Article

Hominini-specific regulation of *CBLN2* increases prefrontal spinogenesis

https://doi.org/10.1038/s41586-021-03952-y

Received: 31 December 2019

Accepted: 25 August 2021

Published online: 01 October 2021

Check for updates

Mikihito Shibata[1], Kartik Pattabiraman[1,2], Sydney K. Muchnik[1,3], Navjot Kaur[1], Yury M. Morozov[1], Xiaoyang Cheng[4,5,6], Stephen G. Waxman[4,5,6] & Nenad Sestan[1,2,3,7,8,9,10] ✉

The similarities and differences between nervous systems of various species result from developmental constraints and specific adaptations[1–4]. Comparative analyses of the prefrontal cortex (PFC), a cerebral cortex region involved in higher-order cognition and complex social behaviours, have identified true and potential human-specific structural and molecular specializations[4–8], such as an exaggerated PFC-enriched anterior–posterior dendritic spine density gradient[5]. These changes are probably mediated by divergence in spatiotemporal gene regulation[9–17], which is particularly prominent in the midfetal human cortex[15,18–20]. Here we analysed human and macaque transcriptomic data[15,20] and identified a transient PFC-enriched and laminar-specific upregulation of cerebellin 2 (CBLN2), a neurexin (NRXN) and glutamate receptor-δ GRID/GluD-associated synaptic organizer[21–27], during midfetal development that coincided with the initiation of synaptogenesis. Moreover, we found that species differences in level of expression and laminar distribution of *CBLN2* are, at least in part, due to Hominini-specific deletions containing SOX5-binding sites within a retinoic acid-responsive *CBLN2* enhancer. In situ genetic humanization of the mouse *Cbln2* enhancer drives increased and ectopic laminar *Cbln2* expression and promotes PFC dendritic spine formation. These findings suggest a genetic and molecular basis for the anterior-posterior cortical gradient and disproportionate increase in the Hominini PFC of dendritic spines and a developmental mechanism that may link dysfunction of the NRXN–GRID–CBLN2 complex to the pathogenesis of neuropsychiatric disorders.

Increasing expansion of the PFC in catarrhine primates is associated with increased dendritic complexity and density of synaptic spines of pyramidal neurons compared to more posterior cortical association and sensory areas, creating an anterior–posterior gradient of synaptic density[28–30]. Moreover, this gradient is exaggerated in humans by the disproportionately high number of spines on pyramidal neurons in the human PFC compared to the PFC of other analysed primates[5]. Early neocortical synaptogenesis begins during the human midfetal period or equivalent developmental age in other mammals[31], proceeding in an anterior–posterior direction[32] and continues at an accelerated rate well beyond birth[33]. In our accompanying study[34], we identified a mid-fetal PFC-enriched gradient of retinoic acid (RA) concentration and pattern of gene expression in the human and macaque neocortex. Among the identified genes was the synaptic organizer *CBLN2*, which had the greatest loading of differentially expressed genes representing the anterior–frontal-posterior–temporal axis[35]. Members of the CBLN family encode secreted neuronal glycoproteins, which serve as excitatory and inhibitory synaptic-organizing molecules[21–27]. Thus, we hypothesize that RA-mediated regulation of *CBLN2* expression in the developing PFC underlies the disproportionate number of dendritic spines in the human PFC.

Divergence of *CBLN2* laminar expression

By analysing RNA sequencing data from eleven neocortical areas across human and macaque development from the BrainSpan and PsychENCODE projects[15,20], we identified a precocious and transient increase in human (1.9-fold change) and macaque (2.0-fold change) *CBLN2* expression in the major areas of the prospective PFC compared to the seven analysed non-PFC areas during midfetal development (post-conception weeks 13–24) or periods 4–6 according to Kang et al.[35] (Fig. 1a, Extended Data Fig. 1a, b). Around the time of late infancy, both human and macaque *CBLN2* expression levels are comparable across analysed neocortical areas, despite modest but statistically significant increased expression in PFC areas compared to non-PFC areas (human periods 7–10: 1.1-fold change; macaque periods 7–10: 1.25-fold change; Fig. 1a, Extended Data Fig. 1a), which continued at later adult ages (human periods 11–14: 1.1-fold change; macaque periods 11–14: 1.2-fold change; Extended Data Fig. 1a, c). Analysis of the spatiotemporal expression of *CBLN2* paralogues within each species revealed that *CBLN1* was upregulated in the human and macaque midfetal PFC, compared to non-PFC areas whereas *CBLN4* exhibited human-specific upregulation (Extended Data Fig. 2). While none of the binding partners

[1]Department of Neuroscience, Yale School of Medicine, New Haven, CT, USA. [2]Yale Child Study Center, New Haven, CT, USA. [3]Department of Genetics, Yale School of Medicine, New Haven, CT, USA. [4]Department of Neurology, Yale School of Medicine, New Haven, CT, USA. [5]Center for Neuroscience and Regeneration Research, Yale University, New Haven, CT, USA. [6]Rehabilitation Research Center, Veterans Affairs Connecticut Healthcare Center, West Haven, CT, USA. [7]Department of Comparative Medicine, Yale School of Medicine, New Haven, CT, USA. [8]Department of Psychiatry, Yale School of Medicine, New Haven, CT, USA. [9]Program in Cellular Neuroscience, Neurodegeneration and Repair, New Haven, CT, USA. [10]Kavli Institute for Neuroscience, Yale University, New Haven, CT, USA. ✉e-mail: nenad.sestan@yale.edu

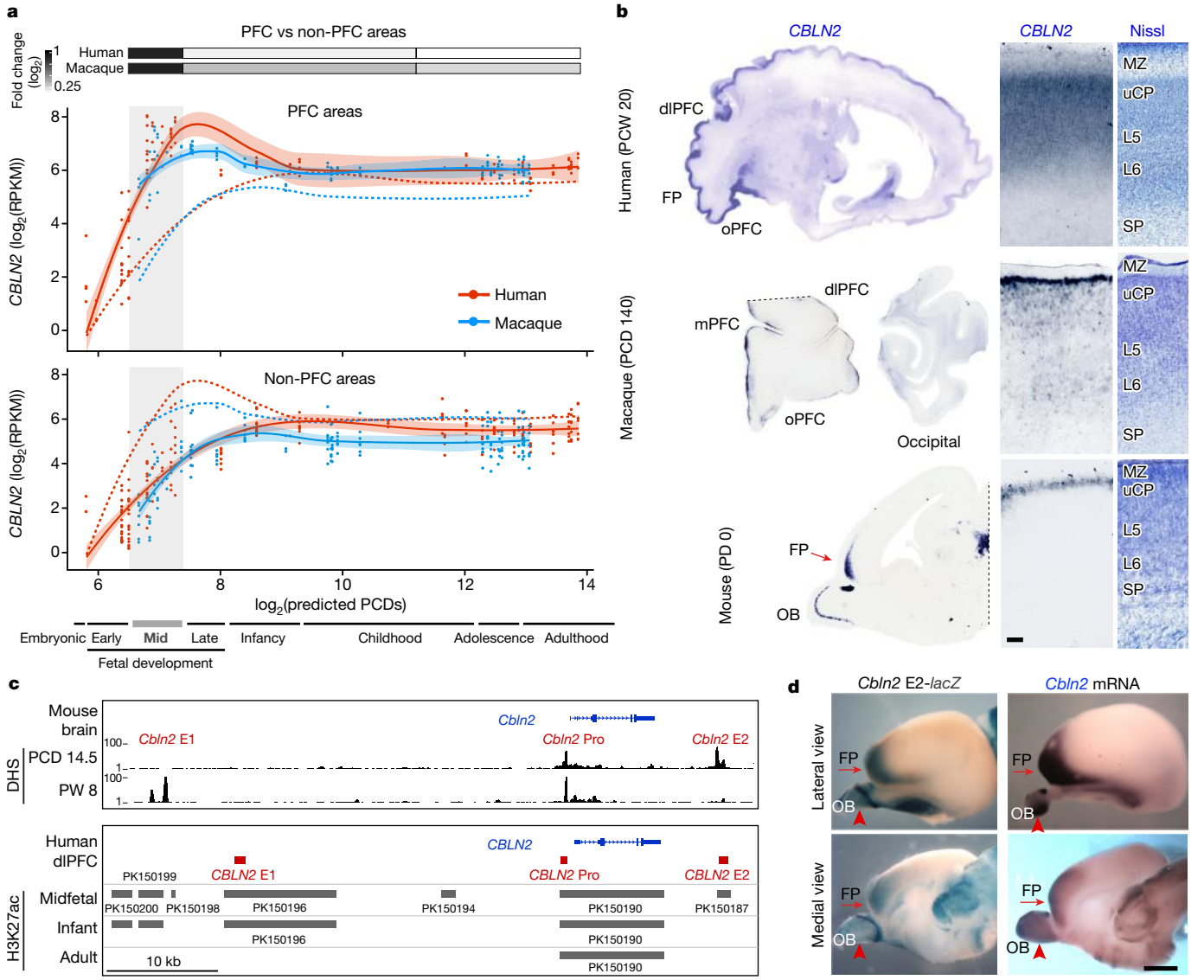


Fig. 1 | Phylogenetic, transcriptomic and regulatory characterization of developmental PFC upregulation of *CBLN2*. **a**, *CBLN2* spatiotemporal expression in human (red lines) and macaque (blue lines) PFC (solid lines) and non-PFC (dotted lines) areas. Quantification of the fold change between PFC and non-PFC expression for periods 4–6 (grey box), 7–10 and 11–14 (two-tailed Mann–Whitney test or two-tailed, unpaired *t*-test $P < 1 \times 10^{-4}$ for all comparisons) is shown above the plots. For all of the plots, the shading around the lines represents the 95% confidence interval. Developmental periods designed by Kang et al.[35] are shown below the plots. RPKM, reads per kilobase of transcript per million reads mapped. **b**, *CBLN2* in situ hybridization on sagittal sections of midfetal human and neonatal mouse and coronal sections of midfetal macaque prefrontal and occipital cortices. The right panels show a laminar *CBLN2* expression pattern in the PFC and Nissl stain of the adjacent section. The human *CBLN2* image is from humanbraintranscriptome.org[18]. Experiments were repeated three times for macaque brain sections and more than five times for mouse brain sections. Scale bar, 500 µm. FP, frontal pole; MZ, marginal zone; OB, olfactory bulb; oPFC, orbital PFC; PCW, post-conception week; SP, subplate; uCP, upper cortical plate. **c**, DNase I-hypersensitive sites (DHSs) in the mouse brain at PCD 14.5 and post-natal week (PW) 8 within 50 kb either upstream or downstream of the *Cbln2* gene (top). H3K27ac ChIP sequencing data from the human dlPFC/DFC at midfetal age, infancy and adulthood is shown (bottom). Putative *cis*-regulatory elements were designated as *CBLN2* E1, E2 and Pro. **d**, β-Galactosidase activity in the transgenic mouse brain carrying mouse *Cbln2* E2 conjugated with the *lacZ* reporter at PCD 17. Endogenous *Cbln2* expression in age-matched brains by in situ hybridization is shown for comparison. The arrows depict the frontal pole, and the arrowheads indicate the olfactory bulb. Scale bar, 1 mm. Experiments were repeated for three brains and two replicates are shown in Extended Data Fig. 3c.

of the CBLN family members, *NRXN1*, *NRXN2*, *NRXN3* and *GRID1*, *GRID2*, showed distinct PFC upregulation during midfetal development, all were expressed in the PFC with increasing levels of expression during development (Extended Data Fig. 2c).

Previous bulk tissue transcriptomic studies have also reported an anterior–posterior gradient of *CBLN2* expression in the human midfetal neocortex[18,36,37]. We confirmed this finding using in situ hybridization (post-conception weeks 21 and 22) and single-cell RNA sequencing data, showing that *CBLN2* was detected broadly in neurons of fetal prospective upper layers (fL) 2–4 and deeper L5 and L6 of the PFC, but faintly posteriorly (Fig. 1b, Extended Data Fig. 1b, d). We also identified an anterior–posterior gradient of *CBLN2* expression in macaque (post-conception day (PCD) 110 and 140) and mouse (post-natal day (PD) 0) at equivalent developmental ages (Fig. 1b, Extended Data Figs. 1b, 3a, b). In macaque, *CBLN2* expression was mostly observed in the prospective upper layers with noticeably weaker expression in the deeper layers (Fig. 1b). In mice, *Cbln2* was mostly restricted to the prospective upper layers (Fig. 1b, Extended Data Fig. 3b). Together, our

analysis found a progressive extension of *CBLN2* expression into the deeper layers of the midfetal PFC in humans compared to macaques, and in macaques compared to mice.

Hominini-specific deletions in *CBLN2* E2 enhancer

To understand regulatory mechanisms of *CBLN2* expression in the midfetal PFC, we analysed publicly available datasets on the regulatory landscape of developing mouse and human brains generated by the ENCODE and PsychENCODE projects[20,38]. Around the mouse *Cbln2* locus, we identified three putative *cis*-regulatory elements marked as DNase I-hypersensitive sites, which we designated as *Cbln2* enhancer 1 (E1; 1,452 bp), enhancer 2 (E2; 1,005 bp) and promoter (Pro; 316 bp; Fig. 1c, top). The sequences of these putative *cis*-regulatory elements are moderately conserved between mouse and human (E1, 76.3%; E2, 84.3%; and Pro, 70.0%). Of note, mouse *Cbln2* E2 showed DNase I-hypersensitivity peaks only in PCD 14.5 samples and *Cbln2* E1 only in 8-post-natal-weeks samples (Fig. 1c, top). Analysis of genomic sites with differential distribution of H3K27ac, a mark of active enhancers, in the developing and adult human dorsolateral PFC (dlPFC (also known as the DFC)) indicates that *CBLN2* Pro is active at all three analysed time points (midfetal, early infancy and adult ages), whereas *CBLN2* E1 is active during midfetal age and infancy, and *CBLN2* E2 is only active during midfetal age (Fig 1c, bottom).

We next assessed the activity of these putative regulatory regions using transgenic mouse lines in which a *lacZ* expression cassette was placed downstream of mouse *Cbln2* E1, *Cbln2* E2 or *Cbln2* Pro. While *Cbln2* E1-*lacZ* and *Cbln2* Pro-*lacZ* transgenic lines exhibited no detectable *lacZ* expression (0 of 10 founders) in the developing cortex between PCD 17 and 18, 50% of *Cbln2* E2-*lacZ* transgenics (3 of 6 founders) showed *lacZ* expression, with two of three independent lines recapitulating endogenous frontal cortex-enriched *Cbln2* expression (Fig. 1d, Extended Data Fig. 3c). These findings indicate that *Cbln2* E2 is an enhancer that drives expression in the mouse neonatal frontal cortex.

Given the presence of *CBLN2* E2 in human, macaque and mouse, we hypothesized that sequence differences between the orthologues may underlie the species-specific expression pattern of *CBLN2*. Comparative genomic and phylogenetic analysis identified three separate deletions (122 bp, 84 bp and 20 bp) in the sequence of the human *CBLN2* E2 orthologue compared to the sequence of mouse *Cbln2* E2 (Fig. 2a). Further comparison across multiple vertebrates revealed that two of the three deleted sequences are jointly absent only in extant members of the Hominini clade (human, common chimpanzee and bonobo) (Extended Data Figs. 4a, c, 5). The first two of these sequences, which we named Hominini-specific deletion 1 and 2 (HSD1 and HSD2; Fig. 2a), are highly conserved among other analysed Haplorhini (gorilla and macaque), Strepsirrhini (lemur) and mouse (Fig. 2a, Extended Data Figs. 4a, c, 5). The third deleted sequence (I3) was detected only in mouse and rat of the Muroidea species analysed (Extended Data Figs. 4a, 5). We did not detect genomic regions orthologous to *Cbln2* E2 in analysed non-placental mammals and non-mammalian chordates, although they do possess the *Cbln2* gene (Extended Data Fig. 5).

In our accompanying study[34], we found that the RA receptors *RXRG* and *RARB* are upregulated in the developing frontal lobe and that *Cbln2* expression was reduced in neonatal *Rxrg* and *Rarb* double knockout mice. Thus, we screened for and identified multiple putative RAR–RXR tandem binding sites in *CBLN2* E2 (Fig 2a, Extended Data Fig. 4a). Consistent with these findings, we found that overexpression of RXRG and RARB in the Neuro2a cell line activated human, chimpanzee, gorilla, macaque and mouse orthologues of *CBLN2* E2 in in vitro luciferase reporter assays more than overexpression of either individually (Fig. 2b, Extended Data Fig. 6a), consistent with previous reports showing that RA receptors exert their effects as heterodimers[39,40].

