ARTICLE

doi:10.1038/nature11094

Cis-regulatory control of corticospinal system development and evolution

Sungbo Shim[1], Kenneth Y. Kwan[1], Mingfeng Li[1], Veronique Lefebvre[2] & Nenad Šestan[1]

The co-emergence of a six-layered cerebral neocortex and its corticospinal output system is one of the evolutionary hallmarks of mammals. However, the genetic programs that underlie their development and evolution remain poorly understood. Here we identify a conserved non-exonic element (E4) that acts as a cortex-specific enhancer for the nearby gene *Fezf2* (also known as *Fezl* and *Zfp312*), which is required for the specification of corticospinal neuron identity and connectivity. We find that SOX4 and SOX11 functionally compete with the repressor SOX5 in the transactivation of E4. Cortex-specific double deletion of *Sox4* and *Sox11* leads to the loss of *Fezf2* expression, failed specification of corticospinal neurons and, independent of *Fezf2*, a *reeler*-like inversion of layers. We show evidence supporting the emergence of functional SOX-binding sites in E4 during tetrapod evolution, and their subsequent stabilization in mammals and possibly amniotes. These findings reveal that SOX transcription factors converge onto a *cis*-acting element of *Fezf2* and form critical components of a regulatory network controlling the identity and connectivity of corticospinal neurons.

The emergence and expansion of the neocortex in mammals has been crucial to the evolution of complex perceptual, cognitive, emotional and motor abilities[1–3]. The neocortex is organized into six layers based largely on the distinct subtypes of excitatory projection (or pyramidal) neurons and their patterns of connectivity[4–9]. Upper-layer (L2, L3 and L4) projection neurons form synaptic connections solely with other cortical neurons. By contrast, the majority of neurons in the deeper layers (L5 and L6) project to subcortical regions. Studies of laminar inversion in *reeler* mice lacking the reelin (RELN) protein[10–14] have shown that neuron identity and connectivity are determined by birth order rather than by laminar position, suggesting that neuronal specification and positioning are largely separately encoded.

The layer-specific pattern of connectivity is dependent on cortical areas. The long-range projections of L5 neurons in somatosensory-motor areas form the corticospinal (CS) system that directly connects the neocortex with various subcortical regions[15–22]. A major component of the system, the CS (or pyramidal) tract, descends through the brainstem and into the spinal cord to provide a high degree of direct control over the precise motor functions affected in many clinical conditions[21,22]. Despite these important functional implications, the genetic programs controlling CS system development and evolution remain unclear.

Phenotypic specification and evolution of neural circuits depend on precise regulation of the timing, location and level of gene expression[23,24]. Transcriptional control via *cis*-regulatory elements has emerged as a crucial mechanism[25–29]. The *cis*-regulatory mechanisms underlying the specification of distinct neuronal cell types and circuits, however, remain poorly understood. Specification of CS neurons and the formation of the CS tract critically depend on *Fezf2*, which encodes a zinc-finger transcription factor highly enriched in early cortical progenitor cells and their deep-layer neuron progenies[30–35]. Inactivation of *Fezf2* disrupts the molecular specification of deep-layer neurons and the formation of corticofugal projections, including the CS tract, without affecting the inside-out pattern of neurogenesis and lamination[33–35]. Misexpression of *Fezf2* in L2 or L3 cortico-cortical projection neurons[35,36] or striatal interneurons[37] alters their molecular profile and induces ectopic subcerebral projections. These findings indicate that the precisely regulated transcription of *Fezf2* is probably critical to the proper specification of distinct types of cortical projection neurons.

In this study, we used bacterial artificial chromosome (BAC) engineering and genetic inactivation in mice to identify and characterize a cortex-specific *Fezf2* enhancer and its *trans*-regulators. We show that three SOX transcription factors converge onto the *Fezf2* enhancer to control CS system development via functional binding sites that emerged in tetrapods. We also found that *Sox4* and *Sox11* are required for *Fezf2*-independent regulation of cortical RELN expression and laminar organization. Thus, these findings reveal novel developmental genetic programs that control layer formation and CS neuron identity, and the regulatory mechanisms by which they may have evolved.

E4 enhancer controls cortical *Fezf2* expression

Previously, it has been shown that the BAC transgenic mouse harbouring 200 kb of the mouse *Fezf2* locus and the *Gfp* reporter gene (*Fezf2-Gfp*; ref. 38) recapitulates the spatio-temporal expression of endogenous *Fezf2* (refs 39, 40). This indicated that the *cis*-regulatory elements required for cortical *Fezf2* expression are located within the BAC sequence. On the basis of the remarkable similarity in cortical expression pattern of *Fezf2* between mouse and human[39,41], we proposed that the regulatory elements are also highly evolutionarily conserved. Comparative sequence analysis revealed several conserved non-exonic elements (CNEEs) within the *Fezf2-Gfp* BAC (Fig. 1a). To test whether the selected CNEEs regulate *Fezf2-Gfp* expression, we generated multiple lines of BAC transgenic mice in which one of the CNEEs within the *Fezf2-Gfp* BAC was deleted (ΔE1 to ΔE4; Fig. 1b and Methods). We analysed GFP expression in whole-mount brain preparations and tissue sections of transgenic mice at embryonic and postnatal ages (Fig. 1c–g and Supplementary Figs 1 and 2). We found no pronounced change in ΔE1, ΔE2 and ΔE3 mutants, which, like

[1]Department of Neurobiology and Kavli Institute for Neuroscience, Yale University School of Medicine, New Haven, Connecticut 06510, USA. [2]Department of Cell Biology and Orthopaedic and Rheumatologic Research Center, Cleveland Clinic Lerner Research Institute, Cleveland, Ohio 44195, USA.

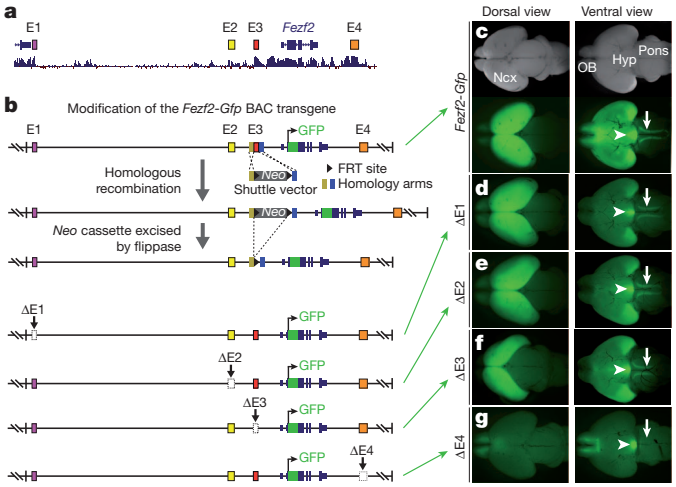


Figure 1 | Identification of a cortex-specific *Fezf2* enhancer. **a**, The locations of CNEEs (E1 to E4) analysed in this study are indicated. **b**, Deletion of each CNEE from the *Fezf2-Gfp* BAC and transgenesis. A positive-selection neomycin cassette (*Neo*) flanked by homology arms was inserted by homologous recombination, resulting in the deletion of each CNEE. After the removal of *Neo* by flippase, the modified BACs (ΔE1 to ΔE4) were used for transgenesis. **c**–**g**, Whole-mount brains of P0 control *Fezf2-Gfp* (**c**) and founder-mutant (**d**–**g**) transgenic mice. $n \geq 3$ founders per mutant line. In *Fezf2-Gfp* mice (**c**), GFP was expressed in the neocortex (Ncx), olfactory bulb (OB), hypothalamus (Hyp) (arrowhead) and CS axons in the pons (arrow). The deletion of E4 (**g**), but not E1 to E3 (**d**–**f**), led to a specific loss of GFP expression in the neocortex and CS axons, but not in the olfactory bulb or hypothalamus.

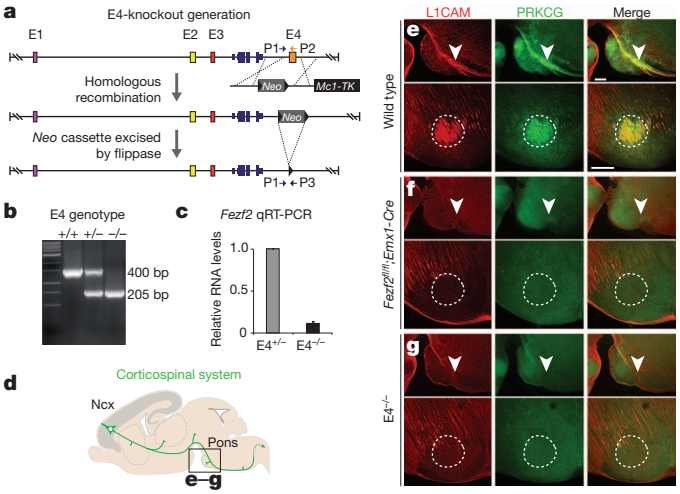


Figure 2 | Loss of neocortical *Fezf2* expression and CS axons in E4-knockout mice. **a**, Generation of E4-knockout mice ($E4^{-/-}$) using the *pGK-Neo/Mc1-TK* selection cassette. **b**, Duplex PCR genotyping of wild-type and mutant alleles using primers P1, P2 and P3. **c**, Analysis of neocortical *Fezf2* expression by quantitative RT–PCR (qRT–PCR). Normalized to *Gapdh*, *Fezf2* messenger RNA levels were significantly reduced in the $E4^{-/-}$ mice compared with heterozygous littermate controls ($E4^{+/-}$). $P = 2.1 \times 10^{-6}$; one-tailed Student's *t*-test; $n = 3$ per genotype. Error bars; s.e.m. **d**, Schematic depiction of the CS system. **e**–**g**, Sagittal (top row) and coronal (bottom row) sections of the pons from wild-type (**e**), cortex-specific *Fezf2*-knockout (**f**) and $E4^{-/-}$ (**g**) mice immunostained for axon marker L1CAM (red) and PRKCG (green). The near-complete loss of CS axons (arrowheads and dashed outlines) in the $E4^{-/-}$ mice is a phenocopy of cortex-specific *Fezf2* deletion. Scale bars, 200 μm.

wild-type *Fezf2-Gfp* mice, expressed GFP in neocortex and pontine CS axons (Fig. 1d–f). By contrast, the deletion of CNEE E4 (ΔE4), which is located 7.3 kb downstream of the *Fezf2* transcription start site, resulted in a drastic loss of GFP expression in the cortex but not in the hypothalamus or the olfactory bulb (Fig. 1g). Moreover, in ΔE4 mice, GFP-positive CS axon fascicles were lost from the ventral surface of the pons (Fig. 1g). Consistent with this finding, E4 was identified as an *in vivo* binding site of the enhancer-associated protein EP300 in mouse embryonic forebrain[27]. Together, these results demonstrate that E4 is a *cis*-regulatory module acting as a cortex-specific enhancer of *Fezf2*.

Loss of *Fezf2* expression and CS axons in $E4^{-/-}$ mice

To test the function of the E4 enhancer directly, we deleted the E4 region, while leaving intact the entire *Fezf2* coding and proximal promoter regions, using homologous recombination in embryonic stem cells (Fig. 2a, b). Mice lacking E4 ($E4^{-/-}$) were viable, fertile and without overt behavioural or motor phenotypes. Quantitative reverse-transcription PCR (RT–PCR) analysis revealed that *Fezf2* expression is drastically downregulated in neocortex of these mice (Fig. 2c). Immunostaining of tissue sections of E4-null mice for PRKCG and L1CAM, two proteins expressed by CS neurons and their axons[11] (Fig. 2d, e), revealed that CS axons were absent from the pons (Fig. 2g), similar to mutant mice with a cortex-specific deletion of *Fezf2* (*Fezf2*$^{fl/fl}$;*Emx1-Cre*) (Fig. 2f). Furthermore, the expression of *Bcl11b* (*Ctip2*), a gene functioning downstream of *Fezf2* (refs 33, 36), was also downregulated in the E4-null neocortex (Supplementary Fig. 3). Taken together, these results indicate that the E4 enhancer is required for neocortical *Fezf2* expression, molecular specification of L5/6 neurons and the formation of CS tract.

SOX4 and SOX11 bind and activate the E4 enhancer

We have previously shown that SOX5, a SOXD member of the large family of SOX transcription factors, binds to and represses the transcriptional activity of the E4 enhancer[39]. Unexpectedly, SOX5 itself is also required for CS tract formation independent of its repression of *Fezf2* (ref. 39). Because different SOX transcription factors are known to compete for a common motif to mediate both activation and repression of regulatory elements[42–44], we proposed that other SOX members may activate the E4 enhancer, perhaps acting competitively against SOX5-mediated repression. Consistent with this hypothesis, our sequence analysis using MATINSPECTOR (Genomatix) revealed eight putative SOX-binding sites in E4. To prioritize which SOX transcription factors may be good candidates for potential *trans*-regulators of the E4 enhancer, we searched for those that are most highly correlated in their spatio-temporal expression pattern with *FEZF2*, using the Human Brain Transcriptome database[45] (http://www.humanbrain-transcriptome.org). The highest-correlated *SOX* genes were three members of the *SOXC* group (*SOX4*, *SOX11* and *SOX12*) and *SOX5* (Supplementary Fig. 4). The transcription factors encoded by these *SOX* genes are crucial in regulating cell fate and differentiation[42–44,46], and a *de novo* deletion of *SOX11* was described in a patient with autism and intellectual disability[47]. Moreover, *Sox4* and *Sox11* act as transcriptional activators[45] and their expression patterns overlap with that of *Fezf2* in developing cortex (Supplementary Fig. 5).

To determine whether SOX4 or SOX11 binds E4, we performed chromatin immunoprecipitation (ChIP)-PCR assays in Neuro-2a cells transiently expressing V5-tagged SOX4 or SOX11 (Methods). Anti-V5 antibodies precipitated E4 DNA, but not E1, E2 or E3 DNA, confirming binding of SOX4 and SOX11 to the E4 enhancer (Fig. 3a). Moreover, recruitment of RNA polymerase II to E4 occurred in the presence of SOX4 or SOX11, which is consistent with increased transcriptional activity. To test the functional consequence of *SoxC*-gene expression on E4, we expressed a luciferase reporter under the control of E4 (pGL4-E4) in Neuro-2a cells (Fig. 3b). Luciferase activity driven by the E4 enhancer was significantly increased by co-transfected *Sox4* or *Sox11*, but not *Sox12*.

