Title

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

Journal

Nature Communications, 17(1)

ISSN

2041-1723

Authors

Hamerlinck, Lisa

D'haene, Eva

Vaughan, Michael B

et al.

Publication Date

2026-06-03

DOI

10.1038/s41467-026-73770-1

nature communications

Article

https://doi.org/10.1038/s41467-026-73770-1

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

Received: 25 March 2025

Accepted: 14 May 2026

Published online: 03 June 2026

Check for updates

Lisa Hamerlinck[1,2,14], Eva D'haene[1,2,14], Michael B. Vaughan[1,2], Nore Van Loon[1,2,3], María del Rocío Pérez Baca[1,2], Sebastian Leimbacher[1,2], Michael Kosicki[4], Lara Colombo[1,2], Lukas Genbrugge[1,2], Lies Vantomme[1,2], Esperanza Daal[1,2], Luiza Lorena Pires Ramos[5,6], Daniela Mircheva Avdjieva-Tzavella[7], Himanshu Goel[8,9], Koen Devriendt[10], Albena Jordanova[5,6,11], Annelies Dheedene[1,2], Axel Visel[4,12,13], Björn Menten[1,2], Bert Callewaert[1,2] & Sarah Vergult[1,2]

The *FOXG1* transcription factor is a crucial regulator of embryonic brain development. Pathogenic *FOXG1* variants cause *FOXG1* syndrome. Although structural variants in the non-coding region downstream of *FOXG1* have been reported in 38 individuals with similar characteristics, the regulatory pathomechanisms remain unknown. Here, we identify two non-coding structural variants in individuals with *FOXG1* syndrome-like features, allowing us to delineate a ~124 kb commonly affected regulatory region. Using epigenomic profiling and in vivo enhancer assays, we characterize and validate regulatory elements within the commonly affected regulatory region and wider *FOXG1* TAD. We see strong activation of previously validated forebrain enhancers, and identify an enhancer cluster and progenitor-specific enhancer region that are strongly activated during forebrain-directed neural progenitor cell differentiation, a process in which *FOXG1* is an important regulator. Perturbation of these elements results in varying degrees of reduced *FOXG1* transcription in forebrain neural progenitor cells and in population shifts within these cells, while removal of the TAD boundary leads to aberrant expression of the neighbouring *PRKD1* gene. Our findings characterize enhancer and architectural elements essential for proper *FOXG1* transcription during neurodevelopment, therefore improving variant interpretation in this region.

The forkhead box G1 (*FOXG1*) transcription factor is highly and specifically expressed in the developing brain, where it acts as a transcriptional repressor essential to early brain development. It regulates neural progenitor cell (NPC) proliferation, triggers forebrain patterning, coordinates migration of maturing neurons, and organizes neuronal circuitry assembly[1]. *FOXG1* haploinsufficiency has been shown to impair GABAergic interneuron differentiation[2,3], while *FOXG1* overexpression has been associated with overproduction of inhibitory neurons in autism spectrum disorder[4]. In schizophrenia, *FOXG1* has also been identified as a risk gene by a regulatory, schizophrenia-associated, single-nucleotide polymorphism (SNP)[5].

Heterozygous aberrations leading to *FOXG1* loss-of-function, as well as microdeletions and duplications, cause a congenital form of Rett syndrome (OMIM #613454), also termed *FOXG1* syndrome to

A full list of affiliations appears at the end of the paper. e-mail: Sarah.Vergult@UGent.be

indicate notable differences with the *MECP2*-associated phenotype[1,6–8]. Characteristic clinical features include severe developmental delay and intellectual disability (ID), microcephaly, hypotonia, seizures, and agenesis or hypoplasia of the corpus callosum. In addition, there is a growing cohort of individuals described in literature with structural variants (SVs) disrupting the noncoding region downstream of the *FOXG1* gene with phenotypic characteristics that closely resemble those associated with *FOXG1* syndrome[9–24]. In this manuscript, we refer to this condition as *FOXG1* syndrome-like disorder. Although the putative regulatory effects have not been functionally characterized, it has been postulated that these copy-number variants (CNVs; deletions, duplications), as well as balanced chromosomal aberrations (BCAs; translocations, inversions) 3′ to *FOXG1* disrupt regulatory mechanisms governing *FOXG1* transcription.