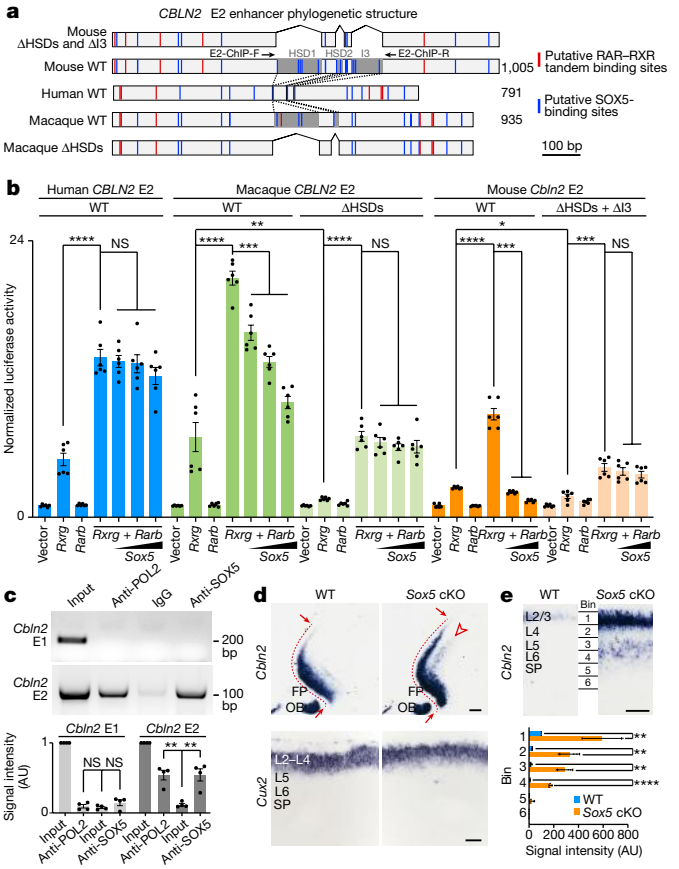


Fig. 2 | SOX5 represses *CBLN2* through Hominini-specific regulatory deletions. a, Schematics of *Cbln2* E2 constructs used for the luciferase assays. The shaded dark grey areas indicate HSD1–HSD2 and I3 (mouse) and HSD1–HSD2 (macaque). All HSDs and I3 were deleted from mouse *Cbln2* E2 in ΔHSDs and ΔI3, and all HSDs were deleted from macaque *CBLN2* E2 in macaque ΔHSDs. The red and blue lines indicate putative RAR–RXR tandem binding sites and SOX5-binding sites, respectively. E2-ChIP-F and E2-ChIP-R represent ChIP–PCR primers (**c**) (Supplementary Table 2). **b**, *Cbln2* E2 luciferase assay in Neuro2a cells with mouse *Rxrg* and *Rarb* and increasing concentrations of *Sox5*. Two-tailed Student's *t*-test; $^{*}P = 0.01$; $^{**}P = 0.001$; $^{***}P = 4 \times 10^{-4}$ (macaque), 4×10^{-4} (mouse), 7×10^{-5} (mouse ΔHSDs + ΔI3); $^{****}P < 1 \times 10^{-5}$; NS, not significant. Centre value, average; error bars indicate s.e.m.; $n = 6$ per condition. **c**, ChIP–PCR assays in the PD 0 mouse neocortex using anti-RNA polymerase II (anti-POL2), IgG, anti-SOX5 and *Cbln2* E1 and E2 primers; the quantifications of signal intensity are below. Two-tailed Student's *t*-test; $^{**}P = 0.001, 0.003$; Centre value, average; error bars indicate s.e.m.; $n = 4$ per condition. AU, arbitrary units. **d**, *Cbln2* expression in PD 0 *Sox5* conditional knockout (cKO) compared to *Sox5*$^{+/+}$; *Emx1–Cre* (WT) cortex. $n = 3$ per genotype. The arrows indicate the posterior spread of *Cbln2* expression and the arrowheads indicate ectopic deep layer expression. Scale bars, 200 µm (top); 100 µm (bottom). **e**, Quantification of cortical *Cbln2* signal intensity in six bins spanning from the pia to the ventricular zone (top) ($n = 3$). Two-tailed Student's *t*-test; $^{****}P = 2 \times 10^{-5}$ (bin 4); $^{**}P = 3 \times 10^{-3}$ (bin 1), 2×10^{-3} (bin 2), 1×10^{-3} (bin 3). Centre value, average; error bars indicate s.e.m. Scale bar, 100 µm.

Repression of *CBLN2* E2 by SOX5

In addition to RA receptors, analysis of *Cbln2* E2 also identified multiple conserved putative SOX5-binding sites, including three in HSD1 and a fourth in HSD2 (Fig. 2a, Extended Data Fig. 4a, b), as well as non-conserved sites in mouse HSD1 and HSD2 orthologues, and in I3 (Fig. 2a). Screening for gained putative human and Hominini-specific transcription factor-binding sites identified only a few generic transcription factor-binding sites (Supplementary Table 1). Thus, we focused on putative SOX5-binding sites given their loss in Hominini E2 and the role of SOX5 in the specification and development of cortical

projection neurons in the deep layers[41,42]. We found that mouse *Cbln2* E2 activation by the RXRG–RARB heterodimer was suppressed by SOX5 in a dose-dependent manner, whereas human and chimpanzee *CBLN2* E2 showed no statistically significant suppression by SOX5 at the same doses (Fig. 2b, Extended Data Fig. 6a). Differential suppression of *Cbln2* E2 activation was not a consequence of sequence differences between mouse and human SOX5, as activation of mouse *Cbln2* E2 was suppressed by both mouse and human SOX5 (Fig. 2b, Extended Data Fig. 6b). Activation of macaque and gorilla *CBLN2* E2 showed intermediate suppression of activity at the same doses of SOX5 (Fig. 2b, Extended Data Fig. 6a). Deletion of HSD1, HSD2 and I3 in mouse, or HSD1 and HSD2 in macaque, abolished suppression by SOX5, suggesting that these regions mediate suppression of SOX5 (Fig. 2b). Testing for direct in vivo binding using chromatin immunoprecipitation followed by PCR (ChIP–PCR) assays found that endogenous SOX5 bound *Cbln2* E2 but not *Cbln2* E1 (Fig. 2c).

To understand the functional importance of SOX5 repression of *Cbln2* E2 in vivo, we examined *Cbln2* cortical expression by creating a cortex-specific *Sox5* conditional knockout (*Sox5* loxP/loxP; *Emx1–Cre*) mouse. As predicted by luciferase assays, the absence of SOX5 led to ectopic expression of *Cbln2* in deeper layers of the neonatal frontal cortex (Fig. 2d, e, Extended Data Fig. 6c). There was also an increase and posterior extension of *Cbln2* expression in the upper layers (Fig. 2d, e, Extended Data Fig. 6c), probably due to previously described transient expression of *Sox5* in these layers[41]. Consistent with these results, the 5′ segment of *Cbln2* E2 lacking the two HSD sequences and I3 (*Cbln2* E2-Fr1) was not suppressed by SOX5 in the luciferase assay (Extended Data Fig. 6d, f). Furthermore, a transgenic mouse line expressing *lacZ* under the transcriptional control of *Cbln2* E2-Fr1 showed deep layer expression of *lacZ* compared to *Cbln2* E2 (Extended Data Fig. 6d, e). Together, these observations suggest that SOX5 directly represses *Cbln2* expression in the deep layers of a neonatal mouse and midfetal macaque (and probably other non-hominin primates) frontal cortex by binding to HSD sites in *Cbln2* E2.

Human *CBLN2* E2 promotes spinogenesis

To test whether human *CBLN2* E2 is sufficient to drive a human-like pattern of *Cbln2* expression in the mouse neonatal frontal cortex, we generated a knock-in mouse in which *Cbln2* E2 was replaced in situ (in the orthologous position) with the corresponding human *CBLN2* E2 using the CRISPR–Cas9 genome editing technique (Extended Data Fig. 7a–c). The engineered mice carrying homozygous h*Cbln2* E2 in both alleles (*Homo sapiens*;*Homo sapiens* (*Hs;Hs*)) were viable and fertile, and the expression of the cortical upper layer-specific marker *Cux2* and SOX5 were comparable to homozygous wild-type (WT) *Cbln2* E2 (*Mus musculus;Mus musculus* (*Mm;Mm*)) mice (Fig. 3c), indicating that genetically humanized h*Cbln2* E2 grossly functions as the mouse orthologue. Neocortical *Cbln2* expression at PD 0 was increased by 21.7% in h*Cbln2* E2 (*Hs;Hs*), as compared to WT by quantitative reverse transcription–PCR (Extended Data Fig. 7d). Analysis of expression by in situ hybridization showed that *Cbln2* was also transiently upregulated in both prospective upper and deeper layers of the h*Cbln2* E2 PFC at PCD 18 and PD 1, but not at PCD 16 (Fig. 3a, b, Extended Data Fig. 8). After PD 1, expression of *Cbln2* extended into the deeper layers of the PFC, coinciding with the downregulation of SOX5 protein expression (Extended Data Fig. 9) in both h*Cbln2* E2 and WT *Cbln2* E2.

Members of the CBLN family have been reported to regulate synapse development and function[21–27]. Therefore, we hypothesized that a transient increase in expression of *Cbln2* in the h*Cbln2* E2 PFC will lead to an increased density of post-synaptic structures and synapses. We quantified both excitatory and inhibitory post-synaptic density in the medial PFC (mPFC), primary somatosensory area (SSp) and primary visual (VISp) area using immunostaining for the post-synaptic proteins PSD95 (also known as DLG4) and gephyrin (GPHN), respectively.

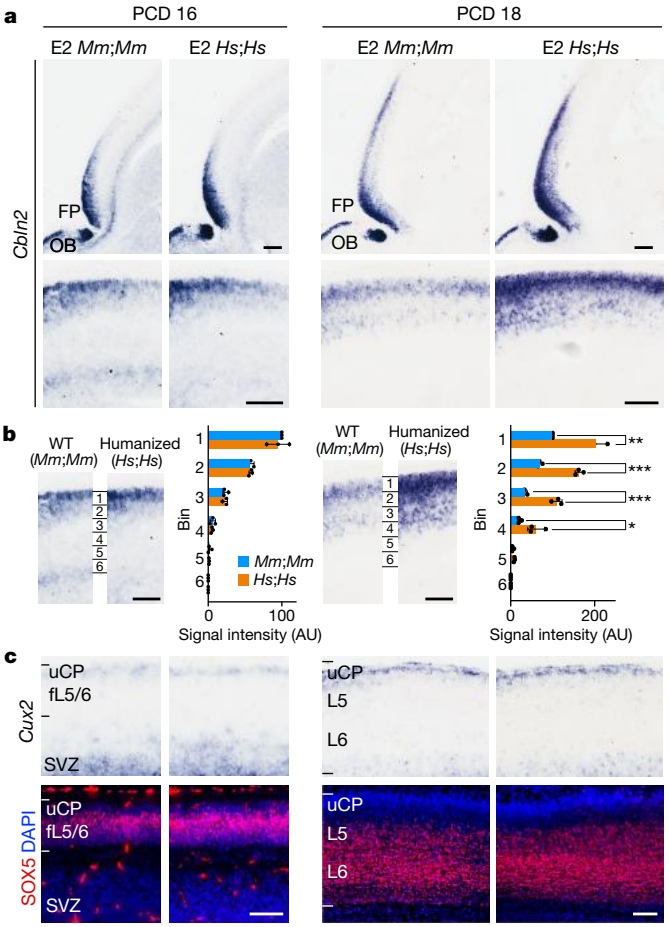


Fig. 3 | In situ genetic humanization of the mouse *Cbln2* enhancer drives upregulation and ectopic laminar *Cbln2* expression in the neonatal frontal cortex. a, *Cbln2* expression in h*Cbln2* E2 knock-in mouse (*Hs;Hs*) brain at PCD 16 and PCD 18 compared to WT mouse (*Cbln2* E2, *Mm;Mm*) with higher magnification panels shown below. Scale bars, 200 µm (top); 100 µm (bottom). Experiments were repeated for three brains and two replicates are shown in Extended Data Fig. 8. **b**, The neocortex was divided into six equal bins spanning from the pia to the ventricular zone (left panels), and the signal intensity of *Cbln2* was quantified for each bin and compared between WT and h*Cbln2* E2 (*Hs;Hs*). Two-tailed Student's *t*-test; $^{*}P = 0.02$ (PCD 18), 0.02, 0.04 (PD 1), $^{**}P = 1 \times 10^{-3}$ (PCD 18), 1×10^{-3}, 3×10^{-3}, 3×10^{-3} (PD 1), $^{***}P = 1 \times 10^{-4}$, 2×10^{-4} (PCD 18); Centre value, average; error bars indicate s.e.m. $n = 3$ brains per genotype. Scale bars, 100 µm. **c**, Expression of the upper layer marker *Cux2* and SOX5 in adjacent tissue sections was detected by in situ hybridization and immunostaining, respectively. Scale bars, 100 µm. $n = 3$ brains per genotype. DAPI, 4′,6-diamidino-2-phenylindole; SVZ, subventricular zone.

In h*Cbln2* E2 (*Hs;Hs*) mice, we observed an increase in the density of PSD95-immunopositive excitatory post-synaptic puncta in both upper (39.5%) and deep (47.9%) layers in the mPFC but not in the SSp and VISp, where *Cbln2* expression was not altered, at PD 0 (Extended Data Fig. 10a, b). A similar increase in PSD95-immunopositive puncta in the mPFC was identified in the *Sox5* conditional knockout mice (58% for upper layers, 40% for deep layers; Extended Data Fig. 10b). We also observed an increase in the density of GPHN-immunopositive inhibitory post-synaptic puncta number in both upper (24.2%) and deep (37.9%) layers at PD 0 in the h*Cbln2* E2 PFC but not in the SSp and VISp cortex (Extended Data Fig. 10c, d). In the *Sox5* conditional knockout mice, anti-GPHN immunostaining revealed a similar increase in the density of inhibitory post-synaptic puncta (27.3% for upper layers, 39.4% for deep layers; Extended Data Fig. 10d). At PD 60, an increase in the density of PSD95-immunopositive post-synaptic puncta was observed in deep

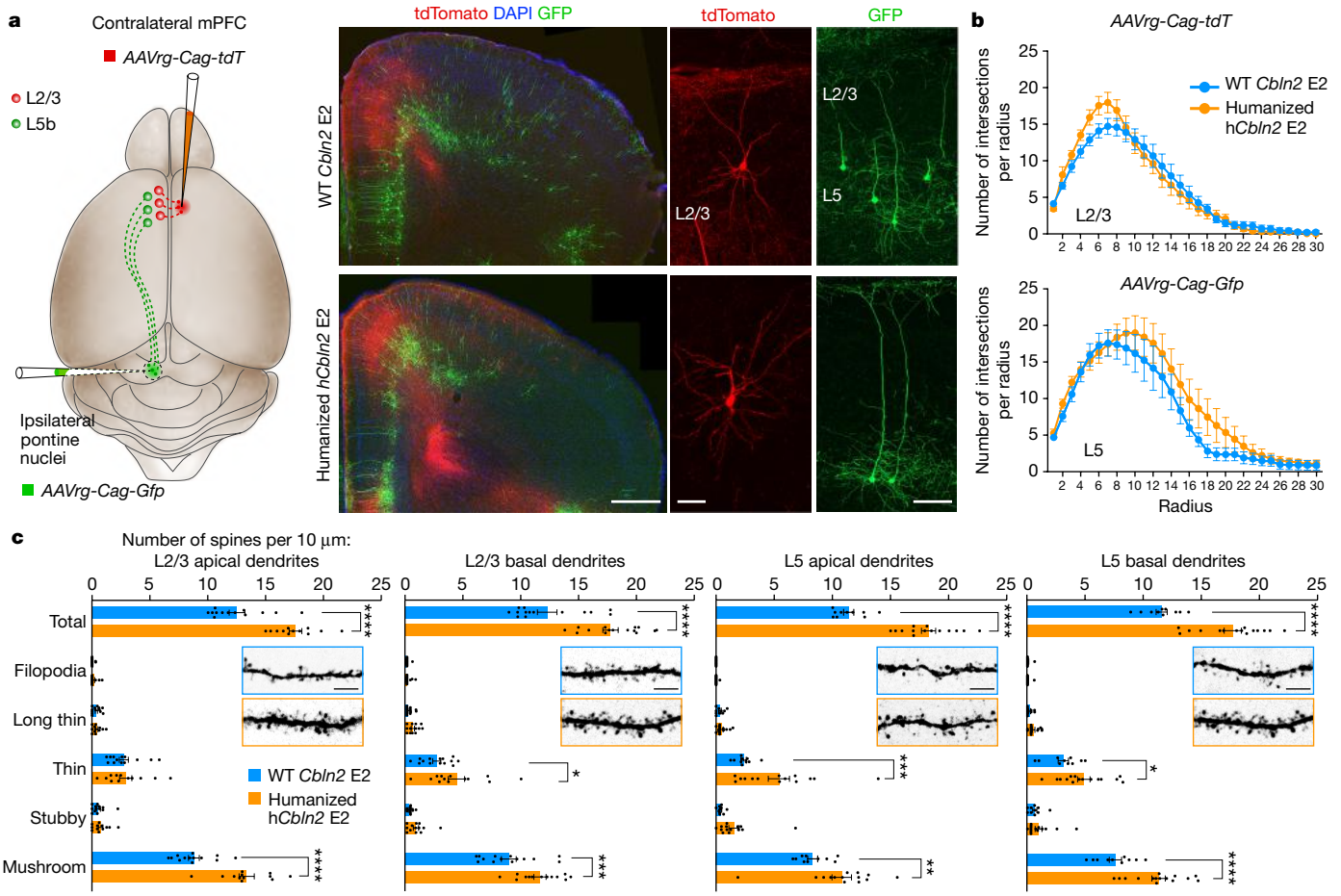


Fig. 4 | Increased density of dendritic spines in the mPFC of h*Cbln2* E2 mice. **a**, Retrogradely labelled callosal L2/3 (red; *AAVrg-Cag-tdT*) and subcerebral L5b (green; *AAVrg-Cag-Gfp*) PD 60 pyramidal neurons. Scale bars, 500 µm (left); 20 µm (right). Three brains from each genotype were injected. **b**, Sholl analysis of retrogradely labelled L2/3 (top) and L5 (below). Two-way analysis of variance (ANOVA) with Sidak's multiple comparison method was used. Centre value, average; error bars indicate s.e.m.; $n = 13$ (WT *Cbln2* E2) and 15 (h*Cbln2* E2) L2/3; $n = 11$ (WT *Cbln2* E2) and 15 (h*Cbln2* E2) L5 ($n = 3$ brains per genotype). **c**, Quantification of dendritic spines. Two-way ANOVA with Sidak's multiple comparison method was used; $^{*}P = 0.03$ (L2/3 basal dendrites), 0.03 (L5 basal dendrites); $^{**}P = 5 \times 10^{-3}$, $^{***}P = 4 \times 10^{-4}$ (L2/3 basal dendrites), 4×10^{-4} (L5 apical dendrites); $^{****}P < 5 \times 10^{-5}$. Centre value, average; error bars indicate s.e.m. $n = 13$ (WT *Cbln2* E2) and 12 (h*Cbln2* E2) L2/3 apical dendrites; $n = 14$ (WT *Cbln2* E2) and 15 (h*Cbln2* E2) L2/3 basal dendrites; $n = 10$ (WT *Cbln2* E2) and 14 (h*Cbln2* E2) L5 apical dendrites; $n = 10$ (WT *Cbln2* E2) and 15 (h*Cbln2* E2) L5 basal dendrites ($n = 3$ brains per genotype). Scale bars, 10 µm.

layers of h*Cbln2* E2 mice (21.4%), and no statistically significant changes were observed in the density of GPHN-immunopositive post-synaptic puncta (Extended Data Fig. 10b, d).