To dissect which sequences within E4 drive *Fezf2* expression, we generated four truncated versions of the E4 sequence (E4F1to E4F4; Fig. 3c). The luciferase activity of the E4F2 fragment was increased by SOX4 and SOX11, but not by SOX12 (Fig. 3d). Using

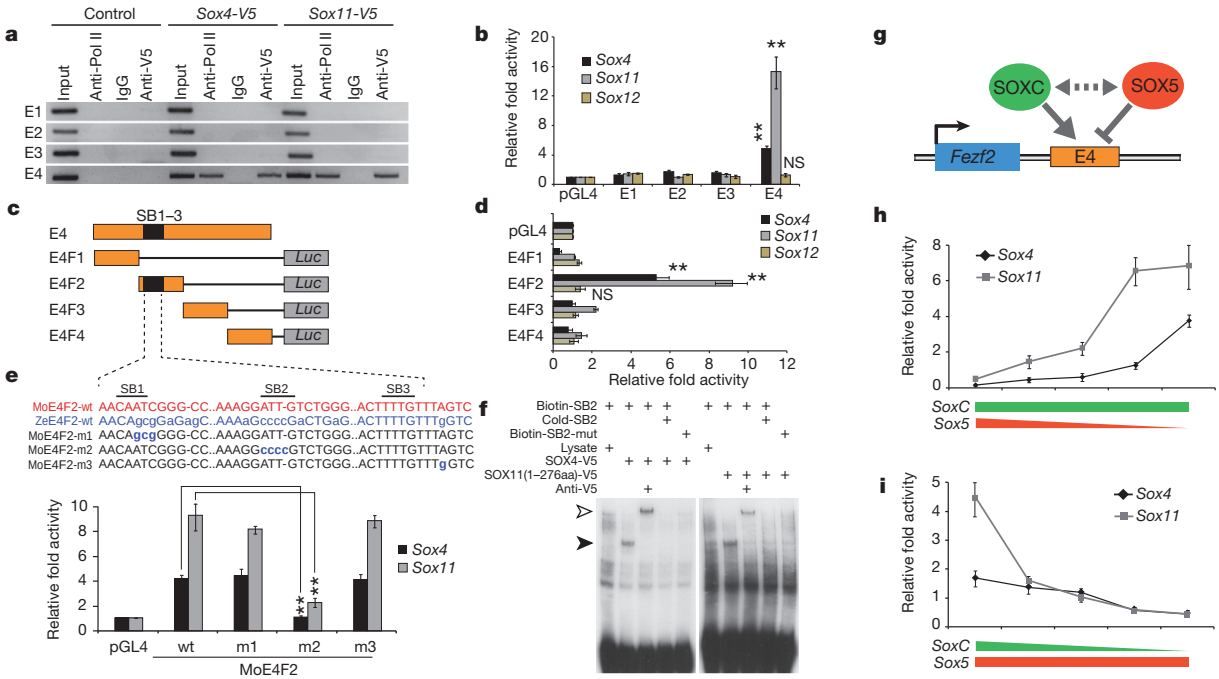


Figure 3 | SOX4 and SOX11 bind to and activate E4 via competition with SOX5. **a**, Chromatin immunoprecipitation (ChIP) from Neuro-2a cells expressing V5-tagged SOX4 and SOX11. Captured DNA was analysed by PCR using primers specific for E1–E4. SOX4 and SOX11 bound and recruited RNA polymerase II (Pol II) to E4 but not E1–E3. **b**, Analysis of SOXC transactivation using empty luciferase vectors (pGL4) or luciferase vectors containing E1–E3. The activity of E4, but not E1–E3, was significantly increased by *Sox4* (≥4-fold) and *Sox11* (≥13-fold), but not by *Sox12* (≤1.5-fold). **c**, **d**, Analysis of functional elements within E4 using luciferase vectors containing deletion fragments of E4 (E4F1 to E4F4) (**c**). E4F2, but not the other fragments, was significantly activated by co-transfection of *Sox4* (≥5-fold) and *Sox11* (≥9-fold), but not of *Sox12* (≤1.5 fold) (**d**). NS, not significant. **e**, Analysis of putative SOX-binding sites (SB1 to SB3) using mouse (Mo) E4F2 luciferase vectors mutagenized by replacement (blue lowercase nucleotides) with zebrafish (Ze) sequence. Targeted mutation of SB2 (MoE4F2-m2) significantly diminished the transactivating ability of SOX4 and SOX11. **f**, Electrophoretic mobility shift assay with biotin-labelled SB2 DNA. SOX4-V5 and SOX11(1–276aa)-V5 shifted wild-type SB2 DNA (arrowhead) but not mutated SB2 DNA. SOXC–SB2 complexes were supershifted (open arrowhead) by an anti-V5 antibody. **g**, Schematic model of E4 regulation by SOX5 and SOXC proteins. **h**, **i**, Analysis of competition between SOX5 and SOXC proteins using the E4-containing luciferase vector. Decreasing concentrations of co-transfected *Sox5* led to a dose-dependent increase in E4 activation in response to *Sox4* and *Sox11* (**h**), whereas decreasing concentrations of co-transfected *Sox4* and *Sox11* led to a dose-dependent increase in *Sox5* repression of E4 (**i**). One-tailed Student's *t*-test; $^{**}P < 0.01$; $n = 4$ per condition. Error bars, s.e.m.

MATINSPECTOR, six potential SOX-binding sites within the E4F2 sequence were predicted. To define the precise base pairs with which SOXC proteins interact, we first tested whether the zebrafish E4F2 sequence, which has 74.2% sequence identity with mouse E4F2, is activated by SOXC proteins. Remarkably, the zebrafish E4F2 sequence was not activated by SOX11 (Supplementary Fig. 6), indicating that the SOXC-interacting sequences of mouse E4F2 lie within the sequences that are divergent in zebrafish. Of the predicted SOX-binding sites in mouse E4F2, three putative SOX-binding sites (SB1 to SB3) are absent from zebrafish E4F2. To test the function and specificity of these sites, we generated mutant versions of E4F2 by substituting SB1, SB2 or SB3 with the zebrafish sequence (Fig. 3e). Only mutations of the mouse SB2 site significantly attenuated the ability of SOX4 and SOX11 to activate the luciferase reporter gene, indicating that this site is crucial to species differences in SOXC-mediated transactivation of E4. Next we assessed the ability of SOX4 and SOX11 to bind to the SB2 site *in vitro*, using an electrophoretic mobility shift assay. We synthesized V5-tagged SOX4 and a higher-affinity, but equally specific, truncated form of SOX11 (SOX11(1–276aa)), because the binding affinity of native SOX11 is weak[43]. Biotinylated SB2 DNA was shifted in the presence of SOX4-V5 or SOX11(1–276aa)-V5 and supershifted to a higher molecular weight by an anti-V5 antibody (Fig. 3f), but not when excess unlabelled or mutated SB2 DNA was used. Therefore, SOX4 and SOX11 directly bind to and activate the transcription activity of E4 via SB2.

To test whether SOX4 and SOX11 functionally compete with the repressor SOX5 in the transactivation of E4 (Fig. 3g), we used the pGL4-E4 luciferase assay. Luciferase activity was significantly increased with increasing amounts of co-transfected *Sox11* and, to a lesser extent, *Sox4*, whereas increased amounts of *Sox5* significantly decreased luciferase activity (Fig. 3h, i). Taken together, our sequence analysis and assays, both *in vitro* and *in vivo*, demonstrate that SOX4 and SOX11 functionally compete with SOX5 repression to activate *Fezf2* transcription via direct binding to sites within E4.

Loss of *Fezf2* expression and CS axons in cdKO mice

To test the phenotypic consequences of *Sox4* and *Sox11* inactivation, we generated cortex-specific deletions of *Sox4* and *Sox11* using the *Emx1-Cre* line (Methods) to circumvent the prenatal lethality associated with constitutive deletions of the two genes[44]. Single conditional knockout mice (cKO: $Sox4^{fl/fl}$;$Sox11^{fl/+}$;*Emx1-Cre* and $Sox4^{fl/+}$;$Sox11^{fl/fl}$;*Emx1-Cre*) were both viable and fertile, whereas the conditional double knockout mice (cdKO: $Sox4^{fl/fl}$;$Sox11^{fl/fl}$;*Emx1-Cre*) died within the first postnatal week (data not shown). Inspection of the cerebrum during the first postnatal week revealed no obvious gross defects in single *Sox4* or *Sox11* cKO animals (Supplementary Fig. 7). However, in the absence of both *Sox4* and *Sox11*, we observed a reduction in the size of the cerebral hemispheres and the olfactory bulb.

To assess the requirement of *Sox4* and *Sox11* in E4-mediated activity, we transfected primary cortical neurons cultured from heterozygous littermate control or *Sox4;Sox11* cdKO embryos with control or E4-containing luciferase constructs (Fig. 4a). In control neurons, the presence of E4 increased luciferase activity by a factor of 2.5 ± 0.3 ($P = 1.9 \times 10^{-4}$; one-tailed Student's *t*-test). This increase was abolished in neurons from *Sox4;Sox11* cdKO (factor of 1.1 ± 0.1; $P = 0.866$), indicating that the two SOXC transcription factors are major activators of the enhancer in cortical neurons.

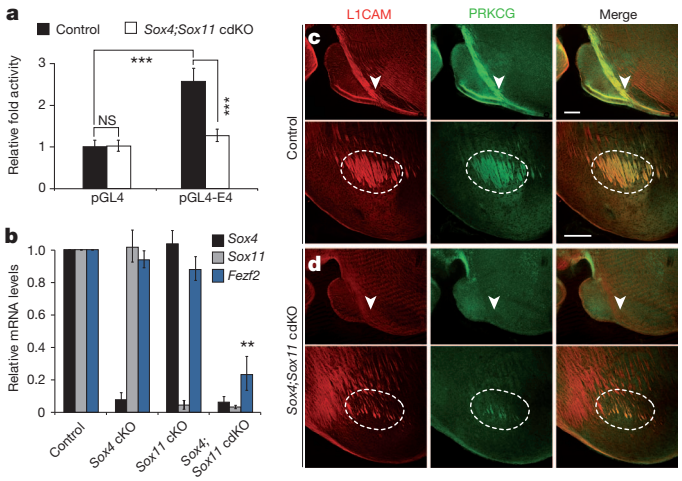


Figure 4 | *Sox4* and *Sox11* are required for *Fezf2* expression and CS tract formation. **a**, The requirement of SOX4 and SOX11 for E4 transactivation was determined in neurons cultured from E14.5 heterozygous littermate control mice and *Sox4*;*Sox11* cdKO mice, and transfected with the E4 luciferase construct. The activity of E4 in double-knockout neurons was not significantly above background (1.1-fold). **b**, Analysis of neocortical *Fezf2* expression by quantitative RT–PCR in cortex-specific *Sox4* and/or *Sox11* mutants. Normalized to *Gapdh*, *Fezf2* messenger RNA (mRNA) levels were drastically reduced in cdKO mice but not in *Sox4* or *Sox11* mutants. **c**, **d**, Sagittal (top row) and coronal (bottom row) sections of the control and cdKO pons immunostained for L1CAM (red) and PRKCG (green). The drastic loss of L1CAM- and PRKCG-positive CS tract axons in the cdKO mice (arrowhead and dashed outline) is similar to that in the cortex-specific *Fezf2* mutant. One-tailed Student's *t*-test; $^{**}P < 0.01$, $^{***}P < 0.001$; $n = 4$ per genotype. Errors bars; s.e.m.; scale bars, 200 μm.

Next, using quantitative RT–PCR, we found a significant reduction in neocortical *Fezf2* expression in cdKO littermates compared with control or single cKO littermates at postnatal day 0 (P0) (Fig. 4b). In addition, immunostaining for L1CAM and PRKCG revealed a drastic loss of CS axons in *Sox4*;*Sox11* cdKO mice but not in single cKO littermates at P0 and P6 (Fig. 4c, d and Supplementary Fig. 8), whereas the organization of other brainstem tracts was not affected. Because the absence of L1CAM and PRKCG labelling could reflect their down-regulated expression, as opposed to loss of CS axons, in the cdKO mice, we confirmed the absence of the CS tract using the CRE-responsive *CAG-Cat-Gfp* transgenic line to express GFP in all cortical projection neurons and their axons[39,40] (Supplementary Fig. 9).

Two additional observations indicate that the loss of *Fezf2* and CS axons in these mice was not due to an absence of L5 neurons. First, only a moderate increase in cell death was detected in the somatosensory-motor areas, in which CS axons originate (Supplementary Fig. 10). Second, many BCL11B-immunopositive L5 neurons were present in the somatosensory-motor areas of cdKO mice, although lightly immunostained (Supplementary Fig. 11), indicating that neurons that would normally give rise to the CS tract were present. Analysis of additional laminar markers further revealed an inversion of cortical layers similar to that of the *reeler* mutant mouse[10–12] (Supplementary Fig. 11), indicating that *Sox4* and *Sox11* are also required for proper laminar positioning of neurons. Previous studies have shown that the CS tract is present in *reeler* mice[12], indicating that this laminar phenotype occurs independently of *Fezf2* and probably in response to defects in the RELN signalling pathway[13,14]. In support of this possibility, RELN was absent from cdKO mouse neocortex (Supplementary Fig. 12). Thus, the combined deletion of *Sox4* and *Sox11* leads to defects in the laminar position of neurons and the molecular specification and connectivity of CS neurons.

