



## Supplementary Information for

# Identifying Key Regulators in Neuronal Transdifferentiation by Gene Regulatory Network Analysis

Li Li<sup>a</sup>, Binglin Zhu<sup>a,b</sup>, Jian Feng<sup>a,b,\*</sup>

<sup>a</sup> Department of Physiology and Biophysics, State University of New York at Buffalo, Buffalo, NY 14203, USA., <sup>b</sup> Veterans Affairs Western New York Healthcare System, Buffalo, NY 14215, USA.

\* To whom correspondence should be addressed:

Jian Feng, Ph.D.,

Email: [jianfeng@buffalo.edu](mailto:jianfeng@buffalo.edu)

### This PDF file includes:

Supplemental methods  
Figures S1-S8

## **Supplemental methods**

### *Lentivirus production*

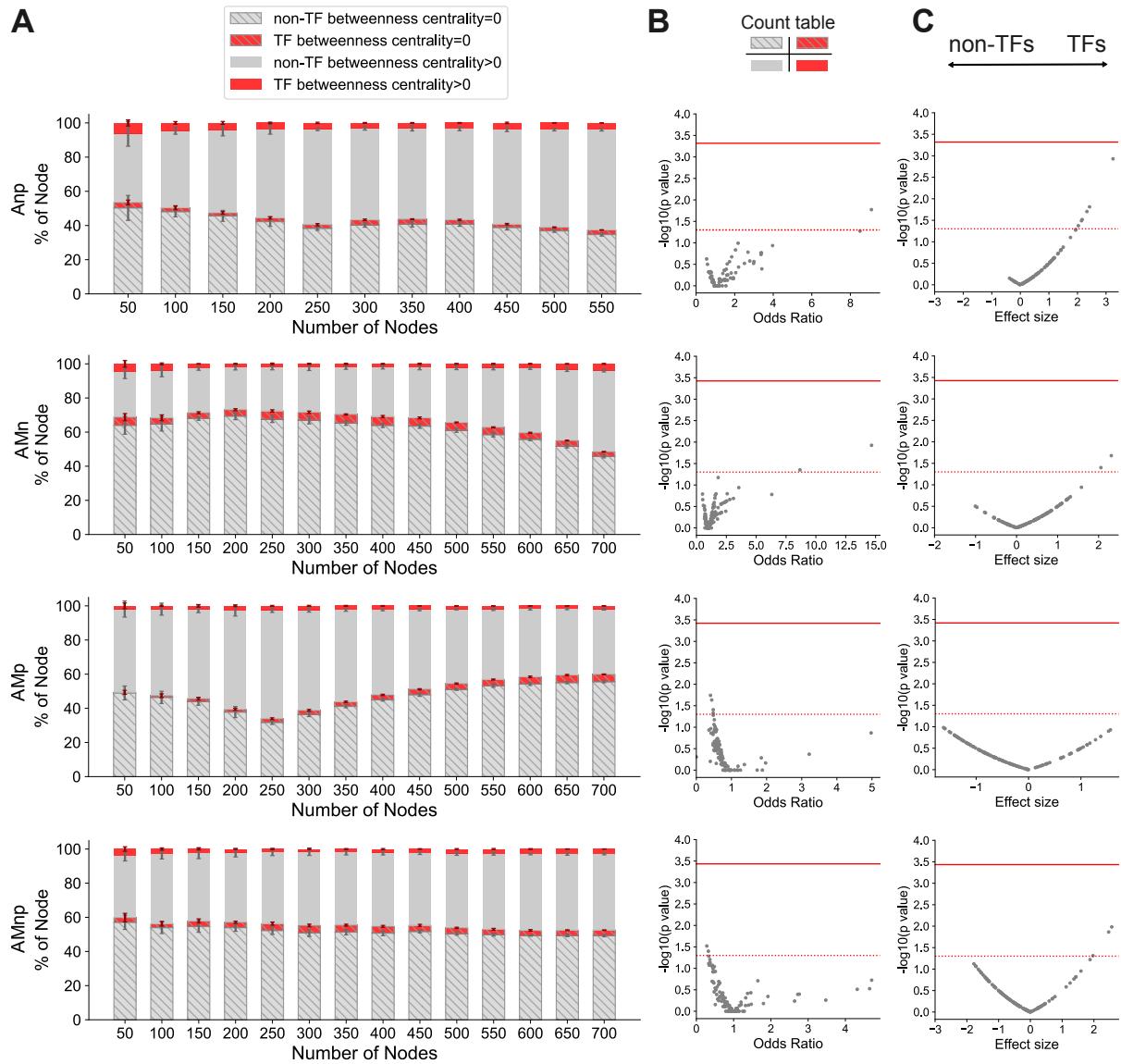
Plasmids for producing lentiviruses include pMD2.G (Addgene, #12259), psPAX2 (Addgene, #12260), pTight-9-124-BclxL (Addgene, #60857), pLKO.1/p53shRNA (Addgene, #19119), pLKO.1/scramble (Addgene, #1864), FUW-M2rtTA (Addgene, #20342) from Addgene, FUW-tetO-LoxP-ASCL1, FUW-tetO-LoxP-LMX1A, pLKO.1/PTBP2shRNA generated previously (7), FUW-tetO-LoxP-OTX2 generated by subcloning OTX2 ORF (GenScript) to FUW-tetO-LoxP, pLKO.1/OTX2shRNA, pLKO.1/LMX1AshRNA generated by subcloning siRNA sequences (OTX2 GCACTGAAACTTTACGACAAA; LMX1A TCAGAACGGGTGATGAGTTGT) to shRNA onto pLKO.1. To produce lentiviruses, HEK293FT cells were transfected with 10 µg vector of interest along with 2.5 µg pMD2.G and 7.5 µg psPAX2 packaging plasmids using 45 µl Lipofectamine 2000 (Invitrogen, 11668019). Lentiviruses were collected at 24 h, 40 h, 48 h post-transfection and titer was determined using p24 ELISA kit (XpressBio, XB-100).