We confirmed the PFC-specific increase in excitatory post-synaptic structures in h*Cbln2* E2 (*Hs;Hs*) mice by quantifying dendritic spines using two different strategies. Using Golgi staining, we identified an increase in mature mushroom-shaped spines and long thin spines in both the upper and the deep layers of the h*Cbln2* E2 (*Hs;Hs*) mPFC (Extended Data Fig. 11a), but not in the SSP or VISp (Extended Data Fig. 11b, c). Similarly, we identified an increase in dendritic spines in apical and basal dendrites of both retrogradely labelled upper layer callosally projecting neurons and L5b subcerebrally projecting neurons of the h*Cbln2* E2 (*Hs;Hs*) mPFC compared to WT littermates (Fig. 4a, c). To assess whether dendritic spines in the WT and h*Cbln2* E2 (*Hs;Hs*) mPFC form synapses, we conducted 3D reconstruction from serial section electron microscopy at PD 60. We found that 91.4% of dendritic spines in the h*Cbln2* E2 mPFC had presynaptic innervation, similar to the *Cbln2* E2 control (Extended Data Fig. 11f, g), thus providing evidence against an increase in non-synaptic dendritic spines in h*Cbln2* E2 (*Hs;Hs*). Furthermore, we also identified a statistically significant increase in overlapping PSD95-immunolabelled post-synaptic and synaptophysin-immunolabelled presynaptic puncta in both upper (15.4%) and deep (38.6%) layers of the early post-natal h*Cbln2* E2 (*Hs;Hs*) mPFC (Extended Data Fig. 11d, e). Finally, as increase in dendritic spines and synapse number could be secondary to increased dendritic complexity, which has also been described in the human PFC[5], we compared the dendritic arbors in WT *Cbln2* E2 with h*Cbln2* E2. We found no statistically significant difference in dendritic complexity in dendritic arbors labelled with Golgi and retrograde viral tracers using Sholl analysis in the PFC, SSp or VISp (Fig. 4b, Extended Data Fig. 11a–c).

Discussion

In this study, we describe a potential CBLN2-based molecular mechanism and underlying genetic basis for the anterior–posterior gradient of neocortical synaptogenesis and disproportionate increase in dendritic spines within the human PFC. In our accompanying paper[34], we identified that loss of RA signalling in the developing mouse mPFC leads to both reduced expression of *Cbln2* and reduced density of excitatory dendritic spines. Regulation of *Cbln2* E2 by RXRG and RARB provides a direct link between these two findings, and may be part of a more complex gene regulatory network regulating development and adaptive changes in PFC connectivity. Furthermore, species-specific regulation of *CBLN2* expression is probably only one aspect of phylogenetic

differences in connectivity, including potential inter-hominin differences[6] and accelerated expression of synaptic genes in the developing human neocortex, compared to macaque[15].

While CBLN2 has been implicated in obsessive compulsive disorder, schizophrenia and Tourette syndrome[43], it serves as ligands for multiple NRXNs and GRID1, which are autism spectrum disorder-associated and schizophrenia-associated proteins, respectively[44,45]. Multiple lines of evidence have shown that autism spectrum disorder and schizophrenia risk genes converge in time, brain region, cell type and cell compartment, with evidence for convergence in synapses as well as cortical projection neurons and the frontal cortex during human midfetal development[46,47]. Consistent with these findings, alterations in the levels of some synaptic proteins and connectivity within the frontal lobe have been described as the possible underlying pathophysiology of both autism spectrum disorder and schizophrenia[43–48], with dysfunction in the NXRN–GRID trans-synaptic complex, which includes CBLNs, being a putative mechanism and target for interventions.

Online content

Any methods, additional references, Nature Research reporting summaries, source data, extended data, supplementary information, acknowledgements, peer review information; details of author contributions and competing interests; and statements of data and code availability are available at https://doi.org/10.1038/s41586-021-03952-y.

1. Finlay, B. L. & Darlington, R. B. Linked regularities in the development and evolution of mammalian brains. *Science* **268**, 1578–1584 (1995).
2. Barton, R. A. & Harvey, P. H. Mosaic evolution of brain structure in mammals. *Nature* **405**, 1055–1058 (2000).
3. Krubitzer, L. & Kaas, J. The evolution of the neocortex in mammals: how is phenotypic diversity generated? *Curr. Opin. Neurobiol.* **15**, 444–453 (2005).
4. Passingham, R. E. & Wise, S. P. *The Neurobiology of the Prefrontal Cortex: Anatomy, Evolution, and the Origin of Insight* (Oxford Univ. Press, 2015).
5. Elston, G. N. et al. Specializations of the granular prefrontal cortex of primates: implications for cognitive processing. *Anat. Rec. A Discov. Mol. Cell. Evol. Biol.* **288**, 26–35 (2006).
6. Semendeferi, K. et al. Spatial organization of neurons in the frontal pole sets humans apart from great apes. *Cereb. Cortex* **21**, 1485–1497 (2011).
7. Kwan, K. Y. et al. Species-dependent posttranscriptional regulation of NOS1 by FMRP in the developing cerebral cortex. *Cell* **149**, 899–911 (2012).
8. Gabi, M. et al. No relative expansion of the number of prefrontal neurons in primate and human evolution. *Proc. Natl Acad. Sci. USA* **113**, 9617–9622 (2016).
9. Caceres, M. et al. Elevated gene expression levels distinguish human from non-human primate brains. *Proc. Natl Acad. Sci. USA* **100**, 13030–13035 (2003).
10. Khaitovich, P. et al. Regional patterns of gene expression in human and chimpanzee brains. *Genome Res.* **14**, 1462–1473 (2004).
11. Uddin, M. et al. Sister grouping of chimpanzees and humans as revealed by genome-wide phylogenetic analysis of brain gene expression profiles. *Proc. Natl Acad. Sci. USA* **101**, 2957–2962 (2004).
12. Konopka, G. et al. Human-specific transcriptional networks in the brain. *Neuron* **75**, 601–617 (2012).
13. Bauernfeind, A. L. et al. Evolutionary divergence of gene and protein expression in the brains of humans and chimpanzees. *Genome Biol. Evol.* **7**, 2276–2288 (2015).
14. Sousa, A. M. M. et al. Molecular and cellular reorganization of neural circuits in the human lineage. *Science* **358**, 1027–1032 (2017).
15. Zhu, Y. et al. Spatiotemporal transcriptomic divergence across human and macaque brain development. *Science* **362**, eaat8077 (2018).
16. Pollen, A. A. et al. Establishing cerebral organoids as models of human-specific brain evolution. *Cell* **176**, 743–756 (2019).
17. Kanton, S. et al. Organoid single-cell genomic atlas uncovers human-specific features of brain development. *Nature* **574**, 418–422 (2019).
18. Johnson, M. B. et al. Functional and evolutionary insights into human brain development through global transcriptome analysis. *Neuron* **62**, 494–509 (2009).
19. Pletikos, M. et al. Temporal specification and bilaterality of human neocortical topographic gene expression. *Neuron* **81**, 321–332 (2014).
20. Li, M. et al. Integrative functional genomic analysis of human brain development and neuropsychiatric risks. *Science* **362**, eaat7615 (2018).
21. Urade, Y. et al. Precerebellin is a cerebellum-specific protein with similarity to the globular domain of complement C1q B chain. *Proc. Natl Acad. Sci. USA* **88**, 1069–1073 (1991).
22. Hirai, H. et al. Cbln1 is essential for synaptic integrity and plasticity in the cerebellum. *Nat. Neurosci.* **8**, 1534–1541 (2005).
23. Uemura, T. et al. Trans-synaptic interaction of GluRδ2 and neurexin through Cbln1 mediates synapse formation in the cerebellum. *Cell* **141**, 1068–1079 (2010).
24. Matsuda, K. et al. Cbln1 is a ligand for an orphan glutamate receptor δ2, a bidirectional synapse organizer. *Science* **328**, 363–368 (2010).
25. Yasumura, M. et al. Glutamate receptor delta1 induces preferentially inhibitory presynaptic differentiation of cortical neurons by interacting with neurexins through cerebellin precursor protein subtypes. *J. Neurochem.* **121**, 705–716 (2012).
26. Wei, P. et al. The Cbln family of proteins interact with multiple signaling pathways. *J. Neurochem.* **121**, 717–729 (2012).
27. Seigneur, E. & Sudhof, T. C. Genetic ablation of all cerebellins reveals synapse organizer functions in multiple regions throughout the brain. *J. Neurosci.* **38**, 4774–4790 (2018).
28. Elston, G. N. Pyramidal cells of the frontal lobe: all the more spinous to think with. *J. Neurosci.* **20**, RC95 (2000).
29. Jacobs, B. et al. Regional dendritic and spine variation in human cerebral cortex: a quantitative golgi study. *Cereb. Cortex* **11**, 558–571 (2001).
30. Bianchi, S. et al. Synaptogenesis and development of pyramidal neuron dendritic morphology in the chimpanzee neocortex resembles humans. *Proc. Natl Acad. Sci. USA* **110**, 10395–10401 (2013).
31. Molliver, M. E. et al. The development of synapses in cerebral cortex of the human fetus. *Brain Res.* **50**, 403–407 (1973).
32. Voigt, T. et al. Synaptophysin immunohistochemistry reveals inside-out pattern of early synaptogenesis in ferret cerebral cortex. *J. Comp. Neurol.* **330**, 48–64 (1993).
33. Rakic, P. et al. Concurrent overproduction of synapses in diverse regions of the primate cerebral cortex. *Science* **232**, 232–235 (1986).
34. Shibata, M. et al. Regulation of prefrontal patterning and connectivity by retinoic acid. *Nature* https://doi.org/10.1038/s41586-021-03953-x (2021).
35. Kang, H. J. et al. Spatiotemporal transcriptome of the human brain. *Nature* **478**, 483–489 (2011).
36. Lambert, N. et al. Genes expressed in specific areas of the human fetal cerebral cortex display distinct patterns of evolution. *PLoS ONE* **6**, e17753 (2011).
37. Miller, J. A. et al. Transcriptional landscape of the prenatal human brain. *Nature* **508**, 199–206 (2014).
38. ENCODE Project Consortium et al. An integrated encyclopedia of DNA elements in the human genome. *Nature* **489**, 57–74 (2012).
39. Chiang, M. Y. et al. An essential role for retinoid receptors RARβ and RXRγ in long-term potentiation and depression. *Neuron* **21**, 1353–1361 (1998).
40. Krezel, W. et al. Impaired locomotion and dopamine signaling in retinoid receptor mutant mice. *Science* **279**, 863–867 (1998).
41. Kwan, K. Y. et al. SOX5 postmitotically regulates migration, postmigratory differentiation, and projections of subplate and deep-layer neocortical neurons. *Proc. Natl Acad. Sci. USA* **105**, 16021–16026 (2008).
42. Shim, S. et al. *Cis*-regulatory control of corticospinal system development and evolution. *Nature* **486**, 74–79 (2012).
43. Clarke, R. A. & Eapen, V. Balance within the neurexin trans-synaptic connexus stabilizes behavioral control. *Front. Hum. Neurosci.* **8**, 52 (2014).
44. State, M. W. & Sestan, N. The emerging biology of autism spectrum disorders. *Science* **337**, 1301–1303 (2012).
45. Sudhof, T. C. Synaptic neurexin complexes: a molecular code for the logic of neural circuits. *Cell* **171**, 745–769 (2017).
46. Willsey, A. J. et al. Coexpression networks implicate human midfetal deep cortical projection neurons in the pathogenesis of autism. *Cell* **155**, 997–1007 (2013).
47. Gulsuner, S. et al. Spatial and temporal mapping of de novo mutations in schizophrenia to a fetal prefrontal cortical network. *Cell* **154**, 518–529 (2013).
48. Lewis, D. A. & Mirnics, K. Transcriptome alterations in schizophrenia: disturbing the functional architecture of the dorsolateral prefrontal cortex. *Prog. Brain Res.* **158**, 141–152 (2006).

Methods

Analysis of human and macaque transcriptomic data

Developing human and macaque brain RNA sequencing data (counts file) with its metadata information was obtained from BrainSpan (brainspan.org) and PsychENCODE (development.psychencode.org and evolution.psychencode.org) projects[15,20]. The timeline of human and macaque development and associated periods were designed by Kang et al.[35]. Predicted ages for macaque samples were calculated via the TranscriptomeAge algorithm described in Zhu et al.[15]. To perform statistical comparisons, samples from various developmental periods were grouped (periods 4–6, 7–10 and 11–14) and a two-tailed Student's *t*-test was used to compare gene expression levels between brain regions and species.

Animals

All experiments using animals were performed in accordance with protocols approved by the Yale University Institutional Animal Care and Use Committee (IACUC). All studies using mice (*M. musculus*) and rhesus macaques (*Macaca mulatta*) were performed in accordance with protocols approved by the Yale University's IACUC and US National Institutes of Health guidelines. The animals were housed, and timed-pregnant pre-natal and post-natal mouse and monkey brains were obtained in-house at the Yale Animal Resource Center. Mice were reared in group housing fewer than five mice per cage at 25 °C and 56% humidity in a 12-h light:12-h dark cycle and provided food and water ad libitum with veterinary care provided by the Yale Animal Resource Center. Animals were maintained on the C57BL/6J background. Both sexes were used and randomly assigned for all experiments. The day on which a vaginal plug was observed in mice was designated as PCD 0.5. Mice carrying the floxed *Sox5* allele were a kind gift from V. Lefebvre[49]. No sample size calculation were performed. Data collection was performed by independent investigators. Before data analysis, all experiments were randomized and analysed by independent, blinded observers. Sex of mouse samples used was not specified in this study.

Plasmid construction

For construction of expression vectors used for luciferase assays, full-length cDNAs (human *SOX5*, clone ID 30343519; mouse *Rarb*, clone ID 30608242; mouse *Rxrg*, clone ID 5707723; all purchased from GE Healthcare) were inserted into the pCAGIG vector (Addgene plasmid #11159). For luciferase reporter plasmids, human, chimpanzee, gorilla, macaque and mouse *Cbln2* E2 fragments were PCR-amplified from genomic DNA of individual species and inserted into the pGL4.24 vector (cat. no. E8421, Promega). The sequences of the PCR primers and synthetic oligonucleotides are listed in Supplementary Table 2.

Post-mortem human and macaque tissue

This study was conducted using post-mortem human brain specimens or RNA sequencing data generated previously[20] from tissue collections at the Department of Neuroscience at Yale School of Medicine, the Brain and Tissue Bank for Developmental Disorders at the University of Maryland, the Clinical Brain Disorders Branch of the National Institute of Mental Health, the Human Fetal Tissue Repository at the Albert Einstein College of Medicine, the Birth Defects Research Laboratory at the University of Washington (R24HD000836), Advanced Bioscience Resources and the Joint MRC–Wellcome Trust Human Developmental Biology Resource (www.hdbr.org; MR/R006237/1). Tissue was collected after obtaining parental or next of kin consent and with approval by the institutional review boards at each institution from which tissue specimens were obtained, the Yale University and the National Institutes of Health. Donated deidentified tissue was handled in accordance with ethical guidelines and regulations for the research use of human brain tissue set forth by the National Institutes of Health (https://oir.nih.gov/sites/default/files/uploads/sourcebook/documents/ethical_conduct/guidelines-biospecimen.pdf) and the WMA Declaration of Helsinki (https://www.wma.net/policies-post/wma-declaration-of-helsinki-ethical-principles-for-medical-research-involving-human-subjects/). All available non-identifying information was recorded for each specimen. No obvious signs of neuropathological alterations were observed in any of the human or macaque specimens analysed in this study. The post-mortem interval was defined as hours between time of death and time when tissue samples were fresh frozen or started to undergo fixation process.

