

Non-coding structural variants disrupt *FOXG1* transcriptional regulation in early neurodevelopment

Hamerlinck L., D'haene E. *et al.*

SUPPLEMENTARY INFORMATION

Supplementary Note 1. Clinical description of del8 included in the manuscript

Supplementary Note 2. Clinical description of bca4 included in the manuscript

Supplementary Figure 1. Identification and finemapping of patient deletion.

Supplementary Figure 2. Identification of translocation breakpoints in bca4 case.

Supplementary Figure 3. Overview of all CNVs observed in patients overlapping the bigger *FOXG1* TAD based on the DECIPHER database.

Supplementary Figure 4. 3D chromatin architecture in bca4 case lymphoblastoid cells (LCLs).

Supplementary Figure 5. Detailed view of the enhancer cluster and TAD boundary.

Supplementary Figure 6. Chromatin interaction frequencies at the *FOXG1* and *PRKD1* locus.

Supplementary Figure 7. TAD structure conservation at the *FOXG1* locus across tissues.

Supplementary Figure 8. CTCF binding at the commonly affected regulatory region (CARR) across ENCODE tissues and cell types.

Supplementary Figure 9. RAD21 binding at the commonly affected regulatory region (CARR) across ENCODE tissues and cell types.

Supplementary Figure 10. Single-cell RNA-seq data from human brain development shows high *FOXG1* expression in ventral telencephalic NPC and neurons.

Supplementary Figure 11. Characterization of neural progenitor cell (NPC) populations using regional marker gene expression during general NPC induction.

Supplementary Figure 12. Regulatory dynamics of neural progenitor cell (NPC) markers during general and forebrain-directed NPC differentiation.

Supplementary Figure 13. Quantification of zebrafish enhancer assay results.

Supplementary Figure 14. Representative images zebrafish enhancer assays.

Supplementary Figure 15. Activation of candidate cis-regulatory elements (cCREs) during mouse embryonic development.

Supplementary Figure 16. *FOXG1* and *PRKD1* normalized RNA expression levels during general neural progenitor cell (NPC) induction of heterozygous knockout models.

Supplementary Figure 17. Successful differentiation of WT and CRISPR KO models for the EC and TAD boundary towards NPCs.

Supplementary Figure 18. *FOXG1*, *PAX6* and *PRKD1* transcription during forebrain neural progenitor cell (NPC) differentiation of heterozygous knockout (KO) models.

Supplementary Figure 19. *FOXG1* binding in the *PAX6* locus.

Supplementary Figure 20. *FOXG1* and *PRKD1* chromatin interactions in TAD DEL^{HET} and EC DEL^{HET} forebrain neural progenitor cells (NPCs).

Supplementary Figure 21. Flow Cytometry gating strategy.

SUPPLEMENTARY NOTES

Supplementary Note 1. Clinical description of del8 included in the manuscript

The proband is a 7-year-old East-African boy. Pregnancy was uneventful. He was born at term with caesarean section because of a narrow birth channel. At birth, weight was 3725g, length 54 cm and head circumference 34 cm. Apgar score was 9 after one and five minutes. Weight and length remained between p25-50, but head circumference evolved from -1.5sd to -4.5sd during the first year of life and followed the curve thereafter. He showed a severely delayed neuromotor development. He was breastfed during one year without swallowing problems, but underwent adenoidectomy at age 3 because of swallowing difficulties for solid foods with good results. At 18 months, he showed a first generalized tonic-clonic epileptic insult with temporal and occipital peak waves on electroencephalography. Valproate was started. Seizure control remained poor and levetiracetam and clonazepam was added over the years, but he still had periods of daily convulsions. At age 5, he did not show major epileptic insults anymore.

Clinical examinations at 18 months showed marked microcephaly. He made eye contact, showed some head control, but could not roll over, sit, or grasp. He developed contractures in the knees and showed increased tension in the upper legs. Peripheral deep tendon reflexes were normal and symmetrical. There was no Babinski sign. He was axially hypotonic, and continuously showed choreoathetosis. Inspection of trunk and limbs was normal. He made sounds, but no syllables. He showed a high anterior hair line, a broad nasal bridge, a short philtrum, a pit at the insertion of the left ear helix, and a small penis. Until last follow-up at age 7, his neurodevelopment stagnated and he did not acquire further neurological milestones.

Brain MRI (at the age of 18 months) showed simplification of the gyral pattern, but no cortical dysplasia or polymicrogyria. The diameters of the ventricles were at the upper normal limit. Myelinisation and differentiation of the white and grey matter was normal. Echocardiography, abdominal and renal ultrasound were all normal. Eye examination showed strabismus, and horizontal nystagmus. Slit lamp examination and fundoscopy were normal.

Supplementary Note 2. Clinical description of bca4 included in the manuscript

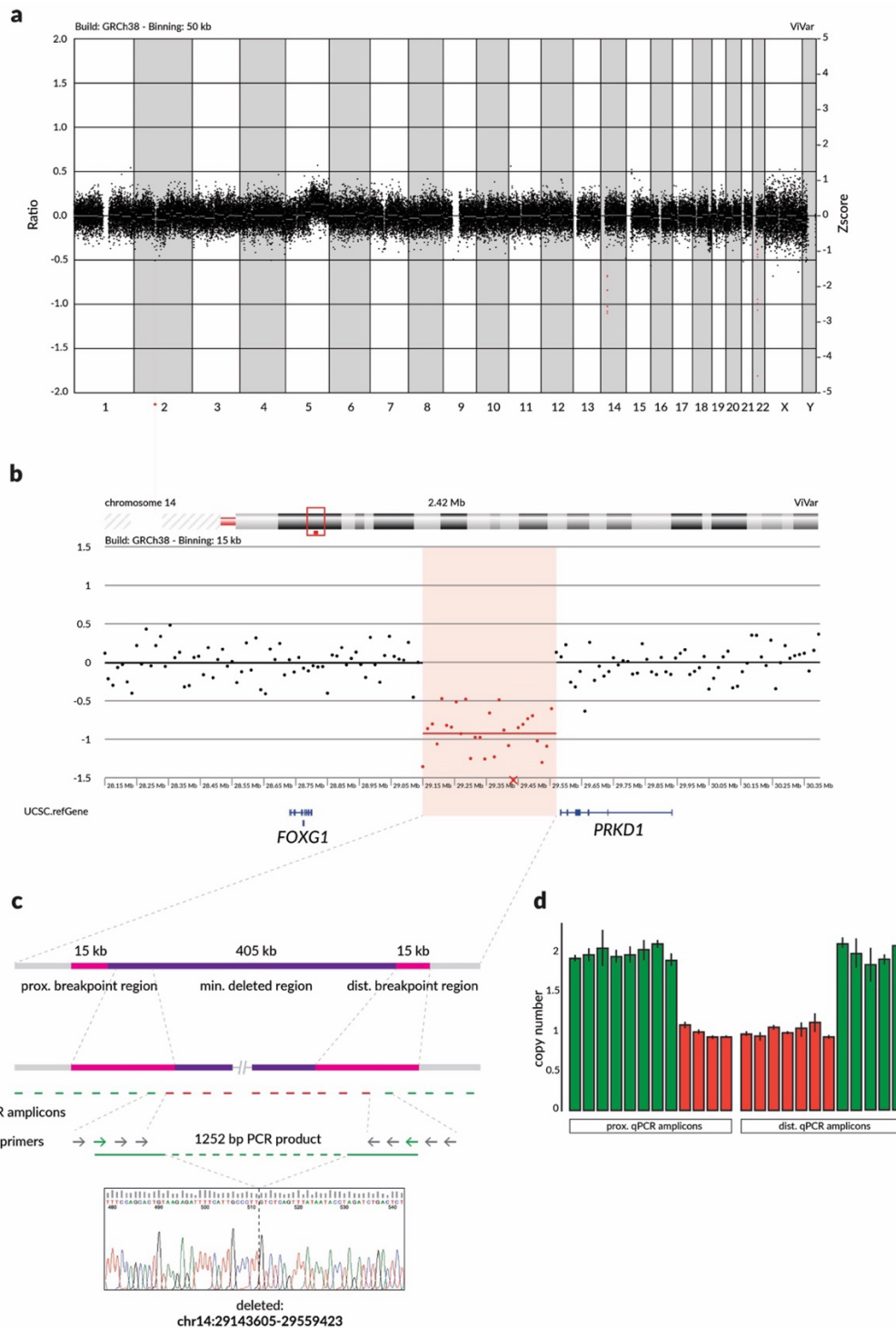
A girl, 2 years and 8 months old, born from a first pregnancy during which intrauterine hypotrophy was detected. Prenatal genetic testing was performed and excluded the presence of: trisomy 21, 18, 13, 9, 16, 22; sex chromosome aneuploidies; 84 types of microdeletions and duplications. She was born at term with caesarean section. At birth, weight was 2130 g, length 42 cm and head circumference 30 cm. With hypoglycemia and feeding difficulties in the neonatal period so she was fed through a naso-gastric tube.

Due to a severely delayed neuromotor development and microcephaly, the following studies were initially performed - karyotype analysis - 46,XX,t(8;14)(q22;q11. 2), female karyotype with reciprocal translocation between 8th and 14th chromosomes, which microscopically appears balanced, MLPA for microdeletions, subtelomeric deletions and duplications - normal result, comparative genomic hybridization (aCGH) did not detect pathological variants, metabolic screening of amino acids and acylcarnitines in blood and organic acids in urine was normal, with normal blood lactate and ammonia levels . Negative results also came from NGS panels including 1089 genes for epilepsy and 3199 genes associated with intellectual deficiency. Echocardiography, abdominal and renal ultrasound were all normal. Both parents have a normal karyotype.

