Supporting Information

Wang et al. 10.1073/pnas.1412569111

SI Text

Comment on ChIP Combined with Deep Sequencing Experiments

The genome-wide Cockayne syndrome group B (CSB) and RNA polymerase II (RNAP II) distribution was investigated by ChIP combined with deep sequencing (ChIP-Seq), using the highly specific GFP antibody to detect GFP-tagged CSB and N20 antibody to detect RNAP II (a list of antibodies used is provided in Dataset S11). CSB and RNAP II peaks across the genome were called using Model-Based Analysis of Chip-Seq (MACS) (1) (further details are provided in *SI Methods*). As mentioned in the main text, peaks of CSB density overlapped significantly with genes that were deregulated in CSB-deficient cells. In total, only 975 CSB ChIP-Seq peaks were identified in both experimental replicates [1,1430 CSB peaks in replicate 1 and 1,271 peaks in replicate 2 were not detected in any of the control CS1AN (a simian virus [sv] 40-transformed patient cell line) replicates (which do not express GFP-CSB)]. No less than 722 of those 975 high-confidence peaks were in the coding region of a gene (in 703 distinct genes, P value of 2.2×10^{-16}). Because protein-coding genes are rare in the genome, this high proportion strongly connects CSB with RNAP II transcription. Of those 703 genes, 60 were also among the 1,244 genes that change expression upon CSB mutation. This overlap between CSB occupancy inside a gene and its effect on the gene's expression is much larger than would be expected by chance (P value of 1.04×10^{-4}). As alluded to in the main text, we also identified genes that were differentially regulated in CSB-deficient cells and also had altered RNAP II density at their transcriptional start site (by ChIP-Seq). Of the 651 genes whose expression was decreased in the absence of CSB, 205 (31.5%) also displayed a decrease in RNAP II density. Again, this overlap is highly significant (P value for overlap of 1.2×10^{-71}).

In the analysis above, we introduced the restriction that a peak had to be called in both experimental repeats of CSB-WT cells and not in either of the two CS1AN (mutant) replicates. If, instead, we only required that a peak should be called in one of the two replicates, but not in the CS1AN control, the overlap was more substantial. For example, there were 9,137 CSB peaks (in 7,165 genes) within the body of genes called in replicate 1, 439 of which were found at one of the 1,242 genes that were dysregulated in CSB mutant cells. This overlap is more impressive in a numerical sense, but the P value for this overlap remains in the same range as that of the more stringent analysis (P value of 5.30×10^{-5}).

It is also important to note that CSB obviously does not need to be present in the coding region of a gene to regulate it. As a matter of fact, most regulation would be expected to occur from the promoter or enhancer(s) of a gene. Unfortunately, present knowledge of which enhancers regulate which genes is limited, and promoter length varies dramatically from gene to gene. We note that while our paper was being prepared for publication, Fan and coworkers (2) published a ChIP-Seq dataset for CSB. These authors reported that CSB occupancy sites are overrepresented at promoters and enhancers compared with the overall genomic distribution (2). Because it is not important for the message of our paper, we have not compared our ChIP-Seq or gene expression datasets with the ChIP-Seq dataset of Fan and coworkers (2).

SI Methods

Cell Lines and Lentivirus Supernatants. The BAC RP11-111H18, containing human CSB, was obtained from Invitrogen. The FLAP (EGFP-IRES-Neo)-tagging cassette (3) (T. Hyman laboratory, Max Planck Institute of Molecular Cell Biology and Genetics, Dresden, Germany) was PCR-amplified using primers carrying 50-nt homology arms to the C terminus of CSB. BAC recombineering and stable transfection into CS1AN of the modified BAC was performed as described (3).

To construct a tetracycline/doxycycline (TET)-responsive plasmid for CSB expression in CS1AN cells, CSB cDNA was cloned into the pcDNA4-TetON-GFP vector (Clontech), downstream of the minimal CMV promoter. Clones expressing high GFP in the presence of doxycycline, and with low leaky expression in its absence, were selected for the establishment of stable cell lines. All cell lines were cultured in DMEM containing 10% (vol/vol) FBS [Tet system-approved FBS for inducible cell lines (Clontech) at 37 °C with 5% CO_2]. Cockayne syndrome (CS) primary fibroblasts (CS19BR, CS21BR, CS2SH, and CS2LI) were immortalized by the transduction of SV40 large T-antigen.

HEK 293T cells were used to generate lentivirus supernatants. Cells were cotransfected with a Trans Lentiviral Packaging Kit (Thermo Scientific) and control-specific or CSB-specific shRNAs, according to the manufacturer's instructions. Lentiviral shRNAs against human CSB [TRCN0000016776, named short hairpin CSB-16776 (shCSB-16776); TRCN0000016777, named shCSB-16777], mouse CSB (TRCN0000090508, named Csb-sh1; TRCN0000173411, named Csb-sh2; TRCN0000173450, named Csb-sh3; TRCN0000194563, named Csb-sh4), and control shRNA (TRC lentiviral pLKO.1 GFP shRNA control, named shGFP) were all purchased from Thermo Scientific. Lentiviral supernatant was collected twice at 48 h and 72 h after transfection, and filtered through a 0.45-μm low-protein–binding filter. For generating stable neuroblastoma cell lines in which endogenous CSB was depleted, human SH-SY5Y cells and mouse Neuro2a cells were transduced with lentiviral supernatant in the presence of 8 μg/mL polybrene, followed by selection with 1 μg/mL puromycin. Knockdown efficiency was assessed by Western blot analysis.

