



Generation of human excitatory forebrain neurons by cooperative binding of proneural NGN2 and homeobox factor EMX1

Cheen Euong Ang^{a,b,c}, Victor Hipolito Olmos^{b,c}, Kayla Vodehnal^{b,c} , Bo Zhou^{b,c,d,e}, Qian Yi Lee^{a,b,c}, Rahul Sinha^c , Aadit Narayanaswamy^{b,c} , Moritz Mall^{b,c,1,2,3} , Kirill Chesnov^{b,c}, Caia S. Dominicus^{f,g}, Thomas Südhof^{d,e}, and Marius Wernig^{b,c,4}

Edited by John Rubenstein, University of California San Francisco, San Francisco, CA; received May 24, 2023; accepted January 24, 2024

Generation of defined neuronal subtypes from human pluripotent stem cells remains a challenge. The proneural factor NGN2 has been shown to overcome experimental variability observed by morphogen-guided differentiation and directly converts pluripotent stem cells into neurons, but their cellular heterogeneity has not been investigated yet. Here, we found that NGN2 reproducibly produces three different kinds of excitatory neurons characterized by partial coactivation of other neurotransmitter programs. We explored two principle approaches to achieve more precise specification: prepatterning the chromatin landscape that NGN2 is exposed to and combining NGN2 with region-specific transcription factors. Unexpectedly, the chromatin context of regionalized neural progenitors only mildly altered genomic NGN2 binding and its transcriptional response and did not affect neurotransmitter specification. In contrast, coexpression of region-specific homeobox factors such as EMX1 resulted in drastic redistribution of NGN2 including recruitment to homeobox targets and resulted in glutamatergic neurons with silenced nonglutamatergic programs. These results provide the molecular basis for a blueprint for improved strategies for generating a plethora of defined neuronal subpopulations from pluripotent stem cells for therapeutic or disease-modeling purposes.

induced neuron | reprogramming | transcription factor | homeodomain | chromatin

The mammalian nervous system is the most diverse organ, comprising a plethora of neurons and glial cells organized along anterior–posterior and dorsal–ventral axes. Those neurons and glial cells differentiate from progenitor cells endowed each with a positional and subtype identity by spatially and temporally defined morphogens, transcription factors, and growth factors gradients (1–4). Various proneural basic helix–loop–helix (bHLH) factors which are responsible to induce neuronal fates in progenitor cells have been described to assume different roles in subtype specifications (5, 6). For example, *Ngn2* and *Ascl1* often have mutually exclusive expression patterns, generating neurons of different subtypes such as glutamatergic and GABAergic neurons in the forebrain and spinal cord (7). In the retina, *Ascl1/Math3* and *NeuroD1/Math3* are important in the development of bipolar cells and amacrine cells, respectively (8). Those proneural factors also partner with cofactors, such as homeodomain and POU domain transcription factors, in different stages of neuronal differentiation (2, 9). One of the well-studied areas is the dorsal–ventral axis specification of the developing spinal cord. The progenitors along the dorsal–ventral axis in the spinal cord are marked by different homeodomain transcription factors, specifying the progenitors destined to give rise different neuronal subtypes (10). In postmitotic neurons, homeodomain transcription factors, such as *Mnx1* and *Crx*, are crucial in motor neurons and cone cells differentiation (11, 12). Such transcription factors also function outside the context of normal development as combinations between proneural and lineage-specific factors can even convert fibroblasts to induced neuronal (iN) cells with different neuronal subtypes (13–18).

We previously developed a protocol to generate functional neurons rapidly and robustly by NGN2 overexpression from human embryonic stem (hES) of induced pluripotent stem (iPS) cells in a serum-free and defined media. Those NGN2-iN cells express pan-neuronal and excitatory neuronal markers, fire repetitive action potentials and form functional synapses, making them a versatile platform to study cell biological processes and pathophysiology in human neurons (13, 19–25). It has been previously reported that NGN2 activity is context dependent: Overexpression of NGN2 in neural tube cell cultures under a low BMP condition promoted a sensory fate while under a high BMP condition promoted an autonomic fate (26). It is thus unclear whether overexpression of NGN2 in hES/iPS cells will generate glutamatergic iN cells with partial neurotransmitter programs. Here, we explored NGN2's molecular and chromatin function and its ability to induce specific neuronal programs in

Significance

Neuronal reprogramming has shown exciting promises, enabling disease modeling to drug screening. While considerable progresses have been made on finding transcription factors that can generate different neuronal subtypes, little work has been performed to understand mechanistically how chromatin environments and homeodomain cofactors can restrict the promiscuous action of proneural factor NGN2. Here, we present a comprehensive analysis using sequencing to dissect the molecular mechanisms of how chromatin environments and cofactors restrict NGN2.

Author contributions: C.E.A. and M.W. designed the research; C.E.A., V.H.O., K.V., B.Z., Q.Y.L., R.S., A.N., M.M., K.C., and C.S.D. performed research; T.S. contributed new reagents/analytic tools; C.E.A., V.H.O., K.V., B.Z., A.N., K.C., C.S.D., T.S., and M.W. analyzed the data; and C.E.A. and M.W. wrote the paper.

Competing interest statement: T.S. and M.W. are co-founders of Neucyte, Inc., and M.W. is a scientific advisor at Bit.Bio; both companies use transcription factor programming methods.

This article is a PNAS Direct Submission.

Copyright © 2024 the Author(s). Published by PNAS. This article is distributed under [Creative Commons Attribution-NonCommercial-NoDerivatives License 4.0 \(CC BY-NC-ND\)](https://creativecommons.org/licenses/by-nc-nd/4.0/).

¹Present address: Cell Fate Engineering and Disease Modeling Group, German Cancer Research Center and Center for Molecular Biology Alliance, Heidelberg 69120, Germany.

²Present address: Hector Institute for Translational Brain Research gGmbH, Heidelberg 69120, Germany.

³Present address: Central Institute of Mental Health, Medical Faculty Mannheim, Heidelberg University, Mannheim 68159, Germany.

⁴To whom correspondence may be addressed. Email: wernig@stanford.edu.

This article contains supporting information online at <https://www.pnas.org/lookup/suppl/doi:10.1073/pnas.2308401121/-/DCSupplemental>.

Published March 6, 2024.

the context of different cell states and different transcription factor combinations using a series of single-cell and bulk genomic sequencing techniques and epigenomic methods.

Results

NGN2 Induces Three Kinds of Glutamatergic Neurons in hES Cells Characterized by a Partial Cholinergic Program. We have previously shown that forced expression of NGN2 in human ES and iPS cells results in an efficient conversion into functionally homogeneous excitatory neurons by synaptic measurements where we saw >90% of the neurons exhibited spontaneous and evoked excitatory postsynaptic currents (25). For best characterization of a functionally homogeneous population and to reduce RNA dropout for lowly expressed transcription factors, we decided to sequence fewer cells but with deep coverage using the standard SMART-seq2 protocol. We performed single-cell RNA sequencing (scRNA-seq) at days 4 and 28 post-NGN2 induction (Fig. 1A). After filtering for cells with expression of at least 2,500 genes and with 200,000 unique paired end reads, we obtained 27 and 62 high-quality cells for the day 4 and 28 time points, respectively (Fig. 1B). We also obtained scRNA-seq data from hES cells and included them into our analysis (day 0 time point) (27). When we performed principal component analysis (PCA), the cells from the three time points separated into three clusters largely corresponding to cells from the three time points with the first principal component corresponding to genes enriched for gene ontology (GO) terms associated with cell proliferation and the second principal component corresponding to genes enriched in nervous system development (Fig. 1C and *SI Appendix, Fig. S1A*) compatible with previous results (20, 25). When we performed hierarchical clustering of genes that are four-fold changed among the different time points, we found that ES cell-specific genes (*POU5F1*, *NANOG*, and *SOX2*) were down-regulated precipitously in the d4 and d28 time points. As expected, transcription factors reported to be downstream of NGN2 (*NEUROD1*, *NHLH1*, *NEUROD4*, and *HES6*) were up-regulated at day 4 and subsequently down-regulated. Mature neuronal markers (*MYT1L*) and neuronal subtype markers (*ISL1* and *PHOX2B*) were induced and maintained during differentiation (Fig. 1D and *SI Appendix, Fig. S1B*). To assess the degree of heterogeneity of day 4 and 28 iN cells, we performed t-distributed stochastic neighbor embedding (tSNE) using genes that are most variable across hES, day4 iN and day28 iN cells ($y_{\text{cutoff}} = 0.75$, $x_{\text{cutoff}} = 0.5$, variable genes = 1,144). As expected, day 28 iN cells homogeneously expressed glutamatergic markers [*VGLUT2* (*SLC17A6*) and *VGLUT1* (*SLC17A7*)] (25). However, the day 28 iN cells formed three relatively distinct clusters each containing about 1/3 of the cells (Fig. 1E–G). One cell cluster was positive for a subset of cholinergic transporters [*vAChT* (*SLC18A3*) and *ChT* (*SLC5A7*)] and the two transcription factors, *ISL1* and *PHOX2B*. Another cluster was characterized by glutamatergic genes and *ISL1*, but not *PHOX2B*. The third cluster expressed glutamatergic markers only (Fig. 1F and H). These findings agree with previously reported scRNA-seq data on NGN iN cells (28). To validate the results of the scRNA-seq data, we performed immunofluorescence for *ISL1* and *PHOX2B* at day 4. Immunofluorescence data confirmed the presence of *ISL1*+ cells among the FLAG-NGN2 infected cells (~50%, $n = 3$). 25% of the *ISL1* positive cells were also *PHOX2B* positive (Fig. 1I–L). Of note, CHAT, the rate-limiting enzyme for acetylcholine synthesis, is not expressed in any of the cells with either of the two cholinergic transporters hinting that the cholinergic identity is incomplete (*SI Appendix, Fig. S1D*).

ISL1 and *PHOX2B* are known to be expressed in spinal/cranial motor neurons and sympathetic/parasympathetic neurons (27, 29). Notably, their overexpression induces cranial or spinal cholinergic

neurons (30). Moreover, *ISL1*, *PHOX2B*, *SLC18A3*, and *ChT* (*SLC5A7*) are not detected by scRNA-seq performed by Allen Brain Atlas in neurons and glial cells in the human cortex (medial temporal gyrus) (*SI Appendix, Fig. S1G*) (31). We therefore found the prominent *ISL1* induction following NGN2 expression noteworthy and asked whether the level of NGN2 transgene expression may affect *ISL1* expression in the hES cell system. However, we found no correlation between the intensity of FLAG staining (indicative of NGN2 level) to that of *ISL1* ($R^2 = 0.0002$, Pearson = -0.01) indicating that the level of transgene NGN2 expression did not correlate with *ISL1* (*SI Appendix, Fig. S1E*). Similarly, little to no correlation was found between *ISL1* and *PHOX2B* expression levels ($R^2 = 0.0534$, Pearson = 0.23) with *PHOX2B*+ cells being a subset of the *ISL1*+ cell population (Fig. 1L and *SI Appendix, Fig. S1F*).

A closer examination on the neurotransmitter genes and transporters or rate-limiting enzymes for GABAergic, monoaminergic, and cholinergic neurons showed that the transporters and enzymes required for the neurotransmitter production and release were not all expressed within the same cell (*SI Appendix, Fig. S1C and D*) signifying that the induction of other nonglutamatergic programs were not complete. This intriguing finding raised the possibility that NGN2 induces multiple incomplete neurotransmitter programs and additional mechanisms must complement NGN2 to accomplish precise neurotransmitter programs (*SI Appendix, Fig. S1D*).