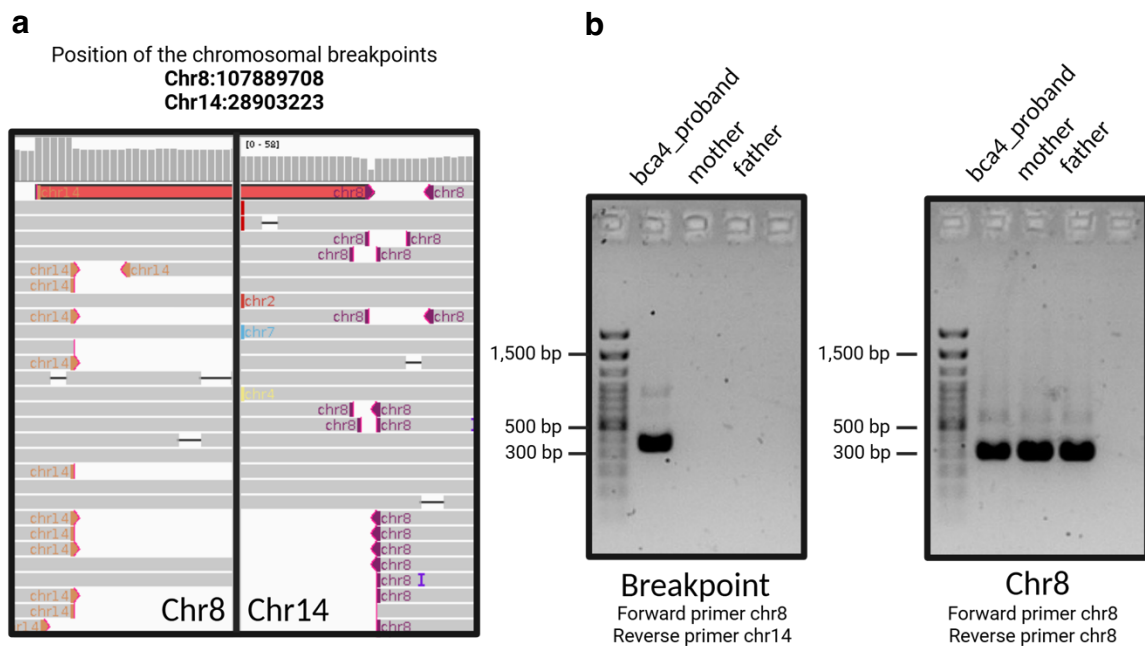
At 12 months, she showed a first generalized tonic-clonic epileptic seizure. Levetiracetam was started. Brain MRI (at the age of 18 months) showed partial hypoplasia of corpus callosum, non-obstructive hydrocephalus and impaired myelination.

Clinical examinations at 2 years and 8 months showed marked microcephaly (HC SDS= -5.32). Weight was also below the age norm (weight SDS= -3.09), height was normal for age (height SDS= -1.17). She showed a small nose with ball-shaped tip, deep philtrum, micrognathia, large and low-set ears. The muscle tone of the limbs was increased. She made eye contact, showed some head control, but could not roll over, sit, or grasp.

SUPPLEMENTARY FIGURES



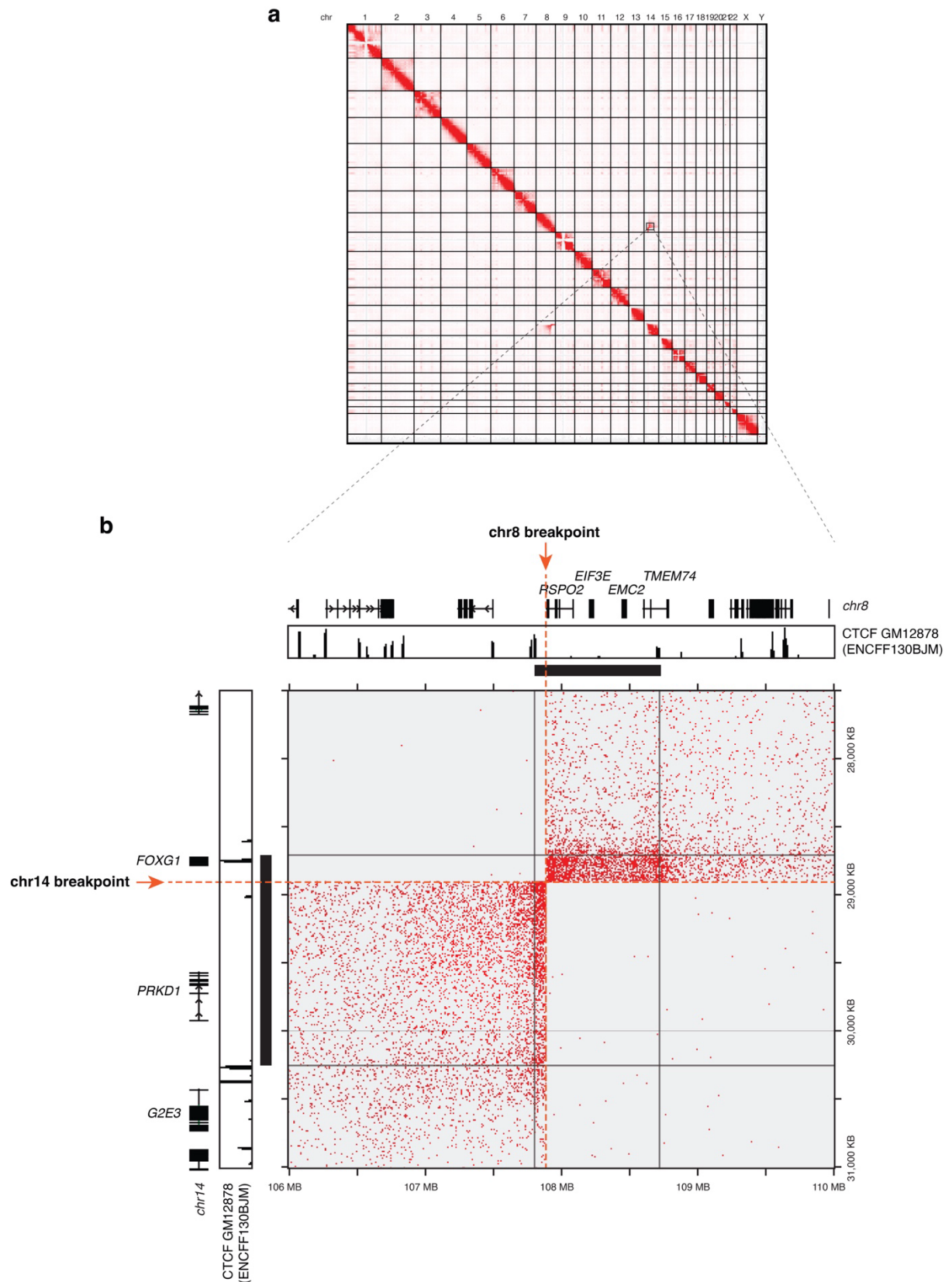
Sup Fig 1. Identification and finemapping of patient deletion (del8). (a) Genome-wide view of copy number profile visualized in ViVar and determined through shallow whole genome sequencing on patient lymphoblastoid cells (50 kb binning). (b) Copy number profile showing patient deletion (15 kb binning). (c) Strategy for finemapping of the deletion breakpoints, including design of CNV-qPCR amplicons and PCR primer across the breakpoint region, as well as the Sanger sequencing trace of the PCR product including the breakpoint. (d) Copy number plot for qPCR amplicons designed across the breakpoint regions used to narrow down the deletion breakpoints.



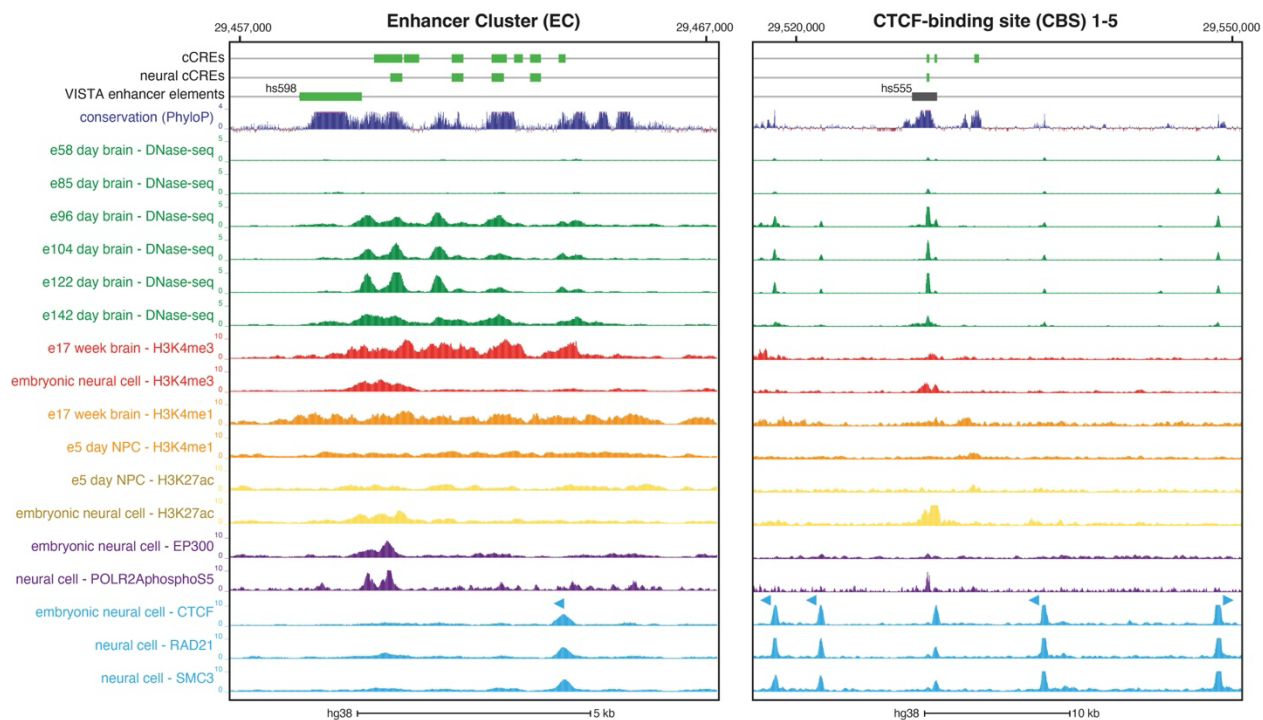
Sup Fig 2. Identification of translocation breakpoints in bca4 case. (a) IGV snapshot of Oxford Nanopore Technologies (ONT) long reads mapping across the chromosome 8 and chromosome 14 breakpoints. (b) PCR validation of the chromosome 8-14 translocation in the bca4 proband and parents.



