

Partial rescue of neuronal genes deregulated in Cornelia de Lange Syndrome by cohesin

Felix D. Weiss^{1,\$}, Lesly Calderon^{1,\$}, Yi-Fang Wang², Radina Georgieva^{1,3}, Ya Guo¹, Nevena Cvetesic³, Maninder Kaur⁴, Gopuraja Dharmalingam², Ian D. Krantz^{4,5,6}, Boris Lenhard^{3,7}, Amanda G. Fisher¹, Matthias Merkenschlager^{1,*}

¹ Lymphocyte Development Group, Epigenetics Section, MRC London Institute of Medical Sciences, Institute of Clinical Sciences, Faculty of Medicine, Imperial College London, London, UK.

² MRC London Institute of Medical Sciences, Institute of Clinical Sciences, Faculty of Medicine, Imperial College London, London, UK.

³ Computational Regulatory Genomics Group, Integrative Biology Section, MRC London Institute of Medical Sciences, Institute of Clinical Sciences, Faculty of Medicine, Imperial College London, London, UK.

⁴ Division of Human Genetics, The Department of Pediatrics, The Children's Hospital of Philadelphia, Philadelphia, Pennsylvania, 19104, USA.

⁵ The Perelman School of Medicine at The University of Pennsylvania, Philadelphia, Pennsylvania, 19104, USA.

⁶ Department of Pathology and Laboratory Medicine, The Children's Hospital of Philadelphia, Philadelphia, Pennsylvania, 19104, USA.

⁷ Sars International Centre for Marine Molecular Biology, University of Bergen, N-5008 Bergen, Norway

Current addresses: School of Life Sciences and Biotechnology, Shanghai Jiao Tong University, Shanghai, China (YG), Research Institute of Molecular Pathology, Campus-Vienna-Biocenter 1, Vienna, Austria (LC), Summit Therapeutics, UK (GD).

\$ These authors contributed equally

* matthias.merkenschlager@lms.mrc.ac.uk

Abstract

Cornelia de Lange Syndrome (CdLS) is a human developmental disorder caused by mutations that compromise the function of cohesin, a major regulator of 3D genome organization. Cognitive impairment is a universal and as yet unexplained feature of CdLS. We characterized the transcriptional profile of cortical neurons from CdLS patients and found deregulation of hundreds of genes enriched for neuronal functions related to synaptic transmission, signalling processes, learning and behaviour. Inducible proteolytic cleavage of cohesin disrupted 3-D genome organization and transcriptional control in post-mitotic cortical mouse neurons. The genes affected belonged to similar gene ontology classes and showed significant numerical overlap with those deregulated in CdLS. Interestingly, gene expression was largely rescued by subsequent reconstitution of cohesin function. These experiments show that cohesin is continuously required for neuronal gene expression and provide a tractable approach for addressing mechanisms of neuronal dysfunction in CdLS.

Introduction

Three-dimensional (3D) genome organisation into compartments, topologically associated domains (TADs), contact domains and chromatin loops spatially compartmentalises genes and enhancers and facilitates transcriptional control by gene regulatory elements¹⁻⁴. 3D genome organisation is achieved through the activity of architectural proteins, including the cohesin complex. Initially identified as essential for chromosomal integrity during the cell cycle, cohesin is now known to cooperate with the DNA binding protein CTCF in 3D chromatin contacts essential for transcriptional control³⁻⁸. Mechanistically, cohesin increases 3D contact probabilities of sequence elements, including enhancers and promoters, within boundaries marked by CTCF binding sites in convergent orientation^{3,5,6}. In addition, a subset of promoters and enhancers are direct targets of CTCF, and genes that contact enhancers via CTCF-based cohesin loops are highly susceptible to deregulation when CTCF or cohesin are perturbed^{3,9,10}.

Mammalian genes have been classified into those that are controlled mainly by their promoters, and those that primarily depend on distal enhancers for transcriptional regulation¹¹. This difference in regulatory 'logic' broadly separates ubiquitously expressed, promoter-centric 'housekeeping' genes from enhancer-controlled tissue-specific genes¹². Dependence on distal regulatory elements may explain why enhancer-associated genes are preferentially deregulated when 3D organisation is perturbed by the loss of cohesin^{7,13}.

Heterozygous or hypomorphic germline mutations of cohesin and associated factors

such as the cohesin loading factor NIPBL are compatible with normal cell cycle progression, yet result in a human developmental disorder known as Cornelia de Lange Syndrome (CdLS)¹⁴. All CdLS patients show a degree of intellectual disability, and 60 to 65% have autism spectrum disorder (ASD), often in the absence of structural brain abnormalities or neurodegeneration¹⁴⁻¹⁶.

Experimental perturbations of neuronal cohesin expression during mouse development, some aimed at modelling CdLS, have shown changes in animal behaviour as well as abnormal neuronal morphology¹⁷, migration¹⁸ and gene expression¹⁷⁻²¹. However, these studies have been unable to discern whether these changes are the result of cohesin deficiency during development, or reflect an ongoing requirement for cohesin in mature neurons. This is of interest, as mouse models of neurodevelopmental syndromes such as Rett Syndrome can benefit from postnatal interventions that correct aberrant gene expression and animal behaviour^{22,23}. Although aberrant gene expression is a likely cause of neuronal dysfunction in CdLS, studies of gene expression in CdLS patient cells have to date been limited to induced pluripotent stem cells (iPSCs)²⁴, cardiac²⁴ and lymphoblastoid cells²⁵. We therefore do not know whether cohesin deficiency in adults perturbs neuronal gene expression and contributes to disease.

In this study we examined gene expression in neuronal nuclei isolated from post-mortem cerebral cortex of CdLS patients. We discovered prominent downregulation of hundreds of genes enriched for important neuronal functions, including synaptic transmission, signalling processes, learning and behaviour, and neuroprotective

sphingolipid metabolism. Deregulated gene expression in CdLS showed significant overlap with ASD (ref. 26). We established experimental models that allow for the inducible degradation of the cohesin subunit RAD21 in post-mitotic primary cortical mouse neurons. These revealed that cohesin is continuously required to sustain neuronal gene expression. A significant number of cohesin-dependent genes were deregulated in CdLS, and, importantly, the expression of these genes was rescued by reintroducing functional cohesin. Our findings show changes in neuronal gene expression that may contribute to neuronal dysfunction in CdLS. Deregulated expression of a subset of genes in cohesin-deficient neurons reflects an ongoing requirement for cohesin, and at least some of these changes may be reversible.

Results

Transcriptomic characterisation of CdLS patient neurons

We sourced frozen post-mortem prefrontal cortex from four CdLS patients aged 19 to 48 years and six age-matched controls (Extended Data Fig. 1a). Three of the patients had heterozygous mutations in *NIPBL*, the gene most frequently mutated in CdLS (ref.14, 27, Fig. 1a, Extended Data Fig. 1a, b). No mutations in *NIPBL* or other CdLS genes such as *RAD21*, *SMC1A* or *HDAC8* were found in the fourth patient, consistent with the lack of identifiable mutations in ~30% of CdLS patients²⁷.

To characterise gene expression in CdLS patient neurons we isolated nuclei from prefrontal cortex, stained with the neuronal marker NeuN (ref. 28), and sorted NeuN positive and NeuN negative nuclei by flow cytometry (Fig. 1b, Extended Data Fig. 1c). We generated ATAC-seq, CAGE, and RNA-seq data from neuronal nuclei. Although ATAC-seq and CAGE signals at enhancers did not show CdLS-specific features due to extensive inter-individual variation (data not shown), CAGE revealed significant deregulation of 766 gene promoters in CdLS neurons (adj $P < 0.05$; 358 upregulated, 408 downregulated; Extended Data Fig. 2a-c, Extended Data Table 1). Gene promoters that were downregulated in CdLS neurons were enriched for the gene ontology (GO) terms synaptic signalling and organization, ion and neurotransmitter transport, axon development, behavior and cognition, while upregulated promoters showed no striking functional enrichment (Extended Data Fig. 2d, Extended Data Table 2). RNA-seq identified 617 differentially expressed genes (adj $P < 0.05$) in CdLS neurons, 310 genes were up- and 307 were downregulated (Figure 1c, Extended Data Table 3). *NIPBL* RNA levels were not significantly deregulated by the *NIPBL* loss of function mutations in these patients (Extended Data Fig. 1d). In close agreement with CAGE, genes that were downregulated by RNA-seq were related to synaptic transmission, signalling processes, learning and behaviour, and sphingolipid metabolism, while upregulated genes showed no striking functional enrichment (Fig. 1d, Extended Data Table 4).

RNAseq and CAGE of CdLS neurons showed highly significant overlap (Odds ratio = 11.86, $P < 2.2\text{e-}16$), and the direction of gene deregulation by RNAseq and CAGE was highly correlated ($R = 0.68$, $P = < 2.2\text{e-}16$ for genes deregulated either by RNAseq or

CAGE, and $R = 0.95$, $P = < 2.2e-16$ for genes deregulated in both RNAseq and CAGE; Extended Data Fig. 2e).

A majority of CdLS patients have ASD in addition to intellectual disability^{14,16}. We therefore asked whether genes implicated in ASD were deregulated in CdLS. Genes found deregulated in prefrontal cortex of patients with idiopathic ASD (ref. 26) were deregulated in CdLS more frequently than expected by chance (Odds ratio = 6.25, $P < 2.2e-16$), and the direction of gene expression changes in CdLS was highly correlated with that in idiopathic ASD ($R = 0.56$, $P < 2.2e-16$ for genes deregulated in idiopathic ASD, $R = 0.85$, $P < 2.2e-16$ for genes deregulated in both idiopathic ASD and CdLS; Fig. 1e, f). Strong overlap of deregulated genes and correlated changes in the direction of gene expression were also seen between CdLS and a form of ASD caused by the duplication of chromosome 15q.11.2-13.1 (ref. 26, $R = 0.85$, $P < 2.2e-16$; Fig. 1g). This data reveal similarity between the transcriptional programs of neurons in CdLS and in idiopathic and syndromic ASD.

Deregulated genes in CdLS neurons included transcription factors, such as the ASD risk gene SOX5, neurotransmitter receptors and associated proteins for glutamatergic (*NETO1*, *NETO2*), GABAergic (*GABRA1*, *GABRG3*) and serotonergic (*HTR1A*, *HTR3B*) signalling, as well as potassium (eg: *KCNA1*, *KCNG1*, *KCNK1*), sodium (*ATP1B*), and chloride (*ANO2*) channels that control electrochemical gradients and neuronal transmission. Components of the Wnt (*FZD1*), Ephrin (*EFNA1*, *EPHA6/7*), EGF (*EGFR*), FGF (*FGF9/17*, *FGFR4*) and TGF-beta (*TGFBR1*) signalling pathways were

deregulated, as were genes involved in neuronal adhesion including protocadherins (*PCDHB9/10*), which had previously been shown to depend on cohesin and CTCF for their correct expression²⁹, and enzymes involved in the synthesis of sphingolipids with neuroprotective functions (*B3GALT2*, *ST8SIA1/2*, Fig. 1h).

RNA-seq of NeuN negative nuclei showed upregulation of immune response genes that was not seen in NeuN positive nuclei, including interferon response, JAK-STAT signalling, and the complement pathway (Extended Data Fig. 3, Extended Data Table 5, 6), suggesting that the deregulation of neuronal genes in CdLS is accompanied by activation of inflammatory genes in non-neuronal cells.

In summary, these data demonstrate that neurons from CdLS patients show deregulated expression of genes that have important neuronal functions, consistent with neuronal dysfunction in CdLS. They raise the question whether deregulated gene expression is a direct consequence of reduced cohesin function in post-mitotic neurons, or secondary to a requirement for cohesin during neural development.