These cases underline that *FOXG1* expression must be tightly regulated to support normal brain development. However, despite its established dosage sensitivity[3], little is known about the cis-regulatory elements controlling *FOXG1* transcription. The gene-poor region downstream of *FOXG1* contains several in vivo validated enhancer elements displaying activity in neural tissues[25], yet their effect on *FOXG1* expression has not been investigated. Additionally, while the 3D chromatin conformation of the locus, including its organization in topologically associating domains (TADs), has been hypothesized to be crucial for regulating *FOXG1* transcription[10,26] the impact of TAD disruption on *FOXG1* gene regulation remains unexplored.

In this study, we provide an updated overview of all previously reported individuals ($n = 38$)[10–24] with an SV overlapping the *FOXG1* TAD, and we describe two additional cases with *FOXG1* syndrome-like disorder, harboring, respectively, a noncoding microdeletion and a balanced translocation 3′ to *FOXG1*. These two additional cases enable us to narrow down the previously identified smallest region of overlap (~400 kb)[10] to a ~124 kb noncoding, commonly affected regulatory region (CARR), located ~665 kb downstream of *FOXG1*. To improve noncoding variant interpretation, we functionally characterize candidate cis-regulatory elements (cCREs) within this region and the wider *FOXG1* locus. Using chromatin interaction mapping and epigenomic assays, we elucidate cCRE and chromatin contact dynamics during in vitro (forebrain) NPC differentiation. Furthermore, we validate cCRE enhancer potential using in vivo enhancer assays in zebrafish and mice, and we assess the impact of enhancer and TAD boundary perturbation on *FOXG1* transcription during in vitro NPC differentiation.

Taken together, our findings improve the functional annotation and validation of regulatory elements at the *FOXG1* locus, which is crucial for (noncoding) variant interpretation.

Results

Noncoding structural variants in individuals with *FOXG1* syndrome-like disorder delineate the commonly affected regulatory region 3′ to *FOXG1*

To gather a comprehensive overview of cases with a *FOXG1*-related phenotype in which the *FOXG1* gene itself is intact, we mined literature and DECIPHER (Database of Genomic Variation and Phenotype in Humans using Ensembl Resources) for SVs within or overlapping the *FOXG1* TAD, not affecting the *FOXG1* coding sequence. (Fig. 1a and Supplementary Data 1). We found 20 individuals with CNVs (1 duplication, 19 deletions), ranging in size from 391 kb to almost 6 Mb, 17 cases with BCA breakpoints downstream of *FOXG1*, as well as one individual with a balanced rearrangement with an excision of 14q12q24 and an extra ring chromosome composed of the excised material[10–24]. Like the previously reported translocations, this aberration relocates a large portion of the noncoding region downstream of *FOXG1* in trans relative to the *FOXG1* promoter.

While most of these SVs occurred de novo (30/38), the duplication (dup1) was maternally inherited. For five deletions (del1, 2, 5, 7, and 16) and two BCAs (bca12 and 18), the inheritance pattern is unknown. Most of these previously reported individuals display hallmarks of *FOXG1* syndrome, with the most frequently reported symptoms including ID, developmental delay, microcephaly, hypotonia, seizures, stereotypic movements, and hypoplasia of the corpus callosum (Fig. 1b and Supplementary Data 1). The duplication (dup1) remains difficult to definitively link to a *FOXG1*-related disease mechanism, as it was maternally inherited, and the phenotype does not align with the other *FOXG1* syndrome-like cases. We also found three additional CNVs (one deletion and two duplications) overlapping the *FOXG1* TAD without affecting the *FOXG1* coding sequence. However, as we were unable to obtain any phenotypic information, they were excluded from the overview (ex.del1; ex.dup1&2) (Supplementary Data 1 and Supplementary Fig. 3).

