Stem Cell Reports
Report
ISSCR
OPEN ACCESS

NGN2 induces diverse neuron types from human pluripotency

Hsiu-Chuan Lin,[1,2,9] Zhisong He,[3,9] Sebastian Ebert,[3,4,9] Maria Schörnig,[4] Malgorzata Santel,[3] Marina T. Nikolova,[3] Anne Weigert,[4] Wulf Hevers,[4] Nael Nadif Kasri,[5,6] Elena Taverna,[4,8] J. Gray Camp,[1,2,4,*] and Barbara Treutlein[3,4,7,*]

[1]Institute of Molecular and Clinical Ophthalmology Basel, Basel, Switzerland
[2]Department of Ophthalmology, University of Basel, Basel, Switzerland
[3]Department of Biosystems Science and Engineering, ETH Zürich, Basel, Switzerland
[4]Max Planck Institute for Evolutionary Anthropology, Leipzig, Germany
[5]Department of Human Genetics, Donders Institute for Brain, Cognition, and Behaviour, Radboudumc, Nijmegen, the Netherlands
[6]Department of Cognitive Neuroscience, Donders Institute for Brain, Cognition, and Behaviour, Radboudumc, Nijmegen, the Netherlands
[7]Max Planck Institute of Molecular Cell Biology and Genetics, Dresden, Germany
[8]Human Technopole, Milan, Italy
[9]These authors contributed equally
*Correspondence: jarrettgrayson.camp@unibas.ch (J.G.C.), barbara.treutlein@bsse.ethz.ch (B.T.)
https://doi.org/10.1016/j.stemcr.2021.07.006

SUMMARY

Human neurons engineered from induced pluripotent stem cells (iPSCs) through neurogenin 2 (*NGN2*) overexpression are widely used to study neuronal differentiation mechanisms and to model neurological diseases. However, the differentiation paths and heterogeneity of emerged neurons have not been fully explored. Here, we used single-cell transcriptomics to dissect the cell states that emerge during *NGN2* overexpression across a time course from pluripotency to neuron functional maturation. We find a substantial molecular heterogeneity in the neuron types generated, with at least two populations that express genes associated with neurons of the peripheral nervous system. Neuron heterogeneity is observed across multiple iPSC clones and lines from different individuals. We find that neuron fate acquisition is sensitive to *NGN2* expression level and the duration of *NGN2*-forced expression. Our data reveal that *NGN2* dosage can regulate neuron fate acquisition, and that *NGN2*-iN heterogeneity can confound results that are sensitive to neuron type.

INTRODUCTION

Human cell types engineered from induced pluripotent stem cells (iPSCs) through transcription factor overexpression are widely used to study the mechanisms controlling cell fate differentiation, to model human diseases, and to identify potential therapies (Guo and Morris, 2017). Human neurons can be generated through the forced expression of the transcription factor neurogenin 2 (*NGN2*) with high efficiency and reproducibility (Zhang et al., 2013). These *NGN2*-induced neurons (*NGN2*-iNs) functionally mature into morphologically complex and electrophysiological active neurons after approximately 3–4 weeks of co-culture with astrocytes. The *NGN2*-iN system has been used extensively to understand neuron development and model disease (Lin et al., 2018). However, the characterization of *NGN2*-iNs so far has generally been limited to functional assays, biomarker expression, and bulk transcriptomics. There is a lack of comprehensive transcriptomic comparison with primary neuron subtypes and it is unclear whether any off-target fate emerges during the differentiation process. Single-cell sequencing methods provide powerful resolution into the heterogeneity of directed differentiation culture systems (Biddy et al., 2018, 2018, 2018; Camp et al., 2018; Karow et al., 2018). Previously, we have used single-cell mRNA sequencing (scRNA-seq) to dissect the differentiation path from mouse embryonic fibroblasts and human pericytes to neurons and identified previously undescribed heterogeneity generated by the overexpression of the pioneer factor *ASCL1* (Karow et al., 2018; Treutlein et al., 2016). Here, we set out to characterize *NGN2*-iNeuron heterogeneity, identify the cell states that are generated during differentiation, and analyze the dynamics of the differentiation process using scRNA-seq.

RESULTS

Heterogeneity of *NGN2*-induced neurons dissected by scRNA-seq

We generated a stable iPSC line expressing *NGN2* that can be induced by doxycycline (Dox) and drives the differentiation toward iNeurons (Zhang et al., 2013). We then performed scRNA-seq (10× Genomics) at multiple time points during directed differentiation (Figure 1A). After filtering the data of astrocytes, multiplets, and cells with insufficient unique molecular identifiers (UMIs), a total of 6,764 cells (day 0 [d0], 1,412 cells; d1, 2,688 cells; d5, 524 cells; d14, 1,515 cells; d28, 625 cells) were included in the analysis. We combined all time course data and reconstructed the differentiation path of *NGN2*-iNs (Figures 1B–1D). Surprisingly, we found at least four transcriptionally distinct cell populations at the d28 time point. One population is marked by *DCN/COL5A1* and we interpret this cluster as

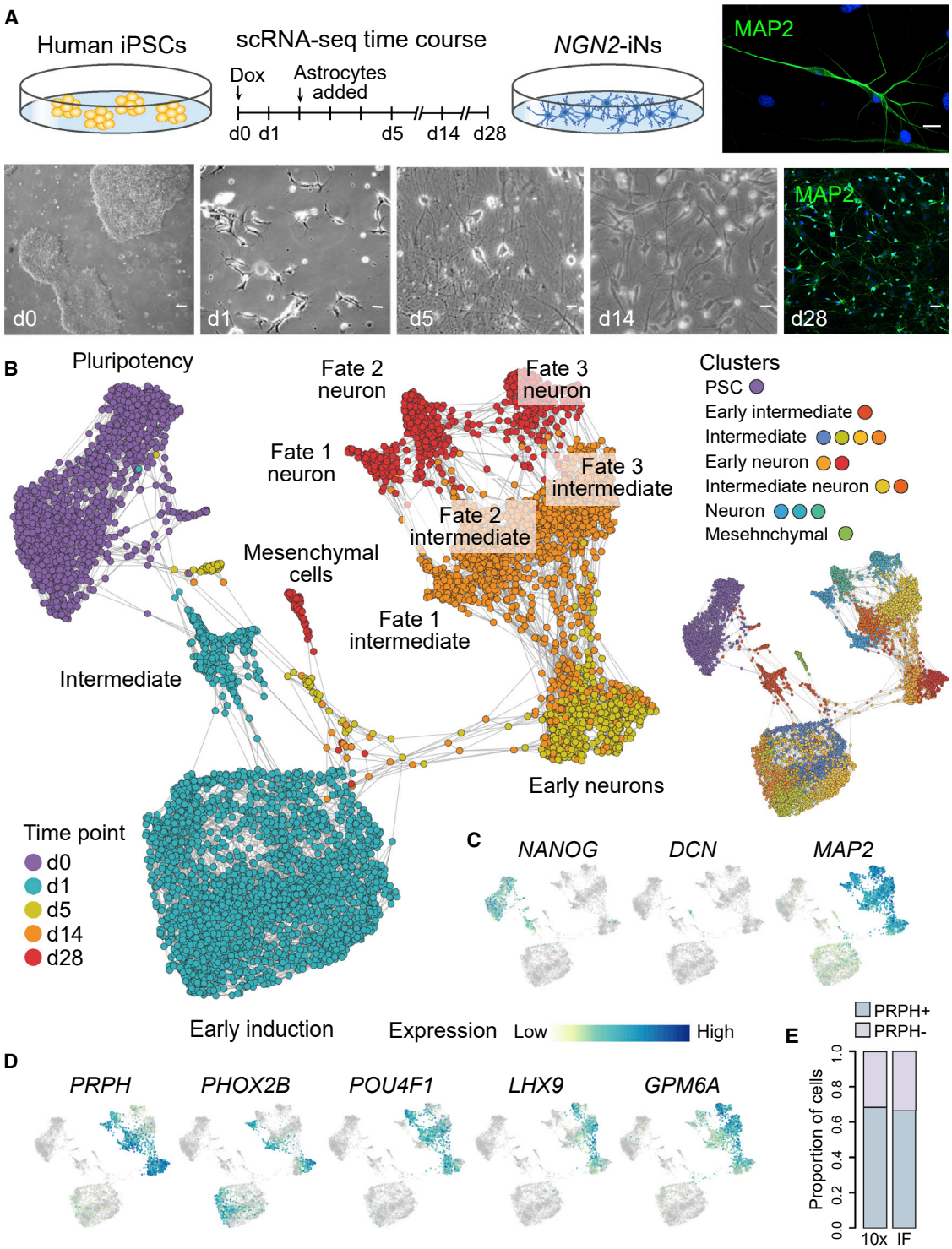


Figure 1. Diverse subpopulations emerge during *NGN2*-directed neuron differentiation into iNs from human iPSCs

(A) Schematic of scRNA-seq time course experiment and representative images from human iPSCs differentiating into *NGN2*-iNs. Cells were analyzed with scRNA-seq (10× Genomics) at multiple time points during differentiation. Immunohistochemical staining of *NGN2*-iNs at d28 with MAP2 (green) and DAPI (blue). Scale bars, 10 μm.

(B) SPRING embedding shows the developmental relationships of 409B2-derived *NGN2*-iNs with cells colored by time points (left) or cluster (right).