ChIP-Seq. ChIP-Seq was performed essentially as described (4). Briefly, cells were harvested by trypsin treatment and cross-linked for 15 min at room temperature by adding formaldehyde to a final concentration of 1%. The cross-linking reaction was quenched with glycine (125 mM final concentration) for 5 min, followed by two washes with ice-cold 1× PBS. Cells were resuspended in ChIP lysis buffer [5 mM Hepes (pH 8.0), 85 mM KCl, 0.5% Nonidet P-40, and protease inhibitor mixtures] and incubated on ice for 5 min. Nuclei were pelleted by centrifugation at 3,900 × g for 5 min at 4 °C. Nuclei were then resuspended and sonicated in ChIP nuclear lysis buffer [50 mM Tris·HCl (pH 8.1), 10 mM EDTA (pH 8.0), 1% SDS, and protease inhibitor mixtures] for 10 cycles at 30 s each in an ice-water bath with 30 s on ice between cycles. Sonicated chromatin was cleared and diluted 1:5 with ChIP dilution buffer [0.01% SDS, 1.1% (vol/vol) Triton X-100, 1.2 mM EDTA (pH 8.0), 16.7 mM Tris·HCl (pH 8.1), 167 mM NaCl, and protease inhibitor mixtures]. Sonicated chromatin was incubated overnight at 4 °C with preblocked magnetic beads bound with antibodies. Beads were washed twice with ChIP low-salt buffer [0.1% SDS, 1% Triton X-100, 2 mM EDTA, 20 mM Tris·HCl (pH 8.1), and 150 mM NaCl], twice with ChIP high-salt buffer [0.1% SDS, 1% Triton X-100, 2 mM EDTA, 20 mM Tris·HCl (pH 8.1), and 500 mM NaCl], one time with ChIP LiCl buffer [10 mM Tris·HCl (pH 8.1), 250 mM LiCl, 1% Nonidet P-40, 1% deoxycholic acid, and 1 mM EDTA], and

one time with TE buffer [10 mM Tris·HCl (pH 8.0) and 1 mM EDTA]. DNA was eluted at 65 °C for 15 min in ChIP elution buffer [50 mM Tris·HCl (pH 8.0), 10 mM EDTA, and 1% SDS]. Cross-links were reversed overnight at 65 °C in the presence of 1 μg/mL RNase A. Protein was digested using Proteinase K (100 μg), and DNA was purified with phenol/chloroform extraction and ethanol precipitation. Purified immunoprecipitated DNA was prepared for deep sequencing according to standard ChIP-Seq library preparation techniques (Illumina), and DNA sequencing was performed on an Illumina GAIIx DNA sequencer. Antibodies used in the ChIP-Seq assay are described in Dataset S11.

ChIP-Seq Data Analysis. Single-ended reads of 36 bp in length were sequenced using a GAIIx DNA sequencer. Reads were aligned against the hg19 genome using BWA version 0.6.1 (5) by utilizing a mismatch of 2 bp and a seed length of 36. Duplicate reads were removed using the MarkDuplicates function, and resulting bam files were sorted and indexed using Picard tools (version 1.81).

CSB peak calling. Peak calling was performed using MACS (version 1.4.2) comparing CSB-GFP or CS1AN-GFP samples against their respective input control using default settings of mfold and p.value. Normalized coverage files were generated from bam files using IGV tools (version 2.1.7).

RNAP II density at transcription start site. Read depth coverage over the region 500 bp upstream to 2,000 bp downstream of the transcription start site (TSS) of all Reference Sequence database protein-coding transcripts was calculated and then normalized such that each sample had a total depth equivalent of 20 million reads. The mean normalized read depth over each base-pair position was used to construct a plot of the average TSS profile.

Reprogramming of Human Fibroblasts into Neurons. ***Plasmid construction and viral preparation.*** Lentiviral shRNA against human PTBP1 (TRCN0000001062, cloned in the pLKO.1 vector) and negative control shRNA (TRC lentiviral pLKO.1 empty vector control) were purchased from Thermo Scientific. Lentiviral constructs expressing microRNA (miR)-9/9*-124 (pLemir9-124), nonspecific miR (pLemiR-NS), and individual neural transcription factors (phASCL1-N106, phMYT1L-N106, and phND2-N174) were purchased from Addgene. Lentiviral preparation was as described by Xue et al. (6) and Yoo et al. (7). Briefly, an individual construct was first packaged into replication-incompetent lentiviral particles in HEK 293T cells by cotransfecting individual lentiviral vectors with the packaging plasmids (Gag-pol and vesicular stomatitis virus glycoprotein). Lentiviral particles were collected twice, 48 h and 72 h posttransfection.

Cell culture. Human fibroblasts were maintained in fibroblast medium [DMEM supplemented with 10% (vol/vol) FBS and penicillin/streptomycin]. Cells were first seeded onto gelatin-coated culture vessels and then transduced the next day with lentiviral particles in the presence of 8 μg/mL polybrene. After overnight incubation, the cells were selected with 0.2 μg/mL puromycin. Two days after selection, the cells were switched to N3 medium, which consists of 1:1 DMEM/F12 + 1% N2 supplement (Invitrogen) and 20 ng/mL basic FGF to enhance the viability of cells. After 2 wk, human BDNF, GDNF, CNTF, and NT3 (10 ng/mL of each; Peprotech) were added to the medium.

Neuroblastoma Differentiation. Human SH-SY5Y cells were seeded onto poly-L-lysine (50 μg/mL)–coated plates at an initial density of 10^4 cells per square centimeter. All trans-retinoic acid (RA; Sigma) was added at a final concentration of 5 μM in N2 medium (1:1 DMEM/F12 + 1% FBS + 1% N2 supplement) the next day after plating. The medium was changed every 3 d. Mouse Neuro2a cells were plated in the same way as SH-SY5Y cells. Cells were incubated in serum-free medium 24 h after plating; with the medium changed every 3 d, neurites can be observed at day 6.

Immunofluorescence. After fixation with 4% (vol/vol) paraformaldehyde in PBS for 15 min at room temperature, cells were permeabilized in 1× PBS containing 0.1% Triton X-100 and blocked with blocking solution [1× PBS containing 0.01% Triton X-100, 10% (vol/vol) FBS, and 3% (wt/vol) BSA] for 1 h. Primary antibodies in blocking solution were then added and incubated for 1 h at room temperature, followed by washing and incubation with fluorophore-conjugated corresponding secondary antibodies. Antibodies used in the immunofluorescence are described in Dataset S11. Cell images were acquired on a Zeiss Axio Imager M1 microscope equipped with an ORCA-ER camera (Hamamatsu) and controlled by Volocity version 5.5.1 software.

Quantification of differentiated SH-SY5Y cells. Fluorescent images of the cells were collected using an objective lens with a magnification of 10×. Differentiated SH-SY5Y cells with neurite outgrowth were defined as Tuj1-positive cells with neurites longer than twice the length of the cell body. The percentage of differentiated cells was normalized to the total cell number depicted by DAPI staining (Fig. 3*D*; $n > 660$ cells from three independent experiments in an shGFP SH-SY5Y cell line and $n > 570$ cells from three independent experiments in an shCSB SH-SY5Y cell line). *P* values were calculated using a two-tailed *t* test.

