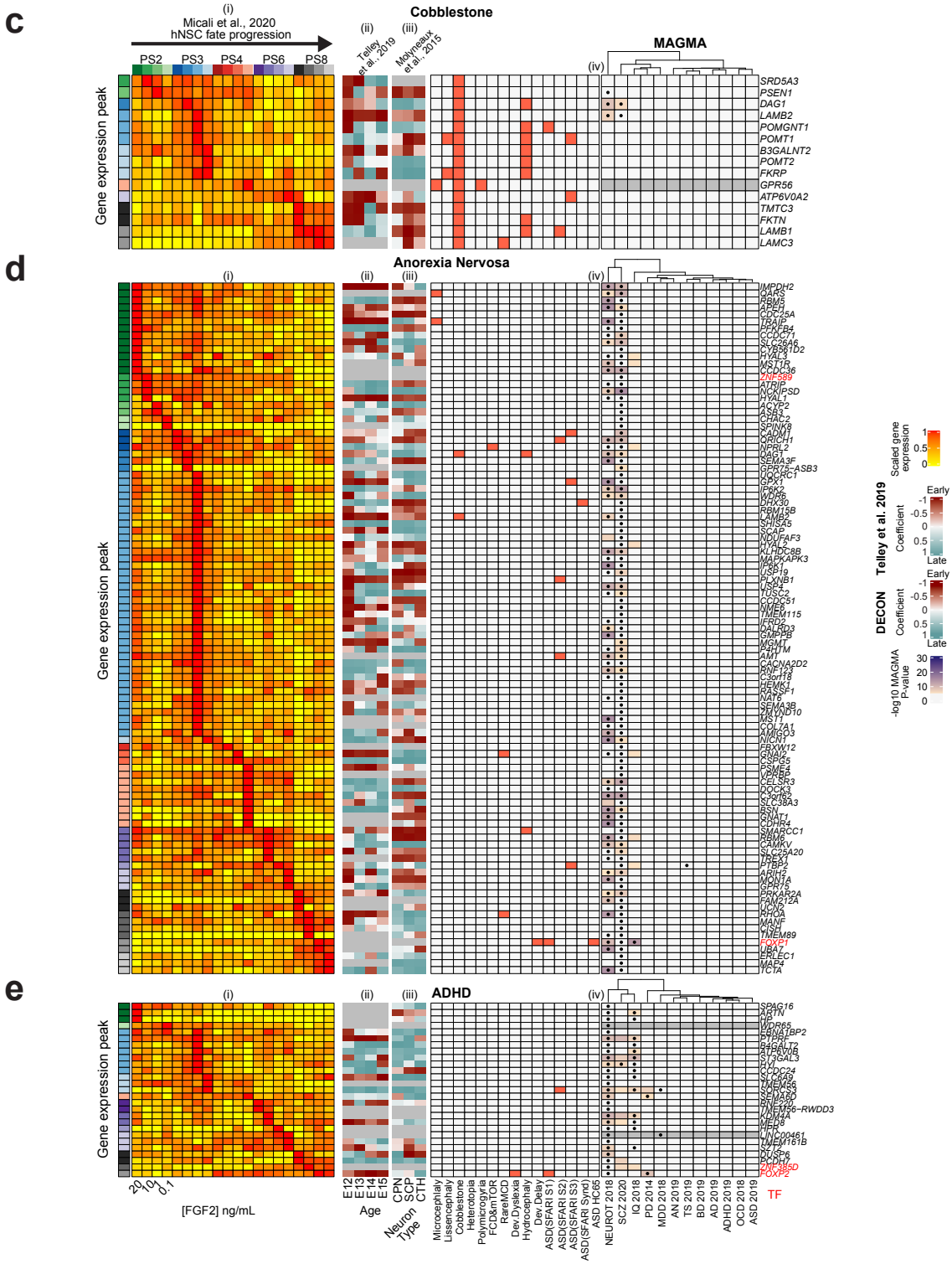
excitatory neurons; MGE: medial ganglionic eminence; GE: ganglionic eminence; nIN: new inhibitory neurons; IN: inhibitory neurons; Astro: astrocytes; OPC: oligodendrocyte progenitor cells; VZ: ventricular zone; SVZ: subventricular zone, i: inner, o: outer; CP: cortical plate; WM: white matter; IZ: intermediate zone; MGZ: marginal zone; SGctx: subgranular cortex; Ctx: cortex. **c and f**) Enrichment analysis of the disease gene sets in GWCoGAPS patterns from panels a (c) and d (f). n.s.: not significant; P: uncorrected P-values at $p<0.05$ (yellow); Padj. Dis: significance correcting by each disease independently (light orange); Padj. AllTest: significance after multiple-testing correction using the whole dataset (orange).

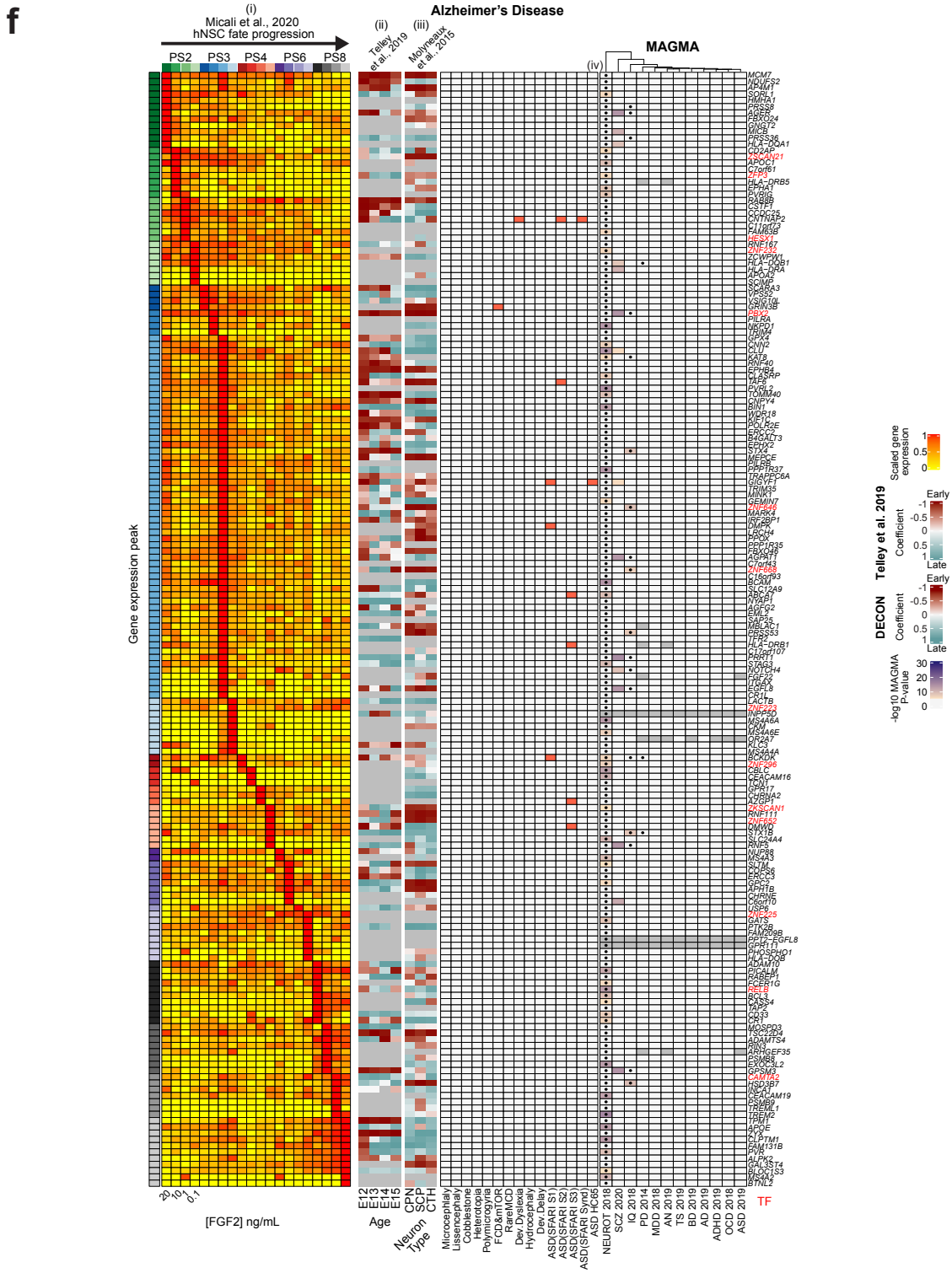


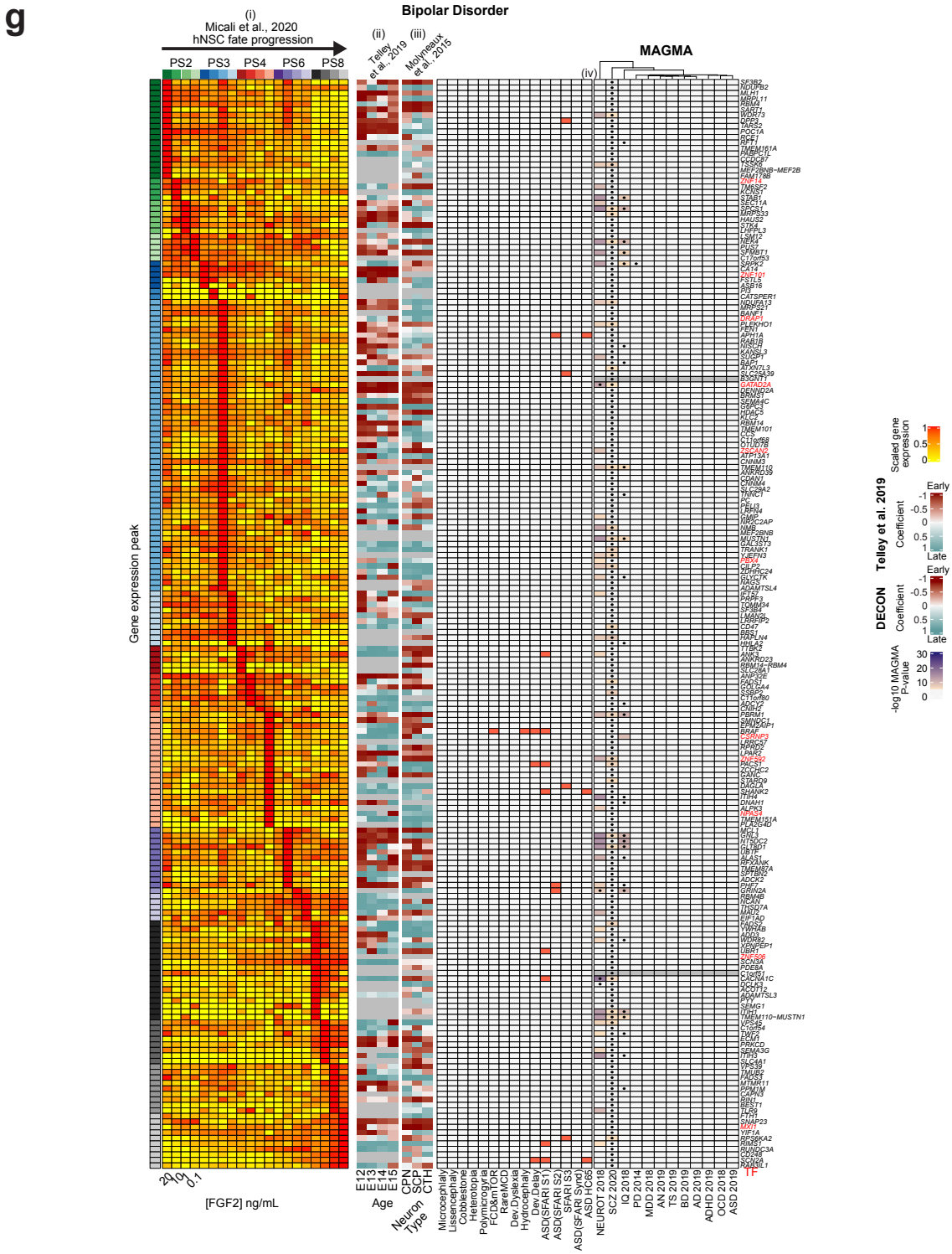
Supplementary Fig. 2. Related to Fig. 1. Sequentially passaged hNSCs recapitulate cell states and transcriptomic dynamics *of* in vivo corticogenesis. a) Scheme of H9 human embryonic stem cell (hESC) differentiation and progression of NSCs across sequential passages (PS) in presence of

different FGF2 doses from Micali et al., 2020 [40], here defined as NSC progression protocol. N2 + LSB (LDN193189 + SB431542) medium was applied at PS1 for 8 days, then hNSCs were serially passaged in N2 + 20 ng/mL FGF2 medium every 6 days up to PS8. In parallel, cultures of NSCs at PS2, 3, 4, 6, and 8 were subjected to FGF2 modulation (20, 10, 1, 0.1 ng/mL) for 6 days during their terminal passage before RNA collection. **b**) Heatmap depicting GWCoGAPS patterns (p1-24) describing progression of hNSCs across passages (PS2-8) and FGF2 dose from Micali et al., 2020 [40]. (*) indicates the patterns defining the strongest changes across passage and FGF2 and distinguishing discrete cortical cells in (c). These selected GWCoGAPS patterns are shown in Fig. 1. **c**) Projection of scRNA-seq data from the developing macaque [37] and human [48] telencephalon, and bulk RNA-seq data from microdissected developing human [49] and macaque [50] cortex into GWCoGAPS patterns from (b). Thresholding within each in vivo dataset was used to show tissues and cells with highest expression of each GWCoGAPS pattern, i.e. dark cells indicate high levels of in vitro transcriptomic patterns in the in vivo data. Abbreviations as in S1. **d**) Projection of scRNA-seq data from developing macaque brain [37] into specific GWCoGAPS patterns from [40]. Early passage patterns (p8 and p2) show high expression in in vivo organizers; mid- and late passage patterns (p5 and p21) have high expression in in vivo neurogenesis, and late passage patterns (p3 and p18) show high expression in both late neurogenesis and gliogenesis in vivo (NeMO/CoGAPS).