Functional and evolutionary implications of E4 sequence variations

The differences in SOXC-mediated transcription between mouse E4F2 and zebrafish E4F2 suggest that species differences in the E4F2 sequence have functional implications. Analysis of SB1, SB2 and SB3 sites within the E4F2 sequence in 23 species revealed that SB2 is conserved in all analysed mammals and the two available non-mammalian amniotes (chicken and lizard) (Supplementary Fig. 13). To investigate the functional consequences of species differences, we used luciferase assays in Neuro-2a cells to analyse the activity of E4F2 sequence from different vertebrates (Fig. 5a, b) in response to SOX11, a more powerful transactivator than SOX4 (Fig. 3). Of the constructs containing an E4F2 sequence of a mammal (human, chimpanzee, macaque, or mouse) or non-mammalian amniote (chicken), which have conserved SB1-3 sequences, the luciferase reporter activity was strongly increased following co-transfection with *Sox11*. The reporter activity of a non-amniote tetrapod (*Xenopus*) E4F2 construct was moderately increased compared with the empty plasmid, but was not as high as that of the mammalian constructs ($P = 3.9 \times 10^{-16}$; one-tailed Student's *t*-test; Fig. 5b). By contrast, the reporter construct containing the zebrafish E4F2 sequence, the SB2 sequence of which is highly divergent from those of mammals, exhibited only basal level activity similar to the empty control luciferase plasmid. To confirm that this is dependent on sequence variations between species, we mutated zebrafish E4F2 by 'murinizing' its SB1, SB2 or SB3 sequence (ZeE4F2-m1, -m2 or -m3, respectively). Reporter activity of the murinized zebrafish SB2 (ZeE4F2-m2), but not SB1 or SB3, was robustly induced by SOX11 (Fig. 5c). ZeE4F2-m2 murinization, however, was insufficient to restore the level of expression observed in the wild-type mouse sequence, suggesting the potential contribution of additional sequences. Thus, evolutionary differences in the E4F2 sequence, especially within SB2, are directly related to functional differences in the ability of SOX11 to activate this regulatory element.

Next, to investigate the functional consequences of species differences in the FEZF2 coding sequence, we tested the ability of zebrafish *fezf2* to rescue the effects of mouse *Fezf2* deficiency. We electroporated *in utero* the neocortical wall of the floxed *Fezf2* (*Fezf2*$^{fl/fl}$) mice[40] with *Cre* and CRE-responsive *Gfp*-expressing constructs with or without *fezf2* at embryonic day 12.5 (E12.5) (Fig. 5d). Analysis of electroporated mice at P0 revealed that *fezf2* was sufficient to rescue the formation of the CS tract in *Fezf2*-null mouse neurons (Fig. 5e). Taken together, these results suggest that sequence variations in the E4 enhancer, but not the *Fezf2* coding region, gave *Fezf2* an essential role in CS tract evolution.

Discussion

Our data provide critical insight into the genes and regulatory components controlling CS system development, centring on the *cis*-regulation of *Fezf2* by SOX4 and SOX11. We show that genetic inactivation of the E4 *cis*-regulatory module results in compartmentalized phenotypic effects largely limited to the loss of cortical *Fezf2* expression and CS system. The spatio-temporal dynamics of cortical *Fezf2* is further controlled by SOX5 and TBR1, two transcription factors expressed post-mitotically in projection neurons[39,40,46,48,49]. On its initial activation in early cortical progenitors, *Fezf2* is highly expressed in deep-layer neurons during early corticogenesis but subsequently repressed in L6 neurons by SOX5 and TBR1 to create a postnatal L5-enriched pattern. The earlier expression of SOX4 and SOX11 is consistent with their role in *Fezf2* activation before the L6 upregulation of SOX5, which is a functional competitor. Despite what is known about the function and regulation of *Fezf2*, a number of key questions have yet to be addressed. First, additional, as yet untested, *trans*-regulators and regulatory elements probably contribute to *Fezf2* regulation in a context-dependent manner. These include the mediators of *Fezf2* repression in L2, L3 and L4 projection neurons. Second, the distinct, and seemingly incongruous, effects of SOX5 in repressing *Fezf2* but being independently required for CS tract formation[39] have not been resolved. Third, the direct transcriptional targets of FEZF2 in CS neurons remain largely unknown.

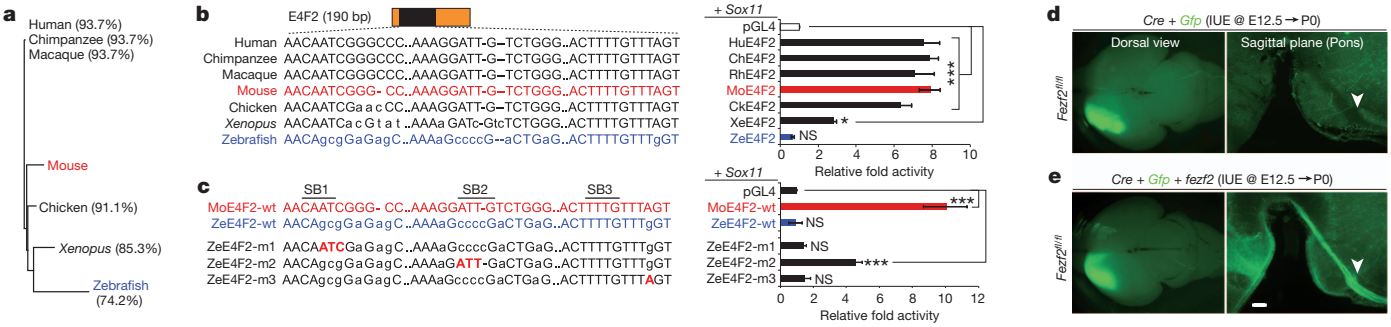


Figure 5 | Functional analysis of species differences in E4 sequence. **a**, Hierarchical clustering of E4F2 sequences from seven vertebrates. Percentage nucleotide identity relative to mouse E4F2 is indicated in parentheses. **b**, Analysis of species differences in *Sox11* activation of E4F2. *Sox11* transactivated the E4F2 sequence of four mammals and chick (≥6.3-fold) and *Xenopus* (2.8-fold). By contrast, the activity of zebrafish E4F2, which is divergent in SB1 and SB2, was not activated by *Sox11*. **c**, Murinization (red uppercase nucleotides) of zebrafish SB2 (ZeE4F2-m2) partly rescued the loss of transactivation of wild-type zebrafish E4F2 (ZeE4F2-wt) by *Sox11*. **d**, **e**, Cell-autonomous rescue of mouse *Fezf2* loss of function by zebrafish *fezf2*. *In utero* electroporation (IUE) of $Fezf2^{fl/fl}$ neocortical wall at E12.5 with *Cre* and CRE-responsive *Gfp* plasmids. *Fezf2*-deficient L5 neurons do not form CS tract at P0 (**d**). Co-electroporation of an *fezf2* plasmid cell-autonomously rescued the formation of CS tract by *Fezf2*-deficient neurons (**e**). One-tailed Student's *t*-test; $*P < 0.05$, $***P < 0.001$; $n = 4$ per condition. Errors bars, s.e.m.; scale bar, 200 μm.

The laminar position and radial organization of projection neurons are key features of cortical development, which rely on RELN signalling[10–14]. We also show that SOX4 and SOX11 are crucial in regulating RELN expression and the inside-out pattern of cortical layer formation, independent of E4 or *Fezf2* and probably involving interactions with distinct regulatory elements. Moreover, SOX4 and SOX11 have additional roles, as in mice lacking both genes, the cortex and olfactory bulb are smaller and cell death is increased. Thus, SOX4 and SOX11 have pleiotropic functions, which are probably mediated by distinct regulatory elements and downstream target genes that are involved in multiple developmental processes.

Our results indicate that, following their emergence in tetrapods, functional SOX-binding sites have retained high conservation through purifying selection in mammals and some amniotes, thus directly linking species variations in regulatory sequences to functional outcomes. E4 sequence substitutions in SB2 may constitute an evolutionary turning point for *Fezf2* function during forebrain development, possibly facilitating the formation of descending telencephalic pathways including the CS system. Whereas minor projections from the ventral (subpallial) telencephalon to the spinal cord are present in amphibians[2], direct dorsal telencephalo-spinal projections resembling the CS tract have been reported only in mammals and some birds[2,50]. We propose that the concurrent emergence of the described regulatory mechanisms and direct telencephalo-spinal projections in early amniotes, together with subsequent changes in genetic programs driving the patterning and expansion of a six-layered dorsal pallium, made possible the evolution of the CS system in mammals.

METHODS SUMMARY

Selected CNEE sequences were deleted from a *Fezf2-Gfp* BAC by recombination-mediated genetic engineering, and GFP expression from modified BACs was analysed by transgenesis. To confirm the requirement of E4 enhancer for *Fezf2* expression, a germline E4 deletion mutant was generated. Putative E4 transactivators SOX4 and SOX11, identified by E4 sequence and co-expression analyses, were tested using ChIP and luciferase reporter assays. To analyse the requirement of *Sox4* and *Sox11* for *Fezf2* expression and cortical development, we generated cortex-specific single and double *Sox4* and *Sox11* null mice. Complete materials and methods are described in Supplementary Information.

Full Methods and any associated references are available in the online version of the paper at www.nature.com/nature.

Received 7 November 2011; accepted 28 March 2012.
Published online 30 May 2012.

1. Northcutt, R. G. & Kaas, J. H. The emergence and evolution of mammalian neocortex. *Trends Neurosci.* **18,** 373–379 (1995).
2. Nieuwenhuys, R., ten Donkelaar, H. J. & Nicholson, C. *The Central Nervous System of Vertebrates* (Springer, 1998).
3. Butler, A. B., Reiner, A. & Karten, H. J. Evolution of the amniote pallium and the origins of mammalian neocortex. *Ann. NY Acad. Sci.* **1225,** 14–27 (2011).
4. O'Leary, D. D. M. & Koester, S. E. Development of projection neuron types, axon pathways, and patterned connections of the mammalian cortex. *Neuron* **10,** 991–1006 (1993).
5. Rash, B. G. & Grove, E. A. Area and layer patterning in the developing cerebral cortex. *Curr. Opin. Neurobiol.* **16,** 25–34 (2006).
6. Molyneaux, B. J., Arlotta, P., Menezes, J. R. L. & Macklis, J. D. Neuronal subtype specification in the cerebral cortex. *Nature Rev. Neurosci.* **8,** 427–437 (2007).
7. Leone, D. P., Srinivasan, K., Chen, B., Alcamo, E. & McConnell, S. K. The determination of projection neuron identity in the developing cerebral cortex. *Curr. Opin. Neurobiol.* **18,** 28–35 (2008).
8. Hansen, D. V., Rubenstein, J. L. & Kriegstein, A. R. Deriving excitatory neurons of the neocortex from pluripotent stem cells. *Neuron* **70,** 645–660 (2011).
9. Kwan, K. Y., Sestan, N. & Anton, E. S. Transcriptional co-regulation of neuronal migration and laminar identity in the neocortex. *Development* **139,** 1535–1546 (2012).
10. Caviness, V. S. & Sidman, R. L. Time of origin of corresponding cell classes in cerebral-cortex of normal and *reeler* mutant mice: autoradiographic analysis. *J. Comp. Neurol.* **148,** 141–151 (1973).
11. Steindler, D. A. & Colwell, S. A. *Reeler* mutant mouse: maintenance of appropriate and reciprocal connections in the cerebral cortex and thalamus. *Brain Res.* **113,** 386–393 (1976).
12. Terashima, T. Anatomy, development and lesion-induced plasticity of rodent corticospinal tract. *Neurosci. Res.* **22,** 139–161 (1995).
13. Bar, I., de Rouvroit, C. L. & Goffinet, A. M. The evolution of cortical development. An hypothesis based on the role of the Reelin signaling pathway. *Trends Neurosci.* **23,** 633–638 (2000).
14. Rice, D. S. & Curran, T. Role of the Reelin signaling pathway in central nervous system development. *Annu. Rev. Neurosci.* **24,** 1005–1039 (2001).
15. Joosten, E. A. J. & Bar, D. P. R. Axon guidance of outgrowing corticospinal fibres in the rat. *J. Anat.* **194,** 15–32 (1999).
16. Martin, J. H. The corticospinal system: from development to motor control. *Neuroscientist* **11,** 161–173 (2005).
17. Canty, A. J. & Murphy, M. Molecular mechanisms of axon guidance in the developing corticospinal tract. *Prog. Neurobiol.* **85,** 214–235 (2008).
18. Lemon, R. N. Descending pathways in motor control. *Annu. Rev. Neurosci.* **31,** 195–218 (2008).
19. Rathelot, J.-A. & Strick, P. L. Subdivisions of primary motor cortex based on cortico-motoneuronal cells. *Proc. Natl Acad. Sci. USA* **106,** 918–923 (2009).
20. Nudo, R. J. & Masterton, R. B. Descending pathways to the spinal-cord. IV. Some factors related to the amount of cortex devoted to the corticospinal tract. *J. Comp. Neurol.* **296,** 584–597 (1990).
21. ten Donkelaar, H. J. *et al.* Development and malformations of the human pyramidal tract. *J. Neurol.* **251,** 1429–1442 (2004).
22. Eyre, J. A. Corticospinal tract development and its plasticity after perinatal injury. *Neurosci. Biobehav. Rev.* **31,** 1136–1149 (2007).
23. Jessell, T. M. Neuronal specification in the spinal cord: inductive signals and transcriptional codes. *Nature Rev. Genet.* **1,** 20–29 (2000).
24. Hobert, O., Carrera, I. & Stefanakis, N. The molecular and gene regulatory signature of a neuron. *Trends Neurosci.* **33,** 435–445 (2010).
25. Wray, G. A. The evolutionary significance of *cis*-regulatory mutations. *Nature Rev. Genet.* **8,** 206–216 (2007).
26. Carroll, S. B. Evo-devo and an expanding evolutionary synthesis: a genetic theory of morphological evolution. *Cell* **134,** 25–36 (2008).
27. Visel, A. *et al.* ChIP-seq accurately predicts tissue-specific activity of enhancers. *Nature* **457,** 854–858 (2009).
28. Davidson, E. H. Emerging properties of animal gene regulatory networks. *Nature* **468,** 911–920 (2010).