### *Transdifferentiation*

Fibroblasts were seeded at  $1.5 \times 10^4/\text{cm}^2$  onto Matrigel (Corning, 354277) coated 12-well plate (Corning, 3513). After 16 h, the appropriate combination of viruses was added at multiplicity of infection (MOI) of 10 with 8 µg/ml polybrene (Sigma-Aldrich, H9268). After 16 h, medium containing viruses was replaced by DMEM/F12 (GIBCO, 11330032) and maintained for 24 h, then the medium was replaced by neuron induction medium containing DMEM/F12 with 0.1mM NEAA, 1 × N2 supplement (Gibco, 17502048), 1 × B27 (Gibco, 12587010), 100 U/ml penicillin and 100 µg/ml streptomycin, 20 ng/ml NGF (PeproTech, 450-01), 20 ng/ml GDNF (PeproTech, 450-10), 20 ng/ml BDNF (PeproTech, 450-02), 0.5 µM dorsomorphin (Tocris, 3093), 2.5 µM SB431542 (Tocris, 1614), 3 µM CHIR99021 (Tocris, 4423), 0.2 mM Vitamin C (Sigma, A4544), 10 µM Y27632 (Tocris, 1254), 1 µM PD0332991 (Tocris, 4786), 0.5 mM dibutyryl-cAMP (ApexBio, B9001) and 1 µg/ml doxycycline (Sigma, 4090) on transdifferentiation day 0. Medium was changed every two days.