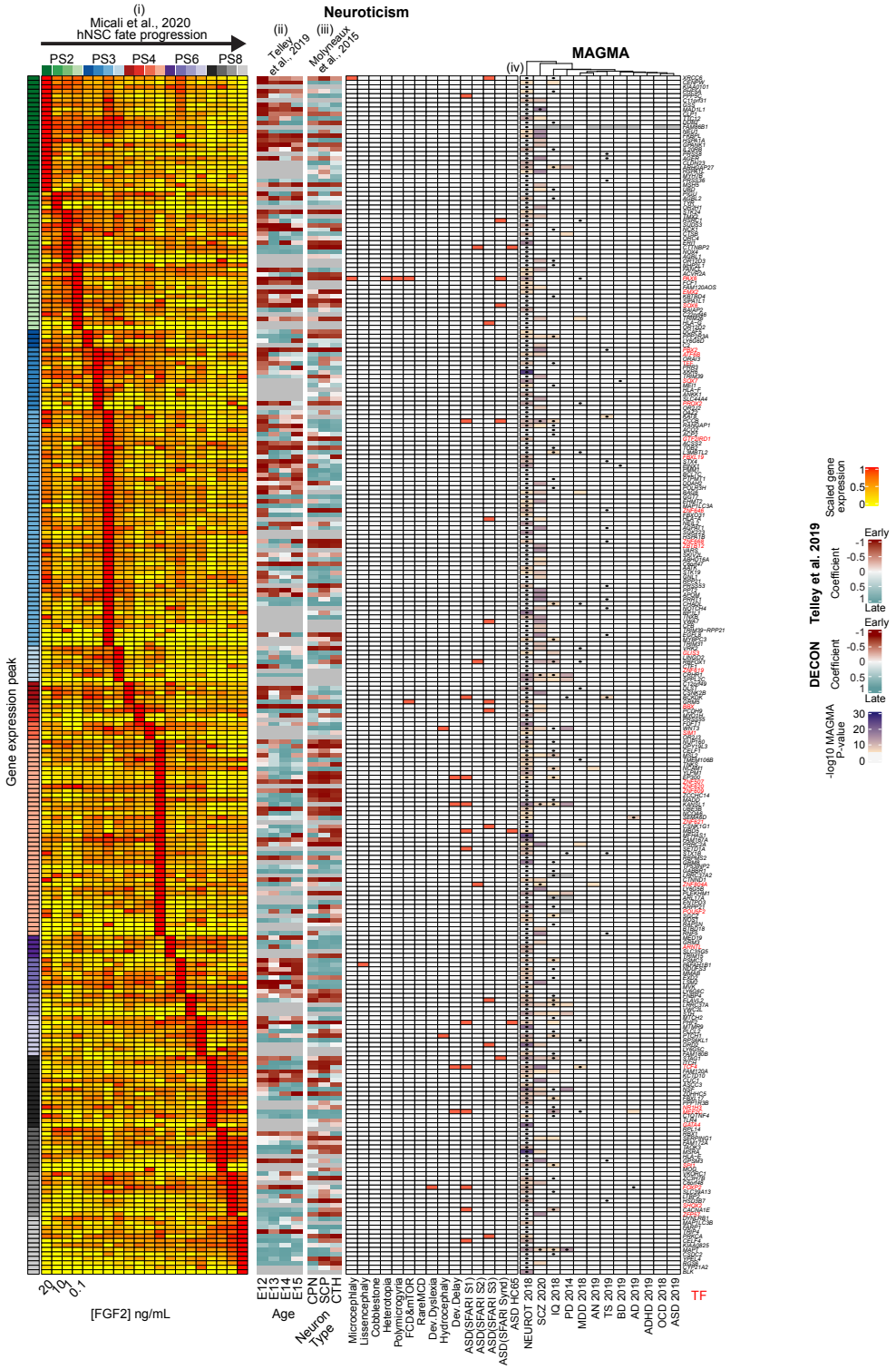


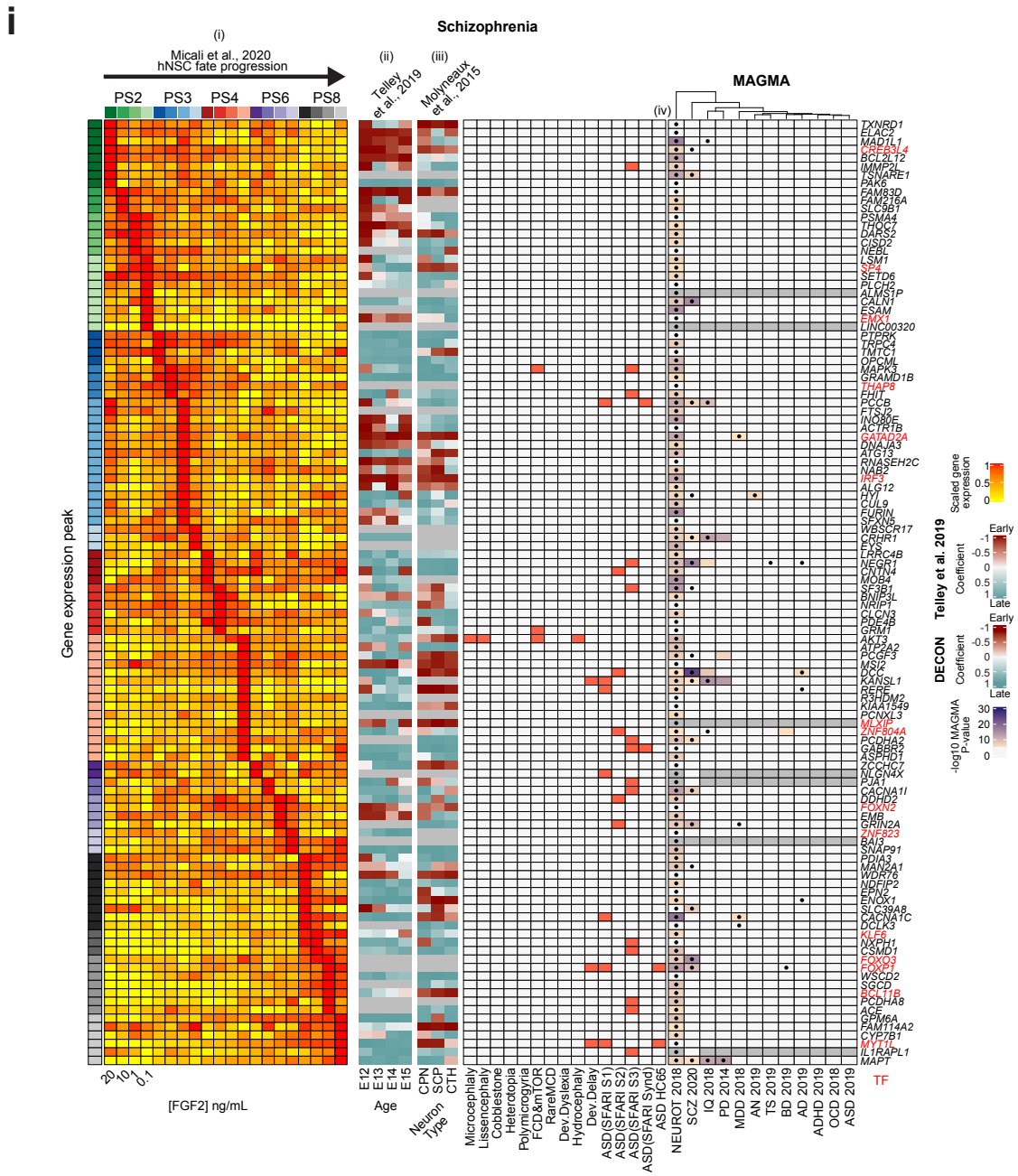


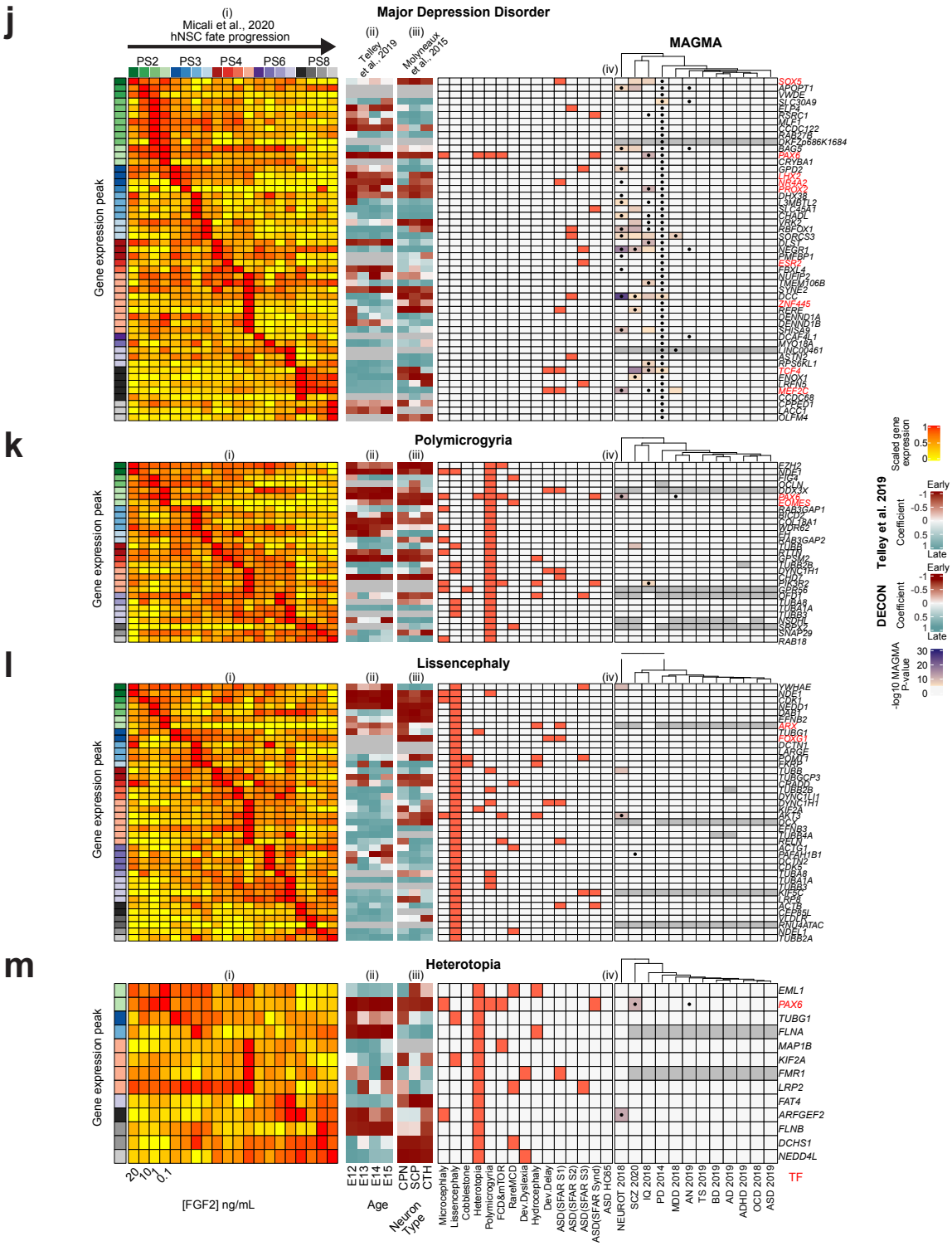


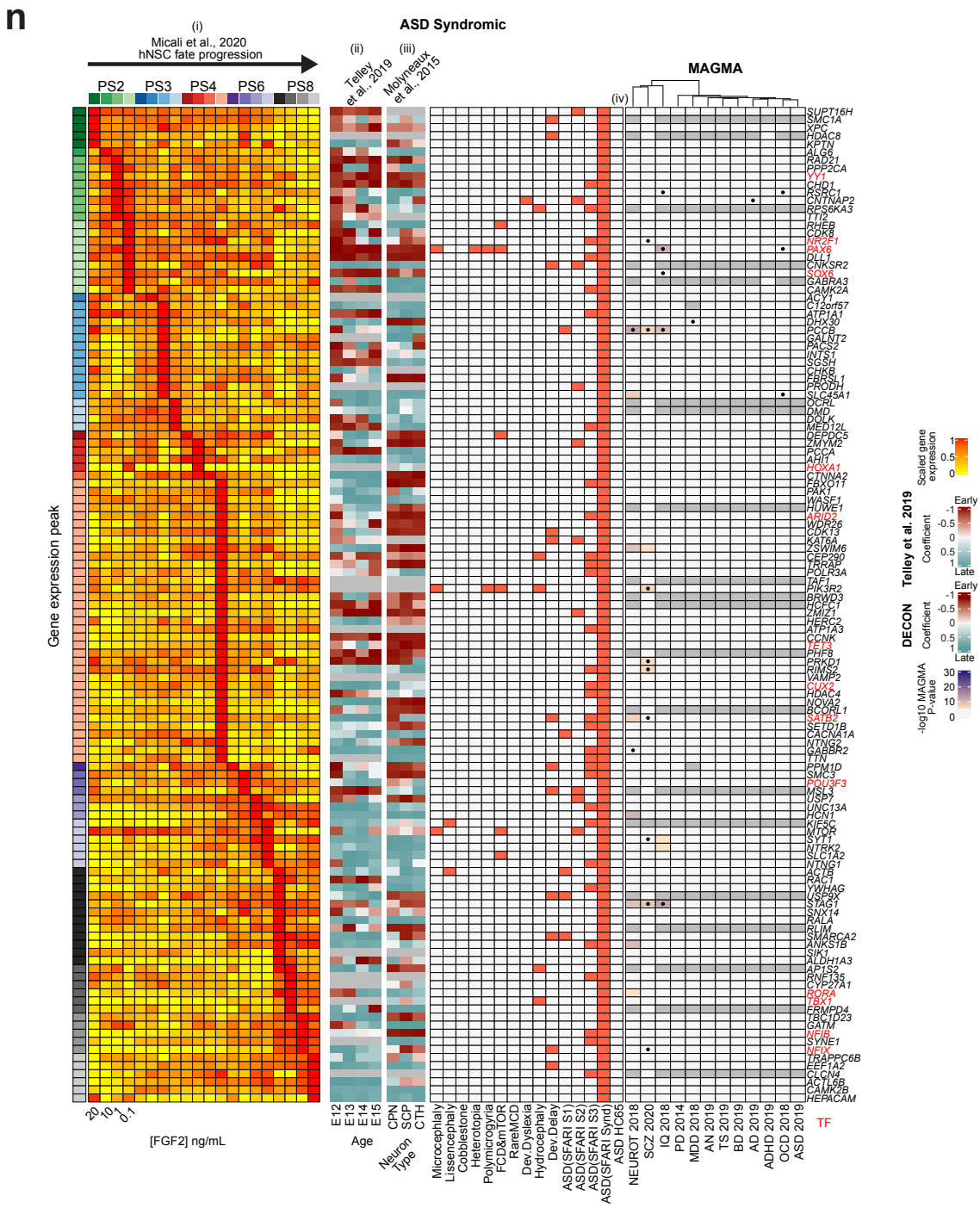


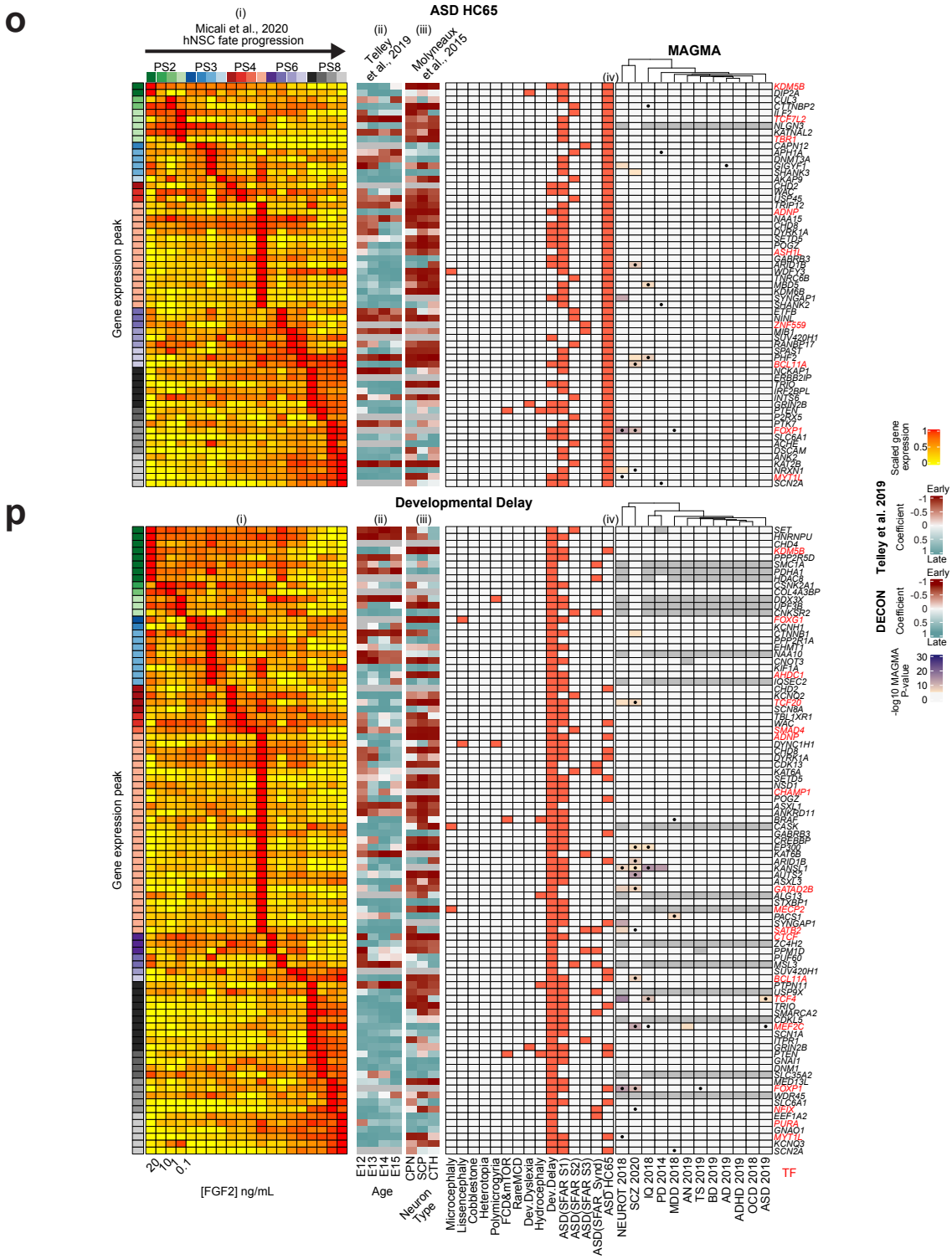
h



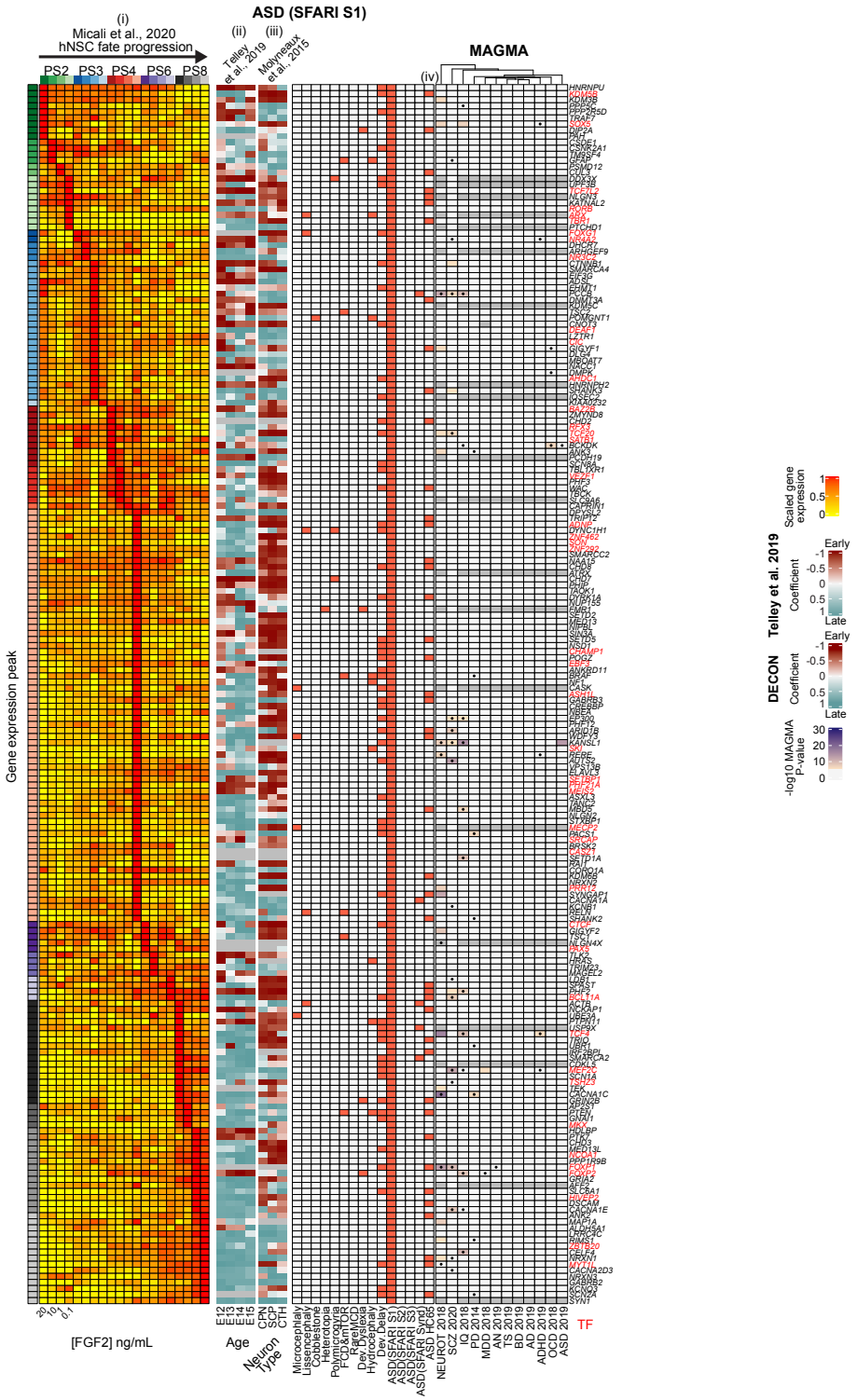




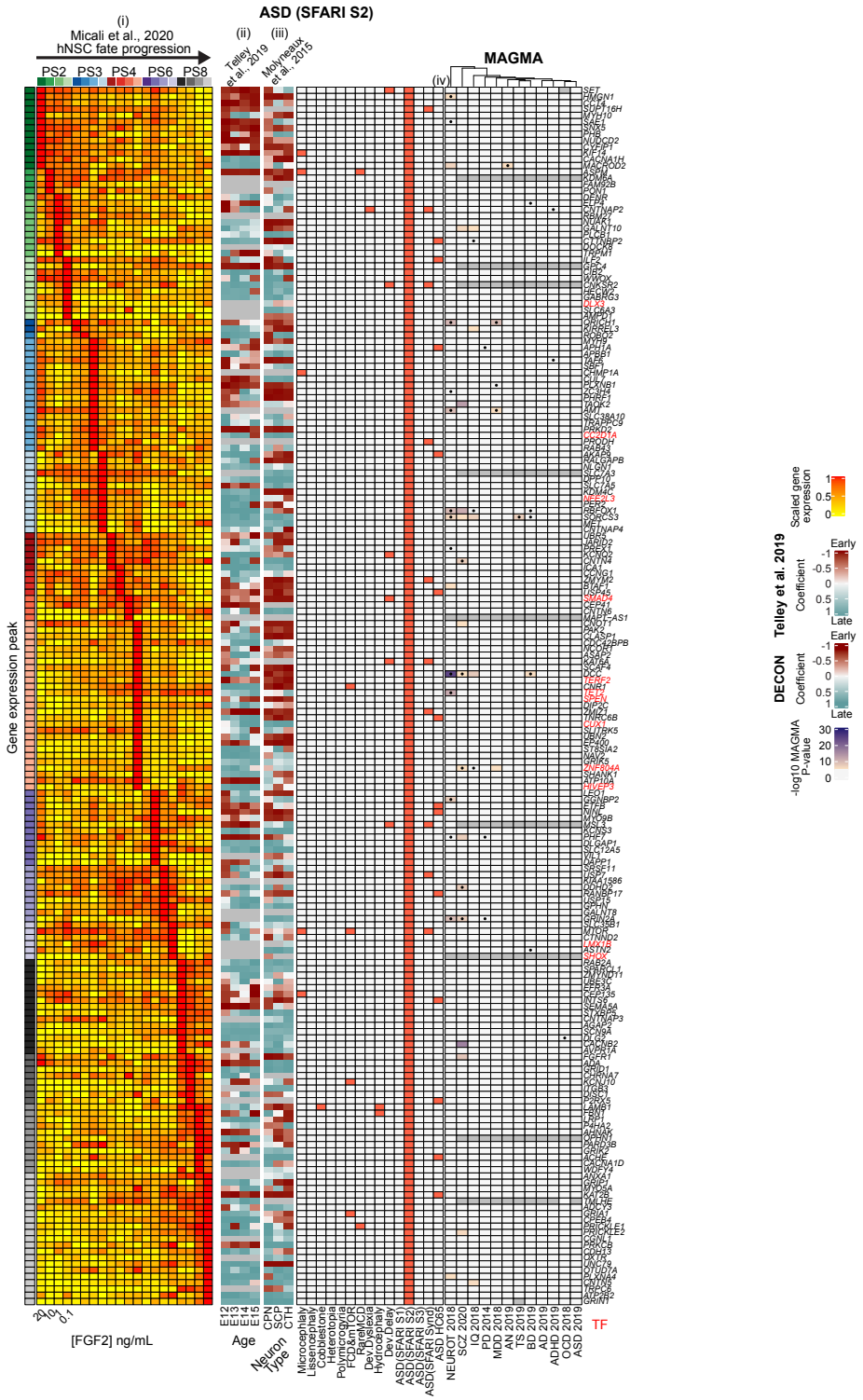




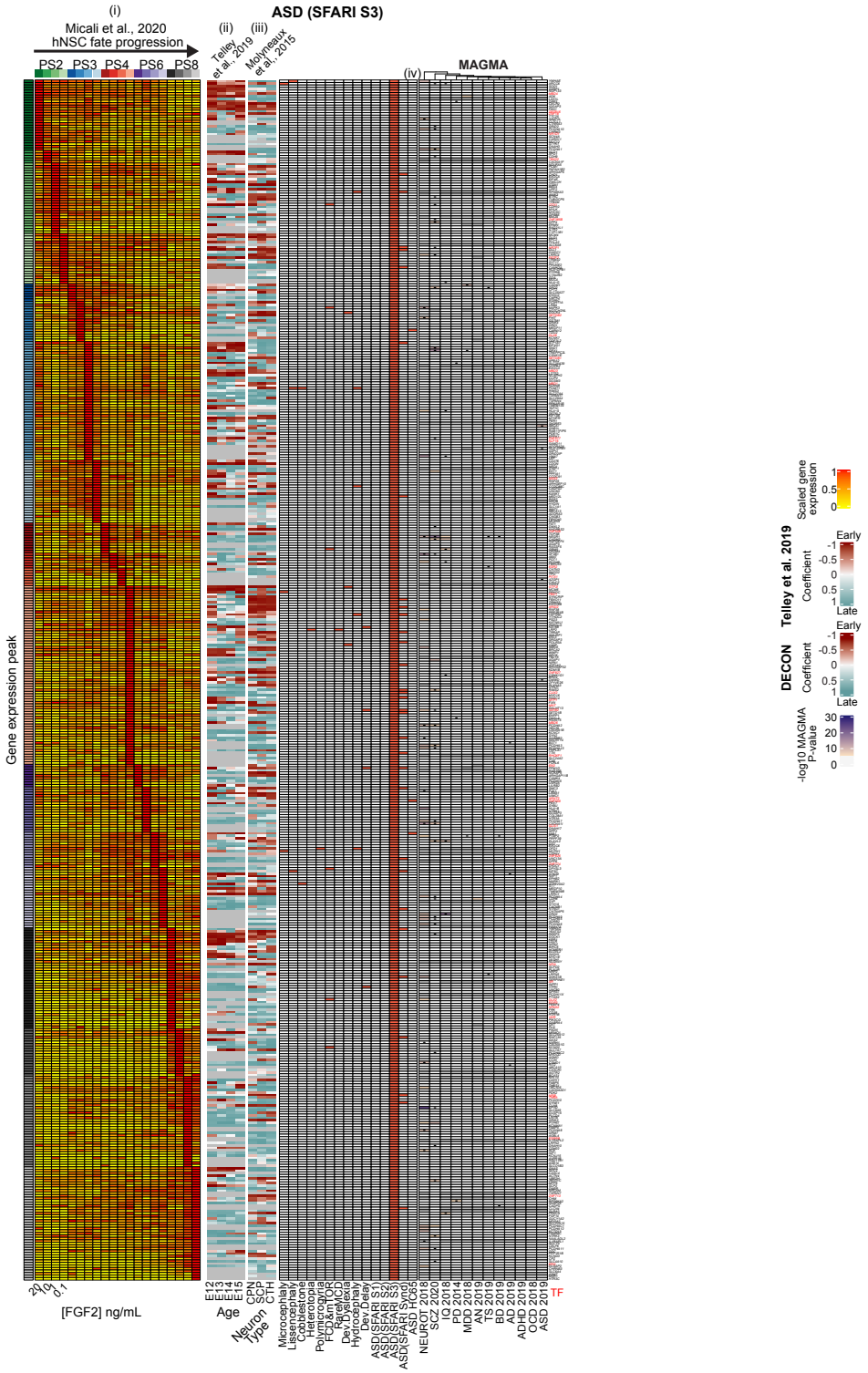
q

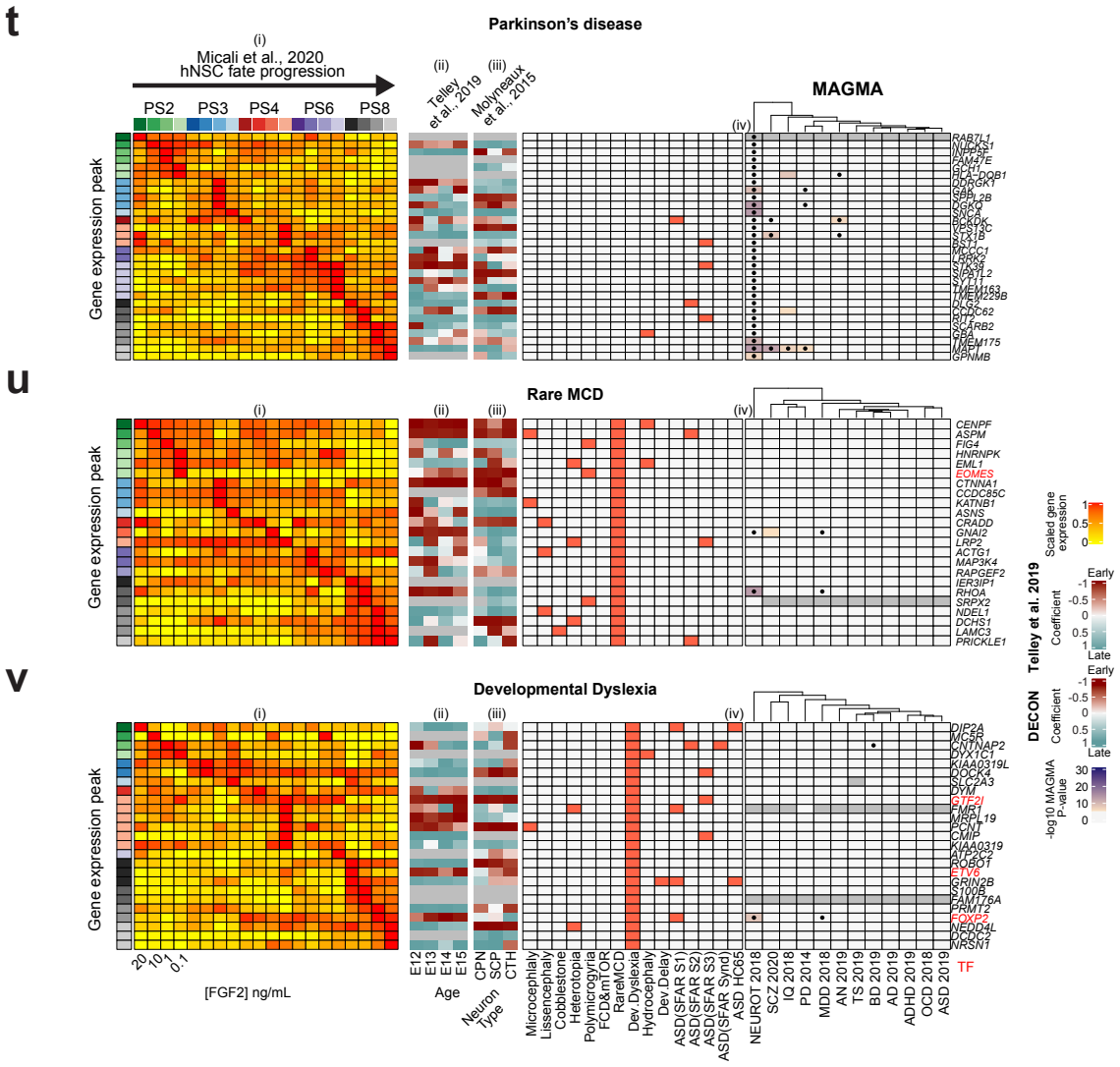


r



S

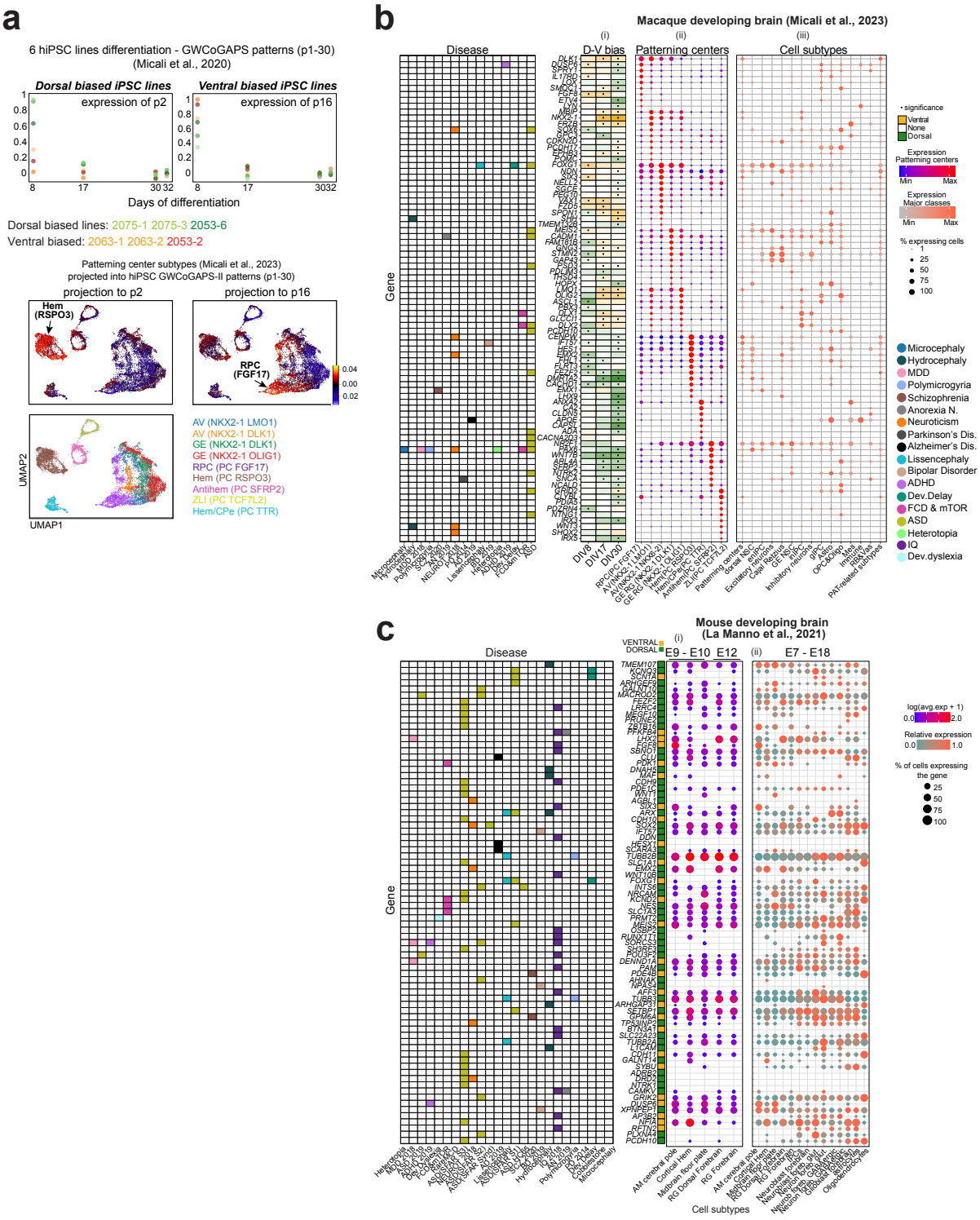




Supplementary Fig. 3. Related to Fig. 1. Expression dynamics of disease-associated risk genes across cortical neurogenesis. a-v) Multi-panel heatmaps displaying: (i) expression levels for each risk gene in the FGF2-regulated hNSC progression across passages [40]. Genes are ordered by the temporal peak of expression. The distribution is represented by the left column colored by passage and FGF2 concentration. (ii) Temporal gene expression change across neuronal differentiation of age specific RG cells from developing mouse cortex [56]. (iii) Temporal expression change across the maturation of the neurons from DeCoN dataset [57]. (iv) Disease-gene association (left panel), and log10 P value of the MAGMA gene-level test of association with each GWAS dataset (right panel). Black dots indicate a top hit gene in the corresponding GWAS publication, based on genome-wide significant loci.

