Stem Cell Reports, Volume *16*

Supplemental Information

NGN2 induces diverse neuron types from human pluripotency

Hsiu-Chuan Lin, Zhisong He, Sebastian Ebert, Maria Schörnig, Malgorzata Santel, Marina T. Nikolova, Anne Weigert, Wulf Hevers, Nael Nadif Kasri, Elena Taverna, J. Gray Camp, and Barbara Treutlein

SUPPLEMENTAL FIGURES

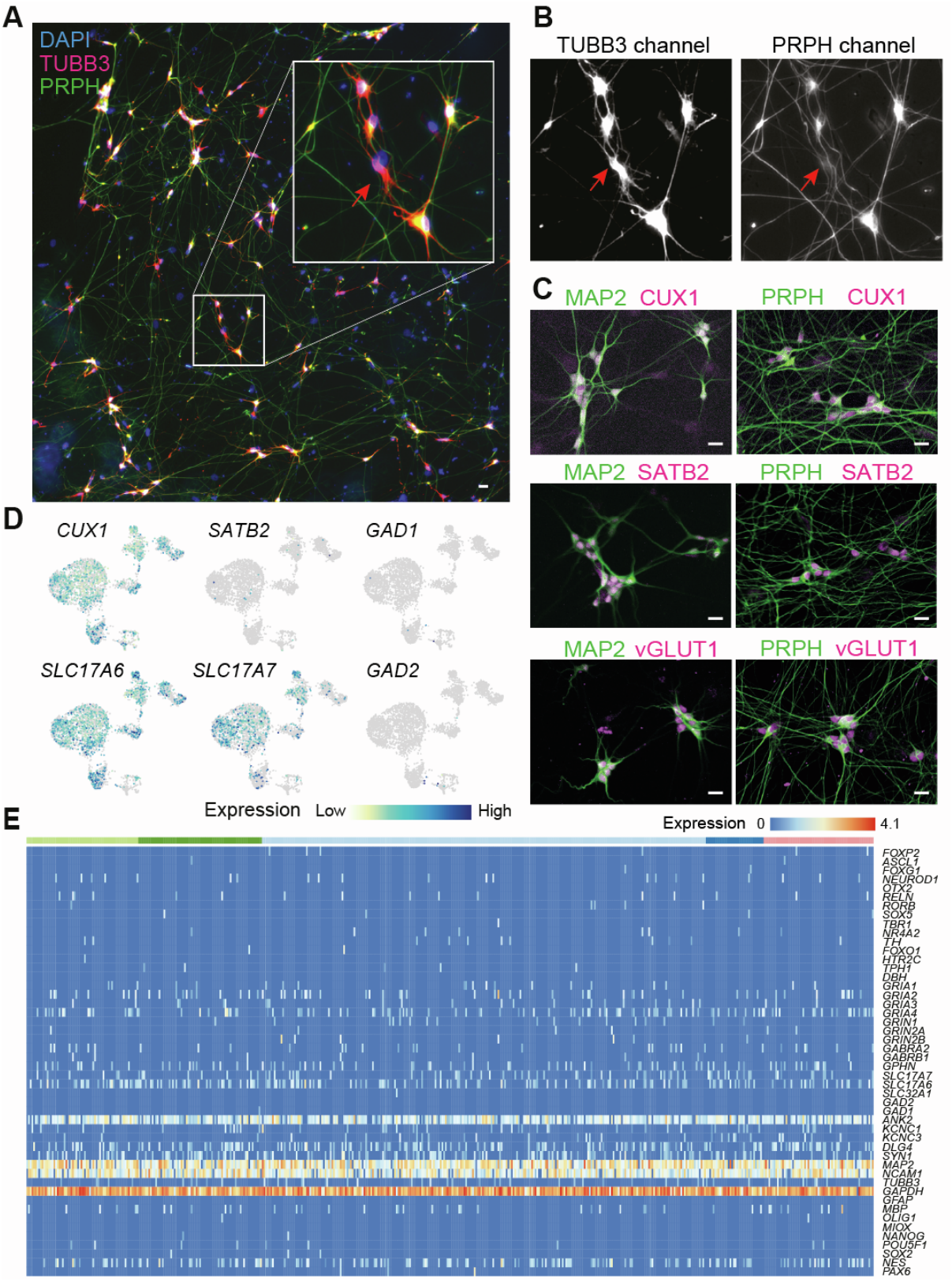
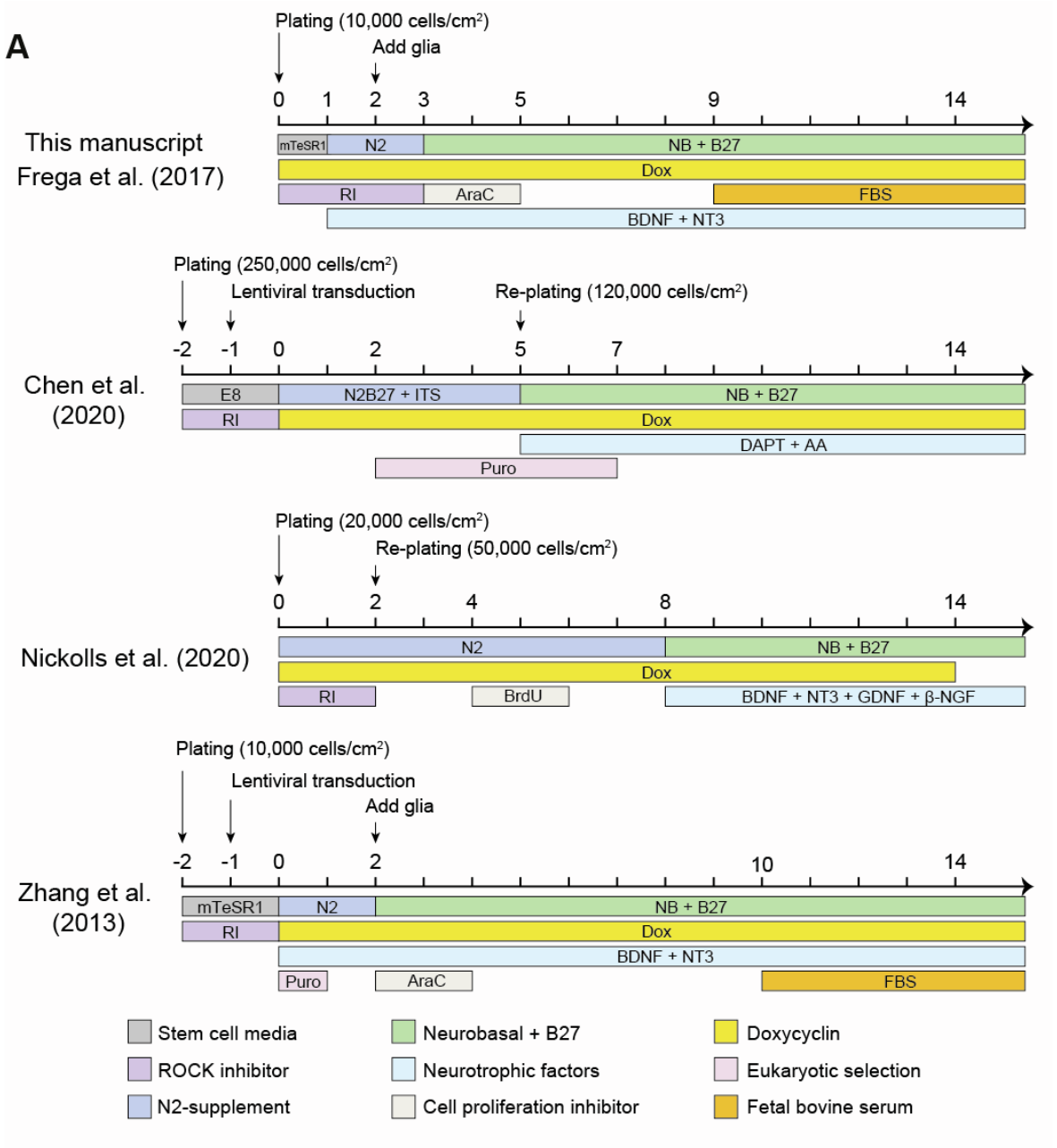


