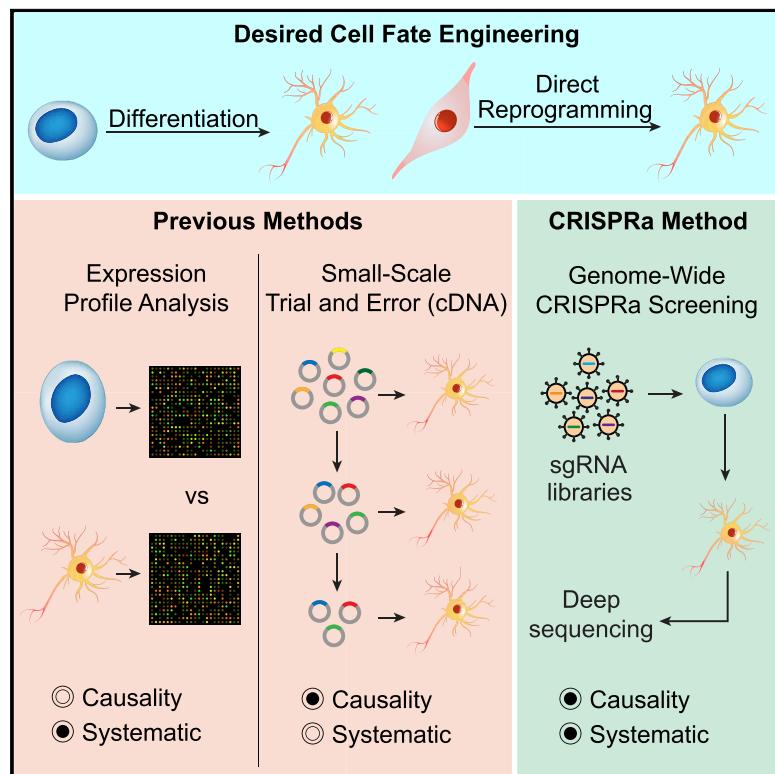


# CRISPR Activation Screens Systematically Identify Factors that Drive Neuronal Fate and Reprogramming

## Graphical Abstract



## Authors

Yanxia Liu, Chen Yu,  
Timothy Patrick Daley, ...,  
Wing Hung Wong, Marius Wernig,  
Lei S. Qi

## Correspondence

stanley.qi@stanford.edu

## In Brief

Liu and colleagues developed a high-throughput CRISPR activation screening approach to systematically identify transcription factors that efficiently promote neuronal fate from ESCs. Some of these single and pairwise factors can further reprogram fibroblasts into neurons, showing that this approach may have broad utility for engineering cell lineages.

## Highlights

- CRISPRa screening identifies factors that promote neuronal fate in an unbiased manner
- Quantitative genetic interaction maps reveal contributions of pairwise factors
- Factor combinations can efficiently and directly reprogram fibroblasts into neurons
- Induced neurons share transcriptional signatures similar to those of endogenous neurons



# CRISPR Activation Screens Systematically Identify Factors that Drive Neuronal Fate and Reprogramming

Yanxia Liu,<sup>1,2,3,11</sup> Chen Yu,<sup>4,11</sup> Timothy Patrick Daley,<sup>1,5,11</sup> Fangyuan Wang,<sup>1,6</sup> William S. Cao,<sup>8</sup> Salil Bhate,<sup>1</sup> Xueqiu Lin,<sup>1,2,3</sup> Chris Still II,<sup>7</sup> Honglei Liu,<sup>1</sup> Dehua Zhao,<sup>1,2,3</sup> Haifeng Wang,<sup>1,2,3</sup> Xinmin S. Xie,<sup>8</sup> Sheng Ding,<sup>4,9</sup> Wing Hung Wong,<sup>5</sup> Marius Wernig,<sup>10</sup> and Lei S. Qi<sup>1,2,3,12,\*</sup>

<sup>1</sup>Department of Bioengineering, Stanford University, Stanford, CA 94305, USA

<sup>2</sup>Department of Chemical and Systems Biology, Stanford University, Stanford, CA 94305, USA

<sup>3</sup>Stanford ChEM-H, Stanford University, Stanford, CA 94305, USA

<sup>4</sup>The Gladstone Institute of Cardiovascular Disease, San Francisco, CA 94158, USA

<sup>5</sup>Department of Statistics & Department of Biomedical Data Science, Stanford University, Stanford, CA 94305, USA

<sup>6</sup>Shanghai Institute of Rheumatology, Ren Ji Hospital, School of Medicine, Shanghai Jiao Tong University, Shanghai 200127, China

<sup>7</sup>Graduate Program of Stem Cell Biology and Regenerative Medicine, Stanford University, Stanford, CA 94305, USA

<sup>8</sup>AfaSci Research Laboratories, Redwood City, CA 94063, USA

<sup>9</sup>School of Pharmaceutical Sciences, Tsinghua University, Beijing 100084, China

<sup>10</sup>Department of Pathology, Institute for Stem Cell Biology and Regenerative Medicine, Stanford, CA 94305, USA

<sup>11</sup>These authors contributed equally

<sup>12</sup>Lead Contact

\*Correspondence: stanley.qi@stanford.edu

<https://doi.org/10.1016/j.stem.2018.09.003>

## SUMMARY

Comprehensive identification of factors that can specify neuronal fate could provide valuable insights into lineage specification and reprogramming, but systematic interrogation of transcription factors, and their interactions with each other, has proven technically challenging. We developed a CRISPR activation (CRISPRa) approach to systematically identify regulators of neuronal-fate specification. We activated expression of all endogenous transcription factors and other regulators via a pooled CRISPRa screen in embryonic stem cells, revealing genes including epigenetic regulators such as Ezh2 that can induce neuronal fate. Systematic CRISPR-based activation of factor pairs allowed us to generate a genetic interaction map for neuronal differentiation, with confirmation of top individual and combinatorial hits as bona fide inducers of neuronal fate. Several factor pairs could directly reprogram fibroblasts into neurons, which shared similar transcriptional programs with endogenous neurons. This study provides an unbiased discovery approach for systematic identification of genes that drive cell-fate acquisition.

## INTRODUCTION

Cell identity is established through the activity of core transcription factors (TFs) and other regulators (Takahashi and Yamanaka, 2006; Vierbuchen et al., 2010; Xu et al., 2015). It has remained a challenge to systematically profile how genetic

drivers determine cell-lineage phenotypes, which is critical to interrogate the regulatory network of cell fate. Approaches that allow us to determine causal contributions of individual factors and their interactions to cell fate on a large scale will greatly advance our understanding of cell-identity regulation and facilitate rational engineering of clinically useful cell types.

Systematic profiling of the gene-cell-fate relationship has relied on comparative genome-wide gene expression analyses across multiple cell types (Cahan et al., 2014; D'Alessio et al., 2015; Heinäniemi et al., 2013). These approaches provide valuable information but cannot establish causality between genes and phenotypes. Small-scale ectopic overexpression of genes that are specifically expressed in desired tissues, or implicated with important roles in relevant developmental processes, is often used to define factors for inducing a desired cell identity. While factor combinations that induce a given cell fate have been identified through trial and error (Takahashi and Yamanaka, 2006; Tsunemoto et al., 2018; Vierbuchen et al., 2010), this approach cannot provide a systematic understanding of how the whole set of TFs and other DNA-binding factors contribute to cell-fate determination.

CRISPR activation (CRISPRa)-based gain-of-function perturbation offers a powerful approach to activate genes on a large-scale in a pooled manner (Black et al., 2016; Chavez et al., 2015; Gilbert et al., 2013, 2014; Konermann et al., 2015; Mali et al., 2013). While screens based on CRISPR-Cas9 gene knockout or CRISPRi/a transcriptional repression/activation have been demonstrated, no CRISPR-based screen has been used to establish a causal relationship between gene expression and cell-fate determination (Koike-Yusa et al., 2014; Parnas et al., 2015; Shalem et al., 2014; Wang et al., 2014).

Here, we developed a CRISPRa approach to profile the contribution of TFs and other DNA-binding factors to cell fate, both individually and in combination. We used serial pooled CRISPRa



screens to first generate an individual factor map, and then a factor genetic interaction (GI) map for genes that promote neuronal differentiation. These maps revealed many factors that previously have been unknown to promote neuronal differentiation. We validated the top factor hits, both when activated alone and in combination, by assessing their neuronal-promoting capacities. Importantly, several discovered factor pairs efficiently induced neuron formation from fibroblasts, including a novel combination of Ezh2 and Mecom. The function of the induced neurons was confirmed by electrophysiology. We performed a large set of whole-genome RNA sequencing (RNA-seq) and found that the differentiated and directly reprogrammed neurons shared very similar transcriptional profiles to endogenous neurons. Our study provides an approach to systematically discover lineage-promoting factors, offering a useful method to study gene functions related to cell-fate determination.

## RESULTS

### Establishing an Efficient CRISPRa System in Mouse Embryonic Stem Cells to Study Differentiation

To enable systematic evaluation of whether individual factors promote neuronal fate, we first generated a stable mouse embryonic stem cell (ESC) line that could express CRISPRa components after doxycycline (Dox) induction. We transduced lentiviral vectors that expressed *Streptococcus pyogenes* nuclease-dead Cas9 (dCas9) fused to the polypeptide SunTag system into E14 ESCs (Tanenbaum et al., 2014) and generated a clonal CRISPRa cell line (named CRISPR-activating mouse ESC, or CamESC) (Figures S1A).

We verified that CamESCs expressed the pluripotency marker OCT4 (Figure S1B) and could efficiently activate previously silenced endogenous genes using single sgRNAs (sgRNAs) (Figure S1C). We next tested whether a sgRNA could induce neuronal differentiation in CamESCs. Using a sgRNA targeting known neuronal-promoting genes *Ascl1* or *Ngn2*, we observed robust induction of neurons on day 12 (Figure S1D). These differentiated neurons were TUJ1 and MAP2 positive, same as the neurons generated by overexpression of *Ascl1* cDNA in ESCs (Figure S1D). All negative controls showed no neuronal morphology or gene expression (Figure S1E), confirming that neurons were indeed induced by CRISPRa-mediated target gene activation. We also observed similar differentiation kinetics of neuronal differentiation when comparing the sg*Ascl1* in CamESCs and *Ascl1* cDNA overexpression in ESCs (Figure S1F).

To facilitate isolation of differentiated neuronal cells, we further inserted a human *CD8* (*hCD8*) gene downstream of endogenous *Tubb3* in the CamESCs (Figure S1G). We verified the *Tubb3-hCD8* knockin CamESCs allowed enrichment of TUJ1<sup>+</sup> cells: after 12 days differentiation of CamESCs using CRISPRa to activate *Ascl1*, a hCD8<sup>+</sup> cell population enriched by magnetic activated cell sorting (MACS) highly expressed neuronal markers *Tuj1* and *Map2* (Figure S1H). The *Tubb3-hCD8* CamESCs were used for both individual and combinatorial CRISPRa screens.

### Genome-wide Profiling of TFs and Other Regulators that Promote Neuronal Fate

We created a sgRNA library (containing 55,561 sgRNAs) that targeted all computationally predicted TFs and other DNA-binding

factors from the Transcription Factor Database TRANSFAC (total number = 2,428) (Wingender et al., 1996). sgRNAs targeting *Ascl1* and *Ngn2* were manually excluded, as we were concerned that including them might drown out signals from novel factors. The library also contained thousands of non-targeting sgRNAs as negative controls (Figure 1A).

We transduced the sgRNA library into CamESCs, induced CRISPRa activity and differentiation (Figures 1A and S2A). After 12 days, the cell population showed expression of neuronal markers TUJ1 and MAP2 (Figure S2B). We performed MACS to sort hCD8<sup>+</sup> and hCD8<sup>-</sup> populations, followed by next-generation sequencing (NGS) to characterize the sgRNA presentation in each population (Figures S2C and S2D).

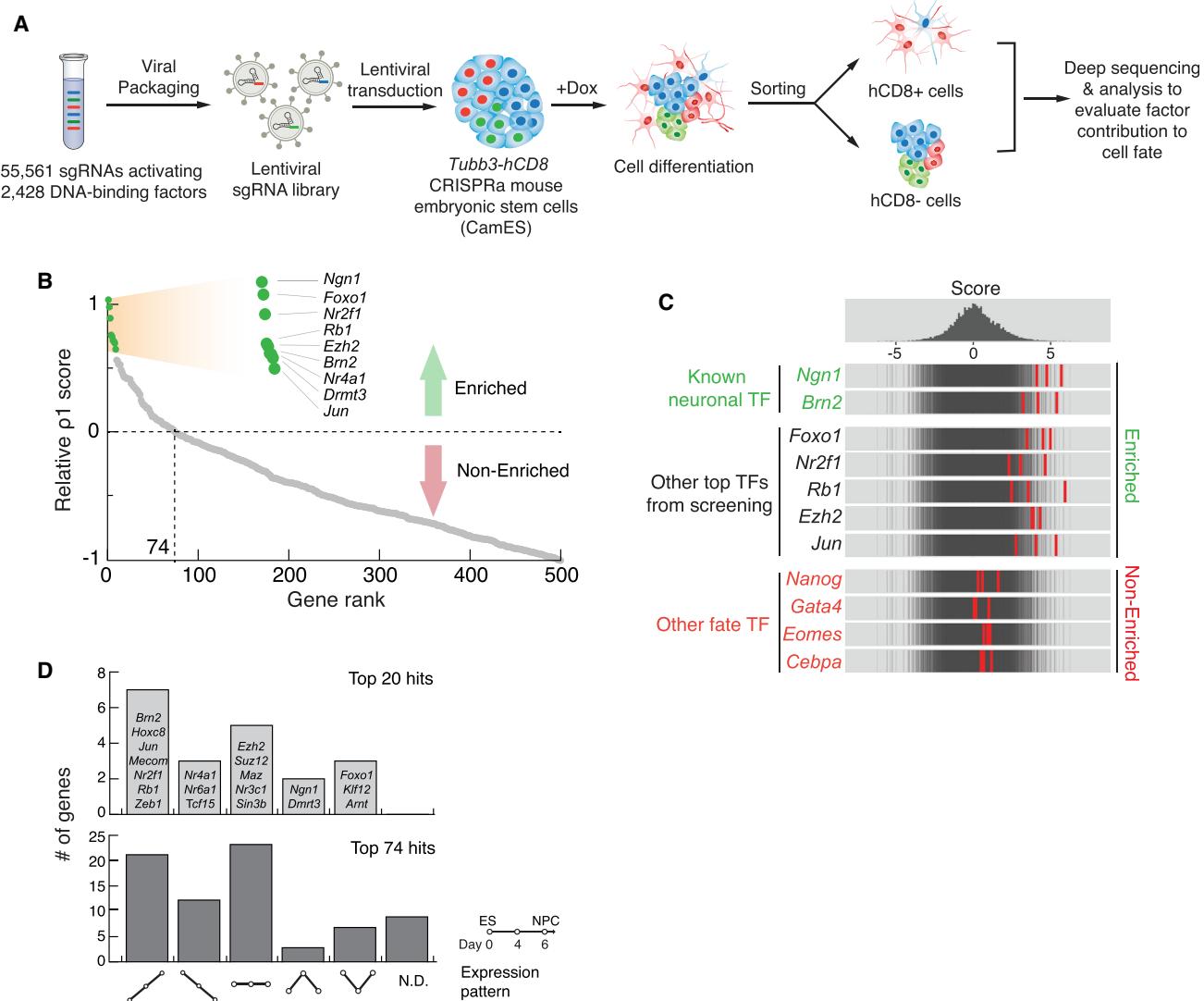
Unlike cell viability/fitness screens that directly compare sgRNA presentation in final and initial populations (Chen et al., 2015; Konermann et al., 2015), we developed an algorithm that accounted for the probabilistic contribution of each factor to neuronal differentiation (Figure S2E). We defined a numerical cell-fate contribution score ( $\rho_1$ ), calculated by the relative enrichment of the top 3 sgRNAs for each gene when comparing the *Tubb3-hCD8<sup>+</sup>* and *Tubb3-hCD8<sup>-</sup>* populations. The  $\rho_1$  score is an enrichment score that represents the probability that one cell becomes a neuron in the presence of the gene-perturbing guide relative to the negative control guides: a higher  $\rho_1$  score represents a greater chance that this factor could induce neuronal fate. Using  $\rho_1 = 0$  as a cutoff value, we identified 74 factors that positively contributed to neuronal fate (Figure 1B; Table S1). Within this 74-factor list, we observed many factors that were known to induce neuronal fate. In contrast, other genes that were known to be important for alternative fates (pluripotency, mesoderm, and endoderm) were not enriched (Figure 1C; Table S1).

