Dysregulation of gene expression as a cause of Cockayne syndrome neurological disease

Yuming Wang*, Probir Chakravartyb, Michael Ranesb, Gavin Kellyb, Philip J. Brooksb, Edward Neilandi, Aengus Stewartha, Giampietro Schiavoro, and Jesper Q. Svejstrupab,1

*Mechanisms of Transcription Laboratory, Cancer Research UK London Research Institute, South Mimms EN6 3LD, United Kingdom; bBioinformatics and Biostatistics Group, Cancer Research UK London Research Institute, London WC2A 3LY, United Kingdom; Laboratory of Neurogenetics, National Institute on Alcohol Abuse and Alcoholism and Office of Rare Diseases Research, National Center for Advancing Translational Sciences, National Institutes of Health, Bethesda, MD 20892; 1Genetics and Metabolism, Boston Children’s Hospital, Harvard Medical School, Boston, MA 02115; and 1Obel Department of Motor Neuroscience and Movement Disorders, Institute of Neurology, University College London, London WC1N 3BG, United Kingdom

Edited* by Alexander Varshavsky, California Institute of Technology, Pasadena, CA, and approved September 2, 2014 (received for review July 3, 2014)

Cockayne syndrome (CS) is a multisystem disorder with severe neurological symptoms. The majority of CS patients carry mutations in Cockayne syndrome group B (CSB), best known for its role in transcription-coupled nucleotide excision repair. Indeed, because various repair pathways are compromised in patient cells, CS is widely considered a genome instability syndrome. Here, we investigate the relationship between the neuropathology of CS and dysregulation of gene expression. Transcriptome analysis of human fibroblasts revealed that even in the absence of DNA damage, CSB affects the expression of thousands of genes, many of which are neuronal genes. CSB is present in a significant subset of these genes, suggesting that regulation is direct, at the level of transcription. Importantly, reprogramming of CS fibroblasts to neuron-like cells is defective unless an exogenous CSB gene is introduced. Moreover, neuroblastoma cells from which CSB is depleted show defects in gene expression programs required for neuronal differentiation, and fail to differentiate and extend neurites. Likewise, neuron-like cells cannot be maintained without CSB. Finally, a number of disease symptoms may be explained by marked gene expression changes in the brain of patients with CS. Together, these data point to dysregulation of gene regulatory networks as a cause of the neurological symptoms in CS.

Significance

Cockayne syndrome (CS) is an autosomal-recessive, multisystem disorder characterized by severe neurological disease, growth failure, developmental abnormalities, photosensitivity, and degeneration of organ systems such as the ear and eye, including cataracts (1, 2). The majority of patients who have CS carry mutations in the gene encoding the SWI/SNF family DNA translocase Cockayne syndrome group B (CSB)/ERCC6 (~80% of patients) or the gene encoding ubiquitin ligase-associated CSA/ERCC8. These proteins are best known for their role in transcription-coupled nucleotide excision repair (TC-NER), a process whereby bulky DNA lesions, such as those generated by UV irradiation, are preferentially removed from the transcribed strand of active genes (3, 4). CS is thus frequently referred to as a TC-NER disease (e.g., ref. 5). However, CS cells are sensitive to a number of additional DNA-damaging agents, and to oxidative damage in particular (6, 7), implicating the CS proteins in other repair pathways as well. Indeed, the idea that CS results from inefficient repair of oxidative DNA damage has gained momentum over the past decade (reviewed in refs. 8, 9). Finally, studies from Weiner and coworkers (10), Egly and coworkers (11, 12), and others have reported evidence of a role for CSB in gene regulation, which might provide an alternative explanation for CS (reviewed in refs. 2 and 12). However, direct evidence for gene expression changes in CS patients has not been reported, and the relationship between deficiencies in molecular pathways affected by CS mutation and patient disease symptoms has generally remained tenuous, or unexplored. Here, we provide evidence from human and mouse cell models, as well as brain tissue from patients with CS, that the involvement of CSB in regulating expression levels of protein-encoding genes may explain several features of CS neurological disease.

