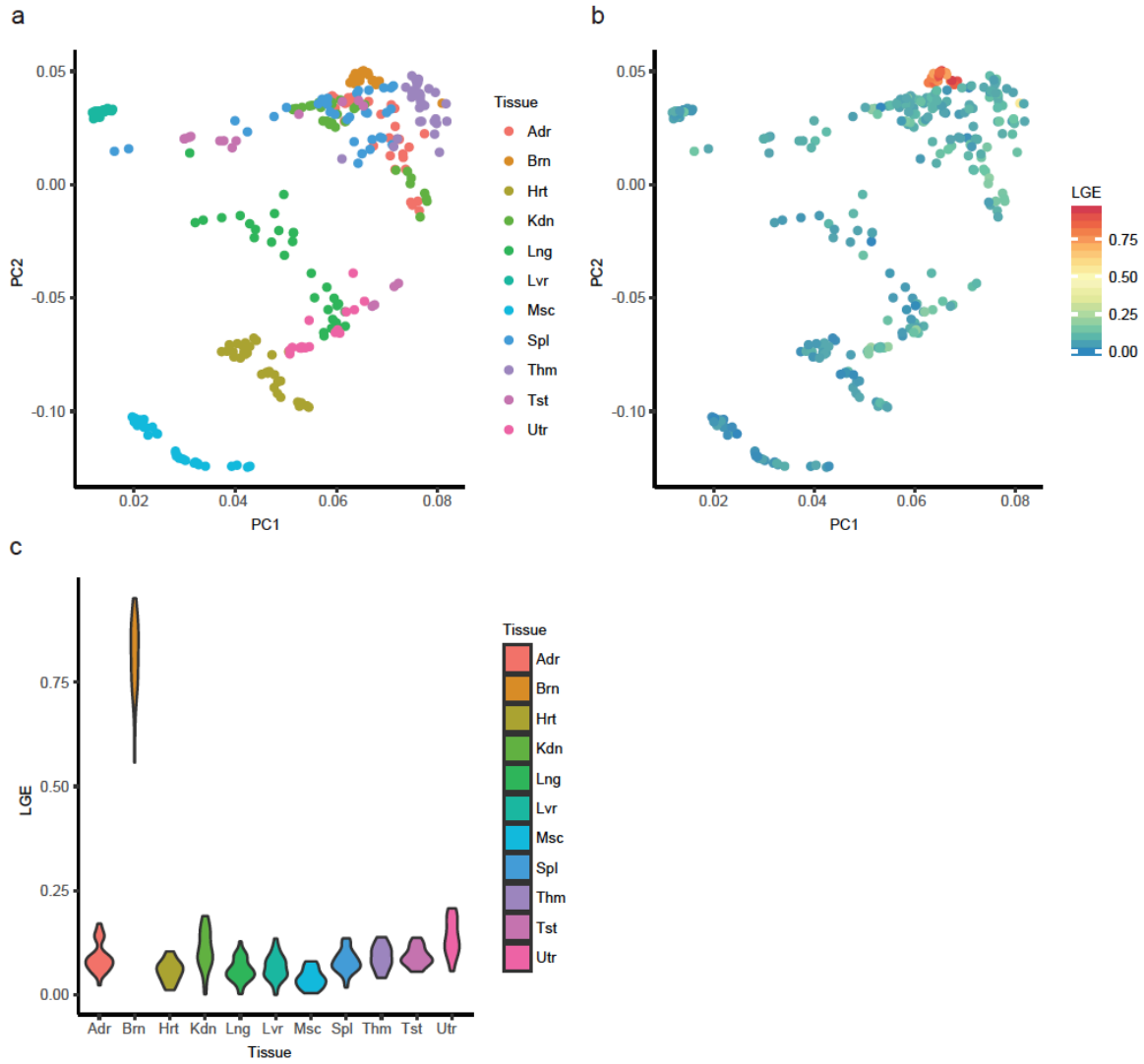
Supplemental Materials

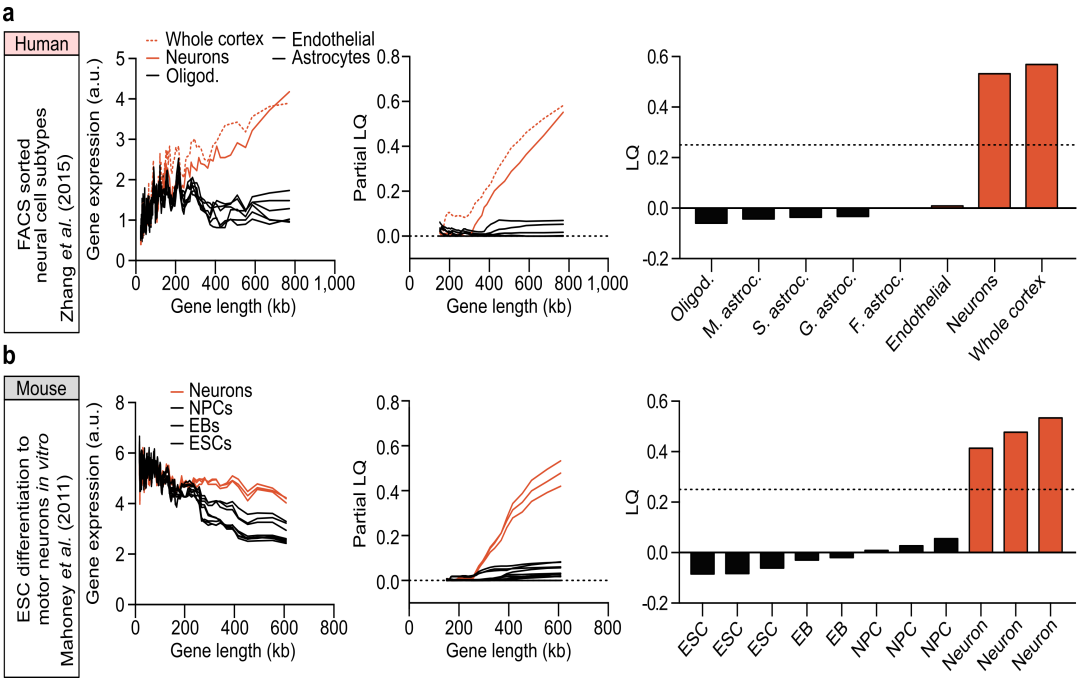
Accession codes

- BioSamples: ND34769 HAF1 (SAMN05340933), ND34769 MSN1 (SAMN05340929), AG04148 HAF2 (SAMN05340931), AG04148 MSN2 (SAMN05340927).
- GEO database (Edgar, et al., 2002): Mazzoni *et al.*, 2011 (GSE19372), Wang *et al.*, 2012 (GSE62936), Treutlein *et al.*, 2016 (GSE67310), Hu *et al.*, 2015 (GSE69480), Zhang *et al.*, 2016 (GSE73721), Xue *et al.*, 2016 (GSE77523), Colasante *et al.*, 2015 (GSE74065).
- Online resources: Tasic *et al.*, 2016 (http://casestudies.brain-map.org/celltax/download_archive), Miller *et al.*, 2014 (http://www.brainspan.org/api/v2/well_known_file_download/278444094), Habib *et al.*, 2016 (http://science.sciencemag.org/highwire/filestream/682212/field_highwire_adjunct_files/5/Table_S5.xlsx)