(legend continued on next page)

an off-target mesenchymal population (Figure 1C). In addition, we observed three different neuronal clusters that express high levels of pan-neuronal genes (*MAP2*, *NCAM1*) yet are molecularly distinct (Figure 1C). Two clusters have high expression of *PRPH*, an intermediate neurofilament that is highly expressed in neurons of the peripheral nervous system and some central nervous system regions that have neural projections toward peripheral structures (Yuan et al., 2012). These two *PRPH*+ clusters segregate into a *PHOX2B*+ cluster and a *POU4F1*+ cluster (Figure 1D). The other neuronal cluster is marked by *GPM6A* expression, which is expressed throughout both the central nervous system (CNS) and the spinal cord during mouse development (Figure 1D) (Diez-Roux et al., 2011).

We characterized the neural identity and presence of molecular heterogeneity in our *NGN2*-iN culture using immunofluorescence of TUBB3 and PRPH (Figures S1A and S1B). The percentage of PRPH+ cells was quantified to compose 67% of neural cells, comparable with the percentage estimated by scRNA-seq (Figures 1E and S1B). We examined the presence of common makers that were used to characterize *NGN2*-iNs and how they overlap with *PRPH* expression (Figures S1C–S1E). Most *NGN2*-iNs and PRPH+ cells express *CUX1* and *VGlut1*, but not *GAD1/2*, supporting their cortical excitatory feature as reported previously (Zhang et al., 2013). However, unlike *PRPH* and other identified cluster markers, common neural markers are not able to resolve the heterogeneity in our dataset. We noticed that the percentage of PRPH+ cells in our dataset is higher than previous reports (Chen et al., 2020; Nickolls et al., 2020; Schörnig et al., 2021). This can be due to differences in protocol, the particular readout of neural identity used in the previous reports, or thresholds for assigning positive staining from immunohistochemistry (Figure S2). Together, these data suggest that *NGN2*-iNs generated from our protocol are comparable with other published reports of induced neurons resulting from *NGN2* overexpression, and differences in the method of readout (selected markers versus whole transcriptome) can influence the interpretations of heterogeneity.

Transcriptome trajectory analysis along the path of *NGN2*-iN development

We further analyzed *NGN2*-iN developmental trajectories after additional integration and clustering of all time course data (Figures S3A and S3B). *NGN2* induction resulted in major gene expression changes early on in programming (d0, d1, and d5), likely driven by the immediate downstream targets of *NGN2*. Based on the observation of rapid transition from iPSCs to cells committed to a neuronal fate (Figures 1B and S3B), we hypothesized that directed differentiation bypasses early transitional states that are usually observed *in vivo* to reach neuronal states. To test the hypothesis, we ordered *NGN2*-iNs in pseudotime based on transcriptome similarities and compared the resulting trajectories with development of neurons in brain organoids (Figures S3C and S3D) (Kanton et al., 2019). We observed an escape from the early developmental stages from pluripotency directly into neural precursor stages, skipping multiple intermediate stages, including neuroectoderm and neuroepithelium induction, supporting a more direct differentiation model (Figure S3D).

To investigate the molecular events underlying the dramatic developmental changes, we identified 3,231 genes with significant expression changes along the course of *NGN2*-iN development and segregated them into six clusters with their expression peak at different stages (Figure S3E). We performed functional enrichment analysis on these differentially expressed (DE) genes and recovered gene ontology terms related to neural development (Figure S3F; Table S1). We cross-referenced DE genes with annotated transcription factors (TFs) (Hu et al., 2019) to identify potential drivers of gene expression changes during *NGN2*-iN development (Figure S1G). We focused on TFs that changed from 6 to 12 h to d1 after Dox induction and constructed a gene-regulatory network (Aibar et al., 2017), incorporating transcription factor binding site prediction in promoters with TF-target co-expression (Figure S3H). *NGN2* was predicted to connect with some TFs with the highest centrality in the constructed regulatory network, including *POU5F1*, *HES6*, and *SOX11*, supporting its role in driving direct reprogramming from iPSCs to induced neurons.

While neural induction begins before d1, the *NGN2*-iN heterogeneity emerges later as the expression of *NGN2*-iN subtype markers were detected after d1 (pseudotime Pt ~0.4) (Figures S3B and S3I). Interestingly, we found that *PHOX2B* and *POU4F1* had divergent expression from the beginning of their activation, while *POU4F1* and *GPM6A* bifurcated later at d5 (pseudotime Pt ~0.75) (Figures S3J and S3K). *PRPH*, on the other hand, was detected only after the *PHOX2B* and *POU4F1* bifurcation, suggesting that it was activated independently in *PHOX2B*- and *POU4F1*-expressing cells (Figure S3I).

(C) Expression feature plots stem cell, mesenchymal, and neural marker genes.
(D) Expression feature plots of *NGN2*-iN clusters.
(E) Proportion of *PRPH*+ and *PRPH*− cells quantified using scRNA-seq (10×) or immunofluorescence.
See also Figures S1A and S1B.

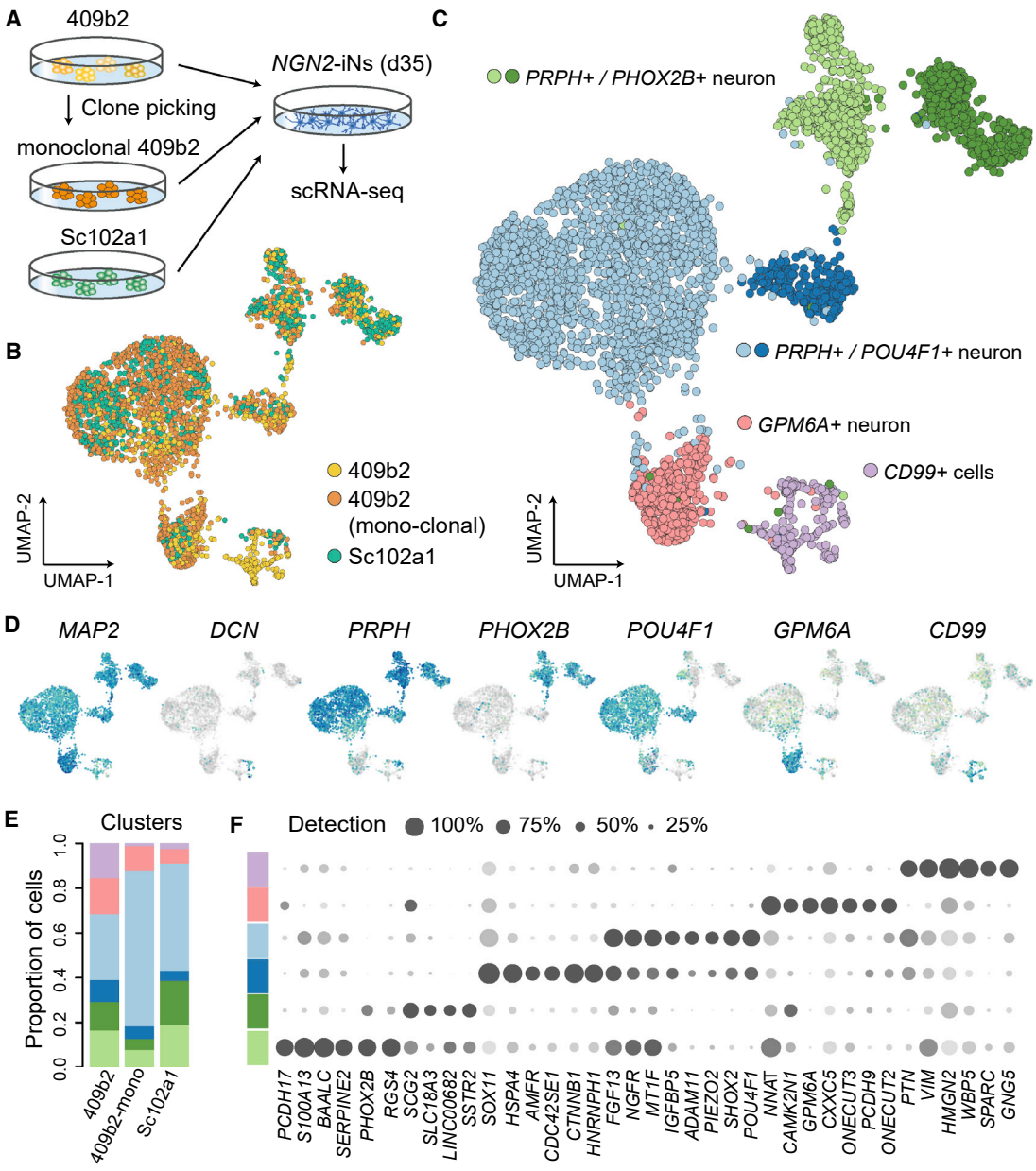


Figure 2. *NGN2*-iN neuron diversity is recapitulated in multiple iPSC lines

(A–C) (A) scRNA-seq was performed on d35 *NGN2*-iNs from polyclonal 409B2, monoclonal 409B2, and polyclonal Sc102a1 iPSCs. (B and C) UMAP embedding of Seurat 3.0 integrated scRNA-seq data, with cells colored by cell source (B) or cluster annotated by marker genes (C).

(D) Feature plots showing the expression of marker genes.

(E) Stacked bar plot showing proportions of clusters in each sample.

(F) Dot plot of marker gene expression patterns and detection rates across clusters.