Results and Discussion

CSB Affects Transcription of Numerous Genes. We used microarray analysis to investigate CSB-dependent gene expression. These experiments were initially performed with CS1ANsv [a simian virus (sv) 40-transformed patient cell line] and two different CSB-reconstituted (WT) counterparts derived from it. In the first, CSB reexpression was achieved by introducing a BAC carrying the normal CSB gene (BAC-CSB). In the other, the CSB protein was expressed to near-normal levels from a tetracycline/doxycycline-regulated promoter (CSB-TetON). As expected, both rescued the UV sensitivity of the CS fibroblasts (Fig. S1 A and B). Even with the conservative requirement that the expression level of a gene had to be statistically significant between CS1ANsv and both kinds of reconstituted WT cells, more than 1,200 genes were markedly (>1.5-fold) deregulated in the CSB-deficient cell line (Fig. 1A, Fig. S1C, and Dataset S1). This level of gene dysregulation is in agreement with earlier work by Weiner and coworkers (10).

Author contributions: Y.W. and J.Q.S. designed research; Y.W. performed all experiments; M.R. contributed new reagents/analytic tools; Y.W., P.C., G.K., A.S., G.S., and J.Q.S. analyzed data; P.J.B. and E.N. genotyped or phenotyped Cockayne syndrome patient material; and Y.W., G.S., and J.Q.S. wrote the paper.

The authors declare no conflict of interest.

*This Direct Submission article had a prearranged editor.

Freely available online through the PNAS open access option.

Data deposition: Microarray data, RNA polymerase II ChIP combined with deep sequenc- ing (ChIP-Seq) data, and CSB ChIP-Seq data reported in this paper are deposited in the Gene Expression Omnibus database (accession no. GSE58071).

1To whom correspondence should be addressed. Email: jesper.svejstrup@cancer.org.uk.

This article contains supporting information online at www.pnas.org/lookup/suppl/doi:10.1073/pnas.1412569111/-/DCSupplemental.
When performing gene ontology analysis of the genes requiring CSB for normal expression, we noticed that three of the top five classes of genes that were less expressed in CSB-deficient cells are connected to neurons (Fig. 1B). This link had not been observed previously (10), possibly because the earlier studies used less gene-rich microarrays. To investigate whether the expression changes were an oddity of the CS1ANsv cell line or more generally observed, gene expression was analyzed in four additional CS patient cell lines. Individual genes affected in CS1ANsv were also affected in the other patient cell lines (Fig. 1C), and microarray analysis of all five CS cell lines together showed a significant degree of overlap, including a consistent down-regulation of neuronal genes (Fig. S1D), indicating that the gene expression defects observed were indeed linked to CSB function.

To examine to what extent the effect of CSB mutation on gene expression is direct and at the level of RNA polymerase II (RNAP II) transcription, we investigated the genome-wide CSB and RNAP II distribution by ChIP combined with deep sequencing (ChIP-Seq); NRG2 and SYT9 are shown as examples in Fig. 1D. Both CSB and RNAP II were detected at these genes, and, importantly, not only CSB but also the polymerase disappeared from the genes in cells lacking CSB. This correlation suggests that CSB has a direct effect on gene expression, at the level of RNAP II transcription, at least in a subset of genes. In general, peaks of CSB density in the genome overlapped significantly with the genes that were deregulated in CSB-deficient cells ($P$ value of $1.04 \times 10^{-5}$). Likewise, genes whose expression was decreased in the absence of CSB also often displayed a decrease in RNAP II density in CS1ANsv cells ($P$ value for overlap of $1.2 \times 10^{-7}$; Dataset S2), further supporting the thesis that the effect of CSB is at the level of transcription (additional discussion and data are provided in SI Text and Dataset S3). Together, these results indicate that CSB has a broad, but gene-specific, effect on RNAP II transcription.

