Supporting Information

Tiwari et al. 10.1073/pnas.1119798109

SI Materials and Methods

Affymetrix Microarray Data Analysis. All 28 microarrays were RMA-normalized using R/Bioconductor (http://www.R-project.org/) (1) and the oligo package version 1.14.0 (2). Probe sets were linked to Entrez Gene IDs using Affymetrix annotation (MoGene-1_0-st-v1.na27.mm9.transcript.csv, downloaded from www.affymetrix.com). Only the probe sets that mapped to a single gene were used for the analysis. When several probe sets mapped to the same gene, the expression levels of these probe sets were averaged. The limma package was used to determine the significance of changes in gene expression in different stages. For each contrast, the set of significantly changing genes was defined as the set of genes with absolute fold-changes of at least 2 and adjusted P value $\leq 10\text{^}(-5)$.

Genomic Coordinates. The July 2007 *Mus musculus* genome assembly (NCBI37/mm9) provided by the National Center for Biotechnology Information (NCBI) (http://www.ncbi.nlm.nih.gov/genome/guide/mouse/) and the Mouse Genome Sequencing Consortium (http://www.sanger.ac.uk/Projects/M_musculus/) was used as a basis for all analyses. Annotation of known RefSeq transcripts was obtained from the University of California, Santa Cruz genome browser (http://hgdownload.cse.ucsc.edu/goldenPath/mm9/database/refGene.txt.gz) on January 10, 2011). A nonredundant, nonoverlapping set of promoters (n = 18,667) was generated starting with 2,000-bp windows centered around RefSeq transcription start sites (TSS) (n = 21,911) and removing all overlapping windows. For Fig. 3*B* in the main text, the entire chromosome 19 was segmented hierarchically into four different types of regions as follows: Promoters were defined as windows of 2,000 bp centered on known RefSeq TSS. Exons were defined as all nonpromoter bases that overlapped exons of RefSeq transcripts, and introns were defined as all nonpromoter/nonexon bases that were flanked by two exons of a single transcript. All remaining bases were assigned to the intergenic region.

ChIP-Chip Data Analysis. NimbleGen array intensity files were read, and log2 enrichments (log2 bound/input ratios) for each individual probe were calculated using the R package Ringo (3). To remove dye artifacts (which were clearly apparent from the differences in log2 bound/input ratios in the low-intensity range between the first and the second replicate, which had been performed with a dye swap), both arrays were loess-normalized using the "normalizeWithinArrays" function from the limma package (4). Probe-level log2 enrichments showed good reproducibility between the two replicates (Pearson's correlation, 0.71). To calculate log2 enrichments for different genomic regions (promoters, exons, introns, and intergenic, as defined above), log2 enrichments of all probes that mapped to a particular region were averaged. Only regions to which at least three probes mapped were used for further analysis. The log2 promoter enrichments showed very good reproducibility between the two replicates (Pearson's correlation, 0.9). For Fig. 3*B* in the main text, only the probes on chromosome 19 were used (because chromosome 19 is fully tiled on the array), and average log2 enrichments per region were determined from the average log2 enrichment of the two replicates. Wiggle tracks were created using the R package rtracklayer (5). For these tracks, probe enrichments were smoothed by replacing the enrichment of each probe by the average enrichment of all of the probes in a 500-nt window centered on the probe. This smoothing is justified because enrichments of single probes are quite noisy, and the resolution of the array readout ultimately is limited by the length of the sonicated fragments, which is roughly 500 nt.

Sequencing Data Analysis. For each sequencing sample, low-complexity reads were filtered out based on their dinucleotide entropy (only a very small percentage of the reads were removed). Alignments to the mouse genome were performed by the software Bowtie (version 0.9.9.1) (6) with parameters -v 2 -a -m 100, tracking up to 100 best-alignment positions per query and allowing no more than two mismatches. To track genomically untemplated hits (e.g., exon–exon junctions or missing parts in the current assembly), the reads also were mapped to an annotation database containing known mouse sequences [miRNA from ftp://ftp.sanger.ac.uk/pub/mirbase/sequences/13.0, rRNA, snRNA, snoRNA and RefSeq mRNA from GenBank http://www.ncbi.nlm.nih.gov/sites/entrez, downloaded July 16, 2009; tRNA from http://lowelab.ucsc.edu/GtRNAdb/; and piRNA from NCBI (accession nos. DQ539889–DQ569912)]. For alignments to the annotation database, all best hits with no more than two mismatches were tracked. Each alignment was weighted by the inverse of the number of hits. When a read had more hits from the annotation database to an individual sequence than to the genome, the number of hits to the individual sequence was selected to ensure that the total weight of a read did not exceed one. All quantifications were based on weighted alignments.