29. Williamson, I., Hill, R. E. & Bickmore, W. A. Enhancers: from developmental genetics to the genetics of common human disease. *Dev. Cell* **21,** 17–19 (2011).
30. Hashimoto, H. *et al.* Expression of the zinc finger gene fez-like in zebrafish forebrain. *Mech. Dev.* **97,** 191–195 (2000).
31. Matsuo-Takasaki, M., Lim, J. H., Beanan, M. J., Sato, S. M. & Sargent, T. D. Cloning and expression of a novel zinc finger gene, *Fez*, transcribed in the forebrain of *Xenopus* and mouse embryos. *Mech. Dev.* **93,** 201–204 (2000).
32. Inoue, K., Terashima, T., Nishikawa, T. & Takumi, T. *Fez1* is layer-specifically expressed in the adult mouse neocortex. *Eur. J. Neurosci.* **20,** 2909–2916 (2004).
33. Molyneaux, B. J., Arlotta, P., Hirata, T., Hibi, M. & Macklis, J. D. *Fezl* is required for the birth and specification of corticospinal motor neurons. *Neuron* **47,** 817–831 (2005).
34. Chen, B., Schaevitz, L. R. & McConnell, S. K. *Fezl* regulates the differentiation and axon targeting of layer 5 subcortical projection neurons in cerebral cortex. *Proc. Natl Acad. Sci. USA* **102,** 17184–17189 (2005).
35. Chen, J. G., Rasin, M. R., Kwan, K. Y. & Sestan, N. *Zfp312* is required for subcortical axonal projections and dendritic morphology of deep-layer pyramidal neurons of the cerebral cortex. *Proc. Natl Acad. Sci. USA* **102,** 17792–17797 (2005).
36. Chen, B. *et al.* The *Fezf2-Ctip2* genetic pathway regulates the fate choice of subcortical projection neurons in the developing cerebral cortex. *Proc. Natl Acad. Sci. USA* **105,** 11382–11387 (2008).
37. Rouaux, C. & Arlotta, P. *Fezf2* directs the differentiation of corticofugal neurons from striatal progenitors *in vivo*. *Nature Neurosci.* **13,** 1345–1347 (2010).
38. Gong, S. C. *et al.* A gene expression atlas of the central nervous system based on bacterial artificial chromosomes. *Nature* **425,** 917–925 (2003).
39. 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).
40. Han, W. Q. *et al.* TBR1 directly represses *Fezf2* to control the laminar origin and development of the corticospinal tract. *Proc. Natl Acad. Sci. USA* **108,** 3041–3046 (2011).
41. Fertuzinhos, S. *et al.* Selective depletion of molecularly defined cortical interneurons in human holoprosencephaly with severe striatal hypoplasia. *Cereb. Cortex* **19,** 2196–2207 (2009).
42. Bergsland, M., Werme, M., Malewicz, M., Perlmann, T. & Muhr, J. The establishment of neuronal properties is controlled by Sox4 and Sox11. *Genes Dev.* **20,** 3475–3486 (2006).
43. Dy, P. *et al.* The three SoxC proteins-Sox4, Sox11 and Sox12-exhibit overlapping expression patterns and molecular properties. *Nucleic Acids Res.* **36,** 3101–3117 (2008).
44. Bhattaram, P. *et al.* Organogenesis relies on SoxC transcription factors for the survival of neural and mesenchymal progenitors. *Nature Commun.* **1,** 9 (2010).
45. Kang, H. J. *et al.* Spatio-temporal transcriptome of the human brain. *Nature* **478,** 483–489 (2011).
46. Lai, T. *et al.* SOX5 controls the sequential generation of distinct corticofugal neuron subtypes. *Neuron* **57,** 232–247 (2008).
47. Lo-Castro, A. *et al.* Deletion 2p25.2: a cryptic chromosome abnormality in a patient with autism and mental retardation detected using aCGH. *Eur. J. Med. Genet.* **52,** 67–70 (2009).
48. Bedogni, F. *et al. Tbr1* regulates regional and laminar identity of postmitotic neurons in developing neocortex. *Proc. Natl Acad. Sci. USA* **107,** 13129–13134 (2010).
49. McKenna, W. L. *et al. Tbr1* and *Fezf2* regulate alternate corticofugal neuronal identities during neocortical development. *J. Neurosci.* **31,** 549–564 (2011).
50. Wild, J. M. & Williams, M. N. Rostral wulst in passerine birds. I. Origin, course, and terminations of an avian pyramidal tract. *J. Comp. Neurol.* **416,** 429–450 (2000).

Supplementary Information is linked to the online version of the paper at www.nature.com/nature.

Acknowledgements We thank W. Han, Y. Imamura Kawasawa, D. Liu and T. Nottoli for technical help; A. Giraldez and A. M. M. Sousa for reagents; and the Sestan laboratory for discussions. This work was supported by the National Institutes of Health (NS054273, MH081896, AR54153), the March of Dimes Foundation and a McDonnell Scholar Award (N.S.).

Author Contributions S.S., K.Y.K. and N.S. designed the research; S.S. performed the experiments; S.S. and K.Y.K. performed the confocal imaging, M.L. analysed coexpression and deep sequencing data; S.S., K.Y.K. and N.S. analysed the other data; V.L. generated mice with floxed *Sox4* and *Sox11* alleles; N.S. conceived the study; and S.S., K.Y.K. and N.S. wrote the manuscript. All authors discussed and commented on the data.

Author Information Reprints and permissions information is available at www.nature.com/reprints. The authors declare no competing financial interests. Readers are welcome to comment on the online version of this article at www.nature.com/nature. Correspondence and requests for materials should be addressed to N.S. (nenad.sestan@yale.edu).

METHODS

Animals. All experiments were performed in accordance with a protocol approved by Yale University's Committee on Animal Research. The generation of mice with the floxed *Sox4* and *Sox11* alleles (*Sox4*$^{fl/fl}$ and *Sox11*$^{fl/fl}$, respectively) was described elsewhere[44,51]. The *CAG-Cat-Gfp* and *Emx1-Cre* P1 artificial chromosome transgenic mice were a gift from Melissa Colbert[52] and Takuji Iwasato[53], respectively. The *Fezf2-Gfp* BAC transgenic mouse was generated by the GENSAT project[38].

Generation of BAC transgenic mouse. To generate BAC reporter constructs harbouring deletions of putative enhancer elements (E1–E4) in the mouse *Fezf2-Gfp* BAC (RP23-141E17-*Gfp*; the GENSAT project[38]), homology arms A and B flanking the sequences targeted for deletion (E1, Chr14: 13,204,475–13,204,944, 470 bp; E2, Chr14: 13,183,314–13,183,991, 678 bp; E3, Chr14: 13,180,797–13,181,283, 487 bp; E4, Chr14: 13,170,100–13,170,960, 861 bp) were amplified by PCR and cloned into the PL451 shuttle vector. SW102 *Escherichia coli* transformed with the *Fezf2-Gfp* BAC was induced for recombination-mediated genetic engineering (recombineering) at 42 °C and electroporated with the linearized shuttle vector. After neomycin selection, *Flpase* expression was induced by arabinose to excise the neomycin cassette[54]. Founder lines were confirmed by PCR using primers corresponding to sequences from E1–E4 (Supplementary Table 1). *Fezf2-Gfp* lines were maintained on a C57BL6/J background and all animals were housed under identical conditions. The intactness of the integrated BACs was confirmed by PCR and deep sequencing of genomic DNA.

Generation of E4-null mice. The targeting vector for the deletion of the E4 enhancer was constructed by a recombineering-based method with an unmodified BAC clone (RP23-141E17). A 6,627-bp fragment containing the E4 enhancer was retrieved from the BAC and inserted into the PL253 vector by recombination in SW102 bacteria. The removal of the E4 sequence was achieved using the strategy described above for the generation of the ΔE4 BAC transgenic mouse. ES cells (C57BL6/J) were electroporated with the resulting targeting construct and expanded under positive (*pGK-Neo*, from the PL451 vector) and negative (*Mc1-TK*, from the PL253 vector) selection. Correctly recombined clones were identified by long-distance PCR, confirmed by sequencing, and used for blastocyst injection and subsequent generation of mice with the targeted allele. The neomycin cassette was subsequently removed by flippase-mediated recombination by breeding with *Flp* mice. The mice were genotyped using the P1/P2 primer set (400 bp) for the wild-type allele and the P1/P3 primer set (205 bp) the E4 deletion allele (Supplementary Table 1).

Immunohistochemistry. P0 and P6 brains were fixed by immersion in 4% paraformaldehyde overnight at 4 °C and sectioned using a vibratome (Leica). Immunostaining was performed as previously described[20]. The following primary antibodies were used: anti-L1CAM (rat, 1:300; Millipore), anti-PRKCG (rabbit, 1:300; Santa Cruz), anti-GFP (chicken, 1:3,000; Abcam), anti-CUX1 (rabbit, 1:150; Santa Cruz), anti-SATB2 (mouse, 1:200; Genway), anti-BCL11B (rat, 1:250; Santa Cruz), anti-ZFPM2 (rabbit, 1:250; Santa Cruz), anti-RELN (mouse, 1:300; Millipore) and anti-CASP3, active (rabbit, 1:200; Millipore). We tested a total of 12 commercially available anti-SOX4 antibodies (Abcam, ab52043, ab86809, ab90696; Abgent, AP2045a, AP2045c; LS Bioscience, LS-B3520; Millipore, AB5803, AB10537; Pierce, PA1-38638, PA1-38639; Santa Cruz Biotechnology, sc-17326; Sigma, HPA029901) and 10 commercially available anti-SOX11 antibodies (Abcam, ab42853; Aviva Systems Biology, ARP33328, ARP38235; LS Bioscience, LS-B1567, LS-C10306; Millipore, AB9090; Pierce, PA5-19728; Santa Cruz Biotechnology, sc-20096, sc-17347; Sigma, HPA000536) on western blots and immunostaining. Three antibodies for each of SOX4 and SOX11 recognized a single band of the expected size on western blots but none of them were suitable for immunostaining or ChIP assay. In addition, a few antibodies showed nuclear immunostaining but also had strong background staining of tissue sections of the knockout brain, so we concluded that they are not specific.

Plasmids. For expression of *SoxC* genes and *Fezf2*, full-length complementary DNAs (mouse *Sox4*, BC052736; mouse *Sox11*, BC062095; mouse *Sox12*, BC067019, zebrafish *fezf2*, BC085677) were inserted into pCAGEN. For ChIP and electrophoretic mobility shift assay, PCR-amplified products (for mouse *Sox4* and *Sox11* full-length complementary DNA and *Sox11(1–276aa)*) were inserted into pcDNA3.1/V5-His TOPO vector (Invitrogen). For luciferase reporter plasmids, PCR-amplified products (for E1 to E4, E4F1 to E4F4 of mouse and E4F2 of additional species) and annealed, 190-bp complementary synthetic oligonucleotides (for MoE4F2-m1-3 and ZeE4F2-m1-3) were inserted into pGL4 (Promega). The sequences of the PCR primers and synthetic oligonucleotides are listed in Supplementary Table 1.

Gene expression analysis. To show which genes in the SOX family are highly correlated with human *FEZF2*, we used the previously generated data[45] for pairwise Pearson comparisons. This data set is generated by exon arrays and is available from the Human Brain Transcriptome database (http://www.humanbraintranscriptome.org) and the NCBI Gene Expression Omnibus under the accession number GSE25219. It covers 16 brain cortical regions over 15 periods, ranging in stage from embryonic development to late adulthood. The evaluation of gene expression in each region and each period was detailed in the study[45]. In this study, we averaged gene expression values of 11 neocortical regions in each of 15 developmental periods represented in the data set, and then performed pairwise Pearson correlation analysis for *FEZF2* and the members of the SOX family. In total, 18 *SOX* genes were represented in the database. The correlation analysis was shown by heatmap and the correlation coefficients were indicated in each cell (Supplementary Fig. 4). Mouse *Fezf2*, *Sox4*, *Sox5*, *Sox11*, and *Sox12 in situ* hybridization images were obtained from the Genepaint database[55] (http://www.genepaint.org).

Chromatin immunoprecipitation. After testing available anti-SOX4 and anti-SOX11 antibodies and finding none to be specific and suitable for ChIP assay, we performed ChIP assay with Neuro-2a cells that has been transfected with pcDNA3-Sox4-V5 or pcDNA3-Sox11-V5 plasmid using Lipofectamine 2000 (Invitrogen). At 36 h after transfection, cells were crosslinked in 1% formaldehyde for 10 min at 37 °C and processed for ChIP assay using the EZ-ChIP kit (Millipore) according to the manufacturer's instructions. Capture of the DNA fragments was tested by PCR using primers 5′-TGGAGAGAAGGCCA ACAAAC-3′ (E1, forward); 5′- GCTGGGGATGGAGAAGAATA-3′ (E1, reverse); 5′-TCACCAAAGCGCCTTTTTAT-3′ (E2, forward); 5′-GTAAGCGG ACATGCCATTTT-3′ (E2, reverse); 5′-TGACTTTCCCCAGCCTTCTA-3′ (E3, forward); 5′-CAACAGCTCACCCACACAAT-3′ (E3, reverse); 5′-ATGCCTA GCCCCAAAGAAAT-3′ (E4, forward) and 5′-TTAACTCCCCCTTTGGC TCT-3′ (E4, reverse).

Electrophoretic mobility shift assay. Double-stranded DNA probes were generated by annealing with complementary single-stranded oligonucleotides. Each DNA probe was biotin-end-labelled with Klenow enzyme and then purified using the G-50 Sephadex Column (Roche). The labelled and unlabelled (cold probe; 100-fold) probes were incubated with SOX4-V5 and SOX11(1–276aa)-V5 fusion proteins, which were produced using a TNT Quick coupled transcription/translation kit (Promega) in a DNA-binding buffer (75 mM NaCl, 1 mM EDTA, 1 mM DTT, 10 mM Tris-HCl (pH 7.5), 6% glycerol, 2 μg BSA, 16 ng poly (dI-dC) and 0.1 μg salmon sperm DNA) at 30 °C for 30 min. For supershift assay, anti-V5 antibody (Invitrogen) was added into the binding reaction and incubated for an additional 10 min. The shift assay was carried out in 6% polyacrylamide gel in ×0.5 TBE buffer. Subsequently, the labelled DNA was transferred on a nylon membrane (Amersham Biosciences). For detection, we used the Chemiluminescent Nucleic Acid Detection Module (Thermo Scientific) according to manufacturers' recommendations. All primer sequences are described in Supplementary Table 1.