Fig S1. Characterization of *NGN2*-iNs directly programmed from iPSCs. (A) Immunofluorescence of *NGN2*-iNs at d14 with DAPI (blue), TUBB3 (red) and PRPH (green). Red arrow in the inset showed neurons with TUBB3 marker but not PRPH. Scale bar is 10

μm. (B) TUBB3 signals and PRPH signals from cells in the inset in (A). See raw images in Image S1-3. (C) Immunofluorescence of *NGN2*-iNs at d14 with MAP2, PRPH, CUX1, SATB2 and vGLUT1 (SLC17A7). Scale bars are 20 μm. (D) Feature plots of excitatory neurons, inhibitory neurons and cortical neurons in d35 *NGN2*-iNs. See also Fig. 2. (E) Expression of various markers in randomly sampled 500 single cells from the 5 neuron clusters in d35 *NGN2*-iNs.



B

	This manuscript	Frega et al. (2017)	Chen et al. (2020)	Nickolls et al. (2020)	Zhang et al. (2013)
Readout	scRNA-seq	Marker IF	Marker IF	Marker IF	Marker IF/WB/qPCR sc-qRT-PCR
NGN2-iN types	CNS + PNS (67% PRPH)	Excitatory upper layer cortical neurons	CNS neuron (10% PRPH)	Neuron with partial PRPH	Glutamatergic cortical layer 2/3 neuron

Fig S2. Comparison of different protocols of *NGN2*-iN generation and corresponding neuron cell type assessments. (A) Different protocols used for generating iNs through *NGN2* overexpression. Stable cell lines expressing dox-inducible *NGN2* are generated either

before (this manuscript, (1, 2) or during induction (3, 4). An eukaryotic selection step is included if *NGN2* is transduced after plating cells for *NGN2* induction by doxycycline. Most protocols have cells seeded in stem cell media containing ROCK inhibitors (RI), followed by a switch to N2-supplement and neurobasal media (NB) containing B27 for various durations. Cell proliferation inhibitors are added in many protocols to remove cells that are not successfully reprogrammed. Different neurotrophic factors are added to the culture at different days after *NGN2* induction. Glial cells are supplemented in some protocols to support the growth and function of *NGN2*-iNs. Fetal bovine serum (FBS) is included in corresponding protocols to support the growth of glial cells. Puro: puromycin; BrdU: Bromodeoxyuridine; BDNF: Brain-derived neurotrophic factor; NT3: Neurotrophin-3; AA: Ascorbic acid; GDNF: glial cell line-derived neurotrophic factor; NGF: Nerve growth factor; ITS: Insulin-Transferrin-Selenium. (B) Comparison of readouts of *NGN2*-iN characterization and identified neuron types in different studies. IF: Immunofluorescence; WB: western blotting.