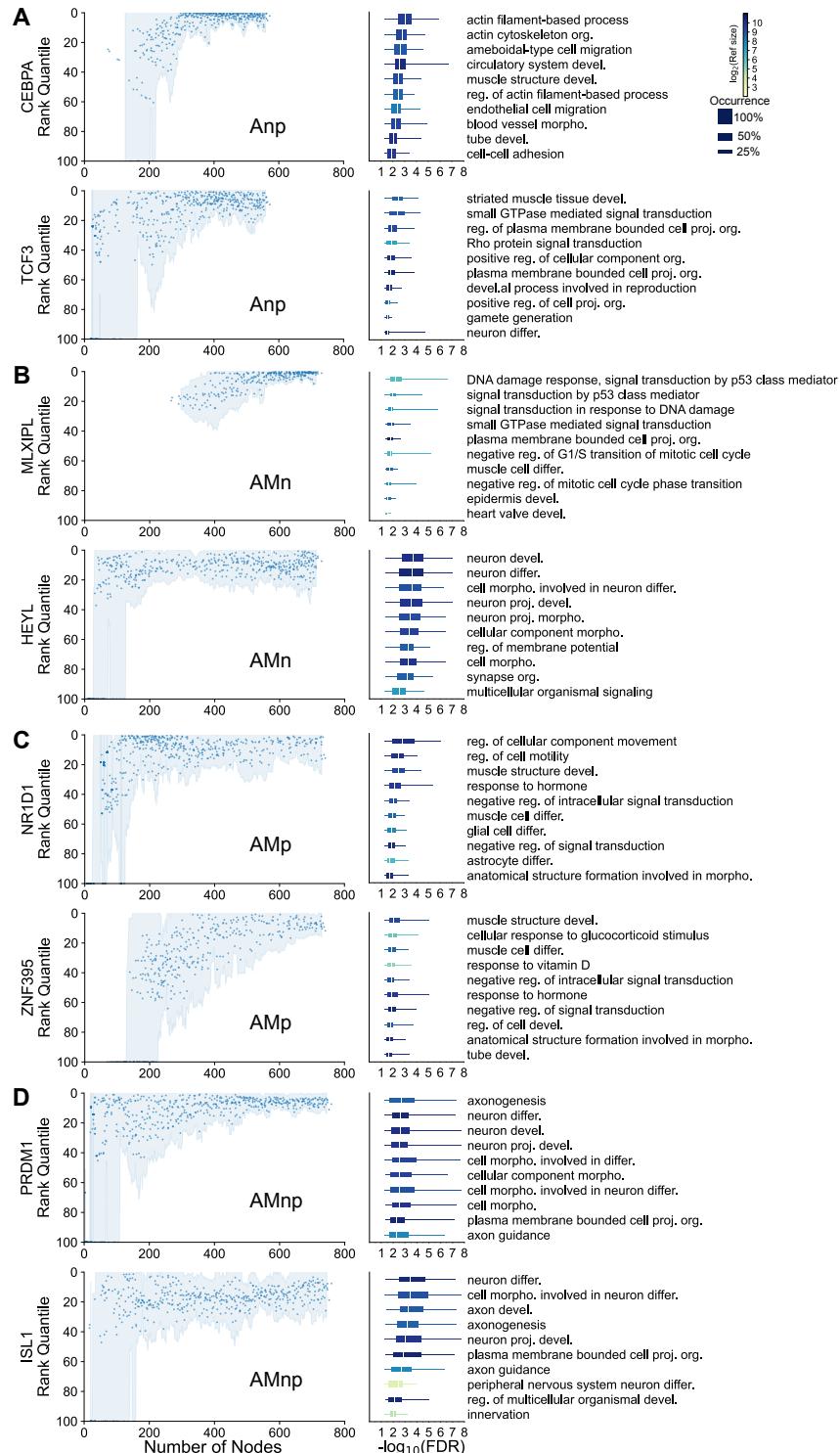
### *Immunohistochemistry and imaging*

Cells were fixed in 4% paraformaldehyde (Sigma, 158127) for 20 min, permeabilized with 0.1% Triton X-100 (VWR, 97062208) for 20 min, and blocked with 3% bovine serum albumin (BSA) (Sigma, A3294) for 1 h at room temperature. The samples were then incubated with primary antibodies (anti-TUJ1, Abcam ab-78078, 1:2000; anti-MAP2, Santa Cruz sc-74421, 1:1000) in 3% BSA at 4°C overnight, and incubated with appropriate secondary antibodies (Alexa Fluor 488, Invitrogen A-21121; Alexa Fluor 546, Invitrogen A-21123) at dilution ratio 1:2000 at room temperature for 2 h. Nuclear staining was performed by incubating with 1µg/ml DAPI (Invitrogen, D1306) for 5 min at room temperature. Stained cells were imaged on Leica AF6000 fluorescence microscope, and manually quantified from 15 frames randomly selected from each sample at 20 × magnification from three independent experiments for each condition.