Here, we further extended this cohort by including two additional cases, respectively, a de novo translocation (bca4) and a de novo deletion (del8) (Fig. 1a, Supplementary Figs. 1 and 2, Supplementary Data 1). Both individuals presented with clinical features consistent with a *FOXG1*-related disease mechanism, including severe neuromotor delay, epileptic seizures, microcephaly and hypotonia (see Supplementary Notes 1 and 2 for full clinical description). The bca4 translocation (46,XX,t(8;14)(q22;q11.2)dn), did not disrupt a protein-coding gene on either of the derivative chromosomes, further supporting a regulatory pathomechanism (Fig. 1a and Supplementary Fig. 2). Consistent with this model, del8 carried a de novo ~416 kb noncoding deletion downstream of *FOXG1* (finemapped to chr14:29,143,605-29,559,423 (hg38)) (Fig. 1a and Supplementary Fig. 1). The distal breakpoint in this individual, enabled refinement of a ~124 kb CARR situated at chr14:29,435,515-29,559,423 (hg38) (Fig. 1a), substantially narrowing down the previously reported smallest region of overlap (~400 kb)[10]. Of note, this individual represents the only reported case within our cohort with a CNV downstream of *FOXG1* that does not overlap the *PRKD1* gene (Fig. 1a and Supplementary Fig. 1), thereby excluding any involvement of direct *PRKD1* aberrations in the phenotype. Heterozygous de novo missense mutations in *PRKD1* have been identified in individuals with syndromic congenital heart defects (pulmonic stenosis, septal defects), with additional clinical features including developmental delay, ectodermal abnormalities, and limb defects (OMIM #617364)[27,28]. Intriguingly, although all 20 previously reported CNVs downstream of *FOXG1* also affect *PRKD1*, cardiac defects were only reported in four cases (del14, del18, del19, and dup1) (Fig. 1a, b, and Supplementary Data 1). The previously reported missense mutations in *PRKD1* were predicted to result in both gain and loss-of-function, although clinical characteristics associated with both types of mutations are similar[27,28]. Yet, the absence of cardiac defects in most cases with CNVs overlapping *PRKD1* suggests that *PRKD1* loss (due to CNVs) and *PRKD1* (in)activation (due to missense mutations) have different phenotypic consequences.

We employed POSTRE[29], an in silico prediction tool, to evaluate whether genes located within the deletions, disrupted by translocation breakpoints, or residing in TADs affected by the SV breakpoints could contribute to the pathogenic mechanisms underlying these SVs. For each candidate gene within the affected TADs, independent pathogenic prediction scores (PS) were derived using cell-type or tissue-specific genomic data, particularly neurodevelopmental datasets (pfcGw15 and pfcGw18), including gene expression profiles, enhancer, and TAD maps. In all cases, *FOXG1* was considered the most likely causal gene (causative gene PS > 0.8, Supplementary Data 2).

Noncoding deletion affects 3D chromatin interactions at the *FOXG1* locus

Noncoding SVs can have an impact on gene regulation, 3D chromatin structure, and ultimately gene expression. To investigate the regulatory consequence of such events at the *FOXG1* locus, we assessed 3D chromatin structure in lymphoblastoid cell lines (LCLs) derived from the two additional cases presented here (i.e., bca4 and del8). We