For the chip-seq samples, promoter levels were defined as log2 (n + 1) where n is the weighted sum of all read alignments to a particular promoter. To determine expression levels from the RNA-seq data, expression levels of RefSeq transcripts were calculated as x = log2(n/len *avg_len + 1), where n is the weighted sum of read alignments to a RefSeq transcript, *len* is the length of the transcript, and *avg_len* is the mean length over all RefSeq transcripts. The wiggle track for K4me2 was created by calculating the average read coverage per base in windows with a length of 100 bp.

Promoter Clustering. To cluster promoters into different subclasses, K4me2, Pol II, and K27me3 levels and top2b enrichment were determined for each promoter as described above. For PolII and K27me3, the different replicates were combined by first normalizing the total read counts of each replicate to the total read count of the smallest library and then averaging the values of the different replicates for each promoter. To allow comparison of ChIPseq levels and Top2β array enrichments, each measure was scaled to have a mean of 0 and SD of 1 across all promoters. This scaling resulted in a matrix of scaled levels/enrichments in which each row represented a promoter and each column represented one of the measures. This matrix then was clustered using k-means clustering using the R function "kmeans" with 100 random starting points and a maximum of 100 iterations. We compared the results of the k-means algorithm with varying numbers of clusters and decided to use five clusters in an attempt to minimize the number of different clusters while retaining the clearest groupings that reoccurred with different total numbers of clusters. The resulting clusters then were sorted by the mean top2b enrichment of each cluster.

1. Gentleman RC, et al. (2004) Bioconductor: Open software development for computational biology and bioinformatics. *Genome Biol* 5:R80.

2. Carvalho BS, Irizarry RA (2010) A framework for oligonucleotide microarray preprocessing. *Bioinformatics* 26:2363–2367.

3. Toedling J, et al. (2007) Ringo—an R/Bioconductor package for analyzing ChIP-chip readouts. *BMC Bioinformatics* 8:221.
4. Smyth GK (2005) Limma: linear models for microarray data. *Bioinformatics and Computational Biology Solutions using R and Bioconductor*, eds Gentleman R, Carey V, Dudoit S, Irizarry R, Huber W (Springer, New York), pp 397–420.
5. Lawrence M, Gentleman R, Carey V (2009) rtracklayer: An R package for interfacing with genome browsers. *Bioinformatics* 25:1841–1842.
6. Langmead B, Trapnell C, Pop M, Salzberg SL (2009) Ultrafast and memory-efficient alignment of short DNA sequences to the human genome. *Genome Biol* 10:R25.