Transcription factor-binding site analysis

The software FIMO from the MEME suite (version 4.12.0) was used to identify putative transcription factor-binding sites (TFBSs) in the human sequence of *CBLN2* E2. For this analysis, a set of transcription factor motifs from the JASPAR 2018 dataset (http://jaspar2018.genereg.net/)[50] were used. Then, MAF (Multiple Alignment Format) files were downloaded from the multiz100way table of the UCSC genome browser. An in-house script was used to extract the sequence of each TFBS in each species from the MAF file and identify mutations in core positions of the motif. TFBSs were considered conserved in a particular species if they had no mutations in core positions and were considered present in a clade if the TFBS was conserved in at least 25% of the species in the clade. To check for potential loss of TFBS regulatory mechanisms due to the deletions of HSD1, HSD2 and I3, we looked for TFBSs that are found in these regions and conserved in multiple species that do not have the deletions. To do so, the software FIMO was used to identify instances of transcription factor-binding motifs from the JASPAR database. This was performed in two species for each deletion (mouse and macaque for HSD1 and HSD2, and mouse and rat for I3). Then, the coordinates of putative TFBSs in the regions of interest in macaque and rat were translated to mm10 coordinates using LiftOver (downloaded on 3 May 2017) and manually compared to the TFBSs identified in the mouse sequence. Instances of the same transcription factor-binding motif at the same location in both datasets were considered putatively conserved. SOX5-binding sites and RA receptor-binding sites of mouse and human were identified using JASPAR 2018 (ref. [51]). The binding sites from other animal species were identified by conservation. RA receptor-binding sites include those of RARA::RXRA, RARA::RXRG, RARA, RARB, RARG, RXRA, RXRB and RXRG.

Generation of transgenic reporter mice

Cbln2 E2 and its related fragments were PCR-amplified from mouse genomic DNA and ligated into pBgn-*lacZ*[52]. The vector was linearized with BglII and KpnI and purified by gel separation followed by phenol/chloroform extraction. The transgene fragment was diluted into microinjection buffer (5 mM Tris-HCl and 0.1 mM EDTA) and injected into the pronucleus of fertilized eggs from B6SJLF1/J mouse purchased from The Jackson Laboratory. Transgenic mice were screened for the presence of transgenes by PCR using the *lacZ* primer set listed in Supplementary Table 2.

Generation of h*Cbln2* E2 knock-in mice

The overall strategy for the generation of h*Cbln2* E2 followed a previously described protocol[53]. The targeting vector carrying mouse h*Cbln2* E2 was constructed as follows: mouse genomic DNA fragments of 1,601 bp (chromosome 18 (chr18): 86,721,709–86,723,309, GRCm38/mm10) and 1,909 bp (chr18: 86,724,521–86,726,429, GRCm38/mm10) flanking the region containing *Cbln2* E2 (chr18: 86,723,310–86,724,520, GRCm38/mm10) were PCR-amplified as left and right arm, respectively, using mouse genomic DNA as template. An 852-bp fragment containing human *CBLN2* E2 (chr18: 72,530,473–72,531,324, GRCh38/hg38) was PCR-amplified using human genomic DNA as a template. These three DNA fragments were ligated into *XhoI*, *HindIII* and *ClaI* sites of the PL451 vector[54]. For the construction of the templates of guidance RNA,

two sets of top and bottom strand oligomers (Supplementary Table 2) directing the double-strand break at the mouse left and right arms were annealed and ligated into the *BbsI* site of the pX330-U6-Chimeric_BB-CBh-hSpCas9 vector[55], which was purchased from Addgene (plasmid no. 42230). After amplification of the insert with T7-tagged primers (Supplementary Table 2), guidance RNAs were synthesized by T7 RNA polymerase. The coding region of Cas9 was PCR-amplified using pX330-U6-Chimeric_BB-CBh-hSpCas9 as a template and inserted into the pSP64 poly(A) vector (cat. no. 1241, Promega). The vectors were digested and linearized with EcoRI. *Cas9* mRNA was synthesized by SP6 RNA polymerase. Guidance RNAs and *Cas9* mRNA were purified by MEGAclear Transcription Clean-Up Kit (cat. no. AM1908, Ambion). The targeting vector, *Cas9* mRNA, and two guidance RNAs were mixed at a concentration of 10 ng, 100 ng, 100 ng and 200 ng μl^{-1}, respectively, in the microinjection buffer (5mM Tris-HCl pH 7.5, and 0.1 M EDTA) and injected into the pronuclei of fertilized eggs from the B6SJLF1/J mouse strain, which was purchased from The Jackson Laboratory. The fertilized eggs were then transferred to the uterus of female CD-1 mice, purchased from Charles River Laboratories. The first generation (F0) mice with recombined allele were identified by long-distance PCR with a primer set (mP3/mP4) designed outside of the targeting vector (Supplementary Table 2, Extended Data Fig. 7a), and confirmed by sequencing. The germline transmission in the F1 generation was confirmed by nested PCR using the primer set of mP3/mP4, followed by hP1/P2 (Extended Data Fig. 7b, c), to exclude the possibility of detecting the targeting vector randomly integrated into the genomic DNA. Mice in the following generation were genotyped by PCR with mP1/mP2 and hP1/P2 as indicated in Extended Data Fig. 7. Mice lines were maintained on the C57BL6/6J background (The Jackson Laboratory).

In situ hybridization

Whole-mount and section in situ hybridization were performed as previously described[56]. Antisense digoxigenin (DIG)-labelled RNA probes were synthesized using DIG or Fluorescein RNA Labeling Mix (cat. no. 11277073910, Roche). Human and mouse *Cbln2* cDNA, and mouse *Cux2* cDNA were obtained from GE Healthcare (Clone ID 5727802, 6412317 and 30532644, respectively). Macaque *CBLN2* cDNA was synthesized using total RNA from the adult macaque dlPFC/DFC region. The PCR fragment was ligated into the pCRII vector (cat. no. K206001, Thermo Fisher Scientific). Fresh frozen archival human (PCW 21 and 22; post-mortem interval < 4 h; sex was not determined) and macaque (PCD 110 and 140; post-mortem interval < 1 h; sex was not determined) brain specimens were used for in situ hybridization in Fig. 1b. Multiple biological replicates were generated for each of the specimens. Images were taken using Aperio CS2 HR Scanner (Leica) and processed by Aperio ImageScope 12.4.3.5008 (Leica).

Immunohistochemistry

Post-mortem brains were dissected and fixed with 4% paraformaldehyde overnight at 4 °C, followed by 30% sucrose/PBS and embedding in OCT Compound (cat. no. 23–730–572, Thermo Fisher Scientific). Brains were sectioned at 15–20 µm by cryostat (CM3050S, Leica) after they were frozen. For antigen retrieval, sections were treated with R-Buffer AG pH 6.0 (cat. no. 62707-10, Electron Microscopy Sciences) for 20 min at 120 °C. The density of excitatory and inhibitory post-synaptic puncta was quantified using immunostaining of PSD95 (also known as DLG4)[57] and GPHN[58], respectively. The sources of primary antibodies were anti-PSD95/DLG4 (1:500; cat. no. 51-6900, Invitrogen), anti-GPHN (1:500; cat. no. 147011, Synaptic Systems), anti-synaptophysin (SYP) (1:2,000, cat. no. SAB4200544, Sigma-Aldrich), anti-SOX5 (1:500; cat. no. ab94396, Abcam) and anti-BCL11B/CTIP2 (1:500; cat. no. ab18465, Abcam). Secondary antibodies: AlexaFluor 488-conjugated, or 594-conjugated AffiniPure donkey anti-rabbit IgG (Jackson ImmunoResearch). For all microscopic analysis, LSM META (Zeiss) and LSM software ZEN2009 were used.

Quantification of post-synaptic and presynaptic puncta marked by immunostaining

For each cortical area, using both the 488-nm channel to detect PSD95 or GPHN and the 594-nm channel to detect BCL11B or SYP, seven serial optical sections at 0.8-µm intervals over a total depth of 5 µm were imaged and the 2nd, 4th and 6th images were eliminated from further analysis to avoid overlap in counting[59]. The area of each image is 0.079 mm^2. The number of PSD95-immunopositive, SYP-immunopositive or GPHN-immunopositive puncta on each image was counted using ImageJ (version 2.0.0-rc-69/1.52p) using the automated Analyze Particles function using a threshold of 985–4,095, determined based on multiple WT *Cbln2* E2 and h*Cbln2* E2 images. BCL11B-immunonegative cells were considered as L2–L4, and BCL11B-immunopositive nuclei were considered as L5. At least three sections from each animal were selected for counting, and at least three animals for each genotype were used.

Golgi staining

A FD Rapid GolgiStain Kit (cat. no. PK401, FD NeuroTechnologies) was used to compare neuronal morphologies approximately five basal dendrites per neuron; approximately four neurons per mouse) throughout the PFC, SSp and VISp of WT *Cbln2* E2 ($n = 3$) and *hCbln2* E2 ($n = 3$) mice aged P60. The manufacturer's protocol was followed, incubating brains in solution A+B for 21 days and solution C for 7 days before sectioning and mounting on TOMO adhesion microscope slides (cat. no. TOM-11, Matsunami). The sections were dried overnight, washed and developed by immersing the slides in solution D and solution E mixed in equal amounts for 10 min. The slides were washed twice with distilled water and dehydrated in grades of ethanol and xylene. After cover-slipped with Permount (cat. no. 15820100, Fisher Scientific), sections were imaged on a Nikon SDC microscope enabled with brightfield.

Retrograde neuronal tracing with adeno-associated viruses

To differentiate between the callosally and subcerebrally projecting neurons in the mPFC, we performed dual retrograde injections into the contralateral mPFC and the ipsilateral basal pontine nuclei, respectively, at PD 30. In brief, animals were anaesthetized by injecting ketamine/xylazine solution and head fixed in a stereotactic frame. Thirty minutes before surgery, buprenorphine was administered. After lubricating the eyes and shaving the fur, an incision of <1 mm was made. A craniotomy was made with the round 0.5-mm drill bit at the desired coordinates (for the mPFC: medial–lateral (ML) ±0.35, anterior–posterior (AP) 1.5, dorsal–ventral (DV) 2.5; and for the contralateral pons: ML ±0.5, AP −4.0 and DV 5.5 from bregma). Using a Hamilton neuros syringe (0.5 ml), we injected 50 nl of *AAVrg-Cag-Tdt* (59462-AAVrg, Addgene) into the mPFC and 150 nl of *AAVrg-Cag-Gfp* (37825-AAVrg, Addgene) into the pons at P30. To prevent the virus from spreading along the injection tract, the needle was held in place for at least 10 min. After injections, the skin was sutured and animals were returned to the cage. Approximately 3 weeks later (PD 60), the animals were killed and their brains were collected. The brains were coronally sectioned on a vibratome to obtain 70-µm-thick sections. After staining the sections with anti-GFP antibody (1:500; cat. no. ab13970, Abcam) and anti-RFP antibodies (1:500, cat. no. ab124754, Abcam), the sections were imaged with the LSM 800 microscope (Zeiss).

Quantification of dendritic spines and dendritic arbor

The images of entire neurons were acquired at ×20 magnification. Nearly 44–50 images per *z*-stack were obtained to cover the entire thickness of dendrites. For spine counts, the *z*-stack images were opened in Reconstruct 1.1.0.0., which is publicly available at http://www.bu.edu/neural/Reconstruct.html[60], and a new series was recreated that enabled us to move across different stacks across *z*-planes in the same image. Whole dendrite was subdivided into segments of

10 µm and the number of spines across whole thickness were traced for length and breadth of each spine. The length and breadth ratio was used to determine the spine subtype as previously described[61,62]. After the analysis for each class of spine, standard deviation and *P* values were calculated using two-way ANOVA with Sidak's multiple comparison method. For Sholl analysis, *z*-stack images were opened in ImageJ (version 2.0.0-rc-69/1.52p) and dendritic arbors were manually traced using the NeuronJ plugin[63]. Dendritic complexity was then quantified and plotted using the Sholl analysis option.

Processing, analysis and image visualization

For select figure images, slides background was removed and pseudocoloured white for improved clarity, as in Fig. 1b. To allow robust visualization and analysis, images depicting immunohistochemistry using antibody to PSD95/DGL4 and GPHN have been inverted and/or pseudocoloured, as in Fig. 4.

Chromatin immunopreciptation assay sequencing data analysis

H3K27ac ChIP sequencing from the developing human brain was obtained and peaks were called by Li et al.[20] that were converted to the mm10 genome build using LiftOver[64] and visualized using the Integrative Genomics Viewer[65]. DNase I hypersensitivity annotation was obtained and visualized via the UCSC Genome Browser[66,67].

Luciferase assays

Neuro2a mouse neuroblastoma cell line was purchased from ATCC. The cell line was authenticated by morphology or genotyping, and no commonly misidentified lines were used. All lines tested negative for mycoplasma contamination, checked monthly using the MycoAlert Mycoplasma Detection Kit (Lonza). Neuro2a cells were transfected using Lipofectamine 2000 (cat. no. 11668019, Thermo Fisher Scientific) with either mouse or human pCAGIG-*Sox5*, *Rxrg*, *Rarb* or empty pCAGIG, together with one of the pGL4.24 luciferase vectors generated with enhancer sequences as described above. pGL4.73 The *Renilla* luciferase plasmid (cat. no. E6911, Promega) was co-transfected to control for transfection efficiency. The luciferase assays were performed 48 h after transfection using the Dual-Luciferase Reporter Assay System (cat. no. E1910, Promega) according to the manufacturer's instructions. Luciferase activity was measured and quantified by GloMax-Multi Detection System (Promega).

Chromatin immunopreciptation assay (ChIP)

The anterior half of cortices from PD 0 mice were dissociated then crosslinked with 1% formaldehyde and processed using EZ-ChIP kit (cat. no. 17-371, Sigma-Aldrich). Chromatin fragments bound by endogenous mouse SOX5 were pulled down by anti-SOX5 (cat. no. ab94396, Abcam), anti-POL2 antibody (1:1,000; cat. no. 05-623, Millipore-Sigma; included in the kit) or random IgG control (included in the kit), then detected by PCR using E2-ChIP-F and E2-ChIP-R or E1-ChIP-F, and E1-ChIP-R primers (Supplementary Table 2).

Quantitative reverse transcription–PCR

Total RNA was isolated using TRIzol (cat. no. 15596026, Thermo Fisher Scientific) from freshly dissected neocortices after removal of the olfactory bulb, hippocampus and striatum. cDNAs were prepared using SuperScript II (cat. no. 18064022, Thermo Fisher Scientific) from more than three independent WT littermates and *hCbln2* E2 brains. Quantitative reverse transcription–PCR was performed as previously described[68] using 7000HT Sequence Detection System (Applied Biosystems). At least three biological replicates per transcript were used for every reaction. The copy number of transcripts was normalized against the housekeeping TATA-binding protein (TBP) transcript level. For *Cbln2* and *Tbp* primer sets, correlation (R^2) was higher than 0.98, and the slope was −3.1 to −3.6 in each standard curve. Primers were designed in a single exon and are listed in Supplementary Table 2.

3D reconstruction of dendrites and dendritic spines using electron microscopy

PD 60 WT *Cbln2* E2 and h*Cbln2* E2 mice ($n = 3$ for each genotype) were intracardially perfused with fixative containing 4% paraformaldehyde and 1% glutaraldehyde in 0.1 M phosphate buffer (pH 7.4). Brains were removed and immersed in the same fixative for 3–4 days at 4 °C. Coronal 60-µm-thick slices through the PFC were cut with a vibratome, post-fixed with 1% OsO_4, dehydrated in ethanol, embedded in epoxy resin Durcupan (ACM; Fluka) and polymerized on microscope slides. Ultrathin sectioning and electron microscopic investigation were performed as previously described[69,70] with minor modifications as follow. Segments of the mPFC were dissected and re-embedded into Durcupan blocks. Serial 60-nm-thick sections from mPFC L2/3 were cut by Leica UC7 ultramicrotome, collected on one-slot grids covered with Butvar B-98 films (EMS) and stained with lead citrate. Series of 48–51 consecutive images were photographed with the Talos L120C electron microscope (Thermo Fisher Scientific) at ×5,300 magnification. For serial imaging, random neuropil segments were chosen, although avoiding cell bodies and blood capillaries. 3D reconstructions were performed using the computer program Reconstruct 1.1.0.0., which is publicly available at http://www.bu.edu/neural/Reconstruct.html[60]. Sixteen-representable dendrite fragments were traced in the serial images and their innervated spines and non-innervated spine-like membrane protrusions were counted. The percentage of innervated spines among all the protrusions were calculated for each animal.

Reporting summary

Further information on research design is available in the Nature Research Reporting Summary linked to this paper.