Cholinergic Gene Induction in hES Cells by NGN2 Is Mediated by *ISL1* and *PHOX2B*. To assess whether *ISL1* and *PHOX2B* may be responsible for the induction of cholinergic genes, we overexpressed *ISL1*, *PHOX2B*, and both genes together with NGN2 and found that all transcription factor combinations produced β -III-tubulin positive cells as early as day 4 (Fig. 2A and B). Quantitative RT-PCR on day 4 iN cells showed that *ISL1* and *PHOX2B* induced expression of two cholinergic genes [*vAChT* (*SLC18A3*) and *ChT* (*SLC5A7*)] (Fig. 2D). CHAT was only induced moderately and only by *ISL1* (Fig. 2D). The same findings were reproducible in another ES cell line (*SI Appendix, Fig. S2C*). We confirmed induction of ChT (*SLC5A7*) on the protein level by western blotting which also revealed that neither did *ISL1* induce additional *PHOX2B* compared to control, nor did *PHOX2B* induce *ISL1* suggesting that the induction of cholinergic genes by those two transcription factors is independent of each other (Fig. 2C). Next, we asked whether *ISL1* and *PHOX2B* would be necessary for induction of cholinergic genes. Using two hairpins specific to *ISL1*, we found that *ISL1* downregulation indeed reduced the expression of CHAT, vAChT (*SLC18A3*), and ChT (*SLC5A7*) (Fig. 2E, G, and H). The downregulation of ChT (*SLC5A7*) could be confirmed by western blotting (Fig. 2F). Unlike *ISL1*, *PHOX2B* knock-down only reduced the induction of ChT (*SLC5A7*), the gene most prominently induced by *PHOX2B* (*SI Appendix, Fig. S2A and B*) while the other two cholinergic genes remained unchanged.

Regionalization of Donor Cells Is Maintained throughout NGN2-Mediated Differentiation but Does Not Resolve Neurotransmitter Identity Blurring. It was previously suggested that a proneural factor, NGN2, may induce subtype specification programs in a cell context-dependent manner, integrating the positional identity and developmental history of the cell (32). Given that, we thus hypothesized we can leverage the prior development biology knowledge and overexpress NGN2 in neural progenitor cells of specific regional identities to restrict subtype specification. To test this idea, we first differentiated hES cells into anterior [treated with SB431542 and LDN193189 (SL)] and posterior [treated with SB431542, LDN193189, and CHIR99021 (SLC)]

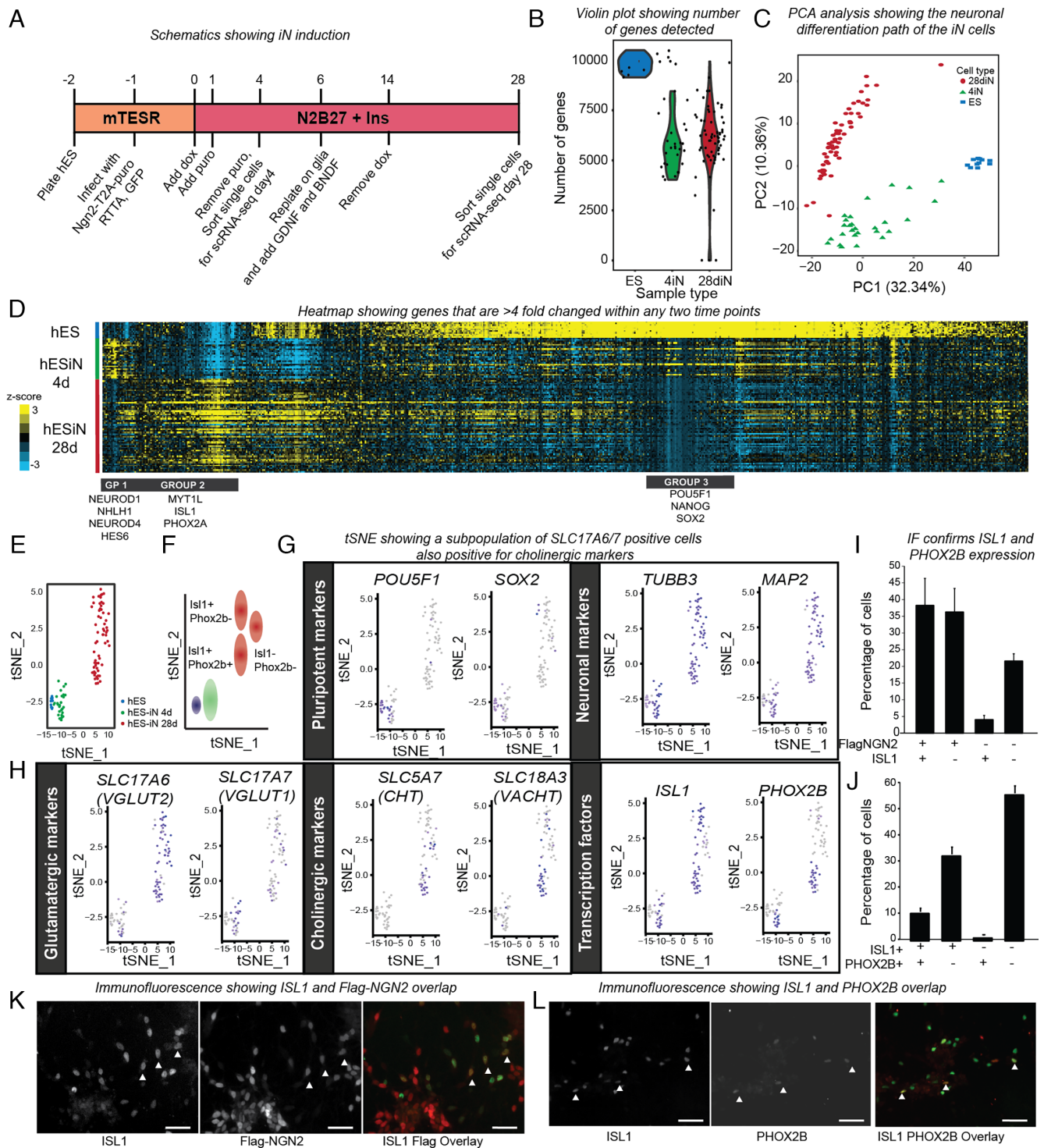


Fig. 1. Single-cell RNA sequencing reveals heterogeneity of *Ngn2* *iN* cells. (A) Differentiation protocol to obtain the day 4 and 28d *NGN2* induced neuronal (*iN*) cells. Day 4 represents immature postmitotic neurons, and day 28 represents mature neurons. In total, we obtained 27 day 4 and 62 day 28 high-quality *iN* cells. (B) Violin plot showing the number of genes detected by single-cell RNA sequencing in undifferentiated ES cells, day 4, and day 28 *iN* cells. Only cells that had 200,000 reads and more than 2,500 genes were considered for subsequent analysis. The ES cell scRNA sequencing (11 cells) was downloaded from NCBI GEO (GSM1964970). (C) Principal component analysis showing the progression of *iN* cell reprogramming. The gene ontology analysis of genes corresponding to the first and second principal components is shown in *SI Appendix, Fig. S1A*. (D) Hierarchical clustering of genes that are changed at least fourfold among the three time points. Group 1 contains direct *NGN2* target genes which tend to be highest in day 4 cells. Group 2 includes mature neuronal markers. Group 3 genes include pluripotency markers. The GO terms for each cluster are listed in *SI Appendix, Fig. S1B*. (E) T-distributed stochastic neighbor embedding (t-SNE) plot for the three different cell populations: ES cells (blue), *NGN2* 4d (green), and 28d (red) *iN* cells plotted using Seurat with the following settings ($y_{cutoff} = 0.75$, $x_{cutoff} = 0.5$, variable genes = 1,144). (F) Key for specific cell populations described in G and H. (G) t-SNE plot for two ES cell markers (*POU5F1* and *SOX2*) and neuronal markers (*TUBB3* and *MAP2*). Gray and purple represent low- and high-expressing cells, respectively. (H) t-SNE plot for two markers for glutamatergic (*SLC17A6* and *SLC17A7*) and cholinergic neuronal markers (*SLC5A7* and *SLC18A3*). The day 28 cells that are positive for both cholinergic markers are also positive for *ISL1* and *PHOX2B*. Gray and purple represent low- and high-expressing cells, respectively. (I) Quantification of day 28 *NGN2* *iN* cells for *FLAG* (*Flag-NGN2*) and *ISL1* immunofluorescence. Double- and single-positive, and double-negative cells are shown as the percentage of all DAPI-positive cells ($n = 3$). Error bars = SEM. (J) Quantification of *ISL1* and *PHOX2B* positive cells as in I ($n = 3$). Error bars = SEM. (K) Immunofluorescence images of *ISL1* (Left) and *FLAG* (Middle) and the overlay (Right). White triangles: *ISL1*:*FLAG* double-positive cells. (Scale bar: 50 μm .) (L) Immunofluorescence images of *ISL1* (Left) and *PHOX2B* (Middle) and the overlay (Right). White triangles: *ISL1* and *PHOX2B* double-positive cells. (Scale bar: 50 μm .)