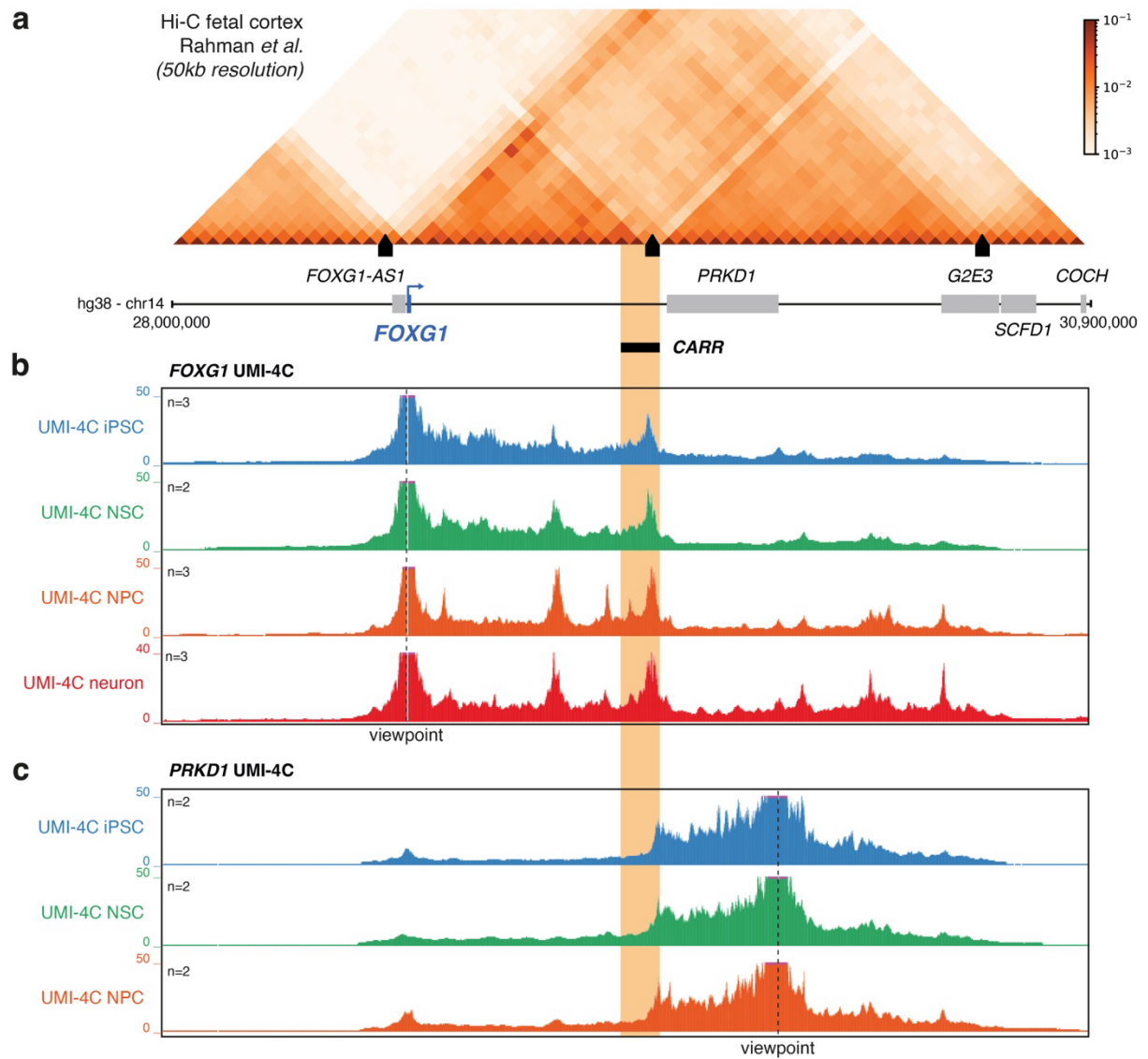
Sup Fig 3. Overview of all CNVs observed in patients overlapping the bigger FOXG1 TAD based on the DECIPHER database. Grey: Overlap FOXG1 gene; blue: duplication; red: deletion.



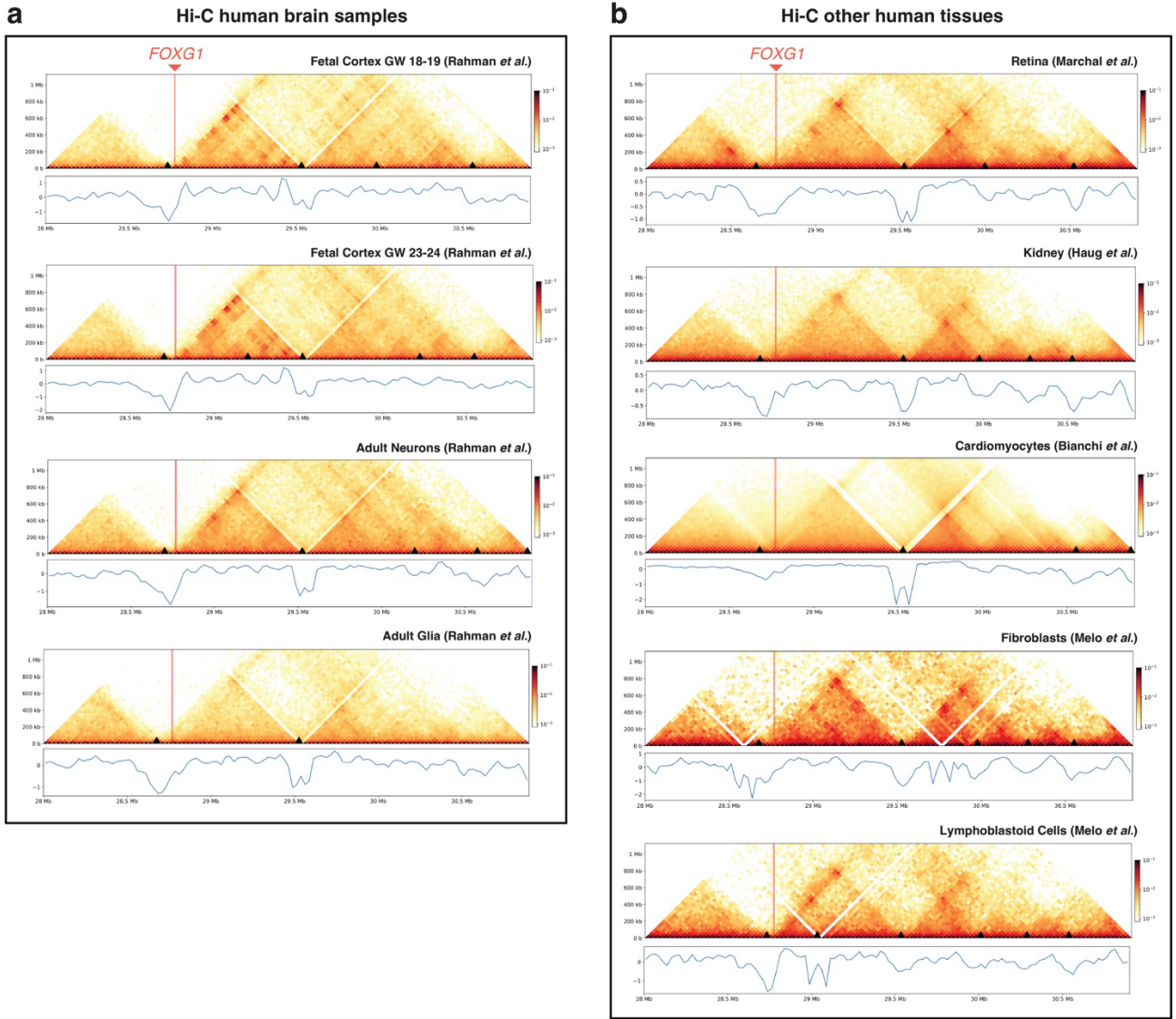
Sup Fig 4. 3D chromatin architecture in *bca4* case lymphoblastoid cells (LCLs). Interaction frequency matrix from Hi-C on *bca4* LCLs. (a) Genome-wide. (b) Zoomed in on translocation breakpoints. Black boxes indicate topologically associated domains (TADs) on chromosome 8 and 14 disrupted by translocation breakpoints.