NGN2-iN heterogeneity is commonly detected

We next determined if the *NGN2*-iN heterogeneity results from heterogeneous iPSC populations used to induce iNs, or if the heterogeneity was specific to the particular iPSC line. We established a single iPSC clone (409B2 monoclonal) from the parent 409B2 line and additionally generated a polyclonal *NGN2*-inducible line from another individual (Sc102a1). We induced these lines and analyzed the resulting transcriptomes at d35 of differentiation (Figure 2A). We found that significant heterogeneity is still observed in an integrated analysis of the scRNA-seq data, with six molecularly distinct clusters containing cells

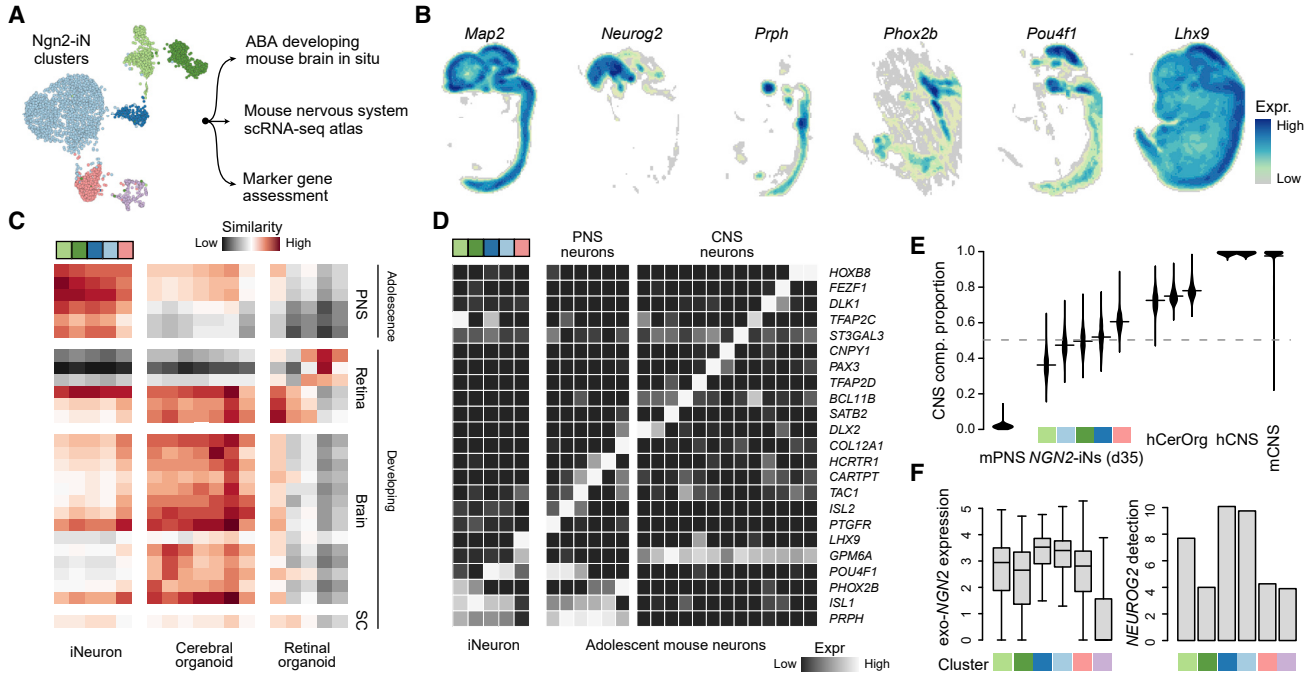


Figure 3. Molecular signatures of *NGN2*-iN compared with primary neuronal cell types mouse reference atlases

(A) *NGN2*-iN subpopulation signatures were compared with diverse reference atlases.

(B) Spatial expression patterns of selected markers as maximum intensity projections across sagittal sections in the embryonic day 13.5 mouse brain from the Allen Developing Mouse Brain Atlas.

(C) Transcriptomic similarities between *NGN2*-iNs, other iPSC-derived neurons, and primary neurons represented as Pearson correlations between expression profiles. SC, spinal cord.

(D) Average expression of various marker genes of primary neuron subtypes in *NGN2*-iN clusters and primary mouse PNS and CNS neuron subtypes.

(E) Proportions of the estimated CNS component in *NGN2*-iNs, cerebral organoid neurons, and human/mouse primary mature PNS/CNS neurons.

(F) Expression of exogenous (left) and endogenous (right) *NGN2* in different *NGN2*-iN clusters. The boxes show the lower and upper quartiles of the distributions. the bars extend to the min/max or 1.5x interquartile range.

from each of the starting iPSC lines. We note that the three main neural subtypes and off-target cells are all observed for each line (Figures 2B–2E). We searched for the top DE genes in the six clusters, and observed distinct gene expression patterns for each cluster (Figure 2F; Table S2). These data suggest that *NGN2*-based neural reprogramming is intrinsically heterogeneous, independent of the purity of the starting cell population, and that the heterogeneity is detected in neurons from multiple iPSC lines.

Molecular features of *NGN2*-iN subpopulations

We next analyzed the molecular signatures that distinguished each *NGN2*-iN cluster. We compared the signatures with *in situ* hybridization (Ravasi et al., 2010) data from the Allen Developing Mouse Brain Atlas, and with single-cell transcriptome atlases containing primary neural cells (Figure 3A). We also assessed the expression of neurotransmitters and other markers of neuron specialization. Based on whole-embryo mouse ISH data (Thompson et al., 2014), we find that *Ngn2* is expressed in progenitor zones in the developing telencephalon and many other brain structures (Figures 3B and S4). *Prph* is expressed in the neural retina, trigeminal nerve, and nuclei within the gray horn of the spinal cord. *Phox2b* is expressed in rhombencephalon/brain stem neurons as well as neurons in the peripheral nervous system (PNS). *Pou4f1* is expressed in the retina, mesencephalon derivatives, trigeminal nerve, and gray horn nuclei. We next compared each *NGN2*-iN cluster with PNS and CNS neurons from primary reference cell atlases (Clark et al., 2019; La Manno et al., 2020; Zeisel et al., 2018) (Figure 3C). Unlike neurons in the iPSC-derived cerebral and retinal organoids, the *NGN2*-iN clusters did not show specific transcriptomic similarity to any CNS neuron subtypes (Figures 3C and S4D). Some *NGN2*-iN clusters were relatively similar to PNS neurons, especially the *PRPH*+ clusters, although they did not specify any of the

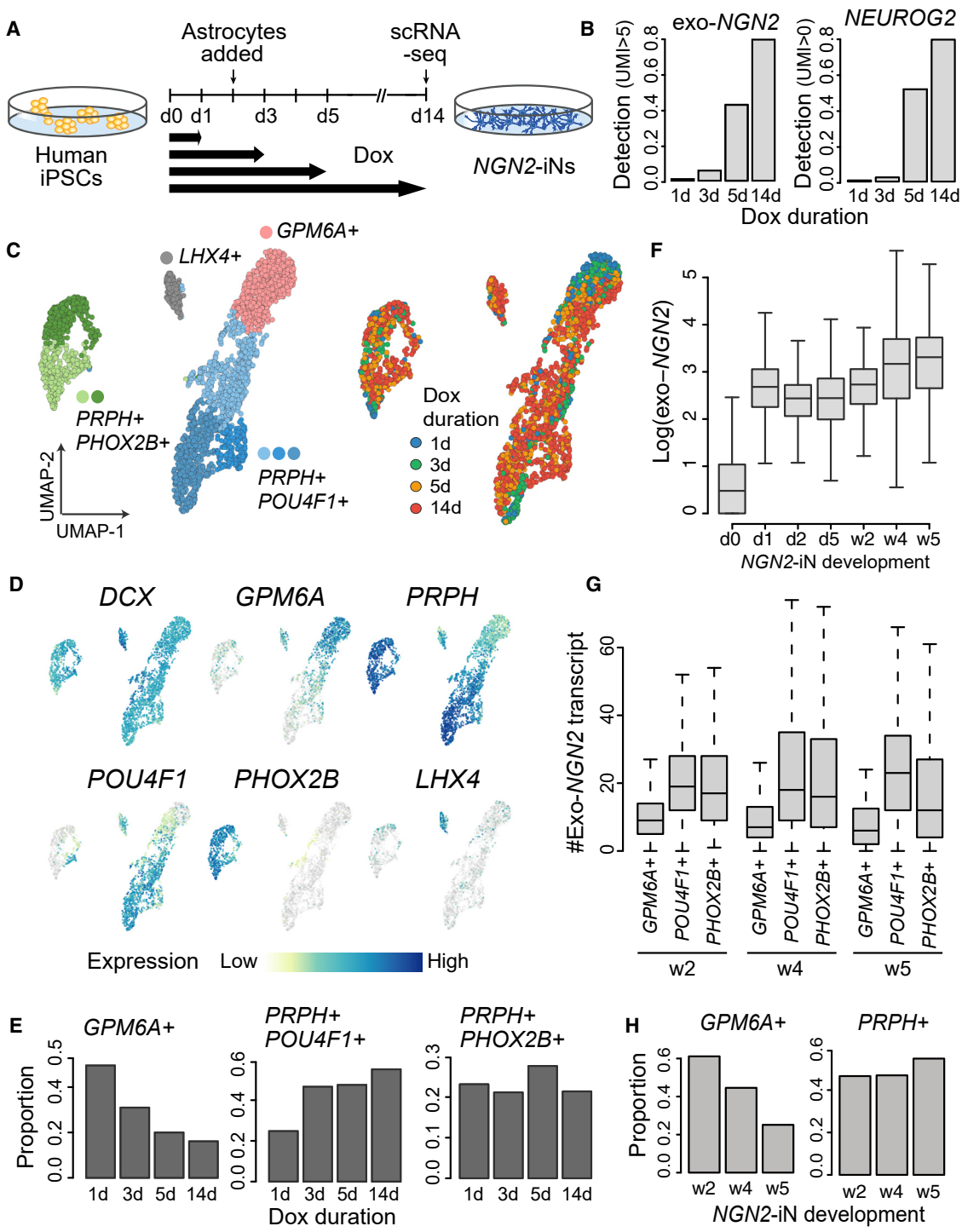


Figure 4. Neuronal fate specification is sensitive to *NGN2* dosage

(A) Schematic for the doxycycline (Dox) treatment duration experiment.