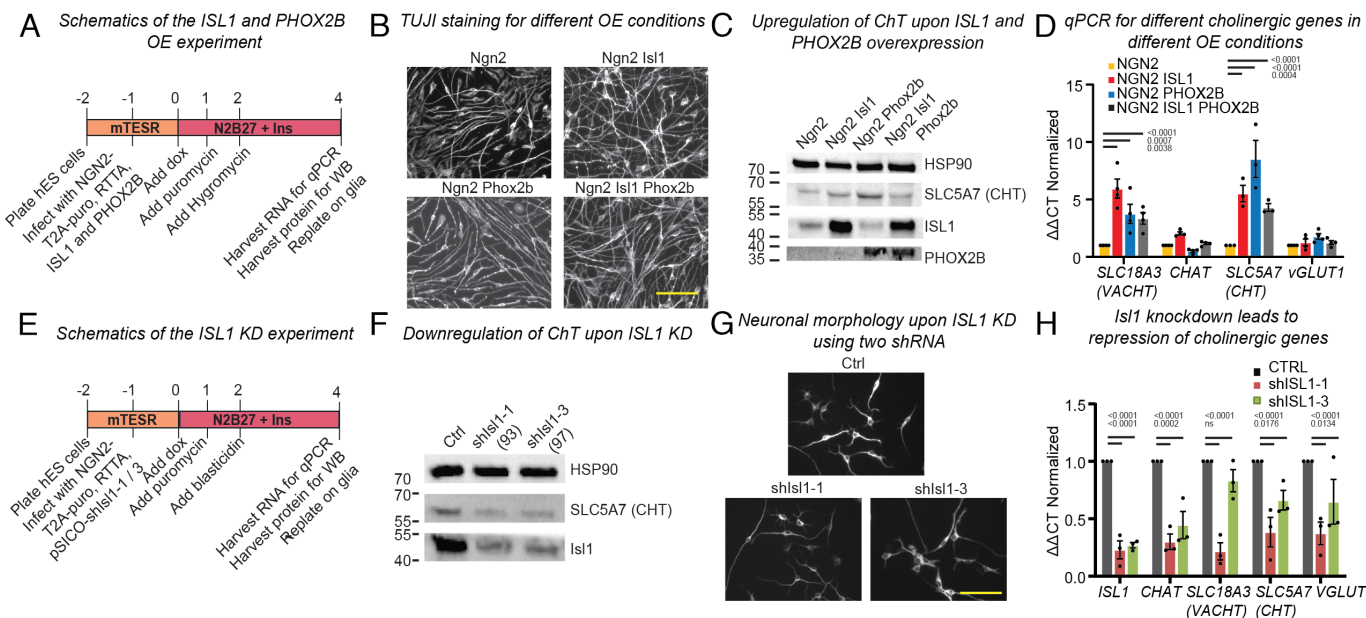


Fig. 2. ISL1 and PHOX2B are sufficient to induce cholinergic genes. (A) Outline to generate day 4 Ngn2 iN cells with additional expression of ISL1, PHOX2B, or ISL1:PHOX2B. (B) Immunofluorescence images of β -III-tubulin (TUJ1) staining of day 4 Ngn2 iN cells alone or coexpressing ISL1, PHOX2B, and ISL1:PHOX2B. (Scale bar: 100 μ m.) (C) Western blot analysis of cells in day 4 NGN2 iN cells alone or coexpressing ISL1, PHOX2B, and ISL1:PHOX2B using antibodies indicated. (D) Quantitative RT-PCR examining the day 4 NGN2 iN cells alone or coexpressing ISL1, PHOX2B, and ISL1:PHOX2B for three cholinergic markers (CHAT, SLC18A3, and SLC5A7) and a glutamatergic marker (VGLUT1). (N = 3, error bars = SEM, ANOVA. Exact adjusted *P*-values are marked in the graph). (E) Outline to generate day 4 iN cells expressing a control or two different ISL1 short hairpins. (F) Western blot analysis of day 4 iN cells expressing a control or two ISL1 hairpins probing for SLC5A7 (ChT), ISL1, and HSP90 as loading control. 14893 (93) and 14897 (97) denote two different hairpins. (G) Morphology of day 4 Ngn2 iN cells infected with a control or two ISL1 hairpins. Shown is GFP immunofluorescence of GFP-expressing iN cells. (Scale bar: 100 μ m.) (H) Repression of cholinergic genes by ISL1 knock-down as shown by qRT-PCR of day 4 Ngn2 iN cells infected with control or two ISL1 shRNAs (N = 3, error bars represent SEM ANOVA. Exact adjusted *P*-values are shown in the graph). 14893 and 14897 denote two different hairpins.

neuroectodermal cells and then expressed NGN2 to induce their differentiation into neurons (Fig. 3A and *SI Appendix, Fig. S3A*) (33). We confirmed that anterior progenitors (SL) were positive for OTX2, and posterior neural progenitors (SLC) were positive for HOXA3 by immunofluorescence (Fig. 3B) and qRT-PCR (*SI Appendix, Fig. S3C*). We picked anterior and posterior neuroectodermal cells since NGN2 generates glutamatergic cortical projection neurons in dorsal forebrain progenitors and induces cholinergic motor neurons in ventral progenitors of the developing spinal cord (7).

All three starting populations [H9-hES, anterior (SL) or posterior (SLC) neural progenitor cells] produced mature iN cells with elaborate processes upon NGN2 induction (*SI Appendix, Fig. S3B*). To investigate how these different starting populations affect subsequent subtype specification, we performed RNA sequencing on the three starting populations (H9-hES, SL, and SLC) and their corresponding day 2 iN cells (immature) and day 28 (mature) iN cells. Principal component analysis revealed an intriguing phenomenon: The stronger first principal component (explaining 69% of the variance) expectedly indicated the cell maturation (Fig. 3C). However, the second, weaker principal component (explaining 8% of the variance) accounted for the different starting populations irrespective of differentiation stage (Fig. 3C). Accordingly, anterior genes were expressed in the anterior population and posterior genes in the posterior population independent of differentiation stage (Fig. 3D and E and *SI Appendix, Fig. S3D*). Nevertheless, we found that NGN2 induces different downstream transcription factors depending on region-specific chromatin configuration. In case of the day 2 NGN2 iN cells derived from anterior cells, anterior progenitor markers LHX2/5 and SIX3 are higher compared to that from H9 and the posterior neural stem cells; meanwhile, for day 2 NGN2 iN cells derived from posterior neural cells, different spinal cord progenitor domain markers IRX3, LBX1, and PAX2 are higher compared to that from H9 and the anterior neural stem cells

(Fig. 3F). With respect to day 28 iN cells, GO term analysis and inspection of key region-specific genes of the RNA sequencing showed that the regional identity of day 28 iN cells was well maintained (Fig. 3G and H). For example, telencephalic development genes (*SLX3* and *FEZF1*) were induced in SL-NGN2-28d iN cells, and inner ear receptor cell genes (*HEY2* and *ATOHI1*) were induced in day 28 SLC-NGN2-28d iN cells.

Next, we addressed our initial hypothesis and investigated the neurotransmitter specification. Contrary to our expectation, the regional patterning did not fundamentally affect the induction of a cholinergic program which remained partial in all three conditions (Fig. 3I). The posterior condition yielded even a decreased than increased expression of cholinergic genes (compare blue or red with green bars in Fig. 3I). Instead, excitatory genes were still prominently expressed, confirmed by RNA sequencing and immunostaining (*SI Appendix, Fig. S3E and G*). Electrophysiology confirmed exclusive generation of excitatory postsynaptic currents (EPSCs) in both SL-NGN2-28d and SLC-NGN2-28d iN cells (*SI Appendix, Fig. S3F*).

To examine whether perturbing major signaling pathways after NGN2 induction would alter neurotransmitter programs given that many signaling pathways have many different temporally defined roles during neuronal differentiation, we systematically added agonists and antagonists targeting major signaling pathways (TGF β , BMP, Wnt, FGF, RA, and SHH) after doxycycline induction for 7 d and examined the level of cholinergic genes of the H9-iN cells at day 7 (34, 35). We observed no significant upregulation or downregulation of ISL1, ChT (SLC5A7), VGLUT1 except for CHIR99021 (Wnt agonist), and trametinib (FGF antagonist) which induced immature/stressed neuronal cells and cell death, respectively (*SI Appendix, Fig. S3H*). Thus, manipulating those specific signaling pathways after neuronal induction had little effect on neuronal subtype specification.

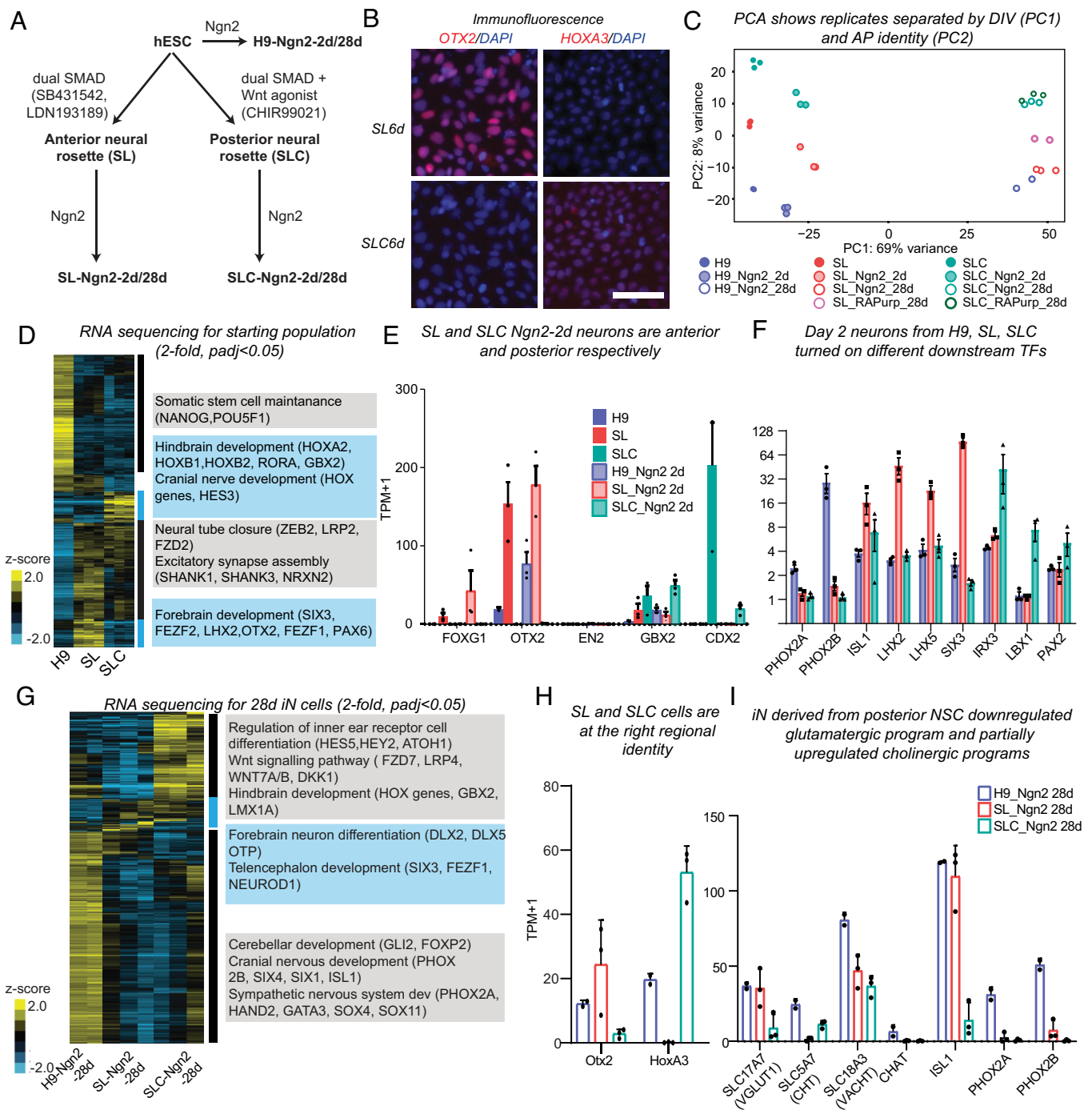


Fig. 3. Regionalization of donor cells is maintained throughout Ngn2-induced neuronal differentiation. (A) Patterning of human ES cells into anterior neural cells with SB431542 and LDN193189 (SL) and posterior neural cells with SB431542, LDN193189, and the Wnt antagonist CHIR99021 (SLC). See also *SI Appendix, Fig. S3A* for details. (B) Immunofluorescence of OTX2 (an anterior marker) and HOXA3 (a posterior marker) validating the positional identity of most anterior and posterior neural cells at day 6 postdifferentiation. (Scale bar: 50 μm .) (C) Principal component analysis of RNA sequencing of the three donor cell populations: ES cells (H9), anterior (SL), and posterior (SLC) neural progenitors and their corresponding day 2 and 28 iN cells. The first principal component separates the samples by the differentiation stage and the second principal component separates the cells derived from different starting populations. (D) Hierarchical clustering of the genes (≥ 2 -fold change and $P_{adj} < 0.05$) of the three starting populations [H9 ($n = 2$), SL ($n = 3$), and SLC ($n = 3$)]. Significant gene ontology terms and the genes contributing to them for each highlighted cluster are listed in order of the highlighted area (PANTHER, $P_{adj} < 0.05$). (E) Expression of positional identity genes (FOXG1, OTX2, EN2, GBX2, and CDX2) in the starting populations (H9, SL, and SLC) and their corresponding day 2 iN cells. Error bars = SEM. (F) Expression of different downstream transcription factors in different starting populations. Error bars = SEM. (G) Hierarchical clustering of significant genes (≥ 2 -fold change and $P_{adj} < 0.05$) of day 28 Ngn2 iN cells from the three starting populations [H9-NGN2-28d ($n = 2$), SL-NGN2-28d ($n = 3$), and SLC-NGN2-28d ($n = 3$)]. Significant gene ontology terms and the genes contributing to them for each highlighted cluster are listed in order of the highlighted area (PANTHER, $P_{adj} < 0.05$). (H) Expression of positional identity genes (OTX2 and HOXA3) in day 28 Ngn2 iN cells derived from different starting populations (H9, SL, and SLC). Error bars = SEM. (I) Expression of a glutamatergic gene (SLC17A7), cholinergic genes (SLC5A7, SLC18A3, and CHAT), and cholinergic transcription factors (ISL1, PHOX2A, and PHOX2B) in day 28 Ngn2 iN cells derived from the three starting populations (H9, SL, and SLC). Error bars = SEM.