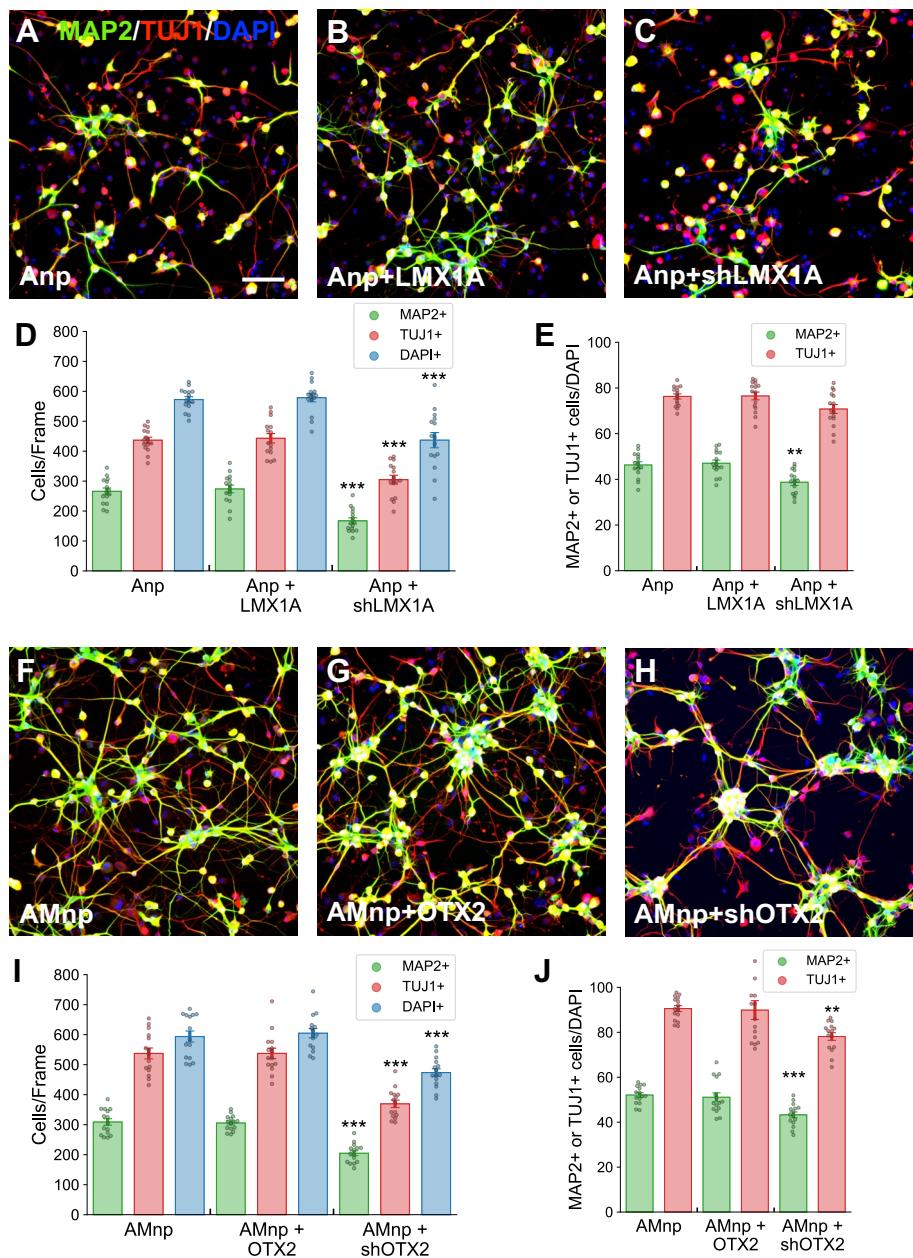


**Figure S1. Comparing betweenness centrality between TFs and non-TFs, associated with Fig. 4.**

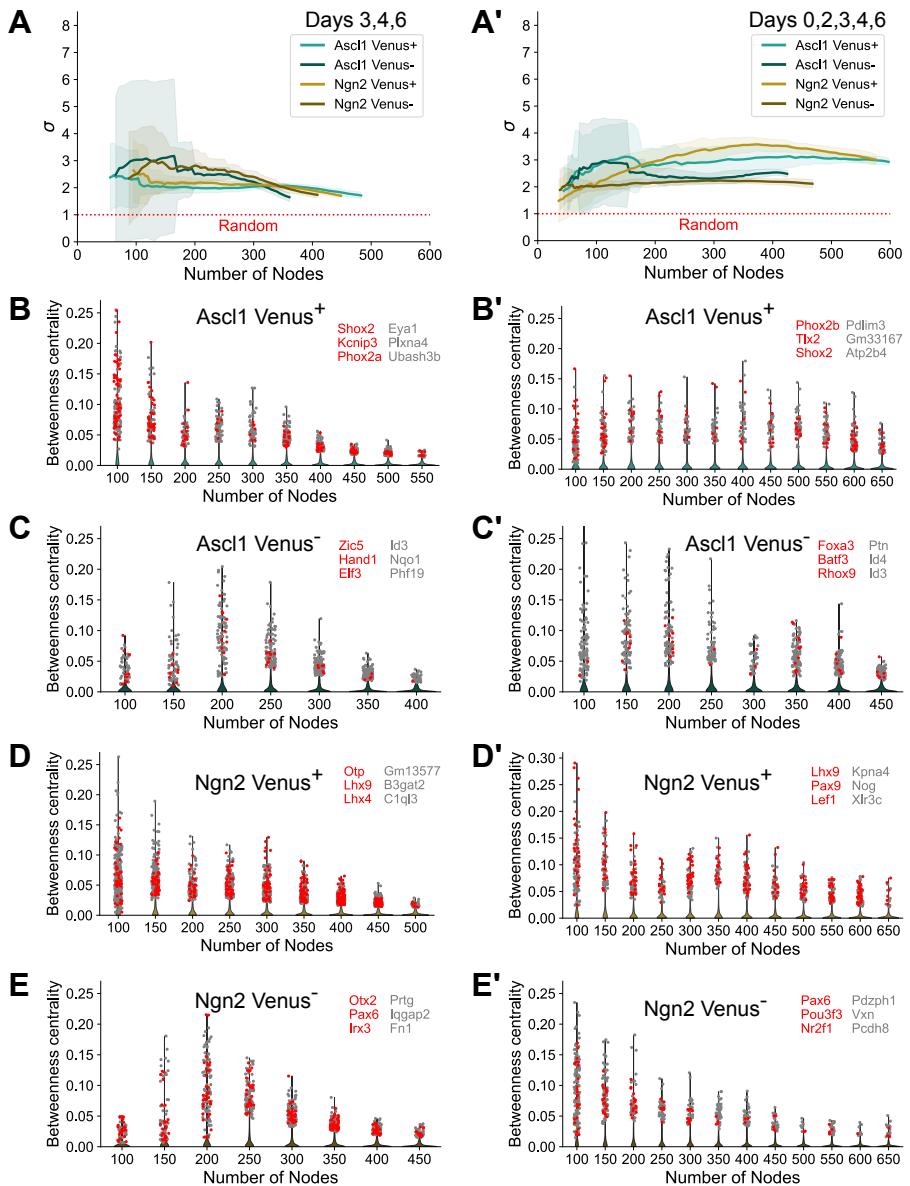
(A) Stacked bars showing the percentage of four types of nodes (TFs with 0 betweenness centrality; non-TFs with 0 betweenness centrality; TFs with positive betweenness centrality; non-TFs with positive betweenness centrality) as labelled. (B) Volcano plots showing odds ratio and p value of Fisher's exact test for each network based on counts of four types of nodes in (A). Significance threshold after Bonferroni correction was marked by red solid line,  $p=0.05$  was marked by red dashed line. (C) Volcano plots showing effect size and p value of Wilcoxon rank-sum test of betweenness centrality between TFs and non-TFs from each network. Positive effect sizes are associated with larger betweenness centrality values in TFs. Significance threshold after Bonferroni correction was marked by red solid line,  $p=0.05$  was marked by red dashed line. Rows from top to bottom are quantifications on Anp, AMn, AMP, AMnp networks, respectively. n=10 runs for GRN construction.