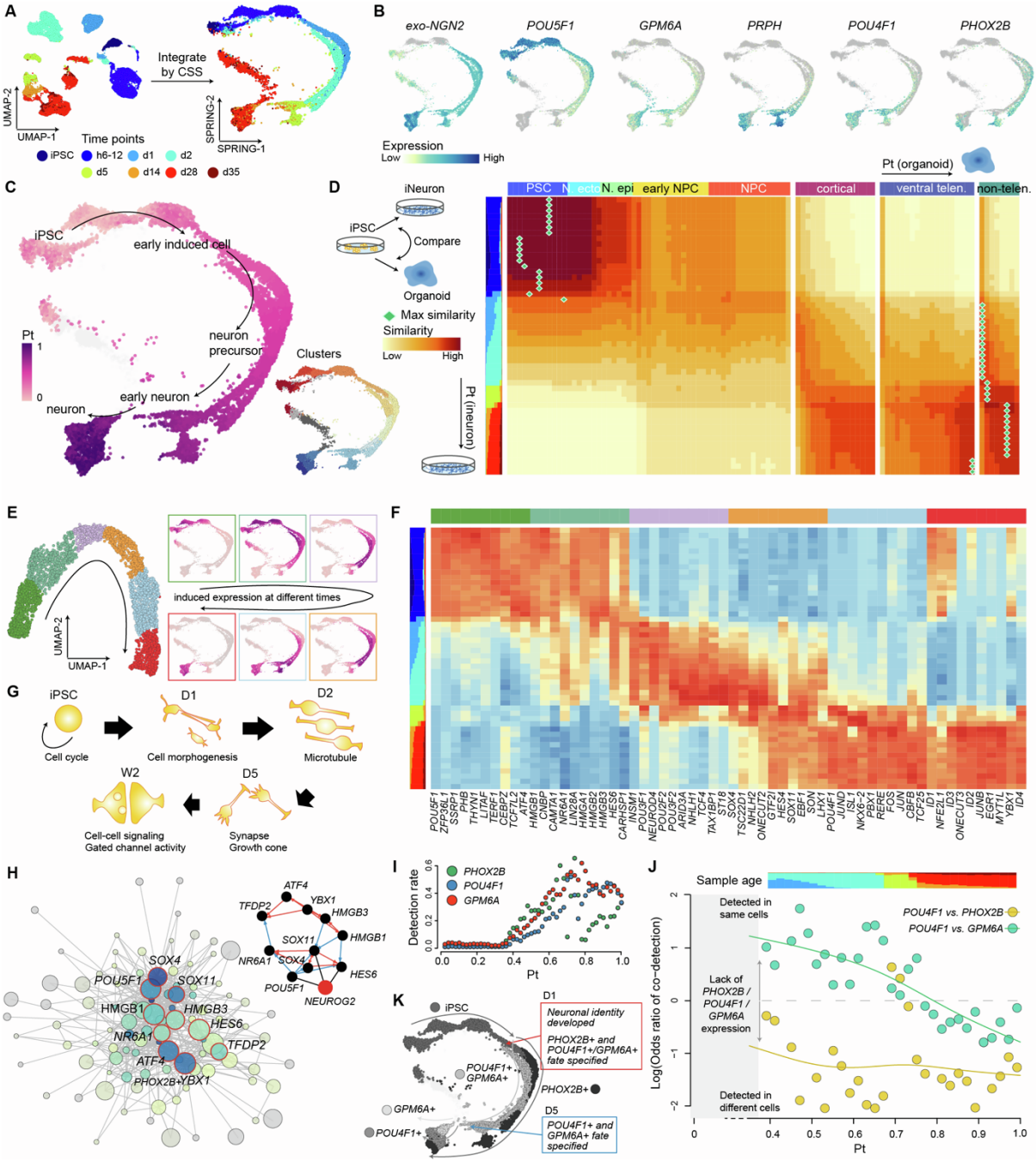


Fig S3. Analysis of *NGN2*-mediated direct reprogramming path from iPSC to induced neurons. (A) *NGN2*-iNs directly reprogrammed from 409B2 iPSCs were harvested at different time points after induced expression of *NGN2* by Dox treatment and analyzed by scRNA-seq. scRNA-seq data was directly analyzed by UMAP clustering or combined by CSS-based integration with differentiation trajectory reconstructed using SPRING. UMAP and SPRING plots were colored by time points. h6-12 is the 1:1 mixture of cells harvested after 6h and 12h Dox induction, respectively. (B) SPRING plots colored by marker gene expression. (C) Each cell in the *NGN2*-iN reprogramming path was ordered in pseudotime based on transcriptome similarities. The derived pseudotime was then colored on the SPRING plot. Cell clusters along the developmental trajectories were also color labelled. (D) Developmental trajectories from iPSCs to *NGN2*-iNs and brain organoids (5) were aligned and compared. (E) 3,231 genes

among the 17,198 genes detected in more than 50 cells along the full time course of *NGN2*-iN development had significant change in expression. These genes were grouped into 6 clusters with their expression colored on the SPRING plot. (F) Heatmap of transcription factors (TFs) in each cluster among the differentially expressed genes identified in (E). (G) Schematic representation of the biological processes identified in the functional enrichment analysis from differentially expressed genes in (E). (H) Gene regulatory networks for TFs with pseudotime-dependent expression. Highlighted are the top-10 TFs with highest expression and most connection with other TFs. (I) Detection of cells with *PHOX2B*, *POU4F1* or *GPM6A* expression along the pseudotime of *NGN2*-iN development. (J) Coexpression of *POU4F1/PHOX2B* and *POU4F1/GPM6A* along the pseudotime of *NGN2*-iN development. (K) Fate specification of the three *NGN2*-iN neural subtypes along the time course of development. The SPRING plot was colored by cell identity.