https://doi.org/10.1038/s41467-026-73770-1

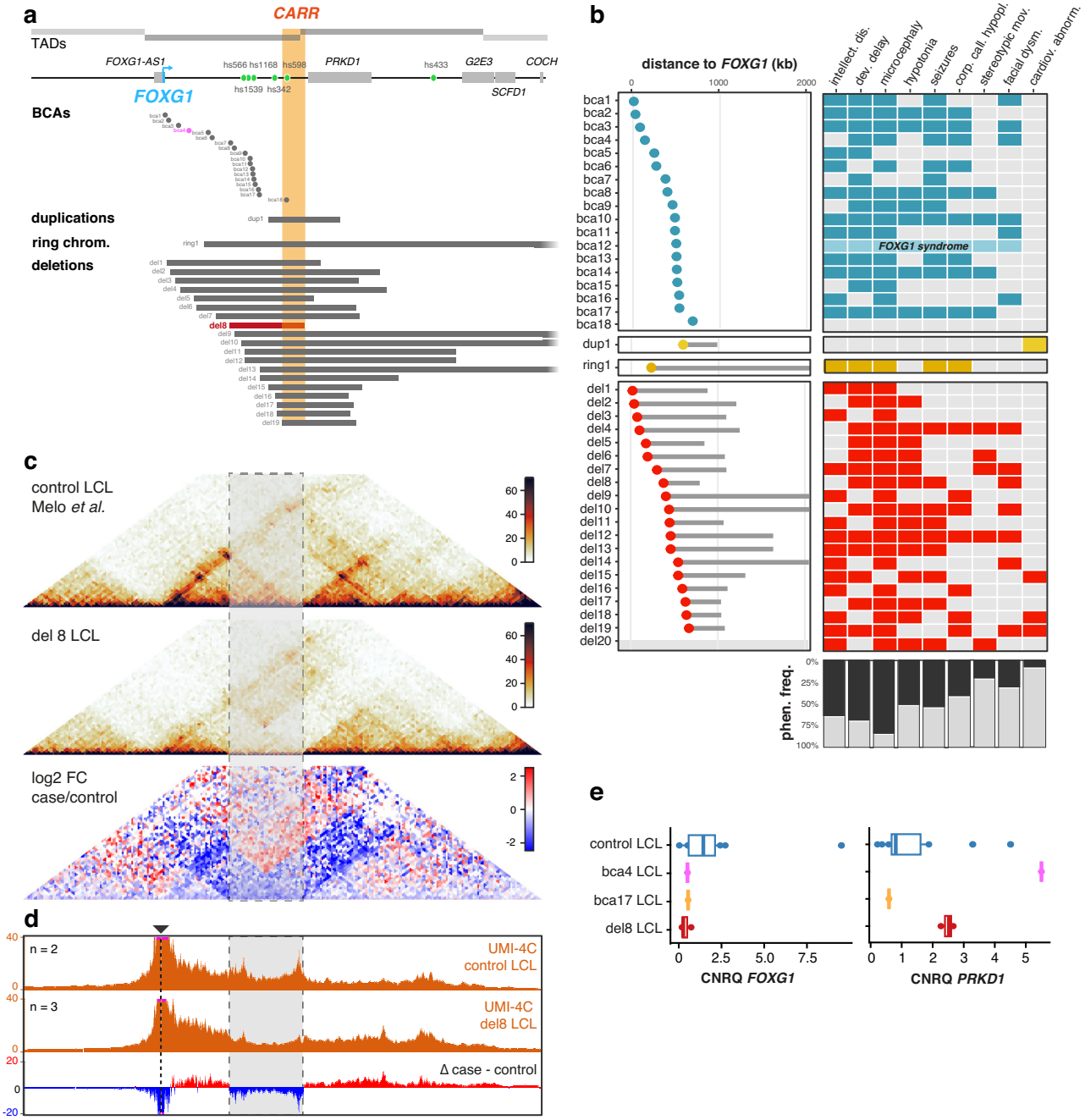


Fig. 1 | Structural variation disrupts the *FOXG1* regulatory region in individuals with *FOXG1* syndrome-like characteristics. a Schematic view of structural variants (SVs) located 3′ to *FOXG1*, obtained from literature, DECIPHER (gray) and the additionally identified translocation (bca4, pink) and deletion case (del8, red). The refined commonly affected regulatory region (CARR) is indicated in orange. **b** Schematic overview of each SV annotated with its distance to the *FOXG1* transcription start site, along with associated phenotypic outcomes. These include neurodevelopmental abnormalities (intellectual disability, developmental delay, microcephaly, hypotonia, seizures, corpus callosum hypoplasia, stereotypic movements, and facial dysmorphisms) primarily linked to aberrant *FOXG1* transcription and cardiovascular abnormalities possibly related to aberrant *PRKD1* transcription. Differential 3D chromatin interactions at the *FOXG1* locus between the del8 case (deletion indicated in gray) and control LCLs using Hi-C **c**, and UMI-4C **d**. **e** RNA expression levels of *FOXG1* (left) and *PRKD1* (right) in control LCLs ($n = 11$), LCLs of two translocation cases (bca4 and bca17; $n = 1$) and one deletion case (del8; $n = 3$). Box plots show the distribution of CNRQ values. The center line indicates the median, the box represents the interquartile range (IQR; 25th–75th percentile) encompassing the middle 50% of the data. Whiskers extend to the most extreme values within 1.5 × IQR from the box, and individual points represent measured values. Source data are provided as a Source data file.

performed UMI-4C[30] (del8) and in situ Hi-C[31] (bca4 and del8) on LCLs from these individuals and compared interaction profiles/matrices to those from a control LCL (previously generated by Melo et al.[13]). Both Hi-C matrices and UMI-4C profiles from the deletion case (del 8) LCLs indicated removal of the distal *FOXG1* TAD boundary and increased interactions with the adjacent TAD (Fig. 1c, d). Hi-C maps from the bca4 LCLs showed that the chromosome 14 translocation breakpoint disrupted the *FOXG1* TAD downstream of *FOXG1* and relocated it to the chromosome 8 TAD containing *RSPO2*, *EIF3E*, *EMC2* and *TMEM74* (Supplementary Fig. 4). As the residual part of the chromosome 8 TAD, now positioned within the *FOXG1* regulatory domain, contained no other genes, we hypothesized that the pathomechanism in this case