Chromatin Configurations Mildly Affect the Genomic Binding of NGN2. We next sought to molecularly explain the context-specific effects of NGN2 and asked whether the chromatin state affects

the physical binding of NGN2. We performed ChIP-seencing of NGN2 and found 2,625, 706, and 770 high-confidence peaks in H9, SL, and SLC cells, respectively, 2 d after NGN2 induction.

Removing the peaks that are also present in rtTA control, we obtained 2018 sites among H9-NGN2-2d, SL-NGN2-2d, and SLC-NGN2-2d. Unexpectedly, while there was a large degree of overlap, NGN2 binding was most widespread when induced in undifferentiated ES cells and more restricted in differentiated population (SL and SLC) and clearly distinct between the three cell types (Fig. 4A). The bHLH motif (CANNTG) was most significantly enriched in all cell types (E-values: 1.0E-676, 1.9E-682, and 8.9E-319, respectively) (Fig. 4B and C), with most peaks harboring more than one bHLH motif (SI Appendix, Fig. S4A). Analyzing different types of E-boxes, we found a preference for CAGATG E-box motif in H9 cells and a preference for CAGCTG motifs in the SL and SLC cells (SI Appendix, Fig. S4B and D). When we compared the peak classification in all three conditions, most of the peaks were in intergenic and intronic regions and only a small subset of the peaks in promoter regions (SI Appendix, Fig. S4C).

Given the differential binding of NGN2, we next sought to better characterize these differences. Hierarchical clustering of NGN2 peak intensity (± 50 bp from the peak summits) showed that in fact most sites (about 60%) are commonly bound by NGN2 in all three cell types (Fig. 4A, marked by black stripe). These common peaks are adjacent to genes that are enriched in GO terms such as Notch binding and regulation of neurotransmitter levels. The remaining 40% sites comprise five clusters that show preferential binding in one or two cell types. Notably, most of such sites comprising the three largest clusters show NGN2 binding in H9 cells with coenrichment in SLC (Fig. 4A, turquoise cluster), coenrichment in SL (Fig. 4A, blue cluster), or no coenrichment in either SL or SLC (Fig. 4A, orange cluster). Only few sites are depleted in ES cells and enriched in both SL and SLC (Fig. 4A, red cluster) or enriched in SLC only (Fig. 4A, yellow cluster). Thus, unlike ASCL1, NGN2's genomic binding is context dependent, still the majority (60%) of sites are shared among three different cell types.

When we performed differential NGN2 binding analysis among the three samples, we found that 186, 509, and 322 peaks were differentially occupied when comparing H9-NGN2 vs. SL-NGN2, H9-NGN2 vs. SLC-NGN2, and SL-NGN2 vs. SLC-NGN2 (FDR < 0.1). For example, NGN2 binds in all three conditions in the proximity of HES6 while it binds to the distal region of the *LHX3* promoter only in the SLC chromatin environment (SI Appendix, Fig. S4E and F).

NGN2 Directly Binds and Regulates *ISL1* and Cholinergic Genes.

The prominent induction of *ISL1* was noteworthy since *ISL1* is expressed only in a subpopulation of NGN2-expressing cells during development. Nevertheless, we found NGN2 bound at the *ISL1* and the combined *CHAT* and *vAChT* (*SLC18A3*) loci (Fig. 4D, right traces). NGN2 binding strength correlated with the expression level of *ISL1*, *CHAT*, and *vAChT* (*SLC18A3*) between the three cell types (Fig. 4D, compare expression left with binding right).

A Combination of Chromatin Accessibility and Signaling Pathway-Induced Transcription Factors May Guide NGN2 Binding.

To understand how chromatin accessibility might affect Ngn2 binding, we performed ATAC sequencing on the three starting populations. We found that there are in total 57,703 sites that are differentially accessible among the three conditions (SI Appendix, Fig. S5H). We next asked whether the differential chromatin accessibility may explain NGN2 binding. Indeed, in many cases, we found an overall correlation between NGN2 binding strength and the ATAC signal (Fig. 4A). For instance, regions strongly bound in all three cell types are generally more accessible (*large parts of the black cluster*). In those clusters that are primarily bound in the neural SL or SLC cells but not ES cells (*yellow and red clusters*), the degree

of chromatin accessibility correlated well with NGN2 binding. In those cases, the process of neuralization may have opened the chromatin configuration allowing NGN2 access.

However, other regions (*green and blue clusters*) cannot be explained by differential chromatin accessibility (Fig. 4A) nor the types of E-boxes alone (SI Appendix, Fig. S4D). We thus performed motif enrichment analysis on these two clusters and found an additional enrichment of the ZNF281 motif and SMAD4 motif adjacent to the bHLH motif in the H9-NGN2 and SL-NGN2 specific peaks (*blue*) and H9-NGN2 and SLC-NGN2 (*green*), respectively (Fig. 4C, **primary motif**). It was previously reported that both TCF and ZNF281 motifs are coenriched in β -catenin ChIP-seq, suggesting ZNF281 might be downstream of Wnt signaling (36). This shows that the downstream effectors TGF β /BMP and WNT (SMAD4 and ZNF281, respectively) may guide NGN2 to these differential sites even though they are in a less accessible state.

The Forebrain Homeobox Factors EMX1 and FOXG1 Cooperate with NGN2 to Induce Forebrain Excitatory Neurons.

Given our inability to accomplish a better cell fate restriction via chromatin pre patterning, we next took a fundamentally different approach and coexpressed the fore-/midbrain transcription factors EMX1, EMX2, OTX1, OTX2, TBR2, LHX2, and FOXG1 with NGN2 in undifferentiated cells (Fig. 5A and SI Appendix, Fig. S5A). Those transcription factors were selected based on their expression during cortical development (37–42). All transcription factors, except OTX1 and TBR2, produced iN cells in combination with NGN2 (SI Appendix, Fig. S5A and C). We first used *ISL1* expression to assess whether the addition of forebrain transcription factors would focus the neurotransmitter program of NGN2-only iN cells. The additional expression of EMX1, EMX2, and FOXG1, but not OTX2, greatly reduced the percentage of *ISL1*-positive neurons from around 80% in NGN2 to 10 to 20% in NGN2 with EMX1, EMX2, and FOXG1 on day 28 after infection (Fig. 5B and SI Appendix, Fig. S5D). Among all conditions, 93 to 97% of iN cells expressed the excitatory marker vGLUT (SI Appendix, Fig. S5B and C).

To understand what the transcriptional changes after the overexpression of individual transcription factors were, we performed RNA sequencing using day 28 induced neurons generated from NGN2 alone or with EMX1, EMX2, and FOXG1 OTX2 (Fig. 5C and D). Importantly, we found that all three cholinergic genes [*ChT* (*SLC5A7*), *vAChT* (*SLC18A3*), and *CHAT*] as well as *ISL1*, *PHOX2B* were repressed in NGN2:EMX1, NGN2:EMX2, and NGN2:FOXG1 compared to NGN2 or NGN2:OTX2 (Fig. 5E). This result was independently confirmed by qRT-PCR in another cell line (SI Appendix, Fig. S5E and H).

Other neurotransmitter genes were also repressed by EMX1, EMX2, and FOXG1 (SI Appendix, Fig. S5I). GAD1, GAD2, and vGAT (*SLC32A1*), the three genes critical for GABAergic identity, were even more reduced by EMX1&2 than in NGN2 only cells. FOXG1 repressed both GAD1&2 but left vesicular GABA transporter (vGAT/*SLC32A1*) unchanged (SI Appendix, Fig. S5I). This finding is compatible with the observation from scRNA-seq performed by Allen Brain Atlas that EMX1 is only expressed in excitatory neurons whereas FOXG1 is expressed in both excitatory and inhibitory neurons (SI Appendix, Fig. S5F) (31). Intriguingly, *EMX1* was properly induced in anterior progenitor cells, but was rapidly down-regulated during neuronal differentiation, explaining the lack of cholinergic gene repression (SI Appendix, Fig. S5G).

Finally, we sought to functionally characterize the excitatory cells. While EMX1, EMX2, OTX2, and FOXG1 are well-characterized developmental regulators, their expression in adult