Code availability

All software and code used in this study are publicly available.

49. Dy, P., Han, Y. & Lefebvre, V. Generation of mice harboring a Sox5 conditional null allele. *Genesis* **46**, 294–299 (2008).
50. Mathelier, A. et al. JASPAR 2016: a major expansion and update of the open-access database of transcription factor binding profiles. *Nucleic Acids Res.* **44**, D110–D115 (2016).
51. Khan, A. et al. JASPAR 2018: update of the open-access database of transcription factor binding profiles and its web framework. *Nucleic Acids Res.* **46**, D260–D266 (2018).
52. Shim, S. et al. Regulation of EphA8 gene expression by TALE homeobox transcription factors during development of the mesencephalon. *Mol. Cell. Biol.* **27**, 1614–1630 (2007).
53. Wang, H. et al. One-step generation of mice carrying mutations in multiple genes by CRISPR/Cas-mediated genome engineering. *Cell* **153**, 910–918 (2013).
54. Liu, P. et al. A highly efficient recombineering-based method for generating conditional knockout mutations. *Genome Res.* **13**, 476–484 (2003).
55. Cong, L. et al. Multiplex genome engineering using CRISPR/Cas systems. *Science* **339**, 819–823 (2013).
56. Wilkinson, D. G. & Nieto, M. A. Detection of messenger RNA by in situ hybridization to tissue sections and whole mounts. *Methods Enzymol.* **225**, 361–373 (1993).
57. Hunt, C. A. et al. PSD-95 is associated with the postsynaptic density and not with the presynaptic membrane at forebrain synapses. *J. Neurosci.* **16**, 1380–1388 (1996).
58. Essrich, C. et al. Postsynaptic clustering of major GABAA receptor subtypes requires the γ2 subunit and gephyrin. *Nat. Neurosci.* **1**, 563–571 (1998).
59. Ippolito, D. M. & Eroglu, C. Quantifying synapses: an immunocytochemistry-based assay to quantify synapse number. *J. Vis. Exp.* **45**, 2270 (2010).
60. Fiala, J. C. Reconstruct: a free editor for serial section microscopy. *J. Microsc.* **218**, 52–61 (2005).
61. Risher, W. C. et al. Rapid Golgi analysis method for efficient and unbiased classification of dendritic spines. *PLoS ONE* **9**, e107591 (2014).
62. Kaur, N. et al. Neural stem cells direct axon guidance via their radial fiber scaffold. *Neuron* **107**, 1197–1211 (2020).
63. Meijering, E. et al. Design and validation of a tool for neurite tracing and analysis in fluorescence microscopy images. *Cytometry A* **58**, 167–176 (2004).
64. Hinrichs, A. S. et al. The UCSC Genome Browser database: update 2006. *Nucleic Acids Res.* **34**, D590–D598 (2006).
65. Robinson, J. T. et al. Integrative genomics viewer. *Nat. Biotechnol.* **29**, 24-26 (2011).
66. Kent, W. J. et al. The human genome browser at UCSC. *Genome Res.* **12**, 996–1006 (2002).
67. Rosenbloom, K. R. et al. ENCODE data in the UCSC Genome Browser: year 5 update. *Nucleic Acids Res.* **41**, D56–D63 (2013).

68. Shibata, M. et al. MicroRNA-9 regulates neurogenesis in mouse telencephalon by targeting multiple transcription factors. *J. Neurosci.* **31**, 3407–3422 (2011).
69. Morozov, Y. M., Ayoub, A. E. & Rakic, P. Translocation of synaptically connected interneurons across the dentate gyrus of the early postnatal rat hippocampus. *J. Neurosci.* **26**, 5017–5027 (2006).
70. Morozov, Y. M., Mackie, K. & Rakic, P. Cannabinoid type 1 receptor is undetectable in rodent and primate cerebral neural stem cells but participates in radial neuronal migration. *Int. J. Mol. Sci.* **21**, 1–19 (2020).
71. Thompson, C. L. et al. A high-resolution spatiotemporal atlas of gene expression of the developing mouse brain. *Neuron* **83**, 309–323 (2014).

Acknowledgements We thank S. Bai, T. Nottoli and X. Xing for technical help with generating transgenic mice and genetically humanized mice; A. Sousa for providing tissue; B. Lorente-Galdos and G. Santpere for helping with data analysis; A. Duque for using equipment from MacBrainResource (MH113257); V. Lefebvre for providing floxed *Sox5* mice; and the members of the Sestan laboratory for comments. This work was supported by the National Institutes of Health (HG010898, MH106874, MH106934, MH110926 and MH116488) and the Simons Foundation Autism Research Institute (SFARI) (572080; to N.S.). Additional support was provided by the National Institutes of Health T32 fellowship (MH18268 to K.P.), the National Science Foundation Graduate Research Fellowship Program (to S.K.M.), the Kavli Foundation and the James S. McDonnell Foundation (to N.S.).

Author contributions M.S., K.P. and N.S. designed the research. M.S. and K.P. performed the overall experiments and analysed the data. S.K.M. analysed the RNA sequencing, ChIP sequencing and genomic sequence data. M.S. generated the transgenic and knock-in mice. N.K. performed retrograde neuronal tracing with the AAV and analysed the data. Y.M.M. performed and analysed the electron microscopy experiments. X.C. and S.G.W. performed and analysed the preliminary electrophysiological analysis. N.S. conceived the study. M.S., K.P. and N.S. wrote the manuscript. All authors discussed the results and implications and commented on the manuscript at all stages.

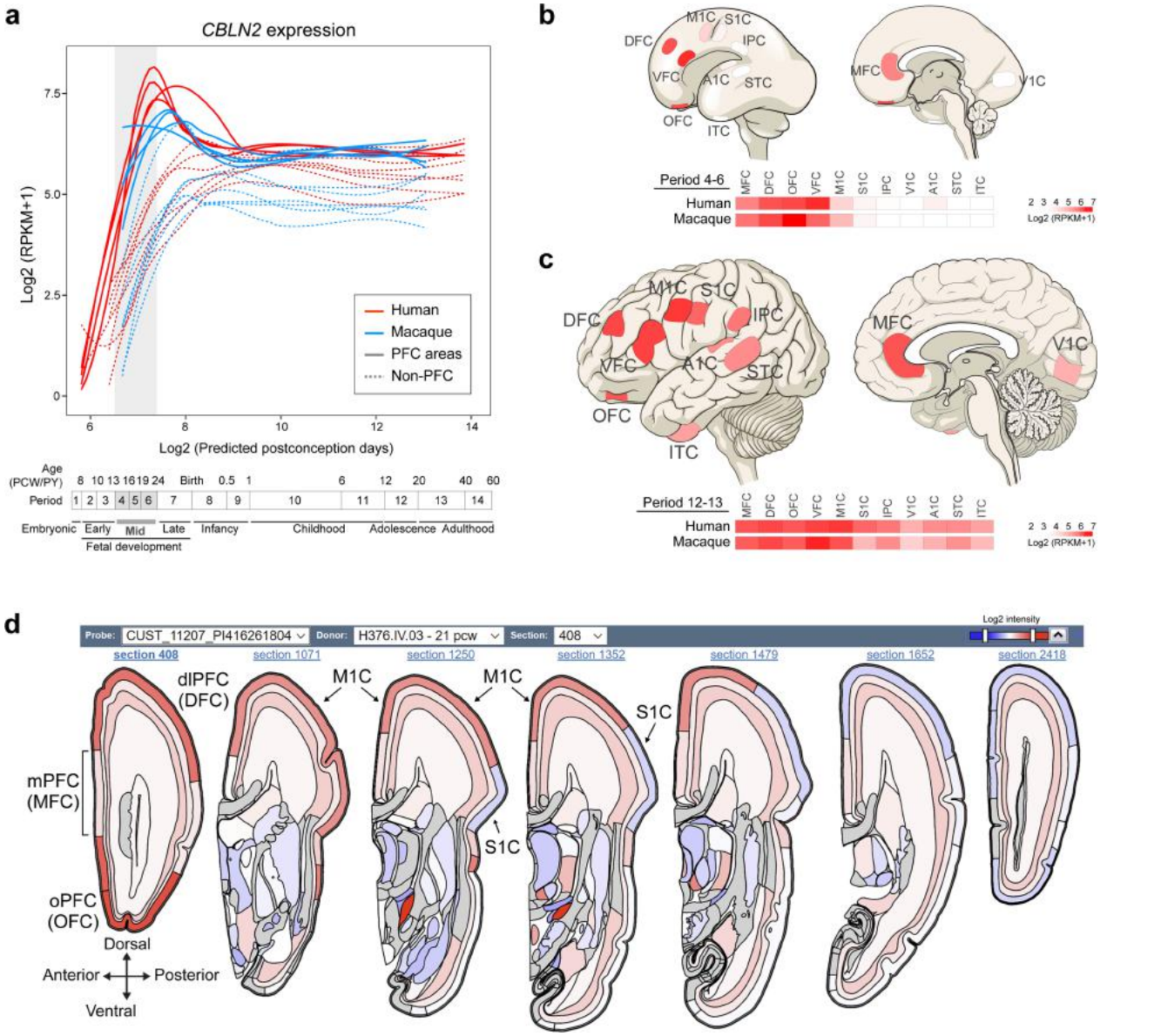
Competing interests The authors declare no competing interests.

Additional information

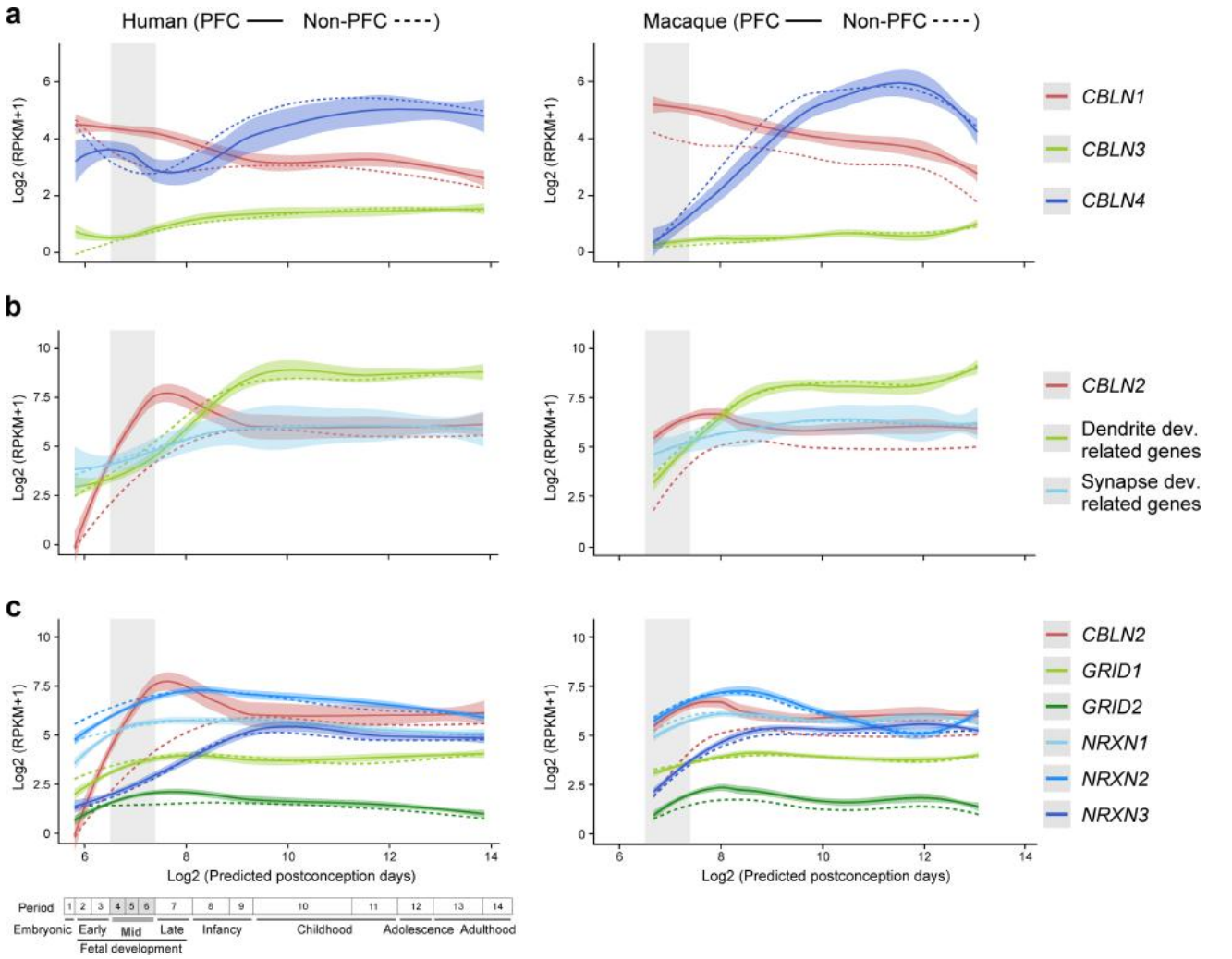
Supplementary information The online version contains supplementary material available at https://doi.org/10.1038/s41586-021-03952-y.

Correspondence and requests for materials should be addressed to Nenad Sestan.

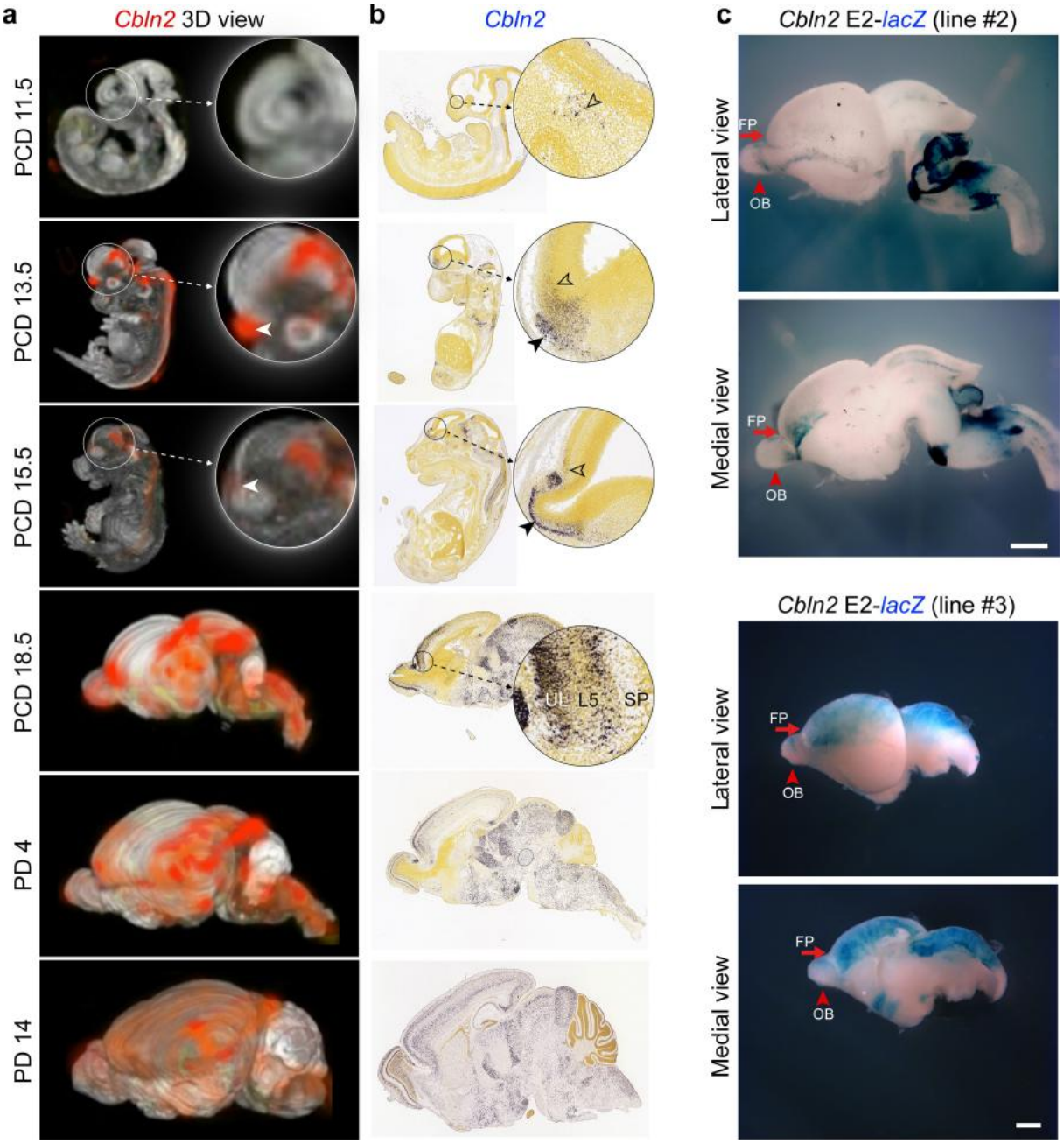
Peer review information *Nature* thanks Alex Pollen and the other, anonymous, reviewer(s) for the peer review of this work.