(B) Detection rates for cells expressing exogenous *NGN2* (>5 UMI) and endogenous *NGN2* (>0 UMI) from each sample.

(C) UMAP embedding of *NGN2*-iN cells from each sample. scRNA-seq data were integrated using cluster similarity spectrum-based integration (He et al., 2020).

(D) UMAP plots colored by marker gene expression. *DCX* marks the neural lineages. *LHX4* marks the identity of the off-target cluster.

(legend continued on next page)

PNS neuron subtypes. We explored if *NGN2*-iN expressed markers of primary neuron subtypes (Figures 3D and S4E). There is no clear *in vivo* neuron population as the counterpart of any *NGN2*-iN cluster. We deconvoluted the ratio of PNS/CNS identity for each cluster and found that all *NGN2*-iNs have mixed signatures of CNS- and PNS-derived neurons without a clearly established identity (Figure 3E). We find that the *GPM6A*+ cluster has more CNS features while the *PRPH*+ clusters have a biased PNS signature, in line with *GPM6A* and *PRPH* showing high expression in CNS and PNS neurons, respectively. Altogether, our data suggested that *NGN2*-iNs have a mixture of neuronal signatures, and we were not able to establish a clear identity of *NGN2*-iN populations. We note that this lack of *in vivo* counterpart could be due to incomplete reference cell atlases, as well as discrepancies between human and mouse neurons.

The expression level of reprogramming factors could affect the outcome of reprogramming (Sommer et al., 2012) and lead in part to the heterogeneity of *NGN2*-iNs. We thus analyzed the relationship between *NGN2* expression and the molecular identity of corresponding *NGN2*-iNs. The *NGN2* expression level and proportion of *NGN2*-expressing cells was indeed lower in the off-target cluster (*CD99*+), in line with previous studies that failed reprogramming is linked to silenced reprogramming factors (Treutlein et al., 2016) (Figure 3F). Among the successfully reprogrammed neural clusters, we observed variable expression levels of *NGN2* and proportion of *NGN2*-expressing cells, prompting us to examine whether *NGN2* dosage affects the *NGN2*-iN reprogramming heterogeneity.

Duration of *NGN2* induction affects *NGN2*-iN subtype configuration

We manipulated *NGN2* expression by shortening the duration of Dox treatment and analyzed the resulting cells at d14 using scRNA-seq (Figure 4A). A total of 2,767 cells (d1, 311 cells; d3, 378 cells; d5, 727 cells; d14, 1,351 cells) were included in the analysis. As expected, the expression level of exogenous *NGN2* is correlated with the duration of Dox treatment (Figure 4B). Endogenous *NGN2* expression is also positively correlated with Dox treatment duration, likely as a result of positive autoregulation (Figure 4B) (Ejarque et al., 2013). Each of the previously identified major clusters were detected in all samples; however, the duration of Dox treatment affects the proportion of samples among each cluster (Figures 4C–4E). Specifically, we found that the *GPM6A*+ population was enriched in samples with shorter Dox treatment, while the *PRPH*+/*POU4F1*+ population was more abundant in samples with increased Dox treatment (Figure 4E). Given the observation that *NGN2* expression could affect *NGN2*-iN fate specification, we revisited the time course data for *NGN2*-iN development. The expression of *NGN2* remained nearly consistent during *NGN2*-iN development, with a slight increase of *NGN2* expression at weeks 4 and 5 (Figures 4F and S3B). Interestingly, the expression level of *NGN2* is consistently lower in *GPM6A*+ cells, independent of the duration of development (Figure 4G). Supporting this data, the proportion of *GPM6A*+ cells also decreased with longer *NGN2*-iN culture (Figure 4H). This could imply that prolonged expression of *NGN2* potentially steers *NGN2*-iN fate away from CNS lineages. Altogether, these data show that the duration of *NGN2* induction impacts the proportion of neuron subtypes that emerge in this single factor reprogramming paradigm.

DISCUSSION

Cell fate engineering of neural subtypes from human iPSCs using defined TFs provides extraordinary new inroads into disease modeling and therapy screening using human cells. Methods to rapidly generate mature human neurons are exciting and transformative for these endeavors. It has been established that the *NGN2*-iN protocol is able to reprogram stem cells to general neural fates, with less heterogeneity and higher consistency across multiple stem cell lines compared with traditional reprogramming strategies mediated by small-molecule inhibitors/activators. However, our analysis suggests that the emergent neuron population is heterogeneous, with the heterogeneity being consistent across different cell lines. We are unable to assign the neuron populations to a particular identity with high confidence. We note that this may be due to the fact that current single-cell and spatial transcriptome reference atlases are incomplete. However, without a specific matrix and guiding molecules it may be expected that neurons are not able to establish the molecular profile observed *in vivo* with high precision. Our data show that multiple *NGN2*-iN subpopulations are more similar to neurons of the PNS than CNS, and it is unclear if this culture paradigm is indicative of CNS functionality. Modifications of the *NGN2*-iN

(E) Proportion of cells per Dox treatment time point in each of the three neural clusters.
(F) Expression level of exogenous *NGN2* from the time course experiment presented in Figure S1A. The boxes show the lower and upper quartiles of the distributions. the bars extend to the min/max or 1.5x interquartile range.
(G) Numbers of exogenous *NGN2* transcripts in the three neural clusters from different time points of *NGN2*-iN development. The boxes show the lower and upper quartiles of the distributions. the bars extend to the min/max or 1.5x interquartile range.
(H) Proportion of cells from different time points of *NGN2*-iN development.

protocol by adding developmental patterning factors to the culture medium can steer neuron differentiation to a desired path (Nehme et al., 2018). Our data support a continued effort into identifying combinatorial transcription factor overexpression systems (Ravasi et al., 2010) and medium conditions that can support precise neuron cell type engineering. Furthermore, comprehensive human nervous system reference cell atlases are required to understand the identity of cell states that emerge in *in-vitro*-engineered neuron systems. Single-cell genomics and comparisons with high-dimensional reference atlases should become a field gold standard to assess the heterogeneity and precision of *in vitro* engineered neurons.

EXPERIMENTAL AND COMPUTATIONAL PROCEDURES

Cell culture

All cells described in this work were incubated at 37°C, 5% CO_2, and 90% humidity unless otherwise stated. 409B2 (RIKEN BRC Cell Bank), Sc102a1 (System Biosciences) stem cells, and corresponding rtTA/*NGN2*-derivatives were cultured in standard feeder-free conditions in mTeSR1 (STEMCELL Technologies) on plates coated with Matrigel (Corning). Primary cortical rat astrocytes (Gibco) were cultured in high-glucose DMEM containing 10% fetal calf serum and 1% pen/strep on plates coated with poly-D-lysine (Sigma-Aldrich). Fresh medium was added to the astrocytes every 4–5 days and passaged once a week with trypsin-EDTA digestion at a standard ratio of 1:2. Astrocytes were used up to passage 10, with passage 0 being the culture of initial isolation. rtTA/*NGN2* double-positive stem cell lines were generated and differentiated into *NGN2*-iNs as described previously (Frega et al., 2017).

scRNA-seq library preparation and sequencing

To prepare scRNA-seq libraries from *NGN2*-iN single-cell suspensions, Chromium Single-cell 3′ Reagent Kits (10× Genomics, Pleasanton, CA, USA) were applied according to the manufacturer's instructions. The Chromium Single-cell 3′ Reagent Kits v.2 was employed on *NGN2*-iN generated from 409B2 time course experiments, monoclonal 409B2 and Sc102a1 iPSCs with approximately 3,000 cells loaded per lane on a 10x microfluidic chip device. Chromium Single-cell 3′ Reagent Kit v.3 was used on Dox treatment duration experiments with nearly 8,000 cells loaded per lane. Quantification and quality control of the 10× library was carried out on a Bioanalyzer (Agilent) using high-sensitivity DNA chips. Libraries prepared from the 10× v.2 kit and v.3 kit were respectively sequenced on the Illumina HiSeq 2500 and Illumina NovaSeq S1 platform.

Data analysis of the scRNA-seq experiments

Cell Ranger was used to demultiplex raw base call files to FASTQ files, align reads to the reference genome and transcriptome with the default alignment parameters, demultiplex human and mouse cells, and generate the count matrices for the human cells. Seurat (v.3.1) was then applied to the human scRNA-seq data for further preprocessing. The scRNA-seq data of all 409b2 cells from iPSCs to *NGN2*-iN was integrated with cluster similarity spectrum (CSS) (He et al., 2020). The scRNA-seq data of *NGN2*-iN cells at day 35 of samples from the 409b2 and Sc102a1 human iPSC lines was integrated with Seurat. Generation of UMAP embeddings, clustering, and pseudotime analysis was done on the integrated spaces. Marker genes of different *NGN2*-iN populations were identified as genes with BH corrected $p < 0.01$ and expression fold change >1.2. The benchmark of *NGN2*-iN populations was done by comparisons with the adolescent mouse nervous system atlas (Zeisel et al., 2018), the developing mouse brain (La Manno et al., 2020), and the developing mouse retina (Clark et al., 2019). The *NGN2*-iN scRNA-seq data with varied Dox treatment durations was processed similarly and integrated with CSS. Details of the computational analysis are described in the supplemental experimental procedures.