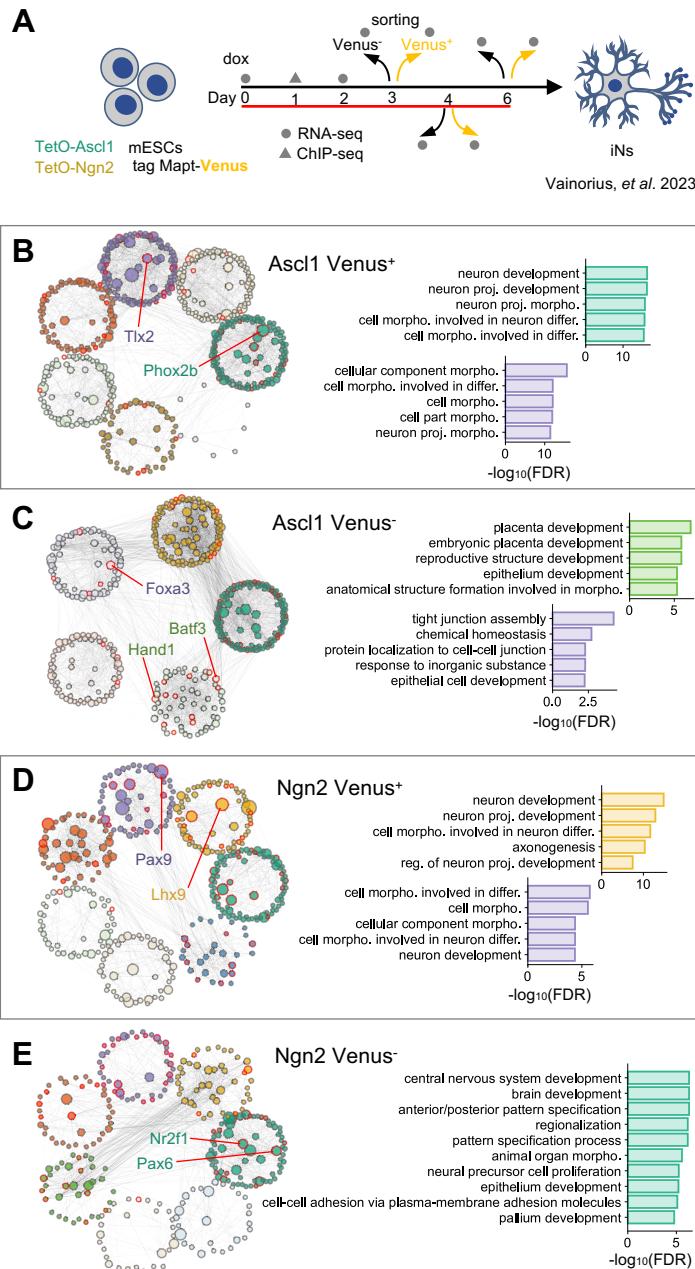
**Figure S2. Top TFs associated with Fig. 5.** (A-D) Left: rank quantile of betweenness centrality for the second and third top TFs identified in the GRN of Anp (A), AMn (B), AMP (C), or AMnp (D) plotted as a blue dot at different network sizes. The light blue shade represents rolling  $\pm$  SD. Right: top 10 replicable GO terms associated with gene communities containing the top TF on the left. Boxes represent the median and interquartile range, and whiskers extend to the lowest and highest FDR ( $FDR < 0.05$ ). The count of observations of the GO term from all communities containing the indicated gene were used to scale the width of the boxes. GO annotation was performed on each community at a given network size but only on communities with at least 20 genes. n=10 runs for GRN construction. devel., development; differ., differentiation; morpho., morphogenesis; org., organization; proj., projection; reg., regulation.