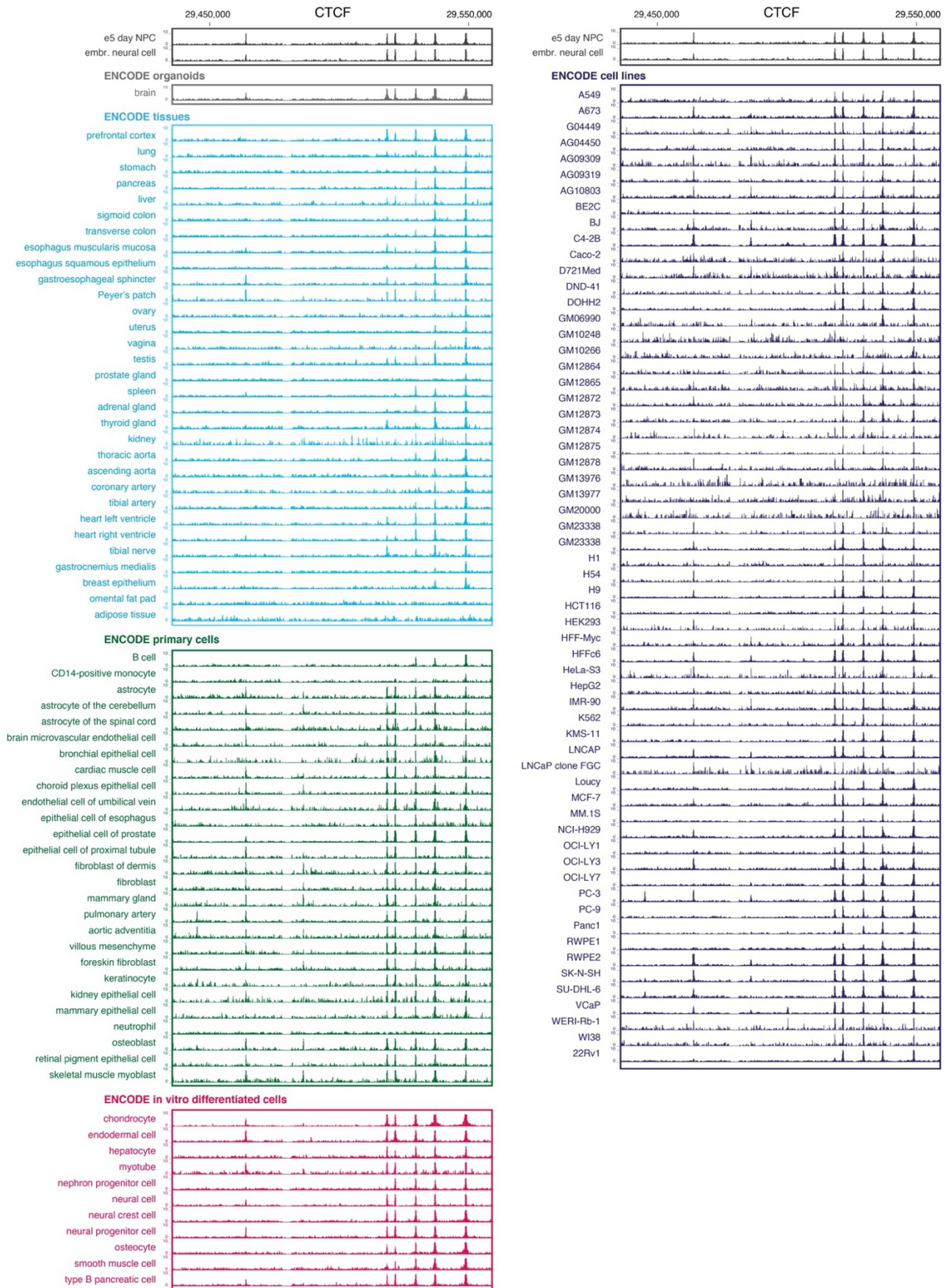
Sup Fig 5. Detailed view of the enhancer cluster and TAD boundary. Plot of candidate cis-regulatory elements (cCREs) and publicly available epigenomics data at the enhancer cluster (EC) and TAD boundary region, including: (neural) enhancer-like cCREs based on human DNase I hypersensitive sites (DHSs)¹ and ENCODE SCREEN² enhancer-like elements; validated VISTA enhancer elements (positive: green, negative: grey)³; PhyloP basewise conservation, accessible chromatin through DNase-seq in embryonic brain, ChIP-seq for regulatory markers H3K4me3, H3K4me1, H3K27ac, as well as active RNA polymerase II (POLR2AphosphoS5) and the P300 cofactor in embryonic brain (embryonic) neural cells and NPCs, ChIP-seq for the architectural protein CTCF and cohesin subunits RAD21 and SMC3 in (embryonic) neural cells.



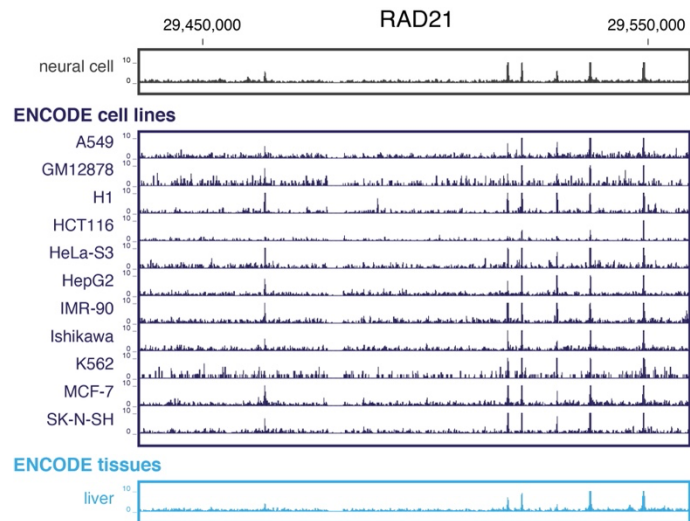
Sup Fig 6. Chromatin interaction frequencies at the *FOXG1* and *PRKD1* locus. (a) TAD structure of the *FOXG1* locus as shown by the Hi-C interaction frequency matrix (fetal cortex, 50kb resolution, Rahman *et al.*⁴). TAD boundaries are indicated by black boxes. The commonly affected regulatory region (CARR) is highlighted in orange. (b) UMI-4C interaction frequency profiles (normalized and smoothed) with the *FOXG1* promoter as viewpoint (dashed line) in induced pluripotent stem cells (iPSCs, n=3), neural stem cells (NSCs, n=2), neural progenitor cells (NPCs, n=3) and neurons (n=3). (c) UMI-4C interaction frequency profiles (normalized and smoothed) with the *PRKD1* promoter as viewpoint (dashed line) in induced pluripotent stem cells (iPSCs, n=2), neural stem cells (NSCs, n=2), neural progenitor cells (NPCs, n=2).



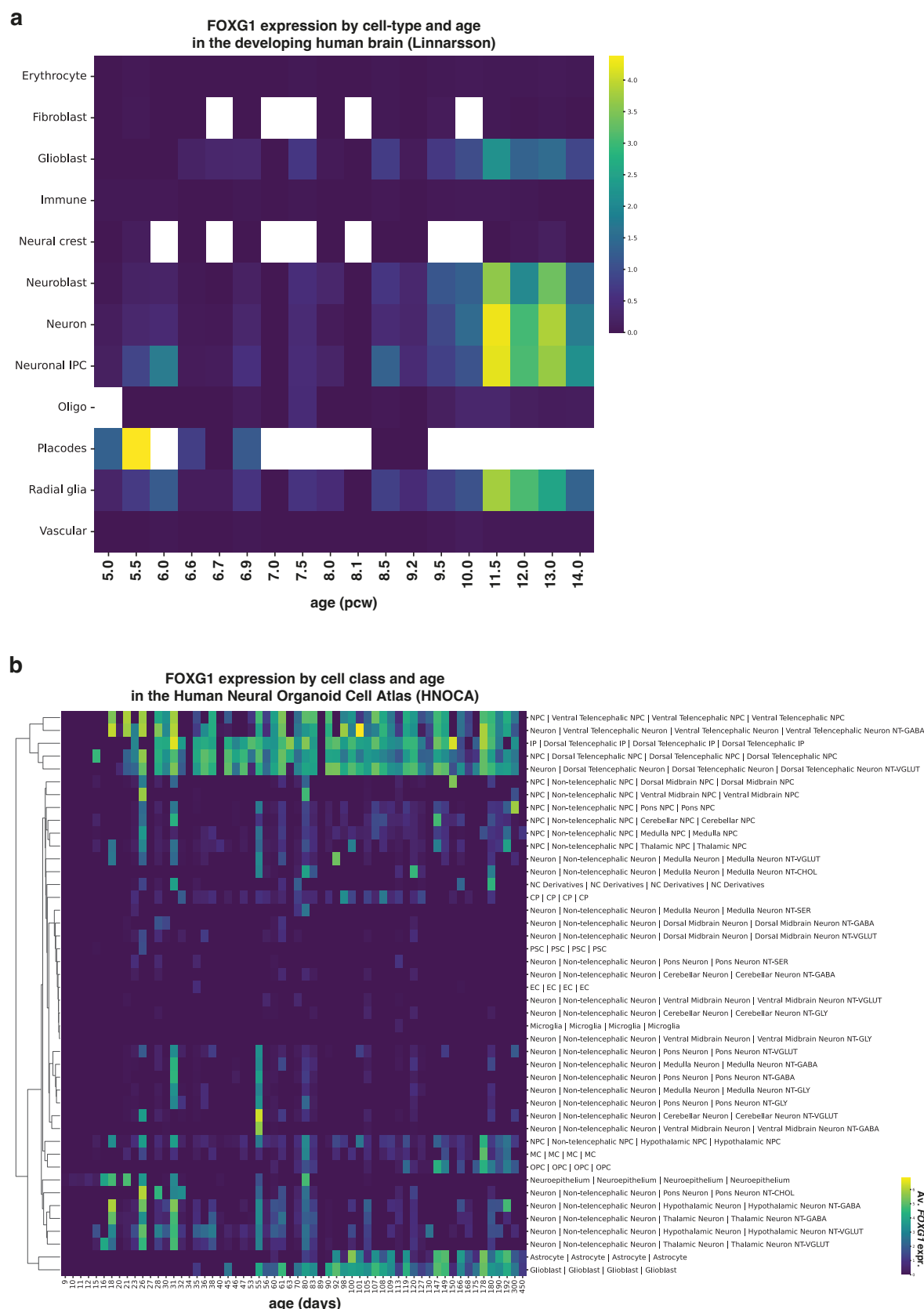
Sup Fig 7. TAD structure conservation at the *FOXG1* locus across tissues. Hi-C interaction frequency matrices and diamond insulation score plots (25 kb resolution) depicting *FOXG1* TAD structure in (a) human fetal and adult brain samples⁵ and (b) other human tissues⁶⁻⁹. TAD boundaries based on insulation score minima are indicated as black boxes.