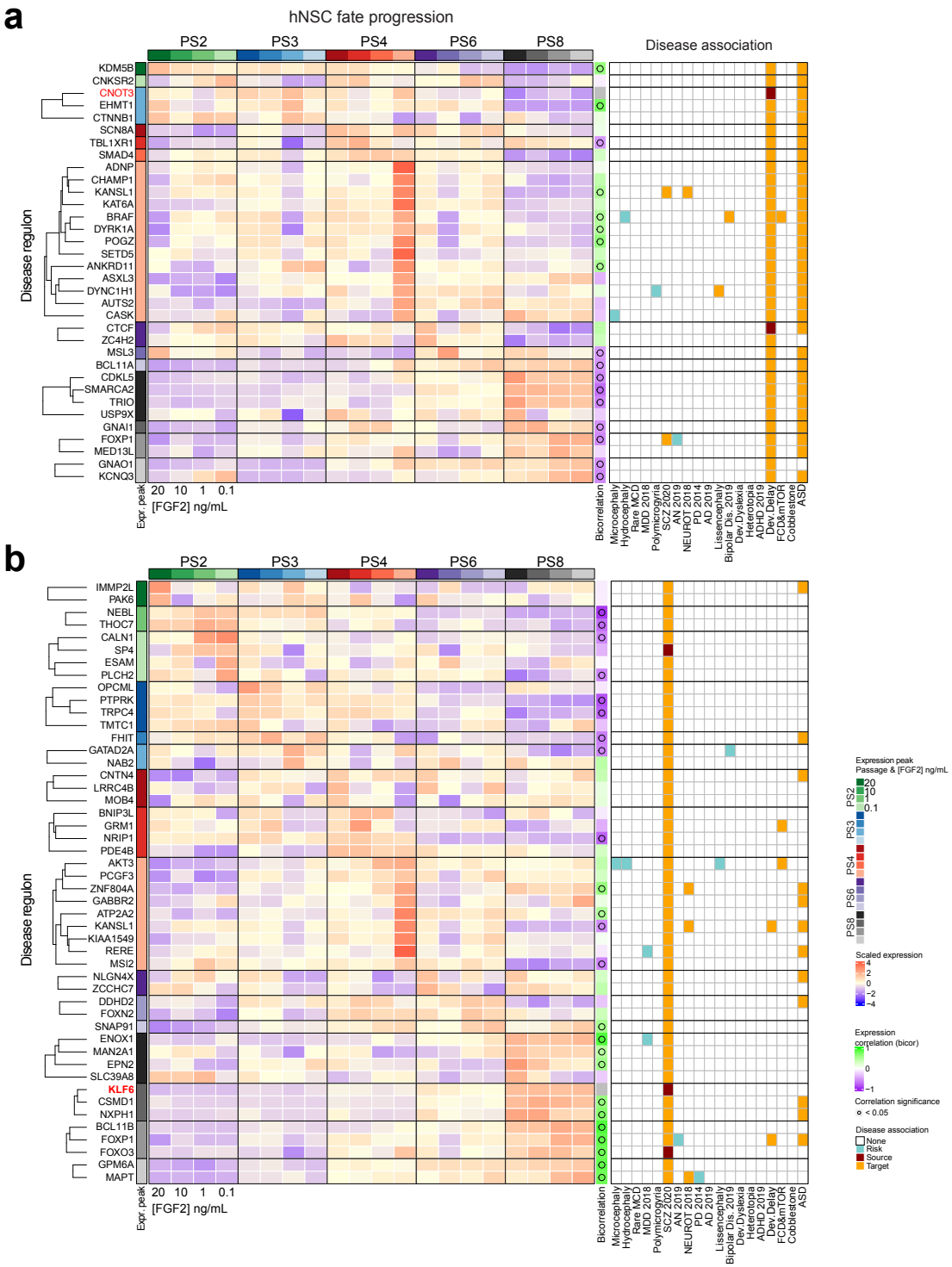
Supplementary Fig. 4. Related to Fig. 1. a) Expression dynamics of brain and non-brain disease-associated genes and FMRP and CHD8 targets across the progression of the hNSCs. Temporal distribution of the human disease-associated genes from DisGeNET across the FGF2-regulated

progression of hNSCs ordered as: PS2 and PS3 high to low, PS4 and PS8 low to high. The majority of the cortical disorder-related gene sets resulted highly enriched at PS4 low FGF2, indicating that this pattern represents a neuronal specific signature which is not shared by non-brain diseases. Glioblastoma risk genes result enriched in the late RG cells of PS8, consistent with previous works [37,78]. Moreover, certain cortical disorders and brain cancers might share similar NSC transcriptomic features. **b**) Barplot summarizing the gene proportion for each disease influenced by FMRP and CHD8. We tested the enrichment of FMRP and CHD8 targets derived from An et al., 2018, [15], Cotney et al., 2015 [79], Darnell et al., 2011 [80], Sugathan et al., 2014 [81], Casingal et al., 2020 [82], including embryonic and adult datasets, in our NDD gene sets and DisGeNET human diseases, across the progression of the in vitro hNSCs (Supplementary Table 2). FMRP targets derived from adult brain showed the highest overlap with all ASD gene sets, except for those GWAS-derived, and Dev.Delay-associated genes. Other diseases including FCD, SCZ, Neuroticism and IQ also showed a significant enrichment in specific sets of targets. LIS risk genes were enriched only in adult FMRP targets, while HET- or HC-associated genes were only enriched in embryonic FMRP targets. CHD8 targets also showed significant associations with non-GWAS ASD gene sets. CHD8 targets identified in NSCs showed significant overlap with cortical disorders caused by genes of earlier expression, for example MIC and several types of cancers. Genes associated with late onset diseases, such as AD and PD, were not enriched in any set of CHD8 and FMRP targets. These data indicate a transcriptional intersection of FMRP and CHD8 with their targets from the early phases of the NSC progression up to neurons. Besides replicating the expected association of ASD with CHD8 and FMRP targets, other cortical disorders also presented significant numbers of targets, with strong dependency on the tissue where they have been identified and developmental time. The findings suggest that the effects of the alteration of FMRP and CHD8 function extend beyond ASD and may implicate genes of earlier expression during telencephalic development involved in MCDs and other NDDs.



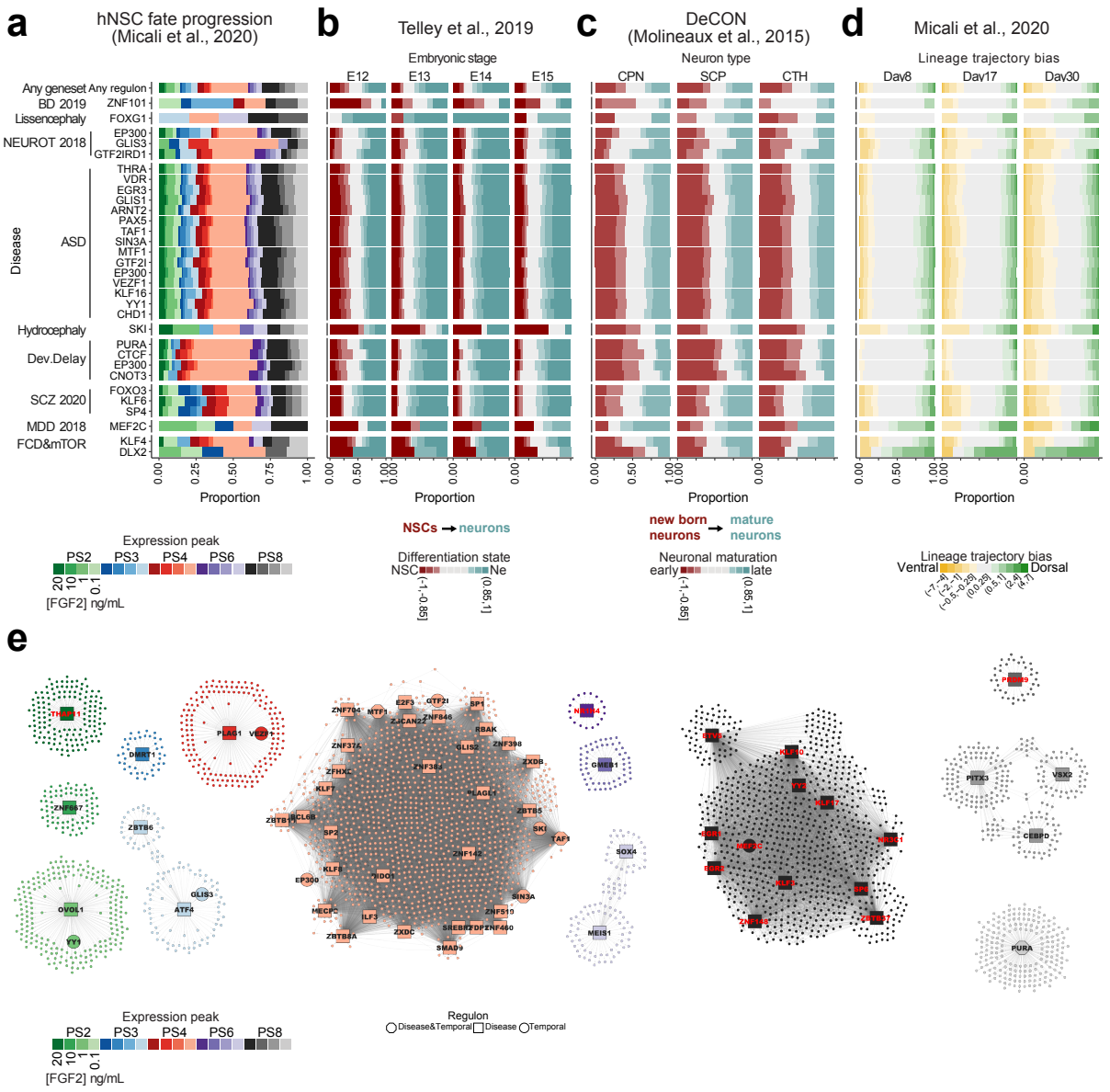
Supplementary Fig. 5. Related to Fig. 2. Expression of risk genes in telencephalic organizers. a) (Top panel) We have previously described 6 hiPSC lines (2063-1, -2; 2053-2, -6; 2075-1, -3) differentiated into forebrain NSCs. RNAseq data were generated across differentiation and decomposed

with GWCoGAPS, which identified 30 patterns of gene expression (p1-30) [40]. In Micali et al., 2020, GWCoGAPS and projection of bulk human developing cortex RNAseq data [49], distinguished 2 groups of cells: 2053-6 and 2075-1, -3, and 2063-1, -2 and 2053- 2, involving dorsal or ventral telencephalic genes across their differentiation trajectory, respectively (Supplementary Fig. 20ci). GWCoGAPS patterns p2 and p16 showed differential expression in these 2 groups of lines at DIV8, as shown in top panel here. Patterns p2 was highly expressed in 2053-6 and 2075-1, -3 lines while patterns p16 was highly expressed in 2063-1, -2 and 2053- 2 lines. The data in Micali et. al, 2020 indicate that cortical hem genes were highly weighted in p2, while anteroventral organizer genes were more represented in p16. Here, we further confirmed the differentiation bias of these 2 groups of cell lines by projection of scRNA-seq data from our recently reported developing macaque brain [37] into the GWCoGAPS patterns (p1-30). (Bottom panel) Projection of patterning center (PC) scRNA-seq data from developing macaque brain into the dorsal telencephalic pattern p2 and the ventral pattern p16 shows enrichment of p2 in monkey cortical hem cluster, and enrichment of p16 in rostral patterning center (RPC) cluster (NeMO/CoGAPSII). UMAP of all the PC clusters (bottom left panel). **b**) (i) Dorsoventral expression bias of PC marker genes from macaque brain single-cell data [37] in NSC lines at DIV 8-30, regardless of disease association and bias significance; gene expression of the same PC markers in (ii) clusters annotated as PCs [RPC (PC FGF17), AV (PC NKX2-1 LMO1), AV (PC NKX2-1 NKX6-1), GE (RG NKX2-1 DLK1), GE (NKX2-1 OLIG1), hem (PC RSPO3), hem/CPe, (PC TTR), antihem (PC SFRP2), and ZLI (PC TCF7L2)], and in (iii) other cell subtypes from the same macaque dataset. Only filtered PC marker genes are displayed, using no more than 15 genes per cell cluster, based on the lowest p-value from the original data [37] (see methods). Gene-disease association on the left. **c**) Dot plots representing risk gene expression in the cell clusters annotated as (i) patterning centers (anteromedial pole, cortical hem and floor plate) and forebrain RGs, and (ii) other neural cell types at different maturation phases from mouse fetal brain single-cell data [58]. Only disease genes with significant dorsoventral expression differences in the 6 hNSC lines at DIV 8 are displayed. Significance in dorsoventral bias is indicated in the left colored column. Gene-disease association on the left. A-M: Antero-medial cerebral pole.

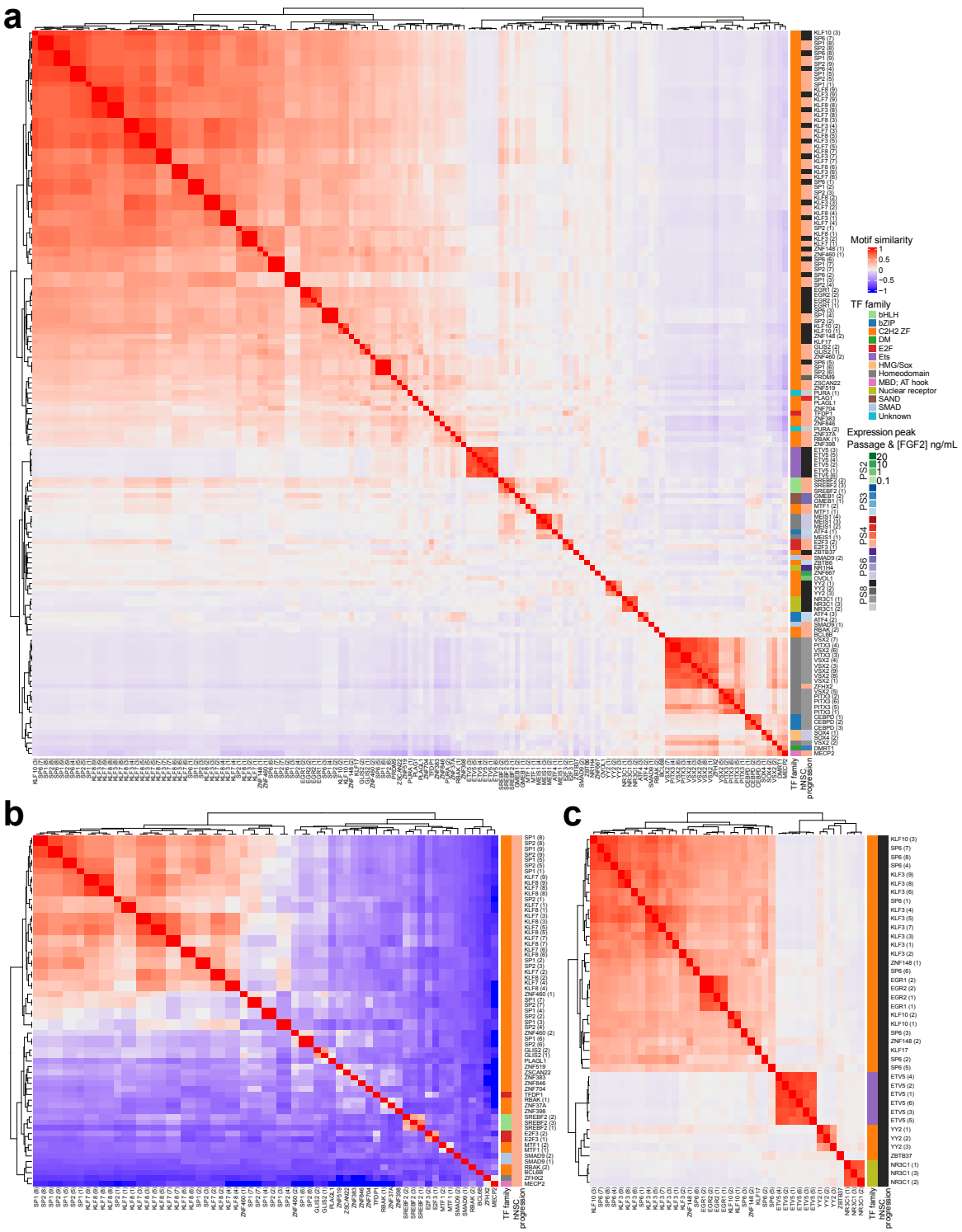


Supplementary Fig. 6. Related to Fig. 3. Expression dynamics of disease-regulon members across hNSC progression. (**a and b**) Expression of core TFs (in red) and target genes of 2 representative regulons, one early (CNOT3) and one late (KLF6), across the progressing hNSCs. The

"Peak Sample" column represents the peak of gene expression. The "Bicorrelation" column shows the correlation between the expression of any target gene and the core TF. The grid on the right indicates gene-disease association, specifying core TFs and target genes in the disease regulons. Each core TF may have a differential expression level and functionality along cortical development. Their target genes can be expressed at multiple phases across the NSC progression diverging from the core TF expression peak. This uncoupling might reflect multiple TF-effector mechanisms (e.g. co-factors' stoichiometry and their combinations, post-translational modifications and subcellular localizations, etc.), a negative regulatory effect, or unspecific associations between TFs and their motifs due to motif similarity.