Luciferase assays. Neuro-2a cells were transfected using Lipofectamine 2000 (Invitrogen) with one of pCAG-Sox5, pCAG-Sox4, pCAG-Sox11, pCAG-Sox12 or empty pCAGEN, together with one of the pGL4 (Promega) luciferase vectors generated with enhancer sequences as described above. A *Renilla* luciferase plasmid (pRL, Promega) was co-transfected to control for transfection efficiency. The luciferase assays were performed 48 h after transfection using the dual-luciferase kit (Promega) according to the manufacturer's instructions. Primary cortical neurons were prepared from E14.5 heterozygous littermate control (*Sox4*$^{fl/+}$;*Sox11*$^{fl/+}$;*Emx1-cre* and *Sox4*$^{fl/+}$;*Sox11*$^{fl/+}$) and *Sox4;Sox11* cdKO (*Sox4*$^{fl/fl}$;*Sox11*$^{fl/fl}$;*Emx1-cre*) cortices and transfected with pGL4 or pGL4-E4 with pRL using the AMAXA Mouse Neuron Nucleofector Kit (VPG-1001; Lonza). At 48 h after transfection, the luciferase assays were performed as described above.

***In utero* electroporation.** Plasmid DNA (4 μg μl^{-1}) mixture (pCAGEN-Cre;pCALNL-Gfp or pCAGEN-Cre;pCALNL-Gfp;pCAGEN-fezf2) was injected into the lateral ventricles of embryonic mice at E12.5 and transferred into the cells of ventricular zone by electroporation (five 50-ms pulses of 40 V at 950-ms intervals) as described elsewhere[35,39]. Brains and tissue sections of electroporated animals were analysed for GFP expression after fixation with 4% paraformaldehyde at P0.

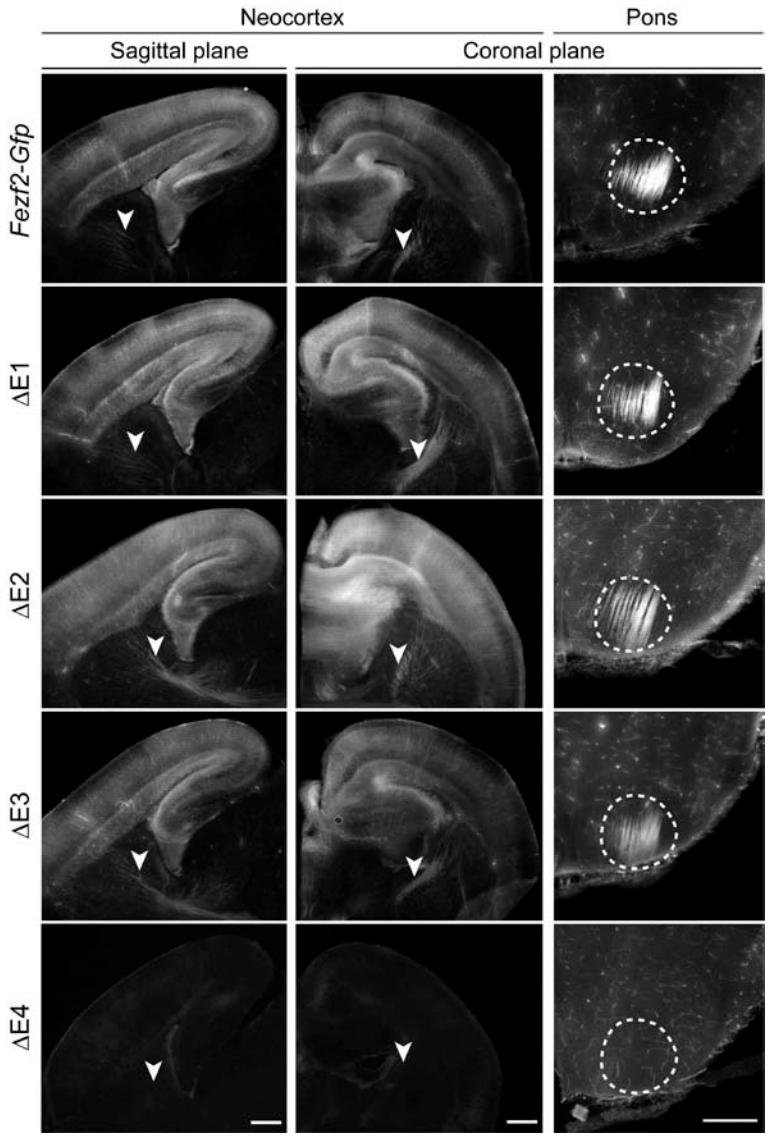
Quantitative RT–PCR. Neocortex was dissected from P0 brain and RNA was isolated using the RNeasy kit (Qiagen). Quantitative RT–PCR was performed using primers listed in Supplementary Table 1 and the SYBR FAST qPCR kit (KAPA Biosystems) according to the manufacturer's instructions. Gene expression levels were normalized to *Gapdh* expression.

51. Penzo-Méndez, A., Dy, P., Pallavi, B. & Lefebvre, V. Generation of mice harboring a Sox4 conditional null allele. *Genesis* **45,** 776–780 (2007).

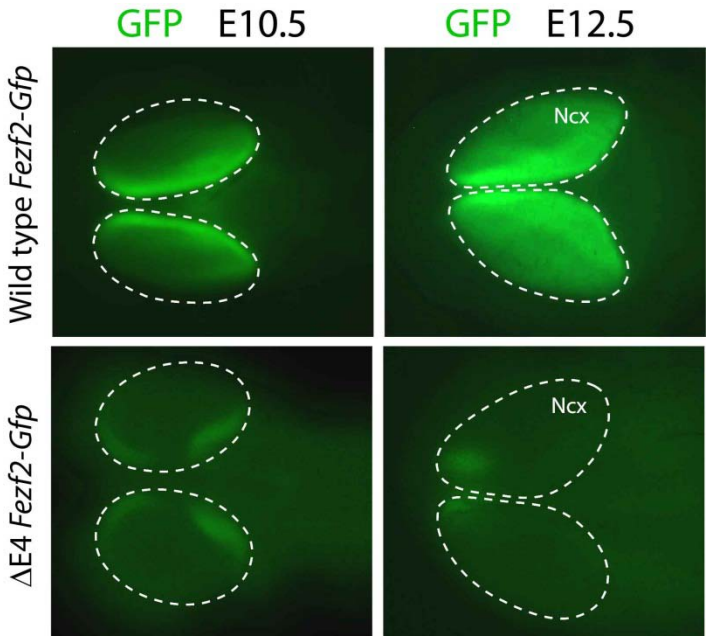
52. Kawamoto, S. *et al.* A novel reporter mouse strain that expresses enhanced green fluorescent protein upon Cre-mediated recombination. *FEBS Lett.* **470,** 263–268 (2000).
53. Iwasato, T. *et al.* Dorsal telencephalon-specific expression of Cre recombinase in PAC transgenic mice. *Genesis* **38,** 130–138 (2004).
54. Liu, P. T., Jenkins, N. A. & Copeland, N. G. A highly efficient recombineering-based method for generating conditional knockout mutations. *Genome Res.* **13,** 476–484 (2003).
55. Visel, A., Thaller, C. & Eichele, G. GenePaint.org: an atlas of gene expression patterns in the mouse embryo. *Nucleic Acids Res.* **32,** D552–D556 (2004).

doi:10.1038/nature11094

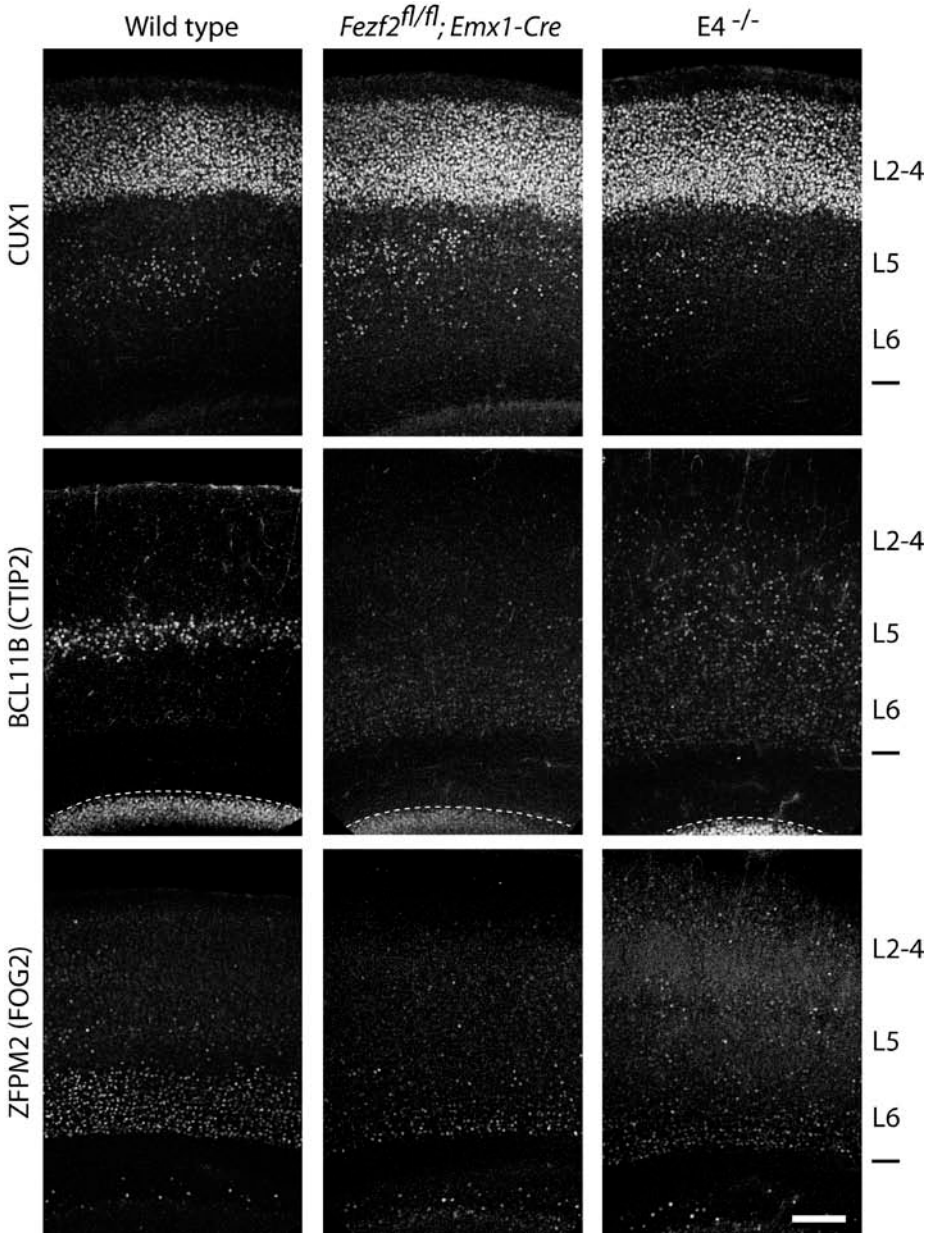
Supplementary Figures



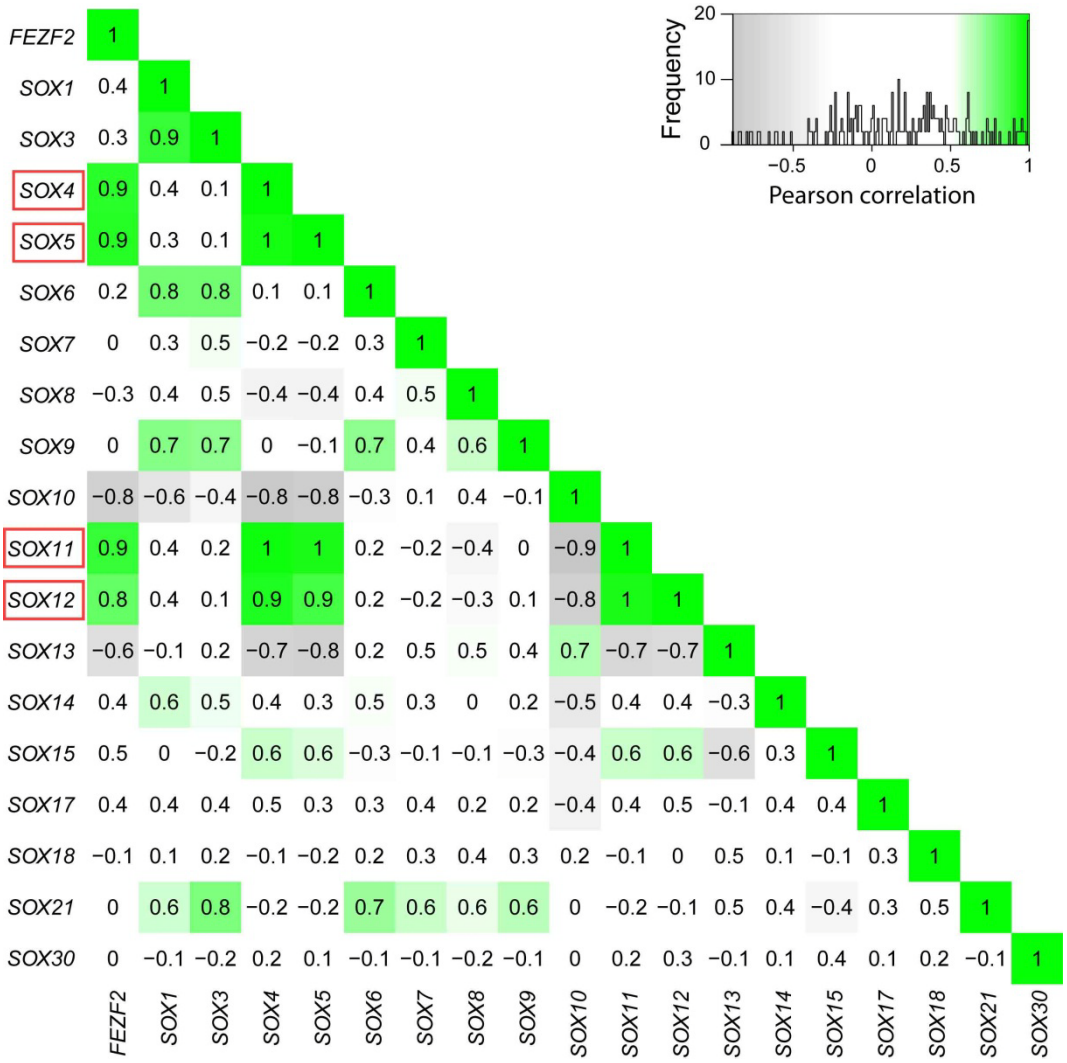
Supplementary Figure 1: Analysis of GFP expression in P0 wild type and mutant (ΔE1-4) *Fezf2-Gfp* BAC transgenic mice. Epifluorescent images of sagittal forebrain and coronal midbrain sections. In the wildtype *Fezf2-Gfp* transgenic brain, GFP was expressed in pyramidal neurons of neocortical L5 and hippocampus. GFP^+ corticofugal axons were present in the striatum (arrowheads) and the pons (dashed outline). Deletion of E1-3 from the BAC transgene (ΔE1-3) did not alter the expression pattern of GFP. Deletion of E4 (ΔE4), in contrast, led to the loss of neocortical and hippocampal GFP expression. GFP^+ corticofugal axons were also lost. Scale bars represent 200 μm.