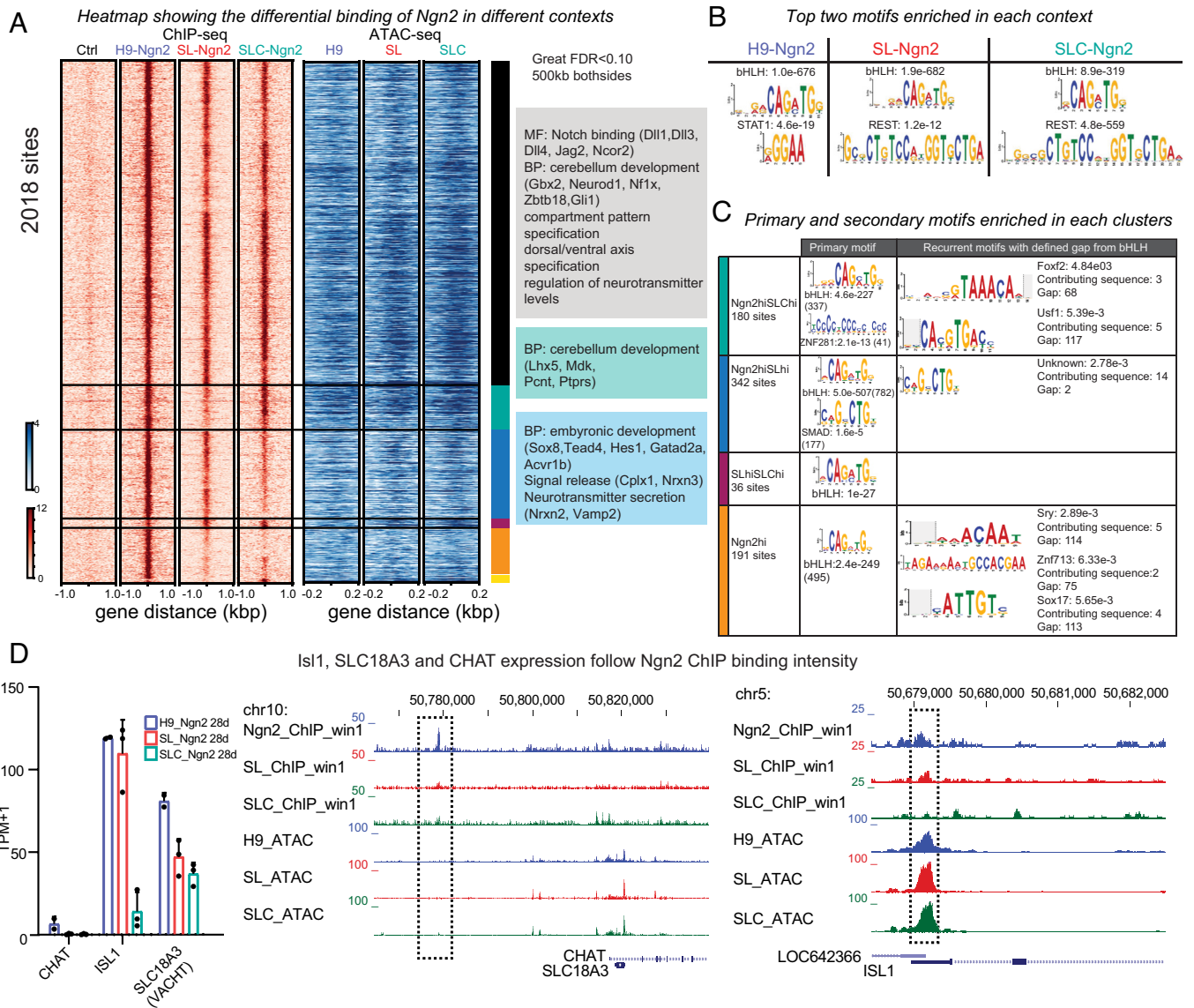


Fig. 4. Genomic binding of NGN2 is chromatin dependent. (A) *Left (red heatmap):* NGN2 ChIP-sequencing profile 2 d after infection with Ngn2 in three different chromatin states (H9, SL, and SLC) or with rtTA only infection (Ctrl) ($n = 2$). Corresponding peaks are displayed ± 1 kb from the peak summit. After removing peaks called in the control condition (rtTA only), we obtained 2,018 significant peaks ($\text{idr} < 0.10$) which includes peaks that are significant in at least one out of the three conditions (H9-Ngn2, SL-Ngn2, and SLC-Ngn2). *Right (blue heatmap):* ATAC sequencing profile for the three starting populations (H9, SL, and SLC) for the corresponding NGN2 ChIP-seq regions. The gene ontology terms were inferred using GREAT using genes within 500 kb from the peaks (FDR P -value < 0.10). (B) Top two motifs significantly enriched in the three populations (H9, SL, and SLC) using sequences ± 50 bp from the peak summit. (C) Primary and secondary motifs with defined gaps from the primary motif significantly enriched in the sequences within ± 250 bp from the peak summit in four clusters (H9-NGN2^{high}SL-NGN2^{high}, H9-NGN2^{high}SL-NGN2^{high}, SL-NGN2^{high}SLC-NGN2^{high}, and H9-NGN2^{high}) highlighted in A. For each secondary motif, the E-value, the number of contributing sequences and the gap between the primary and the secondary motif are listed to the right. The number of sites with the motif was included in brackets. (D) *Left:* RNA-seq expression values (TPM+1) for the day 28 iN cells from the three starting populations (H9, SL, and SLC) for *CHAT*, *ISL1*, and *SLC18A3*. Error bars = SEM. *Right:* Genomic tracks showing the NGN2 ChIP- and ATAC-seq signal at the *CHAT/SLC18A3*, and *ISL1* loci. Note the correlation between the NGN2 ChIP peak heights and expression of corresponding genes shown on the *Left*.

human cortical excitatory neurons is less known. Analyzing data from the Allen brain atlas showed that OTX2 is not expressed well in the adult cerebral cortex but FOXG1 is prominently expressed throughout all neural cell types except oligodendrocytes, EMX1 is quite restricted to excitatory neurons and EMX2 is most strongly expressed in astrocytes (*SI Appendix, Fig. S5F*) (43). Thus, among those 4 genes, only EMX1 and FOXG1 are expressed in cortical excitatory neurons. Electrophysiology confirmed that all iN cells analyzed from the three groups exhibited miniature EPSCs but no IPSCs (Fig. 5F). The EPSC amplitudes were similar between NGN2 and NGN2:EMX1 and slightly decreased in NGN2:FOXG1 iN cells (Fig. 5G). Other intrinsic membrane parameters were similar between the different conditions (Fig. 5G).

EMX1 and FOXG1 Induce Widespread Genomic Redistribution of NGN2. To explore the mechanisms how EMX1 or FOXG1 restrict NGN2's ability to induce neuronal subtypes, we thought of two non-mutually exclusive principal possibilities: i) EMX1/FOXG1 act independently of NGN2 and influence gene expression in an additive manner or ii) they could directly influence the targeting of NGN2 to the chromatin. To distinguish between these two possibilities, we expressed a FLAG-tagged version of NGN2 in hES cells and performed FLAG antibody ChIP-sequencing with and without coexpression of EMX1 and FOXG1 (Fig. 6A). Unlike NGN2 binding in the anterior and posterior neuroectoderm, the NGN2 binding patterns were quite distinct between the three conditions with 1,603, 2,701 and 1,525 condition-specific

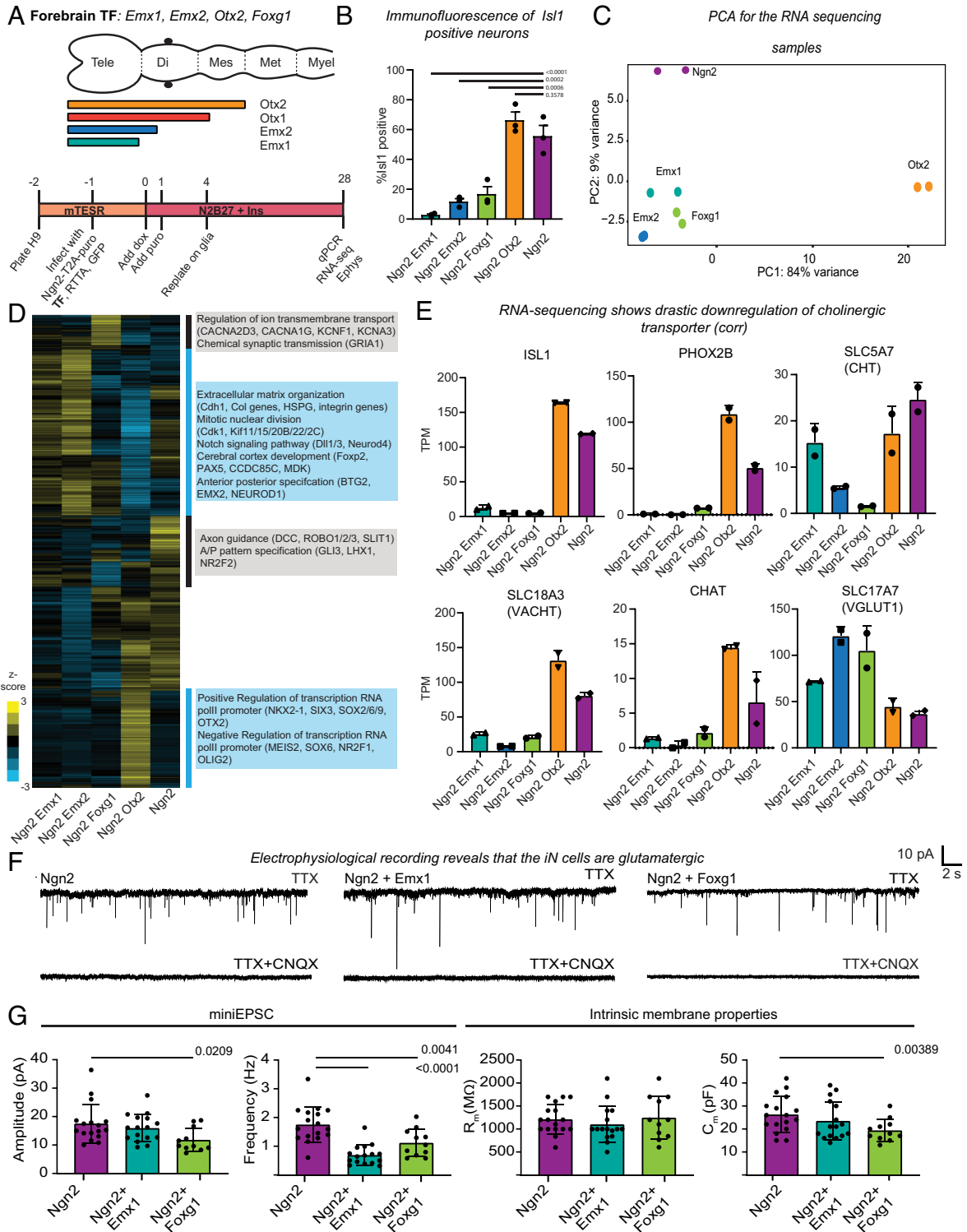


Fig. 5. EMX1 and FOXG1 eliminate heterogeneity of Ngn2 iN cells. (A) *Top*: Nested expression pattern of OTX2, OTX1, EMX2, and EMX1 during development (Tele = Telencephalon, Di = Diencephalon, Mes = Mesencephalon, Met = Metencephalon and Mye = Myelencephalon). *Bottom*: Protocol to generate Ngn2 iN cells coexpressing individual candidate forebrain transcription factors (TF). (B) Quantification of ISL1+ cells among all GFP/NGN2-infected iN cells on day 28. Representative fluorescence images are shown in Fig. 5 SD. (N = 3, error bars = SEM. ANOVA statistical test was used. Exact adjusted P-values are shown above graph). (C) Principal component analysis of NGN2, NGN2:EMX1, NGN2:EMX2, NGN2:OTX2, and NGN2:FOXG1. The first principal component separates NGN2:OTX2 from the other samples. All three NGN2:EMX1, NGN2:EMX2 and NGN2:FOXG1 clustered together. (D) Hierarchical clustering of significantly changing genes in RNA-seq (≥ 2 -fold change and $P_{adj} < 0.05$) within any two of the day 28 iN cells generated with NGN2, NGN2:EMX1, NGN2:EMX2, NGN2:FOXG1, or NGN2:OTX2. Significant gene ontology terms for the corresponding highlighted regions are listed ($P < 0.05$). (E) Bar graphs showing expression values of RNA-seq (TPM) for the two cholinergic transcription factors (ISL1 and PHOX2B), three cholinergic genes (SLC18A3, SLC5A7, and CHAT) and the glutamatergic marker (SLC17A7) of day 28 iN cells generated with NGN2 alone or in combination with EMX1, EMX2, FOXG1, or OTX2. (F) Representative traces of miniature excitatory postsynaptic currents (EPSCs) for day 28 NGN2-, NGN2:EMX1-, and NGN2:FOXG1- iN cells in the presence of the voltage-gated Na⁺-channel blocker TTX to block action potential formation and network activity (*Top* trace). All synaptic events observed are eliminated after addition of the excitatory AMPA receptor inhibitor CNQX (*Bottom* trace). (G) Quantification of miniature EPSC amplitude and frequency and the two intrinsic membrane properties input resistance and capacitance of NGN2, NGN2:EMX1, and NGN2:FOXG1 iN cells (P -values using the ANOVA test are shown above bars. N = 17, 16, 11 cells measured in three independent batches in NGN2, NGN2:EMX1, and NGN2:FOXG1 28d iN cells plated on glia. Error bars = SEM).

binding sites (FDR < 0.1) (Fig. 6A). For example, the distal region of the FOXO6 locus is bound by NGN2 in ES cells only when coinfecting with FOXG1 (SI Appendix, Fig. S6D, green highlight) whereas the NGN2 binding site at the HES6 locus is bound in all conditions (SI Appendix, Fig. S6E, purple highlight). The change in NGN2 binding appears to be meaningful as sites unique in NGN2:FOXG1- and NGN2-only-expressing cells were enriched for relevant GO terms (Fig. 6A).