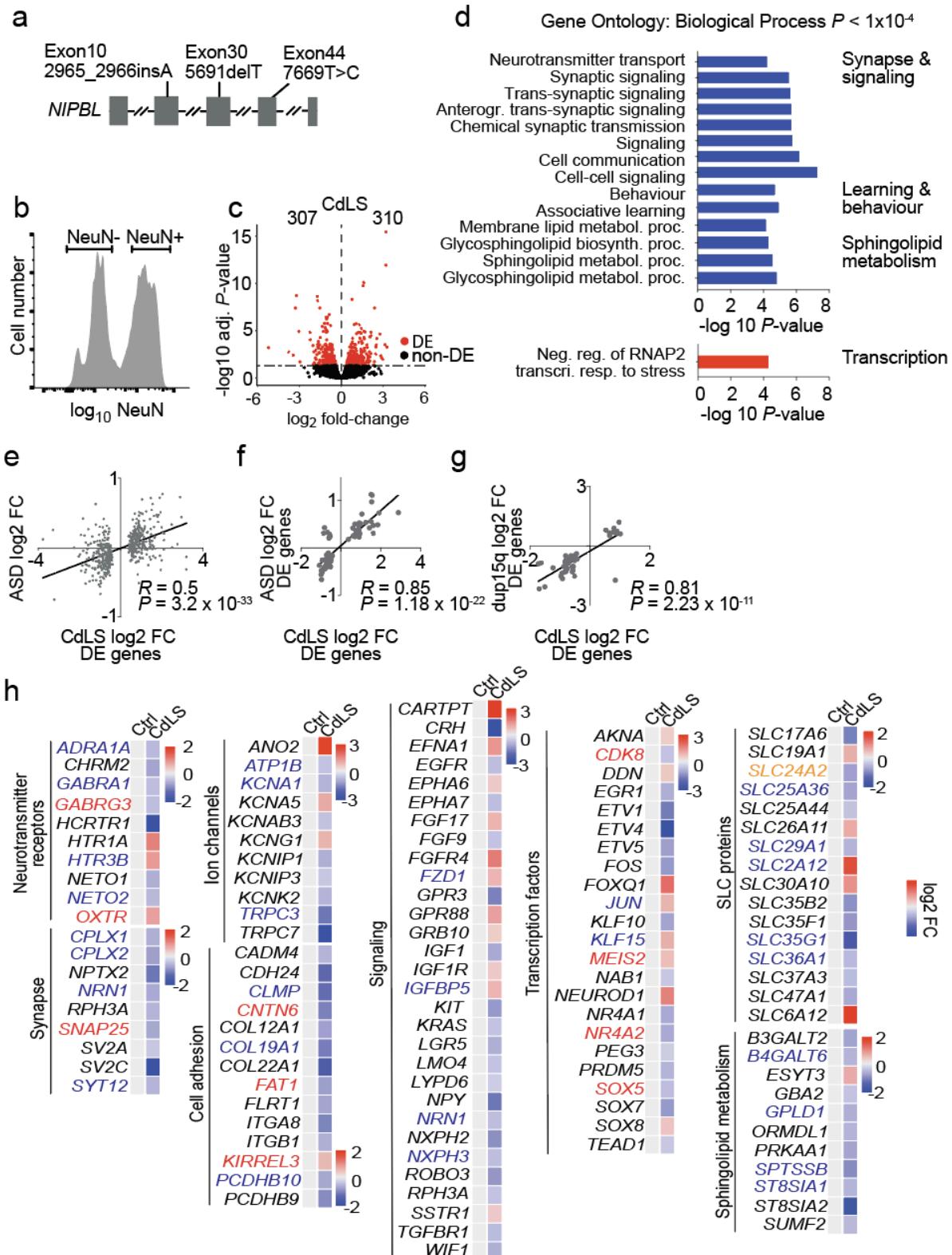


Figure 1. CdLS patients display abnormal neuronal gene expression

- a) Sanger sequencing confirms *NIPBL* mutations in CdLS patients.
- b) Flow cytometry histogram of NeuN-stained nuclei from post-mortem tissue. Sort gates for NeuN-positive and -negative nuclei are indicated.
- c) Volcano plot of gene expression fold-change versus adjusted *P*-value of NeuN-positive nuclei from CdLS patients ($n = 4$). 307 genes were down- and 310 genes were upregulated (RUVg $k=2$, adj. $P < 0.05$, shown in red, DE = differentially expressed).
- d) Bar graph of individual GO terms and broad categories. Terms represented by downregulated genes are shown in blue, those by upregulated genes in red.
- e) Scatter plot of gene expression comparing \log_2 fold-change of differentially expressed genes (adj $P < 0.05$) in idiopathic ASD, and the same genes in NeuN-positive CdLS samples. (R = Pearson correlation coefficient).
- f) Scatter plot of \log_2 fold-change of shared deregulated genes (adj $P < 0.05$) in both ASD and NeuN-positive CdLS samples. (R = Pearson correlation coefficient).
- g) Scatter plot of \log_2 fold-change of genes that were deregulated in both duplication of chromosome 15q.11.2-13.1 (dup15q) and NeuN-positive CdLS samples. (R = Pearson correlation coefficient).
- h) Heatmap of \log_2 fold-changes for deregulated genes in NeuN-positive CdLS samples. Genes are grouped by functional categories. Genes highlighted in red are in the SFARI database, those in blue were deregulated in ASD and/or dup15q, genes highlighted in orange fulfil both criteria.

Temporal control over cohesin levels in neurons

Addressing the question of whether cohesin is directly required for neuronal gene expression requires experimental systems that enable acute cohesin withdrawal from post-mitotic neurons, ideally in a reversible manner. To control the expression of

RAD21, a cohesin subunit associated with CdLS (ref. 14), we utilised $Rad21^{Tev}$, an allele that encodes RAD21 protein cleavable by tobacco etch virus (TEV) protease^{30,31}. $Rad21^{Tev/Tev}$ neurons contain endogenous RAD21-TEV as their sole source of RAD21. We established explant cultures from $Rad21^{Tev/Tev}$ cortices at embryonic day 14.5 (e14.5) under conditions that enrich for post-mitotic neurons and promote neuronal maturation (Fig. 2a, Extended Data Fig. 4a, b).

We used lentivirus to transduce $Rad21^{Tev/Tev}$ neurons with TEV protease fused to a tamoxifen-responsive estrogen receptor hormone binding domain (ERt2-TEV). Constructs were tagged by V5 and t2a-fused GFP (ERt2-TEV) for identification of ERt2-TEV-expressing cells (Fig. 2b, Extended Data Fig. 4c). ERt2-TEV remained cytoplasmic until nuclear translocation of ERt2-TEV was triggered, which occurred within minutes of 4-hydroxy tamoxifen (4-OHT) addition (Fig. 2b, Extended Data Fig. 4c). We monitored the cleavage of RAD21-TEV by western blotting and found that RAD21-TEV expression was reduced to ~15% of pre-treatment levels within 8h of 4-OHT addition, and remained at this level for at least 24h (Fig. 2c). In a second approach, we transduced $Rad21^{Tev/Tev}$ neurons with a lentiviral construct that constitutively expresses Tet-On advanced transactivator (rtTA) and RFP. Addition of doxycycline leads to the expression of TEV protease with an exogenous nuclear localisation sequence (NLS-TEV, Extended Data Fig. 5a, b). Induction of NLS-TEV depleted ~70% of RAD21-TEV within 24 h of doxycycline addition (Extended Data Fig. 5c). These systems establish temporal control over cohesin levels in primary post-mitotic neurons.

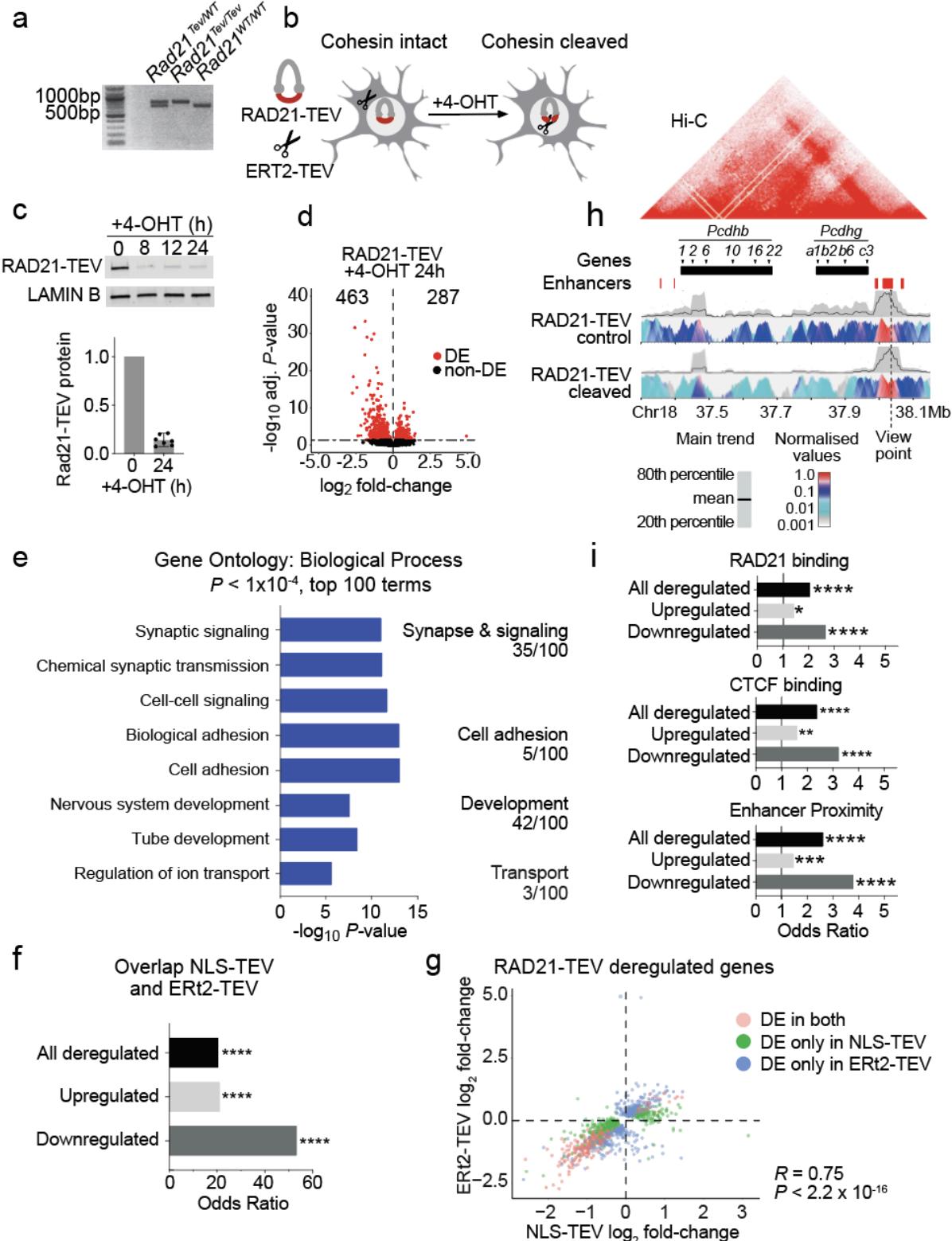


Figure 2. Cohesin is continuously required to sustain neuronal gene expression

- a) PCR analysis of *Rad21* alleles from *Rad21^{Tev/WT}*, *Rad21^{Tev/Tev}* and *Rad21^{WT/WT}* mice.
- b) Schematic of ERt2-TEV-dependent RAD21-TEV degradation
- c) Western blot of RAD21-TEV protein expression over a time course of 4-OHT treatment. Bar plot of RAD21-TEV protein expression normalised to LAMIN B (n = 7, h = hours).
- d) Volcano plot of gene expression fold-change versus adjusted *P* value in RNA-seq of RAD21-TEV neurons transduced with ERt2-TEV and treated with 4-OHT for 24h (n = 3). 463 genes were down- and 287 genes were upregulated (adj *P* < 0.05 shown in red, DE = differentially expressed).
- e) Bar graph of individual GO terms and broad categories in RNA-seq of RAD21-TEV neurons treated as in d). Terms represented by downregulated genes are shown in blue. Upregulated genes showed no GO term enrichment at *P* < 1 x 10⁻⁴.
- f) Enrichment of shared deregulated genes (adj *P* < 0.05) in response to acute cohesin depletion induced by ERt2-TEV and NLS-TEV. Fisher's exact test was applied for the odds ratio and *P* value.
- g) Scatter plot of gene expression, comparing log₂ fold-change of deregulated genes (adj *P* < 0.05) in response to acute cohesin depletion induced by ERt2-TEV and NLS-TEV (DE = differentially expressed, *R* = Pearson correlation coefficient).
- h) 4C analysis of chromatin interactions at the *Pcdhb* locus. Top: Hi-C representation of domain structure at the *Pcdhb* locus in mouse cortical neurons³⁸, with selected genes and enhancers. Bottom: contact profiles of enhancers HS18-20 in control cells (top panel) and RAD21-TEV cleaved cells (bottom panel). A dashed line indicates the enhancer site and viewpoint. A grey band displays the 20th to 80th percentiles and the black line within shows mean values for 40kb windows. The colour panel shows the mean contact intensities for multiple window sizes from 2kb to 5kb, n = 4 independent biological replicates.
- i) Genes significantly deregulated (adj *P* < 0.05) by ERt2-TEV mediated RAD21-TEV degradation are enriched for the binding of cohesin (top) CTCF (center) and proximity to neuronal enhancers (bottom, see methods). Fisher's exact test was applied for the odds ratio and *P* value. * *P* < 0.05, ** *P* < 0.01, *** *P* < 0.001, **** *P* < 0.0001.

Cohesin is continuously required for neuronal gene expression

We utilized temporal control of RAD21-TEV cleavage to quantify the impact of acute cohesin depletion on gene expression in post-mitotic neurons. RNA-seq 24h after 4-OHT-induced nuclear translocation of ERt2-TEV identified 750 deregulated (adj $P < 0.05$) genes, of which 463 were down- and 287 were upregulated (Fig. 2d, Extended Data Table 7). GO term analysis showed that downregulated genes were enriched for biological processes similar to those observed for downregulated genes in CdLS, in particular synapse and signalling, cell adhesion, neuron development and ion transport. As observed in CdLS neurons, no striking enrichment was observed for upregulated genes (Fig. 2e, Extended Data Table 8). Genes deregulated by ERt2-TEV-mediated RAD21-TEV cleavage showed significant overlap with SFARI ASD risk genes³² (Odds ratio = 2.19, $P = 1.16\text{e-}08$). There was limited overlap between genes deregulated by acute cohesin depletion and other mouse models of neuronal dysfunction³³⁻³⁷, with the exception of gene expression in embryonic *Nipbl^{fl/fl}* brain¹⁴ (Extended Data Fig. 6).

RNA-seq 24h after doxycycline-dependent NLS-TEV induction in *Rad21^{TeV/TeV}* neurons identified 570 down- and 150 upregulated genes (adj $P < 0.05$). As observed for the nuclear translocation of ERt2-TEV, downregulated genes were enriched for adhesion, signalling, synaptic function and development (Extended Data Fig. 5d, e, Extended Data Table 9, 10). There was strong overlap for genes deregulated in both systems (Odds ratio = 20.35, $P < 2.2\text{e-}16$), especially for downregulated genes (Odds ratio = 53.22, $P < 2.2\text{e-}16$, Fig. 2f), and the direction of gene deregulation was highly concordant ($R = 14$

0.75, $P < 2.2\text{e-}16$, Fig. 2g). Hence, acute depletion of RAD21-TEV in post-mitotic neurons established that cohesin is continuously required to sustain the expression of genes that mediate important neuronal functions.