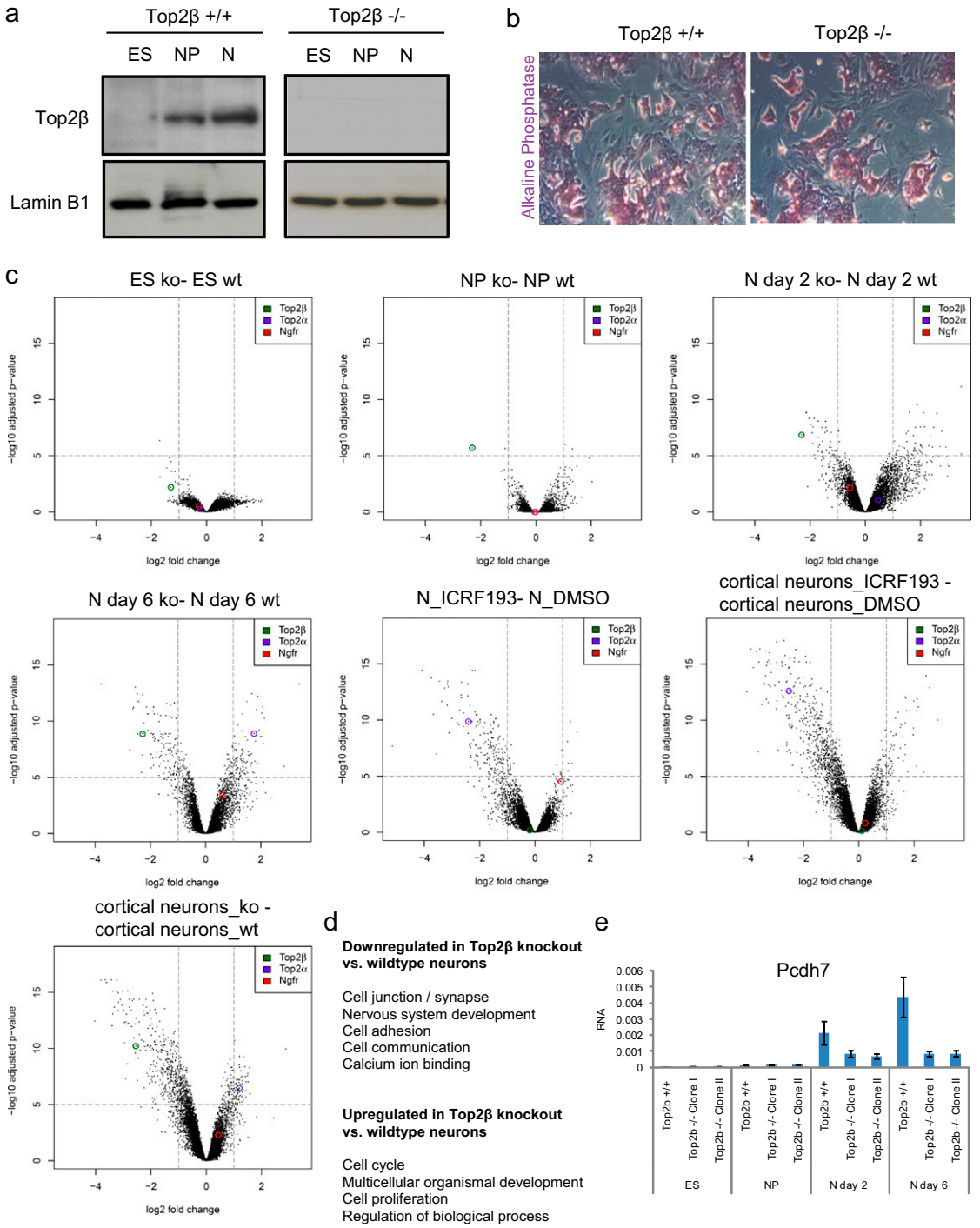

Fig. S1. Impaired neuronal differentiation in ES cells lacking Topoisomerase IIβ (Top2β). (*A*) Top2β detection by Western blot in protein extracts isolated from wild-type ($^{+/+}$) and Top2β-knockout ($^{-/-}$) ES cells, neuronal progenitor (NP) cells, and neurons (N) illustrates lack of Top2β protein in knockout cells. Lamin B1 serves as a loading control. (*B*) Alkaline phosphatase staining (as an indicator of pluripotency) reveals no differences between wild-type and Top2β-knockout ES cells. (*C*) Volcano plot depicting expression changes in wild-type and Top2β-knockout ES cells (*Top Left*), neuron progenitor cells (*Top Center*), and neurons at day 2 (N day 2) (*Top Right*) and day 6 (N day 6) (*Middle Left*); in DMSO- and meso-2,3-bis(2,6-dioxopiperazin-4-yl)butane (ICRF)-treated stem cell-derived neurons (*Middle Center*); in DMSO- and ICRF-treated cortical neurons (*Middle Right*); and in embryonic day 16.5 (E16.5) wild-type and Top2β-knockout cortical neurons (*Bottom*). The *x* axis shows fold changes in expression (log2 scale); the *y* axis shows the associated *P* value. Topoisomerase IIα (Top2α), Top2β, and *Ngfr*

