nature

SUPPLEMENTARY INFORMATION

https://doi.org/10.1038/s41586-018-0103-5

In the format provided by the authors and unedited.

Diverse reprogramming codes for neuronal identity

Rachel Tsunemoto[1,2,8], Sohyon Lee[1,8], Attila Szűcs[3,4], Pavel Chubukov[1], Irina Sokolova[5], Joel W. Blanchard[1], Kevin T. Eade[1], Jacob Bruggemann[6], Chunlei Wu[6], Ali Torkamani[7], Pietro Paolo Sanna[5] & Kristin K. Baldwin[1,2]*

[1]Department of Molecular and Cellular Neuroscience, Dorris Neuroscience Center, The Scripps Research Institute, La Jolla, CA, USA. [2]Neuroscience Graduate Program, University of California San Diego, La Jolla, CA, USA. [3]BioCircuits Institute, University of California San Diego, La Jolla, CA, USA. [4]MTA-ELTE-NAP B Neuronal Cell Biology Research Group, Eotvos Lorand University, Budapest, Hungary. [5]Molecular and Integrative Neurosciences Department, The Scripps Research Institute, La Jolla, CA, USA. [6]Integrative Structural and Computational Biology, The Scripps Research Institute, La Jolla, CA, USA. [7]Scripps Translational Science Institute, Scripps Health and The Scripps Research Institute, La Jolla, CA, USA. [8]These authors contributed equally: Rachel Tsunemoto, Sohyon Lee. *e-mail: kbaldwin@scripps.edu

Supplemental Table 1: List of bHLH, POU, and NR transcription factors tested with their standard gene names/aliases and the non-normalized matrix of pairwise reprogramming factors yielding Tuj1-positive cells. Numbers reflect the percentages of Tuj1-positive cells out of the total number of fibroblasts plated from two separate screen experiments quantified on day 14 post-induction.

Supplemental Table 2: Metadata of RNA-Seq samples. Table of experimental information related to cell sorting, RNA extraction, library preparation and sequencing of each iN and endogenous neuron populations selected for RNA-Seq analyses, with 2-3 biological replicates per population. dpi, days post-induction; RIN, RNA integrity number.

Supplemental Table 3: DESeq2, HOMER and IPA analyses. Table of DESeq2 output from comparing the (1) 35 TauEGFP positive, iN RNA-Seq populations (in duplicate biological samples) to one source MEF RNA-Seq population (in duplicate) and (2) MEFs (n=1 in duplicate) to the sequenced endogenous neuron (n=6 in duplicate, n=2 in triplicate) and whole brain populations (n=1 in duplicate, EndoNs/Brain). Significant differentially expressed genes were those with p-adjusted < 0.05. Enriched genes in iN and EndoNs/Brain samples are those with negative log2 fold changes. Significantly enriched genes and transcription factors (GO:0003700) in EndoNs/Brain populations compared to MEFs that are also significantly enriched in the iN populations when compared to MEFs are listed as shared enriched genes and transcription factors (Sheet 1).

The results of the gene ontology (GO Term) enrichment analysis (PANTHER) and KEGG Pathway enrichment analysis (DAVID) using the shared enriched genes as the input genes (n=2,239 genes from the analysis in Sheet 1), as well as, the results of the GO Term enrichment analysis (DAVID) using the upregulated iN genes (n=3,860, top 10 results) and downregulated iN genes (n=3,467 genes, top 12 results) from the analysis in Sheet 1 as the input genes (Sheet 2).

Results of known motif enrichment in HOMER with the shared enriched genes (n=2,239 genes from analysis in Sheet 1) serving as target genes are also provided. Shared downregulated genes are those enriched in the MEF populations in both DESeq2 comparisons to iNs and EndoNs/Brain populations. Results of upstream transcriptional regulators identified by Ingenuity Pathway Analysis (IPA) with the shared enriched genes serving as target genes are also provided (Sheet 3).

Table of genes not shared between the enriched genes of the EndoN and iN populations when compared to MEFs (using DESeq2 from analysis in Sheet 1) and their corresponding gene ontology enrichment analysis (PANTHER). “All EndoNs” include the EndoN populations (n=6 in duplicate, n=2 in triplicate) and whole brain population (n=1 in duplicate) sequenced for this study. “Defined EndoNs” include a subset of the sequenced EndoN populations (n=3 in duplicate, n=1 in triplicate) that were isolated using previously published, subtype-specific reporter lines (Sheet 4).

Supplemental Table 4: Enriched genes and WGCNA. Table of genes that are enriched across iN populations. This includes a list of genes that were identified as significantly enriched (p-adjusted value < 0.05) in each iN population versus all other iN populations using DESeq2 (Enriched Genes). Of those enriched genes, we filtered out genes whose expression levels were not significantly different from MEFs. Also listed

are the module memberships of genes imputed into WGCNA, which included genes that had greater than 200 raw counts in duplicate samples of a least one iN or MEF population. All merged modules are categorized by module eigengene (ME) expression patterns and include the number of genes in each module, representative hub genes, representative GO Terms, and ME expression bar plots. A1, *Ascl1*; A2, *Ascl2*; A5, *Ascl5*; N1, *Ngn1*; N2, *Ngn2*; N3, *Ngn3*; ND2, *NeuroD2*; B2, *Brn2*; B4, *Brn4*; B3a, *Brn3a*; B3b, *Brn3b*; B3c, *Brn3c*; O4, *Oct4*; P1, *Pit1*.