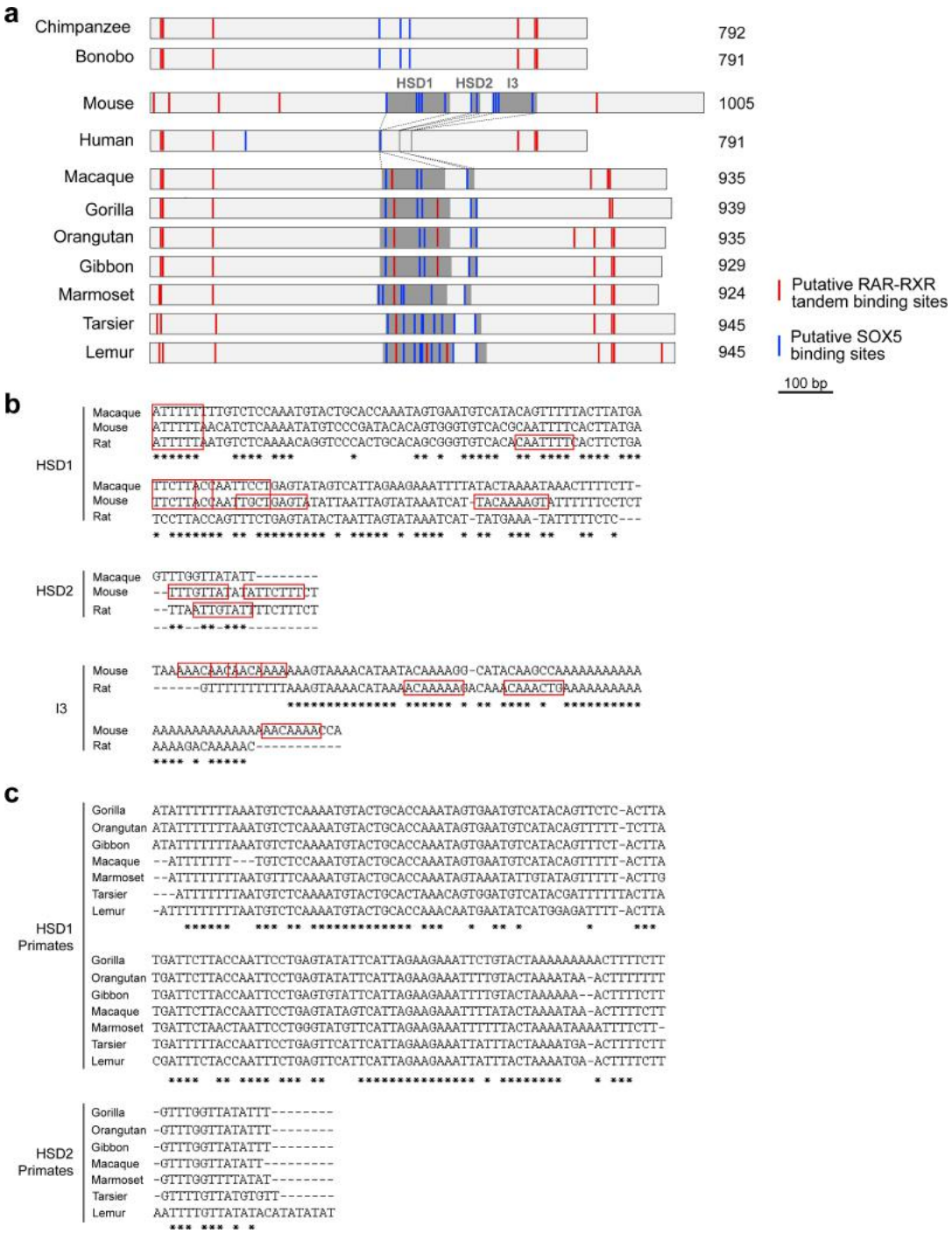
Extended Data Fig. 1 | Spatiotemporal expression of *CBLN2* in the human and macaque neocortex. a–c, Spatiotemporal expression of *CBLN2* expression in the human and macaque cerebral cortex during development base on BrainSpan (brainspan.org) and PsychENCODE (development.psychencode.org, evolution.psychencode.org) human and macaque RNA-seq data[15,20]. The RNA-seq data consisted of tissue-level samples comprising the pial surface, marginal zone, cortical plate (layers 2-6) and adjacent subplate zone, of eleven prospective neocortical areas. Red and blue lines indicate human and macaque PFC *CBLN2* expression, respectively, and dotted lines represent the non-PFC *CBLN2* expression. Vertical grey box demarcates mid-fetal developmental periods. Predicted ages, timeline of human and macaque development and the associated periods are shown below. Visual representation of *CBLN2* in human cortex with heatmaps of regional expression in human and macaque below during period 4-6 (**b**) and period 12-13 (**c**). Darker reds represent high expression levels. **d**, Anteroposterior visual representation of human *CBLN2* expression at PCW 21 from the BrainSpan human prenatal laser microdissection microarray data (brainspan.org)[37].



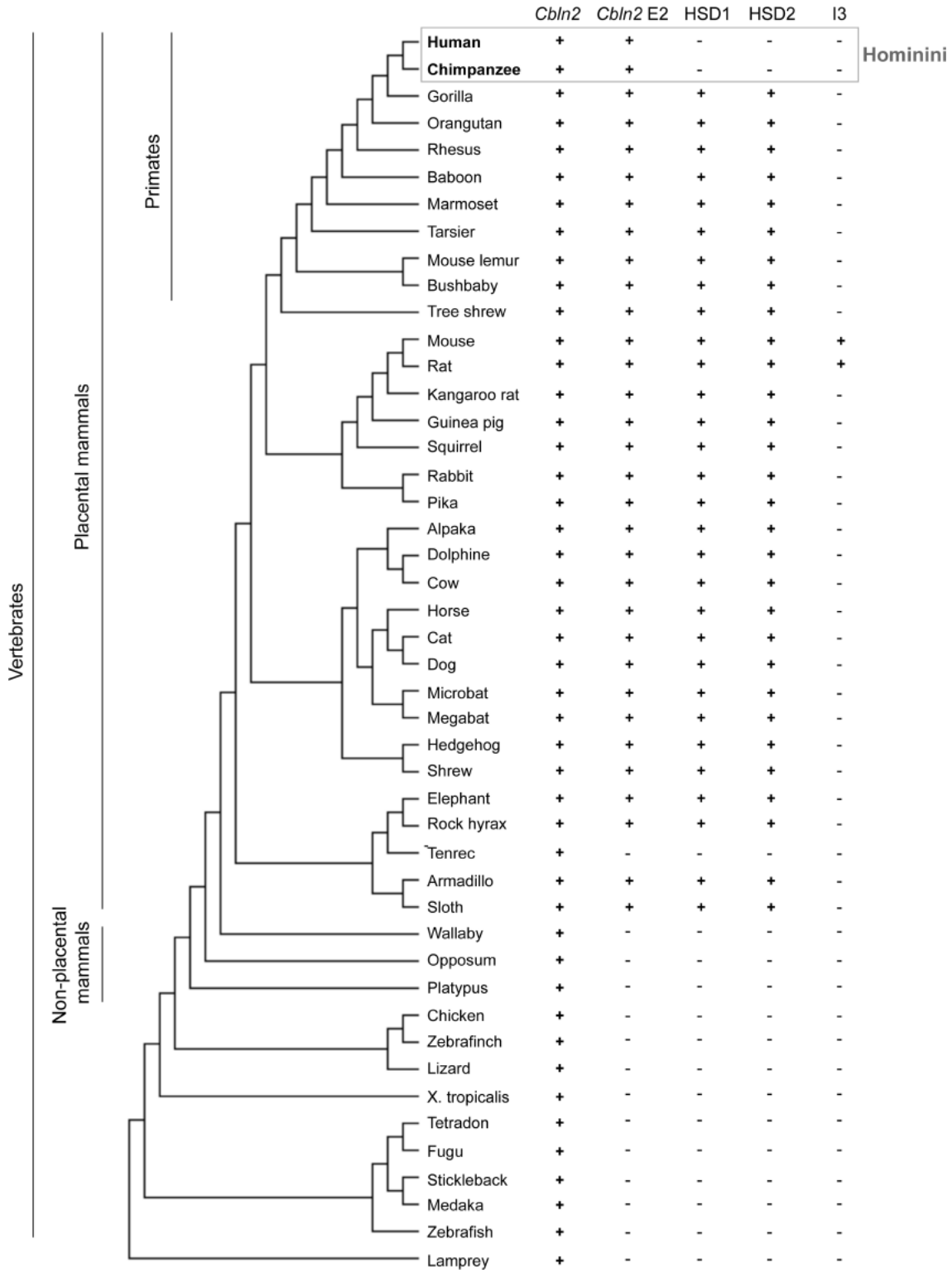
Extended Data Fig. 2 | Expression profile of *CBLN1,3,4* genes related to dendrite and synapse development and CBLN2 binding partners in macaque and human. a, Developmental trajectory of *CBLN2* expression compared to the expression of key genes related to dendrite development (i.e., *MAP1A* and *CAMK2A*) and synapse development genes (i.e., *SYP, SYPL1, SYPL2* and *SYN1*). The lists of genes related to synapse and dendrite development were previously complied and analyzed for their expression trajectories by Kang et al.[35]. **b**, Developmental trajectory of *CBLN2* expression compared to the expression of genes encoding CBLN2 binding partners (i.e., *GRID1, GRID2, NRXN1, NRXN2* and *NRXN3*). Left and right panels show gene expression in human and macaque, respectively. Gene expression in PFC and non-PFC are indicated by solid and dashed line, respectively. Vertical grey box highlights the mid-fetal developmental periods. Predicted ages, timeline of human and macaque development and the associated periods are shown below. Gene expression values are represented as log2(RPKM+1). For all of these plots, the shading around the lines represents the 95% confidence interval. Abbreviations are as described in Fig. 1 legend.



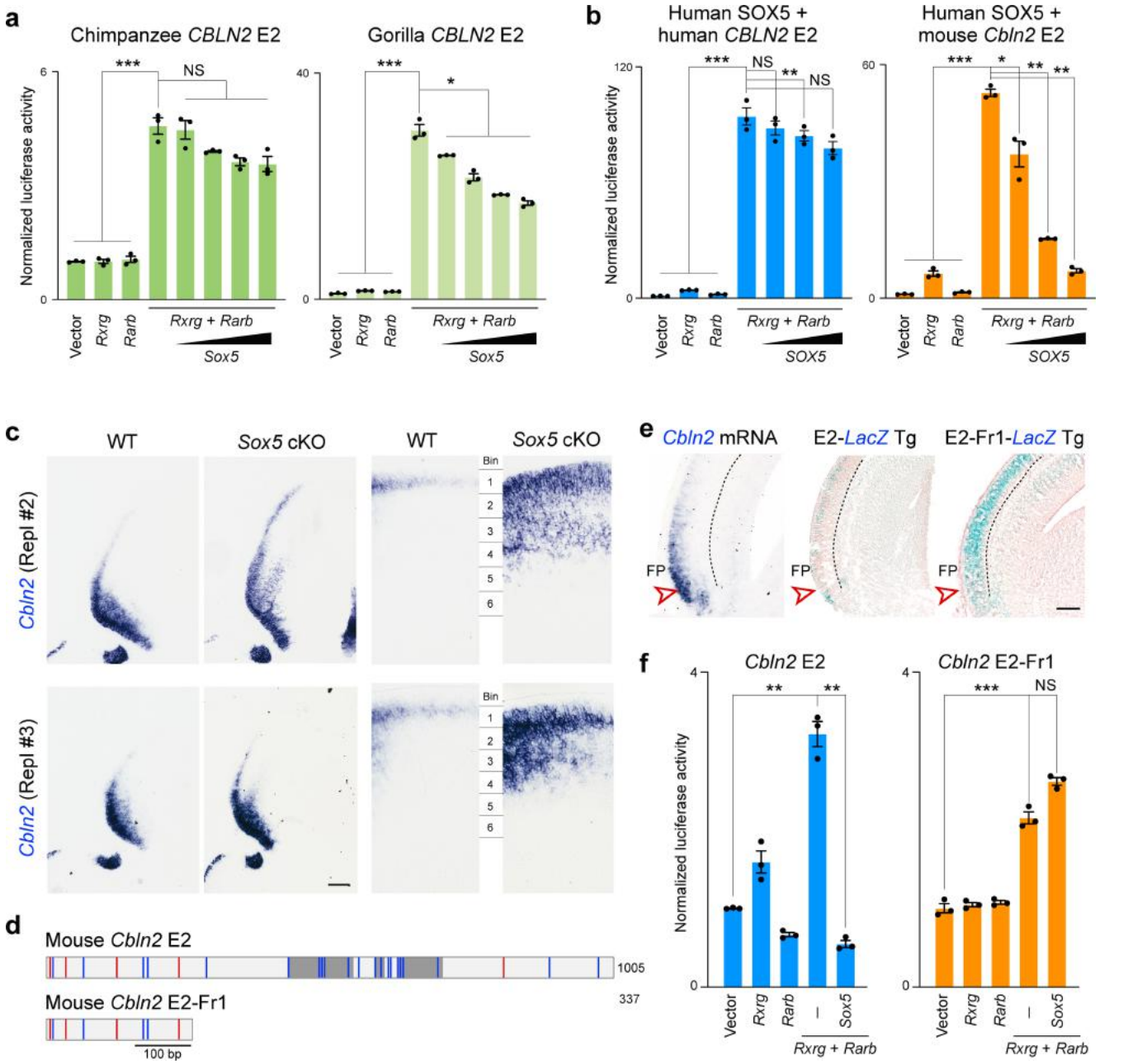
Extended Data Fig. 3 | Mouse *Cbln2* expression at multiple stages and Cbln2 E2 transgenic lines. a, **b**, Visualization of *Cbln2* expression in wholemount (**a**) and sagittal sections (**b**) at PCD 11.5, 13.5, 15.5 18.5 and PD 4, 14 from Allen Brain Atlas developing mouse brain atlas (developingmouse.brain-map.org)[71]. Arrowheads highlight early rostral expression, UL, upper layer; L5, layer 5; SP, subplate. **c**, β-Galactosidase activity in two additional mouse brains from independent transgenic mouse carrying *Cbln2* E2 conjugated with *lacZ* reporter at PCD 17. See Fig. 1c for the third replicate. FP, frontal pole; OB, olfactory bulb. Scale bars, 1 mm.



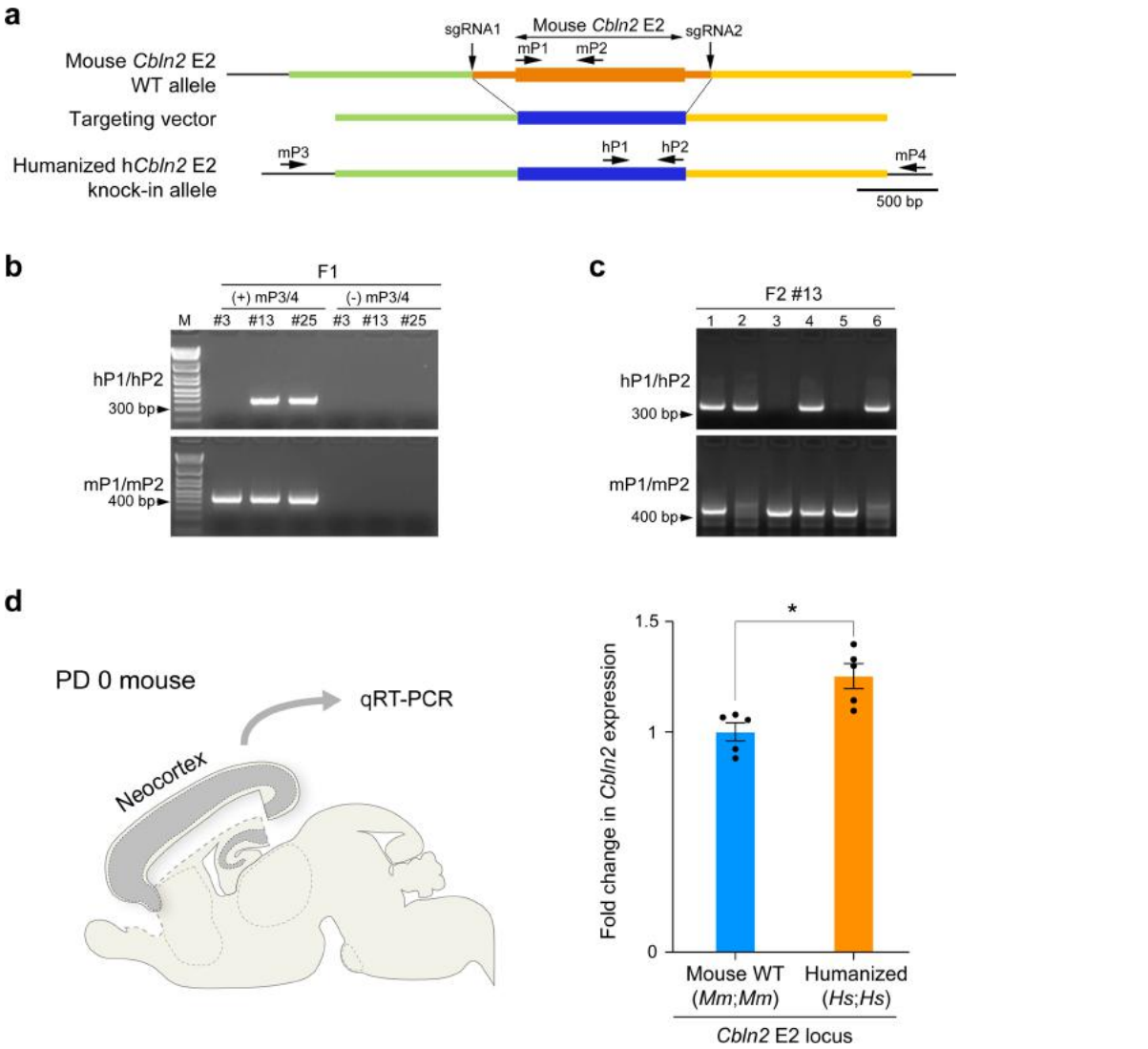
Extended Data Fig. 4 | Comparative analysis of *CBLN2* E2 deletions across mammals. a, Schematic representation of *CBLN2* E2 from mouse and primate species including apes (human, common chimpanzee, bonobo, gorilla, orangutan and gibbon), old world monkey (Rhesus macaque), new world monkey (marmoset) and prosimians (tarsier and lemur). HSD1, HSD2 and I3 are shaded. Putative RAR-RXR tandem binding sites indicated as red lines and putative SOX5-binding sites as blue lines. **b**, Sequence alignments of HSD1, HSD2 and I3 from macaque, mouse, and rat. Putative SOX5 binding sites are shown in red boxes. **c**, Sequence alignments of HSD1 and HSD2 from primates shown in **a**.