RNA Isolation and cDNA Synthesis. Total RNA was extracted from cultured cells with an RNeasy Mini Kit (Qiagen), according to the manufacturer's instructions. The integrity of the RNA was tested on a denaturing agarose gel. RNA quality and quantity were also assessed with a Nanodrop spectrophotometer (Thermo Fisher Scientific). Total RNA from ∼100 mg of frozen human brain tissue [kindly provided by the National Institute of Child Health and Human Development (NIHCD) Brain and Tissue Bank for Developmental Disorders, University of Maryland, Baltimore] was isolated using the Qiagen RNeasy Lipid Tissue Mini Kit. RNA concentration and integrity were assessed by Agilent Bioanalyzer nano-ChIP.

For the microarray analysis, double-stranded cDNA was synthesized from 10 μg of total RNA, using the cDNA synthesis kit according to the NimbleGen user protocol. For quantitative RT-PCR (qRT-PCR) analysis, single-stranded cDNA was synthesized from 200 ng of total RNA using a TaqMan Reverse Transcription Kit (Invitrogen).

DNA Microarray and qRT-PCR. Up to 1 μg of double-stranded cDNA was labeled with Cy3 and hybridized on a NimbleGen 12×125K Human Expression Array (2007-09-12_HG18_opt_expr), followed by washing and drying according to the manufacturer's instructions (NimbleGen–Roche). Arrays were scanned with a NimbleGen MS200 microarray scanning system, and data acquisition was performed with NimbleScan according to the manufacturer's recommendations. Real-time PCR was performed using iQ-SYBR Green Supermix on a CFX96 Real-Time PCR System (BioRad). Reactions were run in duplicate in three independent experiments. The primer sequences are listed in Dataset S12. Expression data were normalized to the housekeeping genes *GAPDH* and 18S rRNA, using the $2^{-\Delta\Delta CT}$ method (8).

DNA Microarray Analysis. ***Identification of differentially expressed genes in microarray data.*** Data were deposited in the Gene Expression Omnibus database (accession no. GSE58071) and analyzed using Bioconductor version 2.12 (www.bioconductor.org) running on R version 3.0.0. Raw pair files (NimbleGen Homo Sapiens HG18 Expression Array) were quantile-normalized in R. Differential gene expression was assessed using an empirical Bayes *t* test (Limma package). *P* values were adjusted for multiple testing using the Benjamini–Hochberg method (9). Any probe sets that exhibited an adjusted *P* value of 0.05 were defined as differentially expressed. In addition, any probe sets that exhibited an

absolute fold change of >2 (Fig. 2*D* and Fig. 6*C*) or a fold change of >1.5 (Figs. 1*A* and 4 *A* and *C* and Figs. S1*D* and S4*C*) were used to generate a heat map.

Samples were clustered using a 1 − Pearson correlation distance-matrix and average linkage clustering. Genes were clustered using a Euclidean distance matrix and average linkage clustering. In the heat map, red indicates higher expression and blue indicates lower expression relative to the mean expression of probes across all samples. In Fig. 4*A*, *k*-means analysis was used to identify six clusters. Average profiles per cluster were generated using the mean expression value of genes per time point.

Defining CSB-dependent temporal gene expression during neuroblastoma differentiation. Time-dependent transcriptional changes that are specific to CSB were identified by two-way ANOVA using Limma (9). Any probe sets that were differentially expressed with a relaxed *P* value of 0.01 were used to generate a 2D hierarchical heat map. Gene probes were clustered using a Euclidean distance matrix and average linkage clustering. Expression values of gene probes at individual time points upon RA treatment were normalized to relative samples at time 0.

To detect the change in gene expression with biological significance, the ∼1,000 gene probes detected from the ANOVA analysis were examined manually; ∼100 genes were selected using the filter that the expression level of a gene at no less than three of five adjacent time points has to show the change in the same direction vs. untreated samples in the WT cell line.

Gene ontology category enrichment. Differentially expressed gene probes were collapsed at the gene level to obtain differentially expressed gene numbers. Differentially expressed genes were used to look for gene set enrichment from gene ontology biological processes/molecular functions using the MetaCore pathway tool (http://thomsonreuters.com/metacore/). The analysis employs a hypergeometric distribution to determine enriched gene sets. All genes on the NimbleGen Expression Array were used as the background for enrichment analyses. The *P* value was corrected using the Benjamini–Hochberg multiple testing correction method.

Correlations between comparisons.

i) A total of 4,130 genes were identified to be differentially expressed between tissue samples from patients with CS and control tissue samples, with an adjusted *P* value of 0.05 and fold change of greater than 1.5 (called "CS patient cerebella" in Fig. S5*B*).

ii) A total of 1,438 genes were found to be CSB-dependent and differentially expressed during time course differentiation in neuroblastoma SH-SY5Y cells, with a relaxed *P* value of 0.01 and fold change of greater than 1.5 (called "Differentiated SH-SY5Y" in Fig. S5*B*).

iii) A total of 1,242 genes were found to be differentially regulated in CSB-deficient fibroblasts vs. reconstituted WT cells, with an adjusted *P* value of 0.05 and fold change of greater than 1.5 (called "CS fibroblasts" in Fig. S5*B*).

iv) A total of 1,512 genes were identified to be CSB-dependent and differentially expressed at day 3 after PTB depletion, with a relaxed *P* value of 0.01 and fold change of greater than 1.5 (called "Transdiff CS fibroblasts" in Fig. S5*B*).

The pbinom function in R was used to determine the chance of obtaining an overlap between any two or more differentially expressed gene lists.

Clonogenic survival assay. Exponentially growing cells were plated in 100-mm tissue culture dishes (1×10^4 cells). After incubation for 24 h, the cells were exposed to UV irradiation from 0 to 10 J/m^2, and the cultures were maintained until surviving cells formed colonies. Cells were fixed and stained with a mixture of 0.5% crystal violet in 50:50 methanol/water for 30 min. Colonies were scored, and the surviving fractions for each dose were calculated.