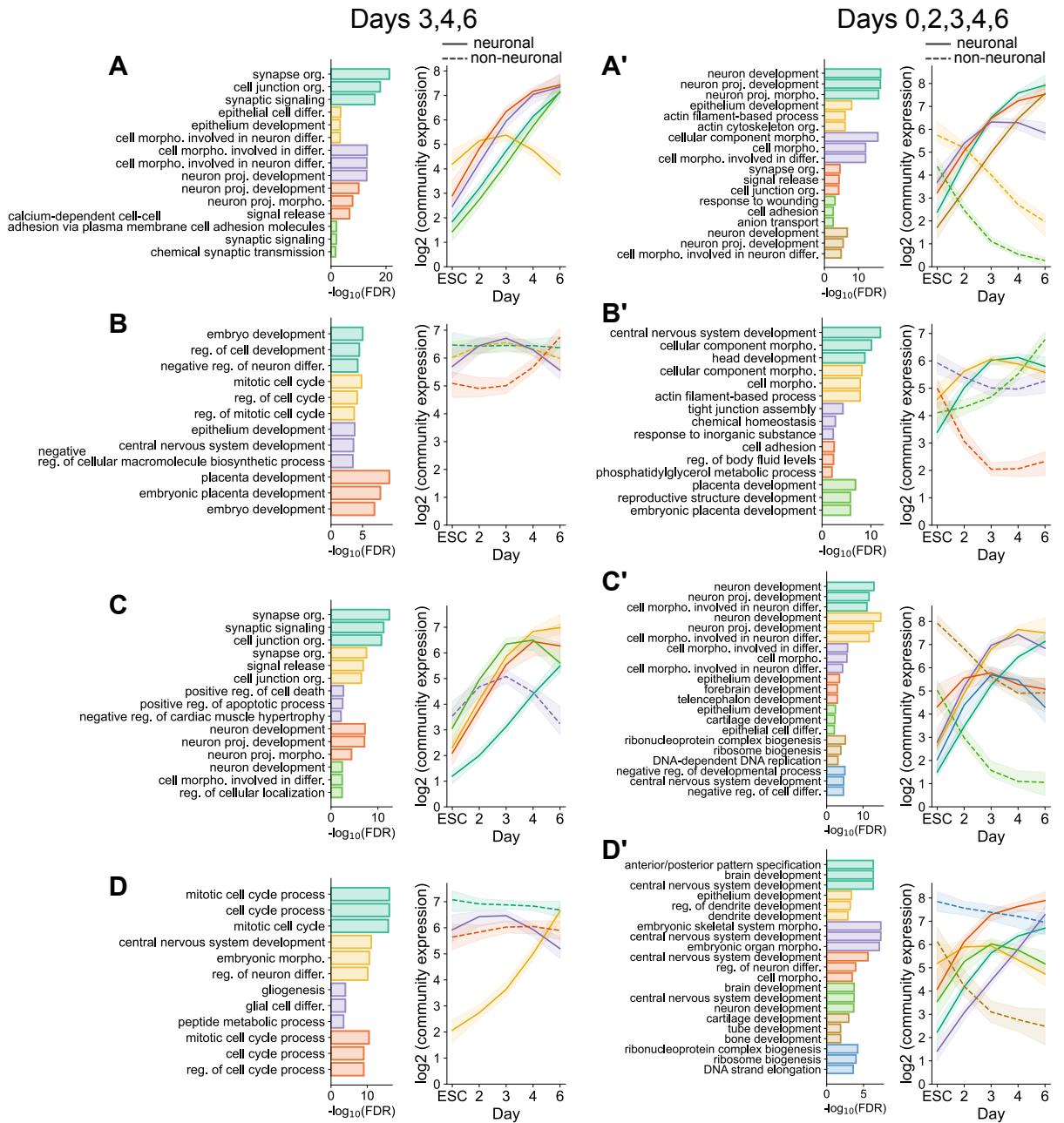
**Figure S3. Transdifferentiation with key regulators associated with Fig. 7.** (A-E) AG22056 human foreskin fibroblasts were converted to iNs with Anp (A), Anp plus LMX1A (B) or Anp plus shLMX1A (C) and co-stained as indicated on day 14. Reprogramming yield in terms of marker-positive cells per frame (D) and reprogramming efficiency in terms of percentage of marker-positive cells among DAPI<sup>+</sup> cells (E) were quantified. (F-J) AG22056 cells were converted to iNs with AMnp (F), or AMnp plus OTX2 (G), or AMnp plus shOTX2 (H) and co-stained as indicated on day 14. Reprogramming yield (I) and reprogramming efficiency (J) were quantified. Scale bar, 100  $\mu$ m. Data are represented as mean  $\pm$  SEM, from three independent experiments, each with 5 random frames per condition. \*\*p<0.01 and \*\*\*p<0.001, vs. the same cell type in Anp (D, E) or AMnp (I, J), one-way ANOVA with Tukey HSD post hoc correction.