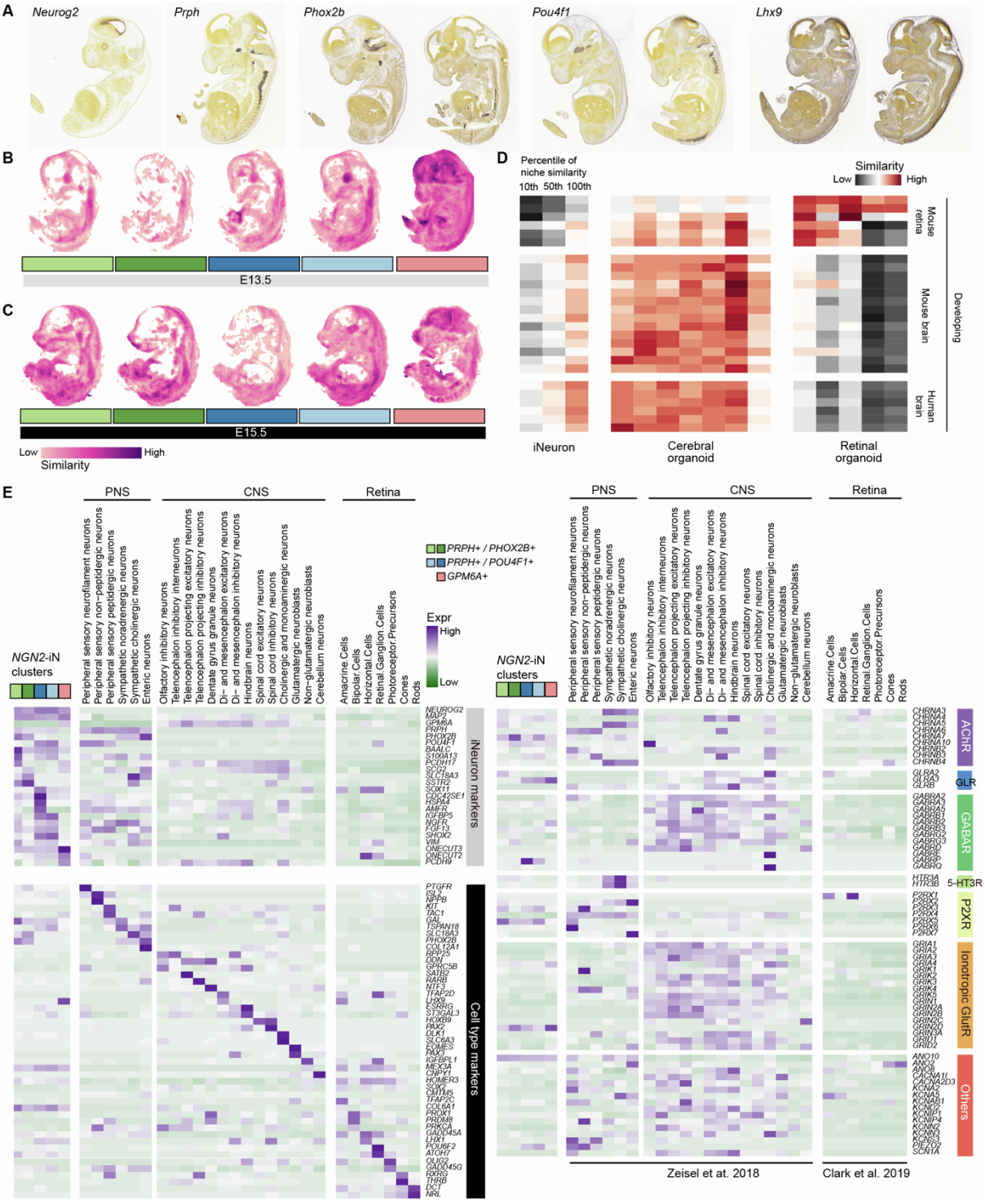


Fig S4. Benchmark of *NGN2*-iNs through transcriptomic comparison to primary references in mouse. (A) Spatial expression patterns of selected markers via in situ hybridization (6) in the E13.5 mouse embryos from the Allen Developing Mouse Brain Atlas. (B-C) Transcriptomic similarity between each of the five *NGN2*-iN populations and E13.5 (B) and E15.5 (C) mouse embryos. It shows the maximum similarity projections across sagittal sections in the mouse embryos from the Allen Developing Mouse Brain Atlas. (D) Transcriptome similarities between iPSC-derived neurons and different sources of primary CNS neuron subtypes including developing mouse retina (7), developing mouse brain (8) and developing human brain (9). Similarities are represented as scaled spearman correlation

between expression profiles. (E) Average expression of various marker genes *NGN2*-iN clusters and primary mouse PNS and CNS neuron subtypes (left), as well as various genes encoding ion channels (right), in *NGN2*-iN clusters as well as primary mouse neuron subtypes.

SUPPLEMENTAL TABLES

Table S1: GO terms enriched in the course of *NGN2*-iN development
Table S2: DE genes of the 6 clusters from d35 *NGN2*-iNs

SUPPLEMENTAL IMAGES

Image S1: Merged image of d14 *NGN2*-iNs immunofluorescence in Fig. S1A.

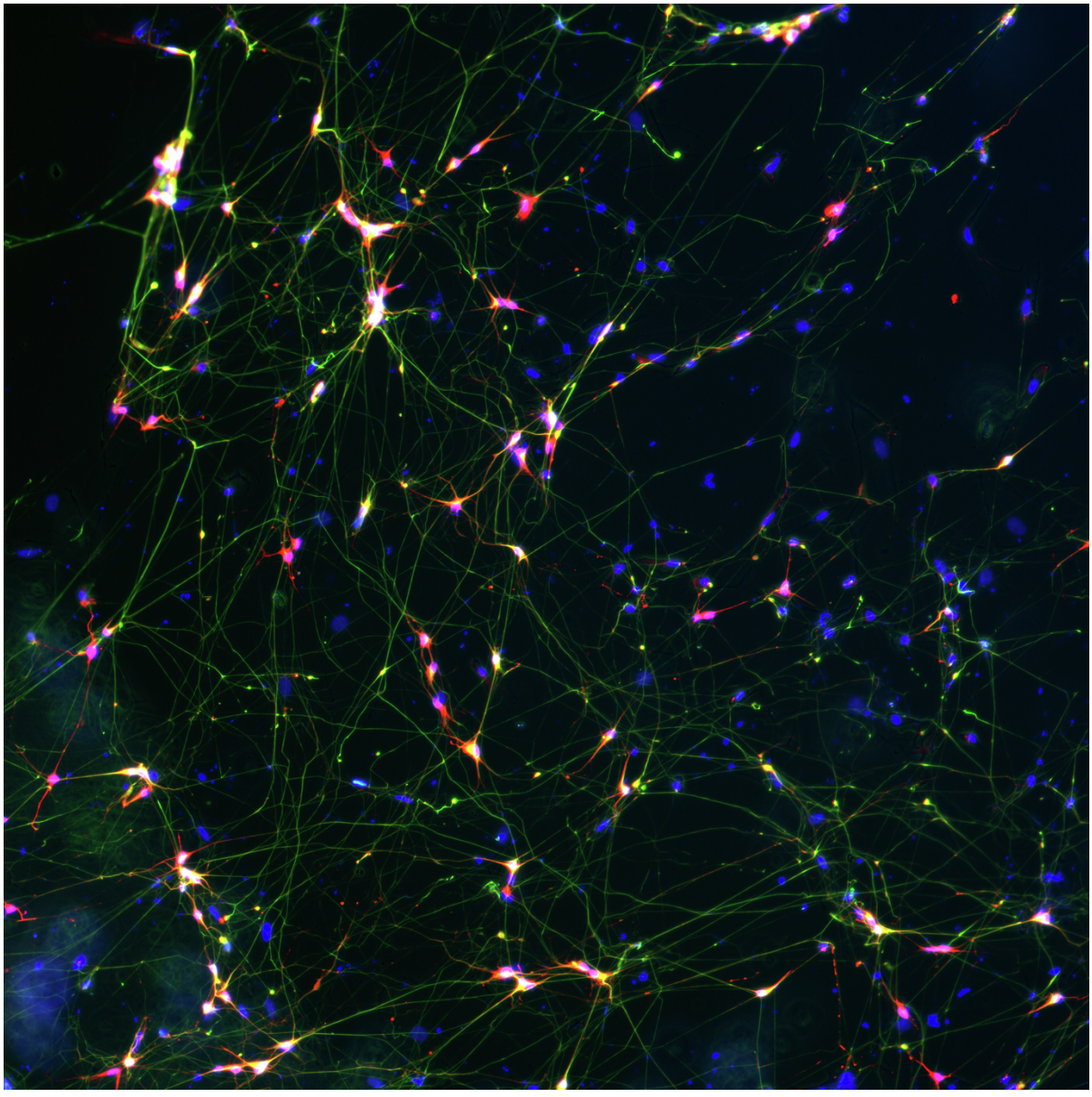


Image S2: Raw image of TUBB3 channel for d14 *NGN2*-iNs immunofluorescence in Fig. S1B.