Cell cycle analysis. A total of 1–5 × 10^5 cells were collected by centrifugation at 300 × *g* for 3 min in an Eppendorf centrifuge 5804R. Cells were then washed and resuspended in 100 μL of 1× PBS containing 1% FBS. Cold methanol (1 mL) was added to the cell suspension drop-wise while vortexing the open tube at minimum speed on a Vortex-Genie 2 vortexer (Scientific Industries, Inc.). Cells were fixed at 4 °C from 1 h up to 1 wk. After this treatment, cells were washed once in 1× PBS + 1% FBS and resuspended in 0.5–1 mL of staining solution (1× PBS, 1% FBS, 25 μg·mL^{-1} propidium iodide, and 10 μg·mL^{-1} RNase A). Samples were incubated at 37 °C for 30 min in the dark. Before acquisition on the flow cytometer cell scanner, cells were filtered through a 0.45-μm mesh.

Western blot. Cells were lysed on ice in TENT buffer [20 mM Tris·HCl (pH 8.0), 2 mM EDTA, 150 mM NaCl, and 0.5% Triton X-100] containing protease inhibitor mixture, sonicated briefly, and then centrifuged at 20,000 × *g* for 20 min at 4 °C. The supernatants were then mixed with Laemmli sample buffer and resolved by SDS/PAGE and Western blot. Antibodies are described in Dataset S11.

Brain tissue samples. All brain tissue samples were kindly provided by the NICHD Brain and Tissue Bank for Developmental Disorders. A brief summary of the clinical history is provided in Dataset S8.

1. Zhang Y, et al. (2008) Model-based analysis of ChIP-Seq (MACS). *Genome Biol* 9(9): R137.
2. Lake RJ, et al. (2014) The sequence-specific transcription factor c-Jun targets Cockayne syndrome protein B to regulate transcription and chromatin structure. *PLoS Genet* 10(4):e1004284.
3. Poser I, et al. (2008) BAC TransgeneOmics: A high-throughput method for exploration of protein function in mammals. *Nat Methods* 5(5):409–415.
4. Saponaro M, et al. (2014) RECQL5 controls transcript elongation and suppresses genome instability associated with transcription stress. *Cell* 157(5):1037–1049.
5. Li H, Durbin R (2009) Fast and accurate short read alignment with Burrows-Wheeler Transform. *Bioinformatics* 26(5):589–595.
6. Xue Y, et al. (2013) Direct conversion of fibroblasts to neurons by reprogramming PTB-regulated microRNA circuits. *Cell* 152(1-2):82–96.
7. Yoo AS, et al. (2011) MicroRNA-mediated conversion of human fibroblasts to neurons. *Nature* 476(7359):228–231.
8. Livak KJ, Schmittgen TD (2001) Analysis of relative gene expression data using real-time quantitative PCR and the 2(-Delta Delta C(T)) Method. *Methods* 25(4):402–408.
9. Smyth G (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.