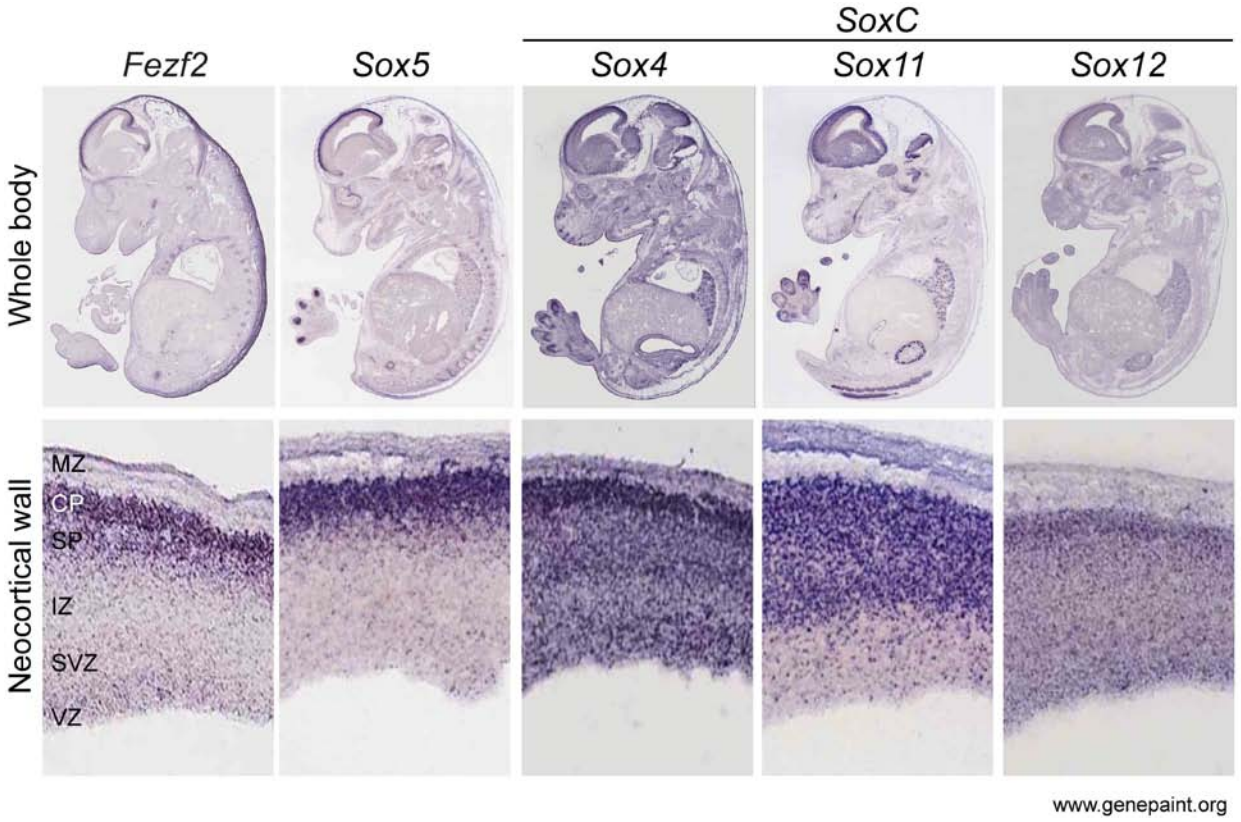
Supplementary Figure 2: Analysis of GFP expression in E10.5 and E12.5 wild type and ΔE4 mutant *Fezf2-Gfp* transgenic mice. Whole-mount brains revealed that deletion of E4 from the BAC transgene led to decreased GFP expression in E10.5 neocortex (Ncx) ventricular zone and loss of GFP expression in E12.5 Ncx. Ventral-medial GFP expression was preserved at E10.5 and 12.5.



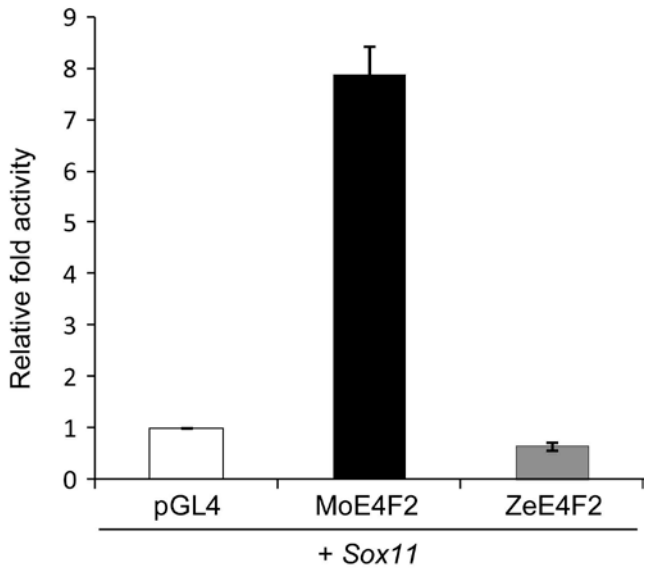
Supplementary Figure 3: Analysis of molecular alterations in cortical laminar identity in cortex-specific *Fezf2* mutant and E4 deletion mutant. Sections of the P0 neocortex were immunostained with the antibodies against proteins indicated on the left. The expression pattern of CUX1, a marker of L2-4 neurons, was not affected by cortical deletion of *Fezf2* (*Fezf2*$^{fl/f}$*;Emx1-Cre*) or deletion of E4 (*E4*$^{-/-}$). In contrast, immunostaining for BCL11B, a marker of subcortically-projecting L5 neurons and, to a lesser extent, L6 neurons, was dramatically reduced in both *Fezf2* and E4-null mice. Weakly BCL11B-immunopositive neurons were present in L6 but some were dispersed to the upper layers. The expression of ZFPM2, a marker of L6 and subplate neurons, was also reduced and only observed in the deep portion of L6 and subplate in both *Fezf2* and E4-null mice. Scale bar represents 200 μm.



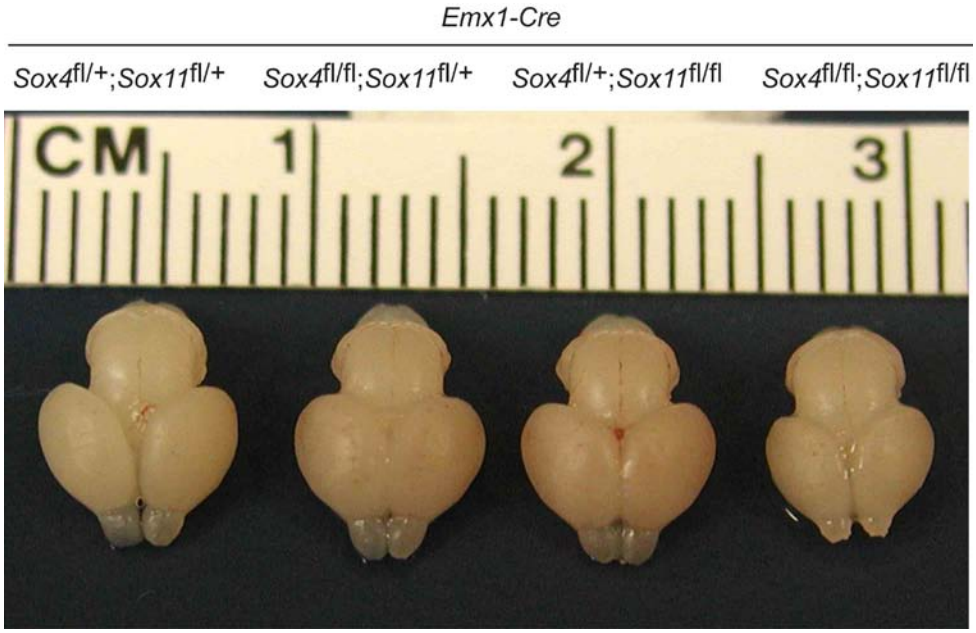
Supplementary Figure 4: Heatmap matrix of pairwise Pearson correlation of spatio-temporal expression dynamics of human *FEZF2* and 18 *SOX* genes in multiple regions of the developing and adult human brain[45]. Gene expression data were generated using exon arrays and retrieved from www.humanbraintranscriptome.org database or from NCBI GEO with the accession number GSE25219. *FEZF2* expression is selectively and highly correlated (correlation coefficient≥ 0.8) with *SOX5* and the three members of the SoxC family (*SOX4*, *SOX11*, and *SOX12*), but not with other *SOX* genes (correlation coefficient ranging from -0.8 to 0.5).



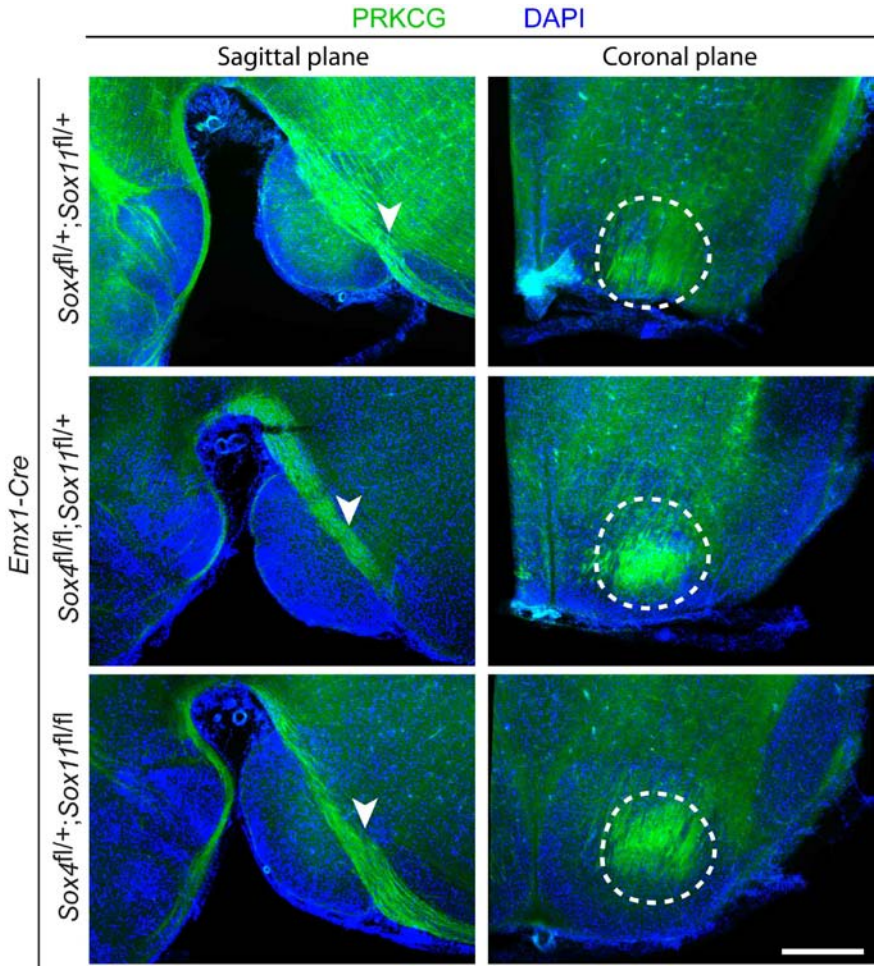
Supplementary Figure 5: The expression patterns of *Fezf2*, *Sox5*, and members of the *SoxC* family during mouse embryonic development. *In situ* hybridization of the E14.5 mouse brain showed that the mRNA patterns of *Sox4* and *Sox11*, but not *Sox12*, significantly overlapped with those of *Fezf2* and *Sox5* in the cortical plate (CP). MZ, marginal zone; SP, subplate; IZ, intermediate zone; SVZ, subventricular zone; VZ, ventricular zone. Images were obtained from the Genepaint database (www.genepaint.org).