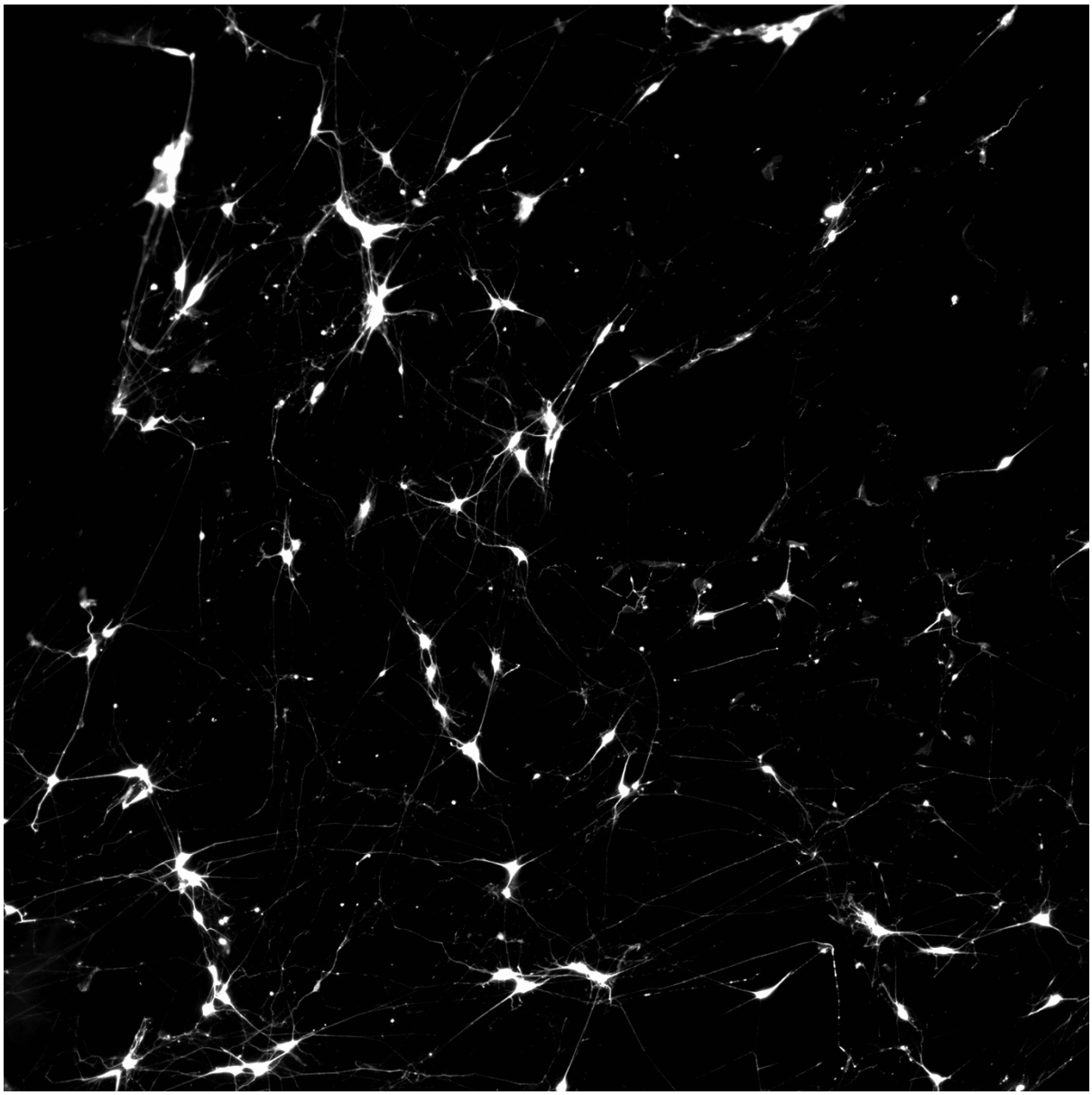
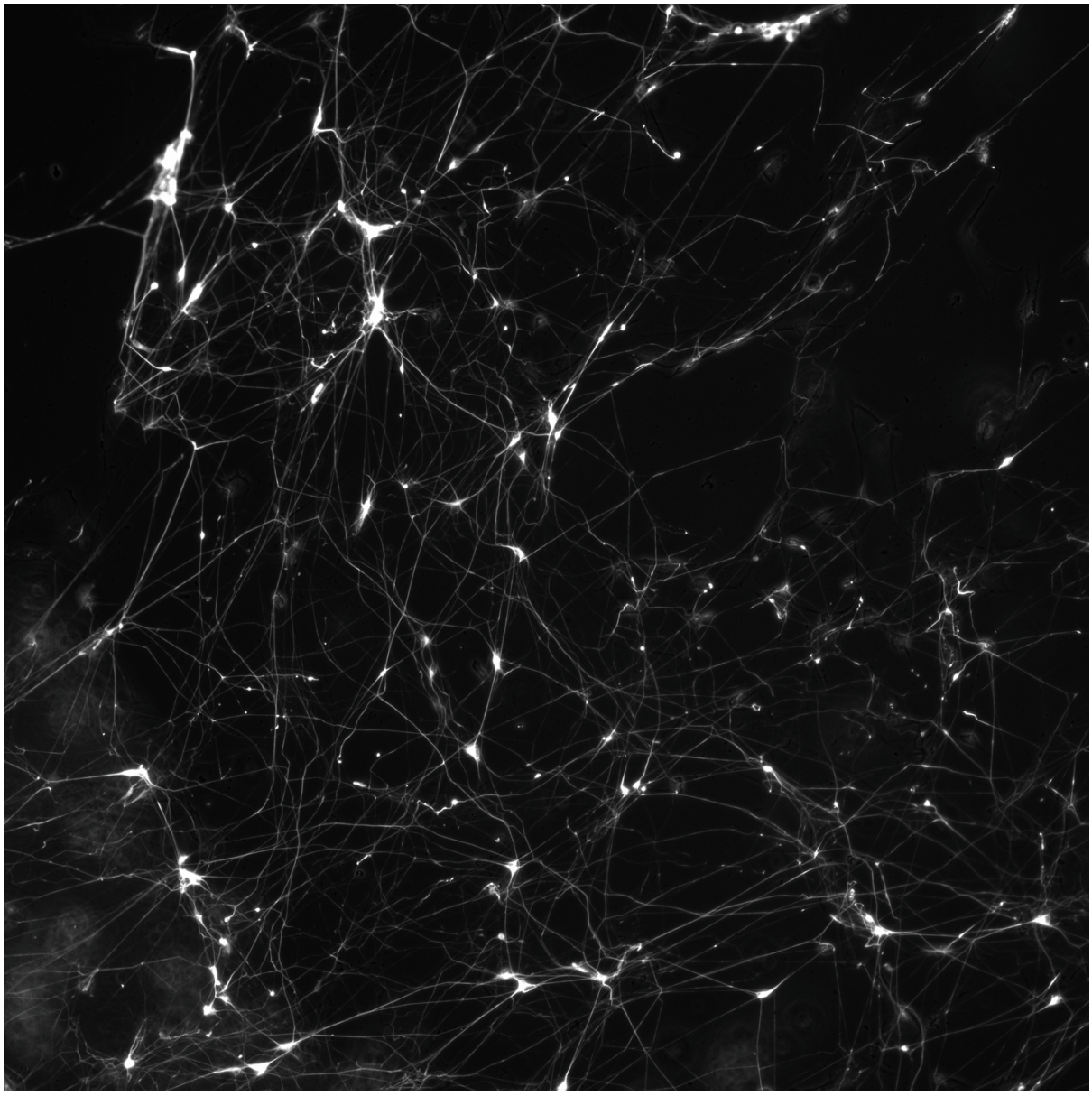