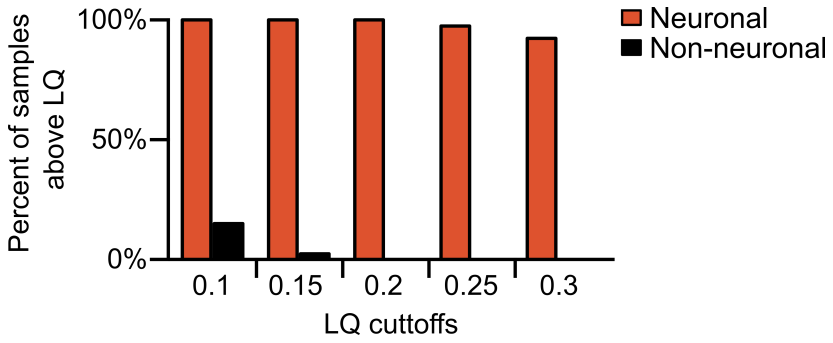
Supplemental Figures



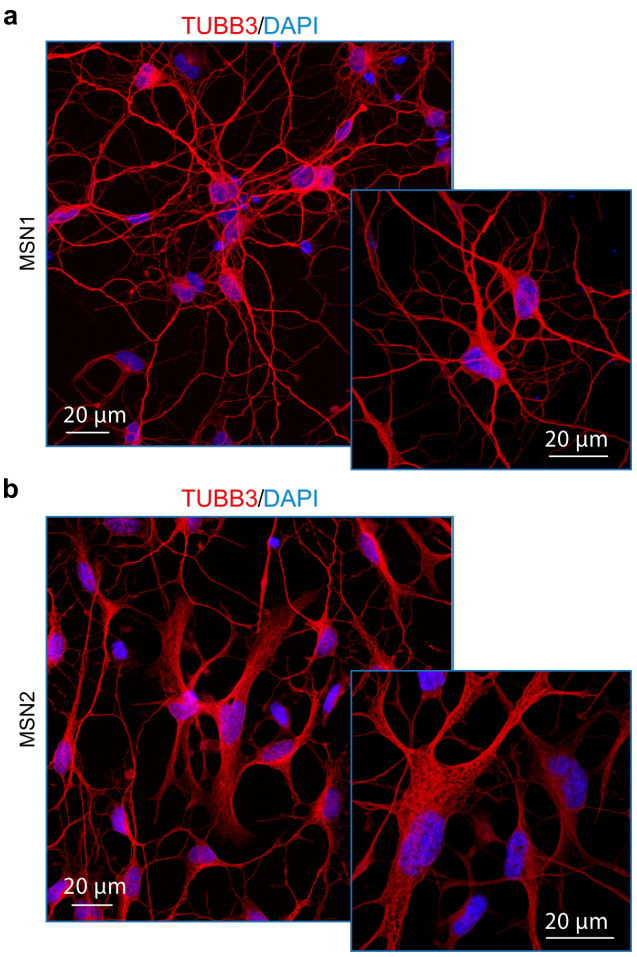
Supplemental Figure 1. LGE identifies neuronal tissues, while PCA only separates tissues. (a) Principle components 1 and 2 of Rat BodyMap principle component analysis (PCA), with each sample colored by tissue of origin. (b) The same PCA as in (a), but now each sample is colored by the scaled correlation of binned gene expression (200 genes per bin; 40 genes per step) versus median gene length, here termed LGE. (c) Violin plot of the LGE values plotted in (b) for each Tissue.