**CSB Is Required for Transdifferentiation of Fibroblasts to Neurons.**

Neuronal genes are typically not highly expressed in fibroblasts. To examine the possible physiological relevance of gene expression deficiencies in CSB-deficient cells, we investigated cellular reprogramming of fibroblasts to neurons (13). This approach has previously been used to study diseases such as Alzheimer’s disease (14) and Parkinson disease (15). We used either shRNA knockdown of PTB (16) (a regulator of microRNA circuits) (Fig. 2A) or overexpression of microRNA (miR)-9/124 combined with three neuronal transcription regulators (17) (Fig. S2A) to obtain induced neurons from the CSB-reconstituted human fibroblasts (Fig. 2B, B, and C, and Fig. S2B). Crucially, the parental CS1ANsv cells could not be converted to neurons by this approach but retained the cellular morphology of fibroblasts (Fig. 2B, B, and G) and lacked expression of neuronal marker MAP2 (Fig. 2B, B, and H). Virtually identical results were obtained in the experiments where neuronal conversion was driven by miR-9/124 expression instead (Fig. S2B). Thus, CSB is required for transdifferentiation of fibroblast to neurons.

The program switch between PTB and its neuron-specific homolog, nPTB, is a key event during neuronal differentiation (18). Upon treatment with PTB shRNA, we detected increased levels of nPTB in BAC-CSB cells (Fig. 2C, compare lanes 1 and 2) but not in the CSB-deficient cells (Fig. 2C, compare lanes 3 and 4). We also characterized gene expression in the early phases of transdifferentiation using microarray analysis. Relative to control cells, 123 genes, of which 32 are involved in neuronal differentiation, were consistently up-regulated already at day 3 after initiating PTB knockdown in BAC-CSB cells. Remarkably, the vast majority of these 32 genes failed to be induced in CS1ANsv cells (Fig. 2D and Dataset S4), providing a likely mechanism for their failure to make the conversion to neuron-like cells.

**CSB Is Required for Neuroblast Differentiation.** The physiological consequence of CSB loss was also investigated in another cell model. The neuroblastoma cell line SH-SY5Y (19) stops proliferating and undergoes neuronal differentiation upon treatment with retinoic acid (RA), extending neurites and expressing neuronal markers (20) (Fig. 3A and Fig. S3A). CSB was deleted from this cell line by treatment with two different shRNAs: 16777 consistently depleted CSB to about 30% of normal levels, whereas 16776 depleted it to below Western blot detection levels (Fig. 3B). Remarkably, CSB depletion effectively blocked cell differentiation and neurite outgrowth, with lower CSB levels correlating with fewer and shorter neurites, or none at all (Fig. 3C, Left, compare $e$ and $f$ with $d$). Staining for the neuronal marker TuJ1 further supported this conclusion (Fig. 3C, Right, compare $e$ and $f$ with $d$; quantified in Fig. 3D). Importantly, CSB was also required for neuronal maintenance: The long neurites
CSB is required for reprogramming of neurons from human fibroblasts. (A) Procedure used for transdifferentiation. (B) Characterization of fibroblast-derived neurons. Black arrowheads in b denote examples of neurite outgrowth, whereas white arrowheads in c indicate the relatively few cells that stained negative for MAP2 in WT cells. (C) Western blot for PTB and nPTB in WT and CS1AN fibroblasts with or without PTB knockdown. (D) Heat map showing neural genes up-regulated greater than twofold (i.e. >1 Log_{foldChange}) in WT cells, and the corresponding levels in CS1AN cells, 3 d after PTB depletion.

present in differentiated SH-SY5Y cells disappeared, and increasing cell death was observed within a few days when CSB was knocked down, but not in control cells (Fig. 3E, compare a–d with e–h, and Fig. S3B). These results were not due to CSB knockdown being toxic, because knockout in proliferating SH-SY5Y cells had little or no effect (Fig. S3C). We conclude that CSB is required not only for neuronal differentiation but also for neuronal maintenance.