Notably, our CRISPRa screening strategy identified potential neuronal-promoting genes that were not biased toward differentially expressed genes (Figure 1D): among the top 20 and 74 hits, 7 and 41 genes respectively showed no differential expression between neurons and ESCs. This likely suggests that CRISPRa screening is a more systematic unbiased approach to identify genes that could promote a certain cell-fate formation compared to the gene expression profiling approach.

### Validating Hit Genes that Promote Neuronal Differentiation

To validate whether the hit genes induced neuronal phenotype, and for simplicity of analyses and manageability of experimental validation, we focused on the top 20 factors as the core factors for validation. For each factor, we cloned the most enriched sgRNA and measured their effects for inducing neuronal differentiation by staining MAP2 (Figure 2A), TUJ1, and NeuN (Figure S2F). Nineteen out of 20 factors showed distinct neuronal marker expression (Figures 2A and 2B). To further confirm the discovered hits and rule out off-target effects, we transduced lentiviral vectors that expressed the cDNA of each of the top 14 genes in E14 ESCs and observed similar neuronal induction (Figure 2C).

Neuronal-promoting factors may directly induce neuron generation or maintain neuronal survival. To deconvolve these two possibilities for our hits, we quantified the percentage of PSA-NCAM<sup>+</sup> neurons for 6 factors (Figure 2D) in a time-course



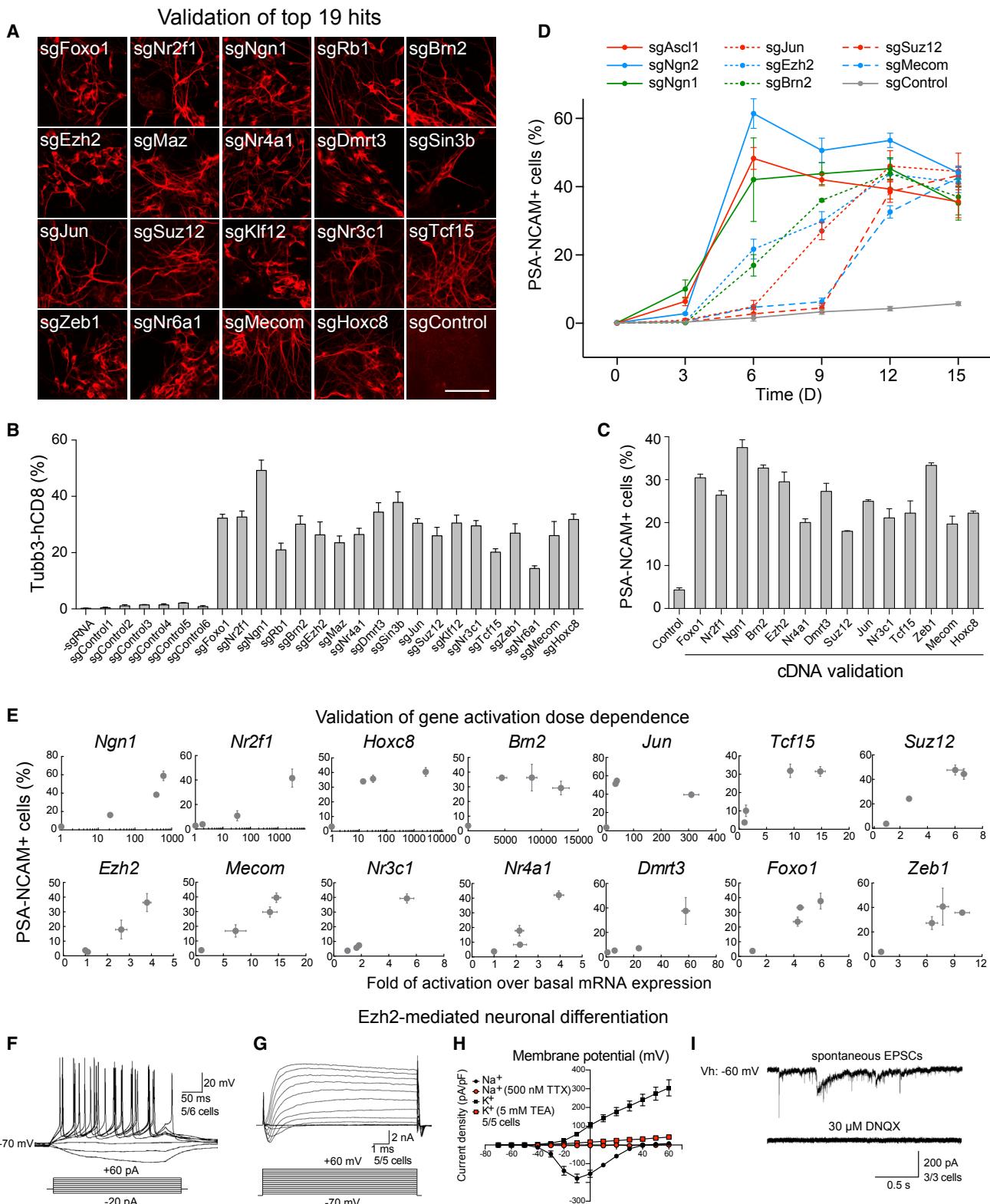
**Figure 1. Systematic Evaluation of Transcription Factors' and Other Regulators' Contributions to Neuronal-Fate Acquisition**

- (A) Schematic representation of our CRISPRa gain-of-function screens to systematically evaluate factors contributing to neuronal differentiation.
  - (B) A map of the contribution of the top 500 factor hits to neuronal fate according to the calculated  $p_1$  score. The top 9 hits are shown in green. The top 74 hits showed enrichment for neuronal differentiation.
  - (C) Frequency histograms of the negative control guides (top), and the top 3 enriched sgRNAs for indicated factors.
  - (D) Analysis of factor expression pattern during neuronal differentiation for the top 20 and top 74 hits according to RNA-seq (W.H.W., unpublished data). Six categories are considered: up, down, no change, up-and-down, down-and-up, and no data (N.D.).
- See also Figures S1 and S2 and Table S1.

experiment. Among these factors, *Ngn1* showed similar neuronal differentiation kinetics as known neuronal-inducing factors *Ascl1* and *Ngn2*, with a high percentage of PSA-NCAM<sup>+</sup> cell population appearing at early time points (Figure 2D). This expression pattern was also observed for other neuronal markers (Figure S2G), which together suggested that *Ngn1* acts in a manner similar to *Ascl1* and *Ngn2* by directly inducing neuronal generation. In contrast, *Suz12* and *Mecom* showed much slower neuronal differentiation kinetics with an abrupt increase of PSA-NCAM<sup>+</sup> cell population after 9 days. These suggested that *Suz12* and *Mecom* increased the number of neurons likely through the promotion of neuronal survival. Other factors,

including *Ezh2*, *Jun*, and *Brn2*, showed intermediate differentiation kinetics, suggesting that their mechanism of action is through both weak neuronal induction and increased neuronal survival (Figures 2D and S2G).

We further investigated the relationship between the efficiency of neuronal differentiation (phenotype) and the activation fold of the factors (genotype). For the above cDNA-validated factors, we tested the 3 most enriched sgRNAs (Figure 2E). qPCR results showed that different genes exhibited wide dynamic ranges of endogenous gene activation. A majority of genes showed positive correlation between neuronal differentiation and gene activation level (Figure 2E). Using 3 sgRNAs to verify the dosage

**Figure 2. Top Hits Are Validated to Drive Neuronal Differentiation**

(A) Validation of 19 top hits for neuronal differentiation. Microscopic images of MAP2-stained neurons transduced with individual sgRNAs for 19 top hits and a negative control sgRNA in CamESCs after 12-day differentiation. Scale bar, 100  $\mu$ m.

(B) Validation of 19 top hits using *Tubb3*-hCD8 $^{+}$  expression. *Tubb3*-hCD8 expression was assayed on day 12 by flow cytometry.

(legend continued on next page)

dependence of neuronal differentiation further confirmed the observed phenotype was not due to off-target CRISPRa effects.

To validate whether our screening approach identifies factors that induce functional neurons, we characterized the electrophysiological functions of neurons induced by CRISPRa-mediated expression of *Ezh2* and *Ngn1*. We observed that differentiated neurons could generate action potentials upon current injection (Figures 2F and S3A). In voltage-clamp mode, we observed both fast, inactivating inward and outward currents, which probably correspond to opening of voltage-dependent potassium ( $K^+$ ) and sodium ( $Na^+$ ) channels, respectively (Figures 2G and S3B). Outward, presumably potassium, currents were inhibited by the addition of tetraethylammonium (TEA), a blocker of voltage-gated  $K^+$  channels. Inward, presumably sodium, currents were inhibited by the addition of tetrodotoxin (TTX), a blocker of voltage-gated  $Na^+$  channels (Figures 2H and S3C). We also observed and recorded spontaneous excitatory postsynaptic currents (EPSCs) in differentiated neurons. These currents could be blocked by the addition of 6,7-dinitroquinoxaline-2,3-dione (DNQX), an AMPA and kainate receptor antagonist (Figures 2I and S3D). These data demonstrated that functional synapses can be formed with differentiated neurons driven by *Ezh2* and *Ngn1*. Together these validations confirmed the function of the discovered factors for neuronal-fate determination.

### Generation of a Quantitative GI Map for Neuronal Fate

The cell-identity regulatory network is governed by a combination of core factors. GI mapping between pairs of factors, which indicates how strongly the function of one factor depends on the presence of a second (Adamson et al., 2016; Bassik et al., 2013; Dixit et al., 2016), offers a powerful approach to reveal molecular interactions in a given regulatory network. We sought to systematically profile how pairs of factors contribute to neuronal phenotype.

We developed a combinatorial CRISPRa gene activation strategy to determine the GIs among the top 19 core factors. We generated a library of lentiviral constructs that co-expressed two sgRNAs (Figures 3A and S4A). We included two sgRNAs for each gene that were validated to activate the target gene to high or low levels. The library also included pairs of a targeting sgRNA and a non-targeting sgRNA, which were used to compare effects of combinatorial factors to individual factors.

We transduced the double-sgRNA library into CamESCs, added Dox to induce differentiation, sorted *Tubb3*-hCD8 positive and negative populations by fluorescence-activated cell sorting (FACS), and characterized the double-sgRNA presentation in each population (Figure 3A). We observed a high correlation score ( $R = 0.902$ ) between two independent experimental replicates (Figure S4B).

To quantify pairwise interactions of factors relative to individual factors for neuronal induction, we generated quantitative interaction scores (defined as  $\rho_2$  and  $\rho_3$  scores) for each factor pair (Figure 3B). The gene-level interaction score  $\rho_2$  was computed under the assumption that a single factor of the pair likely drives the majority of the cells to differentiate toward neurons. This assumption was validated by examining the empirical distribution of interaction scores (Figure 3B, histogram) and verifying that it was centered at zero, corresponding to no interaction. The guide-level interaction score  $\rho_3$  was computed after subtracting off gene level interactions and pooling information by putting a standard normal prior on the expected interactions (Figure S4G).

Hierarchical clustering of factors based on the correlation of their interactions revealed that a number of factors formed positive or negative interactions for neuronal differentiation. Deviations toward the neuronal fate relative to the individual factors were defined as positive GIs, and deviations away from the neuronal fate were defined as negative GIs. Interestingly, a core cluster of 6 genes (*Ngn1*, *Zeb1*, *Tcf15*, *Foxo1*, *Ezh2*, *Brn2*) showed mutually positive effects (Figure 3B). Meanwhile, a few combinations of factors appeared to negate their function as neuronal-fate inducers, for example, co-expression of *Nr4a1* and *Ngn1* reduced neuronal differentiation.

To validate these pairs, we first calculated the local false discovery rate (FDR) for each interaction (Efron, 2012), and then chose 5 interactions that showed the lowest local FDR for experimental validation. Computational analysis of the 4 positive interactions showed additive effects for the pairwise factors compared to individual factors (Figure 3C), while one interaction showed a negative effect (Figure 3C). We generated sgRNA constructs that activated individual or both factors, transduced them into CamESCs, and characterized neuronal differentiation by PSA-NCAM staining (Figure 3D). In parallel, we also transduced lentiviral vectors that expressed the cDNA of these genes in ESCs (Figure 3E). Both methods

(C) Validation of 14 top hits for neuronal differentiation by cDNA overexpression. Quantification of PSA-NCAM<sup>+</sup> neurons was assayed on day 12 by flow cytometry.

(D) Differentiation kinetics of CamESCs transduced with individual hit sgRNAs or a negative control sgRNA. PSA-NCAM percentages were quantified by flow cytometry at different time points with Dox addition.

(E) A quantitative map between neuronal differentiation (phenotype) and gene activation (genotype). PSA-NCAM percentage of cells induced by three individual sgRNAs for each gene and a negative control sgRNA were measured by flow cytometry. The mRNA expression was measured by real-time qPCR, and fold change was calculated by normalizing to the expression level of cells containing a negative control sgRNA.

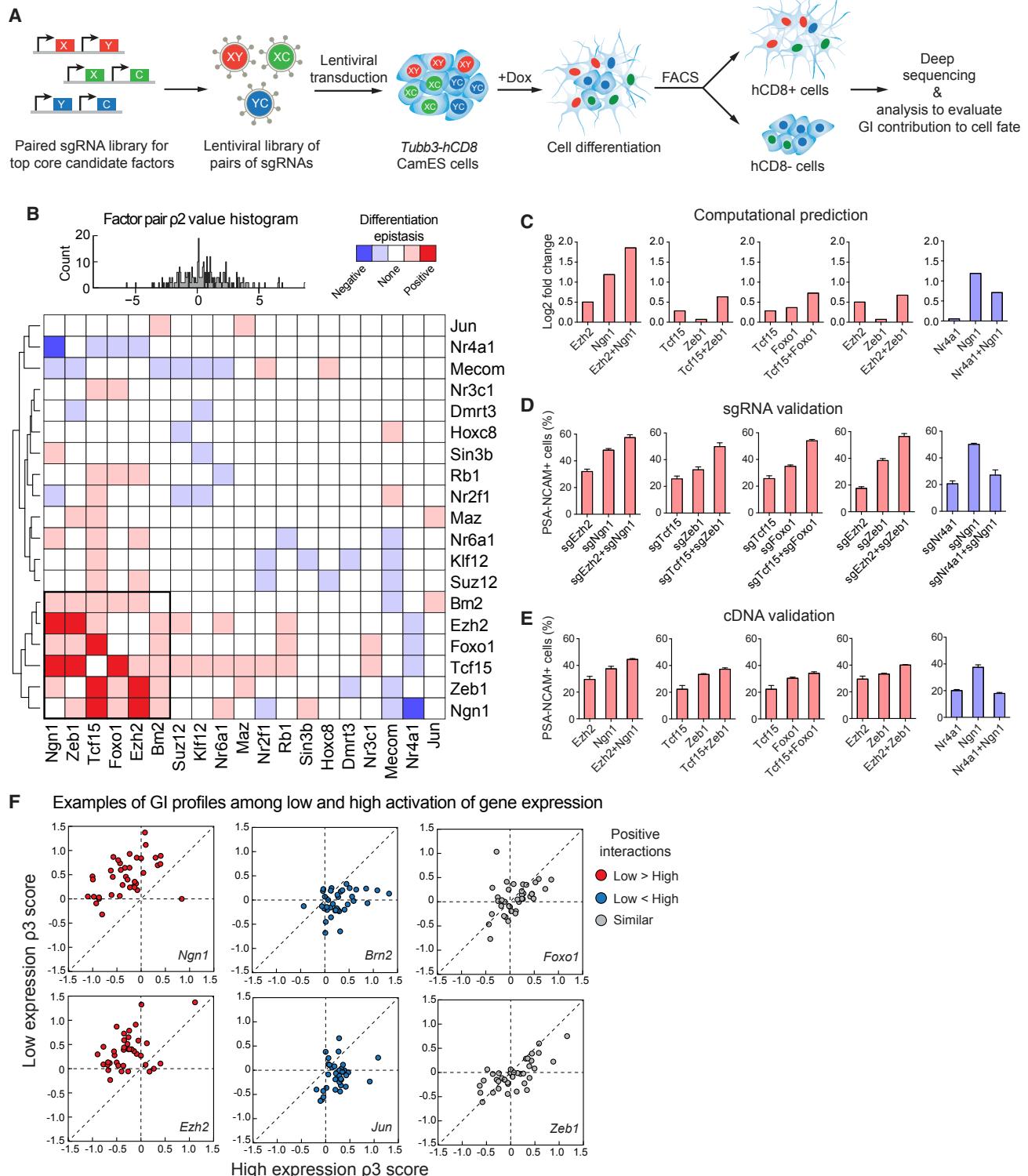
(B–E) Error bars, SD  $\pm$  the mean of three biological replicates.