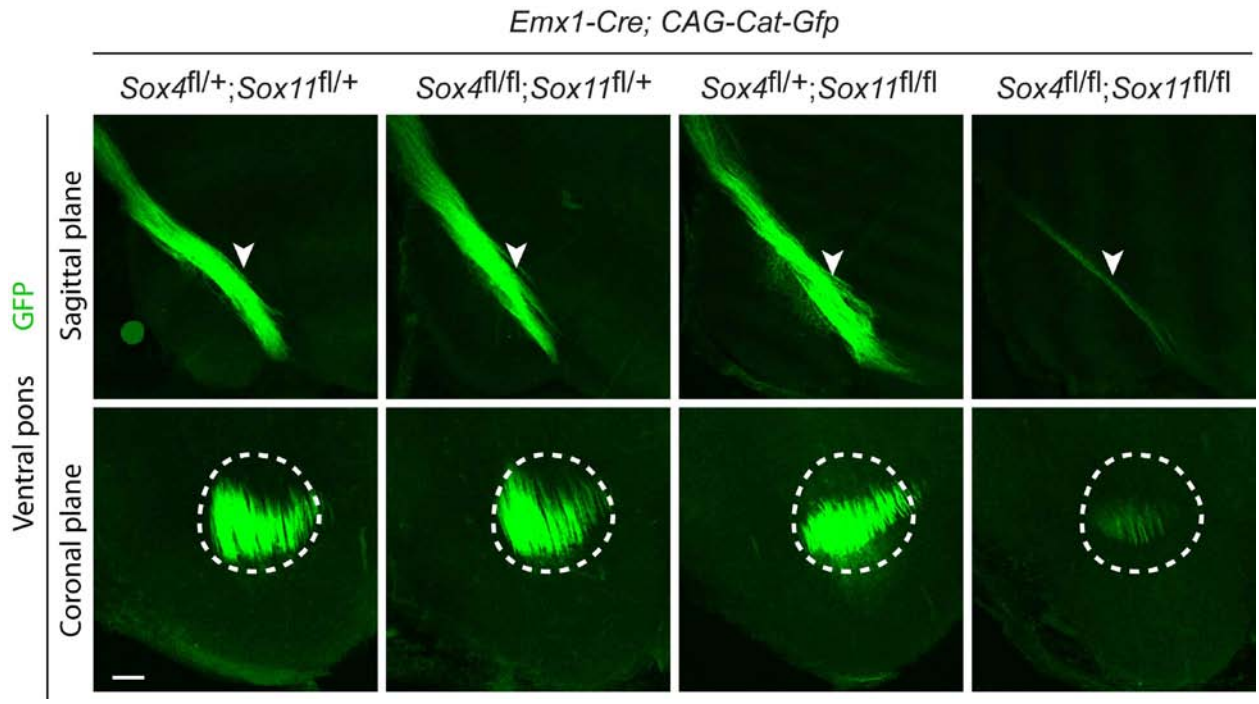
Supplementary Figure 6: Analysis of species-dependence in *Sox11* activation of E4F2 between mouse and zebrafish using a luciferase assay. *Sox11* significantly transactivated mouse (Mo), but not zebrafish (Ze), E4F2 sequence ($P = 2.1\text{x}10^{-6}$, n = 4, one-tailed Student's *t*-test).



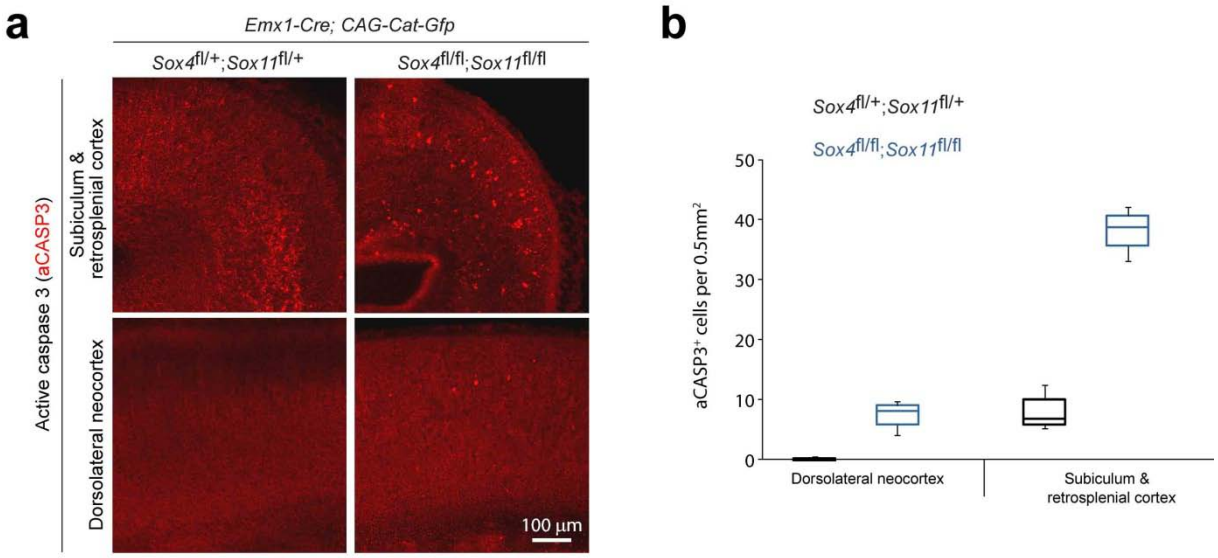
Supplementary Figure 7: Whole-mount brains of P0 *Sox4* and *Sox11* single and double conditional knockout mice (cKO and cdKO, respectively). Cortical single deletion of *Sox4* or *Sox11* did not alter gross morphology of the cerebral cortex. The sizes of the *Sox4*;*Sox11* cdKO cerebral hemispheres and olfactory bulbs were reduced.



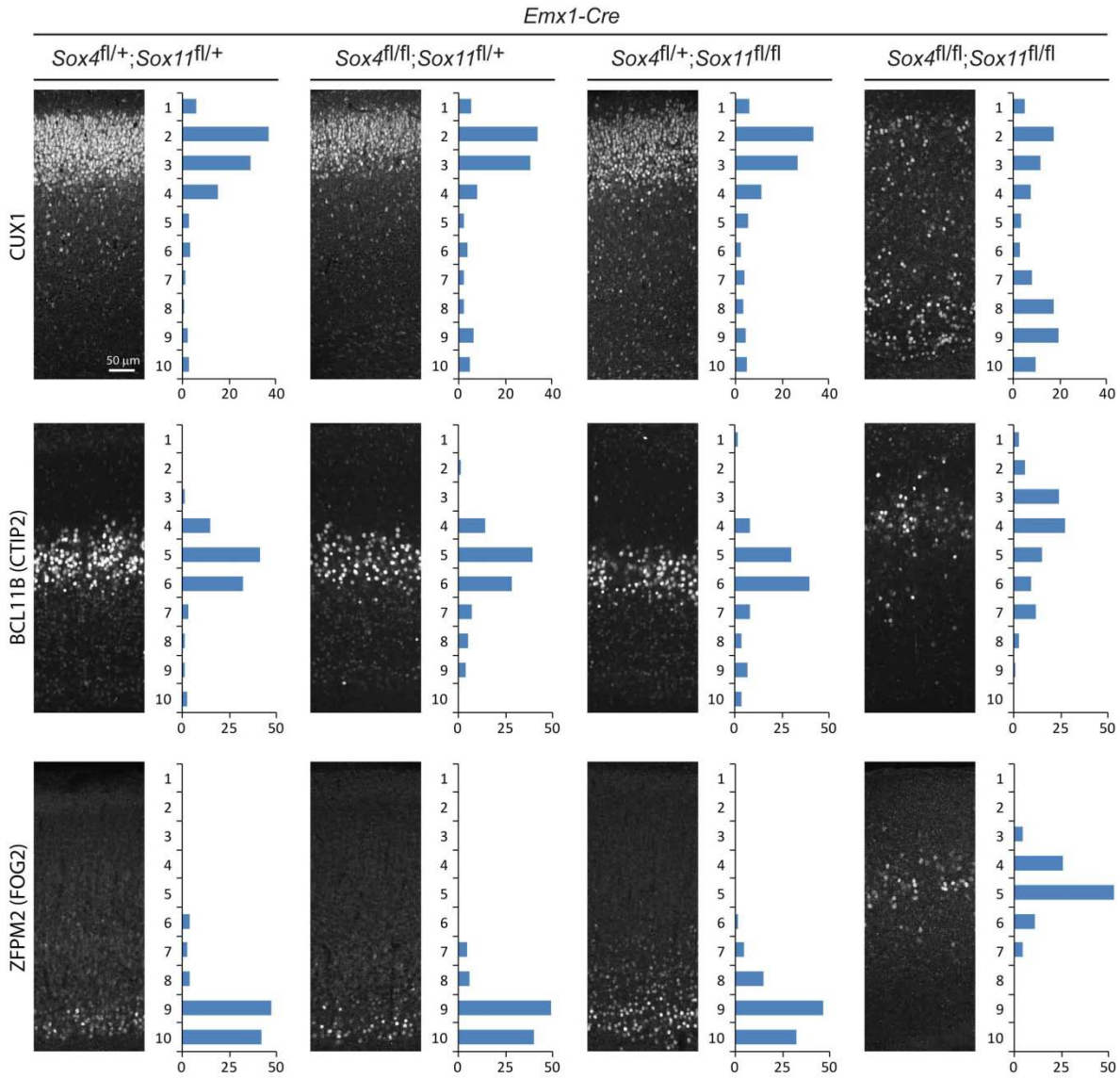
Supplementary Figure 8: Analysis of cortex-specific *Sox4* and *Sox11* single conditional knockout mice. P6 sagittal and coronal sections through the pons were immunostained for PRKCG, a marker of the CS neurons and axons. CS axons (arrowheads in a sagittal plane and dashed outline in a coronal plane) projected normally to the pons in single mutants of *Sox4* (*Sox4*$^{fl/fl}$;*Sox11*$^{fl/+}$; *Emx1-Cre*) or *Sox11* (*Sox4*$^{fl/+}$;*Sox11*$^{fl/fl}$; *Emx1-Cre*). Scale bar represents 200 μm.



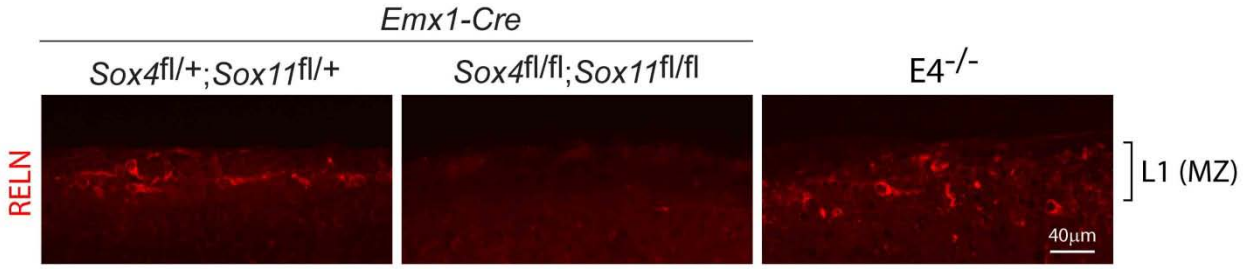
Supplementary Figure 9: Analysis of cortical axons in cortex-specific *Sox4* and *Sox11* single and double knockout mice using a transgenic reporter mouse. Mice doubly transgenic for *Emx1-Cre* and *CAG-Cat-Gfp* express GFP in all cortical projection neurons and their axons, including the CS tract. P0 sagittal and coronal sections through the pons revealed normal CS axons projecting to the pons (arrowheads in a sagittal plane and dashed outline in a coronal plane) in single conditional mutants of *Sox4* (*Sox4*$^{fl/fl}$;*Sox11*$^{fl/+}$;*Emx1-Cre*) or *Sox11* (*Sox4*$^{fl/+}$;*Sox11*$^{fl/fl}$;*Emx1-Cre*). In *Sox4* and *Sox11* double conditional mutants (*Sox4*$^{fl/fl}$;*Sox11*$^{fl/fl}$; *Emx1-Cre*), CS axons were severely diminished. Scale bar represents 100 μm.



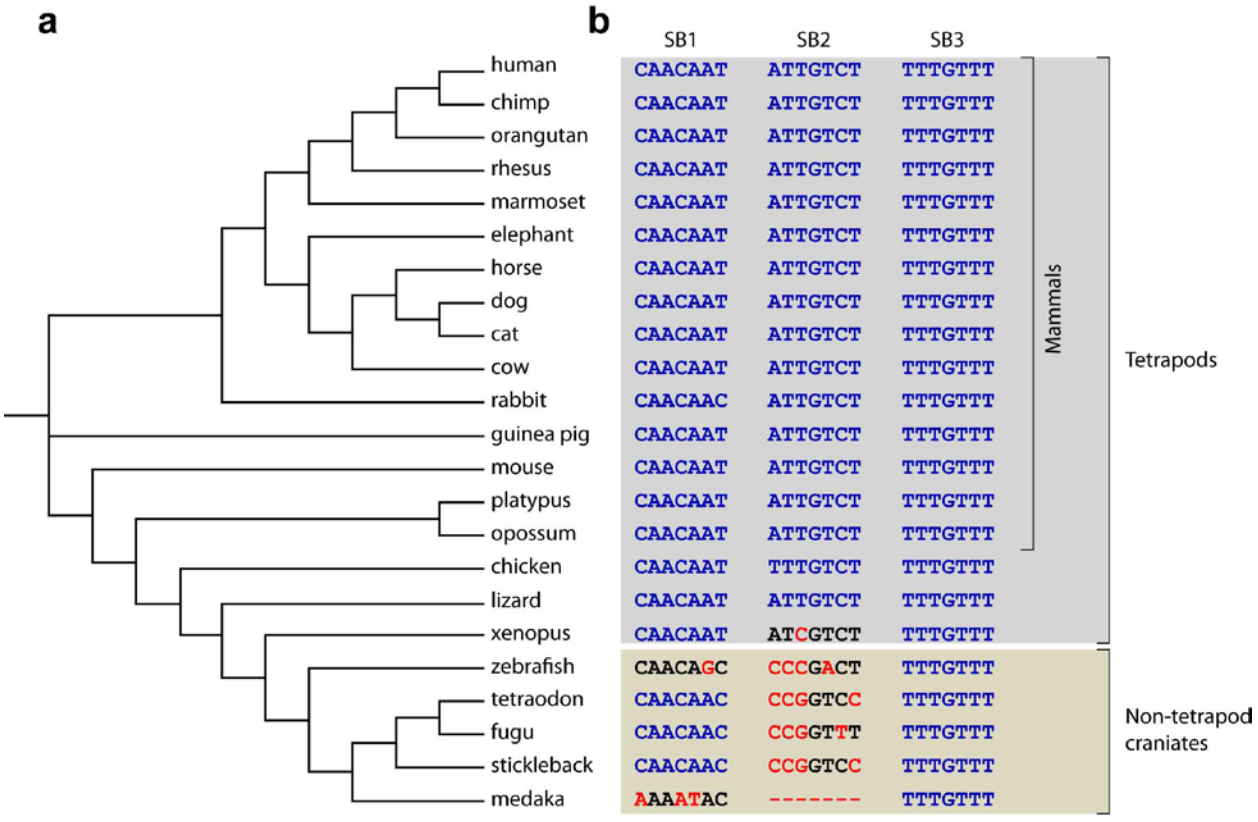
Supplementary Figure 10: Analysis of apoptosis marker expression in cortex-specific *Sox4* and *Sox11* double conditional knockout mice. **a,** Coronal sections of the P0 neocortex were immunostained for activated CASP3. **b,** The expression of activated CASP3 was slightly increased in both retrosplenial cortex/subiculum and dorsolateral fronto-parietal neocortex of cortex-specific *Sox4* and *Sox11* double conditional knockout mice (*Sox4*$^{fl/fl}$;*Sox11*$^{fl/fl}$; *Emx1-Cre*).