Data and code availability

The accession number for the processed scRNA-seq data and computational codes reported in this paper is ArrayExpress E-MTAB-10632.

SUPPLEMENTAL INFORMATION

Supplemental information can be found online at https://doi.org/10.1016/j.stemcr.2021.07.006.

AUTHOR CONTRIBUTIONS

M. Schörnig established the NGN2 iPSC lines. M. Schörnig and S.E. generated iN cultures with assistance from A.W. S.E. and W.H. established the selective single-cell dissociation. S.E. generated the time course and d35 scRNA-seq NGN2-iN scRNA-seq data with support from M. Schörnig, M. Santel, and W.H. M. Schörnig, H.-C.L., and M.T.N., with support from S.E. and A.W., generated IHC data. N.N.K. provided guidance to establish the NGN2 iPSC lines. E.T. provided guidance for IHC and iN culture. H.C.L. generated the NGN2 induction time course scRNA-seq data with support from S.E. and M. Santel. Z.H., S.E., and H.C.L. analyzed the scRNA-seq data. H.C.L., Z.H., S.E., B.T., and J.G.C. designed the study and wrote the manuscript.

CONFLICT OF INTERESTS

B.T. is a member of the Editorial Board of Stem Cell Reports.

ACKNOWLEDGMENTS

We thank the Camp and Treutlein labs for helpful discussions. J.G.C. and B.T. are supported by the Chan Zuckerberg Initiative DAF (grant no. CZF2019-002440), an advised fund of Silicon Valley Community Foundation. J.G.C. is supported by the European Research Council (Anthropoid-803441) and the Swiss National Science Foundation (project grant 310030_84795). B.T. is supported by the European Research Council (Organomics-758877, Braintime-874606), the Swiss National Science Foundation (project grant 310030_192604) and the National Center of Competence in Research Molecular Systems Engineering. H.C.L. is supported by the Human Frontier Science Program (LT000399/2020-L).

Received: September 21, 2020
Revised: July 5, 2021
Accepted: July 6, 2021
Published: August 5, 2021

REFERENCES

Aibar, S., Gonzalez-Blas, C.B., Moerman, T., Huynh-Thu, V.A., Imrichova, H., Hulselmans, G., Rambow, F., Marine, J.C., Geurts, P., Aerts, J., et al. (2017). SCENIC: single-cell regulatory network inference and clustering. Nat. Methods *14*, 1083–1086.

Biddy, B.A., Kong, W., Kamimoto, K., Guo, C., Waye, S.E., Sun, T., and Morris, S.A. (2018). Single-cell mapping of lineage and identity in direct reprogramming. Nature *564*, 219–224.

Camp, J.G., Wollny, D., and Treutlein, B. (2018). Single-cell genomics to guide human stem cell and tissue engineering. Nat. Methods *15*, 661–667.

Chen, M., Maimaitili, M., Habekost, M., Gill, K.P., Mermet-Joret, N., Nabavi, S., Febbraro, F., and Denham, M. (2020). Rapid generation of regionally specified CNS neurons by sequential patterning and conversion of human induced pluripotent stem cells. Stem Cell Res. *48*, 101945.

Clark, B.S., Stein-O'Brien, G.L., Shiau, F., Cannon, G.H., Davis-Marcisak, E., Sherman, T., Santiago, C.P., Hoang, T.V., Rajaii, F., James-Esposito, R.E., et al. (2019). Single-cell RNA-seq analysis of retinal development identifies NFI factors as regulating mitotic exit and late-born cell specification. Neuron *102*, 1111–1126 e1115.

Diez-Roux, G., Banfi, S., Sultan, M., Geffers, L., Anand, S., Rozado, D., Magen, A., Canidio, E., Pagani, M., Peluso, I., et al. (2011). A high-resolution anatomical atlas of the transcriptome in the mouse embryo. PLoS Biol *9*, e1000582.

Ejarque, M., Cervantes, S., Pujadas, G., Tutusaus, A., Sanchez, L., and Gasa, R. (2013). Neurogenin3 cooperates with Foxa2 to autoactivate its own expression. J. Biol. Chem. *288*, 11705–11717.

Frega, M., van Gestel, S.H., Linda, K., van der Raadt, J., Keller, J., Van Rhijn, J.R., Schubert, D., Albers, C.A., and Nadif Kasri, N. (2017). Rapid neuronal differentiation of induced pluripotent stem cells for measuring network activity on micro-electrode arrays. J. Vis. Exp. *119*, 54900. https://doi.org/10.3791/54900.

Guo, C., and Morris, S.A. (2017). Engineering cell identity: establishing new gene regulatory and chromatin landscapes. Curr. Opin. Genet. Dev. *46*, 50–57.

He, Z., Brazovskaja, A., Ebert, S., Camp, J.G., and Treutlein, B. (2020). CSS: cluster similarity spectrum integration of single-cell genomics data. Genome Biol. *21*, 224.

Hu, H., Miao, Y.R., Jia, L.H., Yu, Q.Y., Zhang, Q., and Guo, A.Y. (2019). AnimalTFDB 3.0: a comprehensive resource for annotation and prediction of animal transcription factors. Nucleic Acids Res. *47*, D33–D38.

Kanton, S., Boyle, M.J., He, Z., Santel, M., Weigert, A., Sanchis-Calleja, F., Guijarro, P., Sidow, L., Fleck, J.S., Han, D., et al. (2019). Organoid single-cell genomic atlas uncovers human-specific features of brain development. Nature *574*, 418–422.

Karow, M., Camp, J.G., Falk, S., Gerber, T., Pataskar, A., Gac-Santel, M., Kageyama, J., Brazovskaja, A., Garding, A., Fan, W., et al. (2018). Direct pericyte-to-neuron reprogramming via unfolding of a neural stem cell-like program. Nat. Neurosci. *21*, 932–940.

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. https://doi.org/10.1101/2020.07.02.184051.

Lin, Y.T., Seo, J., Gao, F., Feldman, H.M., Wen, H.L., Penney, J., Cam, H.P., Gjoneska, E., Raja, W.K., Cheng, J., et al. (2018). APOE4 causes widespread molecular and cellular alterations associated with Alzheimer's disease phenotypes in human iPSC-derived brain cell types. Neuron *98*, 1141–1154.e47.

Nehme, R., Zuccaro, E., Ghosh, S.D., Li, C., Sherwood, J.L., Pietilainen, O., Barrett, L.E., Limone, F., Worringer, K.A., Kommineni, S., et al. (2018). Combining NGN2 programming with developmental patterning generates human excitatory neurons with NMDAR-mediated synaptic transmission. Cell Rep. *23*, 2509–2523.

Nickolls, A.R., Lee, M.M., Espinoza, D.F., Szczot, M., Lam, R.M., Wang, Q., Beers, J., Zou, J., Nguyen, M.Q., Solinski, H.J., et al. (2020). Transcriptional programming of human mechanosensory neuron subtypes from pluripotent stem cells. Cell Rep. *30*, 932–946 e937.

Ravasi, T., Suzuki, H., Cannistraci, C.V., Katayama, S., Bajic, V.B., Tan, K., Akalin, A., Schmeier, S., Kanamori-Katayama, M., Bertin, N., et al. (2010). An atlas of combinatorial transcriptional regulation in mouse and man. Cell *140*, 744–752.

Schörnig, M., Ju, X., Fast, L., Ebert, S., Weigert, A., Kanton, S., Schaffer, T., Kasri, N.N., Treutlein, B., Peter, B.M., et al. (2021). Comparison of induced neurons reveals slower structural and functional maturation in humans than in apes. eLife *10*, e59323.

Sommer, C.A., Christodoulou, C., Gianotti-Sommer, A., Shen, S.S., Sailaja, B.S., Hezroni, H., Spira, A., Meshorer, E., Kotton, D.N., and Mostoslavsky, G. (2012). Residual expression of reprogramming factors affects the transcriptional program and epigenetic signatures of induced pluripotent stem cells. PLoS One *7*, e51711.

Thompson, C.L., Ng, L., Menon, V., Martinez, S., Lee, C.K., Glattfelder, K., Sunkin, S.M., Henry, A., Lau, C., Dang, C., et al. (2014). A high-resolution spatiotemporal atlas of gene expression of the developing mouse brain. Neuron *83*, 309–323.

Treutlein, B., Lee, Q.Y., Camp, J.G., Mall, M., Koh, W., Shariati, S.A., Sim, S., Neff, N.F., Skotheim, J.M., Wernig, M., et al. (2016). Dissecting direct reprogramming from fibroblast to neuron using single-cell RNA-seq. Nature *534*, 391–395.

Yuan, A., Sasaki, T., Kumar, A., Peterhoff, C.M., Rao, M.V., Liem, R.K., Julien, J.P., and Nixon, R.A. (2012). Peripherin is a subunit of peripheral nerve neurofilaments: implications for differential vulnerability of CNS and peripheral nervous system axons. J. Neurosci. *32*, 8501–8508.

Zeisel, A., Hochgerner, H., Lonnerberg, P., Johnsson, A., Memic, F., van der Zwan, J., Haring, M., Braun, E., Borm, L.E., La Manno, G., et al. (2018). Molecular architecture of the mouse nervous system. Cell *174*, 999–1014.e22.