(F–I) Electrophysiological properties of *Ezh2*-differentiated neurons. (F) Representative traces showing action potentials elicited. Step current injection protocols were used from  $-20$  to  $+60$  pA at 10-pA intervals (lower panel). (G) Representative traces of whole-cell currents in voltage-clamp mode; cells were held at  $-80$  mV. Step depolarization from  $-70$  mV to  $+60$  mV at 10-mV intervals was delivered (lower panel).

(H) Current density plot of voltage activated currents. The graph shows current density (y axis, pA/pF) for peak outward (potassium, squares) and inward currents (sodium, circles) plotted against voltage step (x axis, mV) in the absence (black) and presence of either TEA (red squares) or TTX (red circles). The holding potential was  $-80$  mV.

(I) Representative traces of evoked EPSCs recorded. Application of DNQX blocked the response of EPSCs (lower panel). Cells were recorded at a holding potential ( $V_h$ ) of  $-60$  mV. Error bars indicate  $\pm$  SEM of cell counts.

See also Figures S2 and S3.

**Figure 3. A Genetic Interaction Map for Core Factors Contributing to Neuronal Fate and Validation**

(A) Schematic representation of our double-CRISPRa screen to evaluate GIs between factors for their contribution to neuronal differentiation.

(B) A GI map of the core factor pairs showing their contribution to neuronal fate according to the computed  $p_2$  scores, and genes were clustered hierarchically based on the correlation of their GIs. A cluster of factors with mutually positive interactions is highlighted in a black box. A  $p_2$  histogram for all interactions and color key are shown on the top.(C) For 5 lowest local FDR GIs, average log<sub>2</sub> fold changes are plotted for the selected genes against negative control guides and the corresponding gene pairs.

(legend continued on next page)

confirmed the double-CRISPRa profiling experimental results and the computed positive and negative interactions between these factor pairs.

We then tested whether differentiated cells exhibited neuronal functions, using Ezh2+Ngn1 cells as an example. Electrophysiological properties such as eliciting action potential, recording inward and outward currents, and firing spontaneous EPSCs demonstrated that differentiated neurons were functional (Figures S4C–S4F). Together these results demonstrated that the combinatorial CRISPRa screens could be used to reveal GIs for cell-fate determination.

Interestingly, our approach also revealed that GI profiles for neuronal-fate determination could be dependent on the quantitative expression level of factors (Figures 3F and S4G). Our combinatorial CRISPRa approach has a unique feature that each gene is activated by different sgRNAs to high or low levels, which allows us to characterize how high or low expression of one factor interacts with another factor. GI profiles revealed genes could have different interactions when highly or lowly expressed (Figures 3F and S4G). For example, high expression of *Ngn1* or *Ezh2* formed weaker interactions with most other factors, likely because their high expression level already induced saturated neuronal differentiation, making their function less dependent on another factor (Figures 3F and S4G). On the other hand, high expression of *Brn2* or *Jun* showed more interactions compared to their low expression, likely because these genes are limiting factors for neuronal differentiation (Figures 3F and S4G). These results provided evidence that the type (positive or negative) and intensity of GIs could be dependent on gene expression levels. Different from the cDNA approach, the CRISPRa screening strategy provides a useful strategy to characterize the relationship between GI and quantitative gene expression levels.

### Characterization of Differentiated Neurons using Transcriptome Profiling

To better identify the characteristics of neurons generated by hit factors, we analyzed their transcriptomes using RNA-seq. We chose four hit factors (*Jun*, *Ngn1*, *Ezh2*, and *Suz12*) and one combination (*Ezh2+Ngn1*), using *Ascl1* as positive control. RNA-seq analysis revealed the differentiated neurons possessed a core neuronal transcriptional program (Figure 4A): the gene ontology (GO) terms associated with neuronal development were highly enriched in commonly expressed genes (Figure 4B), while genes associated with the cell cycle were highly downregulated (Figure 4C).

We sought to find why *Ezh2*, a global epigenetic regulator and a core component of PRC2 complex, promoted neuronal differentiation. To explore this finding, we focused on genes

that expressed specifically in the *Ezh2*-induced cells (Figure S5A). Compared to *Ascl1*- or *Ngn1*-induced neurons, genes related to the development of endodermal and mesodermal lineages were significantly downregulated in *Ezh2*-induced neurons (Figure S5A). These results may suggest that *Ezh2* could enhance neuronal differentiation partly through the inhibition of alternative endodermal and mesodermal lineages. Together, these experiments provided insights into potential roles of the intriguing finding that *Ezh2* promotes neuronal generation.

To explore potential regional separation and subtype specification of neurons generated by the identified factors, we next characterized the regional and subtype specific gene expression (Figures 4D and S5B). We compared the RNA-seq data of differentiated neurons to ESC data and to public available RNA-seq datasets of brain, cortex, and hippocampus samples (Tsunemoto et al., 2018). Overall, our differentiated neurons exhibited a highly similar core transcriptional signature as endogenous neuronal samples (Figure 4D). Diverse subtypes of neurons were generated, and different factors induced different transcriptional signatures. Interestingly, neurons driven by the two PRC2 factors, *Ezh2* or *Suz12*, showed elevated expression of the hindbrain markers, *Gbx2*, *Hoxa2*, and *Hoxa4*. Neurons generated by *Jun* expressed the floor plate markers, *Foxa1*, *Foxa2*, and *Corin*.

As for subtypes, our RNA-seq results showed that none of the factors we tested induced motor neuron and serotonergic neurons. On the contrary, all factors induced a strong glutamatergic neurons transcriptional program, indicating an excitatory phenotype (Figure 4D). Meanwhile, several factors induced GABAergic neuron-specific genes. This inhibitory phenotype was confirmed by the staining of *GAD65*, a marker of GABAergic neurons (Figure 4E). Taken together, these results suggested that our identified factors were able to induce distinct regional and subtype specification of neurons.

We also measured several layer markers to assay cortical layer specification (Figure S5B). *Ngn1* and *Ezh2+Ngn1* neurons had the highest expression level of layer 6 marker *Tbr1* among our hits, although its expression was lower than the positive control *Ngn2* (Figure S5B).

To examine the potential glial specification, we characterized the expression of glial marker genes in differentiated cells. Two hit factors, *Zeb1* and *Nr3c1*, induced strong *GFAP* expression (Figures 4F and S5B). *GLT1*, another astrocyte marker, was also highly expressed in these cells. In addition, *Olig2* and *Sox10* were expressed in *Zeb1* and *Ezh2+Zeb1* cells, which suggested the existence of oligodendrocyte precursors.

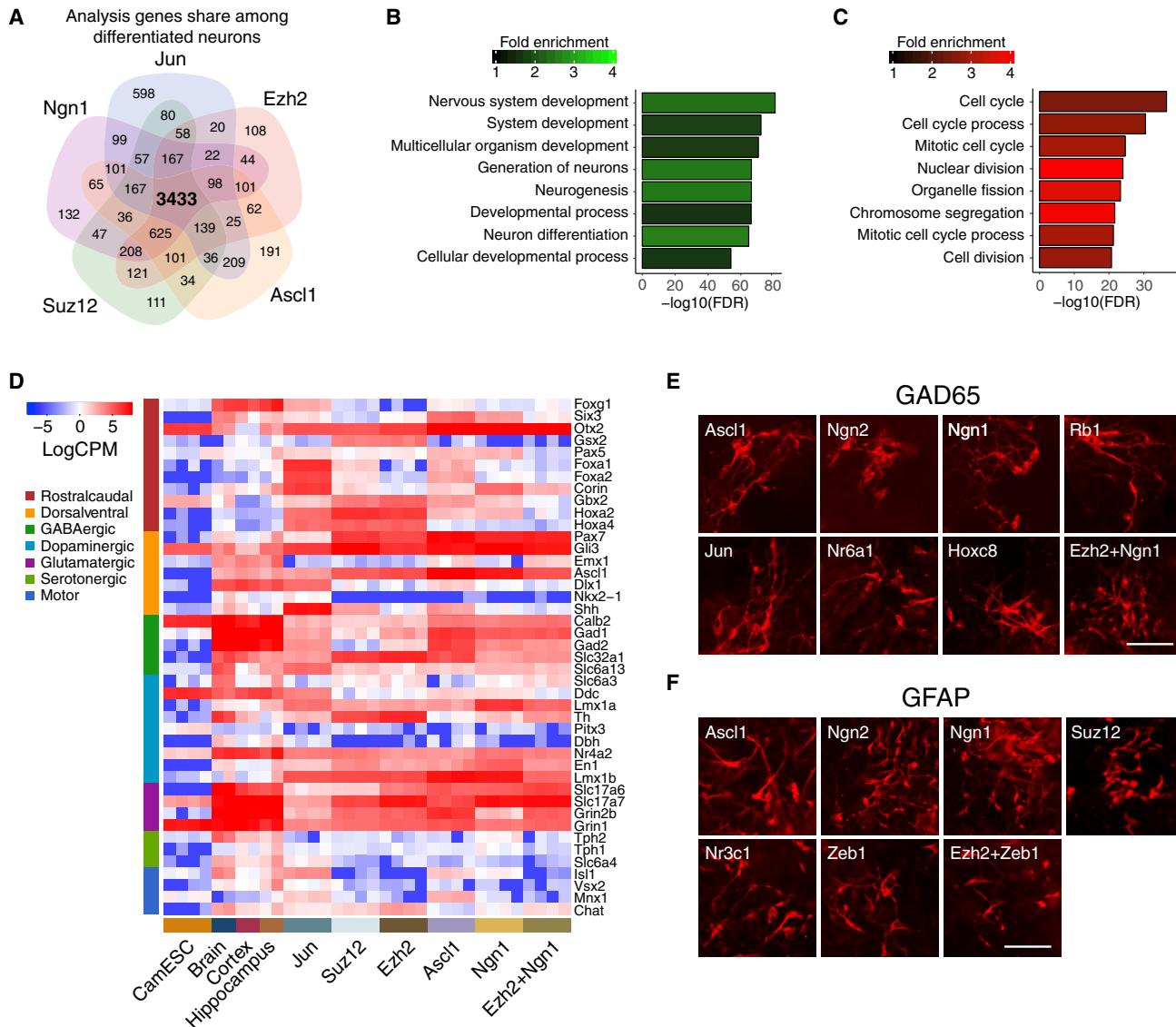
To understand the glial specification in more depth, we measured neural stem cell (NSC) markers, as well as astrocyte

(D) Validation of 4 top positive GIs and 1 negative GI for neuronal differentiation by sgRNAs. Quantification of PSA-NCAM<sup>+</sup> neurons was measured on day 12 by flow cytometry. Error bars, SD ± the mean of three biological replicates.

(E) Validation of 4 top positive GIs and 1 negative GI for neuronal differentiation by cDNAs. Quantification of PSA-NCAM<sup>+</sup> neurons was measured by flow cytometry on day 12 of single or double cDNAs in ESCs. Error bars, SD ± the mean of three biological replicates.

(F) Scatterplots of the p3 score for low expression sgRNAs and high expression sgRNAs of factors paired with the same common sgRNAs. For each dot, its coordinate on horizontal axis represent the p3 score (guide interaction) of GI between this gene's high-expression sgRNA and an independent sgRNA. Its coordinate on vertical axis represent the p3 score of GI between this gene's low-expression sgRNA and the same independent sgRNA. Red, low expression showing more positive GIs; blue, high expression showing more positive GIs; gray, low and high expression showing similar GIs. The dashed line indicates the expectation under no effect of a gene's expression level on GIs with other genes.

See also Figure S4.



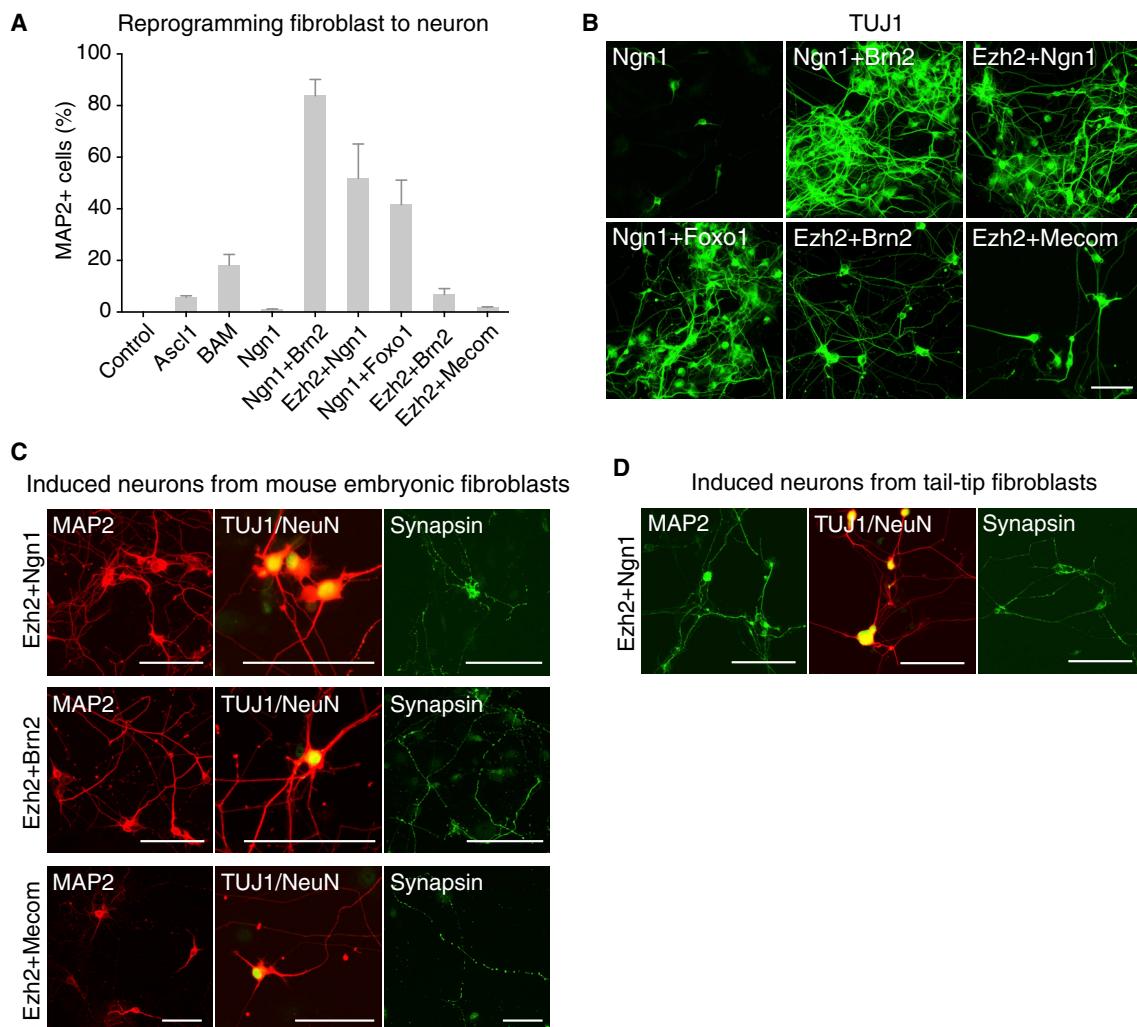
**Figure 4. Transcriptional and Subtype Specification Analysis of Differentiated Neurons Driven by Identified Factors**