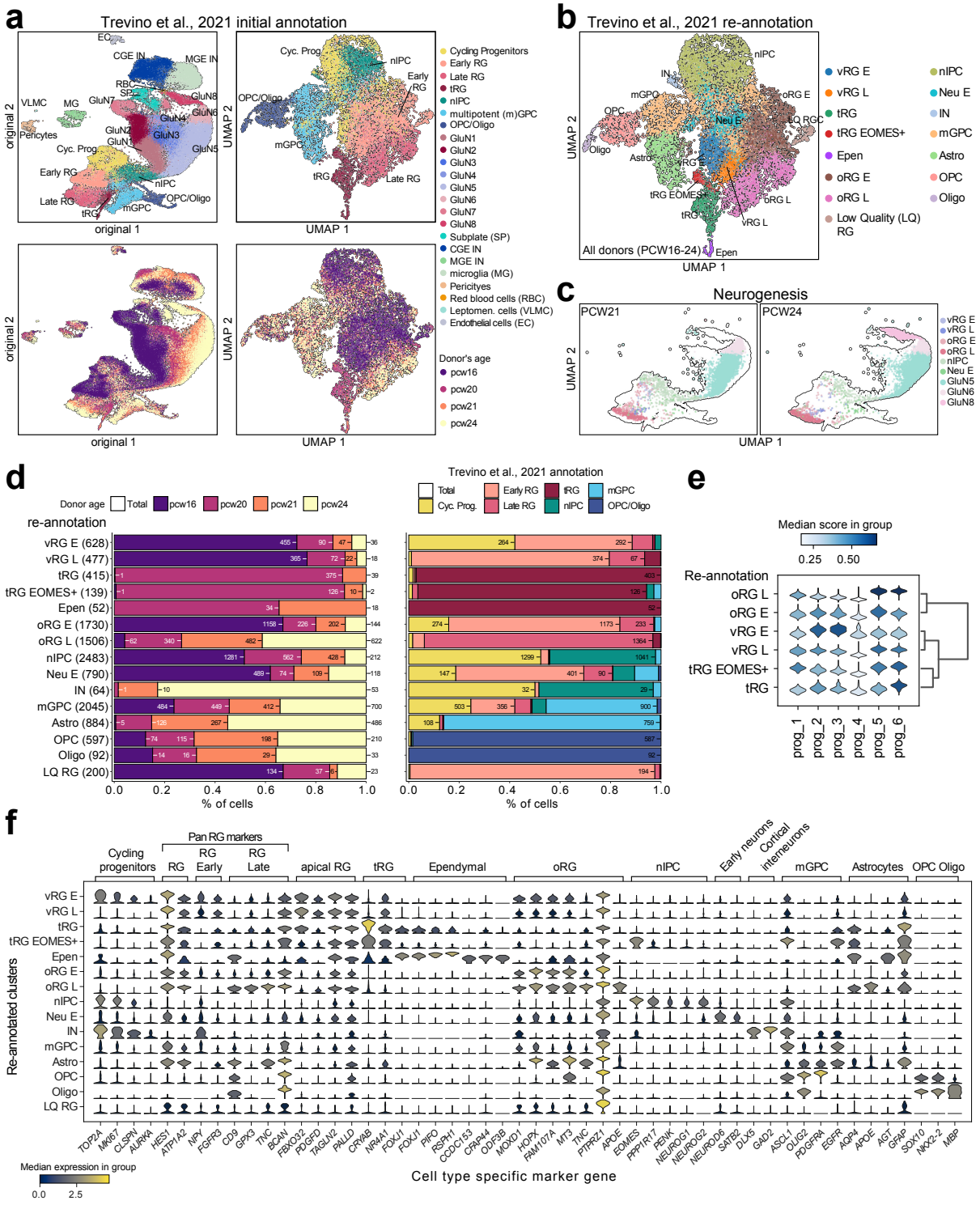


Supplementary Fig. 7. Related to Fig. 3. Disease-regulon genes expressed in vivo and in vitro. Proportion of genes for each disease regulon **a**) sorted by passages and FGF2 conditions according to their maximum expression as explained in Fig. 3a; **b**) classified into different bins of expression fold change across NSC differentiation from [56]; **c**) classified into different bins of expression fold change across the maturation of the neurons from [57]; **d**) classified into different bins of expression fold change in dorsoventral trajectory bias from [40]. **e**) Temporal regulons across passages. Small nodes represent target genes, and bigger, labelled nodes represent core TFs of temporal regulons (squares), disease regulons (circles) or both (octagons).



Supplementary Fig. 8. Related to Fig. 3. Motif similarity of core TFs in temporal regulons. (**a-c**) Similarity of all motifs associated with core TFs of regulons found in the progression of hNSCs. a) All regulons. b) Regulons from Passage 4, 0.1 ng/mL FGF2. c) Regulons from Passage 8, 20 ng/mL

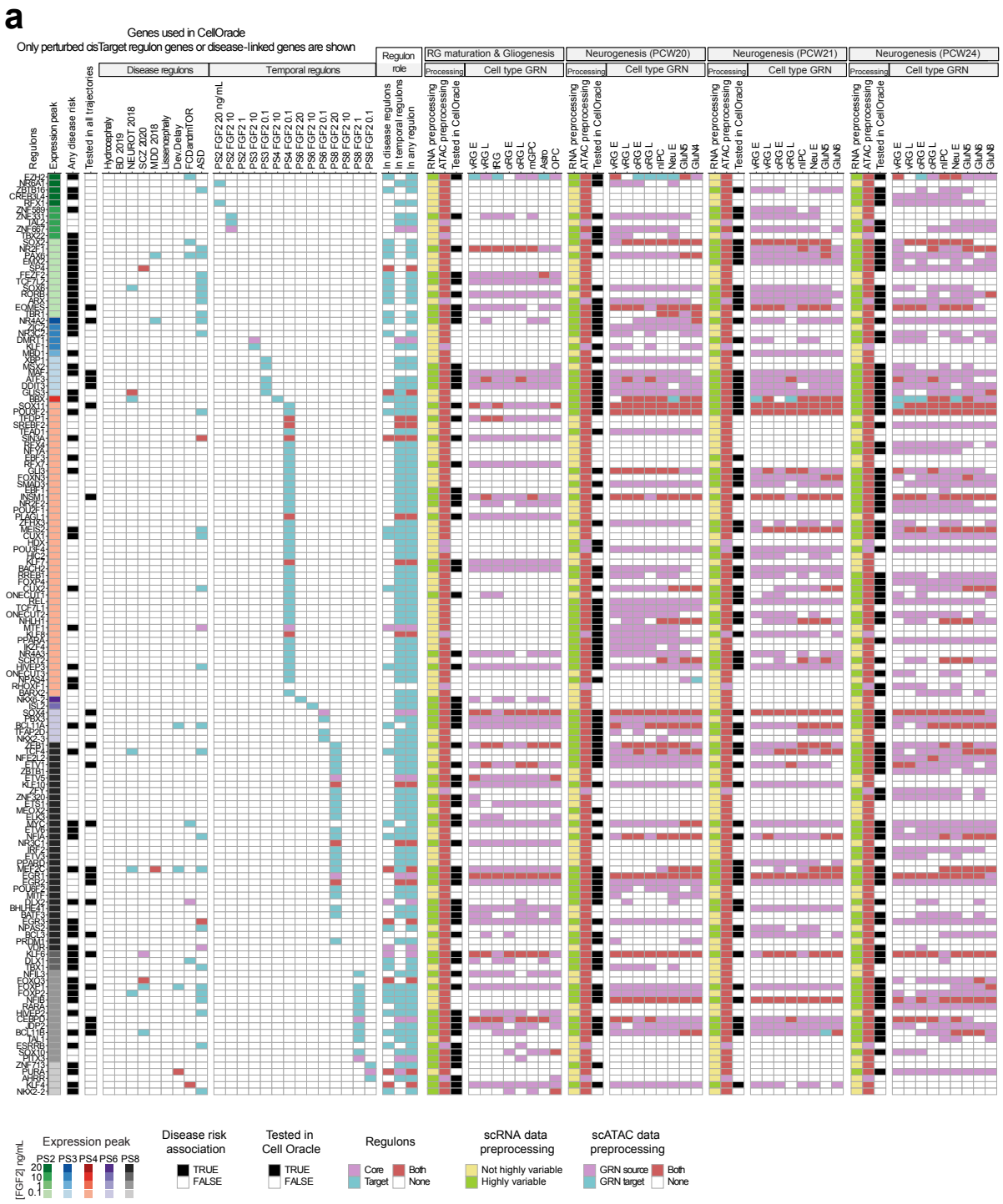
FGF2. Genes associated with multiple motifs are identified in parenthesis. The passage and FGF2 concentration of maximum expression is indicated in the right column “hNSC progression”. Motif similarity and TF-motif association data are from Lambert et al., 2018 [83].



Supplementary Fig. 9. Related to Fig. 4. Re-annotation of single cells from human fetal brain data.

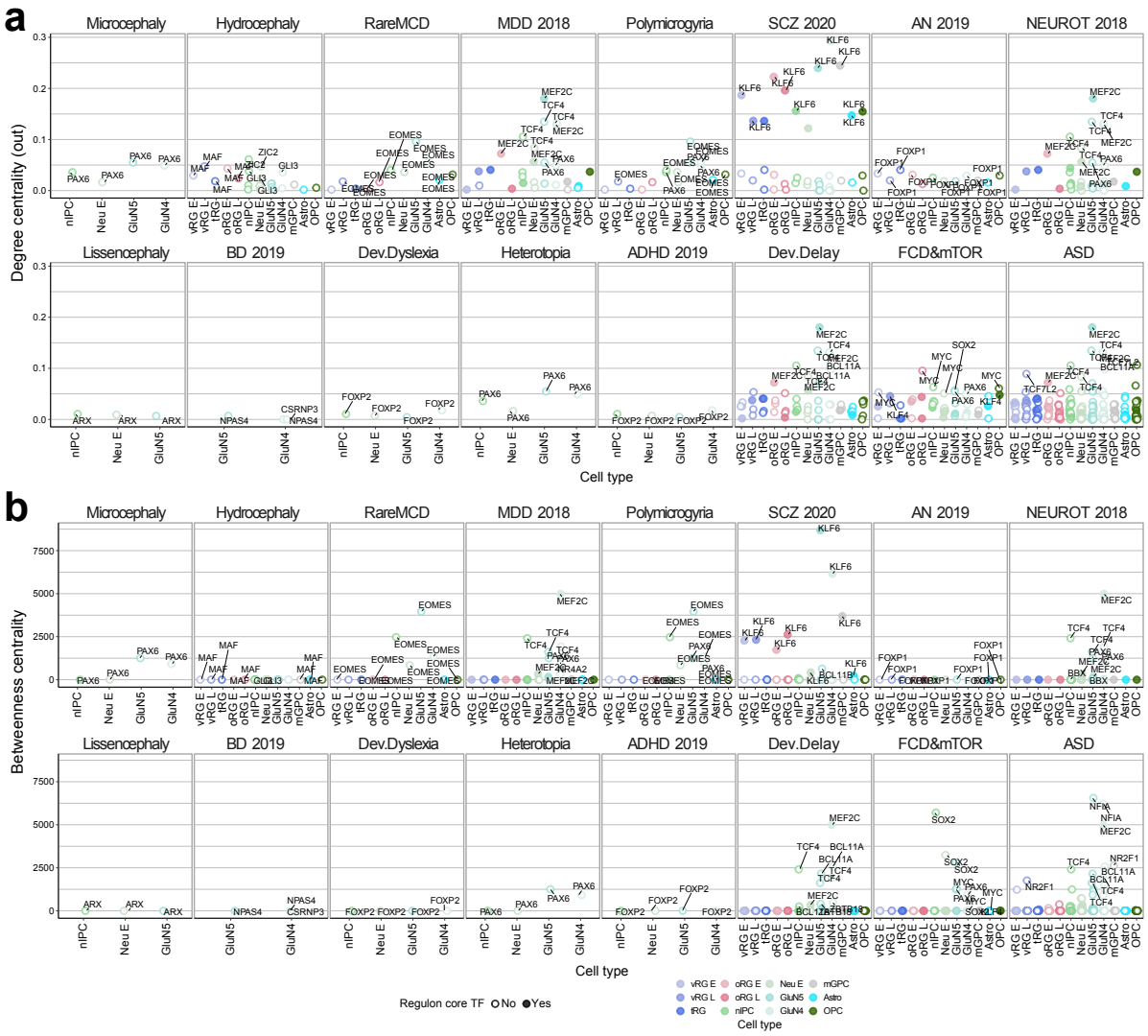
(a) UMAP representation of scRNA-seq data from developing human brain [61]. All cells in the study are represented (left). A subset of progenitor cells of interest for this study was reannotated (right). Colors

show the original cell annotation (top) and the age of the donors (bottom). **b)** UMAP colored by the new annotation of the subset of cells in this study. **c)** Neurogenesis trajectories in Trevino et al., 2021 analyzed in CellOracle. PCW20, PCW21 and PCW24 donors were independently considered. PCW16 donor was not included in the neurogenic lineage since we could not obtain a clear trajectory from NSCs towards differentiated cells in the cell diffusion maps. For the maturation of RGs and gliogenesis, all donors were included. **d)** Size, donor and distribution of the new cell identities. The total number of cells is shown next to the cell type nomenclature. The barplots represent cell distribution in the donors (left) and the distribution of cell types from the original annotation (right) present in the new cell groups. **e)** Expression scores of genes associated with the progression of RG cells. Marker genes associated with the progression of RG cells from Telley et al., 2019 [56] were used to establish the early/late subclasses of reannotated RG cells in Trevino et al., 2021. The gene module “prog_1” includes genes with both early and late expression; gene modules “prog_2” to “prog_6” classify genes according to their early-to-late expression bias. Genes in “prog_2” and “prog_3” have an early expression pattern, while “prog_5” and “prog_6” include genes with a late expression pattern. We observed early and late gene signatures enriched in specific vRG and oRG clusters which we use to label early and late vRG and early and late oRG clusters. **f)** Gene expression of cell type markers in re-annotated cell subtypes. Violin plots representing expression normalized by total counts per cell, in log-scale. The color represents the median gene expression in that group of cells.