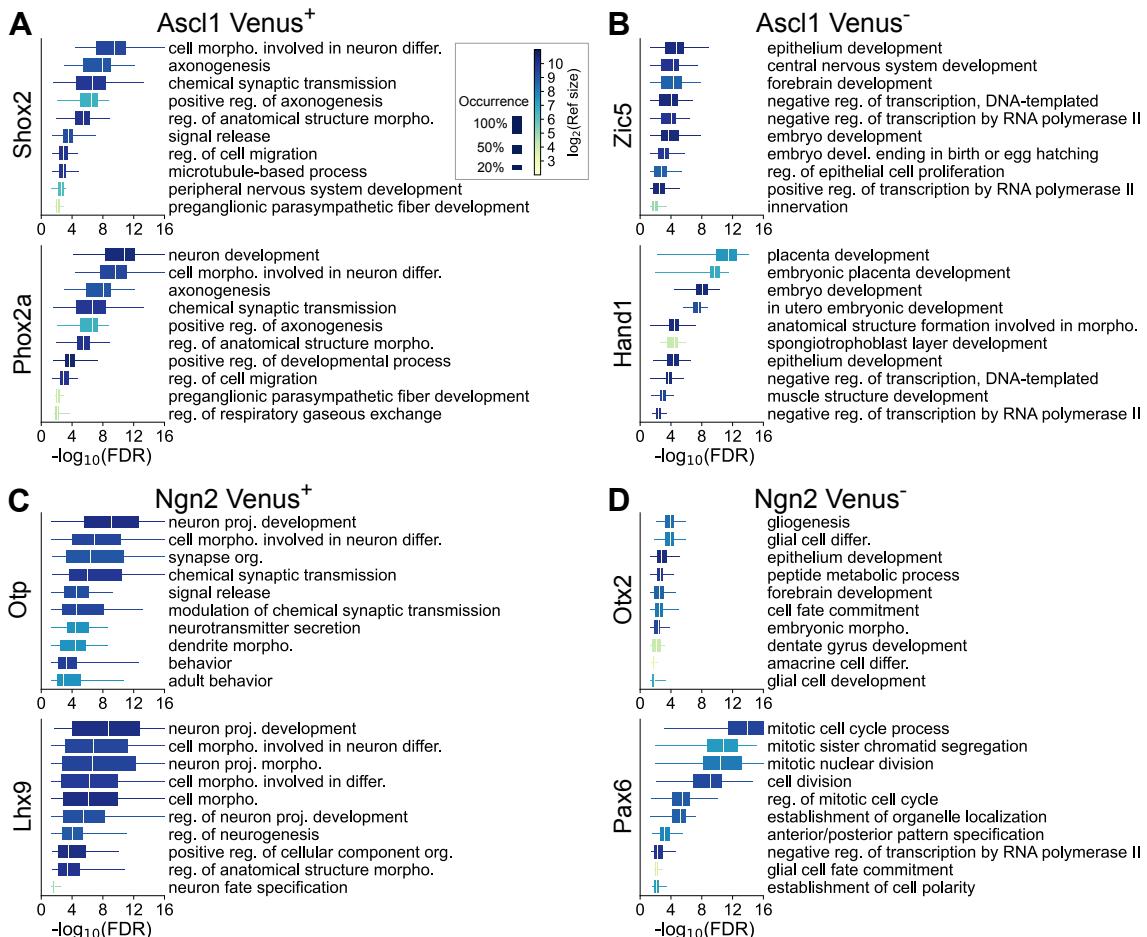
**Figure S4. Topological investigation of GRNs from mESC to neuron differentiation.** (A and A') Small-world coefficient ( $\sigma$ ) against network sizes calculated from GRNs of later days (A) or the entire time series (A'), represented as mean (solid line)  $\pm$  SD (shaded area). The dashed red line represents the threshold between a random ( $\sigma < 1$ ) and small-world ( $\sigma > 1$ ) network. (B-E') Violin plots showing distributions of node betweenness centrality as network size varies from GRNs of later days in Ascl1-Venus<sup>+</sup> (B), Ascl1-Venus<sup>-</sup> (C), Ngn2-Venus<sup>+</sup> (D) and Ngn2-Venus<sup>-</sup> (E), or GRNs of the entire time series in corresponding conditions (B'-E'). Network size is binned into intervals of 50 nodes for the convenience of visualization. The top 10 nodes with highest betweenness centrality in each of the GRNs when network size varies are shown by strip plots. Red and gray represent TFs and non-TFs respectively. Top 3 TFs and non-TFs appearing most frequently in strip plots from each condition are labelled with gene names on the right. n=5 runs for GRN construction.