As an initial response to the treatment with RA, the SH-SY5Y neuroblastoma cells exit the cell cycle and initiate differentiation, which is crucial for the generation of neurites. FACS analysis revealed that after 6 d, both control and CSB-depleted cells exhibited the expected G0/G1-phase cell cycle arrest, with a similarly decreased proportion of cells in the S and G2/M phases, suggesting that CSB is dispensable for the initial cellular response to RA. CSB depletion also did not affect expression of the main RA receptor (Fig. S4A and B). However, the neuronal marker MAP2 failed to be induced when CSB was knocked down (Fig. S3A). To expand on this finding, we performed transcriptome analysis at five time points along a 9-d time course during RA-mediated differentiation (Fig. 4A, Upper). This approach identified genes that changed expression (adjusted \( P < 0.05 \)), and these genes could be grouped into six clusters based on their distinct temporal profiles (k-means clustering) (Fig. 4A, Lower Right; two examples of clusters are shown). Not surprisingly, genes relating to the nervous system were up-regulated, whereas genes driving the cell cycle were down-regulated during differentiation (Fig. 4B). The overall gene expression signature across these ~3,000 genes was not dramatically perturbed in CSB-depleted cells (Fig. S4C and Datasets S5 and S6), indicating that, as expected, CSB controls only a subset of genes and is not required for all gene regulation. The ANOVA method (21) was used to identify genes that were differentially expressed in CSB-depleted cells during differentiation (Dataset S7). Although the change observed upon CSB depletion was relatively subtle in some cases when comparing at an individual time point, it was clear that significant temporal and quantitative dysregulation occurred at more than 100 genes. The expression characteristics of these genes are outlined in Fig. 4C. Seventeen of these differentially regulated genes were in the neuronal gene ontology group, showing that expression of such genes during differentiation of human SH-SY5Y is affected by CSB loss as well, which could provide a mechanism for the lack of neuronal differentiation in the absence of CSB.

Aberrant Gene Expression in Postmortem Cerebella from Patients with CS. The experiments above provide evidence for CSB-mediated gene regulation playing an important role in the differentiation and survival of neuron-like cells in culture. These results predict that gene expression changes might also be detectable in patients with CS. To investigate gene expression in the human brain, RNA was isolated from postmortem patient cerebella from confirmed CSB patients and matched controls in independent replica RNA extractions (Dataset S8). The RNA was then subjected to microarray analysis. Gratifyingly, bioinformatic analysis of the gene expression signatures showed that the six samples derived from patients with CS clustered together, separately from the non-CS control samples (Fig. S4A). Moreover, pairwise comparisons indicated that gene expression patterns are similar between patients with CS, and distinct from those gene expression patterns of the non-CS controls (representative examples are shown in Fig. S5A). Among the 23,266 genes and transcripts identified as expressed in this tissue, 1,320 genes were greater than twofold (and 4,130 were >1.5-fold) differentially expressed in the patients who had CS (Fig. 5B and Dataset S9). As observed in human fibroblasts, genes related to the nervous system were enriched among the down-regulated genes (examples are shown in Fig. 5C). We note, however, that neuronal genes were not simply dysregulated en bloc: Of the 1,832 genes in the neuronal gene ontology group [“neurogenesis/nervous system development” (GO:0007399)] that were detected as being expressed in cerebellum, only 171 were greater than twofold differentially expressed in the CS patients. Importantly, the genes affected by CSB mutation in patients overlapped significantly with those genes affected in fibroblasts and the differentiating cell lines (Fig. S5B and C).

Interestingly, genes encoding components of the protein machinery controlling regulated exocytosis, such as core proteins of synaptic vesicles and dense core granules (e.g., synaptotagmins, synaptotropin, synaptogyrin, SV2B), synaptic SNAREs (syntaxin 1A and 1B), and secreted cargoes (BDNF, WIFI, and IL-16), were among the most down-regulated (Dataset S10). A widespread impairment of regulated secretion induced by a reduction in these proteins, and the decrease in neuronal differentiation and synaptic density that would likely result from it, could potentially help
explain multiple neurodevelopmental defects observed in patients with CS. It is particularly noteworthy that genes encoding synaptotagmins, including SYT9, SYT1, and synaptotagmin-like protein 1 (SYTL1), as well as voltage-dependent calcium channels (VDCCs), were down-regulated in CS cerebella because it suggests that calcium entry in neurons and glia, and/or its detection by the exocytic apparatus, is suboptimal. The dysregulation of SYTL1 in cerebella is interesting because this synaptotagmin-like protein regulates organelle positioning and granule exocytosis in different cell types (22) and has been found to bind neurexin-1 (NRXN1) (23), a cell surface protein crucial for synaptic function and neuronal development. NRXN1 is also significantly down-regulated in CS brains (Fig. 5C).