The *Pcdhb* cluster is a prime example for CTCF-based cohesin-mediated enhancer-promoter connections^{3,29}, and 15 of 17 expressed *Pcdhb* genes were deregulated by inducible RAD21-TEV cleavage. Circular chromosomal conformation capture and sequencing (4C-seq) using the *Pcdhb* enhancers HS18-20 as the viewpoint showed that RAD21-TEV cleavage disrupted 3D contacts, and decreased interactions between *Pcdhb* promoters and enhancers (Fig. 2h). Genes that were downregulated in response to acute cohesin depletion in primary neurons were significantly enriched for the binding of RAD21 (Fig. 2i, Odds Ratio = 2.66, $P = 2.32\text{e-}13$) and CTCF (Fig. 2i, Odds Ratio = 2.38, $P = 3.39\text{e-}11$). Our finding that cohesin deficiency predominantly affects genes related to specialised neuronal functions is consistent with models that distinguish ubiquitously expressed, promoter-centric 'housekeeping' genes from enhancer-controlled tissue-specific genes^{11,12}, and with previous observations that enhancer-associated genes are preferentially deregulated by depletion of cohesin^{7,13}. Accordingly, genes that were downregulated in response to acute cohesin depletion in primary neurons were significantly enriched for proximity to enhancers (Fig. 2i, Odds ratio = 3.78, $P < 2.2\text{e-}16$). These data suggest a mechanism by which RAD21-TEV cleavage disrupts cohesin-mediated chromatin contacts, including enhancer-promoter interactions.

Cohesin-dependent gene regulation in CdLS

We next compared cohesin-dependent genes identified in *Rad21*^{Tev/Tev} neurons with gene expression in CdLS patient neurons. Although gene set enrichment analysis (GSEA) showed no concordance between the two conditions for upregulated genes (NES = 1.28, FDR = 0.13 and NES = +1.04, FDR = 0.35, respectively, data not shown), there was remarkable concordance for downregulated genes. Genes that were downregulated in *Rad21*^{Tev/Tev} neurons 24 hours after ERt2-TEV-induced cohesin depletion showed, as a set, significant downregulation also in CdLS patient neurons (NES = -2.15, FDR = 5.6×10^{-4} , Fig. 3a). The reciprocal test yielded the same result; the set of genes downregulated in CdLS were overall downregulated in *Rad21*^{Tev/Tev} neurons (NES = -2.06, FDR = 0.00, Fig. 3a).

Accordingly, a significant number of genes were deregulated both in CdLS patient neurons and after acute cohesin depletion of mouse neurons (51 in CdLS and ERt2-TEV, 57 in CdLS and NLS-TEV, 81 shared overall, Odds ratio = 2.48, $P = 4.87e-11$, of which 55 were downregulated, Odds ratio = 3.14, $P = 4.06e-11$). These shared genes were related to signaling, neurotransmitter and synapse function, ion channels, and transcription factors (Fig. 3b). The majority of shared genes were deregulated in the same direction (40 of 51 for ERt2-TEV, 42 of 57 for NLS-TEV, Fig. 3b).

There was slight, but statistically significant enrichment for binding of RAD21 and CTCF to genes that were downregulated in CdLS neurons (Fig. 3c). This enrichment increased considerably for genes that were deregulated both in CdLS and by the inducible cleavage of RAD21-TEV (compare Odds ratios in Fig. 3c). Genes that were deregulated both in CdLS and in acutely cohesin-depleted *Rad21*^{Tev/Tev} neurons showed significant enrichment for proximity to human neuronal enhancers⁴⁰ (Fig. 3c).

Gene expression changes in Rett syndrome, but not other human neurological diseases^{34,39-41}, showed significant overlap with *Rad21*^{Tev/Tev} neurons (Odds ratio = 2.06, $P = 4.64\text{e-}10$; Extended Data Fig. 7).

In summary, a subset of genes deregulated in CdLS neurons bore hallmarks of cohesin target genes, and these genes were enriched in the overlap between CdLS and cohesin-dependent genes, as identified by acute depletion of cohesin in *Rad21*^{Tev/Tev} neurons.

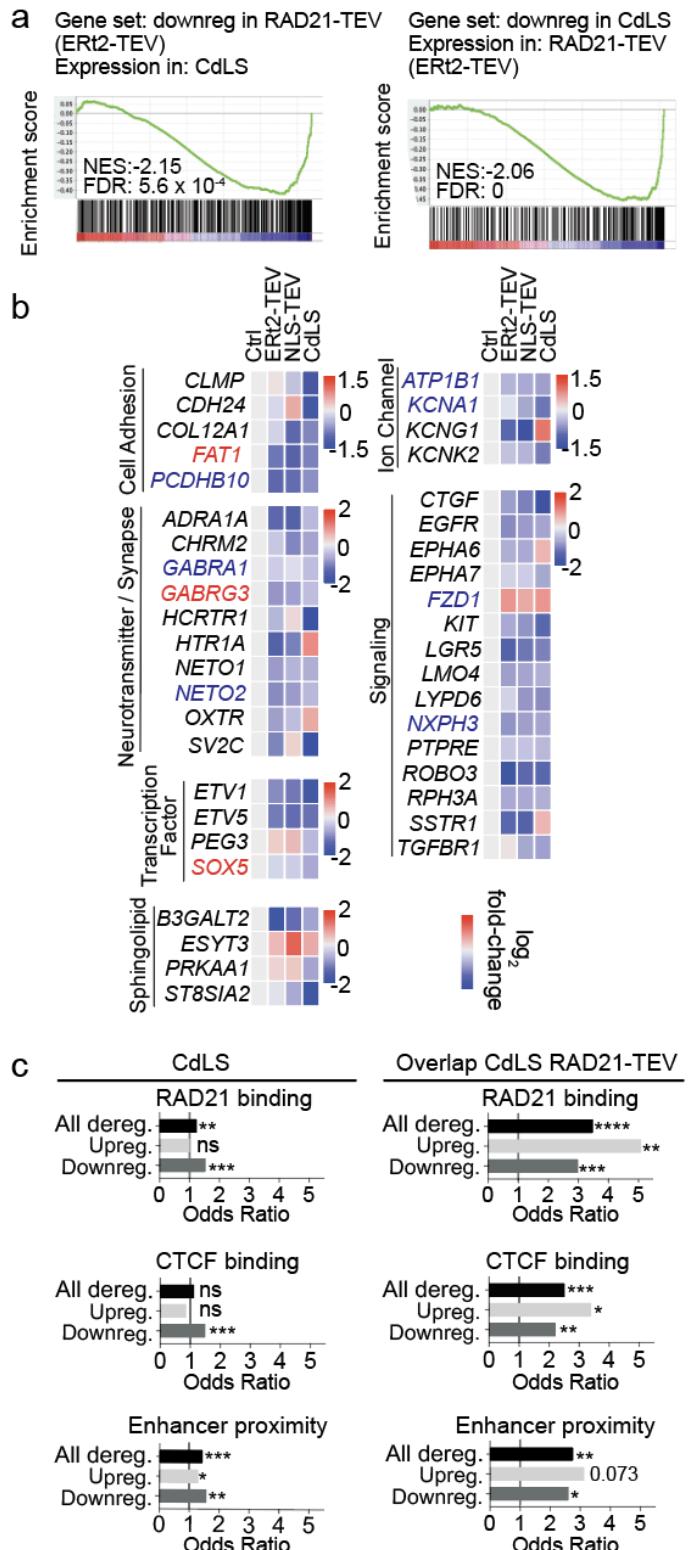


Figure 3. Cohesin-dependent gene regulation in CdLS neurons

a) GSEA of RAD21-TEV downregulated genes in CdLS NeuN-positive RNAseq (left). GSEA of CdLS NeuN-positive downregulated genes in RAD21-TEV (right). NES = normalised enrichment score, FDR = false discovery rate.

b) Heatmap of \log_2 fold-change for selected genes from CdLS NeuN-positive and RAD21-TEV samples separated into broad functional categories. Genes shown were significantly deregulated in CdLS and their mouse orthologs deregulated in at least one RAD21-TEV (ERt2-TEV or NLS-TEV) system. Genes highlighted in red were identified in the SFARI database, those in blue were deregulated in ASD, genes highlighted in orange fulfil both criteria.

c) Comparison of RAD21 (cohesin) binding (top), CTCF binding (center) and enhancer proximity (bottom, see methods) of genes deregulated (adj $P < 0.05$) in NeuN-positive CdLS (left) compared to genes deregulated in both NeuN-positive CdLS and RAD21-TEV (right). Up- and downregulation indicates the direction of deregulation in CdLS NeuN-positive RNAseq. * $P < 0.05$, ** $P < 0.01$, *** $P < 0.001$, **** $P < 0.0001$.

Rescue of cohesin-dependent gene expression

Postnatal intervention can alleviate transcriptional changes and abnormal behaviours in mouse models of Rett Syndrome^{22,23}. In light of these results, we asked whether gene expression changes induced by acute cohesin depletion in *Rad21^{Tev/Tev}* neurons can be reversed when cohesin levels are restored. Doxycycline-induced NLS-TEV expression was reversible, and transient RAD21-TEV cleavage was followed by a return of RAD21-TEV protein to control levels after removal of doxycycline (Fig. 4a, b). Remarkably, comparison of RNA-seq during RAD21-TEV depletion (24h) and after RAD21-TEV rescue (day 7) showed that the vast majority of genes that were found deregulated after 24h (695 of 720 or 96.5%) were fully rescued upon restoration of cohesin levels (Fig. 4c, e; Extended Data Table 11). Interestingly, all 57 genes that were deregulated both in CdLS and by NLS-TEV-mediated cohesin depletion were rescued by restoring cohesin expression. Rescue was near-complete, regardless of the direction and the degree of initial deregulation (Fig. 4d). A small number of residual genes were refractory to rescue, and remained deregulated. These included neuronal signalling (*Gfra1*, *Spon1*, *Scg2*, *Cntnap3*), transcription (*Mycl1*, *Cited2*, *Maml3*, *Lzts1*) and splicing factors (*Rbfox1*).

We identified 62 genes that, although unaffected by acute cohesin depletion at 24 h, were deregulated 7 days later, when cohesin levels were restored to control levels (Fig. 4e). These *de novo* deregulated genes were associated with development and cell communication and were highly enriched for genes that change expression during the maturation of control neurons in explant culture (day 10 to 17, adj $P < 0.05$). Of 62 *de*

novo deregulated genes, 46 were maturation genes (Odds ratio = 6.69, P = 1.14e-12).

Strikingly, the direction of deregulation of these genes was highly correlated with their regulation during neuronal maturation: genes that were upregulated during neuronal maturation in explant culture were also preferentially upregulated as a consequence of transient cohesin depletion ($P = 0.0134$, FDR = 0.0134), while genes that were downregulated during neuronal maturation in explant culture were also preferentially downregulated as a consequence of transient cohesin depletion ($P = 1.51e-04$, FDR = 3.02e-04) (Fig. 4f, g). Hence, while the majority of gene expression changes caused by loss of cohesin can be rescued by restoring cohesin levels, including the expression of genes deregulated in CdLS, the *de novo* deregulation of neuronal maturation genes illustrates that even transient cohesin depletion may have potentially damaging long-term secondary effects on neuronal gene expression.