- (A) Venn diagram of the differentially upregulated genes (Bonferroni-corrected  $p < 0.01$ ) for the neurons driven by identified factors.
- (B) Top enriched GO terms for the genes commonly upregulated in neuron samples derived by 4 identified factors (Ngn1, Jun, Ezh2, and Suz12) and a positive control factor (Ascl1).
- (C) Top enriched GO terms for the genes commonly downregulated in neuron samples derived by 4 identified factors (Ngn1, Jun, Ezh2, and Suz12) and a positive control factor (Ascl1).
- (D) Expression in normalized log count per million (CPM) in the day 12 differentiated CamESCs transduced with individual sgRNAs targeting indicated factors and somatic samples for neuronal regional and subtype-specific genes.
- (E and F) Staining of GAD65 (E) or GFAP (F) in day 12 differentiated CamESCs transduced with individual sgRNAs targeting identified factors. Scale bars, 100  $\mu$ m. See also Figure S5.

and oligodendrocyte markers throughout the differentiation process (Figure S5C). *Nestin* and *Pax6*, two NSC markers, were upregulated early in all samples and showed a downward trend after day 9. *Sox10* expression was greatly increased in Zeb1-cells on day 9, while Zeb1 in combination with Ezh2 induced high *Olig2* expression. These data suggested the transient existence of NSCs during differentiation, with sequential glial specification driven by several hit factors.

#### Validation of Factor Combinations for Neuronal Direct Reprogramming

We asked whether the identified factor combinations were able to drive direct reprogramming of neurons from fibroblasts, which was a more stringent test of neuronal-fate determination (Vierbuchen et al., 2010). We first transduced the top 14 individual factors into mouse embryonic fibroblasts (MEFs) and found only Ngn1 was able to moderately generate MAP2<sup>+</sup> and TUJ1<sup>+</sup>



**Figure 5. Efficient Induction of Neurons from Fibroblasts using Identified Factor Combinations**

(A) Measurement of neuronal direct reprogramming from mouse embryonic fibroblasts (MEFs) using the individual factors and factor combinations. Quantification of MAP2<sup>+</sup> neurons from MEF transduced with individual cDNA or pairwise of cDNAs. Efficiencies are computed by averaging 20 randomly selected visual fields. Error bars, SD  $\pm$  the mean of three biological replicates.

(B) Representative microscopic images of stained TUJ1 neurons 14 days after transduction of Ngn1, Ngn1+Brn2, Ezh2+Ngn1, Ngn1+Foxo1, Ezh2+Brn2, or Ezh2+Mecom in MEFs.

(C) Microscopic images of staining MAP2, TUJ1 (red)/NeuN (green), and Synapsin staining 14 days after transduction of Ezh2+Ngn1, Ezh2+Brn2, or Ezh2+Mecom in MEFs.

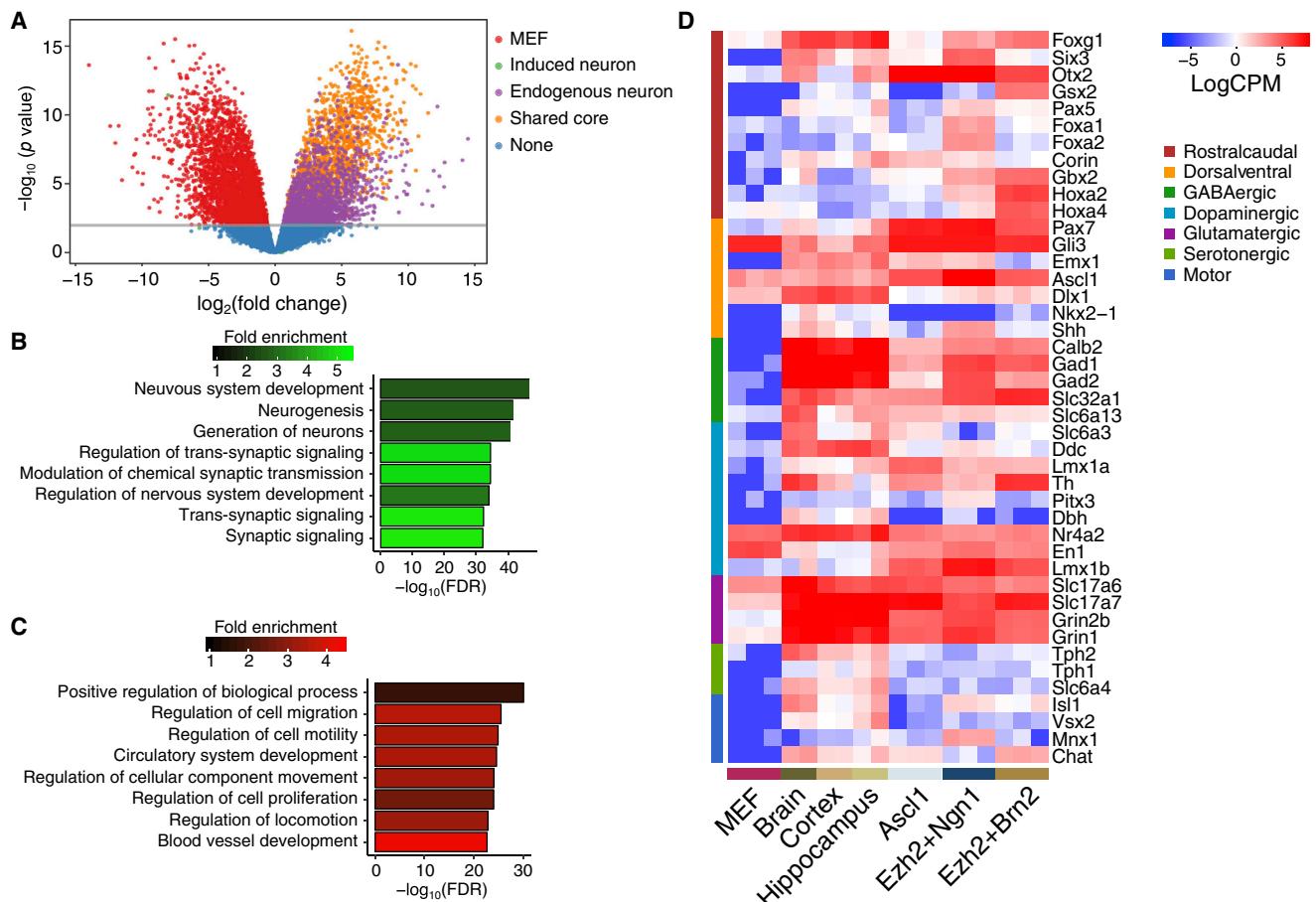
(D) Microscopic images of staining MAP2, TUJ1 (red)/NeuN (green), and Synapsin 26 days after transducing Ezh2+Ngn1 in neonatal tail-tip fibroblast (TTF). Scale bar, 100  $\mu$ m.

See also Figures S6 and S7.

neurons 14 days after infection (Figures 5A and 5B). This suggested individual factors had weak activity for inducing neurons from other somatic cells.

We next validated 15 possible combinations within the cluster of 6 mutually synergistic factors. Notably, four pairs caused significant induction of MAP2<sup>+</sup> and TUJ1<sup>+</sup> neurons from fibroblasts (Figures 5A and 5B): the presence of Brn2, Ezh2, and Foxo1 significantly increased the efficiency of Ngn1 for neuronal direct reprogramming, from 1% to up to 83%. Interestingly, the Ezh2+Brn2 combination also efficiently generated neurons from fibroblasts, whereas neither alone could induce neuronal direct re-

programming. Because of the strong synergistic effects between Ezh2 and Brn2, and our previous findings that Ezh2 promoted neuronal differentiation, we were intrigued to test additional Ezh2 combinations in this direct reprogramming context. We found that Ezh2+Mecom also induced neurons from fibroblasts (Figure 5A). Unlike the effects of factor pairs on neuronal differentiation from ESCs, the factor pairs showed non-linear synergistic effects for direct reprogramming. These distinct behaviors may suggest that the underlying regulatory networks between ESCs and fibroblasts present different barriers to neuron generation, and that a higher degree of synergism between driver genes is



**Figure 6. Transcriptional Analysis of Directly Reprogrammed Neurons Driven by Factor Combinations**

(A) Volcano plot of the log fold change versus minus  $\log_{10}$  Bonferroni-corrected p values for MEFs versus endogenous neuron. Plotted are the enriched common genes among our trans-differentiated neurons (50.3% of all differentially upregulated genes in the endogenous neurons and 61.6% of all trans-differentiated upregulated genes at the 0.01 Bonferroni-corrected level), endogenous-specific genes, MEF-specific genes, and trans-differentiated neuron-specific genes. The line indicates 0.01 Bonferroni-corrected FDR level.

(B and C) Top enriched GO terms for the genes commonly upregulated (B) or downregulated (C) in Ascl1-, Ezh2+Ngn1-, and Ezh2+Brn2-directly reprogrammed neurons and endogenous neurons samples.

(D) Expression in normalized log count per million (CPM) in the directly reprogrammed cells induced by indicated factors and somatic samples for neuronal regional and subtype-specific genes.

often required for more stringent cell-identity conversion (such as from fibroblast instead of stem cells).

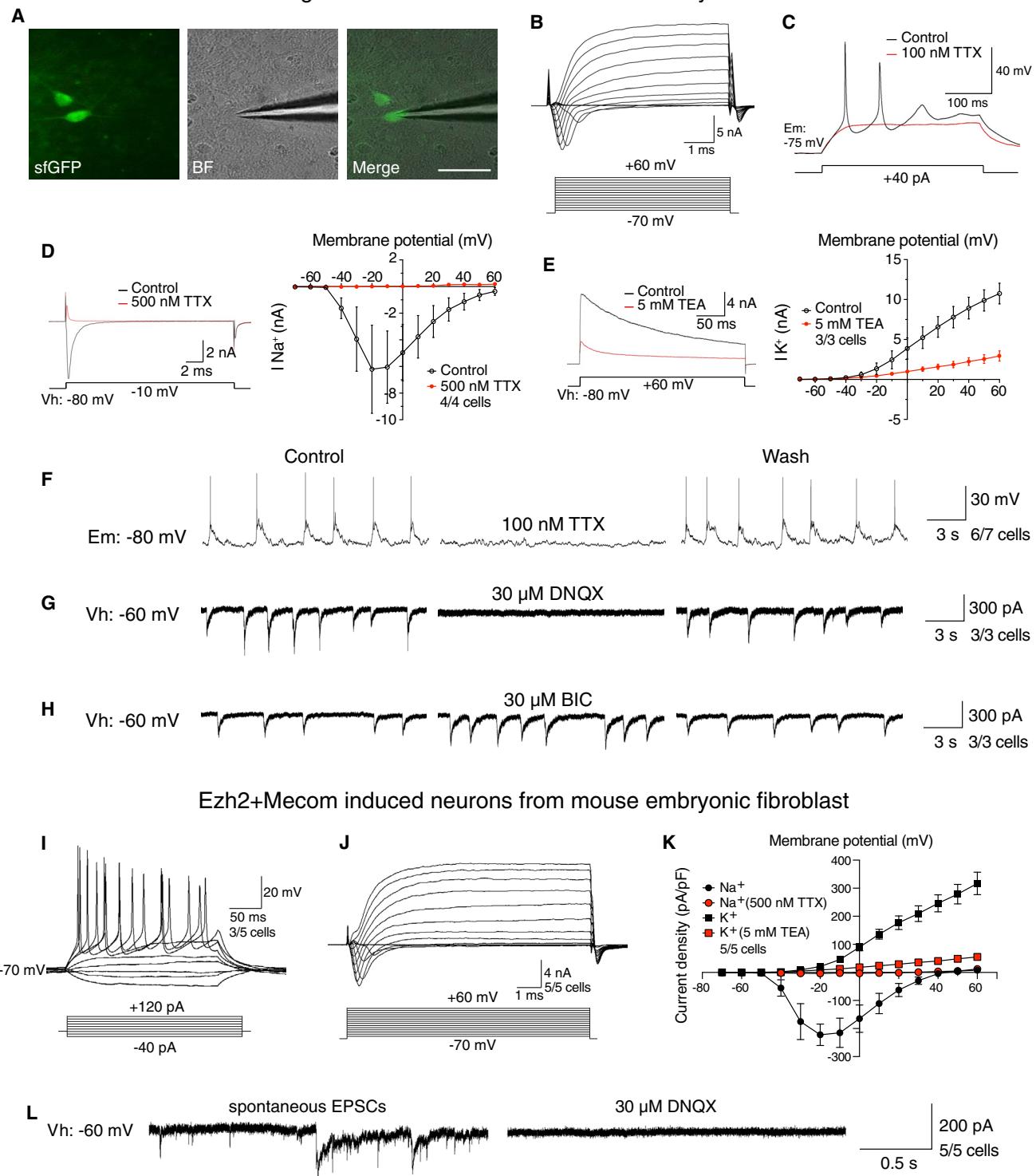
Further staining showed that neurons induced by these combinations expressed other neuronal markers (Figures 5C and S6A). Moreover, we also generated neuronal cells from tail tip fibroblasts using three combinations (Ngn1+Brn2, Ngn1+Ezh2, and Ngn1+Foxo1) after 26 days of infection (Figures 5D and S6B). Together, these results demonstrate that our approach is a useful system to discover factor combinations for direct reprogramming and define gene pairs for more efficient cell-fate engineering.

To characterize our directly reprogrammed neurons, we analyzed their transcriptomes by RNA-seq. As expected, these neurons shared a core regulatory program with endogenous neuron samples (Tsunemoto et al., 2018; Wapinski et al., 2013; Figure 6A). GO analysis of the core gene cluster showed categories associated with neuronal generation and synaptic trans-

mission were highly enriched (Figures 6B and 6C). Comparison of our directly reprogrammed neurons with the brain samples showed that, similar to neurons generated by Ascl1 expression, our neurons showed a strong glutamatergic profile (Figure 6D). Meanwhile, the Ezh2+Brn2 combination induced neurons with high expression of hindbrain markers, *Gbx2*, *Hoxa2*, and *Hoxa4*. These data suggested that our directly reprogrammed neurons acquired a global transcriptional program that resembles the endogenous neurons with subtype diversification.

To more rigorously characterize these induced neurons derived from fibroblasts, we next examined electrophysiological properties of neurons induced by Ezh2+Ngn1, Ezh2+Mecom, and Ngn1+Brn2. We observed and recorded eliciting action potentials, inward and outward currents, and spontaneous EPSCs in these neurons (Figures 7A–7L and S6C–S6H). In addition, in the presence of Bicuculline, a GABA<sub>A</sub> receptor antagonist, the

## Ezh2+Ngn1 induced neurons from mouse embryonic fibroblast

**Figure 7. Electrophysiological Characterization of Directly Reprogrammed Neurons Driven by Identified Factor Combinations**(A) Recording electrode patched onto a sfGFP<sup>+</sup> cell with a stimulation electrode. Scale bar, 10  $\mu$ m.