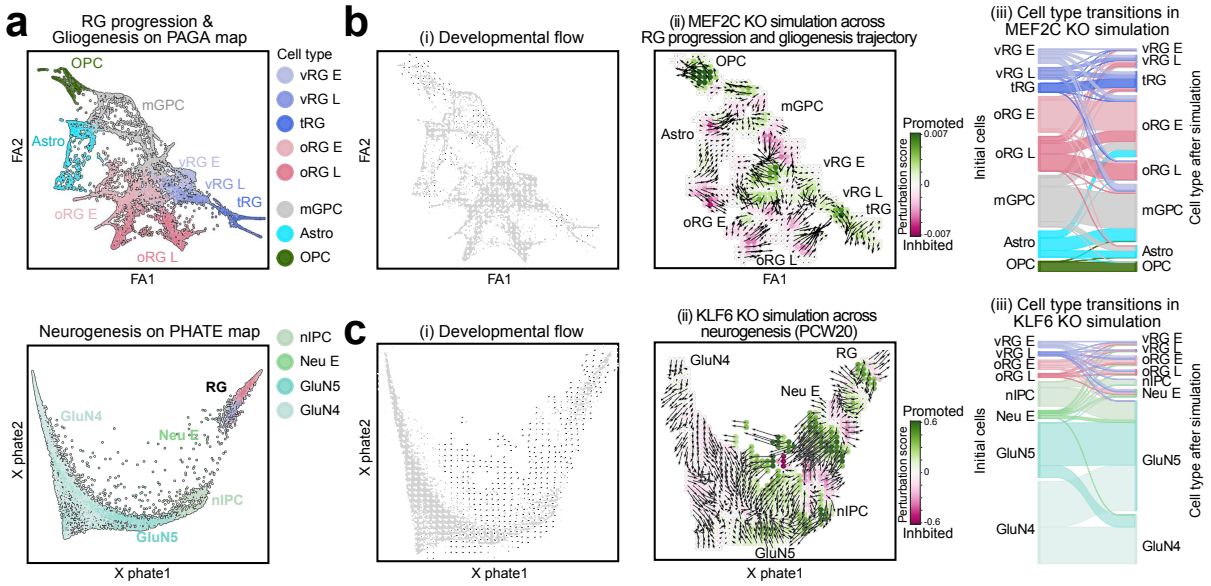


Supplementary Fig. 10. Related to Fig. 4. Overview of the genes tested in CellOracle associated with any disease or regulon. a) From the left to the right column: phase of peak expression across hNSC progression at different passage and FGF2 condition from Micali et al., 2020 i; association to any disease; test of a gene in CellOracle in both neurogenesis and gliogenesis. The columns under disease-

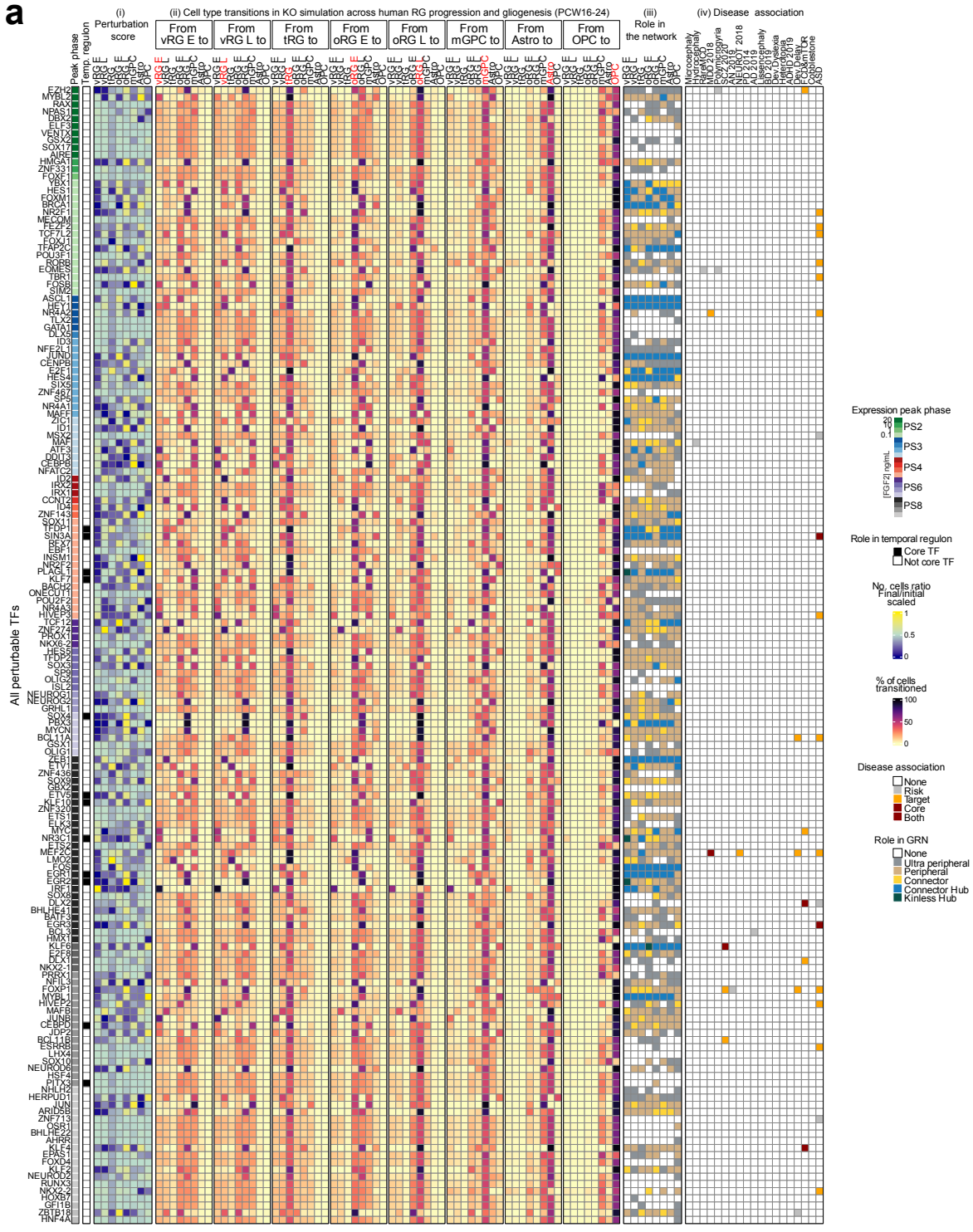
regulons and temporal-regulons show if a gene has been identified as core TF or target gene in the regulons from Fig. 3, derived from disease-associations and from the hNSC progression, respectively. These are summarized in the 'Regulon role' columns. For each subset, it is also shown if a gene passed preprocessing in scRNA-seq data and scATAC-Seq data, if the knock-out (KO) simulation succeeded, and the role (source gene or target gene) in each cell type's GRN (subsetted to the top 2000 regulatory connections in the network). ScRNA-seq preprocessing included removing low-expression genes and considering only the top 3000 highly variable genes (see methods). ScATAC-seq preprocessing included the generation of a base GRN using Cicero [84] to find coaccessible chromatin regions and a motif-based filtering to recover potentially regulatory links between these regions. For every trajectory, the gene expression data is modelled using the GRN derived from the corresponding scATAC-seq data. Expression is modeled individually in each cell type, resulting in a cell type-specific model of gene expression that connects TFs and target genes.



Supplementary Fig. 11. Related to Fig. 4. Network scores for the disease-associated TFs. The columns split the genes based on the associated diseases. The colored dots on the x axis represent the cell types with a score value for the genes on the y axis: **a**) degree centrality (out connections) represents the number of genes regulated by a given TF relative to the network size; **b**) betweenness centrality represents the importance of a given TF for connecting any two genes in the network. For genes and cell types measured in both trajectories (RG maturation and gliogenesis, and neurogenesis), data from RG maturation and gliogenesis are shown.

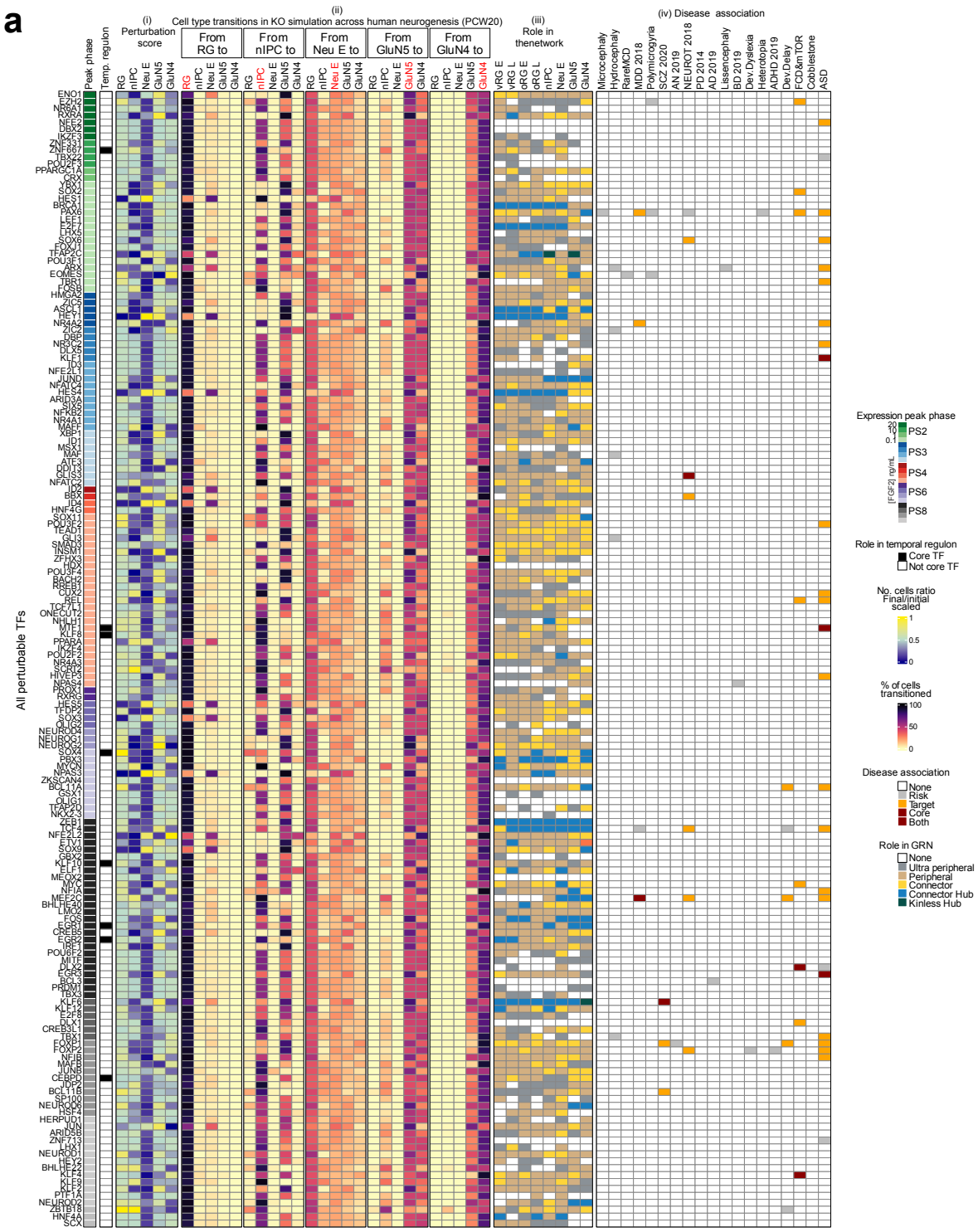


Supplementary Fig. 12. Related to Fig. 4. CellOracle KO simulation of core TFs across RG maturation/gliogenesis and neurogenesis. a) PAGA map of RG maturation and gliogenesis (top), and PHATE map of neurogenesis (bottom). **b and c**) (i) Developmental flow of (b) RG maturation and gliogenesis and (c) neurogenesis. Arrows represent the direction of the progression in the map. ii) KO simulation for *MEF2C* and *KLF6* in RG maturation/gliogenesis and neurogenesis trajectory perturbation, respectively. Arrows show the perturbation of the cell flow that the KO simulation produces, and color represents the flow direction change upon perturbation (green means same direction, i.e., promoted trajectory, and red means opposite direction, i.e., depleted trajectory). iii) Sankey plot representing the cell transitions observed in the perturbation. Original cell identities (left axis) and after KO simulation (right axis) are shown.



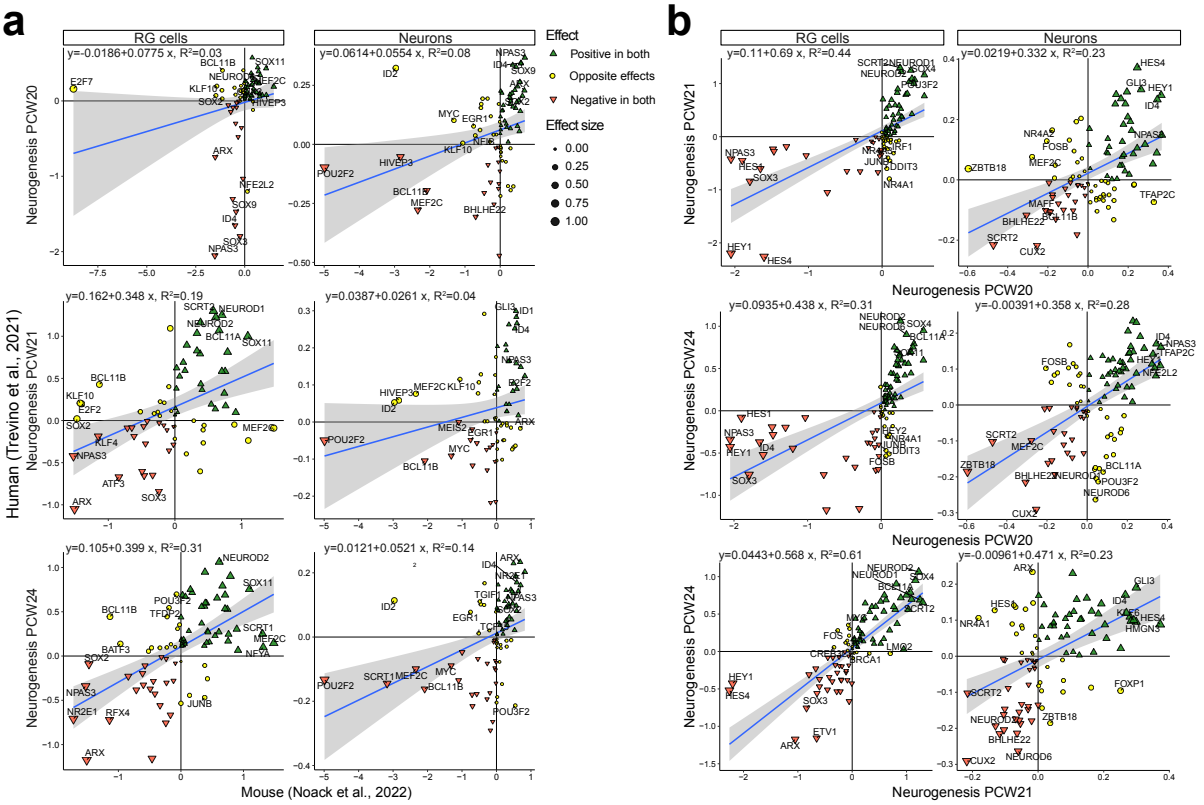
Supplementary Fig. 13. Related to Fig. 4. Comprehensive TF KO simulation across RG maturation and gliogenesis (PCW16-24). a) The peak expression phase across the in vitro hNSC progression from Micali et al., 2020 is shown for all perturbable TFs on the left column. i) Perturbation score indicating

gaining or depletion of a given cell type, measured as the ratio of cells with an identity before and after the perturbation. Values are normalized per cell type. ii) Cell type transitions upon KO simulation. The first bar of the top axis represents the original cell type, and the second bar represents the cell identity resulting from the simulation. The grids represent the fraction of the original cell type (labelled in red) and their final identity. iii) Regulatory role of every gene in each cell type. TFs connecting different GRN modules are considered hubs, and they are ordered according to their connectivity to local modules: from 'Ultra peripheral' to 'Kinless' for non-hubs, and from 'Provincial' to 'Kinless' for hub TFs. iv) Disease association of every gene. TF association to disease can be risk-only, target of a disease-regulon and core of a disease-regulon.