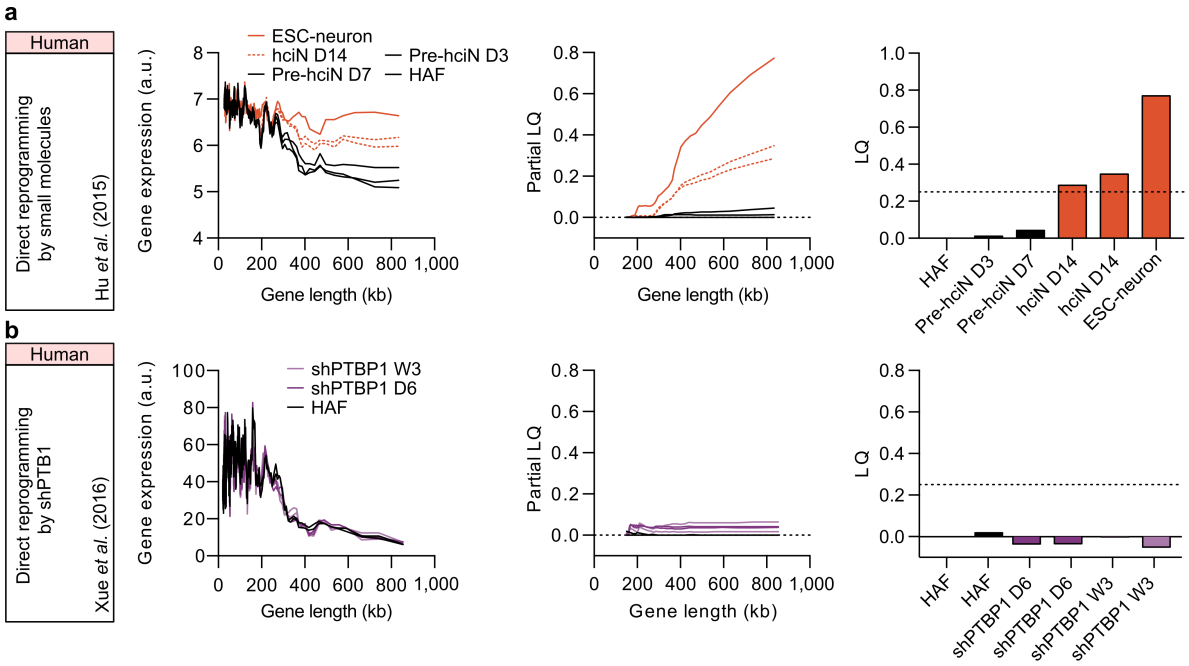
Supplemental Figure 2. LONGO output of neuronal differentiation in vivo and in vitro. From left to right, reference for input data, rolling media of gene expression versus length (200 gene bins, 40 gene step), partial LQ, and final LQ. (a) FACS sorted human neural cell subtypes (Zhang, et al., 2016): glial cells (oligodendrocytes, endothelial cells, S. astrocytes (sclerotic hippocampi astrocytes), G. astrocytes (GBM peri tumor astrocytes), F. astrocytes (fetal astrocytes), and M. astrocytes (mature astrocytes) shown in black; neurons and whole cortex (dotted-line) shown in red. (b) Mouse ESC differentiation to motor neurons in vitro (Mahony, et al., 2011): ESCs (D0), embryoid bodies (EBs; D2), and NPCs (D4) shown in black; motor neurons (D7) shown in red. In all analyzed data, LQ typically exceeds 0.25 in neurons.



Supplemental Figure 3. LQ cutoff of 0.25 stringently distinguishes neurons from non-neuronal cells. Percentage of samples that pass various LQ cutoffs from LQ data in Figures 2b,2c, and 2e for neuronal cells (red bar; n = 39) and non-neuronal cells (black bar; n = 40).



Supplemental Figure 4. Medium spiny neurons reprogrammed with miR-9/9*-124+CDM from fibroblasts of two unrelated adults shows variation in the proper acquisition of neuronal morphological features. (a) MSN1: Female, 68 years old at sampling - Coriell: ND34769 and (b) MSN2: Male, 56 years old - Coriell: AG04148, im-munostained for TUBB3 in red and the nuclear stain DAPI in blue, at 28 days post microRNA-based neuronal induction. In contrast to cells from HAF1 (MSN1), a large fraction of reprogrammed cells from HAF2 (MSN2) retains fibroblastic fea-tures such as expanded cell bodies and a large nucleus, and have limited dendritic trees. Insets for each cell population display morphological characteristics in greater magnification. Our results demonstrate that TUBB3 immunostaining, a commonly used ‘neuronal marker’, does not reliably distinguish neurons from partially repro-grammed cells.



Supplemental Figure 5. LONGO output of small molecule and shPTBP1-induced reprogramming. From left to right, reference for input data, rolling median of gene expression versus length (200 gene bins, 40 gene step), partial LQ, and final LQ. (a) Direct reprogramming of HAFs to neurons (Pre-hciN) by small molecules (Hu, et al., 2015): HAFs, Pre-hciN D3, and Pre-hciN D7 (dotted-line) shown in black; hciN D14 and ESC-derived neurons (dotted-line) shown in red. (b) Direct reprogramming of HAFs to neuron-like cells by shPTBP1 (Xue, et al., 2016): HAFs (black); shPTBP1 for 6 days (dark purple); shPTBP1 for 3 weeks (light purple).