Zhang, Y., Pak, C., Han, Y., Ahlenius, H., Zhang, Z., Chanda, S., Marro, S., Patzke, C., Acuna, C., Covy, J., et al. (2013). Rapid single-step induction of functional neurons from human pluripotent stem cells. Neuron *78*, 785–798.

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

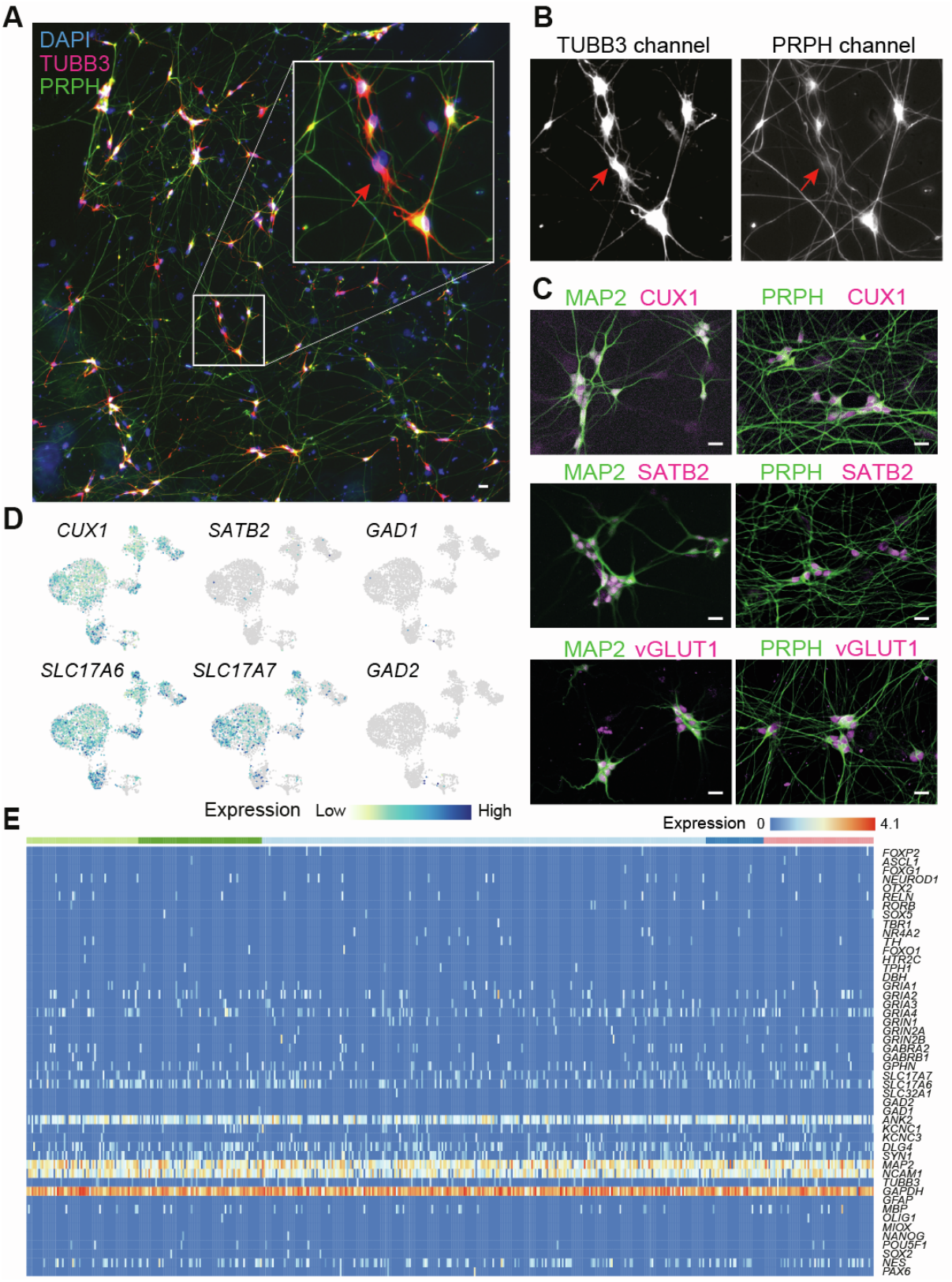
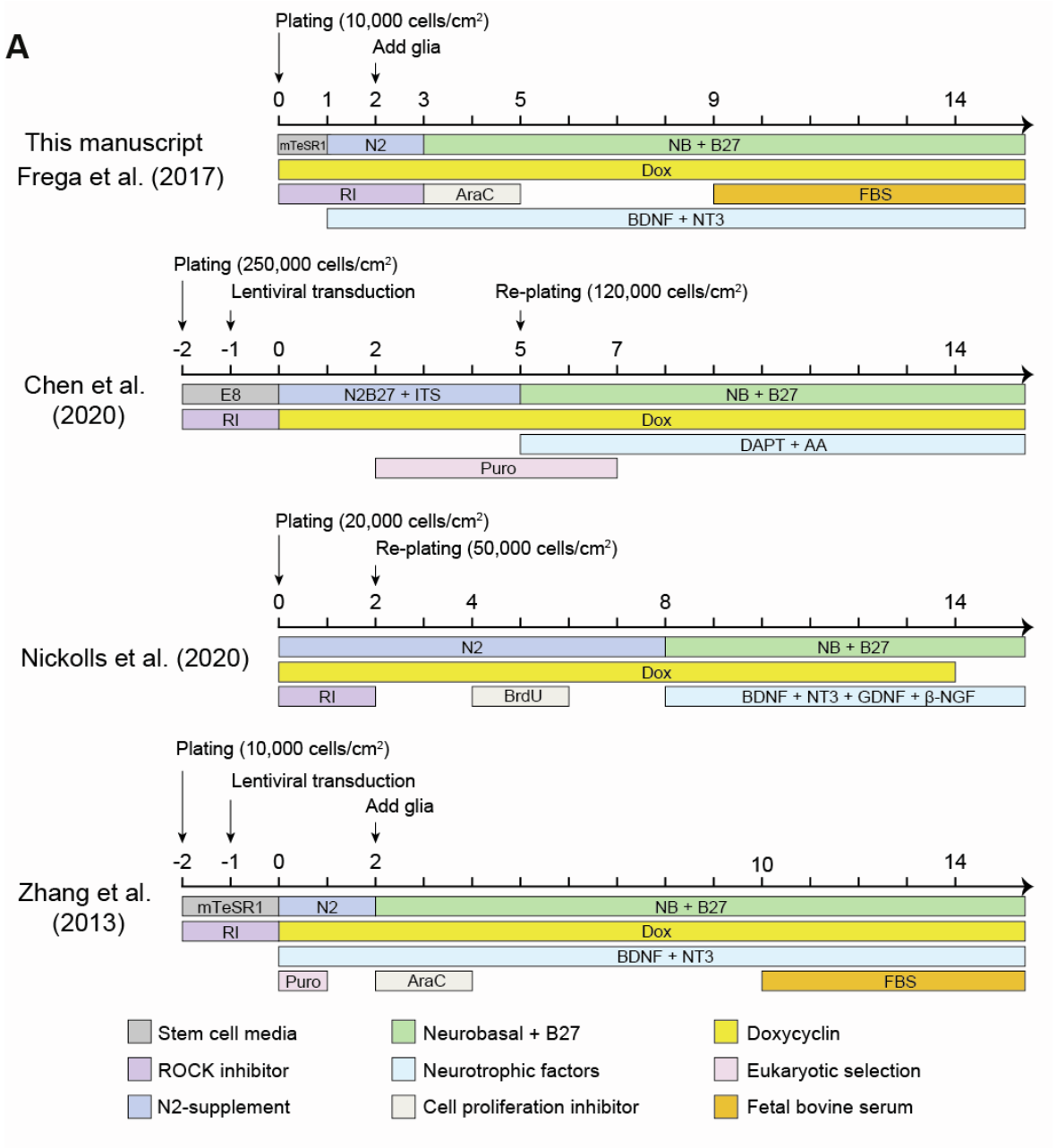


Fig S1. Characterization of *NGN2*-iNs directly programmed from iPSCs. (A) Immunofluorescence of *NGN2*-iNs at d14 with DAPI (blue), TUBB3 (red) and PRPH (green). Red arrow in the inset showed neurons with TUBB3 marker but not PRPH. Scale bar is 10

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.