Supplementary Fig. 14. Related to Fig. 4. Comprehensive TF KO simulation across neurogenesis (PCW20). a) The peak expression phase across the in vitro hNSC progression from Micali et al., 2020 is shown for all perturbable TFs on the left column. i) Perturbation score indicating gaining or depletion

of a given cell type, measured as the ratio of cells with an identity before and after the perturbation. Values are normalized per cell type. ii) Cell type transitions upon KO simulation. The first bar of the top axis represents the original cell type, and the second bar represents the cell identity resulting from the simulation. The grids represent the fraction of the original cell type (labelled in red) and their final identity. iii) Regulatory role of every gene in each cell type. TFs connecting different GRN modules are considered hubs, and they are ordered according to their connectivity to local modules: from 'Ultra peripheral' to 'Kinless' for non-hubs, and from 'Provincial' to 'Kinless' for hub TFs. iv) Disease association of every gene. TF association to disease can be risk-only, target of a disease-regulon and core of a disease-regulon.



c

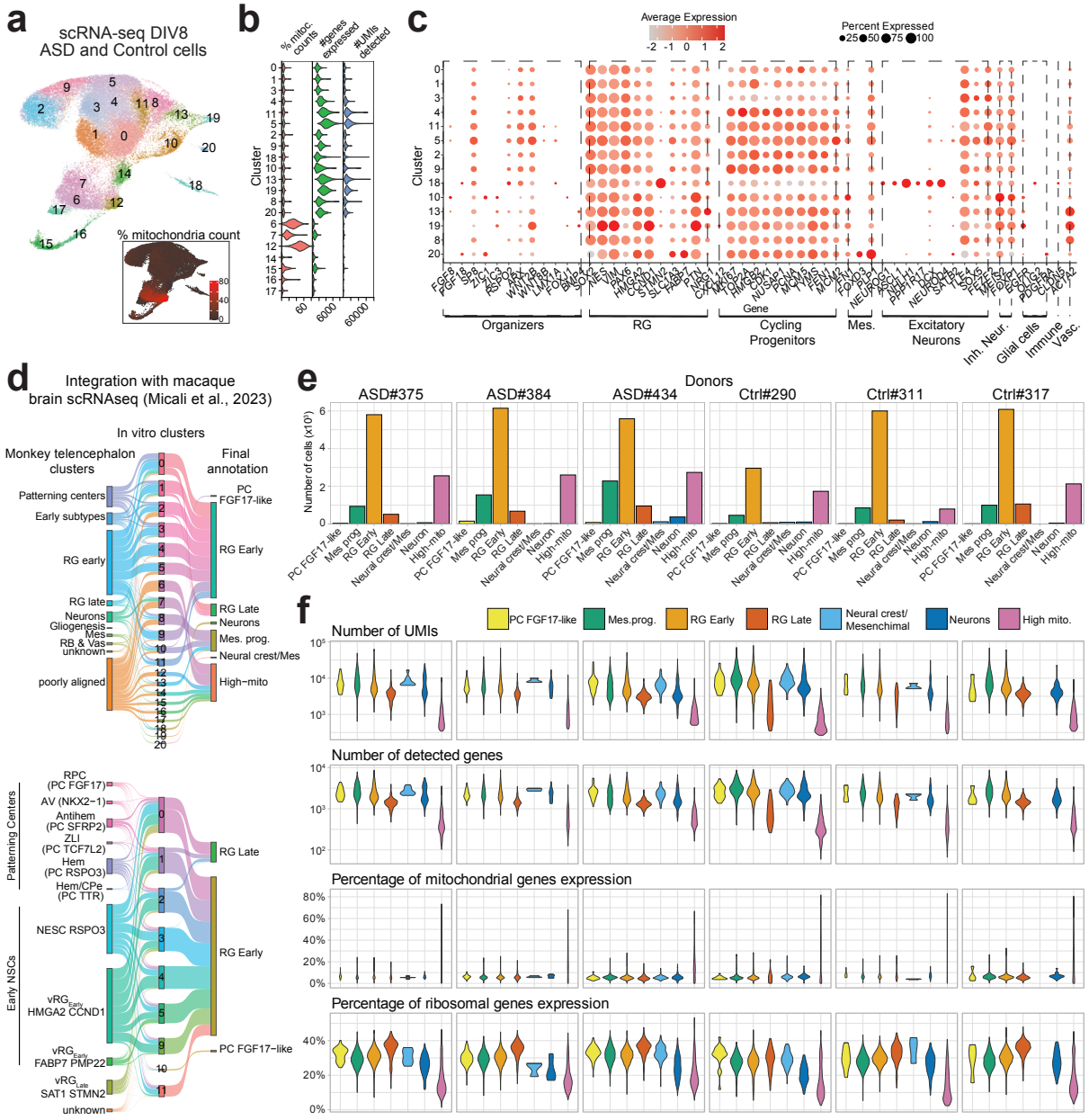
MouseEf.Up HumanEf.Up	MouseEf.Down HumanEf.Up	MouseEf.Up HumanEf.Down	MouseEf.Down HumanEf.Down	Sample	cell.type	p.value	conf.int1	conf.int2	estimate.odds ratio
46	18	3	17	Neurog.PCW20	RG cells	8.302768e-06	4.1210607	Inf	13.960282
37	16	15	16	Neurog.PCW20	Neurons	4.328851e-02	1.0319345	Inf	2.438876
29	13	10	21	Neurog.PCW21	RG cells	1.876281e-03	1.8086762	Inf	4.576463
24	16	17	16	Neurog.PCW21	Neurons	3.119320e-01	0.5835202	Inf	1.405037
46	13	5	23	Neurog.PCW24	RG cells	1.154250e-07	5.4664153	Inf	15.606102
38	9	16	24	Neurog.PCW24	Neurons	9.499809e-05	2.5410847	Inf	6.181935

d

Ef_sample1.Up Ef_sample2.Up	Ef_sample1.Down Ef_sample2.Up	Ef_sample1.Up Ef_sample2.Down	Ef_sample1.Down Ef_sample2.Down	comparison	cell.type	sample1	sample2	p.value	conf.int1	conf.int2	estimate.odds ratio
50	2	22	26	Neurog.PCW20 Neurog.PCW21	RG cells	Neurog.PCW20	Neurog.PCW21	7.499871e-09	7.466614	Inf	28.512630
36	16	24	24	Neurog.PCW20 Neurog.PCW21	Neurons	Neurog.PCW20	Neurog.PCW21	3.927038e-02	1.047344	Inf	2.231334
57	3	15	30	Neurog.PCW20 Neurog.PCW24	RG cells	Neurog.PCW20	Neurog.PCW24	5.950077e-12	11.158840	Inf	36.227311
45	15	19	26	Neurog.PCW20 Neurog.PCW24	Neurons	Neurog.PCW20	Neurog.PCW24	6.523910e-04	1.880342	Inf	4.045168
51	10	5	33	Neurog.PCW21 Neurog.PCW24	RG cells	Neurog.PCW21	Neurog.PCW24	2.143424e-12	11.115958	Inf	31.848012
40	16	11	32	Neurog.PCW21 Neurog.PCW24	Neurons	Neurog.PCW21	Neurog.PCW24	5.659355e-06	3.118874	Inf	7.105299

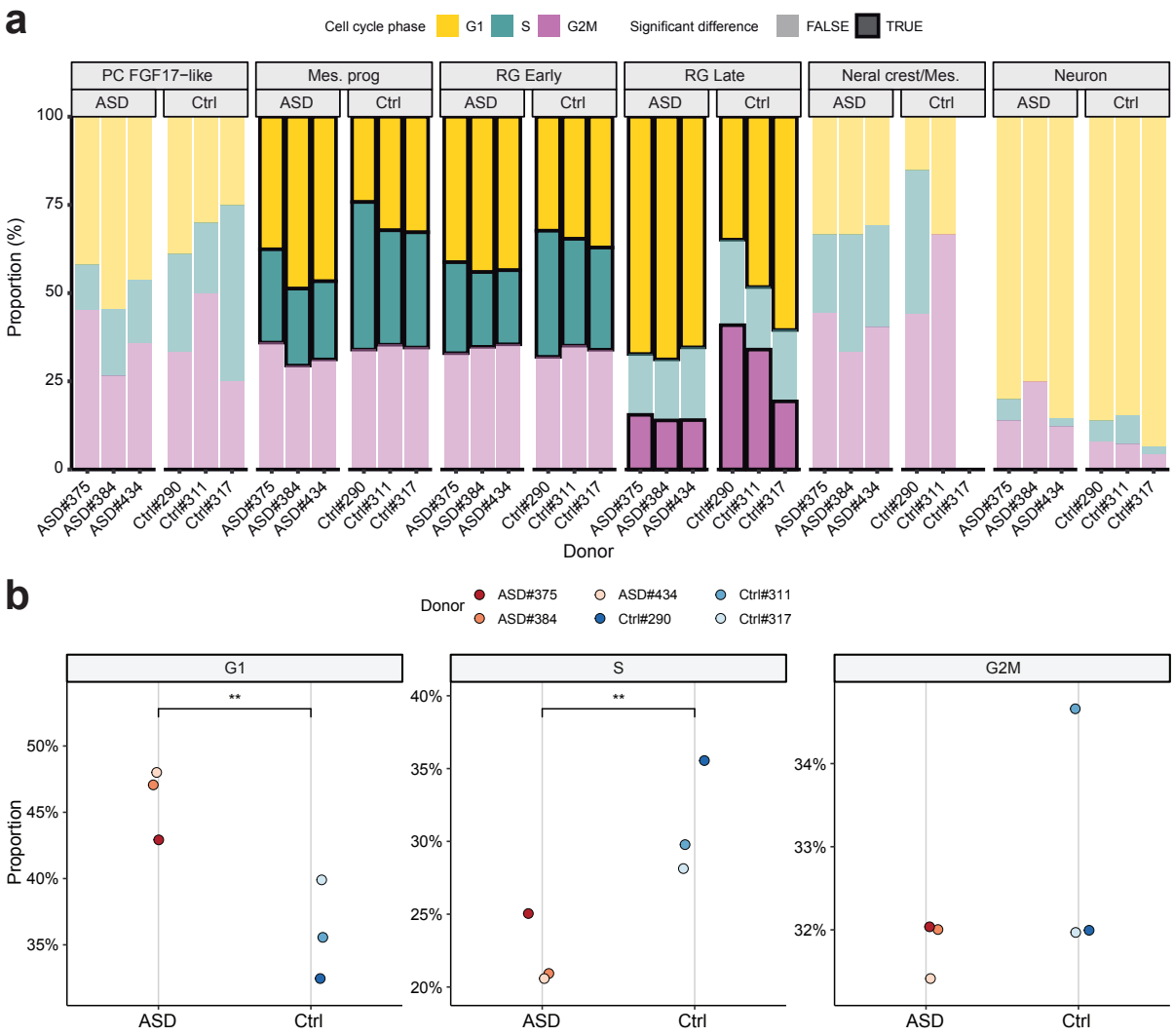
Supplementary Fig. 15. Related to Fig. 4. KO simulation in neurogenesis across species. (**a**) Comparison of KO simulation between mouse (x axis) and human (y axis) neurogenesis leveraging data from Noack et al, 2021 [63] and donors PCW20, PCW21 and PCW24 from Trevino et al, 2021 [61], respectively. The effect of the gene KO simulation on each cell identity group (RG cells and neurons) is

shown as the ratio of final over initial number of cells, in log2 scale. Only genes knocked-out in both datasets are shown. A regression model is overlaid on the plot. The genes with a positive or negative effect in both datasets, or opposite effects after KO simulation, are indicated. **b**) Comparison of KO effects in human donors. Same as in a, one-to-one comparisons in neurogenesis of three human donors. **c and d)** Fisher tests of genes showing the same direction of effect. For each comparison, the table shows the number of genes with a positive or negative effect in both datasets, and opposite effects. The odd ratio of genes with coincident effects, its confidence interval and p-value are also indicated.

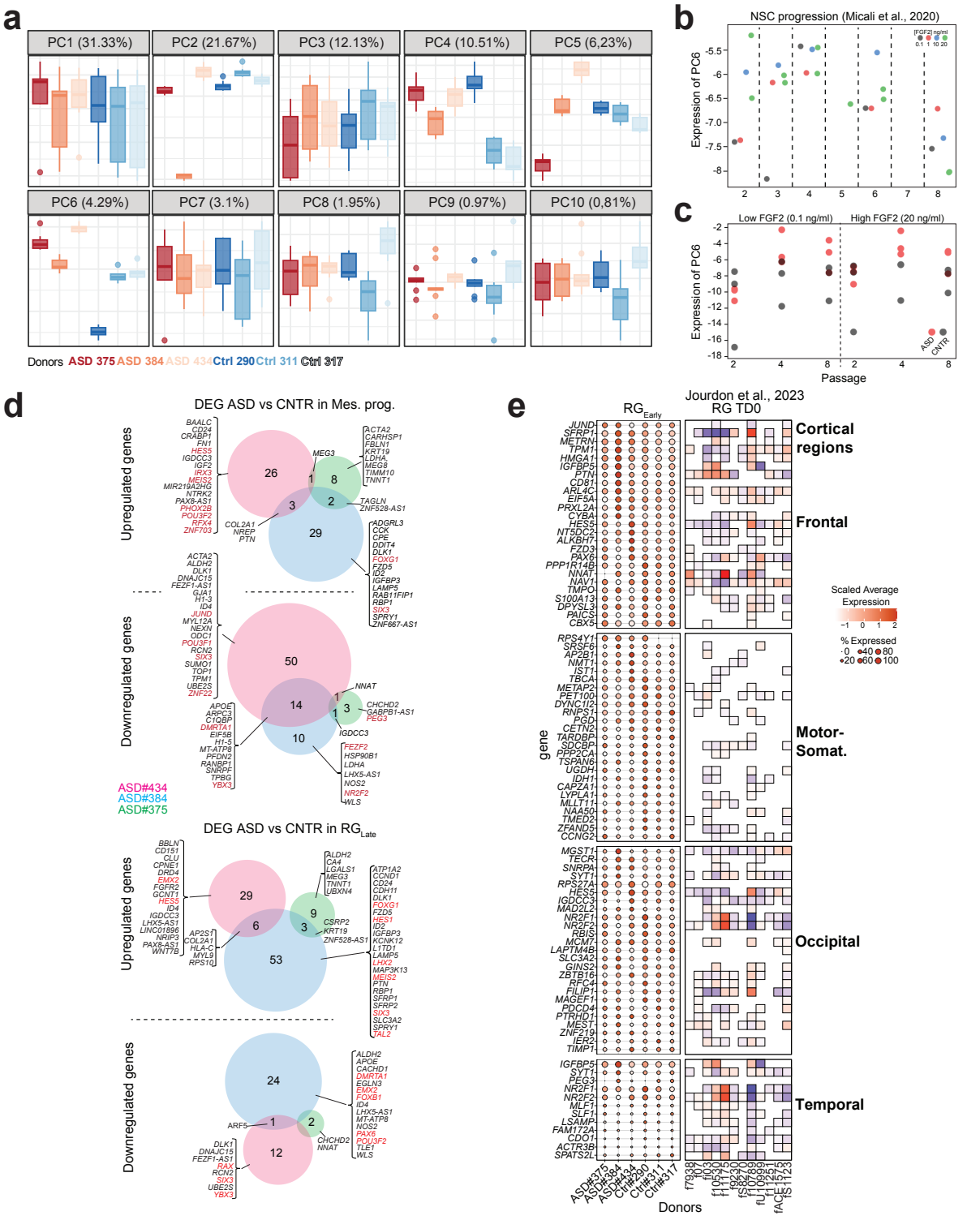


Supplementary Fig. 16. Related to Fig. 5. Quality overview of control- and ASD-derived in vitro NSC scRNA-seq data. a) UMAP showing clusters of cell types identified. **b)** Quality check of the clusters. Violin plots showing the percentage of mitochondrial counts, the number of genes detected and the number of Unique Molecular Identifiers (UMIs) per cluster. **c**) expression level (color gradient) and percentage of cells (dot size) expressing subtype markers for each cluster identified. **d**) Sankey plots showing the integration of the DIV8 scRNA-seq with macaque developing brain dataset [37]. Left, middle and right column showing macaque annotation, raw seurat clusters and final annotation in the DIV8

dataset, respectively. The top panel includes the most representative cells of the macaque dataset and all cells in the DIV8 dataset. The bottom panel includes patterning centers and the ventricular radial glial cell (vRG) subtypes of the macaque dataset, and PC FGF17-like and RG clusters in the DIV8 dataset. **e**) Bar plot showing number of cells for each cluster in each donor. **f**) Number of UMIs, total detected genes, percentage of mitochondrial and ribosomal genes expression in each cluster for each donor.



Supplementary Fig. 17. Related to Fig. 5. Cell cycle phase characterization of control- and ASD-derived cell lines. a) Fraction of cell cycle phases in every cell type in each donor. Significant differences between ASD and control cells are highlighted with black outlines. **b**) Per-donor proportion of total cells in the different cell cycle phases. Significant differences between groups are marked with "**" (adjusted p-value < 0.05).