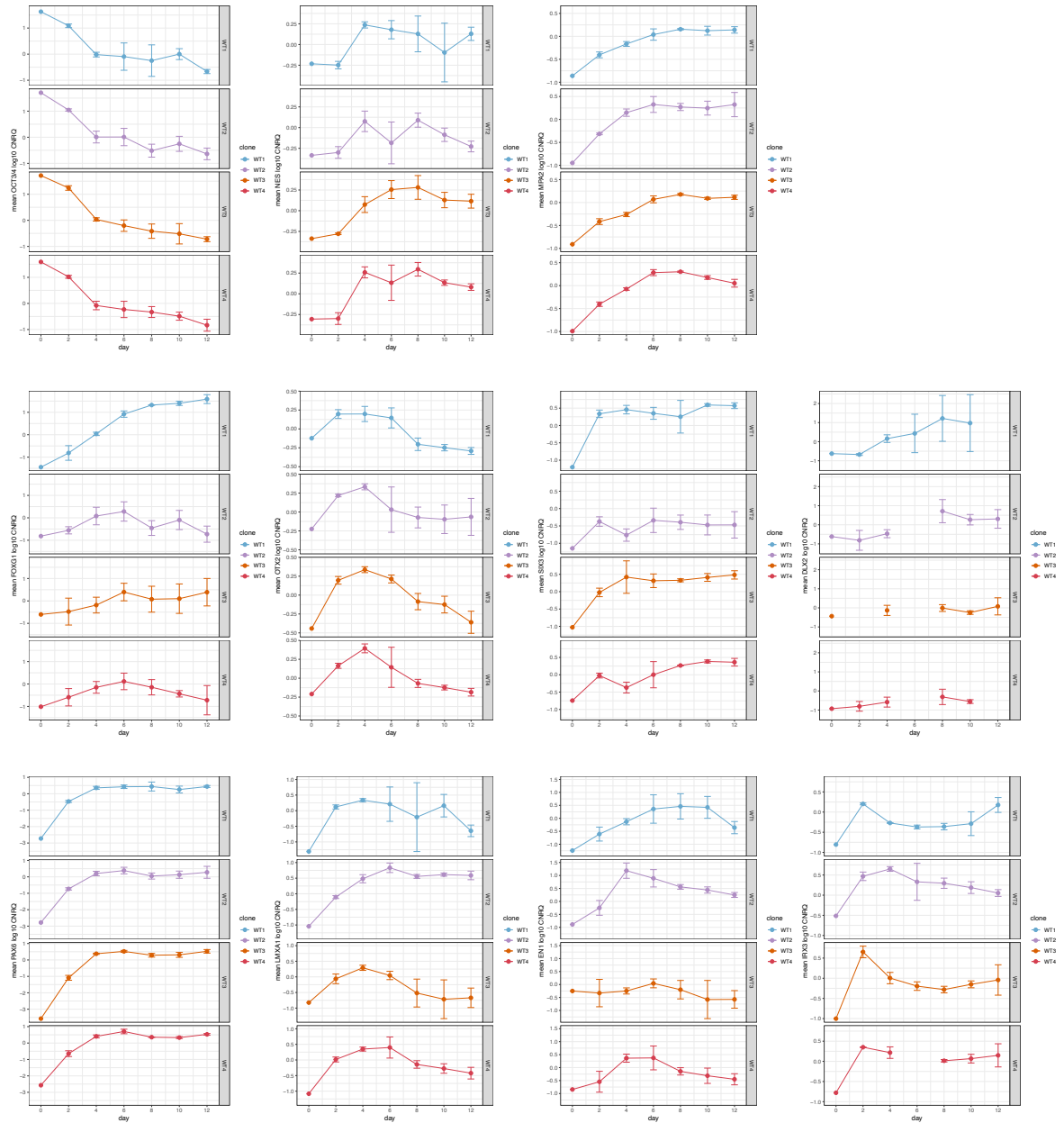
Sup Fig 8. CTCF binding at the commonly affected regulatory region (CARR) across ENCODE tissues and cell types.



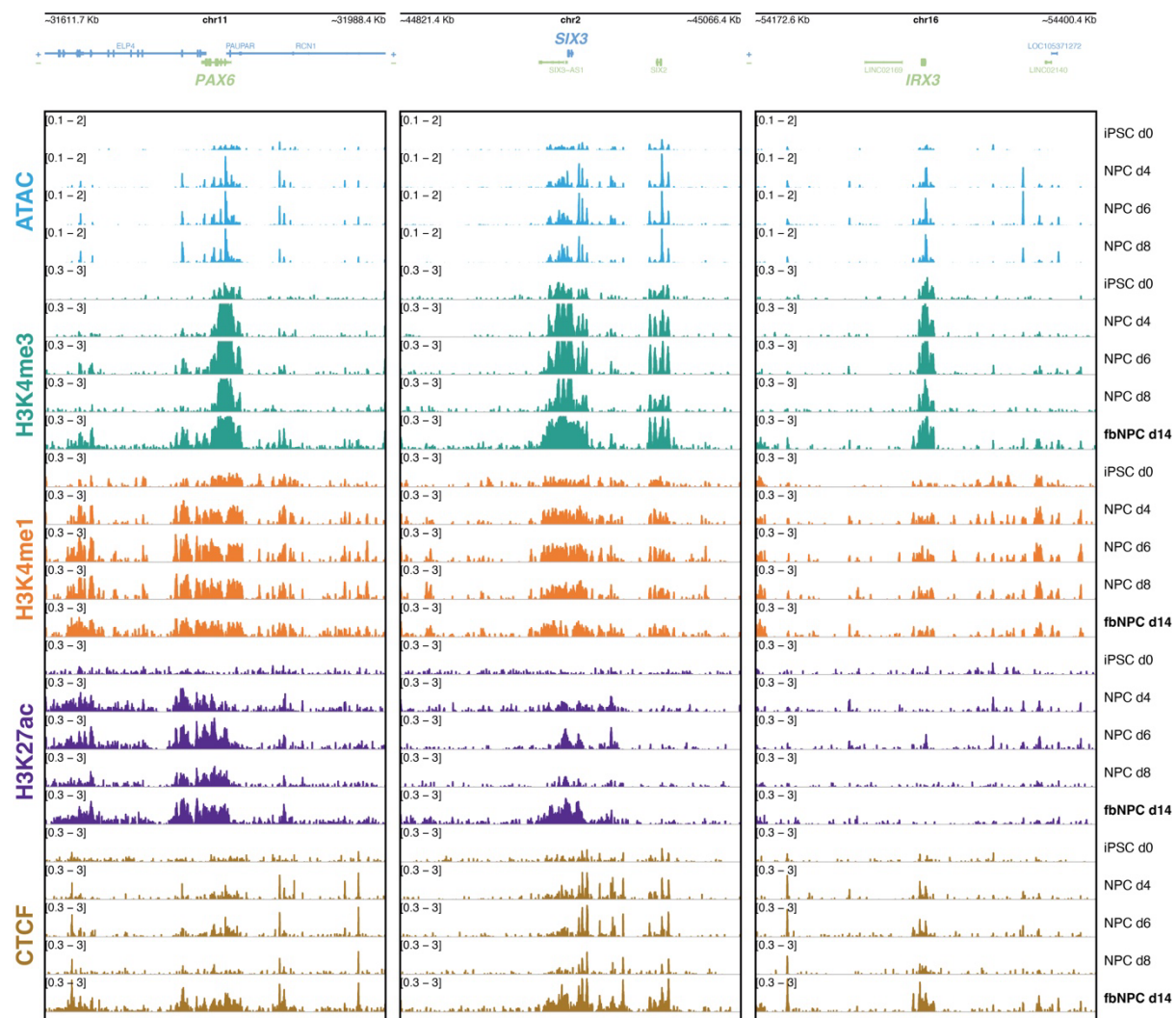
Sup Fig 9. RAD21 binding at the commonly affected regulatory region (CARR) across ENCODE tissues and cell types.

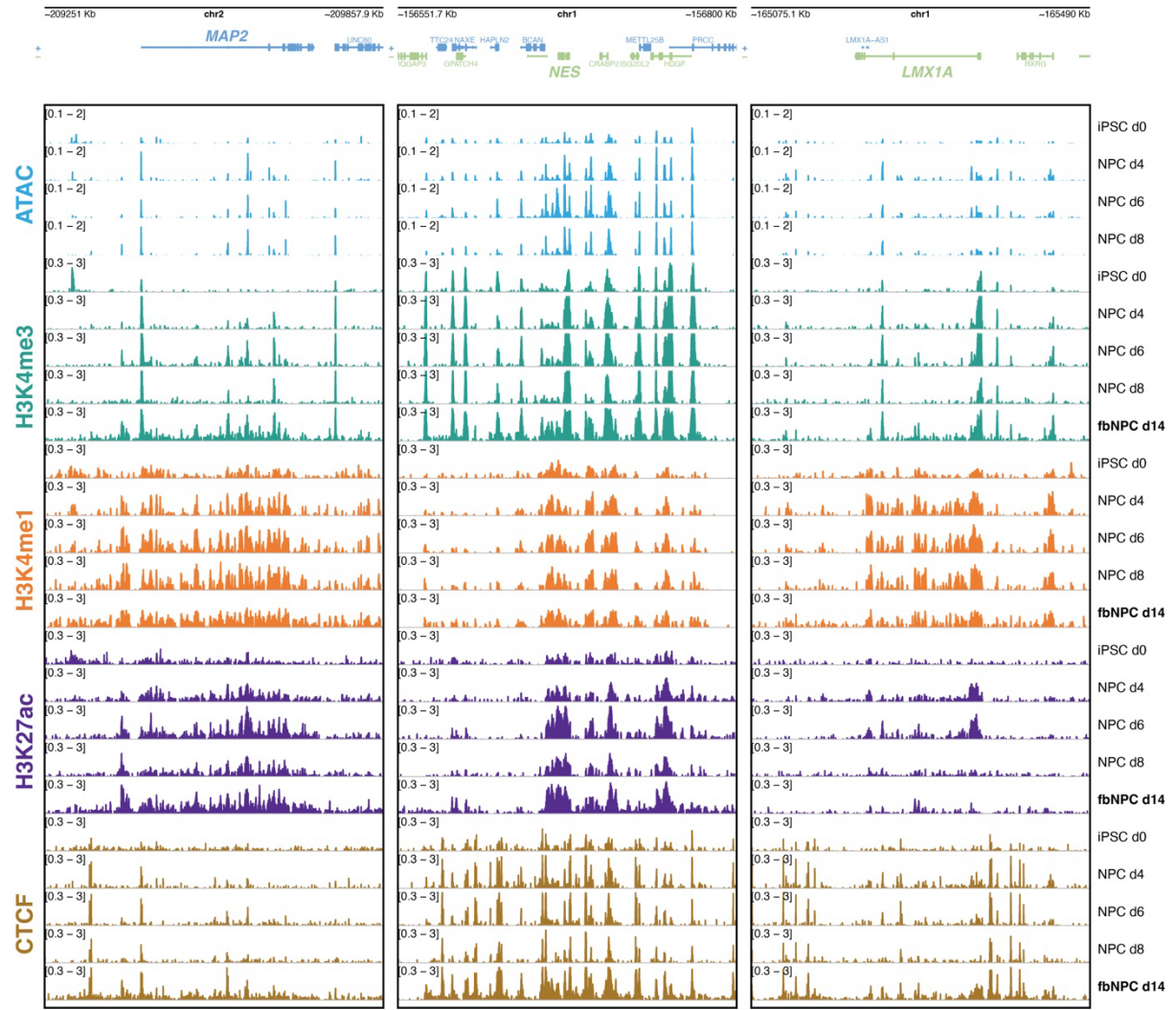


Sup Fig 10. Single-cell RNA-seq data from human brain development shows high FOXG1 expression in ventral telencephalic NPC and neurons. (a) FOXG1 mRNA expression per cell-type and age (in post-conceptional weeks, pcw) from single-cell RNA-seq data by Braun et al.¹⁰ (b) FOXG1 mRNA expression per cell class and differentiation stage from the Human Neural Organoid Cell Atlas (HNOCA)¹¹.