did not involve ectopic gene activation by *FOXG1* regulatory elements, but rather the uncoupling of the *FOXG1* gene from its regulatory domain.

In addition to these structural changes, mRNA expression analysis revealed a general decrease in *FOXG1* transcription in case LCLs (bca4, bca17, and del8) compared to eleven control LCLs (9 independent controls and the parental controls of bca4) (Fig. 1e). The effect on transcription of the neighboring *PRKD1* gene appeared to vary between the different structural variants (Fig. 1e). Increased *PRKD1* in the del8 LCL might be caused by increased ectopic interaction between *PRKD1* and *FOXG1* enhancer elements, while the opposite effect observed for the two bca LCLs could be influenced by the location of the reciprocal translocation breakpoint. It is important to note that *FOXG1* expression in LCLs is low, making reliable quantification challenging.

The commonly affected regulatory region harbors an enhancer cluster

As the delineated CARR downstream of the *FOXG1* gene is disrupted in all individuals with *FOXG1* syndrome-like characteristics included in this study, we hypothesized that this region may play a key role in regulating *FOXG1* transcription during neurodevelopment. To determine whether this region could indeed be involved in *FOXG1* regulation, we mapped chromatin interactions using UMI-4C[30] for the *FOXG1* promoter in induced pluripotent stem cells (iPSCs), neural stem cells (NSCs), NPCs, and neurons. UMI-4C profiles indicated frequent chromatin looping between the *FOXG1* promoter and several previously validated VISTA enhancers, as well as the CARR, in concordance with previously published Hi-C data from human fetal cortex (Fig. 2c)[32]. Interestingly, interaction frequencies varied between cell types, with many interactions becoming more prominent in NPCs and neurons compared to NSCs and iPSCs.