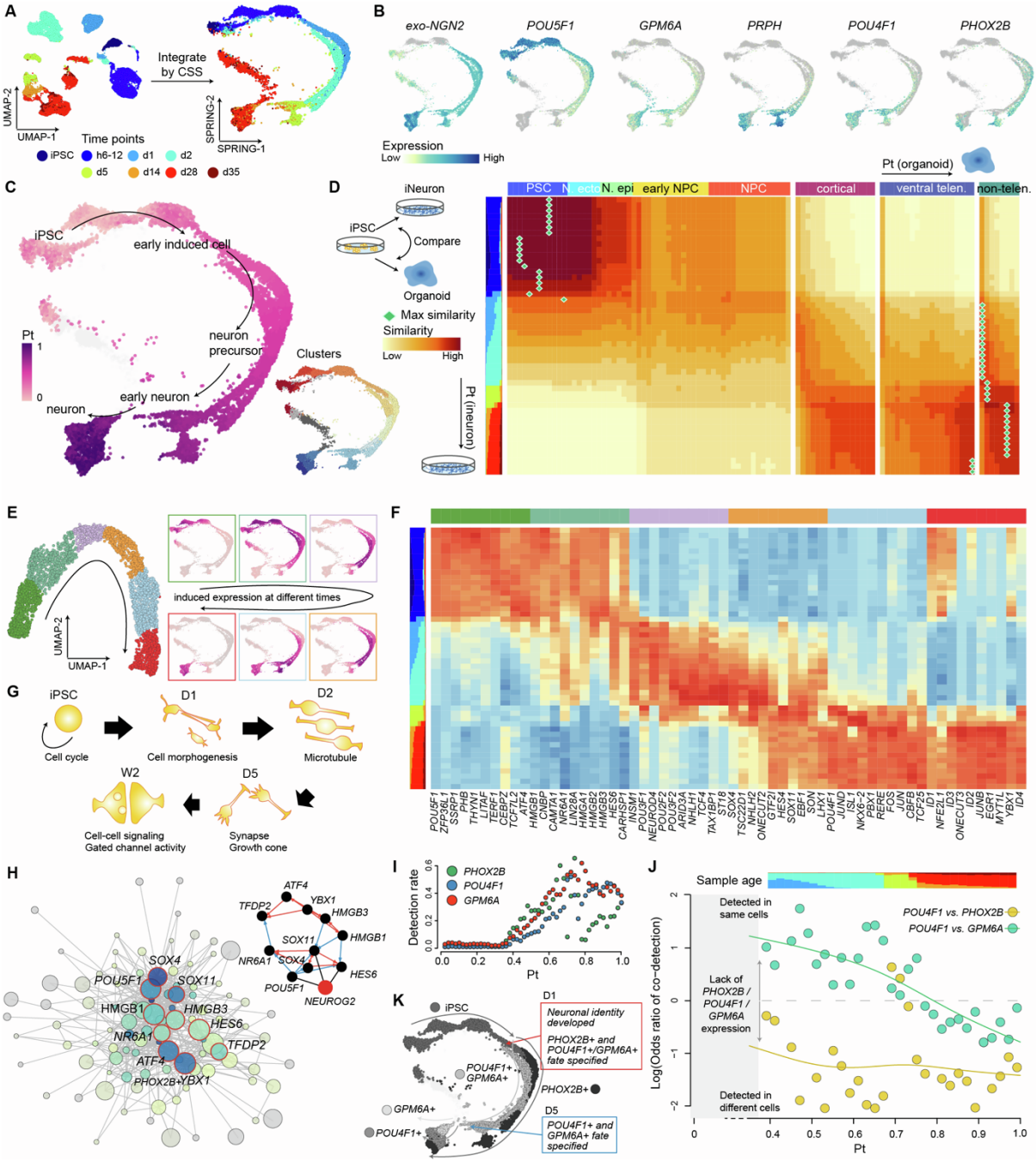


Fig S3. Analysis of *NGN2*-mediated direct reprogramming path from iPSC to induced neurons. (A) *NGN2*-iNs directly reprogrammed from 409B2 iPSCs were harvested at different time points after induced expression of *NGN2* by Dox treatment and analyzed by scRNA-seq. scRNA-seq data was directly analyzed 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.

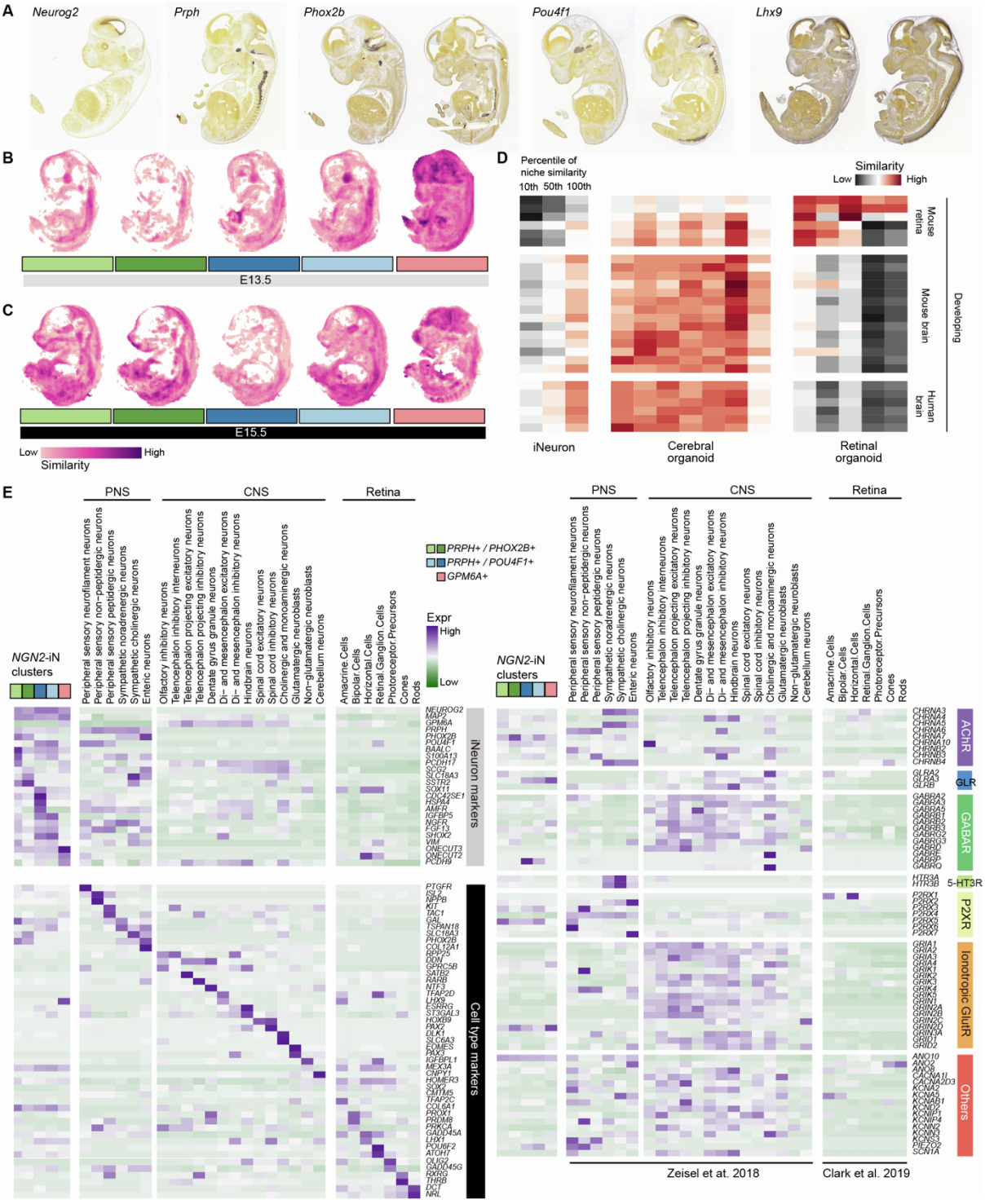


Fig S4. Benchmark of *NGN2*-iNs through transcriptomic comparison to primary references in mouse. (A) Spatial expression patterns of selected markers via in situ hybridization (6) in the E13.5 mouse embryos from the Allen Developing Mouse Brain Atlas. (B-C) Transcriptomic similarity between each of the 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.