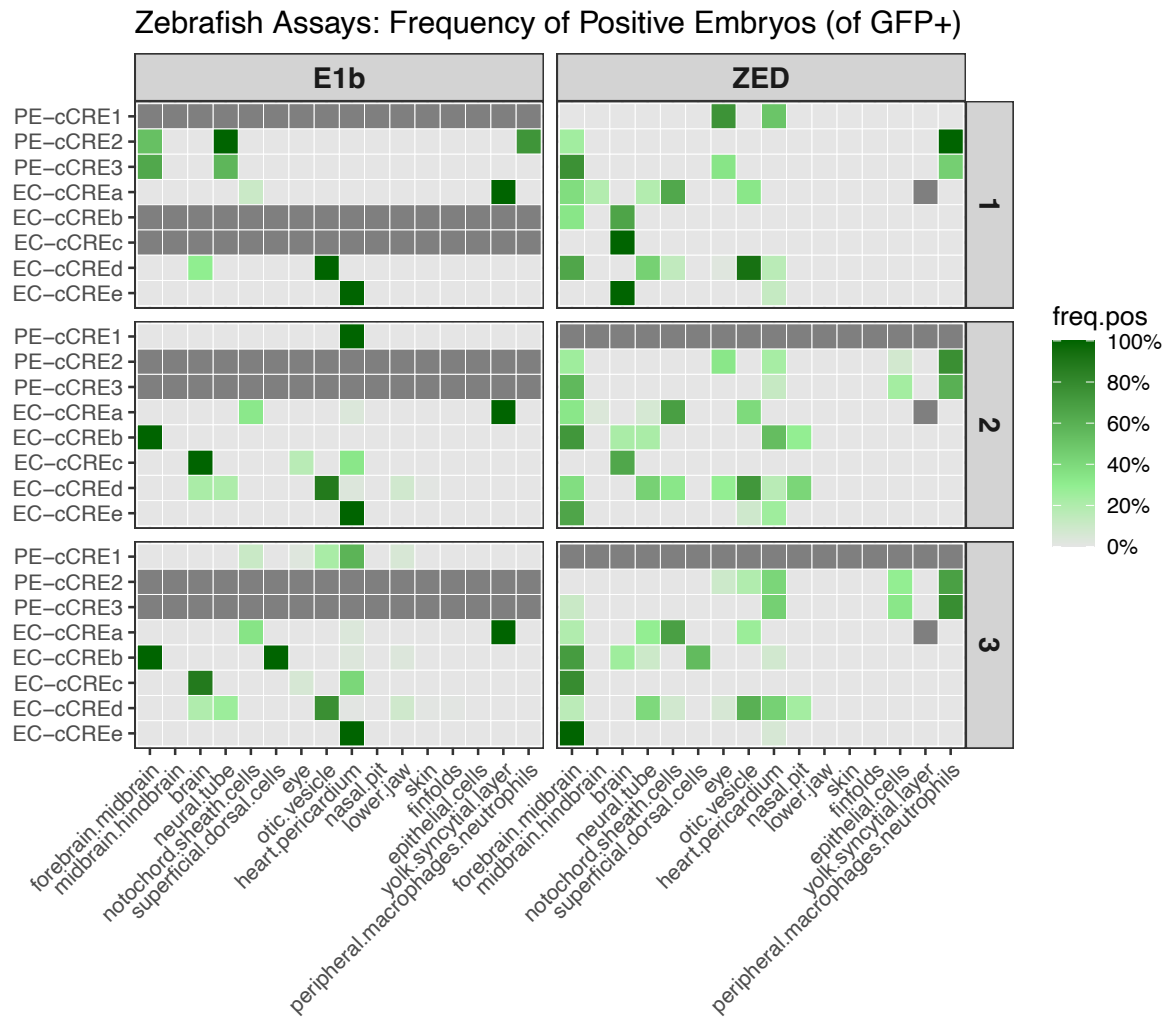


Sup Fig 11. Characterization of neural progenitor cell (NPC) populations using regional marker gene expression during general NPC induction. RNA expression of pluripotency markers (OCT3/4), neural markers (NES, MAP2) and various NPC markers for different brain regions: FOXG1, OTX2, SIX3 and DLX2 (forebrain); PAX6 (forebrain/midbrain); LMX1A (midbrain); EN1 and IRX3 (midbrain/hindbrain) during a 12-day, general NPC differentiation protocol (n=4). Data represents mean $\pm$ SD of three technical replicates.