Image S3: Raw image of PRPH channel for d14 *NGN2*-iNs immunofluorescence in Fig. S1B.



SUPPLEMENTAL METHODS

Generation of rtTA/*NGN2*-iPSCs

rtTA/*NGN2* double positive stem cell lines were generated as previously described (Frega et al., 2017). Briefly, 409B2 or Sc102a1 stem cells were transduced with lentiviruses carrying rtTA (reverse tetracycline-controlled trans-activator)-Neo cassettes and *NGN2*-Puro cassettes. Transduced cells were selected for stable double integration using G418 followed by Puromycin treatment. rtTA/*NGN2*-409B2 cells were either propagated as a non-clonal pool or passed selection by picking a colony derived from a limiting dilution of a single-cell suspension to form a clonal line (monoclonal 409B2). The resulting 409B2, Sc102a1 and monoclonal-409B2 rtTA/*NGN2*-derivatives were subsequently used for differentiating *NGN2*-iNeurons.

Differentiation of *NGN2*-iNs from iPSCs

rtTA-*NGN2*-iPSCs were differentiated into *NGN2*-iNs as previously described (Frega et al., 2017). To differentiate rtTA-*NGN2*-iPSCs into *NGN2*-iNs, rtTA-*NGN2*-iPSCs were seeded as single cells on plates coated with 50 µg/mL poly-L-ornithine and 10 µg/mL Laminin, with *NGN2* expression induced using 4 µg/mL Doxycycline. One day after seeding, the medium was changed to DMEM/F-12 with 4 µg/ml Doxycycline, 1:100 N-2 supplement, 1:100 MEM non-essential amino acid solution (NEAA), 10 ng/µL human BDNF, 10 ng/µL human NT3, 1:1000 ROCK inhibitor and laminin 0.2 µg/mL. Two days after seeding, rat astrocytes were added to the culture. On the third day after seeding, the medium was changed to Neurobasal medium with 4 µg/ml Doxycycline, 1:50 B-27 supplement, 1:100 glutamax supplement, 10 ng/µl human BDNF, 10 ng/µl human NT3 and 2 µM Cytosine-D-arabinofuranoside (Ara-C). On day 5, 7 and 9 after seeding, 50% of the medium was exchanged with Neurobasal-based medium similar as day 3 only without the Ara-C. After day 9, 50% of the medium was changed every second day using the Neurobasal-based medium containing 2.5% FCS. *NGN2*-iNs were kept for the indicated time and dissociated for scRNA-seq.

Immunostaining of iNeurons

iNeurons, grown on coated coverslips (acid-treated, Kleinfeld Labortechnik), were fixed with 2% PFA (prepared freshly, pre-warmed to 37 °C) for 8 min. The coverslips were washed three times with 2 mL PBS and stored in PBS at 4 °C. iNs were permeabilized with 0.05% Triton X-100 in PBS for 10 min and quenched in 1 ml 0.2 M Glycine buffer for 30 min at room temperature. The coverslips were washed with PBS and incubated with 100 µl of the respective primary antibody diluted in IF-buffer (20 mM phosphate buffer containing 0.2% gelatin and 0.05% Triton X-100). The coverslips were washed with IF-buffer five times for 5 min and then incubated with 100 µl of the respective secondary antibody diluted in IF-buffer containing DAPI (dilution 1:1000). After five washes with IF-buffer for 5 min and three quick washes with PBS, the coverslips were mounted on slides with Mowiol 4-88 and stored at 4 °C.

iNs were acquired as confocal Z-stacks. iNs were imaged using an Olympus FV1200 confocal microscope equipped with a 10-x objective or 60-x oil immersion objective (optical section thickness: 1.028 µm, distance between consecutive optical sections: 0.4 µm). We acquired three-dimensional (Z-stack) scans with a number of z-sections ranging from 5 to 30 depending on the cell. Single tiles were 1024 x 1024 pixels.

Antibodies to the following proteins were purchased from the indicated vendors: TUBB3 (MMS-435P, Biolegend), MAP2 (PA1-10005, Invitrogen), PRPH (ab39374, Abcam), CUX1 (MA5-31415, Invitrogen), SATB2 (ab92446, Abcam) and vGLUT1 (AB5905, Sigma).

Dissociation of *NGN2*-iNs for scRNA-seq

To prepare single-cell suspensions for scRNA-seq, *NGN2*-iNs were selectively dissociated from co-cultured rat astrocytes using a mild dissociation procedure. In this procedure, cultured *NGN2*-iNs were washed with DPBS and incubated at 37 °C with a 1% Accutase in EDTA solution for 5 minutes. The digestion was then neutralized using Neurobasal medium with 10% FCS. The media was gently flushed onto the cells using a wide bore tip. The progress of selective neuron detachment was constantly monitored under the microscope. Enriched neural cell suspensions were collected in 15 ml tube and centrifuged for 5 minutes at 300 g. The cell pellets were resuspended and further dissociated in undiluted Accutase under 5-10 minutes incubation at 37 °C. The dissociated cells were neutralized, centrifuged, resuspended in the residual 100 µl medium, triturated 20 times, strained using a 30 um pre-separation filter and quantified using a Countess automated cell counter (Invitrogen). Cell suspensions were then diluted to 500-1000 cells/mL for subsequent scRNA-seq experiment.