Legend continued on following page

are marked in blue, green, and red, respectively. Dotted lines indicate the cut-offs used on the fold change in expression and the associated *P* value. (*D*) The genes that were significantly differentially affected wild-type and Top2β-knockout neurons were queried for top GO Processes using GOstat analysis; the top hits in each of these categories are shown together with the significance (*P* value $< 10^{-20}$). Interestingly, the down-regulated genes predominately contain neurogenesis-associated genes; the up-regulated genes largely comprise genes involved in cell proliferation. (*E*) Expression analysis for the neuron-specific gene *Pcdh7* at various stages of neuronal differentiation comparing wild-type and two different Top2β-knockout clones (I and II). The results indicate that Pcdh7 is not induced in Top2β-knockout neurons, unlike wild-type cells. mRNA levels, as measured by real-time RT-PCR relative to Gapdh, are plotted on the *y* axis. Error bars indicate SEM.

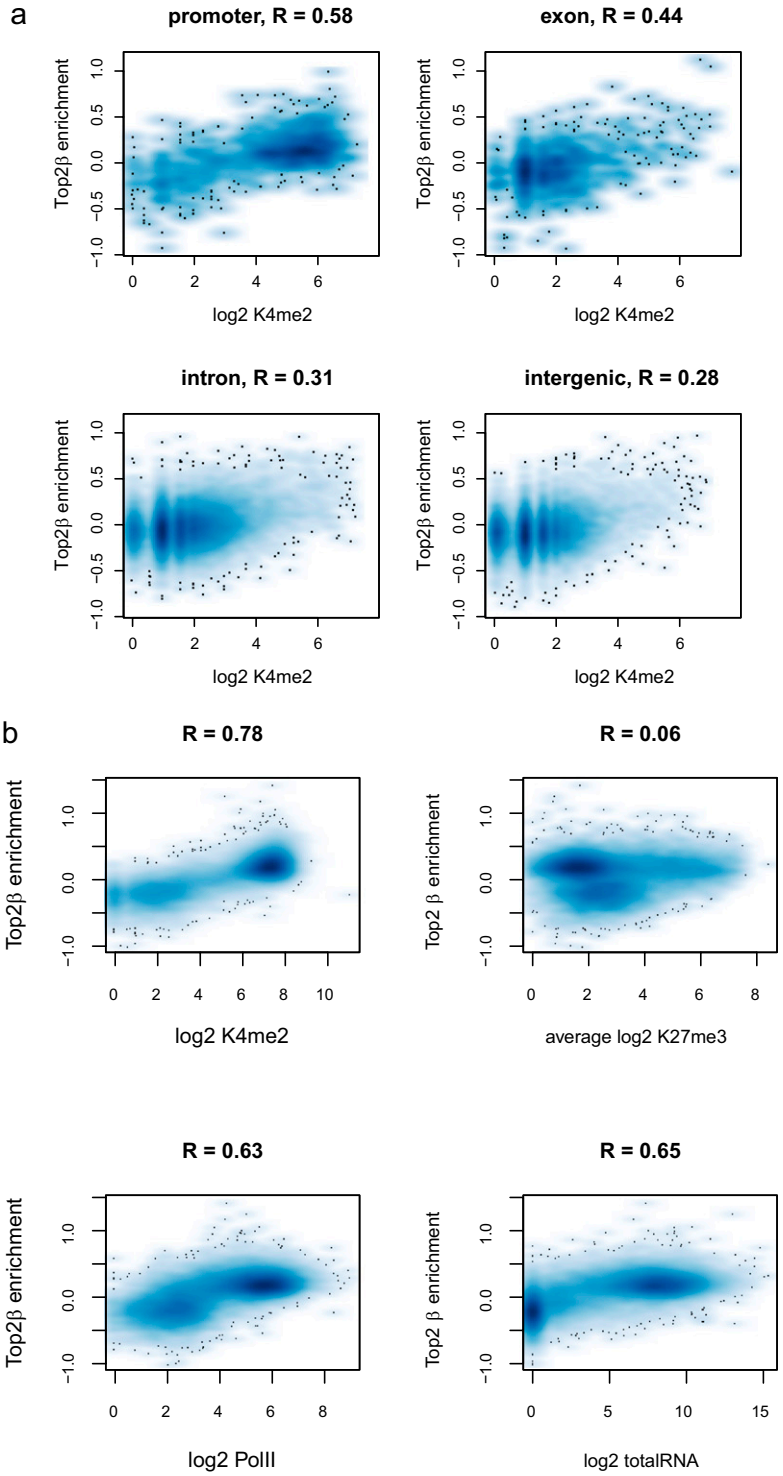

Fig. S2. Top2β enrichments correlate globally with H3K4me2 levels. (*A*) The entire chromosome 19 was divided into consecutive windows of 1,000 bp and annotated as promoter, exon, intron, or intergenic regions. Top2β enrichment and H3K4me2 levels are shown for each region. Top2β globally binds regions