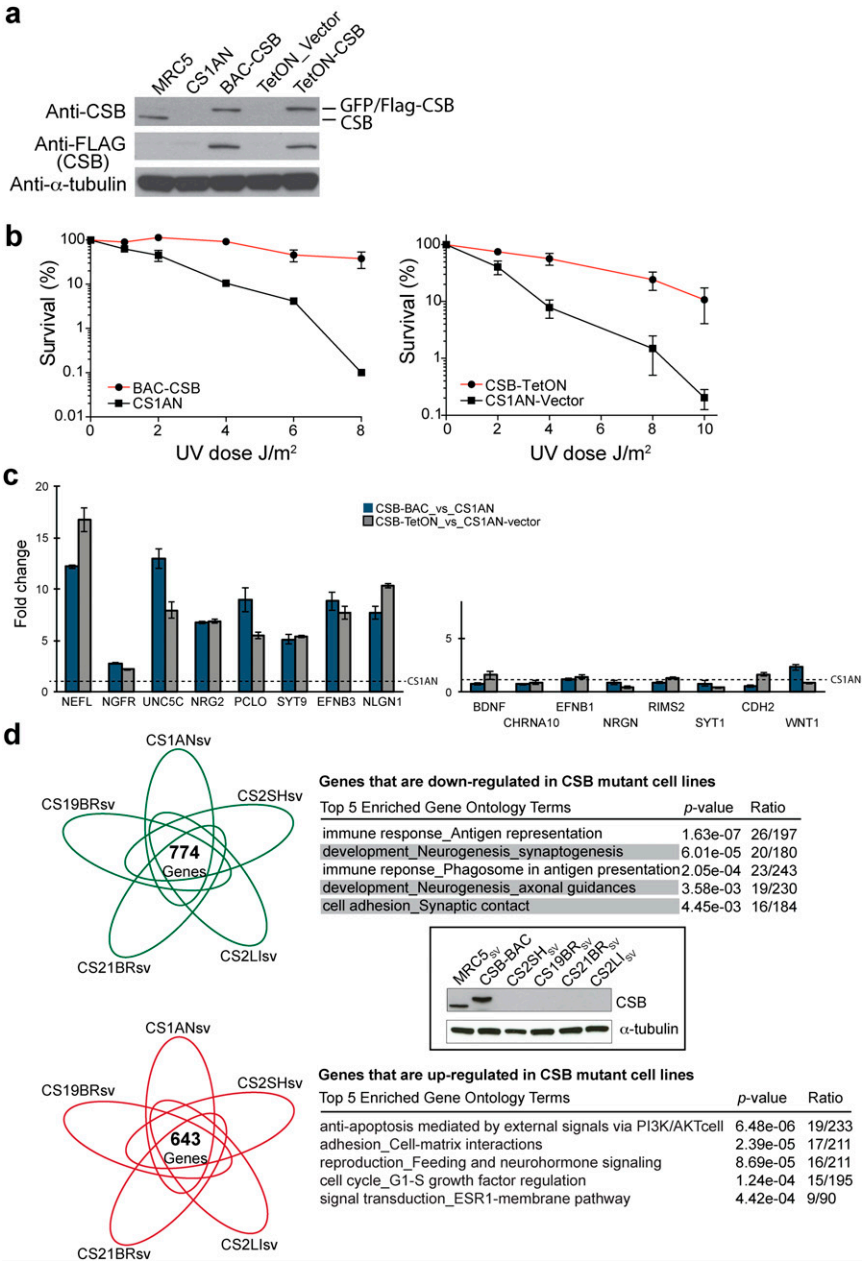


Fig. S1. Analysis of CS1AN cells and their WT reconstituted counterparts. (*A*) Western blot analysis of endogenous CSB protein in human fetal lung fibroblast MRC-5 (MRC5sv) cells and exogenous expression of GFP/Flag-tagged CSB in BAC-CSB and CSB doxycycline-regulated promoter (TetON) cell lines. (*B*) Clonogenic survival assay measuring UV radiation sensitivity of WT [BAC-CSB (*Left*) and CSB-TetON (*Right*)] fibroblasts. Error bars indicate SD of three experiments. (*C*) qRT-PCR validation of selected neuronal genes in WT reconstituted cells (BAC-CSB or CSB-TetON) compared with CS1ANsv cells, confirming the results from the microarray experiment. mRNA levels are shown relative to those mRNA levels seen in CS1AN (stippled line) cells, and were normalized to *GAPDH*. Some genes were called as significantly changed (*Left*), as determined by the microarray analysis, whereas some were not (*Right*). (*D*, *Left*) Venn diagrams showing genes significantly up- and down-regulated in the five different CS fibroblast cell lines compared with BAC-CSB by microarray ($n = 2$). (*D*, *Right*) Top five enriched gene ontology terms of dysregulated genes. (*Inset*) Western blot analysis of endogenous CSB in MRC5sv, exogenous GFP/Flag-tagged BAC-CSB, and lack of detectable CSB in the four additional individual CS fibroblasts.

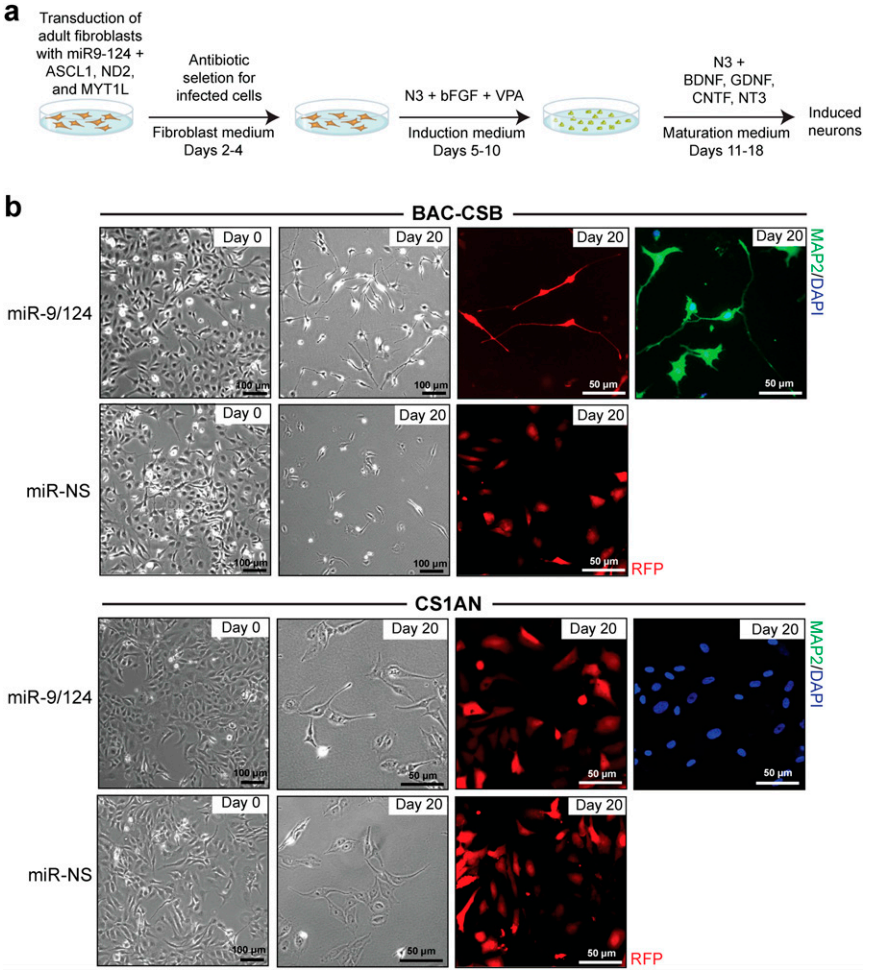


Fig. S2. Conversion of fibroblasts to neurons via miR-9/124 overexpression requires CSB. (*A*) Schematic of experimental approach. CS1AN and CSB-corrected fibroblasts (BAC-CSB) were infected with lentivirus expressing miR-9/124, or nonspecific miR (miR-NS), both marked by RFP, combined with neuronal gene transcription regulators ASCL1, ND2, and MYT1L. (*B*) Images showing morphological changes of fibroblasts induced by microRNAs. CSB-corrected fibroblasts were converted to neuron-like cells (*Upper*), but morphology changes were not observed in CS1AN cells (*Lower*). (*Right*) Neuronal marker MAP2 stained in green and nuclei stained in blue.