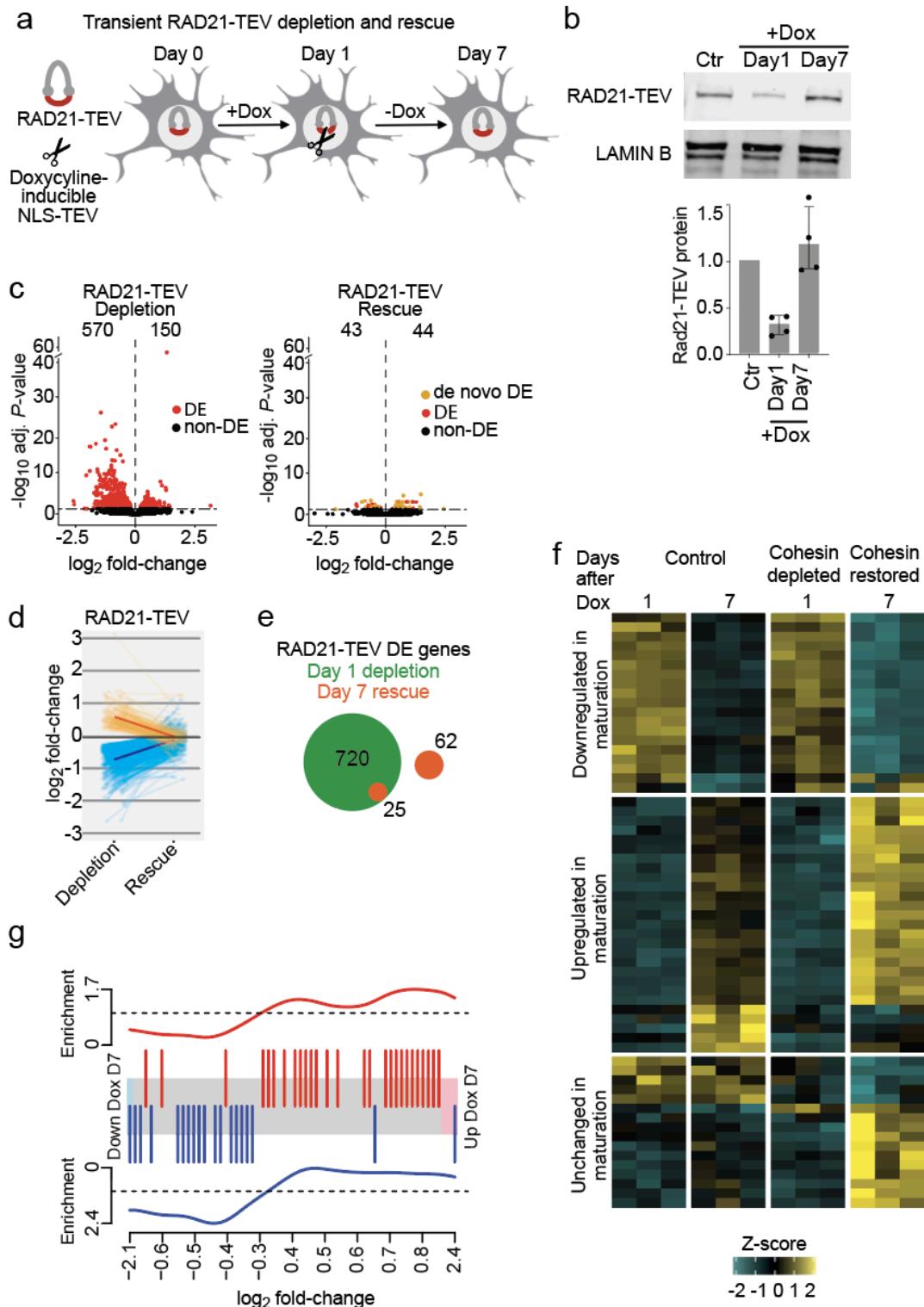


Figure 4. Rescue of cohesin-dependent gene expression

- a) Schematic of NLS-TEV mediated RAD21-TEV degradation and rescue.
- b) Western blot of RAD21-TEV protein expression over a time course of Dox treatment (6h, 100ng/ml). Bar plot of RAD21-TEV protein expression normalised to LAMIN B (n = 4).
- c) Volcano plots of gene expression \log_2 fold-change versus adjusted P-value in RAD21-TEV +Dox treated neurons before and after rescue of RAD21-TEV expression (n = 3). Left: 570 genes were down- and 150 genes were upregulated (adj P < 0.05, shown in red) following RAD21-TEV depletion. Right: 43 genes were down- and 44 genes were upregulated after rescue of RAD21-TEV expression (adj P < 0.05, shown in red and orange). Genes in red are significantly deregulated in both RAD21-TEV depletion and rescue, genes in orange are significantly deregulated only after rescue (DE = differentially expressed).
- d) \log_2 fold-change of significantly deregulated genes (adj P < 0.05) following RAD21-TEV depletion (left) and after rescue (right).
- e) The number of significantly deregulated genes (adj P < 0.05) after RAD21-TEV depletion (in green) and after rescue of RAD21-TEV expression (in orange). DE = differentially expressed.
- f) Heatmap of *de novo* deregulated genes following RAD21-TEV rescue (adj P < 0.05) and their maturation trajectory in control (left) and +Dox treated samples (right).
- g) Barcode plot of *de novo* deregulated genes following RAD21-TEV rescue (adj P < 0.05) and their enrichment for directionality in maturation. Top: significantly *de novo* upregulated genes following rescue are enriched for genes upregulated during neuronal maturation. Bottom: *de novo* downregulated genes after cohesin rescue are enriched for genes that are downregulated during neuronal maturation.

Discussion

We characterized gene expression in primary cortical neurons from CdLS patients and found deregulation of hundreds of genes enriched for important neuronal functions, including synaptic transmission, signalling processes, learning and behaviour, and sphingolipid metabolism. These findings provide experimental support for suggestions that CdLS pathologies, including intellectual disability, may be mediated at least in part by deregulated gene expression¹⁹. The transcriptomic profile of CdLS showed extensive similarities to ASD, both in terms of shared deregulated genes and in the direction of deregulation. This similarity is of particular interest given the high prevalence of ASD in CdLS patients.

We established in vitro models for the acute depletion of cohesin in primary cortical mouse neurons to address the extent to which changes in CdLS gene regulation reflect an ongoing requirement for cohesin in neuronal gene expression. Proteolytic cleavage of the essential cohesin subunit RAD21 disrupted 3D organization and perturbed the expression of neuronal genes. These experiments establish that cohesin is continuously required to sustain neuronal gene expression. Significant overlap with the CdLS transcriptome identified putative cohesin target genes in CdLS. These cohesin-dependent genes were enriched for cohesin binding and enhancer proximity in human neurons. Tissue-specific genes primarily depend on distal enhancers for transcriptional regulation^{11,12}, and enhancer-associated genes are preferentially deregulated when 3D organisation is perturbed by the loss of cohesin^{7,13}. Our finding that cohesin deficiency

predominantly affects enhancer-associated genes related to specialised neuronal functions is consistent with these models. Our analysis supports the idea that deregulation both in CdLS *in vivo* and in novel *in vitro* models for acute cohesin cleavage identifies a core set of cohesin-dependent genes that are enriched for binding of CTCF and cohesin, and that show enhancer-dependent tissue-specific expression. These genes are susceptible to deregulation because they rely on 3D contacts with distal regulatory elements, and their 3D contacts are directly dependent on cohesin.

This core set comprised genes related to signaling, neurotransmitter and synapse components, ion channels, and transcription factors. Interestingly, the great majority of gene expression changes precipitated by acute cohesin depletion was reversible upon restoration of cohesin function, including all 57 genes deregulated in both CdLS and NLS-TEV. Hence, cohesin rescues neuronal genes that are deregulated in CdLS.

In addition to directly cohesin-dependent genes, reduced cohesin function in CdLS neurons also affected genes that were not sensitive to acute cohesin depletion in mouse neurons. These lacked hallmarks of direct cohesin targets. This finding highlighted that cohesin depletion may have knock-on effects on the expression of other genes. Our data point to two mechanisms that may contribute to the indirect deregulation of genes in cohesin-deficient neurons. First, numerous genes related to neuronal maturation were deregulated late in response to acute cohesin depletion, and remained deregulated after cohesin levels were restored. These genes showed a striking pattern of overshoot regulation, where genes upregulated during the maturation

of control neurons were even more highly expressed after transient cohesin depletion and vice versa. Explant cultures contain >90% neurons, and the deregulation of these genes may therefore be driven by factors within, or interactions between neurons. Second, analysis of NeuN negative nuclei from CdLS cortices showed highly significant upregulation of inflammatory genes, consistent with potentially damaging inflammatory responses by glial cells⁴². These data identify an inflammatory component to disrupted gene expression in CdLS, which may contribute to neuronal dysfunction via glia-neuron interactions. Interestingly, one of the gene ontologies found deregulated in CdLS neurons, sphingolipid metabolism, is linked to neuroinflammation⁴³. With a view towards therapeutic strategies for addressing neuronal dysfunction in CdLS, it will be important to investigate whether the expression of indirectly deregulated and inflammatory genes can be rescued, similarly to immediate cohesin targets.

Our data raise the possibility that aspects of the CdLS phenotype may be reversible by postnatal intervention. Precedent for such rescue comes from pioneering studies in mouse models of Rett syndrome, where postnatal correction of *Mecp2* deficiency ameliorates deregulated gene expression and behavioural defects^{22,23}. As the Rett protein MECP2, cohesin exerts its effects on transcription by interfacing with chromatin marking systems⁴⁴, and the dynamics of cohesin on chromatin is controlled in part by acetylation and deacetylation of its SMC subunits⁴⁵. The acetylation/deacetylation cycle is a prime target for pharmacological intervention, as are bromodomain proteins such as BRD4, which connect acetylation to transcription⁴⁶. Mutations in the histone deacetylase HDAC8 and the bromodomain protein BRD4 can cause CdLS-like disease^{45,47}.

Epigenetic drugs^{46,48} may therefore provide alternatives to gene replacement approaches as therapeutics in CdLS.

The data and experimental systems presented here provide a valuable resource for understanding cognitive symptoms in CdLS, an important benchmark for CdLS disease models, and tractable approaches for addressing mechanisms of neuronal dysfunction in CdLS.

Acknowledgements

We thank G. Little, M. Clements, and S. Parrinello, for advice and practical instruction, S. Di Giovanni and A. Schaefer (Icahn School of Medicine, NY) for discussion and comments on the manuscript, B. Patel, J. Elliott, and the LMS/NIHR Imperial Biomedical Research Centre Flow Cytometry Facility for cell sorting, D. Gomez-Cabrero for advice on batch correction of RNA sequencing data, K. Tachibana (Institute of Molecular Biotechnology, Vienna) for providing the NLS-TEV vector, and A. Smith for advice on preparation and sorting of neuronal nuclei. Human brain tissues, and donor-associated de-identified demographic, clinical and phenotypic data used in this study were obtained from the NIH NeuroBioBank. All specimens were collected under the approval of Institutional Review Boards. NIH NeuroBioBank's complete inventory is available at <https://neurobiobank.nih.gov>. The authors thank and acknowledge the Edinburgh Brain and Tissue Bank for providing the tissues used in this study. The Edinburgh Brain Bank is supported by the MRC. Ethical approval was by the NHS-RECSE (16/ES/0084).

This work was funded by the Medical Research Council UK, Wellcome (Investigator Award 099276/Z/12/Z to MM), EMBO and HFSP (HFSP LT00427/2013, EMBO ALTF 1047-2012 TO LC).

Author contributions

FDW, LD, YG, NC and MK did experiments, FDW, LD, YG, Y-FW, RG, YG, NC, MK and GD analysed and curated data, FDW, LD, RG, MK and MM visualised data, FDW, IDK, BL, AGF and MM wrote the paper

Competing interests statement

The authors declare that they have no competing interests

Data availability.

RNA-seq and ATAC-seq data generated for this study have been deposited at GEO under accession number GSE150130.

CAGE data generated for this study has been deposited at ArrayExpress under accession number E-MTAB-9045.

References

1. Bickmore, W. & van Steensel, B. Genome Architecture: Domain Organization of Interphase Chromosomes. *Cell* 152, 1270-1284 (2013).
2. Dekker, J. & Mirny, L. The 3D Genome as Moderator of Chromosomal Communication. *Cell* 164, 1110-1121 (2016).
3. Merkenschlager, M. & Nora, E. CTCF and Cohesin in Genome Folding and Transcriptional Gene Regulation. *Annual Review of Genomics and Human Genetics* 17, 17-43 (2016).
4. McCord, R., Kaplan, N. & Giorgetti, L. Chromosome Conformation Capture and Beyond: Toward an Integrative View of Chromosome Structure and Function. *Molecular Cell* 77, 688-708 (2020).
5. Fudenberg, G. et al. Formation of Chromosomal Domains by Loop Extrusion. *Cell Reports* 15, 2038-2049 (2016).
6. Rao, S. et al. A 3D Map of the Human Genome at Kilobase Resolution Reveals Principles of Chromatin Looping. *Cell* 159, 1665-1680 (2014).
7. Rao, S. et al. Cohesin Loss Eliminates All Loop Domains. *Cell* 171, 305-320.e24 (2017).
8. Schwarzer, W. et al. Two independent modes of chromatin organization revealed by cohesin removal. *Nature* 551, 51-56 (2017).
9. Hyle, J. et al. Acute depletion of CTCF directly affects MYC regulation through loss of enhancer–promoter looping. *Nucleic Acids Research* 47, 6699-6713 (2019).
10. Nora, E. et al. Targeted Degradation of CTCF Decouples Local Insulation of Chromosome Domains from Genomic Compartmentalization. *Cell* 169, 930-944.e22 (2017).
11. Yoshida, H. et al. The cis-Regulatory Atlas of the Mouse Immune System. *Cell* 176, 897-912.e20 (2019).
12. Osterwalder, M. et al. Enhancer redundancy provides phenotypic robustness in mammalian development. *Nature* 554, 239-243 (2018).
13. Ing-Simmons, E. et al. Spatial enhancer clustering and regulation of enhancer-proximal genes by cohesin. *Genome Research* 25, 504-513 (2015).
14. Deardorff, M. A., Noon, S. E. & Krantz, I. D. Cornelia de Lange Syndrome. In GeneReviews® (eds. Adam, M. P. et al.) (University of Washington, Seattle, 2018).

15. Schrier, S. et al. Causes of death and autopsy findings in a large study cohort of individuals with Cornelia de Lange syndrome and review of the literature. *American Journal of Medical Genetics Part A* 155, 3007-3024 (2011).
16. Moss, J., Howlin, P., Magiati, I. & Oliver, C. Characteristics of autism spectrum disorder in Cornelia de Lange syndrome. *Journal of Child Psychology and Psychiatry* 53, 883-891 (2012).
17. Fujita, Y. et al. Decreased cohesin in the brain leads to defective synapse development and anxiety-related behavior. *Journal of Experimental Medicine* 214, 1431-1452 (2017).
18. van den Berg, D. et al. Nipbl Interacts with Zfp609 and the Integrator Complex to Regulate Cortical Neuron Migration. *Neuron* 93, 348-361 (2017).
19. Kawauchi, S. et al. Multiple Organ System Defects and Transcriptional Dysregulation in the Nipbl+/- Mouse, a Model of Cornelia de Lange Syndrome. *PLoS Genetics* 5, e1000650 (2009).
20. Remeseiro, S., Cuadrado, A., Gómez-López, G., Pisano, D. & Losada, A. A unique role of cohesin-SA1 in gene regulation and development. *The EMBO Journal* 31, 2090-2102 (2012).
21. Bottai, D. et al. Modeling Cornelia de Lange syndrome in vitro and in vivo reveals a role for cohesin complex in neuronal survival and differentiation. *Human Molecular Genetics* 28, 64-73 (2018).
22. Guy, J., Gan, J., Selfridge, J., Cobb, S. & Bird, A. Reversal of Neurological Defects in a Mouse Model of Rett Syndrome. *Science* 315, 1143-1147 (2007).
23. Tillotson, R. et al. Radically truncated MeCP2 rescues Rett syndrome-like neurological defects. *Nature* 550, 398-401 (2017).
24. Mills, J. et al. NIPBL+/- haploinsufficiency reveals a constellation of transcriptome disruptions in the pluripotent and cardiac states. *Scientific Reports* 8, (2018).
25. Liu, J. et al. Transcriptional Dysregulation in NIPBL and Cohesin Mutant Human Cells. *PLoS Biology* 7, e1000119 (2009).
26. Parikshak, N. et al. Genome-wide changes in lncRNA, splicing, and regional gene expression patterns in autism. *Nature* 540, 423-427 (2016).
27. Braunholz, D. et al. Hidden Mutations in Cornelia de Lange Syndrome Limitations of Sanger Sequencing in Molecular Diagnostics. *Human Mutation* 36, 279-280 (2015).