Legend continued on following page

that are marked by H3K4me2. The greatest enrichment for regions that are both Top2β-bound and enriched in H3K4me2 is seen in windows that overlap promoters. (*B*) Scatter plots show correlations at TSS in neurons of Top2β enrichment (*y* axis) with H3K4me2 (*Upper Left*), H3K27me3 (*Upper Right*), RNA Pol II (*Lower Left*), and mRNA levels (*Lower Right*). The data suggest that the promoters targeted by Top2β are in an active chromatin state.

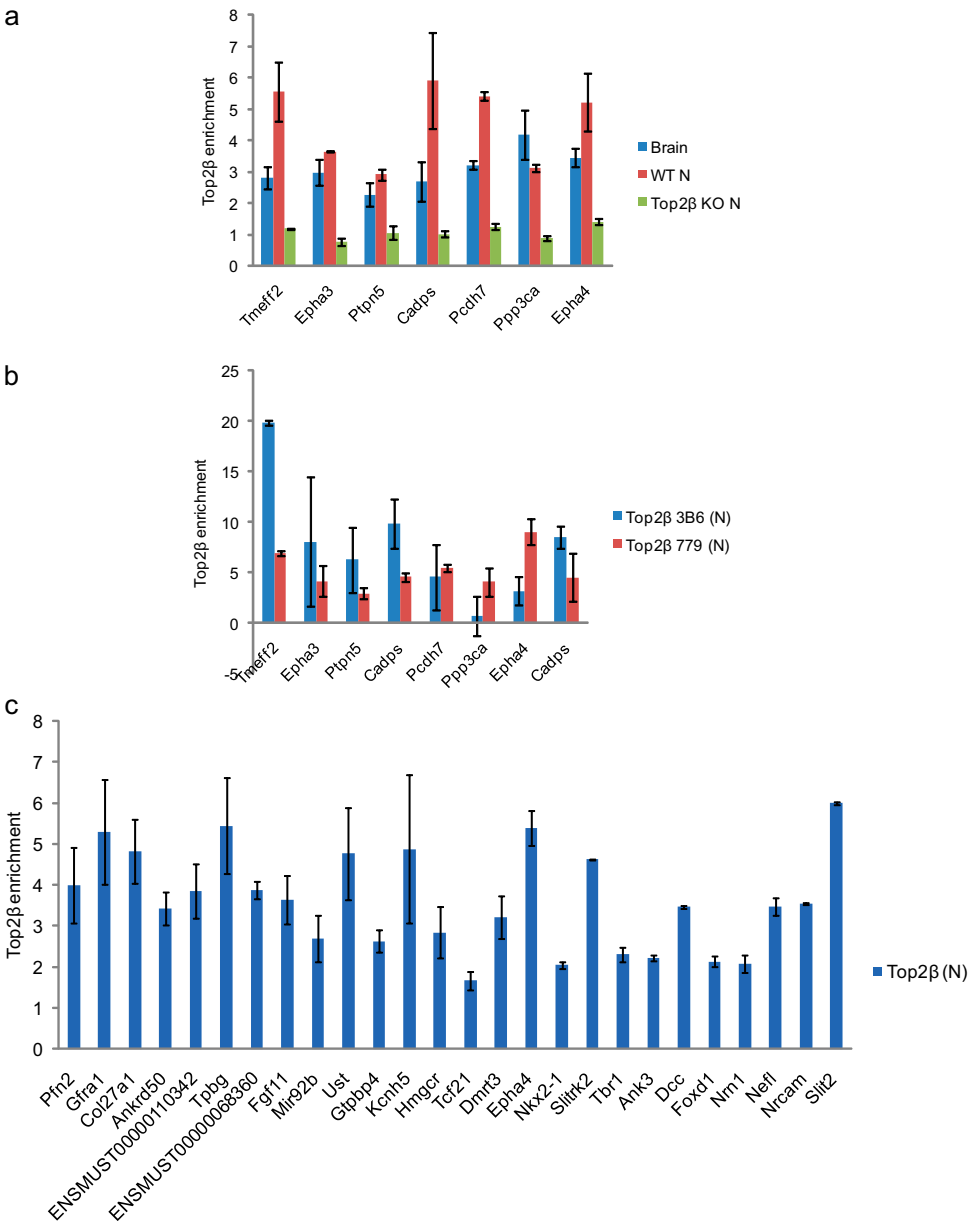

Fig. S3. ChIP-quantitative PCR (qPCR) validation of Top2β targets. (*A*) ChIP-qPCR validation of Top2β enrichment at various identified target promoters using Top2β ChIP samples derived from wild-type and Top2β-knockout neurons (N) and from adult brain. Average enrichments from three separate assays are plotted on the *y* axis as the ratio of precipitated DNA (bound) to the total input DNA and further normalized to a control region. Error bars indicate SEM. (*B*) ChIP-qPCR validation of Top2β enrichment at various identified target promoters using neuronal ChIP samples derived with two different Top2β antibodies (3B6 and 779). The graph is plotted as in *A*. (*C*) ChIP-qPCR validation of Top2β enrichment at several more identified target promoters using ChIP samples derived from neurons. The graph is plotted as in *A*.