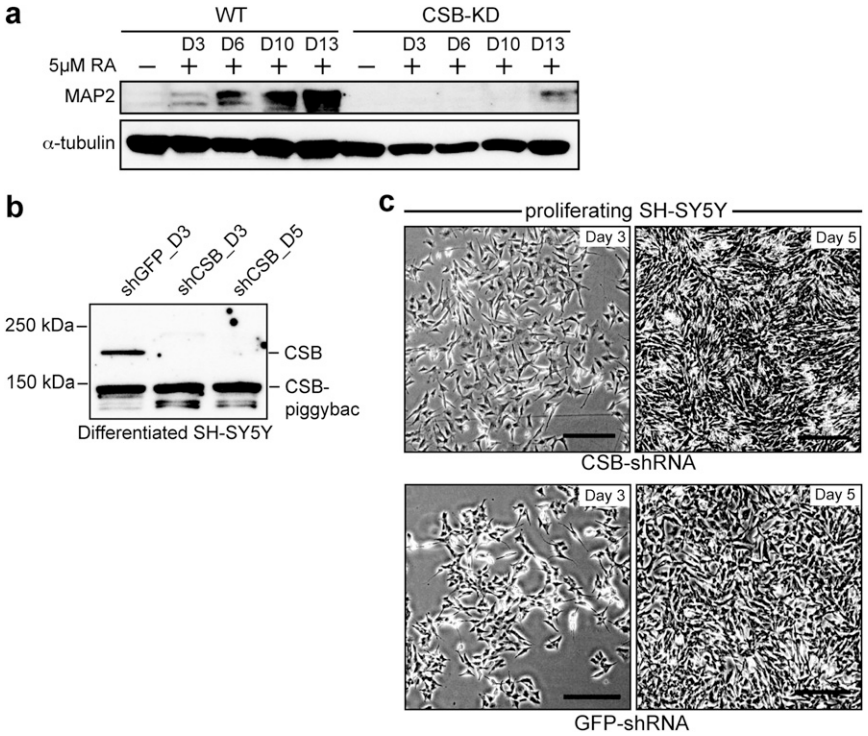


Fig. S3. Analysis of SH-SY5Y cells with or without CSB. (*A*) Induction of MAP2 at different times after RA-induced differentiation, in the presence and absence of CSB. D, day. (*B*) CSB knockdown by shRNA_16776 in undividing SH-SY5Y cells is efficient. (*C*) CSB knockdown does not cause nonspecific cytotoxicity to neuroblastoma cells. SH-SY5Y CSB knockdown (and control) cells were cultured in proliferating medium, and images were from day 3 and day 5. CSB-depleted cells exhibited similar cell density compared with control. (Scale bar: 100 μm.)

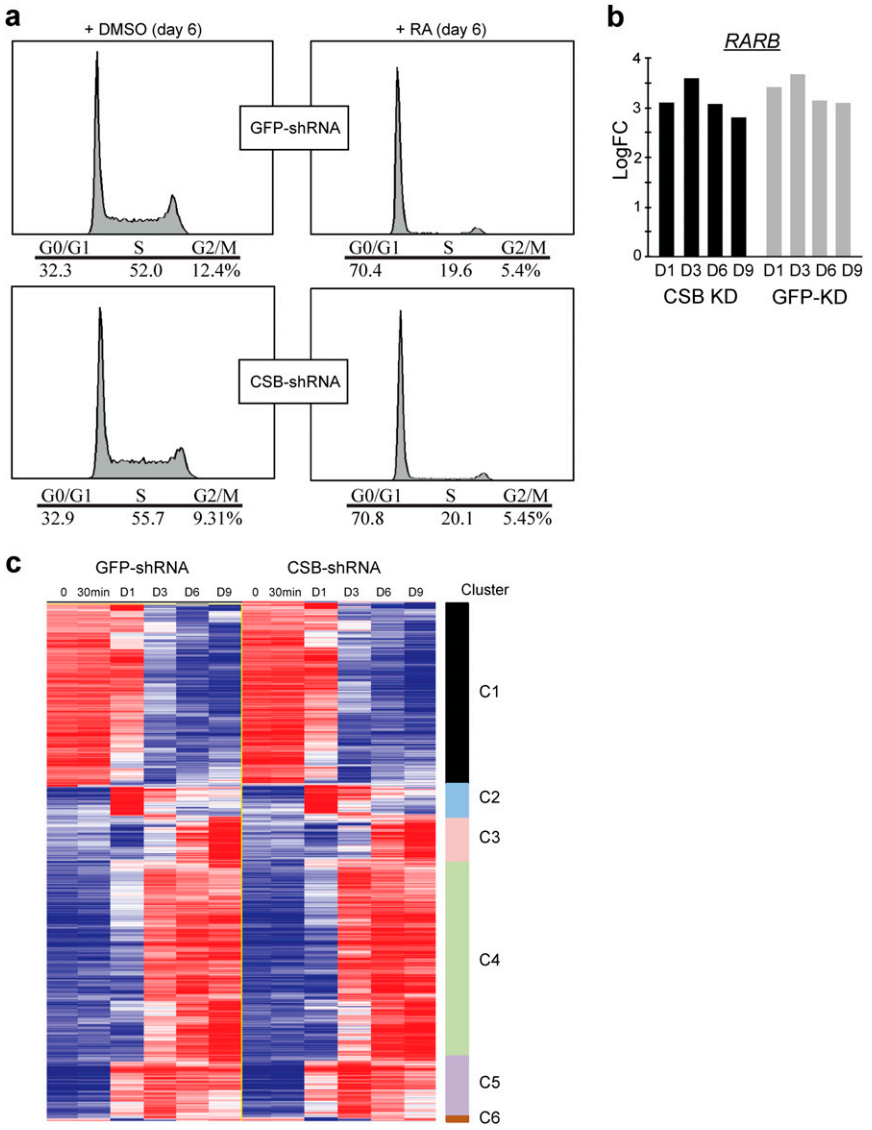


Fig. S4. Effect of CSB knockdown in SH-SY5Y neuroblastoma cells. (*A*) Cell cycle analysis by propidium iodide staining and flow cytometry of GFP-shRNA cells (control, *Upper*) or CSB-shRNA SH-SY5Y cells (*Lower*) without (DMSO) or with 5 μM RA for 6 d. (*B*) Expression of the gene encoding the major RA receptor, *RARB*, is normal in the CSB-depleted cells. The fold change in gene expression was calculated relative to the DMSO-treated (−RA) samples. LogFC, Log fold change. (*C*) Heat map showing ∼3,000 genes that change expression in response to RA treatment at six different time points. Note the overall relatively similar clustering in the two cell types.