(B) Representative traces of whole-cell currents in voltage-clamp mode; cells were held at -80 mV.

(C) Representative trace of evoked membrane potential (black). Application of TTX inhibited the action potential (red).

(D) Inward sodium currents were evoked from an induced neuron (black), and application of TTX inhibited these currents (red). Cells were held at -80 mV (right panel); a representative trace of whole-cell current with and without TTX at -10 mV membrane potential in voltage-clamp mode is shown (left panel).

(legend continued on next page)

observed frequency and amplitude of the spontaneous postsynaptic currents slightly increased (Figures 7H and S6H). These experiments demonstrated that functional synapses could be formed in directly reprogrammed neurons induced by Ezh2+Ngn1, Ezh2+Mecom, or Ngn1+Brn2.

## DISCUSSION

Here, we demonstrated a systematic approach using a series of individual and combinatorial CRISPRa screens to study genetic drivers of neuronal-fate determination. We developed an integrated experimental and computational platform to functionally characterize the full set of TFs and other regulators and the GIs between core factors for neuronal-fate determination using a high-density single-gene profiling and GI maps. Our strategy offers an approach for studying the system biology of cell-identity regulatory networks, providing a useful methodology for understanding cell-fate determination.

Our approach enabled systematic identification of factors for neuronal fates determination, some of which were previously unlinked to neuronal differentiation or direct reprogramming. These findings provide insights into the molecular mechanisms governing cellular reprogramming. For example, Ezh2 and Suz12, two Polycomb-group proteins that are global epigenetic regulators (Margueron and Reinberg, 2011), appeared as top hits and were validated to promote neuronal differentiation. Our transcriptome analyses further suggested that Ezh2 inhibited alternative endodermal and mesodermal lineages. Studying its exact mechanisms in the future might help elucidate how global epigenetic changes favor a specific cell fate and potentially provide drug targets for degenerative diseases.

By pairwise activation of core neuronal-inducing factors, we generated a quantitative GI map for the neuronal-fate decision. The GI map predicted multiple factor pairs that enhance the induction of the neuronal fate. We validated the pairs with the lowest FDR and found that they showed positive effects on neuronal differentiation. These effects for ESC to neuron differentiation were moderate, correlative to that individual single factors already exhibited strong differentiation effects. Compared to the additive effects, several factor pairs showed strong synergistic effects on neuronal direct reprogramming (Figure 5A). For example, the combinations of Ezh2+Mecom and Ezh2+Brn2 efficiently generated neurons from fibroblasts, whereas none of them alone induced neuronal direct reprogramming. The drastic

increase in neuronal direct reprogramming efficiency using two factors compared to a single factor suggests that a higher degree of synergism between driver genes is required in fibroblasts compared to ESCs.

As an enzymatic subunit and core component of the PRC2 complex, Ezh2 participates in histone methylation and is primarily thought to repress gene expression. A recent work demonstrates that Ezh2 controls REST stability independently of PRC2 and acts as an anti-neurogenic barrier during direct neuronal reprogramming of human fibroblasts (Lee et al., 2018). This is in contrast to our study, where overexpression of Ezh2 promoted neuronal differentiation and reprogramming both from ESCs and from fibroblasts. The pro-neuronal effect of Ezh2 is likely PRC2 dependent. First, this is supported by the fact that Suz12, another PRC2 component, was also identified as a pro-neuronal hit from our unbiased screen. In addition, Lee et al. found that the repression of PRC2 components Suz12 and *EED* led to the failed neuronal conversion, also highlighting the importance of PRC2 during neuronal generation. Second, we performed a transcriptome analysis of the most downregulated genes in Ezh2-induced neurons and found that many of these genes were putative PRC2 targets (Figure S5A; Liu et al., 2010). This suggests that overexpression of Ezh2 in our study may have enhanced PRC2-mediated repression, thus promoting neuronal differentiation. Third, in our direct reprogramming experiments, Ezh2 dramatically promoted neuronal generation in the presence of other factors (such as Ngn1, Brn2, and Mecom), suggesting that it may prime the chromatin of fibroblasts and enhance the binding of pro-neuronal factors. Together, our results provide deeper insights on how Ezh2, as a global epigenetic modulator in the PRC2 complex identified from unbiased genetic screens, affects neuronal-fate conversion.

We mapped the identified neuronal-promoting factors, Ezh2, Foxo1, Ngn1, onto the STRING database (Szklarczyk et al., 2015) and found that they directly interacted with  $\beta$ -catenin, a core factor that regulates neuronal differentiation and function (Hoogeboom et al., 2008; Otero et al., 2004; Wisniewska, 2013). Furthermore,  $\beta$ -catenin activates *Ngn1* and *Brn2* as downstream genes (Goodall et al., 2004; Israsena et al., 2004). We incorporated the observations in this study into a hypothetical gene interaction network, in which the Wnt/ $\beta$ -catenin pathway acts as a hub (Figure S7). The direct reprogramming factors, Ezh2 and Foxo1, may directly bind to  $\beta$ -catenin and redirect its transcriptional activity (Hoogeboom et al., 2008; Shi et al.,

(E) Outward potassium currents were evoked from an induced neuron (black), and application of TEA inhibited these currents (red). Cells were held at  $-80$  mV (right panel); a representative trace of whole-cell current with and without TEA at  $+60$  mV membrane potential in voltage-clamp mode is shown (left panel).

(F) Spontaneous action potentials recorded from an induced neuron (left panel). Application of TTX blocked the action potentials (middle panel). Washout of TTX reversed the blockade (right panel).

(G) Representative traces of evoked EPSCs recorded from an induced neuron (left panel). Application of DNQX blocked the response of EPSCs (middle panel). Washout of DNQX reversed the blockade (right panel).

(H) Representative traces of evoked EPSCs recorded from an induced neuron (left panel). Application of BIC (Bicuculline), a GABA receptor antagonist, slightly increased the frequency and amplitude of EPSCs (middle panel). Washout of BIC reversed the increase (right panel).

(I) Representative traces showing action potentials elicited from directly reprogrammed neurons.

(J) Representative traces of whole-cell currents in voltage-clamp mode.

(K) Current density plot of voltage activated currents in directly reprogrammed neurons. The graph shows current density (y axis, pA/pF) for peak outward (potassium, squares) and inward currents (sodium, circles) plotted against voltage step (x axis, mV) in the absence (black) and presence of either TEA (red squares) or TTX (red circles).

(L) Representative traces of evoked EPSCs recorded from an induced neuron (left panel). Application of DNQX blocked the response of EPSCs (right panel). Error bars indicate  $\pm$  SEM of cell counts.

Ngn1+Ezh2 (A–H) and Ezh2+Mecom (I–L) induced neurons were recorded.

2007). Several differentiation factors, such as Tcf15 and Nr4a1, also regulate β-catenin via expression modulation, degradation, or adherent activity (Cheng et al., 2015; Davies et al., 2013; Sánchez-Tilló et al., 2010; Sun et al., 2012). Taken together with previous knowledge on β-catenin, our CRISPRa approach in this study allows us to propose a model with a hypothetic interaction network centered on Wnt/β-catenin, with a number of identified factors to enhance the transcriptional activity of β-catenin and govern neuronal-fate determination.

In sum, the ability to activate a large-set of endogenous genes and rapidly profile GI maps between genes provides a powerful tool to understand the regulatory network of the cell identity, and for discovering factors to promote desired cell-type engineering. This study opens up possibilities in understanding the biology of cellular reprogramming, which could facilitate other strategies to reprogram cells into clinically relevant cell types.

## STAR★METHODS

Detailed methods are provided in the online version of this paper and include the following:

- KEY RESOURCES TABLE
- CONTACT FOR REAGENT AND RESOURCE SHARING
- EXPERIMENTAL MODEL AND SUBJECT DETAILS
  - Cell Culture
  - Mice
  - Rats
  - Primary Cell Culture
  - Construction of the CamES Cell Line
  - Construction of the *Tubb3*-hCD8 CamES Cell Line
- METHOD DETAILS
  - Plasmids Construction
  - sgRNA Library Design
  - sgRNA Library Construction
  - Double sgRNA Library Construction
  - Virus Production
  - High-Throughput Pooled Neuronal Differentiation Screens
  - High-Throughput Pooled Combinatorial Screens
  - Generation of mouse fibroblasts
  - Generation of Induced Neurons
  - Maturation of induced neurons
  - Quantitative RT-PCR
  - Flow Cytometry, Cell Surface Staining and Cell Sorting
  - Immunocytochemistry
  - Electrophysiology
  - Transcriptome Analysis by RNA-seq
- QUANTIFICATION AND STATISTICAL ANALYSIS
  - FACS, Realtime-PCR and Electrophysiology
  - Induced Neurons Efficiency Calculation
  - Bioinformatic Analysis of sgRNA and Gene Hits
  - Bioinformatic Analysis of Double Screens
- DATA AND SOFTWARE AVAILABILITY

## SUPPLEMENTAL INFORMATION

Supplemental Information includes seven figures and seven tables and can be found with this article online at <https://doi.org/10.1016/j.stem.2018.09.003>.

## ACKNOWLEDGMENTS

The authors thank the Qi lab members for technical help and reading the manuscript. We thank Stanford Functional Genomics Facilities (SFGF) for technical help with RNA-seq. This work was supported by NIH Director's Early Independence Award (grant OD017887 to L.S.Q.), NIH 4DN U01 (grant EB021240 to L.S.Q.), Department of Defense Breast Cancer Research Program, Breakthrough Award (W81XWH-17-1-0018), Pew Biomedical Scholarship, Alfred P. Sloan Fellowship, a gift fund from the Kate Li Foundation, and a gift from the Li Ka Shing Foundation (to L.S.Q.).

## AUTHOR CONTRIBUTIONS

Conceptualization, L.S.Q., Y.L., and C.Y.; Investigation, Y.L., C.Y., F.W., W.S.C., C.S., H.L., D.Z., H.W., and X.S.X.; Computational Data Analysis, T.P.D., S.B., and X.L.; Writing—Original Draft, L.S.Q., Y.L., C.Y., and T.P.D.; Writing—Review & Editing; W.H.W. and M.W.; Supervision, L.S.Q., S.D., and W.H.W.

## DECLARATION OF INTERESTS

L.S.Q. and Y.L. are inventors on a patent application (US Provisional Patent # 62/443,401) describing compositions and methods for identifying and using stem cell differentiation markers. All other authors declare no competing interests.

Received: February 28, 2018

Revised: July 23, 2018

Accepted: September 4, 2018

Published: October 11, 2018

## REFERENCES

- Adamson, B., Norman, T.M., Jost, M., Cho, M.Y., Nuñez, J.K., Chen, Y., Villalta, J.E., Gilbert, L.A., Horlbeck, M.A., Hein, M.Y., et al. (2016). A multiplexed single-cell CRISPR screening platform enables systematic dissection of the unfolded protein response. *Cell* 167, 1867–1882.e21.
- Anders, S., Pyl, P.T., and Huber, W. (2015). HTSeq—A Python framework to work with high-throughput sequencing data. *Bioinformatics* 31, 166–169.
- Ashburner, M., Ball, C.A., Blake, J.A., Botstein, D., Butler, H., Cherry, J.M., Davis, A.P., Dolinski, K., Dwight, S.S., Eppig, J.T., et al.; The Gene Ontology Consortium (2000). Gene ontology: Tool for the unification of biology. *Nat. Genet.* 25, 25–29.
- Bassik, M.C., Kampmann, M., Lebbink, R.J., Wang, S., Hein, M.Y., Poser, I., Weibezahn, J., Horlbeck, M.A., Chen, S., Mann, M., et al. (2013). A systematic mammalian genetic interaction map reveals pathways underlying ricin susceptibility. *Cell* 152, 909–922.
- Black, J.B., Adler, A.F., Wang, H.G., D'Ippolito, A.M., Hutchinson, H.A., Reddy, T.E., Pitt, G.S., Leong, K.W., and Gersbach, C.A. (2016). Targeted epigenetic remodeling of endogenous loci by CRISPR/Cas9-based transcriptional activators directly converts fibroblasts to neuronal cells. *Cell Stem Cell* 19, 406–414.
- Cahan, P., Li, H., Morris, S.A., Lummertz da Rocha, E., Daley, G.Q., and Collins, J.J. (2014). CellNet: Network biology applied to stem cell engineering. *Cell* 158, 903–915.
- Carpenter, P., Gelman, A., Hoffman, M.D., Lee, D., Goodrich, B., Betancourt, M., Brubaker, M., Guo, J., Li, P., and Riddell, A. (2017). Stan: A Probabilistic Programming Language. *J Stat Softw* 76, 32.
- Chavez, A., Scheiman, J., Vora, S., Pruitt, B.W., Tuttle, M., P R Iyer, E., Lin, S., Kiani, S., Guzman, C.D., Wiegand, D.J., et al. (2015). Highly efficient Cas9-mediated transcriptional programming. *Nat. Methods* 12, 326–328.
- Chen, B., Gilbert, L.A., Cimini, B.A., Schnitzbauer, J., Zhang, W., Li, G.W., Park, J., Blackburn, E.H., Weissman, J.S., Qi, L.S., and Huang, B. (2013). Dynamic imaging of genomic loci in living human cells by an optimized CRISPR/Cas system. *Cell* 155, 1479–1491.