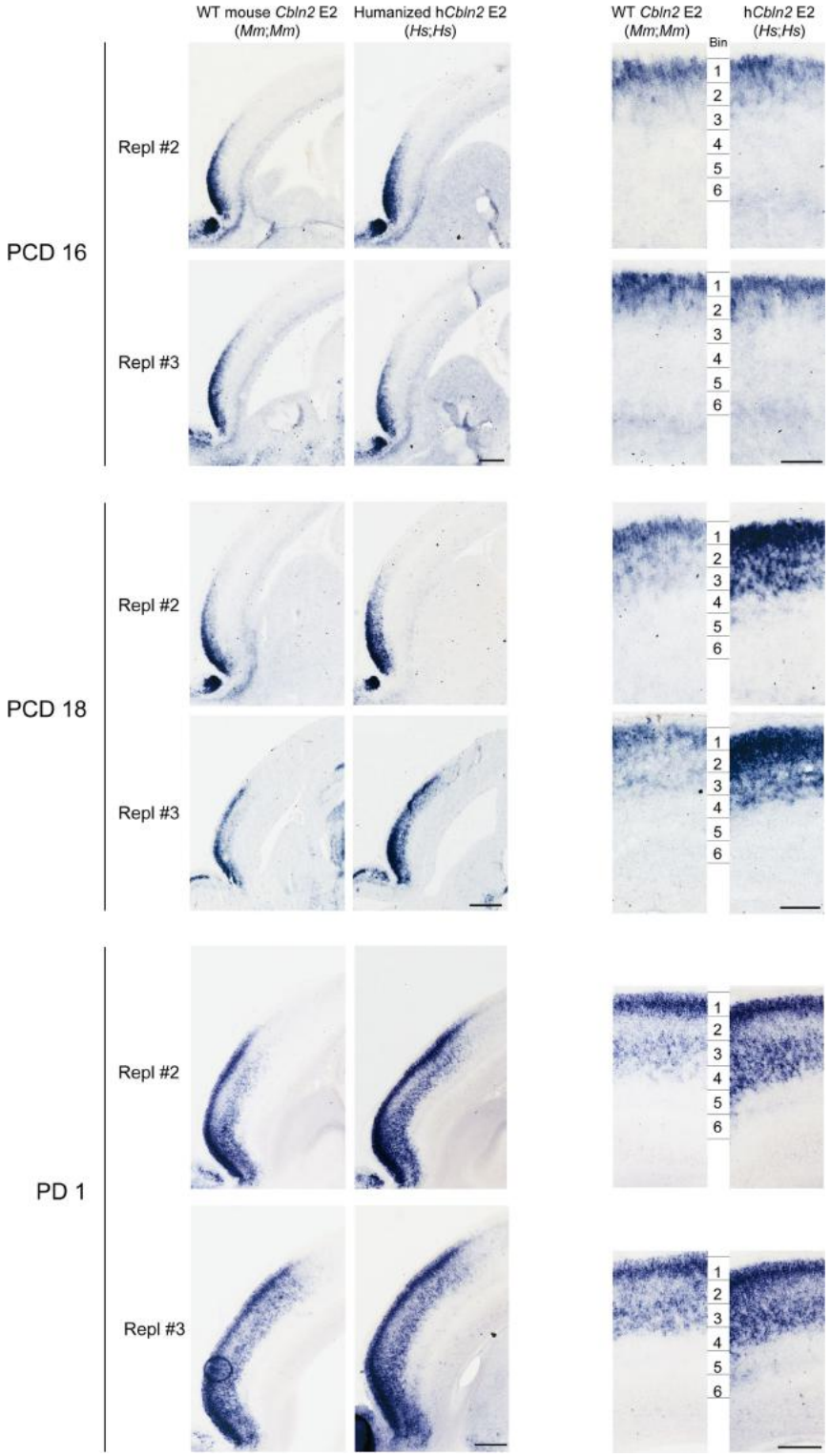
Extended Data Fig. 5 | Conservation of *CBLN2* E2 and deleted regions across species. Phylogenetic tree of selected chordates including placental and non-placental mammals with information about presence of the *CBLN2/Cbln2* gene and *CBLN2* E2/Cbln2 E2. The last three columns describe the presence of HSD1, HSD2 and I3. The mouse E2 sequence was searched in each most updated animal genome browser at UCSC genome browser (as of December 22, 2019). E2 conservation criteria are: 1) identity over 80%; and 2) alignment length over 600 bp compared with mouse E2 (1005 bp).



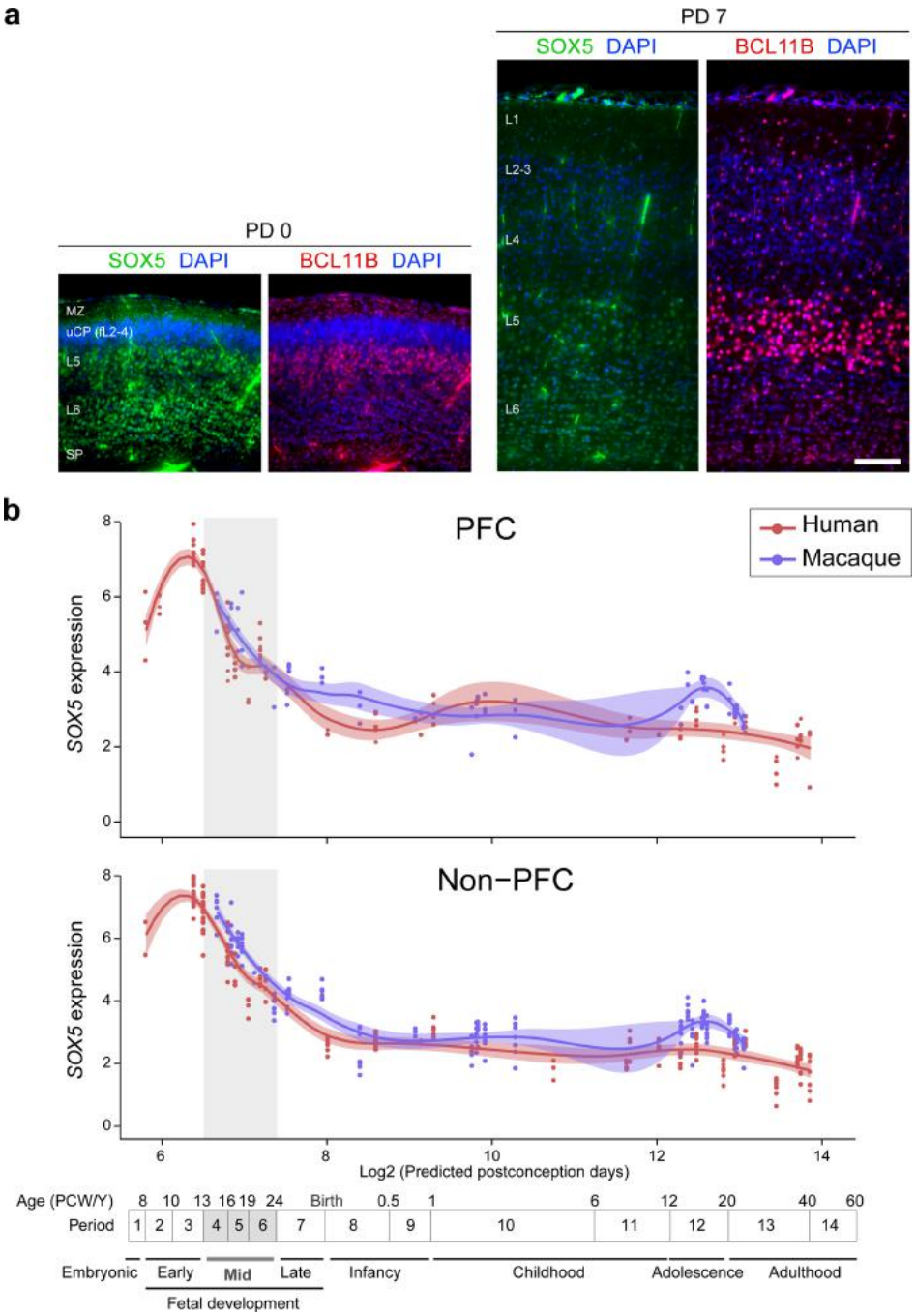
Extended Data Fig. 6 | SOX5 directly suppresses RXRG-RARB responsive human, chimpanzee, gorilla and mouse *Cbln2* E2 enhancer. a, Luciferase assay in Neuro-2a cell line from luciferase reporters conjugated to chimpanzee and gorilla wild *CBLN2* E2. Two-tailed Student's t-test; *P = 0.01; ***P = 9e-5 (Chimpanzee), 1e-5 (Gorilla); NS, not significant. Error bars; S.E.M.; N = 3 per condition. **b**, Overexpression of human *SOX5* exerts a similar effect to mouse *Sox5* on human and mouse *Cbln2* E2 reporters. Two-tailed Student's t-test; *P = 5e-3; ***P = 4e-5; ****P = 2e-6; 2e-6, 1e-6 (Human SOX5 + mouse *Cbln2* E2); NS, not significant; Error bars, S.E.M.; N = 3 per condition. **c**, *Cbln2* expression is upregulated in *Sox5* conditional knockout brain at PD 0. Additional two replicates (Repl #2 and 3) not shown in Fig. 2d are shown. Scale bar, 200 µm (left); 100 µm (right). **d**, Constructs used for luciferase assay and generation of transgenic animals shown in **e**, **f**. **e**, Transgenic mouse brain at PCD 17 carrying *Cbln2* E2 (N = 3) or *Cbln2* E2 Fr1-*lacZ* reporters (N = 3). *Cbln2* expression in the PFC is indicated by arrowheads. Endogenous *Cbln2* expression is also shown for comparison (N = 4). Scale bar, 200 µm. **f**, Luciferase assay for *Cbln2* E2 and *Cbln2* E2-Fr1. Two-tailed Student's t-test; **P = 1e-4 (Cbln2 E2), 2e-4 (Cbln2 E2-Fr1); ***P = 8e-5; NS, not significant; Error bars: S.E.M.; N = 3 per condition.



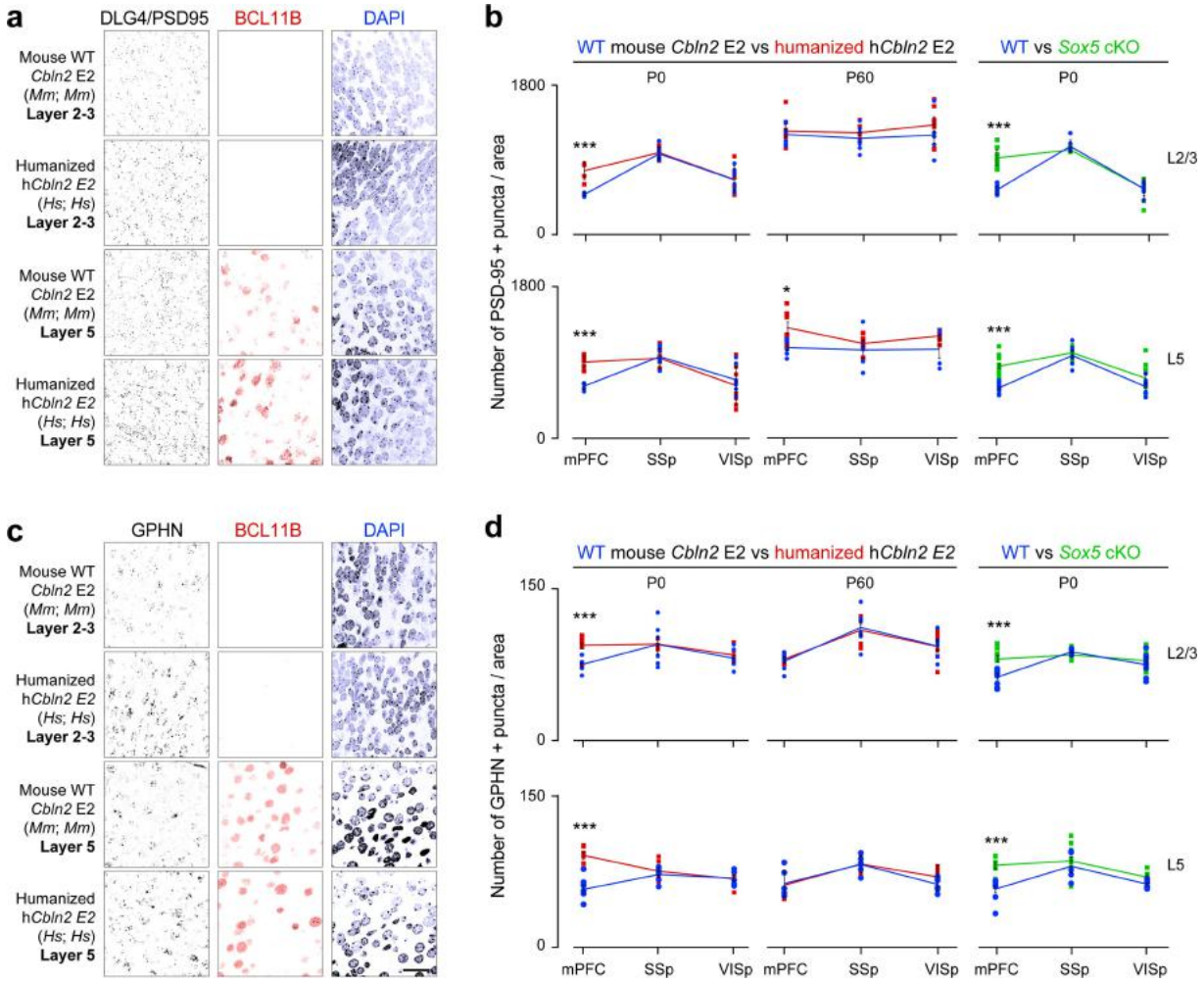
Extended Data Fig. 7 | Humanized *Cbln2* E2 knock-in mouse shows increased *Cbln2* expression in neonatal neocortex. a, Positions of single-guide RNAs (sgRNA1 and 2) to introduce double-strand breaks in the genomic DNA and targeting vector to replace WT mouse *Cbln2* E2 region with that of human *CBLN2* E2 (humanized h*Cbln2* E2) are shown. **b**, Genotyping strategy for F1 mice. Germline transmission in the F1 generation was confirmed by nested PCR using the primer set of mP3/mP4, followed by hP1/P2 as indicated in **a**. Two founders #13 and #25 were obtained. **c**, Mice in the following generation were genotyped by PCR with hP1/P2 and mP1/mP2. An example of genotyping for F2 of line #13 mice is shown (**c**). **d**, Comparison of *Cbln2* expression between WT *Cbln2* E2 (*Mm;Mm*) and h*Cbln2* E2 (*Hs;Hs*) neocortex at PD 0 using quantitative reverse transcription-PCR. RNA was extracted from the neocortex following the removal of hippocampus, olfactory bulb and subpallial regions. Two-tailed Student's t-test; *P = 0.007; Error bars: S.E.M.; N = 5 per genotype. Genotyping were repeated at least two times in **b** and **c**.