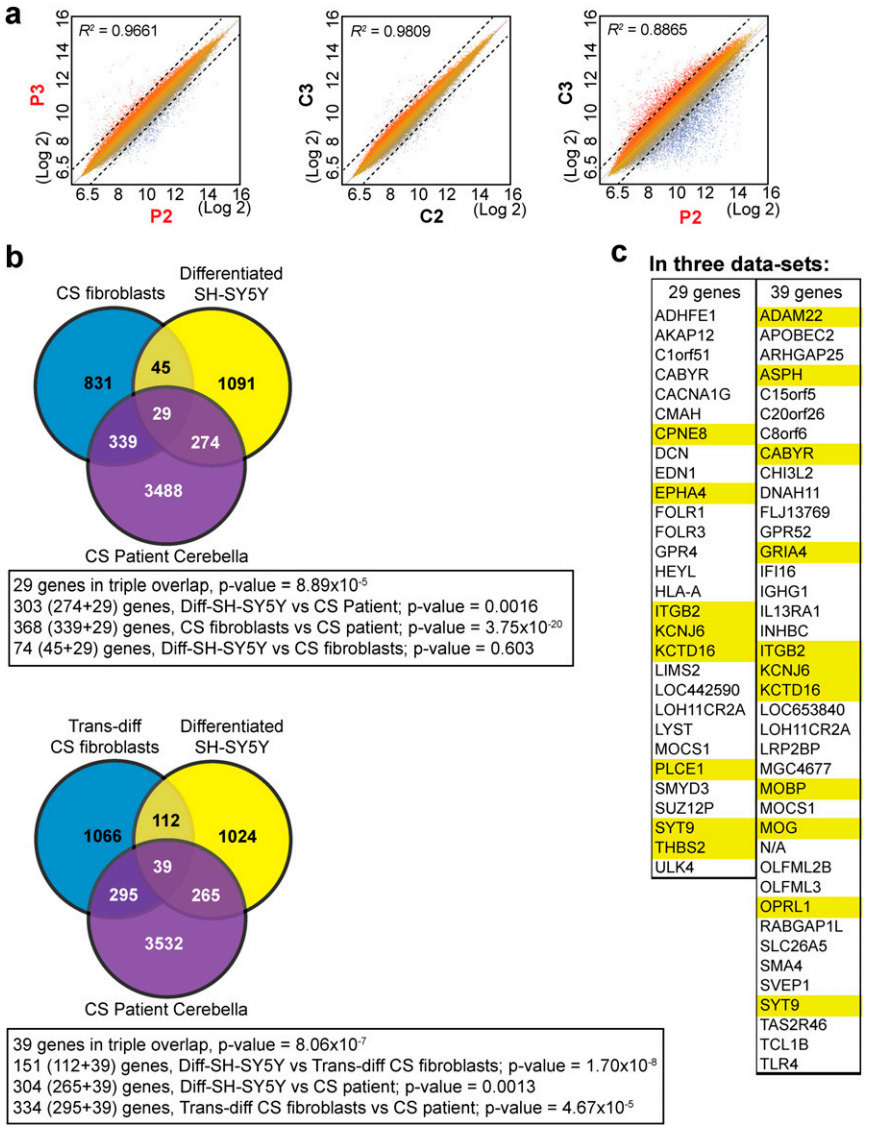
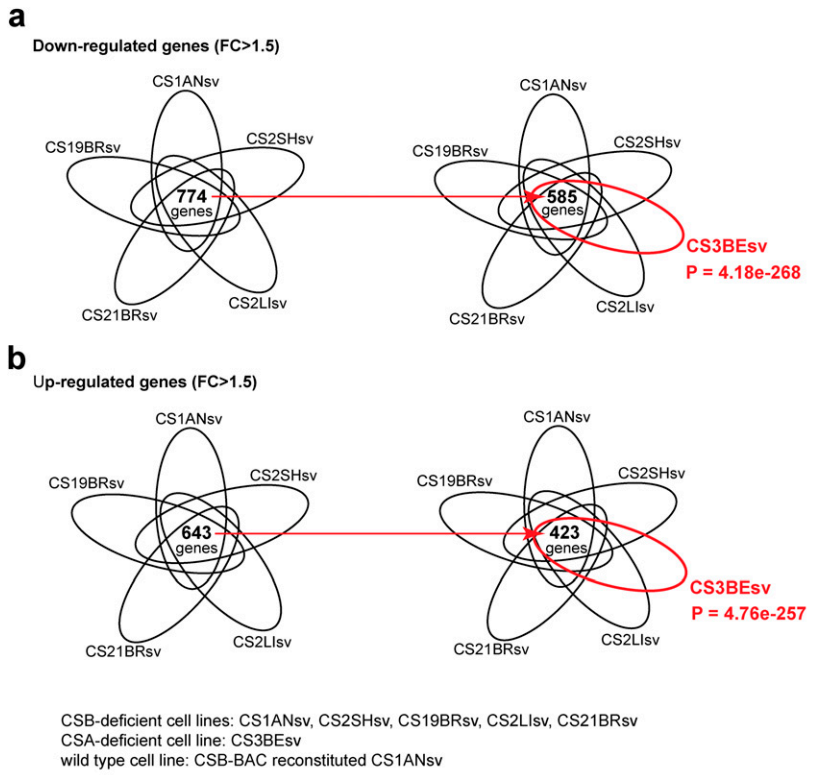


Fig. S5. Genes whose expression is affected by CSB deficiency in different cell contexts. (*A*) Representative scatter plots of pairwise comparisons between patient and nonpatient gene expression levels. Each dot represents an individual gene. Black dashed lines indicate a twofold deviation from the trend line. The Pearson R^2 is shown in left corner. (*B*) Venn diagram showing the number of affected genes and their overlap. *P* values for overlap are listed below. Note that the genes affected in patient cerebella significantly overlap with genes affected in CS1AN fibroblasts and in neuroblastoma cells (SH-SY5Y cells). Trans-diff, Transdifferentiated. (*C*) List of the overlapping genes whose expression changed in all three transcriptomes indicated in *B*. Yellow indicates neuronal genes.



c

Top 1 annotation cluster of the common 585 down-regulated genes in CSB and CSA deficient cell lines

Enriched Gene Ontology Terms	*p*-value	Count
calcium-dependent cell-cell adhesion	2.58e-09	10
synaptogenesis	1.06e-06	9
homophilic cell adhesion	3.16e-05	14
cell-cell adhesion	9.46e-05	20
cell adhesion	1.23e-04	36
biological adhesion	1.23e-04	36
synapse organization	1.56e-04	9
synaptic transmission	2.56e-04	20
transmission of nerve impulse	2.79e-04	22
cell-cell signaling	0.01	26
extracellular structure organization	0.02	10

Fig. S6. Transcriptome analysis of CSA-deficient fibroblasts. (*A* and *B*, *Left*) Venn diagrams showing genes significantly down- and up-regulated in the five different CS fibroblast cell lines compared with BAC-CSB ($n = 2$). (*B*, *Right*) As in *A*, but adding genes deregulated by CSA mutation (red circle). The *P* value for the overlap was calculated as described in *SI Methods*. (*C*) Top 10 enriched gene ontology (GO) terms for down-regulated genes by CSB and CSA mutations. Gene categories related to neurogenesis are highlighted in gray.