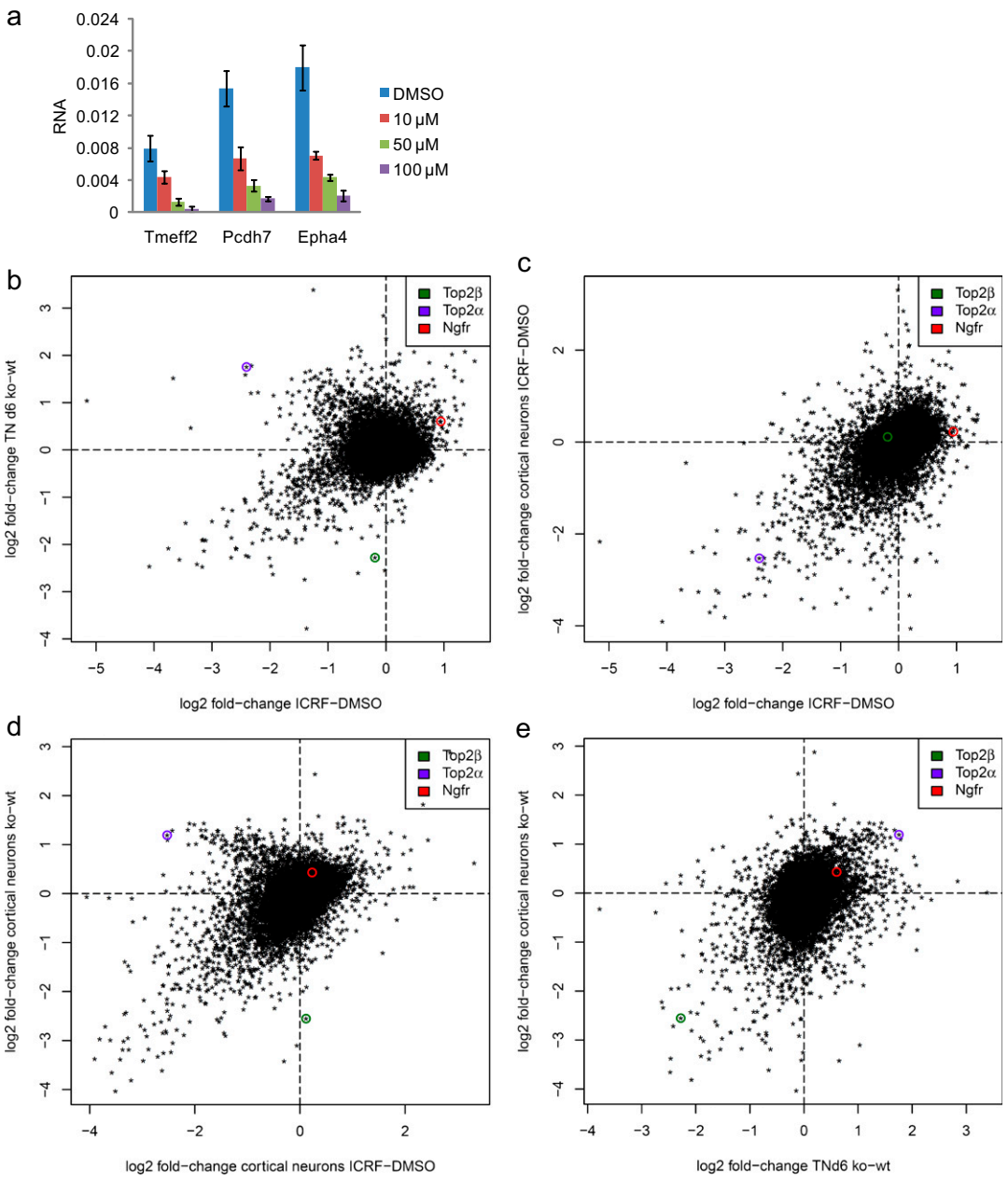

Fig. S4. Lack of the Top2β enzyme or of its activity down-regulates a similar set of genes. (*A*) Down-regulation of Top2β targets by ICRF in a dose-dependent manner. Treatment of cells with ICRF-193 effectively down-regulates Top2β targets in a dose-dependent manner. Shown are real-time RT-PCR analyses of the expression of three neuron-specific Top2β-target genes, *Tmeff2*, *Pcdh7*, and *Epha4*, upon treatment with various dosages of ICRF-193 (10, 50, and 100 μM) compared with control (DMSO). These analyses suggest that ICRF-193 is effective in inhibiting Top2β in postmitotic neurons and that Top2β activity is required for the target-gene expression. Relative mRNA levels detected by RT-PCR were derived after normalization with Gapdh. The average data from three separate assays are plotted on the *y* axis. Error bars indicate SEM. (*B*) Scatter plot comparing expression changes in ICRF-treated neurons (*x* axis) and in stem cell-derived Top2β-knockout neurons (*y* axis). (*C*) Scatter plot comparing expression changes in ICRF-treated cortical neurons (*y* axis) and in ICRF-treated stem cell-derived neurons (*x* axis). (*D*) Scatter plot comparing expression changes in Top2β-knockout cortical neurons (*y* axis) and in ICRF-treated cortical neurons (*x* axis). (*E*) Scatter plot comparing expression changes in Top2β-knockout cortical neurons (*y* axis) and in Top2β-knockout stem cell-derived neurons (*x* axis).

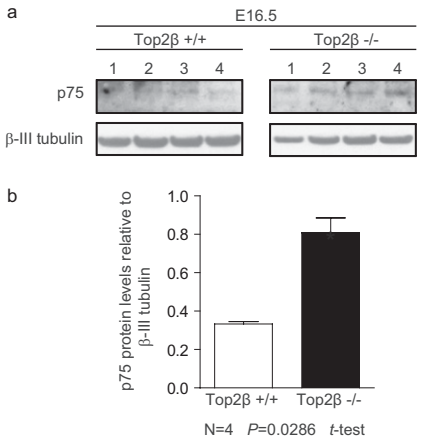

Fig. S5. p75 levels are up-regulated in cortical neurons derived from Top2β-knockout embryos at E16.5. (*A*) p75 protein levels as detected by Western blotting of extracts isolated from cortical neurons of four wild-type and four Top2β-knockout mouse embryos at E16.5 showing the up-regulation of p75 in Top2β-deficient cells. β-III Tubulin served as a loading control. (*B*) Quantification of p75 levels relative to β-III tubulin in the four samples analyzed in *A*. The *t* test suggests a significant up-regulation in p75 levels in Top2β-knockout cortical neurons.

Dataset S1. Genes that are bound by Top2β in wild-type neurons and are misregulated in Top2β-knockout neurons at day 6 in vitro

Dataset S1 (XLSX)