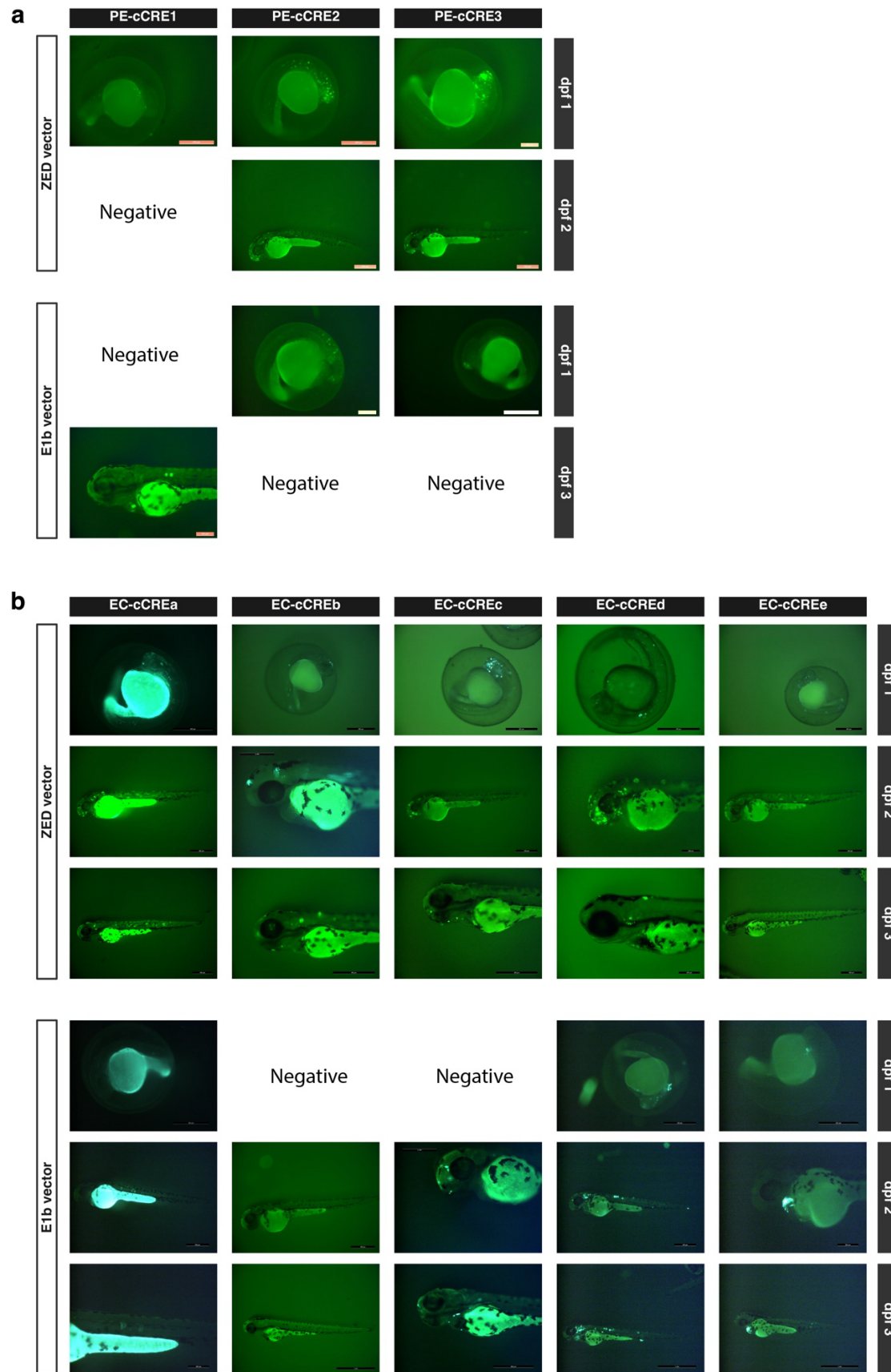




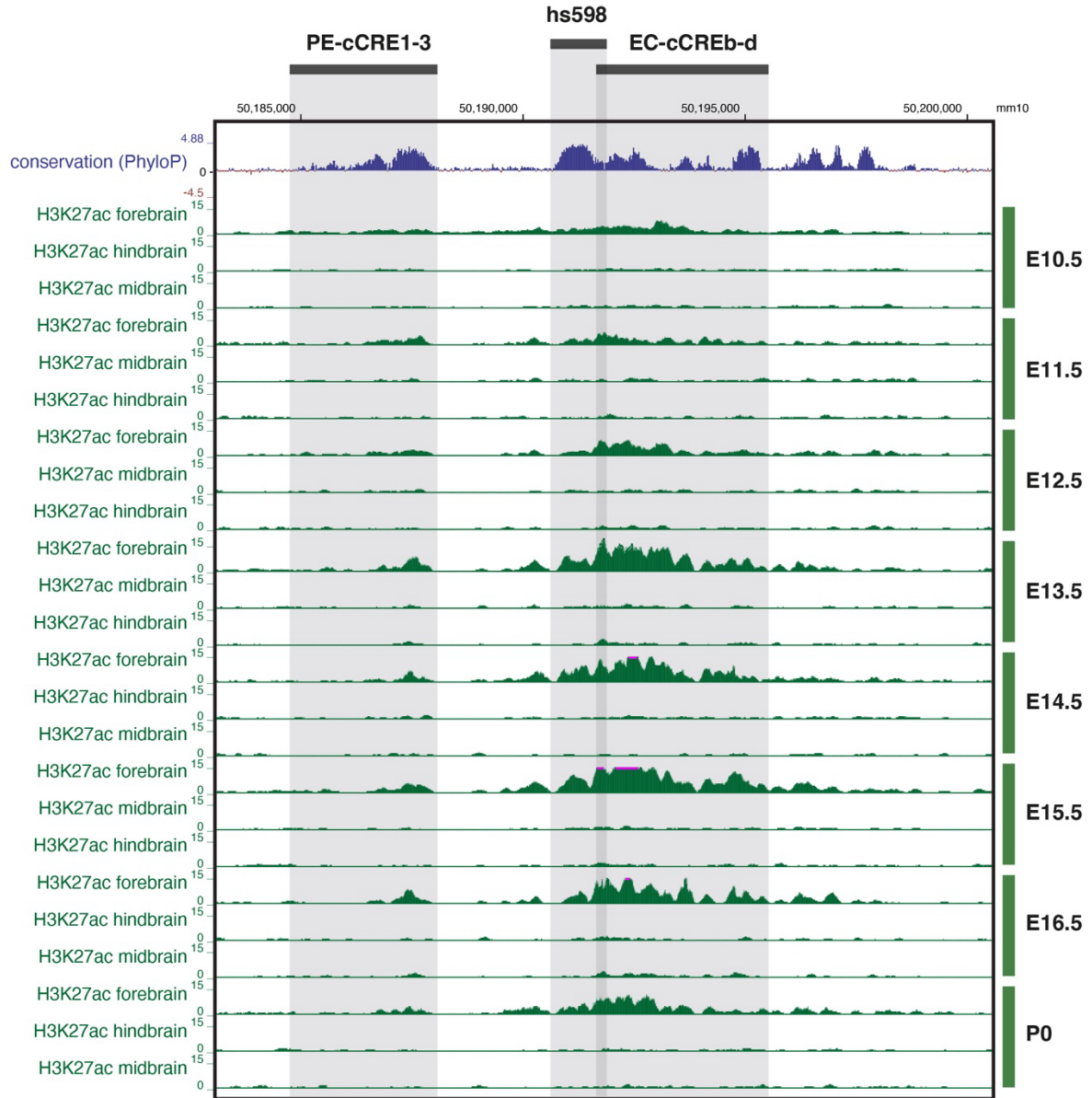
Sup Fig 12. Regulatory dynamics of neural progenitor cell (NPC) markers during general and forebrain-directed NPC differentiation. CUT&RUN signal tracks for H3K4me3 (active promotor marker), H3K4me1 and H3K27ac (active enhancer markers) and CTCF (TAD) for various NPC markers for different brain regions: PAX6 (forebrain/midbrain); SIX3 (forebrain); IRX3 (midbrain/hindbrain); LMX1A (midbrain); MAP2 (neural); NES (pan-neural marker) during an 8-day general NPC differentiation (n=2) and at the endpoint of a 14-day forebrain-directed NPC differentiation (n=3).



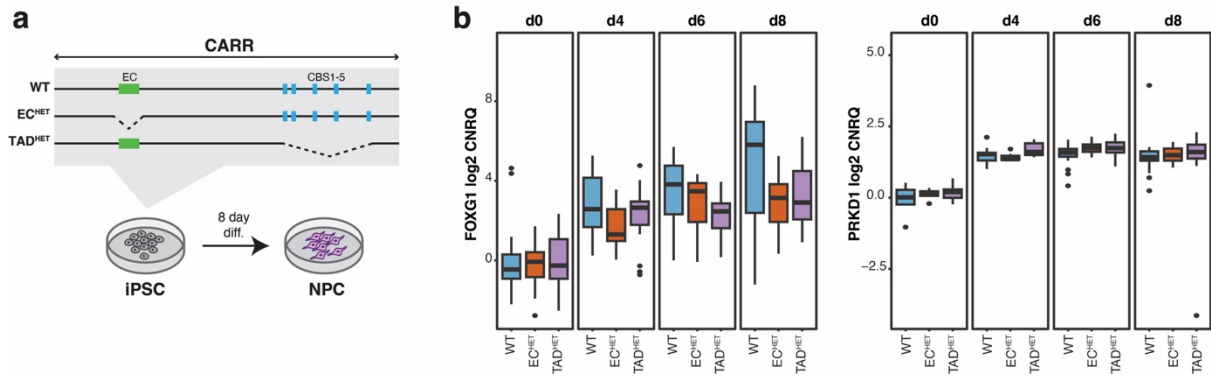
Sup Fig 13. Quantification of zebrafish enhancer assay results for candidate cis-regulatory elements (cCREs) within the Progenitor Element (PE) and Enhancer Cluster (EC) region. Frequency of transgenic F0 zebrafish embryos, obtained using both the E1b¹² and ZED¹³ vector, with tissue-specific GFP signal among total GFP positive embryos at 1, 2 and 3 days post-fertilization (dpf).



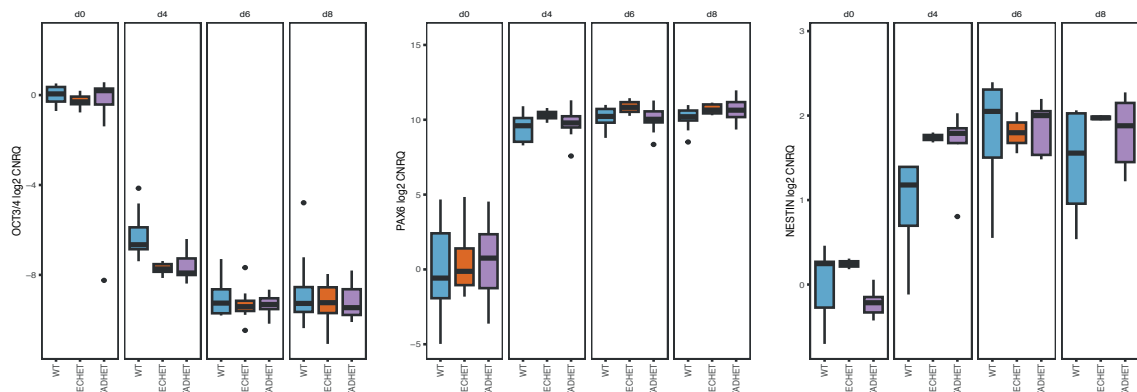
Sup Fig 14. Representative images of zebrafish enhancer assays. GFP reporter expression (using E1b¹² and ZED¹³ vector) in zebrafish embryos at 1, 2 or 3 days post-fertilization (dpf) for each candidate cis-regulatory element (cCRE) in the Progenitor Element (PE, a) and Enhancer Cluster (EC, b) region.



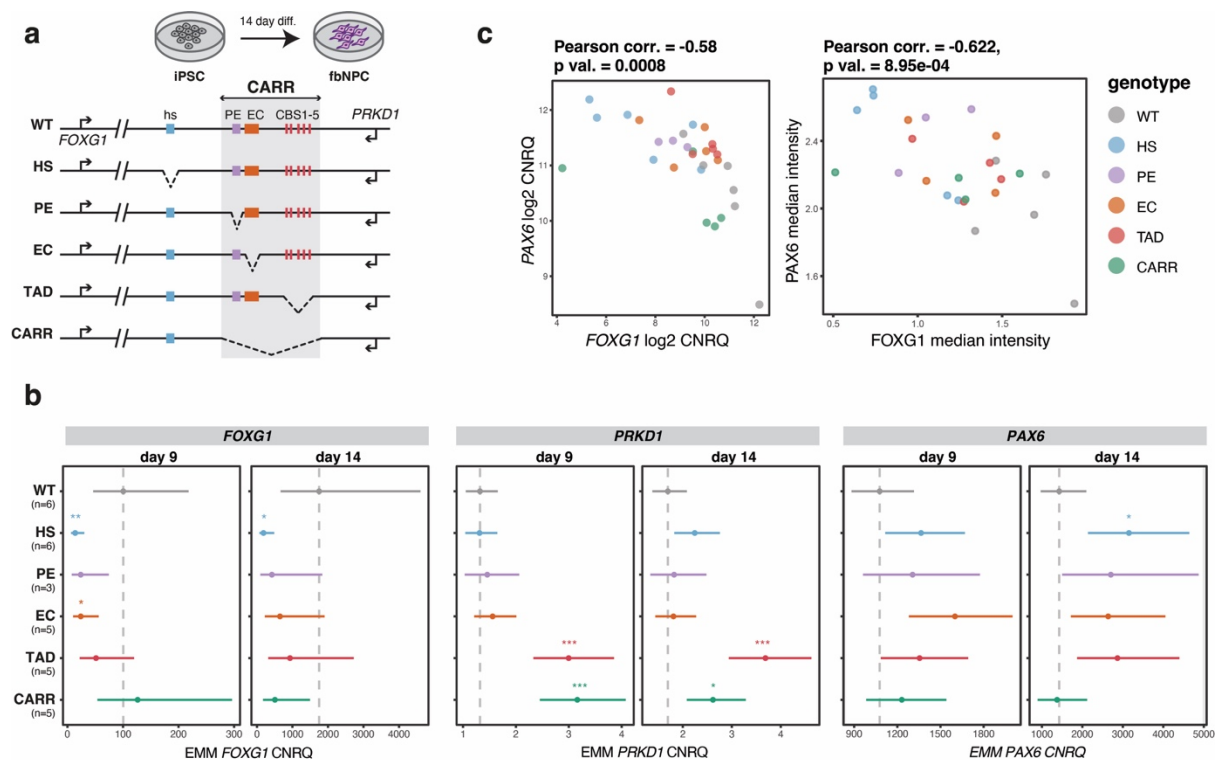
Sup Fig 15. Activation of candidate cis-regulatory elements (cCREs) during mouse embryonic development. ENCODE3² H3K27ac ChIP-seq data from mouse embryonic fore-, mid- and hindbrain, ranging from embryonic day E10.5 to birth (P0). cCREs tested in mouse enhancer assays are highlighted in grey.