28. Mullen RJ, Buck CR, Smith AM. NeuN, a neuronal specific nuclear protein in vertebrates. *Development* 116, 201-11 (1992).
29. Guo, Y. et al. CRISPR Inversion of CTCF Sites Alters Genome Topology and Enhancer/Promoter Function. *Cell* 162, 900-910 (2015).
30. Tachibana-Konwalski, K. et al. Rec8-containing cohesin maintains bivalents without turnover during the growing phase of mouse oocytes. *Genes & Development* 24, 2505-2516 (2010).
31. Cuartero, S. et al. Control of inducible gene expression links cohesin to hematopoietic progenitor self-renewal and differentiation. *Nature Immunology* 19, 932-941 (2018).
32. Banerjee-Basu, S. & Packer, A. SFARI Gene: an evolving database for the autism research community. *Disease Models & Mechanisms* 3, 133-135 (2010).
33. von Schimmelmann, M. et al. Polycomb repressive complex 2 (PRC2) silences genes responsible for neurodegeneration. *Nature Neuroscience* 19, 1321-1330 (2016).
34. Rental, W. et al. Characterization of human mosaic Rett syndrome brain tissue by single-nucleus RNA sequencing. *Nature Neuroscience* 21, 1670-1679 (2018).
35. Korb, E. et al. Excess Translation of Epigenetic Regulators Contributes to Fragile X Syndrome and Is Alleviated by Brd4 Inhibition. *Cell* 170, 1209-1223.e20 (2017).
36. Langfelder, P. et al. Integrated genomics and proteomics define huntingtin CAG length-dependent networks in mice. *Nature Neuroscience* 19, 623-633 (2016).
37. Tilot, A. et al. Neural transcriptome of constitutional Pten dysfunction in mice and its relevance to human idiopathic autism spectrum disorder. *Molecular Psychiatry* 21, 118-125 (2015).
38. Bonev, B. et al. Multiscale 3D Genome Rewiring during Mouse Neural Development. *Cell* 171, 557-572.e24 (2017).
39. Labadorf, A. et al. RNA Sequence Analysis of Human Huntington Disease Brain Reveals an Extensive Increase in Inflammatory and Developmental Gene Expression. *PLOS ONE* 10, e0143563 (2015).
40. Mathys, H. et al. Single-cell transcriptomic analysis of Alzheimer's disease. *Nature* 570, 332-337 (2019).
41. Olmos-Serrano, J. et al. Down Syndrome Developmental Brain Transcriptome Reveals Defective Oligodendrocyte Differentiation and Myelination. *Neuron* 89, 1208-1222 (2016).

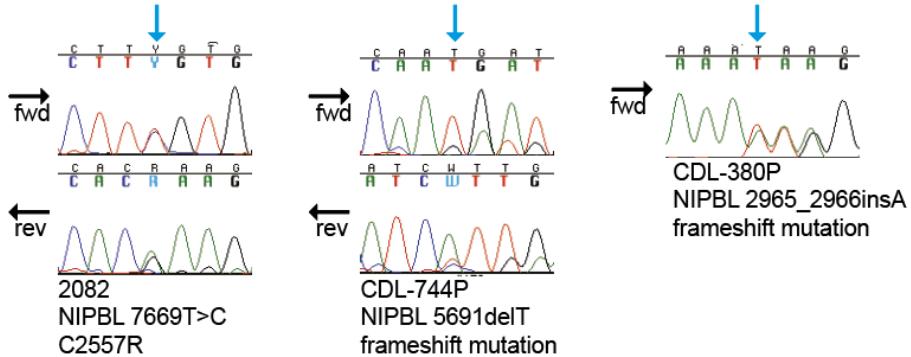
42. Chitnis, T. & Weiner, H. CNS inflammation and neurodegeneration. *Journal of Clinical Investigation* 127, 3577-3587 (2017).
43. Maceyka, M. & Spiegel, S. Sphingolipid metabolites in inflammatory disease. *Nature* 510, 58-67 (2014).
44. Yan, J. et al. Transcription Factor Binding in Human Cells Occurs in Dense Clusters Formed around Cohesin Anchor Sites. *Cell* 154, 801-813 (2013).
45. Deardorff, M. et al. HDAC8 mutations in Cornelia de Lange syndrome affect the cohesin acetylation cycle. *Nature* 489, 313-317 (2012).
46. Tough, D., Tak, P., Tarakhovsky, A. & Prinjha, R. Epigenetic drug discovery: breaking through the immune barrier. *Nature Reviews Drug Discovery* 15, 835-853 (2016).
47. Olley, G. et al. BRD4 interacts with NIPBL and BRD4 is mutated in a Cornelia de Lange–like syndrome. *Nature Genetics* 50, 329-332 (2018).
48. Egger, G., Liang, G., Aparicio, A. & Jones, P. Epigenetics in human disease and prospects for epigenetic therapy. *Nature* 429, 457-463 (2004).

Extended data Figures and Legends

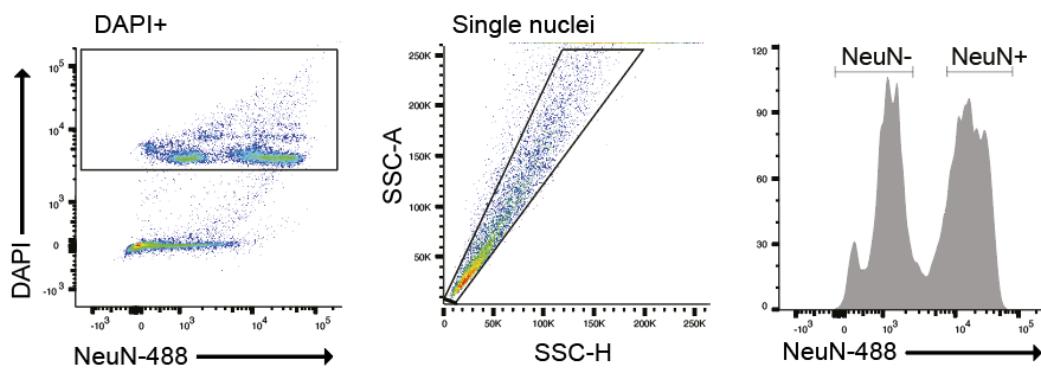
a

| ID | Origin | Condition | Age | Gender | Mutation |
|----------|-----------|-----------|-----|--------|------------------------|
| 5345 | NIH | control | 48 | Female | |
| 4788 | NIH | control | 48 | Female | |
| 1739 | NIH | control | 48 | Female | |
| 2082 | NIH | CdLS | 48 | Female | NIPBL 7669T>C ; C2557R |
| SD047/15 | Edinburgh | control | 19 | Male | |
| SD023/08 | Edinburgh | control | 24 | Female | |
| SD030/11 | Edinburgh | control | 30 | Male | |
| CDL-380P | CHOP | CdLS | 19 | Male | NIPBL 2965_2966insA |
| CDL-744P | CHOP | CdLS | 24 | Female | NIPBL 5691delT |
| CDL-764P | CHOP | CdLS | 31 | Female | none detected |

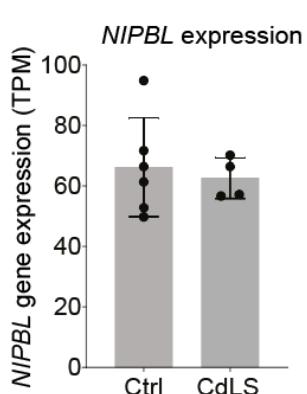
b



c

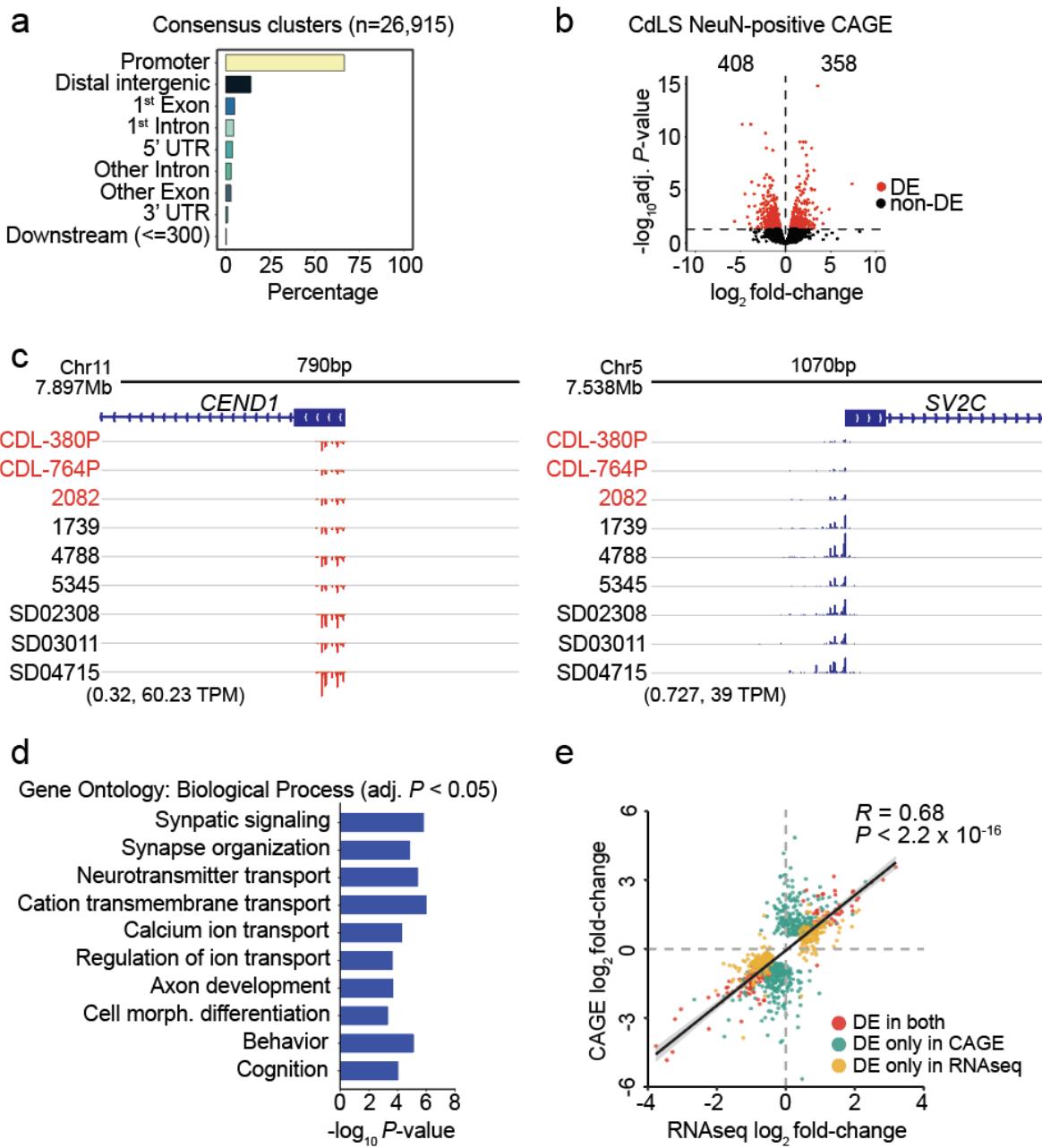


d



Supplemental Figure 1. Characterization of CdLS patients and isolation of nuclei

- a) Metadata for patient samples, including ID, origin, disease status, patient age, patient gender and mutational status if applicable.
- b) Sanger sequencing of CdLS patient samples 2082, CDL-744P and CDL-380P confirming mutations initially identified by RNAseq. Black arrows indicate direction of Sanger sequencing reaction. Blue arrow indicates site of mutation.
- c) Flow cytometric parameters used to purify neuronal nuclei from whole frozen post-mortem tissue. First, we identified DAPI+ nuclei, these were then size selected for single nuclei before being separated on a histogram to identify two distinct populations of neuronal and non-neuronal nuclei for purification.
- d) Mean expression level, measured as TPM (transcripts per million), of *NIPBL* in control and CdLS NeuN-positive RNAseq. (Errors bars = standard deviation, Control, n = 6; CdLS, n = 4).



Supplemental Figure 2. CAGE reveals deregulated gene expression in CdLS neurons

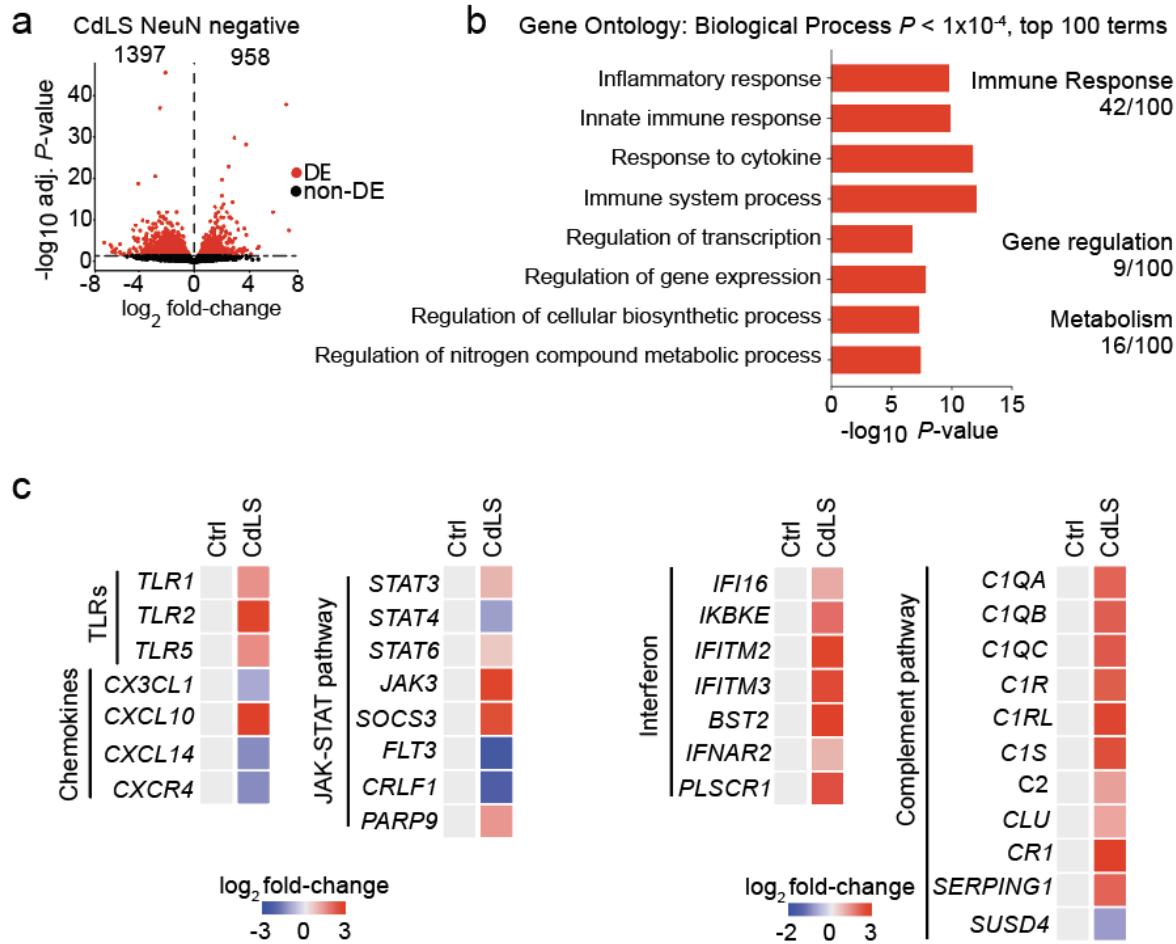
a) Genomic annotation of consensus clusters called by CAGEr (ref. 67). A total of 26,915 consensus clusters were annotated with ChIPseeker⁷⁵ and the percentage of consensus clusters associated with genomic features was quantified. Consensus clusters associated with database-annotated TSSs or 5' UTRs (n = 18,867) were kept for downstream analysis.