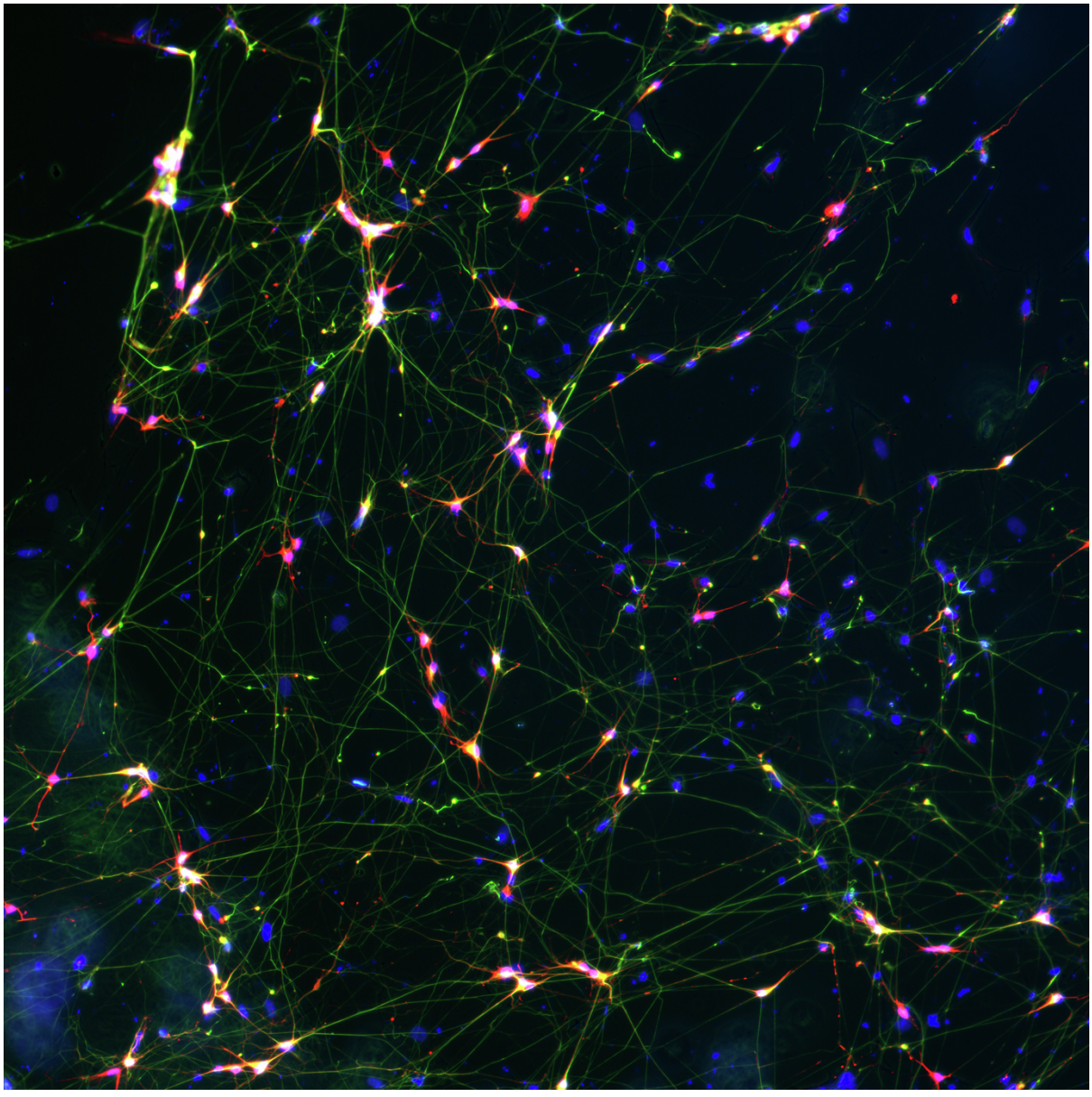


Image S2: Raw image of TUBB3 channel for d14 *NGN2*-iNs immunofluorescence in Fig. S1B.

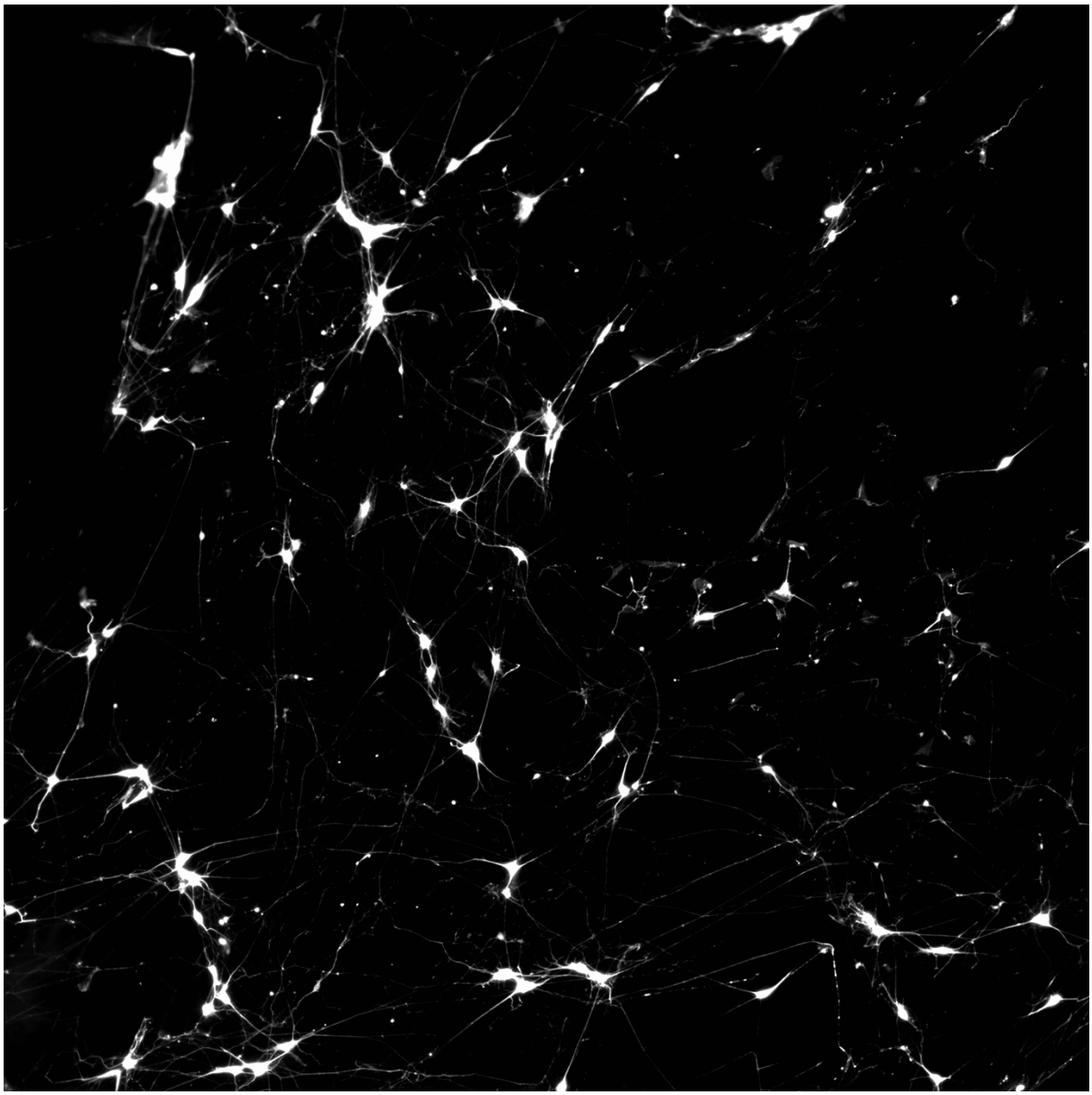
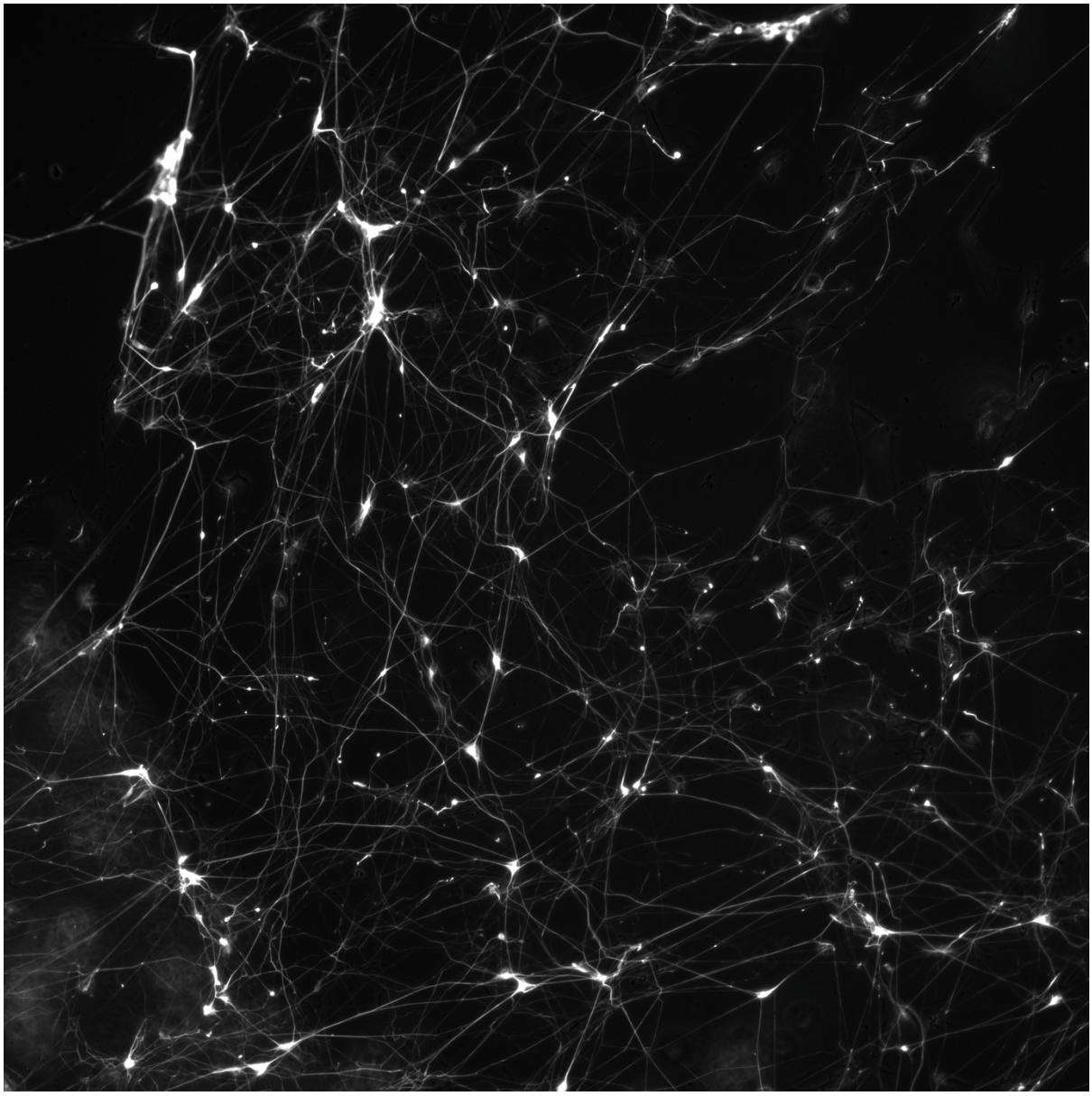


Image S3: Raw image of PRPH channel for d14 *NGN2*-iNs immunofluorescence in Fig. S1B.



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.