Preprocessing of raw sequencing data

We used Cell Ranger, the suggested analytic pipeline by 10x Genomics, to demultiplex raw base call files to FASTQ files and align reads to the mouse-human dual species genome and transcriptome (mm10-hg38, provided by 10x Genomics) with the default alignment parameters. Demultiplexing of human and mouse cells was done based on their read mappability to the two genomes by Cell Ranger. Only human cells were kept for the following analysis. Cell Ranger was used again to map the human reads to the human-only genome and transcriptome (hg38, provided by 10x Genomics). Pooled samples of different human lines (409b2 and Sc102a1) were demultiplexed using demuxlet (Kang et al., 2018), based on the genotyping information of the two lines.

To quantify expression levels of the exogenous *NGN2* in cells, the unmapped reads after mapping to the human genome were extracted and compared to the sequences of the inserted *NGN2*. In brief, for each read a series of substrings were obtained with a sliding window size of 40nt and step size of 10nt. The read was considered as a hit of exogenous *NGN2* transcript if at least 50% of the read substrings were perfectly matched to the inserted *NGN2* transcript sequence.

Analysis of the time course scRNA-seq data of *NGN2*-iN induction

Seurat (v3.1) was applied to the scRNA-seq data for further preprocessing. Quality control (QC) was done by excluding cells with more than 6000 genes detected, as well as those with mitochondrial transcript proportion larger than 10%. A subset of cells from the iPSC samples and those at day 1, 5, 14 and 29 since doxycycline treatment, which have detected gene number between 1500 and 5500 as well as mitochondrial transcript proportion less than 7% was further extracted. Highly variable genes were identified using the ‘vst’ method, and louvain clustering (resolution = 1) was used to do clustering to the cells. A k-nearest neighbor (kNN) graph was generated and visualized by using SPRING (Darmanis et al., 2015) (https://github.com/AllonKleinLab/SPRING).

To observe the molecular changes along the whole time course of *NGN2*-iN induction process from iPSC more comprehensively, we applied Cluster Similarity Spectrum (CSS) (He et al., 2020) to all the cells passing the initial QC, to reduce batch effect between samples of different time points. In brief, 5,000 highly variable genes were firstly identified using the 'vst' method. Their expression levels across cells were z-transformed, followed by principal component analysis (PCA) for dimension reduction. Next, cells from each sample were subset, and louvain clustering (resolution = 0.6) was applied based on the pre-calculated top-20 PCs. Average expression of the pre-defined highly variable genes was calculated for each cluster in each sample. Spearman correlation coefficient was calculated between every cell and every cluster in different samples. For each cell, its correlations with different clusters of each sample were z-transformed. Its z-transformed similarities to clusters of different samples were then concatenated as the final CSS representation. Louvain clustering (resolution = 0.5) was applied to the CSS representation.

Next, pseudocells were constructed, using the similar method as described (Kanton et al., 2019). In brief, cells from the same sample and assigned to the same cluster were grouped into kNN territories, each of which was centered by one randomly selected cell with its kNNs (k=20) surrounded. CSS representation of each pseudocell was calculated as the average CSS representation of its included cells. A constrained kNN network (k=20) of pseudocells was constructed, to only consider pseudocells from the same or nearby time points when screening for nearest neighbours.The kNN network was then visualized by SPRING (Weinreb et al., 2018). The same projection method based on the support vector regression model as described (Kanton et al., 2019) was applied to project the single-cell data to the cell embedding space that was defined for pseudocells.

To better illustrate temporal molecular changes along the *NGN2*-iN induction process, pseudotime analysis was applied to the CSS-represented time course scRNA-seq data using the diffusion map algorithm (implemented in the R package destiny, k=50). Cells in two side-branch clusters (Cl-8, Cl-15) were excluded from the pseudotime analysis. The ranks of DC1 were used as the pseudotimes. An F-test-based test was applied to the expression profile along pseudotimes to identify genes with pseudotime-dependent expression changes. In brief, for each gene, a natural spline linear regression model (df = 5) was constructed for cells along the pseudotime course. The residuals of the variation, which cannot be explained by the model, were then compared to the total variation of the gene by an F-test (Bonferroni corrected $P<0.01$). A kNN network (k=50) was constructed for the identified genes, based on the pairwise Pearson correlation distances. Jaccard indices between genes were calculated based on the resulting kNN network to weight the network, and edges with Jaccard indices less than 1/15 were pruned. UMAP and louvain clustering (resolution = 0.5) was applied to the weighted gene kNN network to construct the gene embeddings and identify gene clusters.

To infer the gene regulatory network (GRN) that contributed to the *NGN2*-iN induction initialization, cells at the samples at induction time point h6-12 and d1 were subset. The pySCENIC pipeline (Aibar et al., 2017) was applied with default settings except for allowing inhibitory regulons to be identified. Transcription factor network was derived from the resulting regulons and the igraph package in R was used to generate the layout of the network for visualization.

Analysis of the scRNA-seq data of *NGN2*-iN of multiple cell lines at week 5

The scRNA-seq data of *NGN2*-iN cells at day 35 of samples from the 409b2 and Sc102a1 human iPSC lines, including those from the monoclonal 409b2 iPSC cells, were log-normalized using Seurat (v3.1). 3,000 highly variable genes were identified for cells from 409b2, 409b2 monoclonal and Sc102a1 iPSCs separately. The data was integrated with Seurat. UMAP and louvain clustering (resolution = 0.3) was applied to the integrated data. During this process, the first 20 components of canonical component analysis (CCA) and PCA were considered. Two resulting clusters (Cl-0 and Cl-1) were merged as no strong signature unique to one of the two clusters was observed. Cluster markers were identified using the R package presto, defined as genes that being detected in at least 50% of cells in the cluster and at least 20% higher than in other cells, with AUC>0.7, expression fold change > 1.2 and Benjamini-Hochberg (BH) corrected two-sided Wilcoxon's rank sum test P<0.01.

To characterize each *NGN2*-iN cluster, we retrieved the scRNA-seq data of three mouse reference data sets: the adolescent mouse nervous system atlas (Zeisel et al., 2018) which includes both the central nervous system (CNS) and peripheral nervous system (PNS) in adolescent mice; the developing mouse brain (La Manno et al., 2020); and the developing mouse retina (Clark et al., 2019). Top-5000 highly variable genes in the adolescent mouse nervous system and the developing mouse brain were identified using the vst method in Seurat and intersected, resulting in 1952 shared variable genes with one-to-one human orthologs according to Ensembl (v84) annotation. Average expression profiles were calculated for each of the annotated neuron subtypes in the three atlases, as well as the 5 *NGN2*-iN clusters, across the shared variable genes. Pearson correlation coefficient was then calculated between average expression profiles of *NGN2*-iN clusters and selected mouse reference neuron subtypes, including 14 subtypes in developing brain, the developing spinal cord motor neuron, 6 subtypes in developing retina, and 6 subtypes in adolescent PNS.