- Chen, S., Sanjana, N.E., Zheng, K., Shalem, O., Lee, K., Shi, X., Scott, D.A., Song, J., Pan, J.Q., Weissleder, R., et al. (2015). Genome-wide CRISPR screen in a mouse model of tumor growth and metastasis. *Cell* 160, 1246–1260.
- Cheng, P., Sun, X., Yin, D., Xu, F., Yang, K., Qin, L., Dong, Y., Guo, F., Chen, A., Zhang, W., and Huang, H. (2015). Nanog down-regulates the Wnt signaling pathway via  $\beta$ -catenin phosphorylation during epidermal stem cell proliferation and differentiation. *Cell Biosci.* 5, 5.
- D'Alessio, A.C., Fan, Z.P., Wert, K.J., Baranov, P., Cohen, M.A., Saini, J.S., Cohick, E., Charniga, C., Dadon, D., Hannett, N.M., et al. (2015). A systematic approach to identify candidate transcription factors that control cell identity. *Stem Cell Reports* 5, 763–775.
- Davies, O.R., Lin, C.Y., Radziszewska, A., Zhou, X., Taube, J., Blin, G., Waterhouse, A., Smith, A.J., and Lowell, S. (2013). Tcf15 primes pluripotent cells for differentiation. *Cell Rep.* 3, 472–484.
- Dixit, A., Parnas, O., Li, B., Chen, J., Fulco, C.P., Jerby-Arnon, L., Marjanovic, N.D., Dionne, D., Burks, T., Raychowdhury, R., et al. (2016). Perturb-Seq: Dissecting molecular circuits with scalable single-cell RNA profiling of pooled genetic screens. *Cell* 167, 1853–1866.e17.
- Dobin, A., Davis, C.A., Schlesinger, F., Drenkow, J., Zaleski, C., Jha, S., Batut, P., Chaisson, M., and Gingeras, T.R. (2013). STAR: ultrafast universal RNA-seq aligner. *Bioinformatics* 29, 15–21.
- Efron, B. (2012). Large-Scale Inference: Empirical Bayes Methods for Estimation, Testing, and Prediction (Cambridge University Press).
- Gilbert, L.A., Larson, M.H., Morsut, L., Liu, Z., Brar, G.A., Torres, S.E., Stern-Ginossar, N., Brandman, O., Whitehead, E.H., Doudna, J.A., et al. (2013). CRISPR-mediated modular RNA-guided regulation of transcription in eukaryotes. *Cell* 154, 442–451.
- Gilbert, L.A., Horlbeck, M.A., Adamson, B., Villalta, J.E., Chen, Y., Whitehead, E.H., Guimaraes, C., Panning, B., Ploegh, H.L., Bassik, M.C., et al. (2014). Genome-scale CRISPR-mediated control of gene repression and activation. *Cell* 159, 647–661.
- Goodall, J., Martinuzzi, S., Dexter, T.J., Champeval, D., Carreira, S., Larue, L., and Goding, C.R. (2004). Brn-2 expression controls melanoma proliferation and is directly regulated by  $\beta$ -catenin. *Mol. Cell. Biol.* 24, 2915–2922.
- Heinäniemi, M., Nykter, M., Kramer, R., Wienecke-Baldacchino, A., Sinkkonen, L., Zhou, J.X., Kreisberg, R., Kauffman, S.A., Huang, S., and Shmulevich, I. (2013). Gene-pair expression signatures reveal lineage control. *Nat. Methods* 10, 577–583.
- Hoogeboom, D., Essers, M.A.G., Polderman, P.E., Voets, E., Smits, L.M.M., and Burgering, B.M.T. (2008). Interaction of FOXO with  $\beta$ -catenin inhibits  $\beta$ -catenin/T cell factor activity. *J. Biol. Chem.* 283, 9224–9230.
- Israsena, N., Hu, M., Fu, W., Kan, L., and Kessler, J.A. (2004). The presence of FGF2 signaling determines whether  $\beta$ -catenin exerts effects on proliferation or neuronal differentiation of neural stem cells. *Dev. Biol.* 268, 220–231.
- Koike-Yusa, H., Li, Y., Tan, E.P., Velasco-Herrera, Mdel.C., and Yusa, K. (2014). Genome-wide recessive genetic screening in mammalian cells with a lentiviral CRISPR-guide RNA library. *Nat. Biotechnol.* 32, 267–273.
- Konermann, S., Brigham, M.D., Trevino, A.E., Joung, J., Abudayyeh, O.O., Barcena, C., Hsu, P.D., Habib, N., Gootenberg, J.S., Nishimasu, H., et al. (2015). Genome-scale transcriptional activation by an engineered CRISPR-Cas9 complex. *Nature* 517, 583–588.
- Lee, S.W., Oh, Y.M., Lu, Y.-L., Kim, W.K., and Yoo, A.S. (2018). MicroRNAs overcome cell fate barrier by reducing EZH2-controlled REST stability during neuronal conversion of human adult fibroblasts. *Dev. Cell* 46, 73–84.e7.
- Li, W., Köster, J., Xu, H., Chen, C.-H., Xiao, T., Liu, J.S., Brown, M., and Liu, X.S. (2015). Quality control, modeling, and visualization of CRISPR screens with MAGeCK-VISPR. *Genome Biol.* 16, 281.
- Liu, Y., Shao, Z., and Yuan, G.-C. (2010). Prediction of Polycomb target genes in mouse embryonic stem cells. *Genomics* 96, 17–26.
- Liu, H., Wei, Z., Dominguez, A., Li, Y., Wang, X., and Qi, L.S. (2015). CRISPERA: A comprehensive design tool for CRISPR-mediated gene editing, repression and activation. *Bioinformatics* 31, 3676–3678.
- Mali, P., Aach, J., Stranges, P.B., Esvelt, K.M., Moosburner, M., Kosuri, S., Yang, L., and Church, G.M. (2013). CAS9 transcriptional activators for target specificity screening and paired nickases for cooperative genome engineering. *Nat. Biotechnol.* 31, 833–838.
- Margueron, R., and Reinberg, D. (2011). The Polycomb complex PRC2 and its mark in life. *Nature* 469, 343–349.
- Mudge, J.M., and Harrow, J. (2015). Creating reference gene annotation for the mouse C57BL6/J genome assembly. *Mamm. Genome* 26, 366–378.
- Otero, J.J., Fu, W., Kan, L., Cuadra, A.E., and Kessler, J.A. (2004).  $\beta$ -catenin signaling is required for neural differentiation of embryonic stem cells. *Development* 131, 3545–3557.
- Parnas, O., Jovanovic, M., Eisenhaure, T.M., Herbst, R.H., Dixit, A., Ye, C.J., Przybylski, D., Platt, R.J., Tirosh, I., Sanjana, N.E., et al. (2015). A Genome-wide CRISPR Screen in Primary Immune Cells to Dissect Regulatory Networks. *Cell* 162, 675–686.
- Robinson, M.D., McCarthy, D.J., and Smyth, G.K. (2010). edgeR: A Bioconductor package for differential expression analysis of digital gene expression data. *Bioinformatics* 26, 139–140.
- Sánchez-Tilló, E., Lázaro, A., Torrent, R., Cuatrecasas, M., Vaquero, E.C., Castells, A., Engel, P., and Postigo, A. (2010). ZEB1 represses E-cadherin and induces an EMT by recruiting the SWI/SNF chromatin-remodeling protein BRG1. *Oncogene* 29, 3490–3500.
- Shalem, O., Sanjana, N.E., Hartenian, E., Shi, X., Scott, D.A., Mikkelsen, T., Heckl, D., Ebert, B.L., Root, D.E., Doench, J.G., and Zhang, F. (2014). Genome-scale CRISPR-Cas9 knockout screening in human cells. *Science* 343, 84–87.
- Shi, B., Liang, J., Yang, X., Wang, Y., Zhao, Y., Wu, H., Sun, L., Zhang, Y., Chen, Y., Li, R., et al. (2007). Integration of estrogen and Wnt signaling circuits by the polycomb group protein EZH2 in breast cancer cells. *Mol. Cell. Biol.* 27, 5105–5119.
- Sun, Z., Cao, X., Jiang, M.M., Qiu, Y., Zhou, H., Chen, L., Qin, B., Wu, H., Jiang, F., Chen, J., et al. (2012). Inhibition of  $\beta$ -catenin signaling by nongenomic action of orphan nuclear receptor Nur77. *Oncogene* 31, 2653–2667.
- Szklarczyk, D., Franceschini, A., Wyder, S., Forslund, K., Heller, D., Huerta-Cepas, J., Simonovic, M., Roth, A., Santos, A., Tsafou, K.P., et al. (2015). STRING v10: Protein-protein interaction networks, integrated over the tree of life. *Nucleic Acids Res.* 43, D447–D452.
- Takahashi, K., and Yamanaka, S. (2006). Induction of pluripotent stem cells from mouse embryonic and adult fibroblast cultures by defined factors. *Cell* 126, 663–676.
- Tanenbaum, M.E., Gilbert, L.A., Qi, L.S., Weissman, J.S., and Vale, R.D. (2014). A protein-tagging system for signal amplification in gene expression and fluorescence imaging. *Cell* 159, 635–646.
- The Gene Ontology Consortium (2017). Expansion of the gene ontology knowledgebase and resources. *Nucleic Acids Res.* 45 (D1), D331–D338.
- Tsunemoto, R., Lee, S., Szűcs, A., Chubukov, P., Sokolova, I., Blanchard, J.W., Eade, K.T., Brüggemann, J., Wu, C., Torkamani, A., et al. (2018). Diverse reprogramming codes for neuronal identity. *Nature* 557, 375–380.
- Vierbuchen, T., Ostermeier, A., Pang, Z.P., Kokubu, Y., Südhof, T.C., and Wernig, M. (2010). Direct conversion of fibroblasts to functional neurons by defined factors. *Nature* 463, 1035–1041.
- Wang, T., Wei, J.J., Sabatini, D.M., and Lander, E.S. (2014). Genetic screens in human cells using the CRISPR-Cas9 system. *Science* 343, 80–84.
- Wapinski, O.L., Vierbuchen, T., Qu, K., Lee, Q.Y., Chanda, S., Fuentes, D.R., Giresi, P.G., Ng, Y.H., Marro, S., Neff, N.F., et al. (2013). Hierarchical mechanisms for direct reprogramming of fibroblasts to neurons. *Cell* 155, 621–635.
- Wingender, E., Dietze, P., Karas, H., and Knüppel, R. (1996). TRANSFAC: A database on transcription factors and their DNA binding sites. *Nucleic Acids Res.* 24, 238–241.
- Wisniewska, M.B. (2013). Physiological role of  $\beta$ -catenin/TCF signaling in neurons of the adult brain. *Neurochem. Res.* 38, 1144–1155.
- Xu, J., Du, Y., and Deng, H. (2015). Direct lineage reprogramming: Strategies, mechanisms, and applications. *Cell Stem Cell* 16, 119–134.
- Yu, C., Liu, Y., Ma, T., Liu, K., Xu, S., Zhang, Y., Liu, H., La Russa, M., Xie, M., Ding, S., and Qi, L.S. (2015). Small molecules enhance CRISPR genome editing in pluripotent stem cells. *Cell Stem Cell* 16, 142–147.

## STAR★METHODS

## KEY RESOURCES TABLE

REAGENT or RESOURCE	SOURCE	IDENTIFIER
<b>Antibodies</b>		
Rabbit polyclonal anti-Oct4	Santa Cruz Technology	Cat# sc-9081; RRID:AB_2167703
Mouse monoclonal anti-Tubulin, beta, Class III (Tuj1)	Covance	Cat# MMS-435P; RRID:AB_2313773
Rabbit polyclonal anti-Map2	Cell Signaling Technology	Cat# 4542; RRID:AB_10693782
Rabbit monoclonal anti-NeuN	Abcam	Cat# ab177487; RRID:AB_2532109
Rabbit polyclonal anti- Glial fibrillary acidic protein (GFAP)	Dako	Cat# Z0334; RRID:AB_10013382
Rabbit polyclonal anti-Olig 2	Millipore Sigma	Cat# AB9610; RRID:AB_570666
Rabbit polyclonal anti- Synapsin I	Abcam	Cat# ab64581; RRID:AB_1281135
Mouse monoclonal anti-GAD65	Abcam	Cat# ab26113; RRID:AB_448989
Mouse monoclonal anti-PSA-NCAM	Miltenyi Biotec	Cat# 130-093-273; RRID:AB_10828799
Mouse monoclonal anti-Human CD8	BD Bioscience	Cat# 555369; RRID:AB_398595
<b>Chemicals, Peptides, and Recombinant Proteins</b>		
Puromycin	Gold Biotech	Cat# P-600-1
Doxycycline hyclate (Dox)	Sigma-Aldrich	Cat# D9891
Polybrene	Sigma-Aldrich	Cat# H9268
Gelatin	Sigma-Aldrich	Cat# G1890
Fetal Bovine Serum (FBS)	Clontech	Cat# 631367
PD0325901	Stemgent	Cat# 04-0006-10
CHIR99021	Stemgent	Cat# 04-0004-10
Leukemia Inhibitory Factor (LIF)	Millipore Sigma	Cat# ESG1107
<b>Critical Commercial Assays</b>		
EasySep Human CD8 Positive Selection Kit	STEMCELL Technologies	Cat# 18053
<b>Experimental Models: Cell Lines</b>		
E14 mouse embryonic stem cells	ATCC	Cat# CRL-1821; RRID:CVCL_9108
HEK293T	ATCC	Cat# CRL-3216; RRID:CVCL_0063
Primary Rat Cortex Neurons	ThermoFisher	Cat# A1084001
CamES	This paper	N/A
<b>Experimental Models: Organisms/Strains</b>		
Mouse: C57BL/6	The Jackson Laboratory	Stock No: 000664
Rat: Sprague Dawley	Charles River	Strain code: 400
<b>Oligonucleotides</b>		
PrimeTime qPCR Primer Assays: Ascl1	Integrated DNA Technologies	Mm.PT.58.42452590.g
PrimeTime qPCR Primer Assays: Ngn2	Integrated DNA Technologies	Mm.PT.58.5623012.gs
PrimeTime qPCR Primer Assays: GFAP	Integrated DNA Technologies	Mm.PT.58.6609337
PrimeTime qPCR Primer Assays: Sox10	Integrated DNA Technologies	Mm.PT.58.42371609
PrimeTime qPCR Primer Assays: Olig2	Integrated DNA Technologies	Mm.PT.58.42319010
PrimeTime qPCR Primer Assays: Cux2	Integrated DNA Technologies	Mm.PT.58.7105048
PrimeTime qPCR Primer Assays: Sox5	Integrated DNA Technologies	Mm.PT.58.31160682
PrimeTime qPCR Primer Assays: Tbr1	Integrated DNA Technologies	Mm.PT.58.42090842
PrimeTime qPCR Primer Assays: Otx1	Integrated DNA Technologies	Mm.PT.58.42606516
PrimeTime qPCR Primer Assays: NG2	Integrated DNA Technologies	Mm.PT.58.29461721
PrimeTime qPCR Primer Assays: En-1	Integrated DNA Technologies	Mm.PT.58.29490484
PrimeTime qPCR Primer Assays: HoxB4	Integrated DNA Technologies	Mm.PT.58.32332549
PrimeTime qPCR Primer Assays: GAD1	Integrated DNA Technologies	Mm.PT.58.10501737

(Continued on next page)

**Continued**

REAGENT or RESOURCE	SOURCE	IDENTIFIER
PrimeTime qPCR Primer Assays: Nestin	Integrated DNA Technologies	Mm.PT.58.5953877
PrimeTime qPCR Primer Assays: Ctip2	Integrated DNA Technologies	Mm.PT.58.31107142
PrimeTime qPCR Primer Assays: GLT1	Integrated DNA Technologies	Mm.PT.58.41489267
sgRNA sequence: See <a href="#">Tables S2</a> and <a href="#">S3</a>	This paper	N/A
Primers used in qPCR: See <a href="#">Table S5</a>	This paper	N/A
Primers used in Hiseq and Miseq: See <a href="#">Table S4</a>	This paper	N/A
Recombinant DNA		
pHR-TRE3G-dCas9-SunTag-mCherry	This paper	N/A
pHR-EF1-alpha-Tet-on 3G	This paper	N/A
pHR-PGK-scFV-sfGFP-VP64	This paper	N/A
pHR-EF1-alpha-puro-T2A-Tet-on 3G_TRE3G_Ascl1	This paper	N/A
Tet-O-FUW-Ascl1	Addgene	Cat# 27150
Tet-O-FUW-Brn2	Addgene	Cat# 27151
Tet-O-FUW-Mty1l	Addgene	Cat# 27152
Tet-O-FUW-Ngn1	This paper	N/A
Tet-O-FUW-Ezh2	This paper	N/A
Tet-O-FUW-Foxo1	This paper	N/A
Tet-O-FUW-Nr4a1	This paper	N/A
Tet-O-FUW-Dmrt3	This paper	N/A
Tet-O-FUW-Jun	This paper	N/A
Tet-O-FUW-Suz12	This paper	N/A
Tet-O-FUW-Nr3c1	This paper	N/A
Tet-O-FUW-Tcf15	This paper	N/A
Tet-O-FUW-Zeb1	This paper	N/A
Tet-O-FUW-Mecom	This paper	N/A
Tet-O-FUW-Hoxc8	This paper	N/A
Tet-O-FUW-Nr2f1	This paper	N/A
Deposited Data		
RNA-seq samples	This paper	GEO: GSE118560
Software and Algorithms		
R	R Core Team, 2017	<a href="https://www.r-project.org/">https://www.r-project.org/</a>
MAGeCK MLE v0.5.6	Li et al., 2015	<a href="https://sourceforge.net/p/mageck/wiki/Home/">https://sourceforge.net/p/mageck/wiki/Home/</a>
stan	Carpenter et al., 2017	<a href="http://mc-stan.org/">http://mc-stan.org/</a>
STAR	Dobin et al., 2013	<a href="https://github.com/alexdobin/STAR">https://github.com/alexdobin/STAR</a>
HTseq	Anders et al., 2015	<a href="https://htseq.readthedocs.io/en/release_0.10.0/">https://htseq.readthedocs.io/en/release_0.10.0/</a>
edgeR	Robinson et al., 2010	<a href="https://bioconductor.org/packages/release/bioc/html/edgeR.html">https://bioconductor.org/packages/release/bioc/html/edgeR.html</a>
geneontology.org	The Gene Ontology Consortium, 2017	<a href="http://www.geneontology.org/">http://www.geneontology.org/</a>
map_paired_guides	This paper	<a href="https://github.com/timydaley/sgrNAdesign">https://github.com/timydaley/sgrNAdesign</a>

**CONTACT FOR REAGENT AND RESOURCE SHARING**

Further information and requests for resources and reagents should be directed to and will be fulfilled by the Lead Contact, Lei S Qi ([stanley.qi@stanford.edu](mailto:stanley.qi@stanford.edu))

## EXPERIMENTAL MODEL AND SUBJECT DETAILS

### Cell Culture

E14 mouse ES cells and CamES cells were maintained on gelatin-coated tissue culture plates with basal medium (50% Neurobasal, 50% Dulbecco modified Eagle medium (DMEM)/Ham's nutrient mixture F12, 0.5% NEAA, 0.5% Sodium Pyruvate, 0.5% GlutaMax, 0.5% N2, 1% B27, 0.1mM  $\beta$ -mercaptoethanol and 0.05 g/L bovine albumin fraction V; all from Thermo Fisher Scientific) supplemented with LIF (Millipore) and 2i (CHIR99021 and PD0325901, Stemgent). Human embryonic kidney (HEK293T) cells (ATCC) were cultured in 10% fetal bovine serum (Thermo Fisher Scientific) in DMEM (Thermo Fisher Scientific).