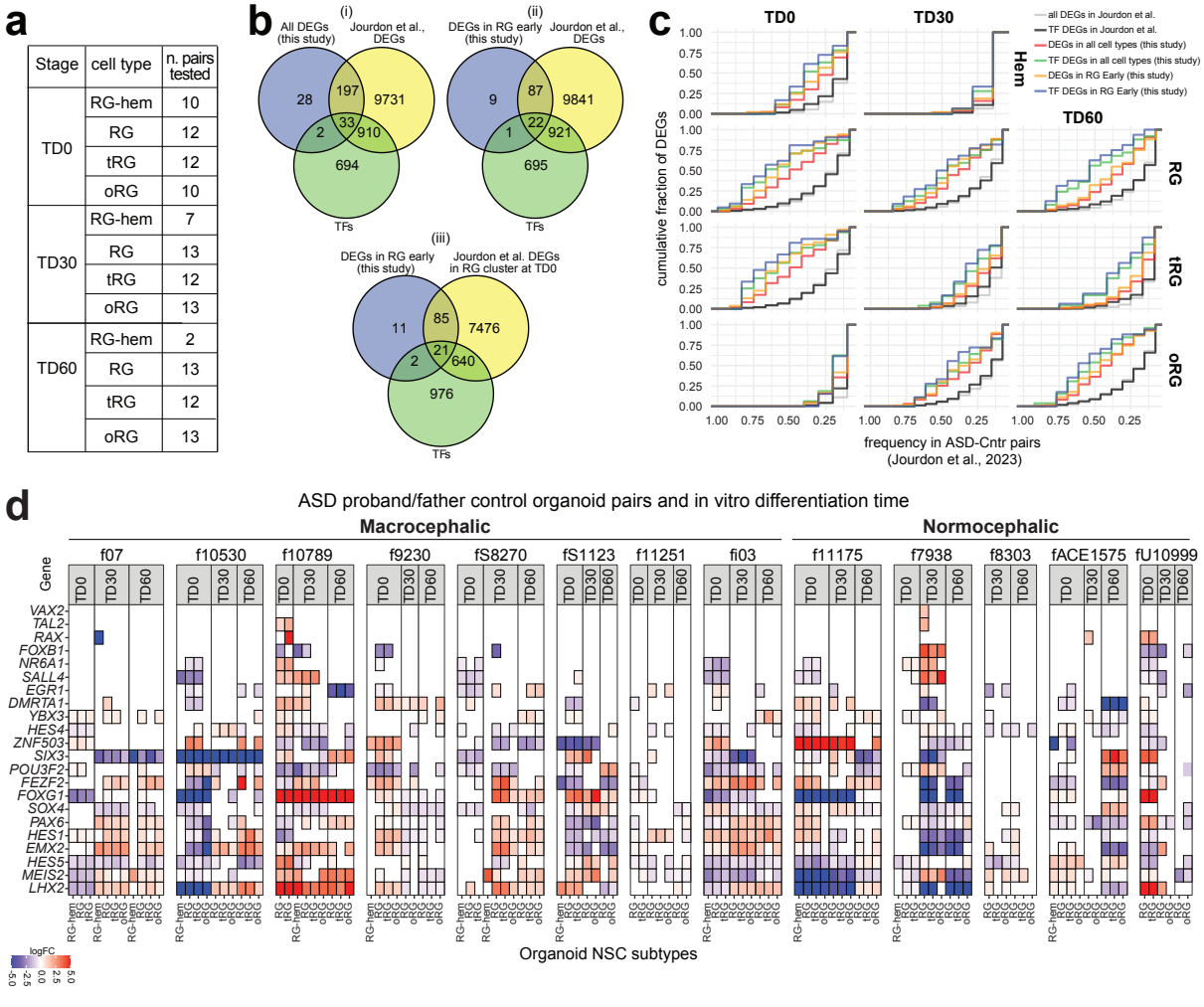


Supplementary Fig. 18. Related to Fig. 5. Differential gene expression analysis of control- and ASD- derived cell lines. a) Principal component analysis of RG_{Early} pseudobulk samples from scRNA-seq data generated from control and ASD hiPSCs at DIV 8 of neural induction. PC6 segregates control

and ASD lines. **b**) Projection of Micali et al., 2020 bulk RNA-seq from sequentially passaged human NSCs into PC6. PC6 shows highest levels in PS4 NSCs (NeMO/PCA). **c**) Projection of bulk RNA-seq data from sequentially passaged control and ASD cell lines in vitro (see Fig. 5g and Supplementary Fig. 20) into PC6. PC6, which segregates control and ASD lines in the scRNA-seq, also shows elevated levels in these same ASD lines in the bulk passaging data (NeMO/PCA). **d**) Venn diagrams of DEGs in mesenchymal (Mes.) progenitor cells and RG_{Late} in individual ASD samples versus grouped controls, evaluated in scRNA-seq data and excluding sex chromosome genes. In Mes. progenitors, 69 genes (9 TFs) up-regulated and 79 genes (9 TFs) down-regulated were found in ASD. In RG_{Late}, 100 genes (8 TFs) up-regulated and 39 genes (8 TFs) down-regulated were found in ASD (Supplementary Table 9). **e**) Expression level (color gradient) and percentage of cells (dot size) expressing cortical region genes in RG_{Early} in each NSC line (left); and corresponding differential expression across ASD-control pairs in cortical organoid RG cluster at TD0, from Jourdon et al., 2023 [34]. The genes were selected from those associated with cortical regions in progenitor cell types in Micali et al., 2023. 25 genes with the highest expression in our data that were expressed maximum in two cortical regions were considered.

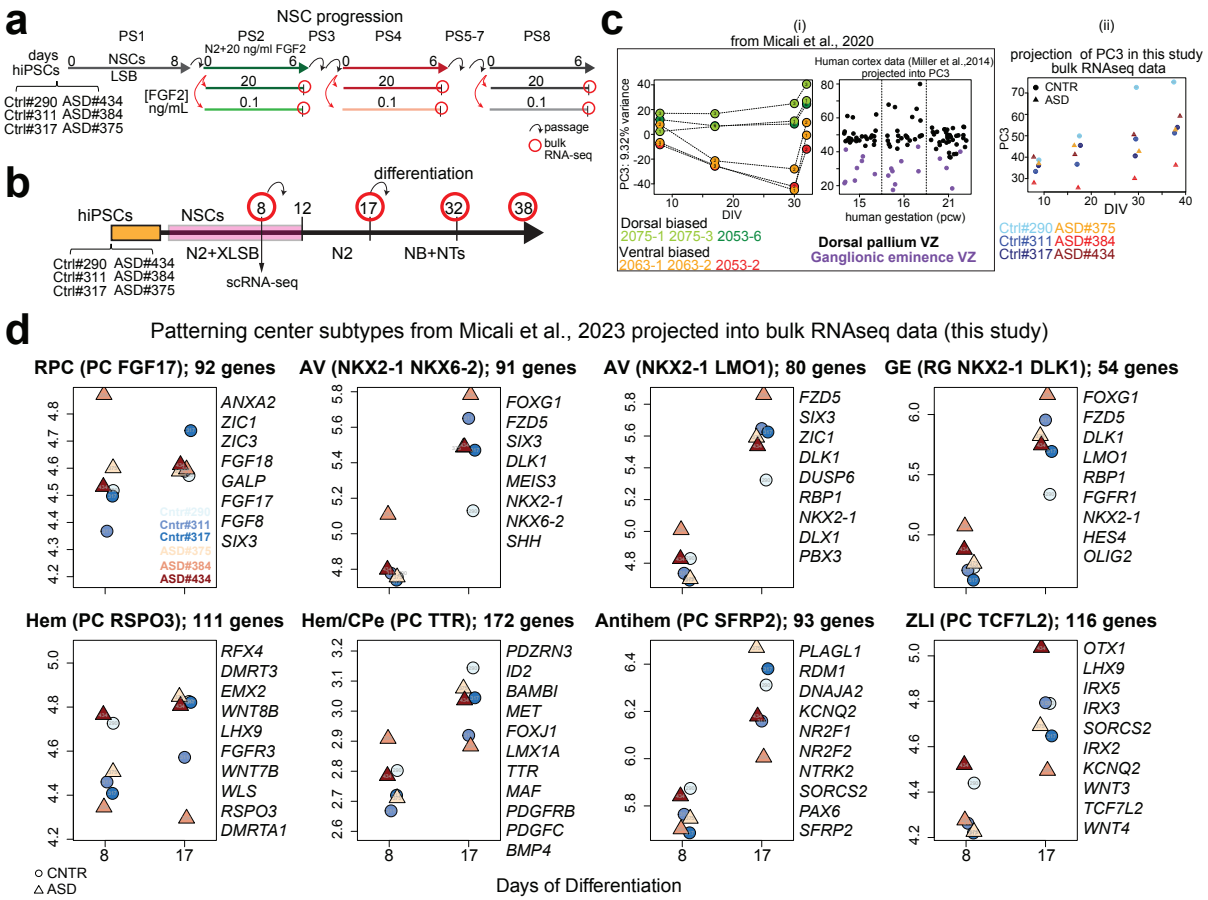
a

Stage	cell type	n. pairs tested
TD0	RG-hem	10
	RG	12
	tRG	12
	oRG	10
TD30	RG-hem	7
	RG	13
	tRG	12
	oRG	13
TD60	RG-hem	2
	RG	13
	tRG	12
	oRG	13



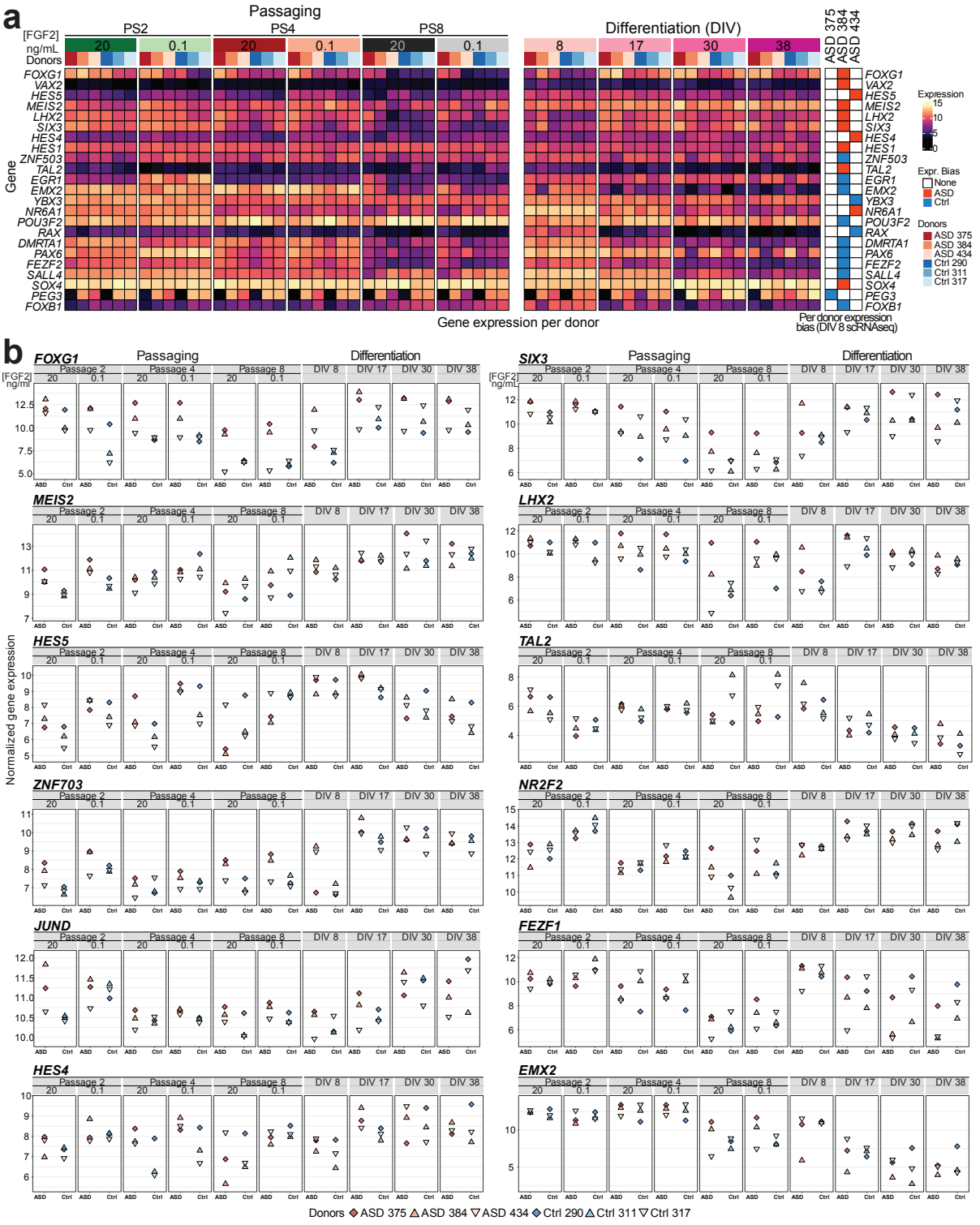
Supplementary Fig. 19. Related to Fig. 5. Intersection with differential gene expression in multiple ASD-control pair derived organoid lines. **a**) Number of ASD-control cortical organoid pairs, per NSC subtype (RG-hem, RG, tRG, oRG) and differentiation stage (terminal differentiation, TD0, TD30 and TD60), tested for differential expression in Jourdon et al., 2023 [34]. In Jourdon et al., 2023, cortical organoids were generated from iPSC lines derived from ASD probands and control fathers from 13 families (or "pairs"); scRNA-seq data were collected at 3 time points (TD0, TD30 and TD60, with TD0 corresponding to the initiation of neurogenesis) and differential expression test was performed independently in each cell type (glmGamPoi, adjusted p-value < 0.01, absoluted log2FC > 0.25). **b**) Venn diagrams showing (i) intersection between all pairwise DEGs in organoid NSC subtypes (RG-hem, RG, tRG, oRG) from Jourdon et al. dataset and ASD versus Control line DEGs in any cluster of DIV8 scRNA-seq data identified in this study (Supplementary Table 9); (ii) intersection between all pairwise

DEGs in organoid NSC subtypes (RG-hem, RG, tRG, oRG) from Jourdon et al. dataset and DEGs from ASD versus Control lines in RG_{Early} cluster from scRNA-seq at DIV8 identified in this study; (iii) intersection only for DEGs in RG_{Early} from this study and from organoid RG cluster of TD0 from Jourdon et al. In each diagram, TFs from both sets were identified using a list of all human TFs (in green) [83]. **c**) Overlap of DEGs from this study (Fig. 5c) and DEGs identified in any of the ASD-control pairs from Jourdon et al., 2023 [34]. The overlap is depicted as the cumulative fraction of DEGs identified in our study (y-axis, grouped into different DEG subsets by color) that are also found to be differentially expressed in varying frequencies among the ASD-control pairs in Jourdon et al. (x-axis). Distribution of all genes/TFs DEG in Jourdon et al. is given as reference (black/grey lines). **d**) Heatmap of pairwise DEGs between ASD proband (8 macrocephalic and 5 normocephalic)- and father control-derived organoids from Jourdon et al. for the TFs differentially expressed in RG_{Early} of the individual ASD lines identified in Fig. 5c. Note that the direction of change, although often concordant across the organoid NSC subtypes at the same stage, varied between pairs, i.e., a same gene may exhibit upregulation or downregulation across pairs. This result suggests that while perturbation of a gene is frequent across ASD lines compared to controls (as shown in S19C), the direction of change may be different among pairs.



Supplementary Fig. 20. Related to Fig. 5. Characterization of the telencephalic identity and trajectory bias of control- and ASD-derived in vitro NSCs. a) NSC progression protocol for six newly-generated hiPSC lines from 3 control (#290, #311, #317)- and 3 ASD (#375, #384, #434)-affected donors showing progression of NSCs across sequential passages (PS) and FGF2 doses, similar to the scheme described in Supplementary Fig. 2. N2 + LSB was applied at PS1 for 8 days, then hNSCs were serially passaged in N2 + 20 ng/mL FGF2 every 6 days up to PS8. In parallel, cultures of NSCs at PS2, 4, and 8 were subjected to FGF2 modulation (20 or 0.1 ng/mL) for 6 days in their terminal passage, before bulk RNA collection and RNA-seq. **b**) Neuronal differentiation protocol. The same 6 iPSC lines were passaged in mTesR + Rock inhibitor, then the day after were switched to N2-B27 + XLSB medium for 12 days to induce NSCs. Cells were terminally differentiated into neurons using Neurobasal (NB) + neurothrophins (NTs) until DIV 38 as described in Micali et al., 2020 [40]. On day 8 and 17, NSCs were passaged. Single cell suspensions for the scRNA-seq were collected at DIV8 from parallel cultures. Days for bulk RNA collection are indicated. **c**) (i) PC3 of 6 human NSC lines (2075-1 and -3; 2063-1 and

-2; 2053-6 and -2) from 3 donors showing dorsal (2075-1 and -3, 2053-6, green) or ventral (2063-1 and -2; 2053-2 orange) neuronal lineage bias, based on the projection of human developing pallium and ganglionic eminence bulk transcriptomic data [49]. These data, previously reported in Micali et al., 2020, show that hiPSC-derived NSC lines have dorsal or ventral telencephalic lineage bias even with the same neuronal differentiation protocol. Lines 2075-1 and -3 exhibit dorsal pallium bias in their differentiation trajectory; 2063-1 and -2 exhibit ventral telencephalic bias; lines 2053-6 and -2, which are 2 replicates of the same donor, show divergent lineage trajectories, i.e. dorsal (2053-6) and ventral (2053-2). (ii) To examine telencephalic differentiation bias, the bulk RNA-seq data generated from 3 control and 3 ASD-derived NSC lines, differentiated using the neuronal differentiation protocol (panel b), were projected into the Micali et al., 2020 PC3, distinguishing dorsal versus ventral NSC fates. The analysis confirmed a ventral identity for ASD#384, a dorsal identity for Cntr#290, and no clear trajectory bias for the other lines. **d**) Projection of macaque patterning centers signatures from Micali et al., 2023 [37] into the bulk RNA-seq of the control- and ASD-derived NSC lines at DIV8 and 17. The data highlight AV telencephalic features in #384 and slight posterior telencephalic organizer features in ASD#434 which express ZLI genes. RPC: rostral patterning center; AV: anteroventral; GE: ganglionic eminence; CPe: Choroid Plexus epitelium; ZLI: zona limitans intrathalamica.



Supplementary Fig. 21. Related to Fig. 5. TF expression across progression and differentiation of control- and ASD-derived NSCs. **a**) Expression of TFs differentially expressed in RG_{Early} in individual ASD samples versus grouped controls from Fig. 5c, across passages and differentiation stages. The

expression bias of these TFs in ASD/Control RG_{Early} from scRNA-seq analysis is on the right. **b)** Normalized gene expression (see Methods) for selected genes from panel (a), and selected DEGs (*ZNF703*, *JUND*, *NR2F2*, *FEZF1*) from Mes progenitors and RG_{Late} clusters, shown in Supplementary Fig. 18d, across passages and differentiation.