To further benchmark the *NGN2*-iN with other iPSC-derived neurons, we additionally retrieved the scRNA-seq data of human cerebral organoids (Kanton et al., 2019) and retinal organoids (Cowan et al., 2020). Similarly, average expression profiles of the seven neuron subtypes in the cerebral organoids, as well as those of the five neuron subtypes in the retinal organoids, were calculated across the shared variable genes obtained above. The resulting profiles were correlated with the mouse references similar to the *NGN2*-iN clusters.

To summarize the CNS/PNS signatures in the *NGN2*-iN cells, a transcriptome deconvolution based on quadratic programming was used to dissect the CNS and PNS neuronal signatures. In brief, differential expression analysis using the presto package was applied to compare transcriptome of CNS and PNS neurons in the mouse nervous system, based on the adolescent mouse brain atlas mentioned above. Differentially expressed genes (DEGs) were identified as genes with BH corrected P<0.01 and expression fold change > 1.2. The average expression levels of those CNS-PNS DEGs were calculated for each annotated cell type in the mouse nervous system (annotation layer 4). Considering the transcriptome of each *NGN2*-iN cell as a mixture of transcriptomes of different cell types, quadratic programing was then used to calculate the relative contribution of each cell type by solving the constrained linear least-square problem:
Here f is the vector of relative cell type contribution, C is the gene expression matrix of the CNS-PNS DEGs in annotated mouse cell types, and x is the expression level of these genes in a cell. Afterwards, the contributions of all CNS neuron subtypes were summed up as the CNS signature score of a cell.

To compare *NGN2*-iN cells with primary fetal and adult neurons, the Fluidigm C1-based single cell RNA-seq data of human fetal and adult cortex was retrieved from SRA (SRP057196)(Darmanis et al.). Average expression levels of each annotated major cell type, including replicating cells (neural progenitor cells) and quiescent cells (immature neurons) in fetal samples, as well as cell types in adult samples (neurons, astrocytes, oligodendrocytes, oligodendrocyte precursor cells, microglia, endothelial cells). Pearson correlations were calculated across expression of highly variable genes in the *NGN2*-iN d35 data set, between each *NGN2*-iN cell and each annotated cortical major cell type.

Analysis of the scRNA-seq data of varied doxycycline treatment duration

The *NGN2*-iN scRNA-seq data with varied doxycycline treatment durations was preprocessed using Seurat similar as above. As quality control, only cells with detected gene numbers between 1370 and 8000 remained in the following analysis. The lower bound of the detected gene number was the minimum threshold to exclude cells with detected gene numbers being at the lower peak of the observed bimodal distribution of detected gene numbers in cells. CSS representation was then calculated to integrate cells from different *NGN2*-iN samples with varied doxycycline treatment duration (cluster resolution = 1). UMAP and louvain clustering (resolution = 0.3) was applied to the CSS representation of cells.

SUPPLEMENTAL REFERENCES

1. Frega M, van Gestel SH, Linda K, van der Raadt J, Keller J, Van Rhijn JR, et al. Rapid Neuronal Differentiation of Induced Pluripotent Stem Cells for Measuring Network Activity on Micro-electrode Arrays. J Vis Exp. 2017(119).
2. Nickolls AR, Lee MM, Espinoza DF, Szczot M, Lam RM, Wang Q, et al. Transcriptional Programming of Human Mechanosensory Neuron Subtypes from Pluripotent Stem Cells. Cell Rep. 2020;30(3):932-46 e7.
3. Zhang Y, Pak C, Han Y, Ahlenius H, Zhang Z, Chanda S, et al. Rapid single-step induction of functional neurons from human pluripotent stem cells. Neuron. 2013;78(5):785-98.
4. Chen M, Maimaitili M, Habekost M, Gill KP, Mermet-Joret N, Nabavi S, et al. Rapid generation of regionally specified CNS neurons by sequential patterning and conversion of human induced pluripotent stem cells. Stem Cell Res. 2020;48:101945.
5. Kanton S, Boyle MJ, He Z, Santel M, Weigert A, Sanchis-Calleja F, et al. Organoid single-cell genomic atlas uncovers human-specific features of brain development. Nature. 2019;574(7778):418-22.
6. Ravasi T, Suzuki H, Cannistraci CV, Katayama S, Bajic VB, Tan K, et al. An atlas of combinatorial transcriptional regulation in mouse and man. Cell. 2010;140(5):744-52.
7. Clark BS, Stein-O'Brien GL, Shiau F, Cannon GH, Davis-Marcisak E, Sherman T, et al. Single-Cell RNA-Seq Analysis of Retinal Development Identifies NFI Factors as Regulating Mitotic Exit and Late-Born Cell Specification. Neuron. 2019;102(6):1111-26 e5.

8. La Manno, G., Siletti, K., Furlan, A., Gyllborg, D., Vinsland, E., Langseth, C.M., Khven, I., Johnsson, A., Nilsson, M., Lönnerberg, P., et al. (2020). Molecular architecture of the developing mouse brain. bioRxiv.
9. Cao J, O'Day DR, Pliner HA, Kingsley PD, Deng M, Daza RM, et al. A human cell atlas of fetal gene expression. Science. 2020;370(6518).
10. Cowan, C.S., Renner, M., Gross-Scherf, B., Goldblum, D., Munz, M., Krol, J., Szikra, T., Papasaikas, P., Cuttat, R., Waldt, A.*, et al.* (2019). Cell types of the human retina and its organoids at single-cell resolution: developmental convergence, transcriptomic identity, and disease map. bioRxiv.
11. Kang, H.M., Subramaniam, M., Targ, S., Nguyen, M., Maliskova, L., McCarthy, E., Wan, E., Wong, S., Byrnes, L., Lanata, C.M.*, et al.* (2018). Multiplexed droplet single-cell RNA-sequencing using natural genetic variation. Nat Biotechnol *36*, 89-94.
12. Weinreb, C., Wolock, S., and Klein, A.M. (2018). SPRING: a kinetic interface for visualizing high dimensional single-cell expression data. Bioinformatics *34*, 1246-1248.