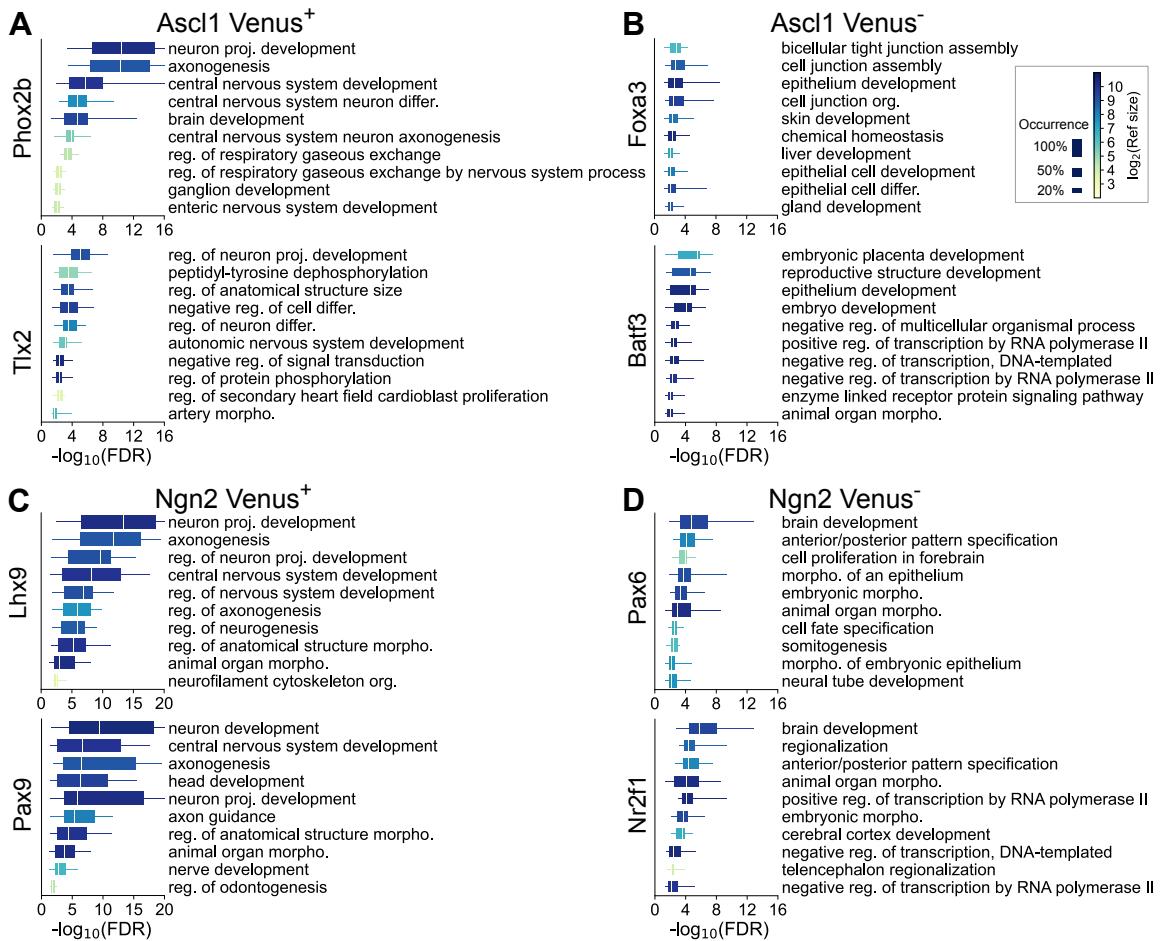
**Figure S5. GRNs of Ascl1 or Ngn2-mediated mESC to iN conversion of the entire time series, associated with Fig. 8.** (A) Schematic representation of mESC differentiation to neuron by induced expression of Ascl1 or Ngn2. The GRNs were constructed by gene expression at days 0, 2, 3, 4 and 6. (B-E) Left, representative GRN of the corresponding condition. Node size reflects betweenness centrality scores of the node. Node color reflects community membership of the gene. Nodes with red outlines represent TFs. Top TFs are labeled with gene names. Communities enriched with neuron-related terms are marked with dark colors, otherwise light colors. Right, bar plot showing the most enriched GO terms in the color-coded communities. differ., differentiation; morpho., morphogenesis; org., organization; proj., projection; reg., regulation. Results were from Ascl1-Venus<sup>+</sup> (B), Ascl1-Venus<sup>-</sup> (C), Ngn2-Venus<sup>+</sup> (D) and Ngn2-Venus<sup>-</sup> (E).



**Figure S6. Community identity and expression comparison of GRNs in mESC to neuron conversion, associated with Fig. 8 and Fig. S5.** Left, bar plot showing top 3 GO terms from corresponding communities in GRNs. Right, longitudinal community expression of corresponding communities in the left panel. The line represents the mean of locally estimated scatter plot smoothing (LOESS) regression on genes from corresponding communities for the entire time series. The shaded area represents 95% confidence interval from 200 LOESS fits. Results were from GRNs of later days in Ascl1-Venus<sup>+</sup> (A), Ascl1-Venus<sup>-</sup> (B), Ngn2-Venus<sup>+</sup> (C) and Ngn2-Venus<sup>-</sup> (D), same color scheme as in Fig. 8 (B-E). Results were from GRNs of the entire time series in corresponding conditions (A'-D'), same color scheme as in Fig. S5 (B-E).



**Figure S7. GO of top TFs from mESC to neuron conversion on later days (days 3, 4, 6).** (A-D) Top 10 replicable GO terms associated with gene communities containing the top TF identified in GRNs of Ascl1-Venus<sup>+</sup> (A), Ascl1-Venus<sup>-</sup> (B), Ngn2-Venus<sup>+</sup> (C) and Ngn2-Venus<sup>-</sup> (D). Boxes represent the median and interquartile range, and whiskers extend to the lowest and highest FDR ( $FDR < 0.05$ ). The count of observations of the GO term from all communities containing the indicated gene were used to scale the width of the boxes. GO annotation was performed on each community at a given network size but only on communities with at least 20 genes. differ., differentiation; morpho., morphogenesis; org., organization; proj., projection; reg., regulation.



**Figure S8. GO of top TFs from mESC to neuron conversion of the entire time series (days 0, 2, 3, 4, 6).** (A-D) Top 10 replicable GO terms associated with gene communities containing the top TF identified in GRNs of Ascl1-Venus<sup>+</sup> (A), Ascl1-Venus<sup>-</sup> (B), Ngn2-Venus<sup>+</sup> (C) and Ngn2-Venus<sup>-</sup> (D). Boxes represent the median and interquartile range, and whiskers extend to the lowest and highest FDR ( $FDR < 0.05$ ). The count of observations of the GO term from all communities containing the indicated gene were used to scale the width of the boxes. GO annotation was performed on each community at a given network size but only on communities with at least 20 genes. differ., differentiation; morpho., morphogenesis; org., organization; proj., projection; reg., regulation.