Supplementary Figure 11: Analysis of laminar defects in cortex-specific *Sox4* and *Sox11* single and double conditional knockout mice. Sections of the P0 neocortex were immunostained with antibodies against proteins indicated on the left. The expression and distribution of the examined markers were unaltered in the cortex-specific *Sox4* (*Sox4*$^{fl/fl}$;*Sox11*$^{fl/+}$;*Emx1-Cre*) or *Sox11* (*Sox4*$^{fl/+}$;*Sox11*$^{fl/fl}$;*Emx1-Cre*) knockout mice. In the *Sox4* and *Sox11* double mutants (*Sox4*$^{fl/fl}$;*Sox11*$^{fl/fl}$; *Emx1-Cre*), the distribution of neurons expressing CUX1 was shifted towards deep parts of the neocortex, where as the distribution of neurons expressing BCL11B (CTIP2) or ZFPM2 (FOG2) were shifted more superficially. Y axis: bin number, X axis: percentage of immunopositive cells in a bin. Scale bar represents 50 µm.



Supplementary Figure 12: Analysis of RELN expression in *Sox4*, *Sox11* double mutants and E4 mutant. Sections of the P0 neocortex were immunostained for RELN (red), which is normally expressed in the Cajal-Retzius neurons of the layer (L) 1/marginal zone (MZ) in wild-type and E4 mutant ($E4^{-/-}$). RELN-expression is absent from the marginal zone (MZ) of cortex-specific *Sox4* and *Sox11* double mutants ($Sox4^{fl/fl}Sox11^{fl/fl}$; *Emx1-Cre*).



Supplementary Figure 13: Analysis of evolutionary conservation of *Fezf2* coding sequence and SOX-binding sites within the E4 element in craniates. **a,** Dendrogram of *Fezf2* coding sequence aligned using ClustalW and drawn using TreeView. **b,** Analysis of SOX binding sites SB1-3 using TRANSFAC algorithm. Blue type indicates predicted functional binding site whereas black type indicates non-predicted sites. Red type indicates substitutions incompatible with predicted binding. Deletions (-) are also indicated.

Supplementary Table 1

Primers for the generation of BAC transgenic mutant constructs		
Primer name		Sequence
ΔE1 A-arm	Forward	5'- ATCTTGCTTTTTGCCTTAT -3'
	Reverse	5'- CAACAGTGTCAAGAATGCT -3'
ΔE1 B-arm	Forward	5'- TGTTCTCAACAGAGCAGCT -3'
	Reverse	5'- TCCCAATTGCTGGGATTAG -3'
ΔE2 A-arm	Forward	5'- TGTAGCCAGGTATACACT -3'
	Reverse	5'- TCCCAGTTATTTAACTTG -3'
ΔE2 B-arm	Forward	5'- TATCTGCCTCAGGCTGAAG -3'
	Reverse	5'- TTCCAGGTATCGCACAGT -3'
ΔE3 A-arm	Forward	5'- CTGGCACACAACAACATGC -3'
	Reverse	5'- TTTATTTTCCCAGGACAC -3'
ΔE3 B-arm	Forward	5'- GAGGTAGTTCTGCTTTGT -3'
	Reverse	5'- TAATATCAGTTCACCTGC -3'
ΔE4 A-arm	Forward	5'- CCAAAGCCACAGCCTTGTAC -3'
	Reverse	5'- TTTACCGCCTGCTCACCTGA -3'
ΔE4 B-arm	Forward	5'- TGAAAAGCAGGCTCTCCAC -3'
	Reverse	5'- CCAAGCCAAGTGTGAAC -3'

E4 enhancer primers		
Primer name		Sequence
E1	Forward	5'- ggactagt GAGAGACCGCCTAG -3'
	Reverse	5'- ggactagt TGTTGAGAACAACT - 3'
E2	Forward	5'- ggactagt GCTTAACTCCTGCTC - 3'
	Reverse	5'- ggactagt GGGGGTGGGGCGGTG - 3'
E3	Forward	5'- ggactagt AATAAATTTGGAGG - 3'
	Reverse	5'- ggactagt TCTCTCTCTCTCTCTCCCAT - 3'
E4	Forward	5'- ggactagt TGGGCTGGAGGGAG - 3'
	Reverse	5'- ggactagt TCTACTTTTCTAAC - 3'
E4F1	Forward	5'- gaagatct TGGGCTGGAGGGAG - 3'
	Reverse	5'- ggggtacc TCTCGCTGTGTTTTG - 3'
E4F2	Forward	5'- gaagatct AAGATTAACTGGAC - 3'
	Reverse	5'- ggggtacc ATCTGAAGCTATCC - 3'
E4F3	Forward	5'- gaagatct TTTTTTTAAAAAGAG - 3'
	Reverse	5'- ggggtacc GCTCATTAACTCCC - 3'
E4F4	Forward	5'- gaagatct AGCTGACAAACTAC - 3'
	Reverse	5'- ggggtacc TGGCCTGTTGGGTTC - 3'
Human, chimp, macaque-E4F2	Forward	5'- gaagatct AAGATTAAATGGACAAAAGC - 3'
	Reverse	5'- ggggtacc ATCTGAAGCTATCCTGTTAA - 3'
Chick-E4F2	Forward	5'- gaagatct CTCTGAAGCTATCCTGTTAA - 3'
	Reverse	5'- ggggtacc GAGATTAAATGGACAAAAGC - 3'
Xenopus-E4F2	Forward	5'- gaagatct AAGATTAAATGGACAAAAGT - 3'
	Reverse	5'- ggggtacc ATCTGAGGCTATCCTGTTAA - 3'
Zebrafish-E4F2	Forward	5'- gaagatct GATTATAGGGACAAGAGCGCAT - 3'
	Reverse	5'- ggggtacc TGCTGACGCTACCCCCGTTA - 3'

MoE4F2-m1	Forward	5'- gatctAAGATTAACTGGACAAAAGCGTAGTAATGAACTGTCAACAGCGGGG CCATCCACAGAACCAGCATCACTCAAGAACTCGGCATTTACAGGCAACAAA GGATTGTCTGGGACAGGAAGAAAAGACGACCTAGACTTT TGTTTAGTCAAC ACAATTGATTACAAATGAGAGATTAACAGGATAGCTTCAGATggtac -3'
	Reverse	5'- cATCTGAAGCTATCCTGTTAATCTCTCATTTGTAATCAATTGTGTTGACTAA ACAAAAGTCTAGGTCGTCTTTTCTTCCTGTCCCAGACAATCCTTTGTTGCCTGT AAATGCCGAGTTCTTGAGTGATGCTGGTTCTGTGGATGGCCCCGCTGTTGACA GTTCATTACTACGCTTTTGTCCAGTTAATCTTa -3'
MoE4F2-m2	Forward	5'-gatctAAGATTAACTGGACAAAAGCGTAGTAATGAACTGTCAACAATCGGGC CATCCACAGAACCAGCATCACTCAAGAACTCGGCATTTACAGGCAACAAAGG CCCCGTCTGGGACAGGAAGAAAAGACGACCTAGACTTTTGTTTAGTCAACAC AATTGATTACAAATGAGAGATTAACAGGATAGCTTCAGATggtac -3'
	Reverse	5'- cATCTGAAGCTATCCTGTTAATCTCTCATTTGTAATCAATTGTGTTGACTAA ACAAAAGTCTAGGTCGTCTTTTCTTCCTGTCCCAGACGGGGCCTTTGTTGCCT GTAAATGCCGAGTTCTTGAGTGATGCTGGTTCTGTGGATGGCCCGATTGTTGA CAGTTCATTACTACGCTTTTGTCCAGTTAATCTTa -3'
MoE4F2-m3	Forward	5'- gatctAAGATTAACTGGACAAAAGCGTAGTAATGAACTGTCAACAATCGGGC CATCCACAGAACCAGCATCACTCAAGAACTCGGCATTTACAGGCAACAAAGG ATTGTCTGGGACAGGAAGAAAAGACGACCTAGACTTTTGTTTGGTCAACACA ATTGATTACAAATGAGAGATTAACAGGATAGCTTCAGATggtac -3'
	Reverse	5'- cATCTGAAGCTATCCTGTTAATCTCTCATTTGTAATCAATTGTGTTGACCAA ACAAAAGTCTAGGTCGTCTTTTCTTCCTGTCCCAGACAATCCTTTGTTGCCTGT AAATGCCGAGTTCTTGAGTGATGCTGGTTCTGTGGATGGCCCGATTGTTGACA GTTCATTACTACGCTTTTGTCCAGTTAATCTTa -3'
ZeE4F2-m1	Forward	5'-gatctGGGATTATAGGGACAAGAGCGCATTAATGAACTGTCAACAATCGAGA GCATGTACACGACAGAGATCAGCCGAAATTGAGGCATTTACAAGCGACAAAA GCCCCGACTGAGACAGGAAGAAAAAAAGAGCTAGACTTTTGTTTGGTCAACT CAATTGATTGTAAATGGAAGATTAACGGGGGTAGCGTCAGCAggtac -3'
	Reverse	5'- cTGCTGACGCTACCCCCGTTAATCTTCCATTTACAATCAATTGAGTTGACC AAACAAAAGTCTAGCTCTTTTTTTCTTCCTGTCTCAGTCGGGGCTTTTGTCGCT TGTAAATGCCTCAATTTCGGCTGATCTCTGTCGTGTACATGCTCTCGATTGTTG ACAGTTCATTAATGCGCTCTTGTCCCTATAATCCCa -3'
ZeE4F2-m2	Forward	5'-gatctGGGATTATAGGGACAAGAGCGCATTAATGAACTGTCAACAGCGGAGA GCATGTACACGACAGAGATCAGCCGAAATTGAGGCATTTACAAGCGACAAAA GATTGACTGAGACAGGAAGAAAAAAAGAGCTAGACTTTTGTTTGGTCAACTC AATTGATTGTAAATGGAAGATTAACGGGGGTAGCGTCAGCAggtac -3'
	Reverse	5'- cTGCTGACGCTACCCCCGTTAATCTTCCATTTACAATCAATTGAGTTGACC AAACAAAAGTCTAGCTCTTTTTTTCTTCCTGTCTCAGTCAATCTTTTGTCGCTT GTAAATGCCTCAATTTCGGCTGATCTCTGTCGTGTACATGCTCTCCGCTGTTG ACAGTTCATTAATGCGCTCTTGTCCCTATAATCCCa -3'
ZeE4F2-m3	Forward	5'-gatctGGGATTATAGGGACAAGAGCGCATTAATGAACTGTCAACAGCGGAGA GCATGTACACGACAGAGATCAGCCGAAATTGAGGCATTTACAAGCGACAAAA GCCCCGACTGAGACAGGAAGAAAAAAAGAGCTAGACTTTTGTTTAGTCAACT CAATTGATTGTAAATGGAAGATTAACGGGGGTAGCGTCAGCAggtac -3'
	Reverse	5'- cTGCTGACGCTACCCCCGTTAATCTTCCATTTACAATCAATTGAGTTGACTA AACAAAAGTCTAGCTCTTTTTTTCTTCCTGTCTCAGTCGGGGCTTTTGTCGCTT GTAAATGCCTCAATTTCGGCTGATCTCTGTCGTGTACATGCTCTCCGCTGTTG ACAGTTCATTAATGCGCTCTTGTCCCTATAATCCCa -3'

qRT-PCR primers		
Fezf2	Forward	5’- ACACCGGAGCTAGACCGTTT -3’
	Reverse	5’- TCCCTTTTTGGTGAAAGCCT -3’
Sox4	Forward	5’- CGTCTACAAGGTGCGGACTC -3’
	Reverse	5’- CGCGCTTCACTTTCTTGT -3’
Sox11	Forward	5’- GCCGGCTCTACTACAGCTTC -3’
	Reverse	5’- GCTGGATGAGGAGGTGGAC -3’
Gapdh	Forward	5’- GATCAACACGTACCAGTGCAA -3’
	Reverse	5’- TGCCTCGATGGACAGATAGA -3’
Oligonucleotide sequence for EMSA		
SB2-wt	Forward	aaaaAACAAAGGATTGTCTGGGACAGGAAGAAAAGA
	Reverse	aaaaTCTTTTCTTCCTGTCCCAGACAATCCTTTGTT
SB2-mut	Forward	aaaaAACAAAGGCCCCTCTGGGACAGGAAGAAAAGA
	Reverse	aaaaTCTTTTCTTCCTGTCCCAGAGGGGCCTTTGTT