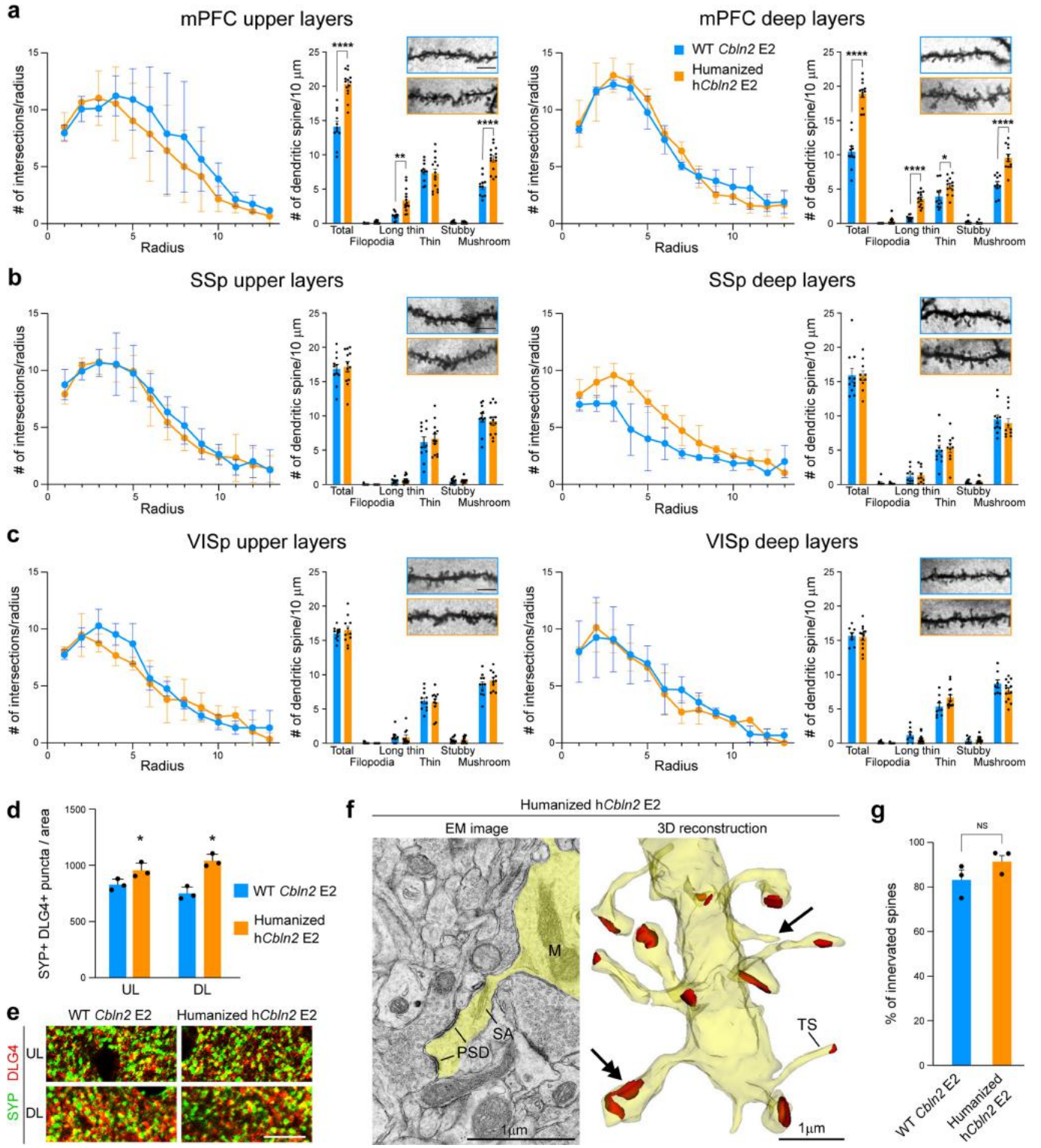
Extended Data Fig. 8 | Additional replicates. Additional replicates from Fig. 3. **a**, **b**, The neocortex of the humanized h*Cbln2* E2 knock-in prenatal and neonatal mice exhibits upregulated *Cbln2* in both upper and deeper layers compared to mice carrying WT *Cbln2* E2. Additional two replicates (Repl #2 and 3) not shown in Fig. 3a are shown for PND 16 and PND18. Scale bars, 200 µm (left); 100 µm (right). All analysis and three replicates for PD1. Neocortex was divided into six equal bins spanning from pia to ventricular zone, and *Cbln2* signal intensity was quantified for each bin and compared between WT and h*Cbln2* E2 (*Hs;Hs*). Two-tailed Student's t-test; *P = 0.02 (PCD18), 0.02, 0.04 (PD1), **P = 1e-3 (PCD18), 1e-3, 3e-3, 3e-3 (PD1), ***P = 1e-4, 2e-4 (PCD18); N = 3 per genotype. Scale bars, 100 µm. **c**, Expression of the upper layer marker, *Cux2*, and SOX5 in adjacent tissue sections were detected by *in situ* hybridization and immunostaining, respectively. Scale bars, 100 µm. N = 3.



Extended Data Fig. 9 | Expression of *Sox5* in the developing mouse and human neocortex. a, Double immunofluorescent staining for SOX5 and BCL11B in PD 0 and PD 7 mouse neocortex. N = 3 per condition. Scale bar, 100 µm. **b**, SOX5 expression in the PFC and non-PFC areas of the cerebral cortex of human and macaque across development. Red and blue lines indicate human and macaque, respectively. Vertical grey box demarcates mid-fetal developmental periods. For all of those plots, the shading around the lines represents the 95% confidence interval. Predicted ages, timeline of human and macaque development and the associated periods are shown below.



Extended Data Fig. 10 | Quantification of excitatory and inhibitory postsynaptic puncta. a–**d**, PD 0 WT and h*Cbln2* E2 mPFC L2-3 and L5 (BCL11B-immunopositive) immunostained for PSD-95/DLG4 or GPHN. Density of PSD-95/DLG4+ excitatory and GPHN+ inhibitory postsynaptic puncta in PD 0 and PD 60 mPFC, primary somatosensory area (SSp), and primary visual area (VISp) of WT (blue), h*Cbln2* E2 (orange), and *Sox5* cKO mice (green). Two-tailed Student's test; *P = 0.018; ***P = 4e-4, 1e-4 (**b**, WT vs. humanized hCbln2 E2 L2-3, and L5, respectively), 1e-4, 7e-5, 4e-4, 2e-4 (**d**, WT vs. humanized hCbln2 E2 L2-3 and L5, respectively; WT vs. Sox5 cKO L2-3 and L5, respectively); ****P < 1e-5. Error bars: S.E.M.; N = 5 per PD 0 genotype; N = 3 per PD 60 genotype. Scale bar, 25 µm.



Extended Data Fig. 11 | Analysis of dendritic spines and dendritic complexity. a–**c**, Quantification of Sholl analysis and representative images (inset) of Golgi stained dendrites and dendritic spines per 10 µm in layer 2-3 and 5 of mPFC (**a**), SSp (**b**) and VISp (**c**) in WT *Cbln2* E2 and humanized hCbln2 E2. Two-way ANOVA with Sidak's multiple comparison method was applied; *P = 0.04; **P = 1e-3, ****P < 0001; Error bars: S.E.M.; N = 16 (WT *Cbln2* E2), and 16 (humanized hCbln2 E2) of PD 60 brains (N = 3). Scale bars: 5 mm. **d**, **e**, Quantification and representative images of juxtaposed synaptophysin (SYP) and DLG4/PSD-95 immunolabelled puncta in upper (UL) and deep layer (DL) mPFC of WT *Cbln2* E2 and humanized hCbln2 E2. Two-tailed Student's t-test was applied * P = 0.03; Error bars: S.E.M. (N = 3); Scale bar: 10 mm. **f**, Electron microscopy (EM) image shows dendrite emitting spine with postsynaptic density (PSD) and spine apparatus (SA) (left) in PD 60 humanized hCbln2 E2 mouse. 3D reconstruction (right) of the dendrite with numerous spines showing spine heads and synaptic contacts (red). Double arrow points to the spine shown in EM image. Thin spines without head (TS) and not innervated thin extensions (arrow) are also detected. Abbreviation: M, mitochondria. Sixteen-representable dendrite fragments were traced in the serial images (see Methods). **g**, Quantification of percentage of innervated spines in mPFC of WT *Cbln2* E2 and humanized hCbln2 E2. Two-tailed Student's t-test was applied. NS, Not significant. Error bars: S.E.M. (N = 3).

nature research

Corresponding author(s): Nenad Sestan

Reporting Summary

Nature Research wishes to improve the reproducibility of the work that we publish. This form provides structure for consistency and transparency in reporting. For further information on Nature Research policies, see Authors & Referees and the Editorial Policy Checklist.

Statistical parameters

When statistical analyses are reported, confirm that the following items are present in the relevant location (e.g. figure legend, table legend, main text, or Methods section).

n/a	Confirmed	
☐	☒	The exact sample size (*n*) for each experimental group/condition, given as a discrete number and unit of measurement
☐	☒	An indication of whether measurements were taken from distinct samples or whether the same sample was measured repeatedly
☐	☒	The statistical test(s) used AND whether they are one- or two-sided *Only common tests should be described solely by name; describe more complex techniques in the Methods section.*
☐	☒	A description of all covariates tested
☐	☒	A description of any assumptions or corrections, such as tests of normality and adjustment for multiple comparisons
☐	☒	A full description of the statistics including central tendency (e.g. means) or other basic estimates (e.g. regression coefficient) AND variation (e.g. standard deviation) or associated estimates of uncertainty (e.g. confidence intervals)
☐	☒	For null hypothesis testing, the test statistic (e.g. *F*, *t*, *r*) with confidence intervals, effect sizes, degrees of freedom and *P* value noted *Give P values as exact values whenever suitable.*
☒	☐	For Bayesian analysis, information on the choice of priors and Markov chain Monte Carlo settings
☒	☐	For hierarchical and complex designs, identification of the appropriate level for tests and full reporting of outcomes
☒	☐	Estimates of effect sizes (e.g. Cohen's *d*, Pearson's *r*), indicating how they were calculated
☐	☒	Clearly defined error bars *State explicitly what error bars represent (e.g. SD, SE, CI)*

Our web collection on statistics for biologists may be useful.

Software and code

Policy information about availability of computer code

Data collection	MEME suite (ver. 4.12.0), and Liftover was downloaded on May 3rd, 2017.
Data analysis	All software/code utilized for the analysis of data is publically available and described in the manuscript. Aperio ImageScope (ver. 12.4.3.5008 (Leica)), ZEN2009, and ImageJ (ver. 2.0.0-rc-69/1.52p), Reconstruct (ver.1.1.0.0. (Boston, MA, USA)) were used for basic image processing which included normalization by size and optical setting. Reconstruct was used for spine counting and 3D reconstruction of spines.

For manuscripts utilizing custom algorithms or software that are central to the research but not yet described in published literature, software must be made available to editors/reviewers upon request. We strongly encourage code deposition in a community repository (e.g. GitHub). See the Nature Research guidelines for submitting code & software for further information.

Data

Policy information about availability of data

All manuscripts must include a data availability statement. This statement should provide the following information, where applicable:

- Accession codes, unique identifiers, or web links for publicly available datasets
- A list of figures that have associated raw data
- A description of any restrictions on data availability

Sequencing data were deposited at http://psychencode.org and NCBI dbGAP Accession phs000755.v2.p1

Field-specific reporting

Please select the best fit for your research. If you are not sure, read the appropriate sections before making your selection.

☒ Life sciences ☐ Behavioural & social sciences ☐ Ecological, evolutionary & environmental sciences

For a reference copy of the document with all sections, see nature.com/authors/policies/ReportingSummary-flat.pdf

Life sciences study design

All studies must disclose on these points even when the disclosure is negative.

Sample size	No sample size calculation was performed. The size of the samples was determined by previously established sample sizes from other murine neurodevelopmental studies and by the availability of samples.
Data exclusions	No data were excluded from the analysis.
Replication	All experiments were performed in this manuscript with 3 biological replicates with the following exceptions: CBLN2 in situ hybridization of PCW 20 human and PCD 140 macaque were performed with 2 biological replicates (Fig. 1b); ChIP-PCR assays in PD 0 mouse neocortex were performed with 4 biological replicates per genotype (Fig. 2c); Density of PSD-95/DLG4+ excitatory and GPHN+ inhibitory postsynaptic puncta in PD 0 were performed with 5 biological replicates per genotype (Fig. 4b, d); Comparison of Cbln2 expression between WT Cbln2 E2 (Mm;Mm) and hCbln2 E2 (Hs;Hs) neocortex at PD 0 were performed with 5 biological replicates per genotype (Extended Data Fig. 7). All experimental findings were reproducible.
Randomization	Randomization was not relevant to this study as controls and mutant mouse lines did not receive different treatments and human/primate studies were descriptive studies.
Blinding	Group allocation was conducted as described above. Data collection was performed by independent investigators. Prior to data analysis, all experiments including immunohistological and histological were randomized and analyzed by independent, blinded observers as described in the methods section.

Reporting for specific materials, systems and methods

Materials & experimental systems

n/a	Involved in the study
☒	☐ Unique biological materials
☐	☒ Antibodies
☐	☒ Eukaryotic cell lines
☒	☐ Palaeontology
☐	☒ Animals and other organisms
☐	☒ Human research participants

Methods

n/a	Involved in the study
☒	☐ ChIP-seq
☒	☐ Flow cytometry
☒	☐ MRI-based neuroimaging

Antibodies

Antibodies used	anti-DLG4/PSD95 (1:500; Invitrogen, Cat. 51-6900), anti-GPHN (1:500; Synaptic Systems, Cat. 147011), anti-SYP (1:2000, Sigma-Aldrich, Cat. SAB4200544), anti-SOX5 (1:500; Abcam, Cat. ab94396, anti-BCL11B/CTIP2 (1:500, Abcam, Cat. ab18465, Alexa Fluor 488-conjugated AffiniPureanti-IgG (1:200, Jackson ImmunoResearch, Cat. 715-545-150), Alexa Fluor 594-conjugated AffiniPure Donkey anti-IgG 1:200, (Jackson ImmunoResearch, Cat. 711-585-152), anti-RNA Polymerase 2 (1:1000, Millipore-Sigma, Cat.

05-623), anti-GFP antibody (1: 500, Abcam, Cat. Ab13970), anti-RFP antibodies (1: 500, Abcam, Cat. ab124754).

Validation

anti-DLG4/PSD95 (1:500; Invitrogen, Cat. 51-6900, rabbit polyclonal IgG, IHC,Velasco et al., 2019 (PMID: 31168097)), anti-GPHN (1:500; Synaptic Systems, Cat. 147011, mouse monoclonal IgG1 mAb7a, IHC, Lorenzo et al., 2020 (PMID: 32054836)), anti-SYP (1:2000, Sigma-Aldrich, Cat. SAB4200544, mouse monoclonal SVP-38, IHC, D'Cruz et al., 2012 (PMID: 22970294)), anti-SOX5 (1:500; Abcam, Cat. ab94396, Rabbit polyclonal IgG, ChIP, IHC, Shibata et al. 2011 (PMID: 21368052)), anti-BCL11B/CTIP2 (1:500, Abcam, Cat. ab18465, rat monoclonal IgG [25B6], IHC, Shim et al. 2012 (PMID: 22678282)), Alexa Fluor 488-conjugated AffiniPureanti-IgG (1:200, Jackson ImmunoResearch, Cat. 715-545-150, Donkey polyclonal IgG, IHC, Sousa et al. 2017 (PMID: 29170230)), Alexa Fluor 594-conjugated AffiniPure Donkey anti-IgG 1:200, (Jackson ImmunoResearch, Cat. 711-585-152, Donkey polyclonal IgG, IHC, Sousa et al. 2017 (PMID: 29170230)), anti-RNA Polymerase 2 (1:1000, Millipore-Sigma, Cat. 05-623, mouse monolonal IgG (CTD4H8), ChIP, Kwan et al. 2005 (PMID: 18840685)), anti-GFP antibody (1: 500, Abcam, Cat. Ab13970, Chicken polyclonal IgY, IHC, Shim et al. 2012 (PMID: 22678282)), anti-RFP antibodies (1: 500, Abcam, Cat. ab124754, Rabbit polyclonal IgG, IHC, Rho et al., 2019 (PMID: 31006651))

Eukaryotic cell lines

Policy information about cell lines

Cell line source(s)	Neuro-2a, which has been originated from mouse brain was purchased from ATCC.
Authentication	Morphology or genotyping.
Mycoplasma contamination	All lines tested negative for mycoplasma contamination, checked monthly using the MycoAlert Mycoplasma Detection Kit (Lonza).
Commonly misidentified lines (See ICLAC register)	No commonly misidentified lines were used.

Animals and other organisms

Policy information about studies involving animals; ARRIVE guidelines recommended for reporting animal research

Laboratory animals	B6SJLF1/J (The Jackson Laboratory) and CD-1 (Charles River laboratories) mice were used for the generation of transgenic reporter mice, and humanized Cbln2 E2 knock-in mice. Mice lines maintained on the C57BL6/6J (The Jackson Laboratory) background, were used at the following time points: E16, E18 and postnatal day P0, P7, P30, and P60. Information on housing conditions for the mice was added in Method section. All relevant details of primate samples (Rhesus macaque and human) were described and put in the Source data file. Gender of mouse samples used was not specified in this study.
Wild animals	The study did not involve wild animals
Field-collected samples	The study did not involve samples collected from the wild.

Human research participants

Policy information about studies involving human research participants

Population characteristics	*Describe the covariate-relevant population characteristics of the human research participants (e.g. age, gender, genotypic information, past and current diagnosis and treatment categories). If you filled out the behavioural & social sciences study design questions and have nothing to add here, write "See above."*
Recruitment	*Describe how participants were recruited. Outline any potential self-selection bias or other biases that may be present and how these are likely to impact results.*