Although we can only speculate that alterations in calcium homeostasis triggered by VDCC dysregulation may contribute to the severe cerebral calcification observed in the brain of CS patients (24), it is very likely that deficits in calcium dynamics have physiological repercussions at the level of the neurons and glia. For example, the down-regulated CACNA1E gene encodes the core subunit of type-R voltage-gated calcium channels (25), which are expressed in the cerebellum, brainstem, and telencephalon by neurons and glial cells. In particular, CACNA1E is localized on the paranodal wraps and myelin sheets of oligodendrocytes (26), which are the main myelin-producing cells in the CNS. CACNA1E channel function is important to signal myelination (26), so its down-regulation in CS cerebella is a possible cause of hypomyelination, a primary pathological feature of CS (24, 27). Similarly, the down-regulated ADAM22 gene encodes a transmembrane protein important for forming the protective myelin sheath around Schwann cells (28), which are the main myelin-forming cells in the peripheral nervous system. Overall, we note that the concurrent down-regulation of so many genes involved in regulated exocytosis is likely to signify disruption of an entire gene regulatory network encompassing these genes. In any case, our data indicate that CSB is required for transcription of a large number of genes in human cerebella, and that a significant fraction of these genes are involved in neuronal development and/or survival.

Since the first tentative connections to DNA repair and transcription were reported in the late 1970s and early 1980s (29–31), a large number of theories have been proposed for the molecular causes of CS (reviewed in ref. 2). Here, we have presented evidence from human fibroblasts, neuroblastomas, and patient postmortem cerebella that defective regulation of RNAP II genes underlies CS neurological disease. Like other Swi/Snf family chromatin remodelers, CSB is a facilitator of transcription and affects different genes in different cell types. CSB is not embryonic-lethal, so it cannot be absolutely required for expression of any individual essential gene. Instead, it “optimizes” gene expression, quantitatively affecting expression...
of thousands of genes, in a cell type-specific manner. Indeed, although statistically significant overlaps were detected between the transcriptomes we measured, and neuronal genes were always overrepresented, the genes affected by CSB mutation/deletion in fibroblasts, neuroblasts, and patient cerebella were not invariably the same. How CSB is directed to genes and why some, but not other, genes depend on it in different cell types is an important subject for future study. Recently published data (32) raise the possibility that cell type- and sequence-specific transcription factors might help recruit CSB to genes in some cases. CSB also affects transcript elongation (33), all in all suggesting that it affects gene expression at several levels.

Given that CS is caused not only by mutation in CSB but also by mutation in CSA, it might be expected that CSA mutation would affect gene expression in a manner similar to CSB mutation. Indeed, transcriptome analysis of the CS3BE [CSA-mutated (34)] fibroblast cell line shows a striking overlap between genes affected by CSB and CSA mutation (Fig. S7 A and B), with gene ontology analysis of the overlapping genes again uncovering neuronal functions (Fig. S6 C).

Interestingly, the neurological disease that characterizes patients who have CS is not recapitulated in csh−/− mice, where severe neuropathology is not observed (reviewed in ref. 2). We believe that the lack of severe neuropathology may be explained by CSB not being critical for optimal expression of the same neuronal gene networks in the mouse. Indeed, of eight tested neuronal genes that showed a marked CSB requirement in human fibroblasts, only one showed the same requirement in mouse fibroblasts (Fig. S7 A). Moreover, mouse neuroblasts in which CSB was knocked down were fully capable of differentiating and growing neurites (Fig. S7 B–D), further supporting this contention.

Finally, we note that although CS has often been described as a neurodegenerative disorder (e.g., ref. 35), a recent review argued that it is more appropriately designated a neurodevelopmental disorder (2). In light of this discussion, it is important to emphasize that the gene expression defects in CSB-deficient cells described here can support both scenarios. Indeed, although much of our work focused on the effect of CSB on neuronal differentiation/development, neuronal maintenance was also affected: Loss of CSB caused the differentiated neuron-like cells to lose their neurites gradually, and an accompanying decrease in cell viability was observed. We also note that although our data argue that gene expression defects may underlie the neurological symptoms of CS, DNA repair deficiencies might obviously also contribute to the etiology of this severe disease.