b) Volcano plot of \log_2 fold-change versus adjusted *P*-value obtained from CAGE analysis of 3 patient and 6 control NeuN-positive samples. 408 down- and 358 upregulated genes were identified (adj. *P*-value < 0.05). Red indicates differential expression (DE).

c) Screenshots from the Integrative Genomics Viewer (IGV) showing CAGE-defined TSSs (CTSSs) and their corresponding normalized signal (in TPM) across 3 patient (top) and 6 control (bottom) samples. Left: normalized CAGE signal within the *CEND1* promoter region. *CEND1* (cell cycle exit and neuronal differentiation protein 1) is not differentially expressed by DESeq2. Right: normalized CAGE signal within the *SV2C* promoter region. *SV2C* (synaptic vesicle glycoprotein 2C) is significantly downregulated by DESeq2.

d) GO terms represented in the top 25 most enriched (adj. *P*-value < 0.05, *q*-value < 0.05) biological processes for downregulated CAGE genes identified with clusterProfiler (ref. 77) using 11.634 genes active in control samples as background.

e) Scatter plot of \log_2 fold-change values for genes significantly deregulated in CAGE (green), RNA-seq (orange), or both (red). *R* = Pearson's correlation coefficient. *R* and *P*-value were calculated using the "cor.test" function in R. The linear regression trendline is shown in black and, confidence bands are displayed in grey.

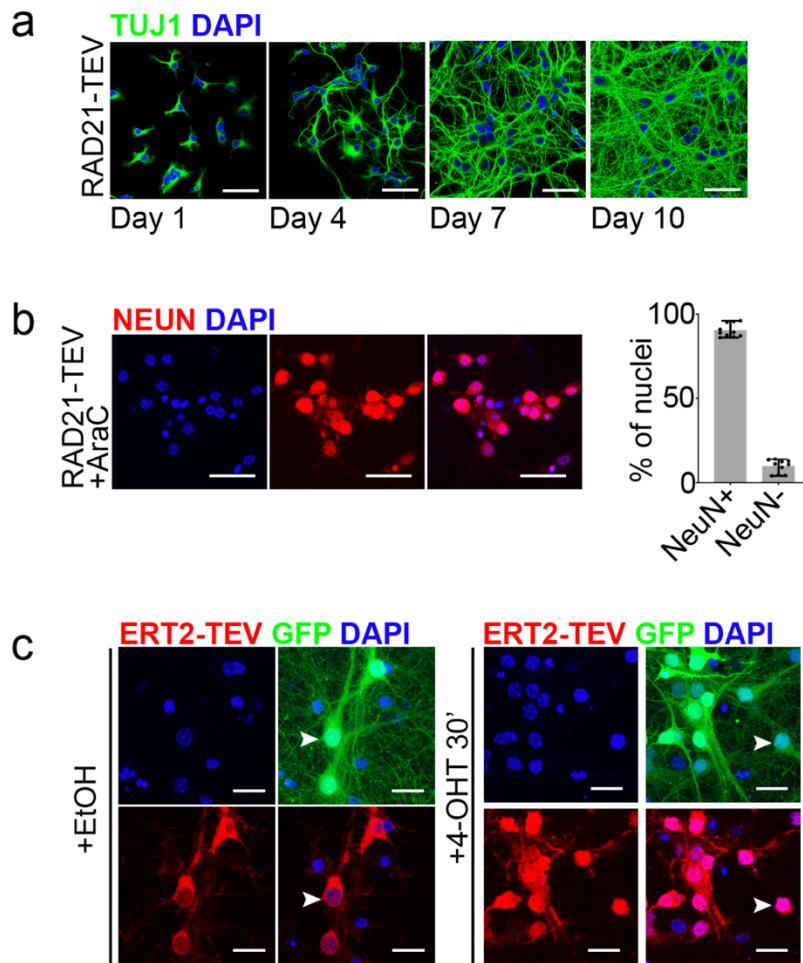


Supplemental Figure 3. RNA-seq of NeuN-negative nuclei from CdLS cortex

a) Volcano plot of gene expression fold-change versus adjusted *P* value of up- and downregulated genes in NeuN-negative nuclei from CdLS patients (n = 3). 1397 genes were down- and 958 genes were upregulated (RUVg k=2, adj. *P* < 0.05, shown in red, DE = differentially expressed)

b) Bar graph of individual GO terms within broad categories. Terms represented by upregulated genes are shown in red.

c) Heatmap of log₂ fold-changes for representative deregulated genes in NeuN-negative CdLS samples.

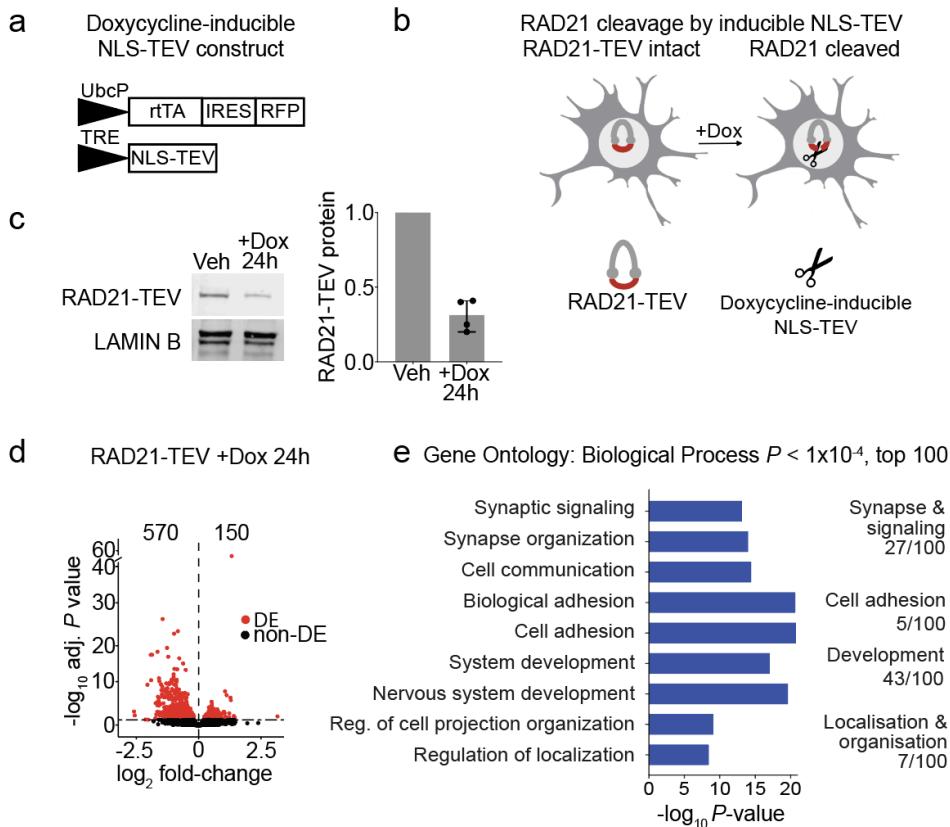


Supplemental Figure 4. Validation of RAD21-TEV neuronal culture and ERt2-TEV translocation

a) Immunofluorescence staining for TUJ1 and DAPI in explant RAD21-TEV neurons showing neuronal maturation over the course of ten days (scale bar = 50 μ m).

b) Immunofluorescence staining for NeuN and DAPI in ten-day old explant RAD21-TEV neuron cultures treated with AraC at day 5. Bar graph plots mean % of NeuN+ and NeuN- nuclei in ten-day old explant RAD21-TEV neuron cultures treated with AraC at day 5. (n =10, error bars = range, scale bar = 50 μ m).

c) Immunofluorescence staining for ERt2-TEV and DAPI, with GFP expressed from ERt2-TEV lentiviral transduction. Left: ERT2-TEV is retained in the cytoplasm with vehicle treatment. Right: ERt2-TEV translocates to the nucleus following 4-OHT exposure. White arrow heads indicate example cells. (Scale bar = 25 μ m).



Supplemental Figure 5. Acute RAD21 depletion by NLS-TEV triggers changes in neuronal gene expression

a) Schematic of lentiviral construct containing an all-in-one doxycycline inducible NLS-TEV system. Tet-On advanced transactivator (rtTA) and RFP are constitutively expressed under control of the ubiquitin promoter. NLS-TEV expression is in turn controlled by tet response element (TRE) promoter upon addition of doxycycline.

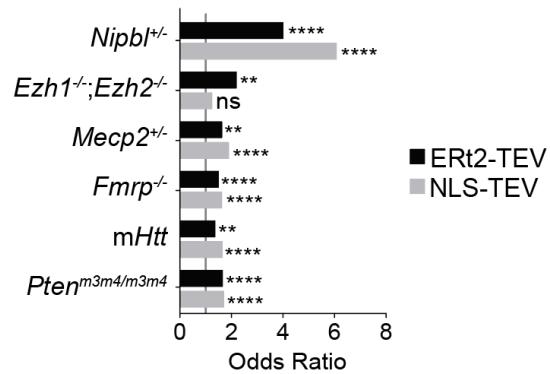
b) Schematic of doxycycline dependent RAD21-TEV degradation by NLS-TEV.

c) Western blot of RAD21-TEV protein expression 24 hours after Dox pulse (6h, 100ng/ml). Bar plot of RAD21-TEV protein expression normalised to LAMIN B 24 hours after Dox exposure, ~30% RAD21-TEV protein remained ($n = 4$, $h = 6$ hours).

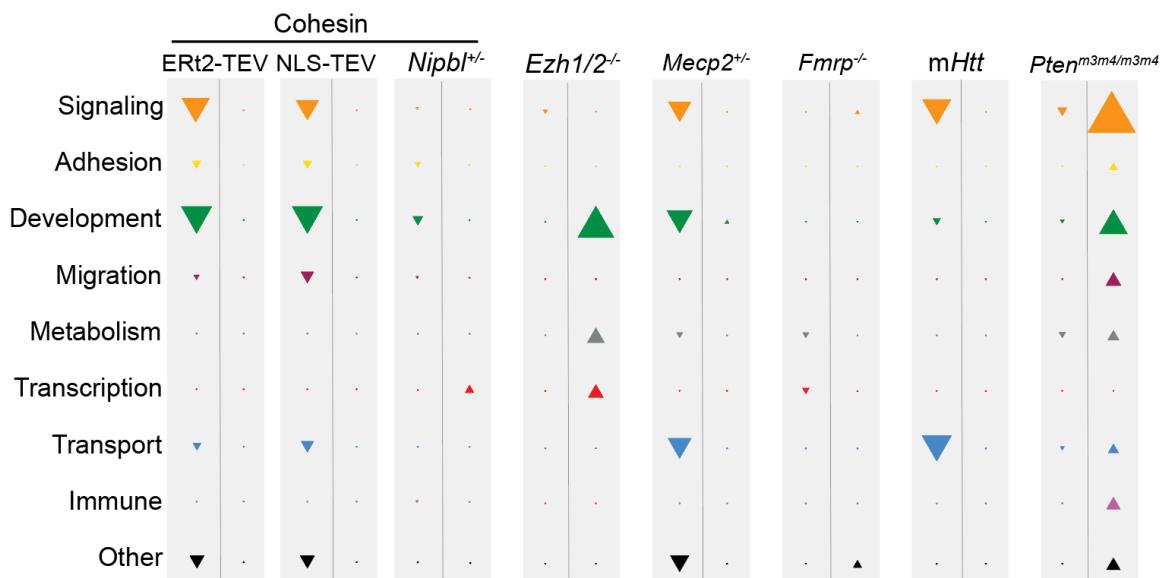
d) Volcano plot of gene expression \log_2 fold-change versus adjusted P value in RAD21-TEV +Dox (24 hours after 6-hour 100ng/ml pulse, $n = 3$). 570 genes were down- and 150 genes were upregulated (adj $P < 0.05$, shown in red, DE = differentially expressed).

e) Top individual GO terms and broad categories for downregulated genes (blue). There was no significant enrichment for upregulated genes.

a Overlap with acute RAD21-TEV depletion



b Gene ontology: Biological process, $P < 1 \times 10^{-4}$



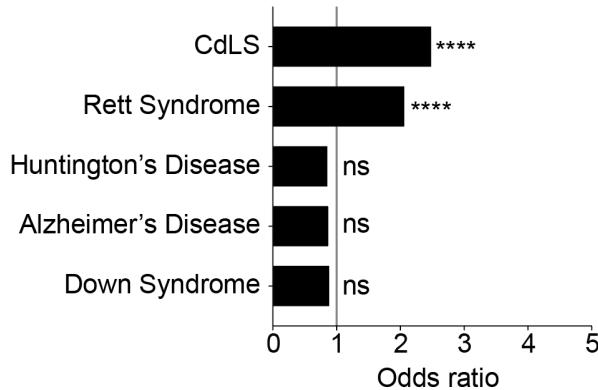
Supplemental Figure 6. Comparison of gene deregulation in acute RAD21-TEV depletion and mouse models of neuronal dysfunction

a) Bar graph of overlap between deregulated genes in RAD21-TEV neurons (NLS-TEV and ERt2-TEV combined) and other animal models of neuronal dysfunction including *Nipbl*^{+/−} (ref. 19), deletion of PRC2 components *Ezh1* and *Ezh2* (*Ezh1*^{−/−}; *Ezh2*^{−/−})³³, *Mecp2*^{+/-} (ref. 34), *Fmrp*^{−/−} (ref. 35), mutant *Htt* (*mHtt*), a mouse model of Huntington's disease³⁶ and mouse mutant *Pten*^{m3m4/m3m4} (ref. 37). * $P < 0.05$, ** $P < 0.01$, *** $P < 0.001$, **** $P < 0.0001$, Fisher's exact test.

b) Bubble plot of number of represented gene ontologies in different mouse models of neuronal dysfunction and RAD21-TEV depleted neurons separated into broad

categories. GO analysis conducted on both down- and upregulated genes for each condition indicated in the respective columns. For each condition either all significant GO ($P < 1 \times 10^{-4}$) or only top 100, where more than 100 terms reach significance, are plotted. Triangle size is proportional to the number of terms in that category, and direction of point indicates direction of deregulation.