Next, we sought potential explanations for the differential NGN2 binding. To understand whether differences in binding among the groups can be explained by chromatin accessibility alone, we plotted the ATAC sequencing signal in hES cells centering at the flagNGN2 peak summits and found that the different clusters are all similarly accessible (Fig. 6A, blue). Thus, the unique NGN2 binding sites in NGN2:EMX1 and NGN2:FOXG1 are not simply due to inaccessibility in ES cells.

We then performed motif and peak distribution analysis of the NGN2 peaks. In all NGN2 ChIP-seq datasets a very similar bHLH motif (CAGATG) was enriched and most NGN2 peaks harbored at least one bHLH motif and often multiple (Fig. 6A and SI Appendix, Fig. S6A). When we further subdivided bHLH motifs enriched in the NGN2:EMX1 and NGN2:FOXG1 cells, we found that NGN2:EMX1 cells had a higher percentage of the CAGCTG E-box motif and NGN2:FOXG1 cells had a higher percentage of the CAGATC E-box motif like NGN2 alone (Fig. 6C and SI Appendix, Fig. S6B). The peak distribution was similar between the samples (SI Appendix, Figs. S4C and S6C).

The differential E-Box enrichment among different NGN2 peaks suggested potentially additional sequence similarities among the different NGN2-bound clusters. Indeed, unbiased de novo motif search analysis also produced the canonical FOXG1 motif among the NGN2:FOXG1-unique NGN2 binding sites (Fig. 6B). Reverse motif search showed that the *Emx1* motif was four times more enriched under the NGN2 peaks specific in the NGN2:EMX1 cells and the *Foxg1* motif was eight times more enriched within the NGN2 peaks specific for the NGN2:FOXG1 cell compared to the peaks specific for the NGN2-only infected cells (Fig. 6C). Remarkably, the *Emx1* motif was also more enriched in NGN2:FOXG1 specific peaks and the *Foxg1* motif in NGN2:EMX1 specific peaks. Even though to date, EMX1 or FOXG1 are not known to physically interact with NGN2, these data suggest that EMX1 and FOXG1 may recruit NGN2 to their own binding sites, even when they contain less preferred E-box motifs.

To explore this idea further, we performed ChIP-sequencing for flagEMX1 in ES cells coinfecting with flagEMX1 and NGN2. We obtained 1,393 significant peaks ($n = 2$, $\text{idr} < 0.10$) (Fig. 6D) and most peaks contained a homeobox motif (SI Appendix, Fig. S6G and H). The peak classification distribution of EMX1 was slightly different than NGN2 with a much-reduced promoter region localization (SI Appendix, Fig. S6I, compared to SI Appendix, Fig. S6C). When plotting the flagEMX1 ChIP-seq signal from genomic sites that are bound by NGN2, we could indeed see an enriched EMX1 binding at NGN2 peaks unique in the NGN2:EMX1 infected cells (dark green cluster in Fig. 6A). This finding demonstrates that EMX1 recruits NGN2 to new genomic sites.

EMX1 Represses Posterior and Cholinergic Genes Independent of NGN2. We next explored whether EMX1 may have NGN2-independent functions, in addition to directly influencing NGN2 chromatin binding. We were wondering whether EMX1 may repress the promiscuous activation of neurotransmitter programs. Therefore, we first explored the direct transcriptional effects of EMX1. We identified the putative EMX1 target genes (genes within 10 Kb of EMX1 binding) and plotted their average expression level in ES cells

infected with NGN2 or NGN2:EMX1. We found a lower average fold change and narrower distribution of fold change of these EMX1 target genes when the cells were coinfecting with EMX1 compared to NGN2 alone (Fig. 6E). This indicates that EMX1 might have repressive functions. Repressed genes include posterior genes like HOX genes, ISL1, and PHOX2A/B genes, involved in cranial nerve, hindbrain formation, and spinal and sympathetic neuron development. On the other hand, anterior genes involved in forebrain development were up-regulated upon EMX1 coexpression (Fig. 6F). In support of this notion, we found that EMX1 binds the *ISL1* locus when expressed in ES cells (Fig. 6D). Thus, EMX1 acts independently of NGN2 to repress posterior genes but also influences NGN2 binding to achieve a forebrain glutamatergic neuronal identity.

Discussion

Previously, we found that NGN2 overexpression in human ES cells gave rise to induced neuronal cells that exclusively form glutamatergic functional synapses (25). Despite this perceived homogeneity with respect to neurotransmitter specification, our single-cell characterization of NGN2-iN cells revealed glutamatergic neurons with variable and partial expression of cholinergic and monoaminergic programs. The cells can be grouped into i) ISL1⁺PHOX2B⁺, ii) ISL1⁺PHOX2B⁻, and iii) ISL1⁻PHOX2B⁻ glutamatergic neurons. We observed that key cholinergic effector ISL1 is directly bound and induced by NGN2 in human ES cells (Fig. 4D). Thus, to achieve better subtype specification, NGN2 chromatin binding needs to be better targeted to a single program. Since NGN2 operates in different, regionalized neural progenitors during normal development, one would expect that the chromatin strongly influences NGN2 binding (7). Thus, we first tested the hypothesis that changing the chromatin landscape by differentiating ES cells into anterior and posterior neural progenitor cells may accomplish a more proper NGN2 relocalization, thus eliminating unwanted neurotransmitter programs (7). However, we found that NGN2 binding is only relatively mildly changed (Fig. 4D). NGN2 was further restricted in different neural progenitors but was not recruited to new sites. Some of the lost binding sites could be attributed to chromatin accessibility, others possibly to interaction with downstream effectors of TGF β /BMP and WNT pathways (Fig. 4A and C). Accordingly, the cholinergic and monoaminergic programs were modified but not eliminated. It is certainly possible that further refinements in chromatin prepatterning protocols may further restrict neuronal subtype specification, but generation of more defined progenitor cells in homogeneity may be challenging but if accomplished eliminate the advantage of transcription factor programming (44).

Unlike neurotransmitter phenotype specification, we found that the regional identity of mature iN cells is maintained from the starting population (Fig. 3E and H). This finding is significant as it opens the NGN2 iN cell platform to generate glutamatergic neurons of different regional identities by merely changing the starting population and will likely extend to other reprogramming factors, as they are not known to influence regional identity (16, 22, 45) (SI Appendix, Fig. S4G). Also, those NGN2 iN cells of different regional identities later up-regulated different downstream TF expressions. For example, forebrain-specific transcription factors (LHX2) were up-regulated 16-fold in the iN cells derived from anterior neural stem cells versus control ES or posterior neural stem cells and spinal cord transcription factor (IRX2) was up-regulated eightfold in the iN cells derived from posterior neuronal stem cells relative to those from the other two (Fig. 3F).

Second, we overexpressed forebrain-specific transcription factors with NGN2 hypothesizing that they might reinforce forebrain