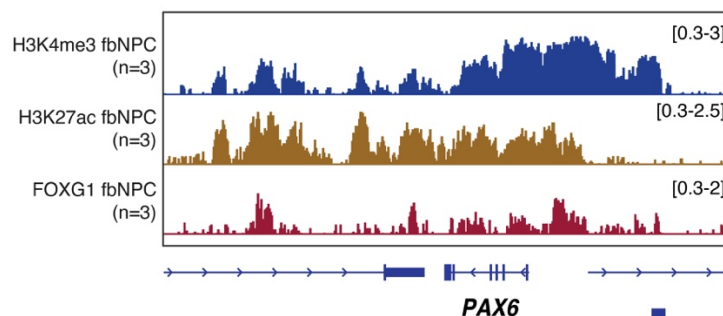
Sup Fig 16. *FOXG1* and *PRKD1* normalized RNA expression levels during general neural progenitor cell (NPC) induction of heterozygous knockout models. (a) Schematic overview of the generation of heterozygous knockout (KO) models for the enhancer cluster (EC) and TAD boundary in iPSCs, through CRISPR using dual ribonucleoprotein (RNP) complexes, which are further differentiated towards NPCs. (b) RNA expression of *FOXG1* and *PRKD1* during an 8-day NPC differentiation for 3 genotypes: wildtype (WT), heterozygous deletion of the enhancer cluster (ECHET) and heterozygous deletion of the TAD boundary (TADHET). Each genotype encompasses at least 3 biological replicates (WT n=6; ECHET n=3; TADHET n=6). RNA expression data was normalized to reference gene expression (UBC, SDHA) and scaled to WT iPSCs. Boxplots display the distribution of expression values per genotype, with the center line indicating the median, box limits representing the interquartile range (IQR; 25th–75th percentiles), whiskers extending to 1.5 × IQR, and points beyond the whiskers indicating outliers.



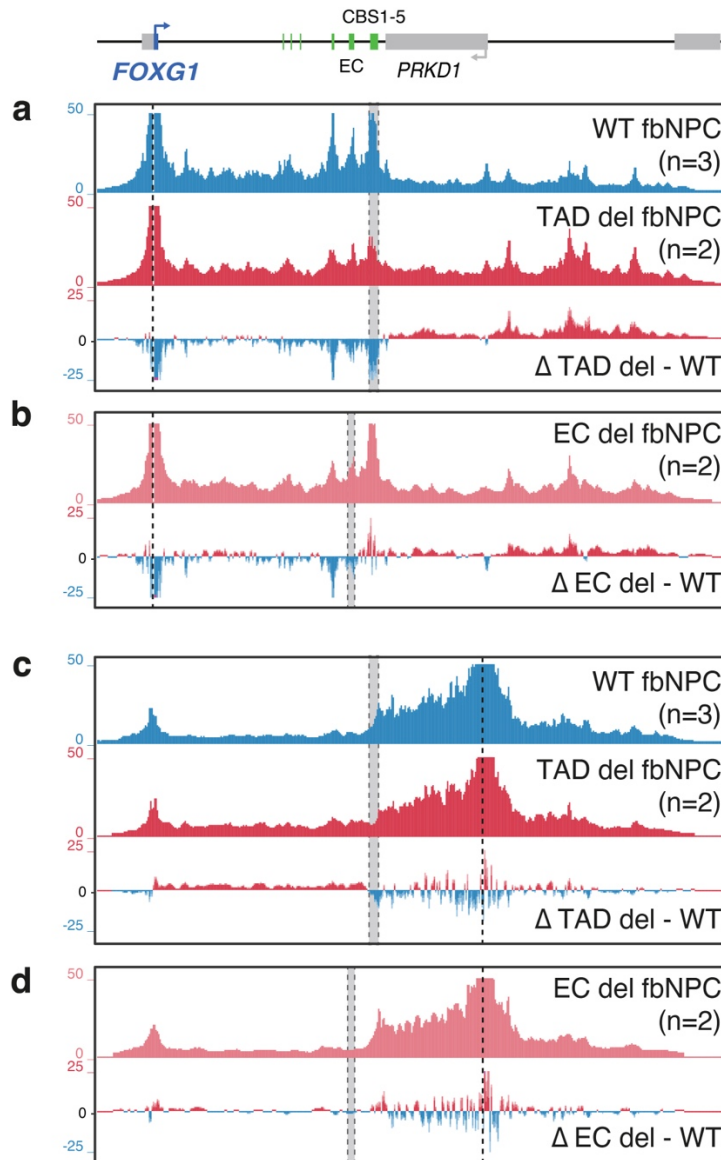
Sup Fig 17. Successful differentiation of WT and CRISPR KO models for the EC and TAD boundary towards NPCs. RNA expression of OCT3/4 (iPSC marker) and PAX6 (NPC marker) and NES (neural marker) during an 8-day NPC differentiation for 3 genotypes: wildtype (WT), heterozygous deletion of the enhancer cluster (ECHET) and heterozygous deletion of the TAD boundary (TADHET). Each genotype encompasses at least 3 biological replicates (WT n=6; ECHET n=3; TADHET n=6). RNA expression data is visualized in log2, scaled to iPSCs. Boxplots display the distribution of expression values per genotype, with the center line indicating the median, box limits representing the interquartile range (IQR; 25th–75th percentiles), whiskers extending to 1.5 × IQR, and points beyond the whiskers indicating outliers.



Sup Fig 18. *FOXG1*, *PAX6* and *PRKD1* transcription during forebrain neural progenitor cell (NPC) differentiation of heterozygous knockout (KO) models. (a) Schematic overview of heterozygous, iPSC-derived fbNPC KO models, including targeted deletions of the hs1523/342 (HS) elements, progenitor element (PE), enhancer cluster (EC), TAD boundary (TAD) and the entire commonly affected regulatory region (CARR). (b) Estimated Marginal Means (EMM) and 95% confidence intervals of *FOXG1*, *PAX6* and *PRKD1* RNA expression at day 9 and day 14 of fbNPC differentiation for all KO models. Three to six independent clones from each genotype were included. Linear mixed-effects model with two-sided Dunnett's adjusted p-values were calculated using the EMM contrasts and are indicated as follows: *** (p-value < 0.001), ** (p-value < 0.01), * (p-value < 0.05). d9_HS DEL^{HET}_FOXG1 (p-val = 0.0034), d9_EC DEL^{HET}_FOXG1 (p-val = 0.0438), d14_HS DEL^{HET}_FOXG1 (p-val = 0.0034), d9_TAD DEL^{HET}_PRKD1 (p-val = 0.0004), d9_CARR DEL^{HET}_PRKD1 (p-val = 0.0002), d14_TAD DEL^{HET}_PRKD1 (p-val = 0.0002), d9_CARR DEL^{HET}_PRKD1 (p-val = 0.0291), d14_HS DEL^{HET}_PAX6 (p-val = 0.0324). (c) Pearson correlation at fbNPC differentiation day 14 between *FOXG1* and *PAX6* normalized RNA expression levels and *FOXG1* and *PAX6* median intensities from Flow Cytometry across all genotypes.



Sup Fig 19. *FOXG1* binding in the *PAX6* locus. CUT&RUN data for the *FOXG1* transcription factor at the *PAX6* promoter.



Sup Fig 20. *FOXG1* and *PRKD1* chromatin interactions in TAD DEL^{HET} and EC DEL^{HET} forebrain neural progenitor cells (NPCs). UMI-4C interaction frequencies for the *FOXG1* (a-b) and *PRKD1* (c-d) promoters (viewpoints indicated by dotted line) at day 14 of forebrain NPC (fbNPC) differentiation in wildtype (WT) versus TAD DEL^{HET} (a, c) and EC DEL^{HET} (b, d) lines. The deleted regions are demarcated by grey boxes. Subtraction ('delta') profiles highlight differential interactions.

SUPPLEMENTARY REFERENCES

1. Meuleman, W. *et al.* Index and biological spectrum of human DNase I hypersensitive sites. *Nature* **584**, 244–251 (2020).
2. Moore, J. E. *et al.* Expanded encyclopaedias of DNA elements in the human and mouse genomes. *Nature* **583**, 699–710 (2020).
3. Visel, A., Minovitsky, S., Dubchak, I. & Pennacchio, L. A. VISTA Enhancer Browser--a database of tissue-specific human enhancers. *Nucleic Acids Res.* **35**, D88–92 (2007).
4. Rahman, S. *et al.* Lineage specific 3D genome structure in the adult human brain and neurodevelopmental changes in the chromatin interactome. *Nucleic Acids Res.* **51**, 11142–11161 (2023).
5. Rahman, S. *et al.* Lineage specific 3D genome structure in the adult human brain and neurodevelopmental changes in the chromatin interactome. *Nucleic Acids Res.* **51**, 11142–11161 (2023).
6. Marchal, C. *et al.* High-resolution genome topology of human retina uncovers super enhancer-promoter interactions at tissue-specific and multifactorial disease loci. *Nat. Commun.* **13**, 5827 (2022).
7. Bianchi, V. *et al.* Detailed Regulatory Interaction Map of the Human Heart Facilitates Gene Discovery for Cardiovascular Disease. 705715 Preprint at <https://doi.org/10.1101/705715> (2019).
8. Haug, S. *et al.* Multi-omic analysis of human kidney tissue identified medulla-specific gene expression patterns. *Kidney Int.* **105**, 293–311 (2024).
9. Melo, U. S. *et al.* Hi-C Identifies Complex Genomic Rearrangements and TAD-Shuffling in Developmental Diseases. *Am. J. Hum. Genet.* **106**, 872–884 (2020).
10. Braun, E. *et al.* Comprehensive cell atlas of the first-trimester developing human brain. *Science* **382**, eadf1226 (2023).
11. He, Z. *et al.* An integrated transcriptomic cell atlas of human neural organoids. *Nature* **635**, 690–698 (2024).
12. Birnbaum, R. Y. *et al.* Coding exons function as tissue-specific enhancers of nearby genes. *Genome Res.* **22**, 1059–1068 (2012).
13. Bessa, J. *et al.* Zebrafish enhancer detection (ZED) vector: A new tool to facilitate transgenesis and the functional analysis of cis-regulatory regions in zebrafish. *Dev. Dyn.* **238**, 2409–2417 (2009).