Overlap between genes deregulated in RAD21-TEV and human disease



Supplemental Figure 7. Comparison of gene deregulation in acute RAD21-TEV depletion and in human disease

Overlap of deregulated genes between RAD21-TEV neurons (NLS-TEV and ERt2-TEV combined) and neuronal or whole cortical tissue RNAseq of human neurological diseases including NeuN-positive CdLS, Rett Syndrome³⁴, Huntington's disease³⁹, Alzheimer's disease⁴⁰, and Down Syndrome⁴¹. * $P < 0.05$, ** $P < 0.01$, *** $P < 0.001$, **** $P < 0.0001$, Fisher's exact test.

Methods

Mice. Mouse work was done under a project licence issued by the UK Home Office, UK following review by the local Ethics Committee as required by the Animals (Scientific Procedures) Act. For timed pregnancies the day of the vaginal plug was counted as day 0.5. *Rad21*^{Tev/Tev} mice have been described^{30,31}.

Neuronal cultures. Mouse cortices were dissected and dissociated on e14.5 as described⁴⁹. Neurons were maintained in Neurobasal medium with B27 supplement (Invitrogen), 1mM L-glutamine, and 100U/ml penicillin/streptomycin. Cells were plated at a density of $1.25 \times 10^5/\text{cm}^2$ on 0.1mg/ml poly-D-lysine (Millipore) coated plates, and half the media was replaced every 3 days. Cultures were treated with 5 μM Cytosine β -D-arabinofuranoside (Ara-C, Sigma) at day 5. Neurons were plated on 12mm coverslips coated with poly-D-lysine (VWR) for immunofluorescence staining.

Lentivirus construction and packaging. Cloning was performed by PCR amplification of cDNA using Phusion taq (Invitrogen) or Q5 polymerase (NEB), restriction digestion, gel purification of fragments, and ligation into the target vector. To generate the ERt2-TEV lentiviral construct, Cas9 was removed from the lentiviral plasmid FUG-T2A-CAS9, which was a gift from Dr. Bryan Luikart (Addgene plasmid #75346)⁵⁰, by restriction enzyme digestion and replaced with v5-ERt2-TEV. To generate the NLS-TEV lentiviral construct, Gibson assembly was performed using Lenti-iCas9-Neo backbone and an NLS-TEV construct kindly provided by Dr. Kikue Tachibana. Lentivirus was generated as described previously⁵¹ with minor modifications. Briefly, 293T cells were co-transfected with the expression vector, together with packaging plasmids pCMV-VSV-G and pCMV-delta8.9 using PEI (Polysciences Inc.) 48 hours after transfection, supernatant containing viral particles was collected and concentrated by ultracentrifugation. The titer of lentivirus was measured by transducing 5×10^4 293T cells per well of a 24-well plate at serial dilutions, and quantification of GFP/RFP-positive cells by flow cytometry after 3 days. $\text{TU} = (\text{P}^* \text{N} / 100^* \text{V})^* 1 / \text{DF}$. ($\text{p} = \%$ GFP+ or RFP+ cells, $\text{N} =$ number of cells at time of transfection, $\text{V} =$ volume of dilution added to each well, $\text{DF} =$ dilution factor).

Inducible cleavage of RAD21-TEV. For cleavage of RAD21-TEV, neurons were plated as described above and transduced at day 3 with lentivirus containing either ERt2-TEV or NLS-TEV at a multiplicity of infection of 3. For ERt2-TEV dependent RAD21-TEV degradation, neurons were treated on culture day 10 with 500nM 4-hydroxytamoxifen (4-OHT) or vehicle (ethanol). For NLS-TEV dependent RAD21-TEV degradation, neurons were treated on culture day 10 with a 6-hour doxycycline (100ng/ml) pulse or

vehicle (water). Cells were then rinsed, and media replaced with conditioned neuronal media.

Isolation of nuclei from post-mortem tissue. Isolation of nuclei was performed as previously described⁵² with minor modifications. Briefly, 50-250mg of pre-frontal cortical grey matter was homogenised in homogenisation buffer (250mM sucrose, 25mM KCl, 5mM MgCl₂, 10mM Tris buffer pH 8.0, 1µM DTT, 1X Proteinase Inhibitor w/o EDTA (Roche), 0.4U µl-1 RNaseIn (ThermoFisher) 0.2U µl-1 Superasin (ThermoFisher), 1µM DAPI) and centrifuged through an iodixanol gradient (Sigma). Pelleted nuclei were washed and then stained with NeuN antibody (Abcam, ab190195) in staining buffer (PBS, 1%BSA, 0.2U/µl RNaseIn (Thermofisher), NeuN antibody (1:200) for one hour at 4°C. For negative controls antibody was excluded. Nuclei were then sorted on BD Fusion, and collected in wash buffer (PBS, 1%BSA, 0.2U/µl RNaseIn (Thermofisher) before RNA extraction.

Protein analysis. Whole cell extracts were prepared by suspending cells in PBS, centrifugation and resuspension in protein sample buffer (50mM Tris-HCl pH6.8, 1% SDS, 10% glycerol) followed by quantification using Qubit (Invitrogen). Following quantification, 0.001% bromophenol blue, and 5% beta-mercaptoethanol were added. Sodium dodecyl sulphate-polyacrylamide gel electrophoresis (SDS-PAGE) was carried out with the Bio-Rad minigel system. Resolved gels were blotted on to polyvinylidene fluoride transfer membrane, followed by 1 hour incubation in fluorescent blocker (Millipore), and then incubated in primary antibody diluted to the appropriate amount in fluorescent blocker (Millipore) overnight at 4°C. Primary antibodies were goat polyclonal to LAMIN B (1:5000, Santa Cruz Biotechnology, sc-6216) and mouse monoclonal to LAMIN B (1:5000, Santa Cruz Biotechnology, sc-374015), and mouse monoclonal anti-myc tag for detection of RAD21-TEV (1:500, Santa Cruz biotechnology, sc-40). Blots were then incubated in secondary antibody diluted in fluorescent blocker (Millipore) for 1 hour at room temperature. Secondary antibodies were donkey anti-goat IgG (H+L) Alexa Fluor 680 (1:10,000, Thermofisher, A-21804) and goat anti-mouse IgG (H+L) Alexa Fluor 680 (1:10,000 Thermofisher, A-28183). Stained membranes were imaged on an Odyssey instrument (LICOR).

Immunocytochemistry. Neurons plated on coverslips were fixed with PBS containing 4% paraformaldehyde and 4% sucrose warmed to 37°C for 10 minutes at room temperature. Neurons were then permeabilised using 0.3% Triton X-100 for 10 minutes at room temperature, and blocked in blocking solution (10% normal goat serum, 0.1% Triton X-100 in PBS) for one hour at room temperature. Samples were incubated with primary antibodies diluted in staining solution (0.1% Triton X-100, 2% normal goat serum in PBS) in a humidified chamber overnight at 4°C. Primary antibodies were

mouse monoclonal TUJ1 (1:500, Biolegend, 801213), mouse monoclonal V5 (1:250, Sigma-Aldrich, V8012) and mouse monoclonal NeuN (1:1000, Abcam, ab104224). Samples were then washed with PBS and incubated in secondary antibodies diluted 1:500 in staining solution (0.1% Triton X-100, 2% normal goat serum in PBS) in a humidified chamber for one hour at room temperature. Secondary antibodies were, goat anti-mouse IgG (H+L) Alexa Fluor 488 (ThermoFisher, A-1101) and goat anti-mouse IgG (H+L) Alexa Fluor 568 (ThermoFisher, A-11004). Cells were mounted in Vectashield medium containing DAPI (Vector Labs). Samples were visualised using a TCS SP5 Leica laser scanning confocal microscope. Images were processed using Leica Confocal Software and FIJI. For quantification of NeuN-positive cells, DAPI-identified nuclei that colocalised with NeuN signal were counted as neuronal, and DAPI-identified nuclei without NeuN were counted as non-neuronal. Samples were quantified using a processing pipeline developed in CellProfiler (version 2.2, www.cellprofiler.org).

Identification of human mutations. *NIPBL*-mutation screening was performed by PCR of genomic DNA and Sanger Sequencing. Genomic DNA was isolated from brain cortical tissue using DNeasy blood & tissue extraction kit (Qiagen), according to the manufacturer's protocol. The entire *NIPBL* coding region (exons 2–47) was screened for mutations. Primer pairs were designed using ExonPrimer to amplify exons, exon/intron boundaries, and short flanking intronic sequences. All PCR reactions were performed in a 25- μ l reaction volume containing 60 ng of genomic DNA, 1 U of AmpliTaq Gold (Applied Biosystems), 20 pmol of each primer, 75 μ M of each dNTP, 10X PCR buffer II (Applied Biosystems), and 1.0 mM or 1.5 mM of MgCl₂ (Applied Biosystems). Prior to sequencing all the PCR products were purified with USB ExoSAP-IT PCR Product Cleanup (Affymetrix) following the manufacturer's instructions, and subsequently sequenced using BigDye Terminator v3.1 cycle sequencing and analyzed on an ABI 3730 (Applied Biosystems, Carlsbad, CA). Sequence analysis was performed using Sequencher software version 5.2.4 and Mutation Taster⁵³.

RNA extraction and sequencing. RNA was extracted from both cells and nuclei using Picopure RNA Isolation Kit (Thermofisher) according to manufacturer's instructions. Residual DNA was eliminated using RNase-Free DNase Set (Qiagen). Sorted nuclei were pelleted before proceeding with downstream RNA isolation. Cultured neurons were treated with extraction buffer (Picopure RNA Isolation Kit, Thermofisher) directly in the well before proceeding with downstream RNA isolation. RNA was assessed for quality (Bioanalyzer, Agilent) and quantity (Qubit, Invitrogen). Libraries for RNA-sequencing of RAD21-TEV neurons were prepared using 200ng of RNA and NEBNext Ultra II Directional RNA Library Prep Kit for Illumina with polyA enrichment for ERt2-TEV, and ribozero ribosomal RNA depletion by NLS-TEV. Libraries for RNA-sequencing of human cortical nuclei were prepared using 40ng of RNA and NEBNext Ultra II

Directional RNA Library Prep Kit for Illumina using ribozero ribosomal RNA depletion. Library quality and quantity were assessed on a bioanalyser and Qubit, respectively. Libraries were sequenced on an Illumina HiSeq2500 (v4 chemistry), generating 40-million paired-end 100bp reads per sample.

SLIC-CAGE. SLIC-CAGE libraries were made using the published SLIC-CAGE protocol^{54,55}. Briefly, 50 ng of nuclear RNA was used per sample and mixed with 4.95 µg of the RNA carrier mix. The library fragments derived from the carrier RNA were digested using I-CeuI and I-SceI homing endonucleases, and the library fragments derived from the sample RNA were PCR-amplified (14-15 amplification cycles). Eight samples with different barcodes were standardly multiplexed prior to sequencing and the resulting library mixes sequenced on a HiSeq2500 in a single-end 50 bp mode.

ATAC-seq. ATAC-seq was performed on 5×10^5 nuclei isolated from control (n = 5) and CdLS patients (n = 4) per sample, following the omni-ATAC protocol⁵⁶. Omni-ATAC is an improvement of the original ATAC protocol, with the inclusion of several additional detergents, NP-40, Tween-20 and digitonin in the nuclei digestion. After 30 minutes of incubation, samples were purified using the Zymo ChIP DNA Clean and Concentrator (Zymo Research) and libraries prepared as previously described⁵⁷.

4C-seq. 4C was performed as previously described²⁹ with modifications. Briefly, neuronal cells were cross-linked in PBS with 1% formaldehyde for 10 min at RT and nuclei were isolated in lysis buffer (10 mM Tris-HCl pH 7.4, 150 mM NaCl, 0.5% NP-40, 1 x protease inhibitors). After digestion using HindIII, digested products were ligated by T4 DNA ligase. 3C templates were purified and then digested again using a second enzyme, NlaIII. Final 4C sequencing libraries were purified using a Roche High-Pure PCR Product Purification Kit, and then sequenced using an Illumina HiSeq 2500 platform. Sequence data was analysed using the 4Cseqpipe software suite⁵⁸ with the setting of "-stat_type mean -trend_resolution 40000". 4C-seq primers can be found in Extended Data Table 12.