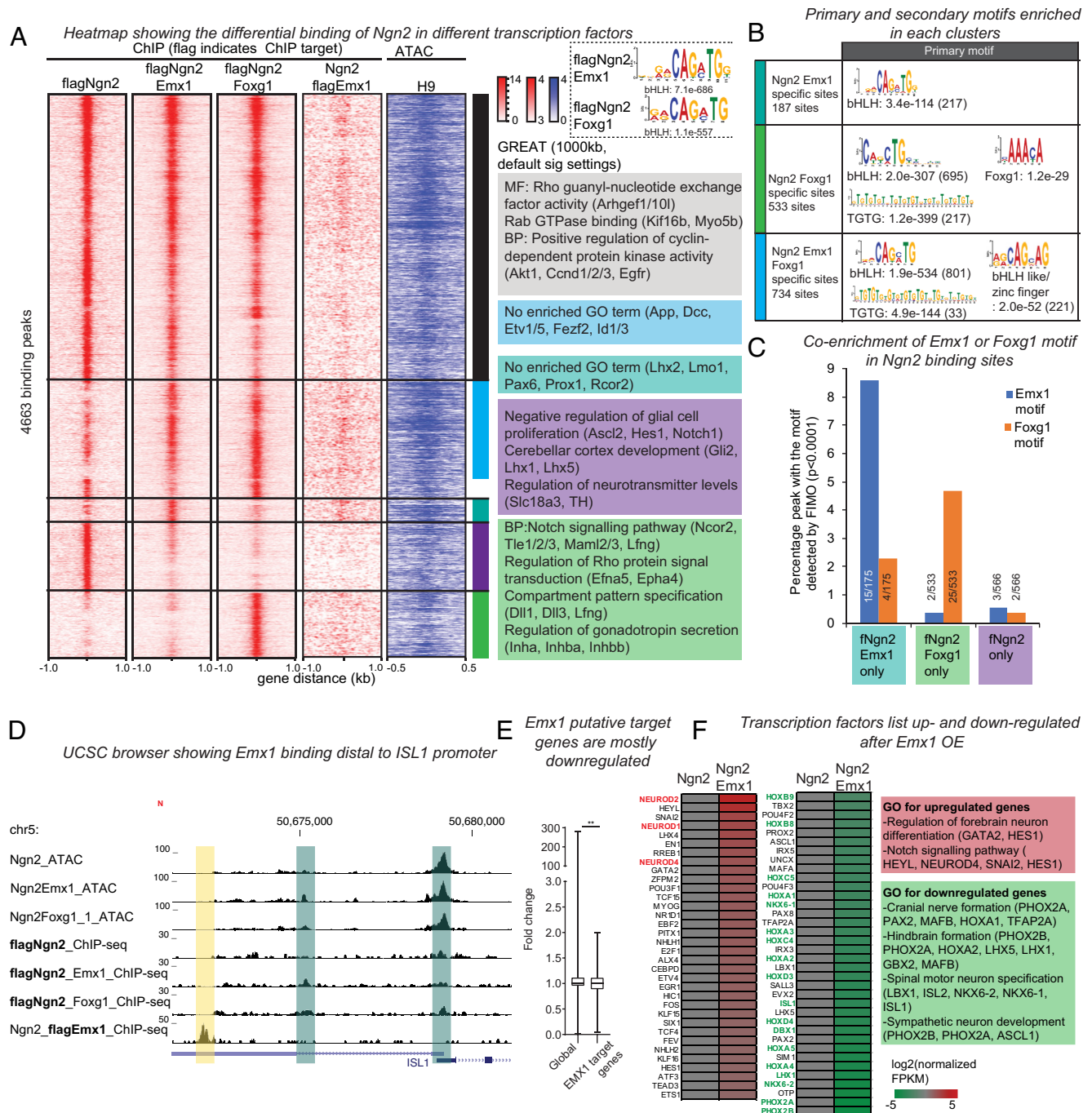


Fig. 6. EMX1 and FOXG1 change NGN2 chromatin targeting. (A) *Left three columns:* flagNGN2 ChIP-seq profile within ± 1 kb from the peak summit (red signal) in hES cells 2 d after infection with flagNGN2 alone or with EMX1 or FOXG1 ($n = 2$). We obtained 4,663 peaks ($\text{idr} < 0.10$) which includes peaks that are significant in at least one out of the three conditions (flagNgn2, flagNgn2 Emx1, and flagNgn2 Foxg1) 4th *column:* Corresponding genomic regions as in left 3 columns showing the flagEMX1 ChIP-seq profile 2 d after infection with flagEMX1 and NGN2. All ChIP-seq peaks are displayed ± 1 kb from the peak summit. 5th *column (blue signal):* Corresponding genomic regions showing ATAC-seq signal in ES cells (H9). The ATAC-seq peaks are displayed ± 500 bp from the respective flagNGN2 ChIP peak summit. Gene ontology terms for various genomic clusters called significant by GREAT analysis are shown using genes within 1,000 kb from the peaks. Blue and turquoise clusters did not significantly enrich any GO term. *Dotted box:* Position weight matrix of the bHLH motif enriched by de novo motif search analysis of the flagNGN2 ChIP-seq of flagNGN2:EMX1 and flagNGN2:FOXG1 infected cells considering all significant peaks (± 50 bp from the peak summit). (B) The top motif significantly enriched by motif search analysis of three clusters color-indicated in A: NGN2 peaks specific to NGN2:EMX1 infected cells (turquoise cluster); NGN2 peaks specific to NGN2:FOXG1 infected cells (green cluster); NGN2 peaks specific to NGN2:EMX1 and NGN2:FOXG1 infected cells (blue cluster). Sequences ± 500 bp from the peak summit were used. The number of sites with the motif was included in brackets. (C) Reverse motif search for EMX1 and FOXG1 motifs in three different clusters showing differential NGN2 binding: NGN2 peaks unique to cells infected with flagNGN2 and EMX1 (turquoise cluster); NGN2 peaks unique to cells infected with flagNGN2 and FOXG1 (green cluster); and NGN2 peaks unique to cells infected with flagNGN2 alone (purple cluster). Shown are percentages of motifs among the total number of genomic sites in the respective cluster. Numbers in bars show the actual numbers. The positional weight matrix used for EMX1 and FOXG1 motifs was obtained from Jaspur. (D) Genomic browser tracks showing the ATAC-seq (day 2 NGN2, NGN2:EMX1, and NGN2:FOXG1 cells) and ChIP-seq of against flag tagged NGN2 (day 2 NGN2, NGN2:EMX1, and NGN2:FOXG1) and ChIP-seq against flag tagged EMX1 in day 2 NGN2:EMX1. Note that there is an EMX1 peak upstream of the *ISL1* promoter (yellow) upstream of two Ngn2 peaks (green). (E) Boxplot showing the expression fold change of all genes and predicted Emx1 target genes by RNA-seq in NGN2 vs. NGN2: EMX1 infected cells. Plotted are $\text{TPM}+1_{\text{NGN2:EMX1}}/\text{TPM}+1_{\text{NGN2}}$ of all genes and genes within 10 kb of an EMX1 peak. Predicted EMX1 target genes are significantly repressed (average fold change of all genes=1.01; average fold change of predicted EMX1 targets=0.89; P value = 0.0027). (F) Heatmap showing transcription factors regulated by EMX1 in the context of NGN2 expression. Boxes to the right show the GO terms enriched in either group and the genes included in the GO terms.

glutamatergic targets and redistribute NGN2 away from cholinergic targets. Even though sequence-specific transcription factors are not known to generally interact with many other transcription factors, we reasoned that they could potentially modulate NGN2 binding. Indeed, among five factors tested, EMX1 and FOXG1 blocked cholinergic gene expression by 50 to 90% and altered NGN2 chromatin targeting (40) (Figs. 5E and 6A). There are three times more differential NGN2 binding in NGN2:EMX1 and NGN2:FOXG1 than that with patterning (Figs. 4E and 6B). Unlike patterning, significantly more NGN2 sites were gained (Figs. 4A and 6A). NGN2 gained sites were enriched for EMX1 or FOXG1 motifs and enriched for EMX1 binding demonstrating that NGN2 is recruited to new chromatin targets likely by EMX1 and FOXG1 (Fig. 6D). These findings are compatible with the notion that a supportive transcription factor milieu at enhancers may facilitate transcription factor recruitment to new sites (46, 47).

In addition to influencing NGN2 binding, we observed that EMX1 also has NGN2-independent functions. During mouse development, EMX1 expression is induced in the anterior neuroectoderm at around E9.5 (39, 48). It was previously shown in knockout studies that EMX1 knockout mice exhibited corpus colosum agenesis. EMX1/EMX2 double knockout mice exhibited more severe phenotypes than either of the individual knockouts with multiple features of cortical development impacted. This suggests that EMX1 and EMX2 cooperatively regulate cortical development and EMX2 can partially compensate for EMX1 function in EMX1 knockout mice (49). On average, we found that EMX1 repressed its direct target genes which included many nonglutamatergic and posterior genes (Fig. 6F and *SI Appendix, Fig. S6J*). From single-cell sequencing data, only EMX1 remains expressed primarily in glutamatergic neurons and astrocytes in the cortex, hippocampus, and olfactory bulb of the adult brain (*SI Appendix, Fig. S5F*). These observations agree with EMX1 having the role of a “spatial selector” which is defined as a factor that instructs/restricts the regional identity of multipotent progenitors and later limits the subtypes of mature neurons generated from that progenitor domain. This function resembles the “many-but-one” cell type specification function of MYT1L. Just like MYT1L limits the programs of many non-neuronal lineages, EMX1 represses nonglutamatergic neurotransmitter programs to ensure that the cortical excitatory neurons retain the glutamatergic program throughout adulthood (50). In perfect agreement, neurodevelopment studies already demonstrated a transcriptional repressor function of EMX1 such as repressing posterior genes. The reduction in Wnt signaling in the rostral forebrain by SIX3 leads to the expression of FOXG1 and EMX1 which counteracts the expression of OTX2 (51). Frowein and coworkers reported that ectopic expression of EMX1 instructed a neuroepithelial identity instead of choroid plexus identity and that coincided with the downregulation of OTX2 (52).

FOXG1 expression begins around E8.5 in the telencephalic primordium which later becomes dorsal and ventral telencephalon (41). The dorsal telencephalon includes the neocortex and hippocampus and generates predominantly glutamatergic neurons. Ventral telencephalon includes medial, lateral, and caudal ganglionic eminences and generates various populations of mostly GABAergic neurons in the cortex, striatum, basal ganglia, and olfactory bulb. It may therefore not be surprising to see an additional incomplete GABAergic program, indicated by the SLC32A1 (or VGAT) expression, in the NGN2 FOXG1 day 28 iN cells. Given that *EMX1* and *FOXG1* remain to be expressed in adult forebrain projection neurons, we reckon that in an in vitro setting EMX1 and FOXG1 might also have terminal selector functions to actively promote the correct neurotransmitter genes and limit the promiscuous neurotransmitter genes (Fig. 6F and *SI Appendix, Fig. S6J*) (53, 54).

The human brain contains innumerable neuronal subtypes. Attempting to develop conventional differentiation protocols to generate them from pluripotent stem cells may seem daunting. The use of proneuronal transcription factors offers a versatile opportunity to generate neurons in a defined and reproducible manner but their authenticity and best approach to yield pure subtypes has not been established yet. In this paper, we outlined the input-output codes of the proneuronal factor NGN2 by modulating the chromatin landscape or by pairing with subtype-specific homeodomain transcription factors. Our results revealed that transcription factor coexpression in undifferentiated cells resulted in better neurotransmitter specification than prepatterning the chromatin, e.g., we found NGN2:EMX1 as powerful combination to efficiently induce pure and functionally mature glutamatergic forebrain neurons. By extrapolation, these data predict that many different defined neuronal subtypes can be generated by combining proneuronal bHLH factors like NGN2 with regionally restricted homeobox factors. Thus, our results provide a blueprint and molecular rationale for the educated generation of specific neuronal subtypes from pluripotent stem cells with defined regionalization and neurotransmitter phenotype further generalizing the utility of NGN2 iN cell protocols (17).

Materials and Methods

Reprogramming of Human Embryonic Stem Cells to Induced Neurons (iN).

The experiments were performed in accordance with California State Regulations, CIRM Regulations and Stanford's Policy on Human Embryonic Stem Cell Research. We followed the protocol previously described (19). Human embryonic stem cells (H1 and H9, University of Wisconsin) and an iPS KOLF2.1J line (55) were plated single cell in a serum-free and defined mTESR media and infected the next day with TetO-FLAG-NGN2-T2A-PURO^R and FUW-rTA. Doxycycline was added to the wells the next day. To select for only NGN2 transducing cells, puromycin (final concentration: 2 µg/mL, Sigma) was added in addition to doxycycline the next day and kept for 3 d. For prolonged culture, the cells were dissociated using Accutase and replated on mouse glia at day 4. For single-cell RNA sequencing, doxycycline was added to 14 d and removed for the last 14 d.

For the knock-down experiment, shRNAs obtained from Sigma (ISL1: TRCN0000014893 and TRCN0000014897; PHOX2B: TRCN0000358499 and TRCN0000358500) were packaged into lentiviruses and coinfecting with TetO-FLAG-NGN2-T2A-BLAST^R and FUW-rTA. Doxycycline was added to the wells the next day. To select for only NGN2 transducing cells, puromycin (final concentration: 2 µg/mL, Sigma) and blasticidin (final concentration: 10 µg/mL, Sigma) were added in addition to doxycycline the next day and kept for 3 d.

For the forebrain transcription factors experiments (EMX1, EMX2, FOXG1, and OTX2), they were first cloned into TetO-IRES-HYGRO^R plasmid and coinfecting with TetO-FLAG-NGN2-T2A-PURO^R and FUW-rTA. In the case of the forebrain transcription factor experiment, hygromycin (150 µg/mL, Roche) was added to select for the additional transcription factor.

For all NGN2 ChIP-sequencing experiments (NGN2 in different chromatin landscapes and NGN2 with EMX1/FOXG1), FUW-rTA and TetO-FLAG-NGN2-T2A-PURO^R were used. For the EMX1 ChIP-sequencing experiment, FUW-rTA, pTight-NGN2-PGK-puro, and TetO-flagEMX1-IRES-HYGRO^R were used instead. The cells were dox induced for 2 d before they were harvested and used for ChIP-sequencing. Differentiation, sequencing library preparation and analysis, western blotting, qRT-PCR, immunofluorescence, and electrophysiology protocols are available in *SI Appendix*.

Data, Materials, and Software Availability. scRNA sequencing for H9 cells used: SRR2977655, SRR2977656, SRR2977657, SRR2977658, SRR2977659, SRR2977660, SRR2977661, SRR2977662, SRR2977663, SRR2977664, and SRR2977665 are available through GEO accession number GSE75748 (<https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE75748>) (27). The datasets generated during and/or analyzed during the current study are available in the GEO repository under the accession number GSE181019 (56).

ACKNOWLEDGMENTS. This study was supported by a training grant from the California Institute of Regenerative Medicine (CIRM, TGR-01159), Siebel Foundation and Stanford Graduate Fellowship to C.E.A. Q.Y.L. was supported by the Singapore Agency for Science, Technology and Research (A*STAR). This project was also supported by a New York Stem Cell Foundation Robertson award (to M.W.), a Howard Hughes Medical Institute Faculty Scholar award (to M.W.), and a Tashia and John Morgridge Faculty Scholar award (to M.W.) by the Child Health Research Institute (CHRI) at Stanford. K.V. is a Bertarelli Fellow through The Foundation of Bertarelli Graduate Fellowship Fund, a Rita Levi Montalcini Fellow through Fondazione Dompé, and was supported by T32 training grant 5T32MH020016-25 through the interdepartmental

Stanford Neuroscience PhD program. We thank members of the Wernig lab and Kyle Loh for insightful discussions. The sequencing in the paper was performed with help from the Stanford Functional Genomic Facility and Stanford PAN facility.