### Mice

Mice used in this study were C57BL/6 background. We did not sort genders because the sex of fetuses and neonates is difficult to be determined. The mice were housed in the specific-pathogen-free (SPF) animal facility at the Gladstone Institutes, University of California, San Francisco. The mice were housed under controlled conditions (temperature ( $21 \pm 2^\circ\text{C}$ ), humidity ( $55 \pm 10\%$ ), 12 h light/dark cycle). The animals were not involved in any previous procedures and were test naive. All procedures involving mice described in the present study were approved by the Institutional Animal Care and Use Committee at the University of California, San Francisco.

### Rats

The rat pups used in this study were of the Sprague-Dawley strain. All rats were maintained in a controlled, SPF environment (temperature:  $21.5 \pm 4.5^\circ\text{C}$  / relative humidity: 35%–55%) under a standard 12-hour light / 12-hour dark lighting cycle (lights on at 06:00). The rat pups were born in-house at the AfaSci animal facility from female timed-pregnant rats (Charles River). The rat pups were allowed to remain with the lactating female rat until the pups were used for cell culture (~postnatal day 1). We did not sort genders because the sex of neonates is difficult to be determined. The timed-pregnant females as well as the pups were not involved in any previous procedures and were tested naive. Procedures involving rats were approved by the Institutional Animal Care and Use Committee (IACUC) of AfaSci Research Laboratory.

### Primary Cell Culture

Primary cultures of cortex neurons were prepared from postnatal day 1 Sprague Dawley rats. Rats were decapitated, and their brains were removed in pre-cooled physiological saline. The cortex was dissected. Tissues were slightly minced and placed into a Papain Dissociation solution (Worthington Biochemical Corporation) containing 20 units/ml papain and 0.005% DNase in Earle's Balanced Salt Solution (Thermo Fisher Scientific). The solution was equilibrated in 95% O<sub>2</sub>, 5% CO<sub>2</sub> before the tissue was incubated at 37°C for 1 hour. After incubation, the tissue and solution mixture was triturated. Undissociated tissue was allowed to settle and the cloudy suspension was removed and centrifuged at 300 x g for 5 minutes. The supernatant was then discarded and the cell pellet was resuspended in a DNase/albumin-inhibitor solution. A discontinuous density gradient was prepared by gently layering the cell suspension on top of an albumin-inhibitor solution in a centrifuged tube. The mixture was centrifuged at 145 x g for 5 minutes. The supernatant was discarded and the neurons were resuspended in Neurobasal (Invitrogen) medium containing 2% B27 supplement, 2 mM glutamine and 0.5% penicillin/streptomycin. A total of 250,000 cells were plated onto a well of 24-well plates that had been pre-treated with 12.5 µg/ml poly-D-lysine (Sigma). The plates were incubated at 37°C in a 5% CO<sub>2</sub>/95% air incubator and half of the medium was changed every three days. One week after culturing primary neurons, the induced neurons were gently removed from the dishes by trypsin dissociation and were replated onto the primary neurons. Electrophysiological recordings were performed between day 14 and day 21 after replating.

### Construction of the CamES Cell Line

Mouse ES cells were co-transduced with lentiviral constructs that express dCas9-SunTag from a TRE3G promoter, scFV-sfGFP-VP64 from the PGK promoter, and reverse tetracycline-controlled transactivator (rtTA) from the EF1a promoter. After adding Doxy-cycline, polyclonal cells were sorted by flow cytometry using a BD FACS Aria2 for GFP+ and mCherry+ cells. After verification of endogenous gene activation for a few genes (Etv2, Nobox, Hoxc11, Brn2, Lhx4) using individual sgRNAs, we sorted monoclonal cells, and chose one efficient cell line as the final CamES cell line for screening.

### Construction of the *Tubb3-hCD8* CamES Cell Line

2.5 µg pSLQ1654-sgTubb3, 12.5 µg Tubb3-IRES-hCD8 template DNA in 100 µL Nucleofector solution (Amaxa) were electroporated into  $1 \times 10^6$  CamES cells using program A-030. Both plasmids were maxiprepped using the Endofree Maxiprep Kit (QIAGEN). After 3 days of culture, sorted single cells were seeded in a 96-well plate with one cell per well. All clonal cell lines were analyzed using PCR and sequencing (Yu et al., 2015). Primers used: 5' homologous arm F: 5'-aaaagtgccacactgacactcgtcttagatgtcggtgg-3'. 5' homologous arm R: 5'-tcacttggccctggct-3'. IRES-human CD8 F: 5'-caggggcccaagtgaacttagataattcgcccctctccctc-3'. IRES-human CD8 R: 5'-cagtcgcgacaacttaacctgcaaaaaggaggcagttaaagg-3'. 3' homologous arm F: 5'-agtgcgtcgacgtgggt-3'. 3' homologous arm R: 5'-agctggagaccgtttctgactggatacaggcat-3'. The pX330-derived pSLQ1654 encoding the nuclease Cas9 and an optimized sgRNA sequence was linearized by a BbsI digest and gel purified. Two primers sgTubb3 F (5'-caccgcctaagtgaagtggctcgac-3') and sgTubb3 R (5'-aaacgcgtcgagcaacttcactggc-3') were phosphorylated, annealed, and ligated to the linearized vector

pSLQ1654 to generate pSLQ1654-sgTubb3. The Tubb3-IRES-hCD8 vector (pSLQ1760) was assembled with three fragments (5' homologous arm of Tubb3, IRES-hCD8, and 3' homologous arm of Tubb3) and a modified pUC19 backbone vector by using Gibson Assembly Master Mix (New England Biolabs). Both 5' and 3' homology arms were PCR amplified from the genomic DNA extracted from mES cells with Herculase II Fusion DNA polymerase (Agilent). The IRES-hCD8 was PCR amplified from pSLQ1729. The backbone vector was linearized by digestion with Pmel and Zral. All DNA fragments and the backbone vector were gel purified.

## METHOD DETAILS

### Plasmids Construction

To clone individual sgRNAs, the lentiviral vector (pSLQ1373) expressing an optimized sgRNA was linearized and gel purified (Chen et al., 2013). New sgRNA sequences were PCR amplified from pSLQ1373 using different forward primers and a common reverse primer, gel purified and ligated to the linearized pSLQ1373 vector using In-Fusion cloning (Clontech). Primers used to construct individual sgRNAs are shown in [Table S2](#).

The double-sgRNA expression vectors were assembled by a two-step cloning procedure ([Figure S4A](#)). Individual sgRNA sequences were PCR amplified, and ligated onto the pSLQ5004 vector that contained a modified human U6 promoter (hU6) linearized by BstXI and Xhol. The same individual sgRNA sequences were also cloned onto pSLQ1373 that contained a modified mouse U6 promoter (mU6). The double-sgRNA expression vectors were constructed by inserting PCR products, amplified from pSLQ5004 containing one sgRNA, onto Nsil-linearized pSLQ1373 vector with another sgRNA. Primers (Integrated DNA Technologies) used to construct individual sgRNAs are shown in [Table S3](#).

### sgRNA Library Design

Putative TFs were selected according to the TRANSFAC database (Wingender et al., 1996), and transcription start site (TSS) for each gene was determined using the Gencode (Mudge and Harrow, 2015) and refFlat databases. We regard different isoforms of TF genes as different TFs. We designed sgRNAs primarily targeting a region of -3kb to 0 upstream of TSS. Using the CRISPR-era algorithm (Liu et al., 2015), we computed the targeting sequences of sgRNAs adjacent to a NGG protospacer adjacent motif (PAM), starting with a G (for better U6 promoter activity), and containing a protospacer length of 20 bp. The sgRNAs containing homopolymers spanning greater than 3 nucleotides (nt) were discarded. To avoid off-target effects, every sgRNA sequence was aligned to the whole mouse genome using the short read aligner Bowtie, and those with 0, 1, or 2 mismatches with another genomic region were excluded. We also excluded Ascl1 and Ngn2 (known neurogenic factors) from our library because we used them to validate the efficiency of CRISPRa for neuronal differentiation ([Figure S1](#)), and also because we were concerned that the signal from Ascl1 and Ngn2 induced neurons would drown out the signal of novel factor hits if they were not excluded from the library. Furthermore, sgRNA sequences that contained BstXI and Bpl restriction sites were excluded for construction convenience. We kept only sgRNAs with a GC content between 30% and 70%. An average of about 20 sgRNAs were selected for each factor. Sequences for non-targeting negative control sgRNAs were generated using a scrambled yeast chromosome region and selected using the same rules as described above.

### sgRNA Library Construction

The oligonucleotide library (Custom Array) was PCR amplified, gel purified, and ligated into the linearized pSLQ1373 vector digested with BstXI and Bpl using In-Fusion cloning. Libraries and the parental vector will be made available on <http://www.addgene.org>.

### Double sgRNA Library Construction

We designed a library of 44 sgRNAs including a set of 19 genes by using the top predicted hits from the single screens and six non-targeting sgRNAs. We excluded any sgRNAs containing Nsil restriction sites, which were used for combinatorial sgRNA library construction. We synthesized individual oligos encoding sgRNAs in a 96-well format (IDT), and cloned each sgRNA into pSLQ1373 individually as previously described. We also cloned the same individual sgRNAs into pSLQ5004 as previously described. After sequencing validation, we manually mixed all pSLQ1373-sgRNA constructs and all pSLQ5004-sgRNA constructs separately in equal amounts for double sgRNA library construction. To generate the pooled double sgRNA library, the sgRNA sequence were PCR amplified from pooled pSLQ5004-sgRNA constructs, gel purified, and ligated onto the Nsil-digested pooled pSLQ1373-sgRNA constructs using In-Fusion cloning (Clontech). The double sgRNA-library pools were prepared in Stellar competent cells (TaKaRa) and purified with a Plasmid Maxi Kit (QIAGEN). Primers (IDT) used to construct individual sgRNAs are shown in [Table S3](#). The representation of each of the double-sgRNA constructs was then quantified by NGS with the primers listed in [Table S4](#).

### Virus Production

HEK293T cells were seeded at ~30% confluence one day before transfection. Lentivirus was produced by co-transfecting with pH plasmids and encoding packaging protein vectors (pMD2.G and pCMV-dR8.91) using TransIT-LT1 transfection reagents (Mirus). Viral supernatants were collected 3 days after transfection and filtered through a 0.45 μm strainer. Supernatant was used for transduction immediately or kept at -80°C for long-term storage.

### High-Throughput Pooled Neuronal Differentiation Screens

The differentiation screens were performed in two independent replicates.

- 1 On day -2,  $10^8$  CamES cells were seeded at 40,000 cells/cm<sup>2</sup> to gelatin-coated plate.
- 2 On day -1, cells were transduced with pooled lentiviral sgRNA library with an MOI of 0.3 in basal medium supplemented with LIF and 2i.
- 3 On day 0, cells were switched to basal medium with Doxycycline at 100 ng/mL and puromycin at 1 µg/mL.
- 4 24 hours after puromycin selection, 25 million cells were harvested as the initial population of the screen.
- 5 Fresh medium with Doxycycline was changed every day starting day 2 (puromycin selection ended).
- 6 On day 12, cells were harvested, stained with hCD8, and sorted for hCD8+ and hCD8- cells using EasySep human CD8 isolation kit (STEMCELL Technologies).
- 7 Genomic DNA was harvested from all samples; the sgRNA-encoding regions were then amplified by PCR using HiSeq forward and reverse primers and then sequenced on an Illumina HiSeq-4000 using HiSeq custom primer with previously described protocols at high coverage ([Bassik et al., 2013](#); [Gilbert et al., 2014](#)). Primers used are summarized in [Table S4](#).
- 8 For the individual sgRNA validation experiments, we used a similar protocol except that CamES cells were transduced with a high MOI. Top hits that show enriched neuronal fate are summarized in [Table S1](#).

### High-Throughput Pooled Combinatorial Screens

The double CRISPRa neuronal differentiation screens were performed as two independent replicates.

- 1 On day -2, 6 million CamES cells were seeded at 40,000 cells/cm<sup>2</sup> to gelatin-coated plate.
- 2 On day -1, cells were transduced with pooled lentiviral sgRNA library with an MOI of 0.3 in basal medium supplemented with LIF and 2i.
- 3 On day 0, cells were switched to basal medium with Doxycycline at 100 ng/mL and puromycin at 1 µg/mL.
- 4 Fresh medium with Doxycycline was changed every day starting day 2 (puromycin selection ended).
- 5 On day 12, cells were harvested, stained with hCD8 and sorted for hCD8+ and hCD8- cells using Aria II cell sorter (BD Biosciences).
- 6 Genomic DNA was harvested from all samples; the double sgRNA-encoding regions were then amplified by PCR using MiSeq forward and reverse primers and then sequenced on an Illumina MiSeq using HiSeq custom primer, which for the first sgRNA, and MiSeq custom primer, which for the second sgRNA. The method to sequence the double sgRNA regions of the constructs is shown in [Figure S4A](#). Primers used are summarized in [Table S4](#).
- 7 For the double sgRNA validation experiments, we used a similar protocol except that CamES cells were transduced with a high MOI.

### Generation of mouse fibroblasts

Embryonic skin-derived fibroblasts were isolated from E13.5 embryos of C57BL/6 mice as previously described ([Vierbuchen et al., 2010](#)). Isolated fibroblasts were cultured and expanded in MEF media (Dulbecco's Modified Eagle Medium, Life Technologies) containing 10% Fetal Bovine Serum (Life Technologies), non-essential amino acids (Life Technologies), and sodium pyruvate (Life Technologies)) for 2 passages before use. Tail tip fibroblasts were isolated from the bottom third of tails from 4-day-old pups (both female and male). Tail tip cells were expanded for 2 passages in MEF media before use.