Methods

BAC-CSB recombineering was performed as described (36). For doxycycline-induced CSB expression, csb-cDNA was cloned into pTRE3G-TetON-GFP (Clontech). Lentiviral shRNAs and cDNA constructs were purchased from Thermo Scientific, whereas lentiviral constructs expressing microRNAs and neural transcription factors were from Addgene. Transdifferentiation was performed essentially as described (16, 17). Neuroblastoma SH-SYSY cells were seeded onto poly-L-lysine–coated plates. To induce their differentiation, all trans-RA (Sigma) was added in N2 medium. Antibodies used in immunofluorescence and Western blots are described in Dataset S11. For gene expression analysis, total RNA was extracted from cultured cells with the RNeasy Mini Kit (Qiagen), or from ∼100 mg of frozen brain tissue using the...
Qiagen RNeasy Lipid Tissue Mini Kit. Quantitative RT-PCR (qRT-PCR) analysis used single-stranded cDNA synthesized from a total of 200 ng of RNA using a TaqMan reverse transcription kit (Invitrogen). Primers used in qRT-PCR are listed in Dataset S12. Microarray analysis used double-stranded cDNA synthesized with a cDNA synthesis kit (NimbleGen). Single-dye labeling of this DNA, NimbleGen array hybridization, and data acquisition were performed according to the manufacturer's instructions. Microarray data were analyzed using Bioconductor version 1.9 (www.bioconductor.org) running on R version 2.8.0. ChiP-Seq analysis was performed essentially as described (37). Details are described in SI Methods.

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 (RNAPII) 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 RNAPII (a list of antibodies used is provided in Dataset S11). CSB and RNAPII 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 72 of those 975 high-confidence peaks were in the coding region of a gene (in 703 distinct genes, P value of 2.2 × 10^−10). Because protein-coding genes are rare in the genome, this high proportion strongly connects CSB with RNAPII 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^−3). As alluded to in the main text, we also identified genes that were differentially regulated in CSB-deficient cells and also had altered RNAPII density at their transcriptional start site (by ChIP-Seq). Of the 651 genes whose expression was decreased in CSB-deficient cells, 205 (31.5%) also displayed a decrease in RNAPII density. Again, this overlap is highly significant (P value for overlap of 1.2 × 10^−21).

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, because there were 9,137 CSB peaks (in 7,165 genes) with ChIP-Seq resolution. Of those 7,165 genes, 60 were also among the 1,244 genes that change expression upon CSB mutation. This overlap is more impressive in a numerical sense, because there were 9,137 CSB peaks (in 7,165 genes) with ChIP-Seq resolution.

Comment on ChIP Combined with Deep Sequencing Experiments

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% CO2]. 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 (TRCN0000095008, 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 trypsination 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 Heps (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 nucleus 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

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).

Wang et al. www.pnas.org/cgi/content/short/1412569111
1 of 11
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, phMYTTL-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 with the packaging plasmids (Gag-pool and vesicular stomatitis virus glycoprotein). Lentiviral particles were collected over 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 10^5 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 1x PBS containing 0.1% Triton X-100 and blocked with blocking solution [1x 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 fluorescent-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 10x. 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. 3D; 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 12x125K 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 2ΔΔCT 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. 2D and Fig. 6C) or a fold change of >1.5 (Figs. 1A and 4A and C and Figs. S1D and S4C) 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. 4A, 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. S5B).

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 “Differen-
tiated SH-SY5Y” in Fig. S5B).

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. S5B).

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. S5B).

The phenom 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 × 10⁴ cells). After incubation for 24 h, the cells were exposed to UV irradiation from 0 to 10 J/m², 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 × 10⁶ 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.

Fig. S1. Analysis of CS1AN cells and their WT reconstituted counterparts. (A) Western blot analysis of endogenous CSB protein in human fetal lung fibroblast MRC5 (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 the 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.
**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<sup>−/−</sup> 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). α-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 ~ 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)