RNAseq analysis. Paired-end 100bp sequencing reads were aligned against mouse genome (mm9) or human genome (hg38) with Tophat2 (2.0.11)⁵⁹. Reference sequence assembly and transcript annotation that was obtained from Illumina iGenomes. Gene based read counts were obtained using featureCounts function from Rsubread Bioconductor package (1.24.2)^{60,61}. Normalisation was performed in DESeq2 Bioconductor package (1.24.0)⁶² and data was rlog transformed to allow for visualisation by PCA and heatmaps. Differentially expressed gene (DEG) analysis was also performed with DESeq2 and DEGs were defined with Benjamini-Hochberg adjusted $P < 0.05$. GO term analysis was performed with goseq Bioconductor package (1.24)⁶³. After converting mouse gene symbol to human gene symbol using Ensembl Biomart⁶⁴, Gene

Set Enrichment Analysis (GSEA 2.2.0)^{65, 66} was performed with GseaPreranked tool using Hallmarks gene set. For RAD21-TEV RNAseq data, polyadenylated genes were annotated using PolyA_DB (ref.67) and non-polyadenylated genes were removed from the analysis. Glial genes were defined by control NeuN-negative vs control NeuN-positive RNAseq data with adj $P < 0.05$ and fold-change > 20 . We processed CdLS NeuN-positive RNAseq data after removing glial genes. In both human CdLS NeuN-positive and NeuN-negative RNAseq dataset, we only focused on autosomes to mitigate sex specific differences. Unwanted batch effects were controlled in CdLS NeuN-positive and NeuN-negative RNAseq by using R Bioconductor package RUVseq (ref. 68) with RUVg function and k=2 for both. For CdLS NeuN-negative RNAseq, sample CDL-744P was excluded. Heatmaps were generated using R Bioconductor package ComplexHeatmap (ref. 69) and GraphPad Prism 8.0.

SLIC-CAGE analysis. Sequenced libraries were demultiplexed using CASAVA allowing zero mismatches for barcode identification. Demultiplexed CAGE tags (47 bp) were mapped to a reference GRCh37/hg19 human genome using Bowtie2 (--phred33-quals -U)⁷⁰ with default parameters allowing zero mismatches per 22 nucleotide seed sequence. The mapped reads were then sorted using Samtools v 1.10 (ref. 71) and uniquely mapped reads kept for downstream analysis in R graphical and statistical computing environment (<http://www.R-project.org/>). The mapped and sorted unique reads were imported into R as bam files using the standard workflow within the CAGER package v1.20 (ref. 72). All 5'ends of reads are CAGE-supported transcription start sites (CTSSs) and the number of each CTSS (number of tags) reflects the transcript expression levels. Raw tags were normalised using a referent power-law distribution (alpha=1.48) and expressed as normalized tags per million (TPMs)⁷³. The highest expressed CTSS is termed the 'dominant CTSS'. Correlation of samples on CTSS level was estimated based on TPM values. This analysis led to the removal of sample CDL-744P due to its low correlation with the rest of the samples. CTSSs with sufficient support (at least 0.1 TPM in at least 1 sample) were further used to construct promoter regions. First, sample-specific clustering of CTSSs was performed to define tag clusters, allowing a maximum distance of 20 bp between any 2 CTSSs in the same tag cluster. In order to allow for between-sample expression profiling, tag clusters with sufficient support (at least 5 TPM) within 100 bp of each other were aggregated across all samples to define promoter regions (consensus clusters). To allow comparison with RNA-seq data, genomic coordinates of consensus clusters were converted into GRCh38/hg38 coordinates using the liftOver tool⁷⁴. Consensus clusters were annotated using ChIPseeker v.1.22.1 (ref. 75) with the transcript annotation *TxDb.Hsapiens.UCSC.hg38.knownGene* to define TSSs of known genes and the *org.Hs.eg.db* annotation to allocate Ensembl IDs to gene symbols. A consensus cluster was assigned to the promoter of a known gene if it was within a region 500 bp upstream and 100 bp downstream of the annotated TSS or within its 5' UTR. This resulted in a set

of 18,867 promoter-associated consensus clusters. To reduce the effect of sex differences, we excluded consensus clusters located on chromosomes X and Y. Consensus clusters were further filtered to exclude those associated with glial-specific genes, resulting in a final set of 18,103 consensus clusters and corresponding Ensembl IDs. DESeq2 v.1.26.0 (ref. 62) was used to define differentially expressed genes, accounting for treatment and age effects. Genes with BH-adjusted *P*-value < 0.05 and log₂ fold-change values of > 0 or < 0 were considered as significantly up- or downregulated. Principal component analysis was implemented in DESeq2 on rlog-transformed read counts. PCA plots were created using the top 10,000 consensus clusters with the highest row variance. Variance associated with age group was removed from the rlog-transformed data with limma v.3.42.0 (ref. 76) for visualisation and plotted as described above. Functional enrichment analysis of significantly up- or downregulated genes was performed using GO over-representation test with clusterProfiler v.3.14.3 (ref. 77). Enriched GO terms associated with biological processes were determined based on a BH-adjusted *P*-value cutoff = 0.05 and a *q*-value cutoff = 0.05 against a background of genes expressed in control samples (mean CAGE signal > 2 TPM, *n* = 11,634). CAGE data was visualized using the Integrative Genomics Viewer (Broad Institute)⁷⁸.

ATAC-seq analysis. After demultiplexing, raw reads were trimmed with Trim Galore! (trim_galore_v0.4.4;https://www.bioinformatics.babraham.ac.uk/projects/trim_galore) with arguments --trim-n --paired. The trimmed reads were then aligned to hg38 using bowtie2 (bowtie2/2.2.9, with -p8 -t –very-sensitive -X 2000)⁷⁰. Samtools 1.2 (ref. 71) was applied for sorting the reads and picardtools MarkDuplicates (1.90; Broad Institute 2019) was used for marking duplicated reads. R bioconductor package ATACseqQC (ref. 79) was applied for removing duplicated reads, reads without properly mapped mates, and reads mapped to chrMT. alignmentSieve function from deeptools⁸⁰ was applied for shifting ATACseq reads⁵⁷ and keeping nucleosome-free reads (fragment length <=100 bps) with argument --ATACshift --minFragmentLength 0 --maxFragmentLength 100. We used the nucleosome-free bin for the downstream analysis. Peak calling was proceeded with MACS2 peakcalling function with -f BAMPE -g hs (ref. 81). Peaks overlapped with the black list hg38 v2 (ref. 82) were removed and only those that appeared in more than any 2 samples were kept. Peak-based read counts were then obtained using the featureCounts (ref. 61) function from Rsubread Bioconductor package. Differential accessibility analysis was proceeded with DESeq2 Bioconductor package⁶⁰. Peaks were annotated with ChIPseeker Bioconductor package⁷⁵.

Cohesin binding and enhancer proximity. Binding of genes by cohesin and CTCF was defined by RAD21 (ref.17, ENCODE: ENCSR198ZYJ) or CTCF ChIP-seq (ENCODE: ENCSR822CEA) peaks overlapping with gene bodies.-Proximity of genes to

enhancers was defined by whether genes were the nearest neighbour to a neuronal enhancer^{83,84}. R Bioconductor package GeneOverlap (ref. 85) was applied for one-tailed Fisher's exact test and reported the odds ratio and *P* value.

Supplementary references

49. Beaudoin, G. et al. Culturing pyramidal neurons from the early postnatal mouse hippocampus and cortex. *Nature Protocols* 7, 1741-1754 (2012).
50. Williams, M. et al. A Retroviral CRISPR-Cas9 System for Cellular Autism-Associated Phenotype Discovery in Developing Neurons. *Scientific Reports* 6, (2016).
51. Auer, S. et al. Silencing neurotransmission with membrane-tethered toxins. *Nature Methods* 7, 229-236 (2010).
52. Krishnaswami, S. et al. Using single nuclei for RNA-seq to capture the transcriptome of postmortem neurons. *Nature Protocols* 11, 499-524 (2016).
53. Schwarz, J., Cooper, D., Schuelke, M. & Seelow, D. MutationTaster2: mutation prediction for the deep-sequencing age. *Nature Methods* 11, 361-362 (2014).
54. Cvetescic, N. et al. SLIC-CAGE: high-resolution transcription start site mapping using nanogram-levels of total RNA. *Genome Research* 28, 1943-1956 (2018).
55. Cvetescic, N., Pahita, E. & Lenhard, B. Transcription Start Site Mapping Using Super-low Input Carrier-CAGE. *Journal of Visualized Experiments* (2019). doi:10.3791/59805
56. Corces, M. et al. An improved ATAC-seq protocol reduces background and enables interrogation of frozen tissues. *Nature Methods* 14, 959-962 (2017).
57. Buenrostro, J., Giresi, P., Zaba, L., Chang, H. & Greenleaf, W. Transposition of native chromatin for fast and sensitive epigenomic profiling of open chromatin, DNA-binding proteins and nucleosome position. *Nature Methods* 10, 1213-1218 (2013).
58. van de Werken, H. et al. Robust 4C-seq data analysis to screen for regulatory DNA interactions. *Nature Methods* 9, 969-972 (2012).
59. Kim, D. et al. TopHat2: accurate alignment of transcriptomes in the presence of insertions, deletions and gene fusions. *Genome Biology* 14, R36 (2013).
60. Liao, Y., Smyth, G. & Shi, W. The Subread aligner: fast, accurate and scalable read mapping by seed-and-vote. *Nucleic Acids Research* 41, e108-e108 (2013).

61. Liao, Y., Smyth, G. & Shi, W. featureCounts: an efficient general purpose program for assigning sequence reads to genomic features. *Bioinformatics* 30, 923-930 (2013).
62. Love, M., Huber, W. & Anders, S. Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. *Genome Biology* 15, (2014).
63. Young, M., Wakefield, M., Smyth, G. & Oshlack, A. Gene ontology analysis for RNA-seq: accounting for selection bias. *Genome Biology* 11, R14 (2010).
64. Hunt, S. et al. Ensembl variation resources. *Database* 2018, (2018).
65. Mootha, V. et al. PGC-1 α -responsive genes involved in oxidative phosphorylation are coordinately downregulated in human diabetes. *Nature Genetics* 34, 267-273 (2003).
66. Subramanian, A. et al. Gene set enrichment analysis: A knowledge-based approach for interpreting genome-wide expression profiles. *Proceedings of the National Academy of Sciences* 102, 15545-15550 (2005).
67. Wang, R., Nambiar, R., Zheng, D. & Tian, B. PolyA_DB 3 catalogs cleavage and polyadenylation sites identified by deep sequencing in multiple genomes. *Nucleic Acids Research* 46, D315-D319 (2017).
68. Risso, D., Ngai, J., Speed, T. & Dudoit, S. Normalization of RNA-seq data using factor analysis of control genes or samples. *Nature Biotechnology* 32, 896-902 (2014).
69. Gu, Z., Eils, R. & Schlesner, M. Complex heatmaps reveal patterns and correlations in multidimensional genomic data. *Bioinformatics* 32, 2847-2849 (2016).
70. Langmead, B. & Salzberg, S. Fast gapped-read alignment with Bowtie 2. *Nature Methods* 9, 357-359 (2012).
71. Li, H. et al. The Sequence Alignment/Map format and SAMtools. *Bioinformatics* 25, 2078-2079 (2009).
72. Haberle, V., Forrest, A., Hayashizaki, Y., Carninci, P. & Lenhard, B. CAGEr: precise TSS data retrieval and high-resolution promoterome mining for integrative analyses. *Nucleic Acids Research* 43, e51-e51 (2015).
73. Balwierz, P. et al. Methods for analyzing deep sequencing expression data: constructing the human and mouse promoterome with deepCAGE data. *Genome Biology* 10, R79 (2009).
74. Kuhn, RM., Haussler, D., & Kent, WJ. The UCSC genome browser and associated tools. *Brief Bioinform* 14: 144-161 (2013).

75. Yu, G., Wang, L. & He, Q. ChIPseeker: an R/Bioconductor package for ChIP peak annotation, comparison and visualization. *Bioinformatics* 31, 2382-2383 (2015).
76. Ritchie ME, Phipson B, Wu D, Hu Y, Law CW, Shi W, Smyth GK. limma powers differential expression analyses for RNA-sequencing and microarray studies. *Nucleic Acids Research* 43: e47 (2015).
77. Yu, G., Wang, L., Han, Y. & He, Q. clusterProfiler: an R Package for Comparing Biological Themes Among Gene Clusters. *OMICS: A Journal of Integrative Biology* 16, 284-287 (2012).
78. Robinson, J. et al. Integrative genomics viewer. *Nature Biotechnology* 29, 24-26 (2011).
79. Ou, J. et al. ATACseqQC: a Bioconductor package for post-alignment quality assessment of ATAC-seq data. *BMC Genomics* 19, (2018).
80. Ramírez, F., Dündar, F., Diehl, S., Grüning, B. & Manke, T. deepTools: a flexible platform for exploring deep-sequencing data. *Nucleic Acids Research* 42, W187-W191 (2014).
81. Zhang, Y. et al. Model-based Analysis of ChIP-Seq (MACS). *Genome Biology* 9, R137 (2008).
82. Amemiya, H., Kundaje, A. & Boyle, A. The ENCODE Blacklist: Identification of Problematic Regions of the Genome. *Scientific Reports* 9, (2019).
83. Malik, A. et al. Genome-wide identification and characterization of functional neuronal activity-dependent enhancers. *Nature Neuroscience* 17, 1330-1339 (2014).
84. Nott, A. et al. Brain cell type-specific enhancer–promoter interactome maps and disease-risk association. *Science* 366, 1134-1139 (2019).
85. Shen L, Sinai M (2019). GeneOverlap: Test and visualize gene overlaps. R package version 1.22.0, <http://shenlab-sinai.github.io/shenlab-sinai/>.