### Generation of Induced Neurons

- 1 Isolated MEFs were cultured and expanded in MEF media for 2 passages before use. Tail tip fibroblasts were expanded for 2 passages in MEF media before use.
- 2 Matrigel (growth factors reduced; BD Biosciences) was thawed on ice according to the manufacturer's instruction and diluted in pre-chilled PBS with a ratio of 1: 30.
- 3 Diluted Matrigel was added to 24-well plates in sufficient quantity to cover the entire growth surface of the plates. Plates were kept at 37°C for 30 minutes until ready for use.
- 4 Passage 1-2 MEFs were thawed and seeded into the Matrigel-coated plates at a density of 25,000 cells per well of a 24-well plates. Cells were grown in MEF media for 4-5 days until confluent.
- 5 When MEFs were confluent, cells were infected with lentiviruses containing expression constructs of rtTA (driven by ubiquitin promoter) and additional lentiviruses overexpressing Ascl1, Ngn1, Ezh2, Foxo1, Brn2, Nr4a1, Dmrt3, Jun, Suz12, Nr3c1, Tcf15, Zeb1, Mecom, Hoxc8, or Nr2f1 (driven by a Tet-on promoter) in the presence of polybrene (8 mg/ml).
- 6 The next day, media was exchanged with basal medium containing Doxycycline (1 mg/ml).
- 7 Culture media was refreshed every 3-4 days during the induction period.

### Maturation of induced neurons

About 14 days after lentivirus transduction (extensive neurites outgrowth should be observed in this stage), the induced cells are able to be progressed for further maturation. To do so, they were re-plated and co-cultured directly with primary neurons.

- 1 Mouse primary postnatal cortical neurons were isolated and cultured for about 6 days before re-plating the induced cells.
- 2 The induced cells were dissociated by using 0.05% trypsin on the culture plate.
- 3 Cells were centrifuged for 3 min at 1000 rpm at room temperature.
- 4 After carefully discarding the supernatant, fresh maturation medium was added (basal media with addition of 200  $\mu$ M ascorbic acid, 2  $\mu$ M db-cAMP, 25 ng/ml BDNF, 25 ng/ml NT3, and 50 ng/ml GDNF) to gently re-suspend the cells and re-plate them to co-cultured with pre-existing primary neurons.
- 5 The next day, media was exchanged with basal medium containing Doxycycline (1 mg/ml).
- 6 Re-plated cells were co-cultured for about 14 days or longer (depending on the maturation process of the induced cells, which can be observed based on the development of the extensive neuritis outgrowth) to become functionally mature. Half of the basal medium was changed every 3-4 days.

### Quantitative RT-PCR

Cells were harvested using Accutase (STEMCELL), and total RNA was isolated using the RNeasy Plus Mini Kit (QIAGEN), according to manufacturer's instructions. Reverse transcription was performed using iScript cDNA Synthesis kit (Bio-Rad). Quantitative PCR reactions were prepared with iTaq Universal SYBR Green Supermix (Bio-Rad). Reactions were run on a LightCycler thermal cycler (Bio-Rad). Primers used are summarized in [Table S5](#). For the below genes, the primers were ordered from IDT: *Ascl1*, *Ngn2*, *GFAP*, *GLT1*, *Sox10*, *Olig2*, *Cux2*, *Ctip2*, *Sox5*, *Tbr1*, *Otx1*, *En-1*, *HoxB4*, *GAD1*, *Nestin*, and *NG2*.

### Flow Cytometry, Cell Surface Staining and Cell Sorting

The antibody CD8-APC was purchased from BD Biosciences, and Anti-PSA-NCAM-APC was from Miltenyi Biotec. Cells were harvested, washed, and adjusted to a concentration of  $10^6$  cells/mL in ice cold PBS with 2% FBS. Cells were stained and incubated with diluted primary antibodies at 4°C for 30 mins in Eppendorf tubes. After staining, cells were washed three times by centrifugation at 400 g for 5 mins and resuspended in 500  $\mu$ L to 1 mL in ice cold PBS. Cells were kept in dark on ice and analyzed using BD Accuri C6 Cytometer. Cell sorting was performed by using Aria II cell sorter (BD Biosciences).

### Immunocytochemistry

Experiments were performed on cells seeded on a plate (IBIDI) that had been coated with gelatin (0.1%) overnight at 37°C. Cells were washed twice with PBS, fixed in 4% Paraformaldehyde (Wako) for 15 mins at room temperature, permeabilized and blocked with 0.1% Triton X-100, 5% donkey serum in PBS (blocking buffer) for 1 h at room temperature. After three washes with PBS, cells were incubated with primary antibodies at 4°C for overnight, then washed three times with PBS. After staining with corresponding secondary antibodies in blocking buffer for 1 hour at room temperature, cells were washed three times with PBS and stained with DAPI (Vector Labs) for 5 mins. Washed cells were examined using a Nikon Spinning Disk Confocal microscope with TIRF.

### Electrophysiology

For differentiation, eight days after sgRNA transduction and induction with Dox, cells were re-plated neurons onto rat neonatal cortical neurons using basal medium. Electrophysiological functions were tested and recorded after 7-10 days.

For direct reprogramming, after 14 days' co-infection of cDNAs and a superfold GFP (sfGFP) reporter, cells were re-plated onto rat neonatal cortical neurons using basal medium that had been cultured for 7 days *in vitro*. Patch-clamp electrophysiological recordings were performed one week after on sfGFP positive fibroblast-induced neurons. sfGFP positive neurons were located using a Lambda DG-4 illumination system and Q Imaging Fast 1394 Rolera-Mgi Plus camera controlled by Micro-Manager (Version 1.4) mounted on an Olympus BX51WI fluorescence microscope. Whole-cell responses were recorded using an MultiClamp 700 B (Molecular Devices) amplifier and headstage and low-pass filtered at 10 KHz before digitization using a DigiData 1440 data acquisition system (Molecular Devices). Data was stored on a PC running pClamp software (Version 10.4, Molecular Devices). Patch-pipettes were fabricated from 1.5 mm OD borosilicate capillary glass (Warner Instruments) using a microipette puller (Sutter Instrument, Model P-87) to give tip resistances of 2-4 M $\Omega$ . The series resistance for all recordings was under 10 M $\Omega$  (Mean: 5.62 M $\Omega$ , SEM: 0.38, n = 12). Capacitance transients and series resistance errors were compensated for (70%) using the amplifier circuitry.

The sodium and potassium currents were recorded in the voltage-clamp configuration at a holding potential of -80 mV. Spontaneous postsynaptic currents were recorded in the voltage-clamp configuration at a holding potential of -60 mV or -70 mV. Spontaneous action potentials were recorded in neurons held at -60 mV to -80 mV. Action potentials were also evoked by applying depolarizing current.

All experiments were performed at ambient room temperature (25°C). The external solution contained (in mM): NaCl (130), HEPES-Na (10), KCl (5), CaCl<sub>2</sub> (2), Glucose (10). For voltage-gated sodium currents, tetraethylammonium (TEA, 5 mM) was added to the external solution and the internal solution contained (in mM): CsF (120), HEPES (10), EGTA (11), CaCl<sub>2</sub> (1), MgCl<sub>2</sub> (1), TEA-Cl (10),

KOH (11). For voltage-gated potassium currents, tetrodotoxin (TTX, 500 nM) was added to the external solution and the internal solution contained (in mM): KF (120), HEPES (10), EGTA (11), CaCl<sub>2</sub> (1), MgCl<sub>2</sub> (1), KCl (10), KOH (11). For current clamp recordings of action potentials, 2 mM MgATP was added to the internal solution. All recording solutions had pH values of 7.3–7.4 with osmolality of 290–300 mOsm/kg. Drug applications were administered via local perfusion approximately 200 μm from the recorded cells at a flow rate of 0.2 ml/min and solutions were continually withdrawn from the recording chamber by vacuum aspiration. Drugs were applied until responses reached a steady-state level. Electrophysiological data were analyzed offline using Clampfit 10.4 and data was plotted using Graphpad Prism software.

### Transcriptome Analysis by RNA-seq

CamES-derived neuronal cells driven by sgRNAs were purified with PSA-NCAM at day 12. MEF-derived neuronal cells driven by the overexpression of cDNAs were purified with PSA-NCAM at day 12. Total RNA was purified using RNeasy Plus Mini Kit (QIAGEN). Library preparation and next-generation sequencing were performed by Novogene Corporation, Chula Vista, CA.

Reads were mapped with the STAR aligner (Dobin et al., 2013) to the mm10 genome using the ENSEMBL GRCm38.83 annotations. Gene counts were then obtained with HTseq (Anders et al., 2015) with the “intersection-nonempty” option to prevent counting of ambiguous reads. Differential expression analysis was performed with edgeR (Robinson et al., 2010). We analyzed the differentiation samples and the trans-differentiation samples separately. For the differentiation samples we set the CamES cells to be the baseline in the design matrix and encoded the presence of each factor as an indicator variable, as well as the interactions to investigate epistasis. For the trans-differentiation samples, we set MEF as the baseline and encoded each trans-differentiation condition as an indicator variable. We collapsed all endogenous neurons and brain samples into a single category to ensure identifiability in the design matrix and to calculate differentially expressed genes for this group as a whole. To compare samples across studies we also encoded an indicator variable of the study and subtracted these estimated batch effects from all estimated expression values. Gene Ontology results were obtained from Gene Ontology Consortium (Ashburner et al., 2000; The Gene Ontology Consortium, 2017).

## QUANTIFICATION AND STATISTICAL ANALYSIS

### FACS, Realtime-PCR and Electrophysiology

For quantification of differentiation efficacy by FACS (Figures 2B–2E, 3D, 3E, and 5A), values represent the average value of three biologically independent experiments. Error bars represent standard deviations that are calculated in Microsoft EXCEL. For quantification of gene expression by realtime-PCR (Figures 2E, S1C, S1H, S1G, S5B, and S5C), gene expression levels were normalized to the housekeeping gene RiboL7. Data are represented as mean from 3 biological replicates, each individually performed in technical triplicates. Error bars represent standard deviations that are calculated in Microsoft EXCEL. For electrophysiology assays, n represents the cell numbers recorded (Figure 2H n = 5, 7D n = 4, 7E n = 3, 7K n = 5, S3C n = 3, S4E n = 4, S6D n = 4, S6E n = 3). Error bars represent standard error of the mean (SEMs) calculated in Microsoft EXCEL.

### Induced Neurons Efficiency Calculation

The following method was used to calculate the efficiency of neuronal direct reprogramming: The total number of Map2+ cells with a neuronal morphology, defined as cells having a circular, three-dimensional appearance that extend a thin process at least three times longer than their cell body, were quantified 14 days after transduction into MEFs or 26 days after transduction into TTFs. The Map2+ and DAPI+ cells were counted from at least 20 randomly selected images at 20 × magnification for each experiment of every condition. The Map2+ cell number was divided by the number of DAPI+ cells to get a percentage of induced neuron cells. Values represents the average value of three biologically independent experiments, Error bars represent standard deviations that are calculated in Microsoft EXCEL.

### Bioinformatic Analysis of sgRNA and Gene Hits

Data processing was conducted with custom scripts. Reads were mapped allowing for a mismatch for the first and last base pair of the spacer, which uniquely identified each sgRNA. We normalized each sample by the total read count (Table S6). This gave us a frequency for each sgRNA:

$$f_{\text{sgRNA}} = \frac{\text{sgRNA counts}}{\sum \text{sgRNA counts}}$$

We used the paired Tubb3-hCD8+ and Tubb3-hCD8- to compute the enrichment scores. Specifically, we computed frequencies as above, and discarded sgRNA with less than 1 count in the Tubb3-hCD8- library. We computed enrichment for each sgRNA in each replicate as the log<sub>2</sub> fold-change from the Tubb3-hCD8- sample to the Tubb3-hCD8+ libraries. We averaged enrichment across replicates and used this as E<sub>sg</sub> in subsequent analysis.

For each gene, we computed an enrichment score (E<sub>gene</sub>) from the sgRNA enrichment above, as follows. We computed an unnormalized enrichment score (E<sub>gene,top3</sub>) by averaging E<sub>sg</sub> for the 3 sgRNA with highest E<sub>sg</sub>. We normalized E<sub>gene,top3</sub> by the distribution of nontargeting sgRNA as follows.

Suppose a gene had N targeting sgRNA. We drew 10000 bootstrap samples of size N from the nontargeting sgRNA. For each sample of size N, we computed E<sub>sample,top3</sub> as above. This gave us an empirical estimate of the distribution of E<sub>gene,top3</sub> if the all the sgRNA

targeting that gene had been negative control sgRNA. For our final normalized gene enrichment score ( $\rho_1$ ), we divided the unnormalized enrichment score by the 0.9 quantile of this empirical distribution:

$$\rho_1 = \frac{E_{\text{gene,top3}}}{\text{quantile}_{\text{samples}}(E_{\text{samples,top3}}, 0.9)}$$

After ranking genes by ES, we selected the most enriched sgRNA for each gene to subsequently validate.

### Bioinformatic Analysis of Double Screens

We calculated the count matrix by exact match for both ends of each paired end read, throwing all other reads out (Table S7). The correlation of counts between replicates of the same condition was high (0.942 – 0.992), indicating high reproducibility of the double screen. Effect sizes for each guide pair was calculated using MAGeCK MLE (Li et al., 2015).

Suppose the null hypothesis that the guide pair of genes  $i$  and  $j$  have an effect size equal to the maximum of the individual effect size. This will be the case if one gene is the primary driver of neuronal differentiation. We verified this assumption by examining the empirical distribution and verifying that it is centered at zero. Note that the coefficients estimated by MAGeCK ( $\beta_{ij}$  for genes  $i$  and  $j$ , in that order) arise from a generalized linear regression and should, if the model posited by MAGeCK is correct, be normally distributed.

Consider the null hypothesis  $H_0$ : the effect of guide targeting two genes is less than the maximal effect of guides targeting either gene. Let sign  $m(x, y)$  be the function that returns the sign of the larger of the absolute values of the two guides compared to the negative controls. Under the null hypothesis  $\beta_{ij} = \text{signm}(\beta_{i0}, \beta_{0j}) \max(|\beta_{i0}|, |\beta_{0j}|)$ .

To this end, note that the standard deviation of  $\beta_{ij}$  is bounded above by

$$\sqrt{s_{\beta_{i0}}^2 + s_{\beta_{0j}}^2}.$$

Therefore the difference  $\beta_{ij} - \text{signm}(\beta_{i0}, \beta_{0j}) \max(|\beta_{i0}|, |\beta_{0j}|)$  has standard error bounded above by

$$\sqrt{s_{\beta_{ij}}^2 + \sqrt{s_{\beta_{i0}}^2 + s_{\beta_{0j}}^2}}$$

We can then define the gene level interaction score ( $\rho_2$ ) of genes  $i$  and  $j$  as

$$\rho_2(i, j) = \frac{1}{2} \frac{\beta_{ij} - \text{signm}(\beta_{i0}, \beta_{0j}) \max(|\beta_{i0}|, |\beta_{0j}|)}{\sqrt{s_{\beta_{ij}}^2 + \sqrt{s_{\beta_{i0}}^2 + s_{\beta_{0j}}^2}}} + \frac{\beta_{ij} - \text{signm}(\beta_{i0}, \beta_{0j}) \max(|\beta_{i0}|, |\beta_{0j}|)}{\sqrt{s_{\beta_{ij}}^2 + \sqrt{s_{\beta_{j0}}^2 + s_{\beta_{0i}}^2}}}$$

$\rho_2$  is constructed to be similar to a z-score for testing interactions, taking variability into account, but will not necessarily have an exactly normal distribution due to the high correlation between estimates (since we are testing all gene-gene pairs). Therefore, we use an empirical Bayes approach to estimate the null distribution and determine significant genes while appropriately controlling the false discovery rate (Efron, 2012).

Guide level interaction scores (guide-level  $\rho_3$ ) were computed by MCMC in the stan language with a modified James-Stein estimator by subtracting off average gene level interactions and putting a standard normal prior on the guide-level interaction scores. This allows us to pool information across the whole experiment and account for specific differences between high and low expression guides.

### DATA AND SOFTWARE AVAILABILITY

The accession number for the data reported in this paper is NCBI GEO: GSE118560.