Author affiliations: ^aDepartment of Bioengineering, Stanford University, Stanford, CA 94305; ^bDepartment of Pathology, Stanford University, Stanford, CA 94305; ^cInstitute of Stem Cell and Regenerative Medicine, Stanford University, Stanford, CA 94305; ^dHHMI, Stanford University, Stanford, CA 94305; ^eDepartment of Molecular and Cellular Physiology, Stanford University, Stanford, CA 94305; ^fWellcome Sanger Institute, Hinxton, Cambridgeshire CB10 1SA, United Kingdom; and ^gOpenTargets, Hinxton, Cambridgeshire CB10 1SA, United Kingdom

- D. W. Allan, S. Thor, Transcriptional selectors, masters, and combinatorial codes: Regulatory principles of neural subtype specification. *Wiley Interdiscip. Rev. Dev. Biol.* **4**, 505–528 (2015).
- F. Guillemot, Spatial and temporal specification of neural fates by transcription factor codes. *Development* **134**, 3771–3780 (2007).
- A. Kicheva, J. Briscoe, Control of tissue development by morphogens. *Annu. Rev. Cell Dev. Biol.* **39**, 91–121 (2023).
- I. Holguera, C. Desplan, Neuronal specification in space and time. *Science* **362**, 176–180 (2018).
- C. Huang, J. A. Chan, C. Schuurmans, "Proneural bHLH genes in development and disease" in *Current Topics in Developmental Biology* (Elsevier, 2014), pp. 75–127.
- F. Guillemot, B. A. Hassan, Beyond proneural: Emerging functions and regulations of proneural proteins. *Curr. Opin. Neurobiol.* **42**, 93–101 (2017).
- C. M. Parras *et al.*, Divergent functions of the proneural genes Mash1 and Ngn2 in the specification of neuronal subtype identity. *Genes Dev.* **16**, 324–338 (2002).
- T. Inoue *et al.*, Math3 and NeuroD regulate amacrine cell fate specification in the retina. *Development* **129**, 831–842 (2002).
- M. Reilly, C. Cros, E. Varol, E. Yemini, O. Hobert, Unique homeobox codes delineate all the neuron classes of *C. elegans*. *Nature* **584**, 595–601 (2020).
- A. Sagner, J. Briscoe, Establishing neuronal diversity in the spinal cord: A time and a place. *Development* **146**, dev182154 (2019).
- C. L. Freund *et al.*, Cone-rod dystrophy due to mutations in a novel photoreceptor-specific homeobox gene (CRX) essential for maintenance of the photoreceptor. *Cell* **91**, 543–553 (1997).
- T. M. Jessell, Neuronal specification in the spinal cord: Inductive signals and transcriptional codes. *Nat. Rev. Genet.* **1**, 20–29 (2000).
- C. E. Ang, M. Wernig, Induced neuronal reprogramming. *J. Comp. Neurol.* **522**, 2877–2886 (2014).
- M. Caiazzo *et al.*, Direct generation of functional dopaminergic neurons from mouse and human fibroblasts. *Nature* **476**, 224–227 (2011).
- U. Pfisterer *et al.*, Direct conversion of human fibroblasts to dopaminergic neurons. *Proc. Natl. Acad. Sci. U.S.A.* **108**, 10343–10348 (2011).
- E. Y. Son *et al.*, Conversion of mouse and human fibroblasts into functional spinal motor neurons. *Cell Stem Cell* **9**, 205–218 (2011).
- R. Tsunemoto *et al.*, Diverse reprogramming codes for neuronal identity. *Nature* **557**, 375–380 (2018).
- N. Yang *et al.*, Generation of pure GABAergic neurons by transcription factor programming. *Nat. Methods* **14**, 621–628 (2017).
- C. E. Ang *et al.*, The novel lncRNA Inc-NR2F1 is pro-neurogenic and mutated in human neurodevelopmental disorders. *Elife* **8**, e41770 (2019).
- S. Chanda *et al.*, Direct reprogramming of human neurons identifies MARCKSL1 as a pathogenic mediator of valproic acid-induced teratogenicity. *Cell Stem Cell* **25**, 103–119.e6 (2019).
- S. Koneermann *et al.*, Transcriptome engineering with RNA-targeting type VI-D CRISPR effectors. *Cell* **173**, 665–676.e14 (2018).
- R. Nehme *et al.*, Combining NGN2 programming with developmental patterning generates human excitatory neurons with NMDAR-mediated synaptic transmission. *Cell Rep.* **23**, 2509–2523 (2018).
- C. Pak *et al.*, Human neuropsychiatric disease modeling using conditional deletion reveals synaptic transmission defects caused by heterozygous mutations in NRXN1. *Cell Stem Cell* **17**, 316–328 (2015).
- F. Yi *et al.*, Autism-associated SHANK3 haploinsufficiency causes Ih channelopathy in human neurons. *Science* **352**, aaf2669 (2016).
- Y. Zhang *et al.*, Rapid single-step induction of functional neurons from human pluripotent stem cells. *Neuron* **78**, 785–798 (2013).
- S. E. Perez, S. Rebelo, D. J. Anderson, Early specification of sensory neuron fate revealed by expression and function of neurogenins in the chick embryo. *Development* **126**, 1715–1728 (1999).
- L.-F. Chu *et al.*, Single-cell RNA-seq reveals novel regulators of human embryonic stem cell differentiation to definitive endoderm. *Genome Biol.* **17**, 173 (2016).
- H.-C. Lin *et al.*, NGN2 induces diverse neuron types from human pluripotency. *Stem Cell Rep.* **16**, 2118–2127 (2021).
- J. Ericson, S. Thor, T. Edlund, T. M. Jessell, T. Yamada, Early stages of motor neuron differentiation revealed by expression of homeobox gene *islet-1*. *Science* **256**, 1555–1560 (1992).
- E. O. Mazzone *et al.*, Synergistic binding of transcription factors to cell-specific enhancers programs motor neuron identity. *Nat. Neurosci.* **16**, 1219–1227 (2013).
- Z. Yao *et al.*, A taxonomy of transcriptomic cell types across the isocortex and hippocampal formation. *Cell* **184**, 3222–3241.e26 (2021).
- J. F. Brunet, A. Ghysen, Deconstructing cell determination: Proneural genes and neuronal identity. *Bioessays* **21**, 313–318 (1999).
- Z.-W. Du *et al.*, Generation and expansion of highly pure motor neuron progenitors from human pluripotent stem cells. *Nat. Commun.* **6**, 6626 (2015).
- A. Sagner, J. Briscoe, Morphogen interpretation: Concentration, time, competence, and signaling dynamics. *Wiley Interdiscip. Rev. Dev. Biol.* **6**, e271 (2017).
- M. V. Chao, Neurotrophin receptors: A window into neuronal differentiation. *Neuron* **9**, 583–593 (1992).
- R. A. S. Kjolby, R. M. Harland, Genome-wide identification of Wnt/ β -catenin transcriptional targets during *Xenopus* gastrulation. *Dev. Biol.* **426**, 165–175 (2017).
- A. Simeone, D. Acampora, M. Gulisano, A. Stornaiuolo, E. Boncinelli, Nested expression domains of four homeobox genes in developing rostral brain. *Nature* **358**, 687–690 (1992).
- J. A. Gorski *et al.*, Cortical excitatory neurons and glia, but not GABAergic neurons, are produced in the Emx1-expressing lineage. *J. Neurosci.* **22**, 6309–6314 (2002).
- M. Yoshida *et al.*, Emx1 and Emx2 functions in development of dorsal telencephalon. *Development* **124**, 101–111 (1997).
- J. M. Hébert, G. Fishell, The genetics of early telencephalon patterning: Some assembly required. *Nat. Rev. Neurosci.* **9**, 678–685 (2008).
- W. Tao, E. Lai, Telencephalon-restricted expression of BF-1, a new member of the HNF-3/fork head gene family, in the developing rat brain. *Neuron* **8**, 957–966 (1992).
- C. Englund *et al.*, Pax6, Tbr2, and Tbr1 are expressed sequentially by radial Glia, intermediate progenitor cells, and postmitotic neurons in developing neocortex. *J. Neurosci.* **25**, 247–251 (2005).
- E. S. Lein *et al.*, Genome-wide atlas of gene expression in the adult mouse brain. *Nature* **445**, 168–176 (2007).
- J. L. Fowler, L. T. Ang, K. M. Loh, A critical look: Challenges in differentiating human pluripotent stem cells into desired cell types and organoids. *Wiley Interdiscip. Rev. Dev. Biol.* **9**, e368 (2020).
- B. Treutlein *et al.*, Dissecting direct reprogramming from fibroblast to neuron using single-cell RNA-seq. *Nature* **534**, 391–395 (2016).
- L. Luna-Zurita *et al.*, Complex interdependence regulates heterotypic transcription factor distribution and coordinates cardiogenesis. *Cell* **164**, 999–1014 (2016).
- Q. Y. Lee *et al.*, Pro-neuronal activity of Myod1 due to promiscuous binding to neuronal genes. *Nat. Cell Biol.* **22**, 401–411 (2020).
- P. Briata *et al.*, EMX1 homeoprotein is expressed in cell nuclei of the developing cerebral cortex and in the axons of the olfactory sensory neurons. *Mech. Dev.* **57**, 169–180 (1996).
- K. M. Bishop, S. Garel, Y. Nakagawa, J. L. R. Rubenstein, D. D. M. O'Leary, Emx1 and Emx2 cooperate to regulate cortical size, lamination, neuronal differentiation, development of cortical efferents, and thalamocortical pathfinding. *J. Comp. Neurol.* **457**, 345–360 (2003).
- M. Mall *et al.*, Myt1l safeguards neuronal identity by actively repressing many non-neuronal fates. *Nature* **544**, 245–249 (2017).
- G. V. Ermakova, E. A. Solovieva, N. Y. Martynova, A. G. Zaisky, The homeodomain factor Xanf represses expression of genes in the presumptive rostral forebrain that specify more caudal brain regions. *Dev. Biol.* **307**, 483–497 (2007).
- J. von Frowein, A. Wizenmann, M. Götz, The transcription factors Emx1 and Emx2 suppress choroid plexus development and promote neuroepithelial cell fate. *Dev. Biol.* **296**, 239–252 (2006).
- P. Arlotta, O. Hobert, Homeotic transformations of neuronal cell identities. *Trends Neurosci.* **38**, 751–762 (2015).
- O. Hobert, Regulatory logic of neuronal diversity: Terminal selector genes and selector motifs. *Proc. Natl. Acad. Sci. U.S.A.* **105**, 20067–20071 (2008).
- C. B. Pantazis *et al.*, A reference human induced pluripotent stem cell line for large-scale collaborative studies. *Cell Stem Cell* **29**, 1685–1702.e22 (2022).
- C. E. Ang, M. Wernig, Data from "Generation of human excitatory forebrain neurons by cooperative binding of proneural NGN2 and homeobox factor EMX1." NCBI GEO. <https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE181019>. Deposited 28 July 2021.