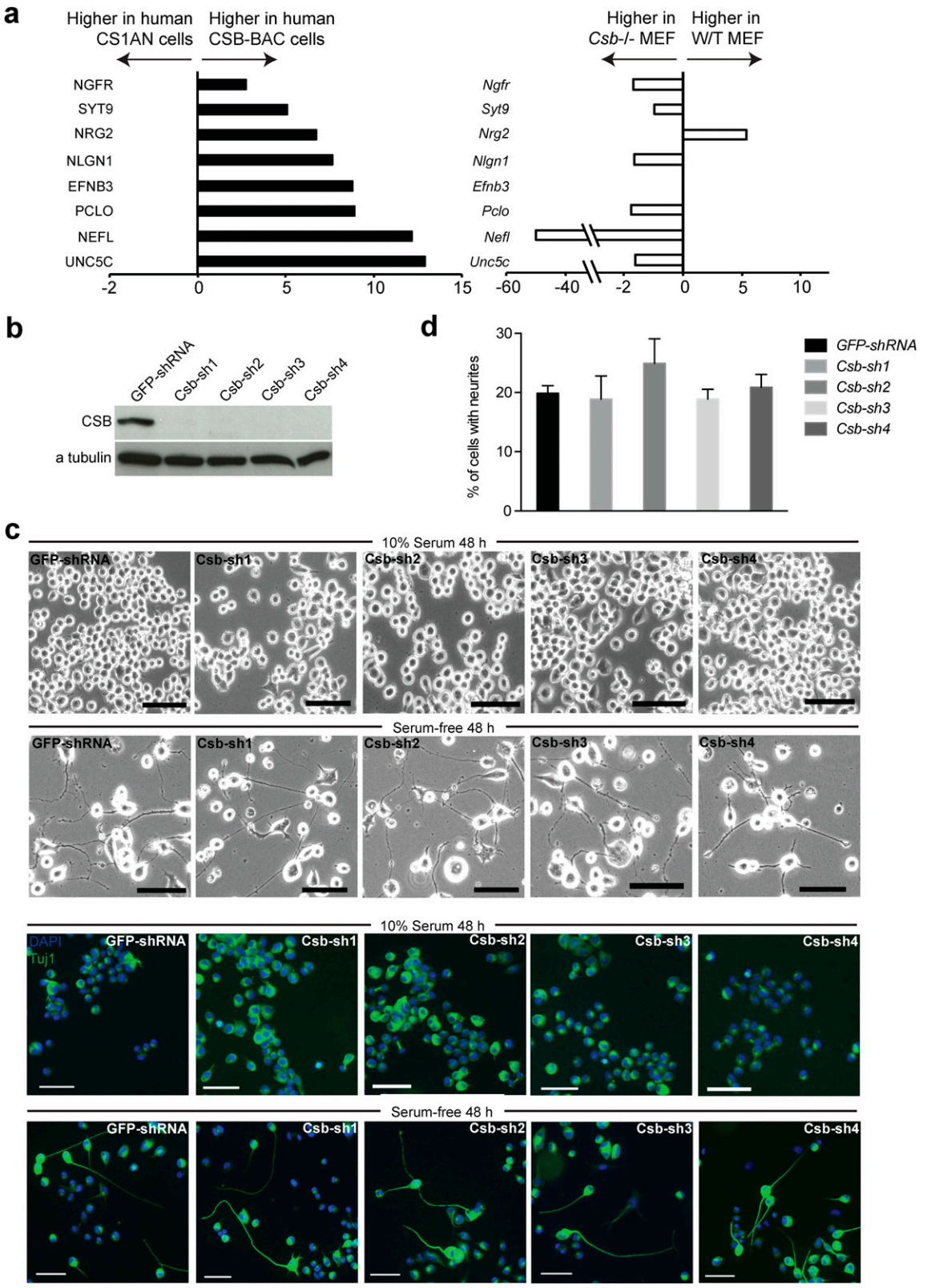


Fig. S7. Characterization of mouse fibroblasts and neuroblastoma cells in the absence of CSB. (*A*) Bar graphs summarizing the representative differentially expressed genes between human (black bars) and mouse (white bars) fibroblasts. (*Left*) CSB-deficient human fibroblasts have lower expression of all of the tested neuronal genes. (*Right*) However, the same genes in $Csb^{-/-}$ mouse embryonic fibroblasts (MEFs) are differently expressed relative to the WT control MEFs. (*B*) Western blot analysis of whole-cell extracts shows the protein expression of *Csb* from mouse neuroblastoma Neuro2a cells transduced with lentiviruses carrying *shRNAs* against either a nontargeting sequence (GFP-*shRNA*) or *Csb*. Four independent *Csb*-targeting *shRNAs* were used (Csb-sh1, Csb-sh2, Csb-sh3, and Csb-sh4). a tubulin, α-tubulin. (*C*) Csb is dispensable for Neuro2a differentiation in vitro. Neuro2a cells were subjected to serum withdrawal for 48 h.

Legend continued on following page

Immunofluorescence staining of Neuro2a cells cultured in complete or serum-free medium with anti-Tuj1 antibody (green) and DAPI (blue) is shown. (Scale bars: 70 μm.) (*D*) Graph represents the quantification of morphologically differentiated cells. The length of neurites was measured using ImageJ software (National Institutes of Health), and a differentiated cell was defined as a cell with a neurite greater than two cell bodies in length. The population of differentiated cells was calculated as cells with neurites relative to total cell number ($n \sim 1{,}000$ per cell line). Error bars represent mean ± SD.

Other Supporting Information Files

Dataset S1 (XLSX)
Dataset S2 (XLSX)
Dataset S3 (XLSX)
Dataset S4 (XLSX)
Dataset S5 (XLSX)
Dataset S6 (XLSX)
Dataset S7 (XLSX)
Dataset S8 (XLSX)
Dataset S9 (XLSX)
Dataset S10 (DOCX)
Dataset S11 (XLSX)
Dataset S12 (XLSX)