Next, we sought to assess the regulatory potential of the CARR. Despite the presence of multiple validated brain enhancer elements 3′ to *FOXG1*, the only validated enhancer within the CARR is element hs598 (Fig. 2b)[25]. Given that its activity pattern appears to be restricted to the neural tube, this enhancer element likely contributes only partially to establishing the full *FOXG1* expression pattern (Fig. 2d, e). Therefore, we compiled a list of candidate cis-regulatory elements (cCREs) present within the *FOXG1* TAD based on tissue-wide maps of human DNase I hypersensitive sites (DHSs) generated by Meuleman et al.[33]. Specifically, we cross-referenced neural DHS maps with ENCODE's SCREEN[34] registry of candidate regulatory elements to identify all neural DHSs with an enhancer-like epigenomic signature. This resulted in the identification of 99 candidate neural enhancers within the *FOXG1* TAD (Fig. 2b), of which eight were located within the CARR (Fig. 2f). Visualization of publicly available epigenomic datasets from ENCODE[34], including DNase-seq data from different stages of human embryonic brain development and ChIP-seq for several regulatory markers, indicated that five of the neural cCREs cluster together in a conserved region with strong regulatory potential, which we termed the enhancer cluster (EC) region (chr14:29458444-29465794, hg38; Fig. 2f and Supplementary Fig. 5). Notably, the EC region also included the previously validated hs598 enhancer, which contains an ultraconserved noncoding element (UCNE)[35]. However, it neither overlaps with the strongest epigenetic signals within the EC region, nor any of the neural cCREs, indicating that additional, untested sequences within the EC might have strong enhancer potential (Fig. 2f and Supplementary Fig. 5).

The commonly affected regulatory region plays a role in establishing TAD structure at the *FOXG1* locus

In addition to the presence of a putative EC, there are several indications that the CARR plays an important role in establishing the 3D chromatin structure at the *FOXG1* locus. Inspection of Hi-C data from human fetal cortex and UMI-4C interaction data for the *FOXG1* and *PRKD1* promoters suggested that *FOXG1* is located near the centromeric boundary of a megabase-scale TAD, containing two distinct subdomains (Fig. 2a and Supplementary Fig. 6). Insulation score analysis on the Hi-C matrix indeed indicated the presence of a TAD boundary within the CARR. Consulting published Hi-C datasets from different human tissues and cell types, including embryonic stem cells, NPCs, and prefrontal cortex, the TAD structure of the *FOXG1* locus appeared to be invariant (Fig. 2c and Supplementary Fig. 7).

In addition, ChIP-seq data for binding of the architectural protein CTCF, as well as cohesion subunits RAD21 and SMC3, indicated the presence of a cluster of five CTCF-binding sites (CBS1–5) within the CARR, constituting the *FOXG1* TAD boundary (Fig. 2f, Supplementary Figs. 8 and 9). CTCF motifs at three out of the five CBSs are convergently orientated to the centromeric TAD boundary proximal to *FOXG1*, indicating these CBSs are likely involved in establishing interactions within the *FOXG1* TAD through a loop extrusion mechanism. CTCF binding at these sites appears to be invariant across tissues and cell types, in concordance with the conserved 3D genome structure of the locus (Supplementary Fig. 8). Interestingly, the presence of enhancer marks and ultraconserved sequences at CBS3 might be indicative of an underlying enhancer element (Fig. 2f and Supplementary Fig. 5). This element (hs555) lacked enhancer activity in in vivo enhancer assays[25]. However, Won et al. found that a 500 bp sequence surrounding a schizophrenia-associated SNP (rs1191551) approximately 1.5 kb downstream of CBS3 showed enhancer activity in luciferase assays, and its deletion decreased *FOXG1* expression by 20–30% in HEK293 cells[5]. This suggests that enhancer potential at the CBS region cannot be fully excluded yet.

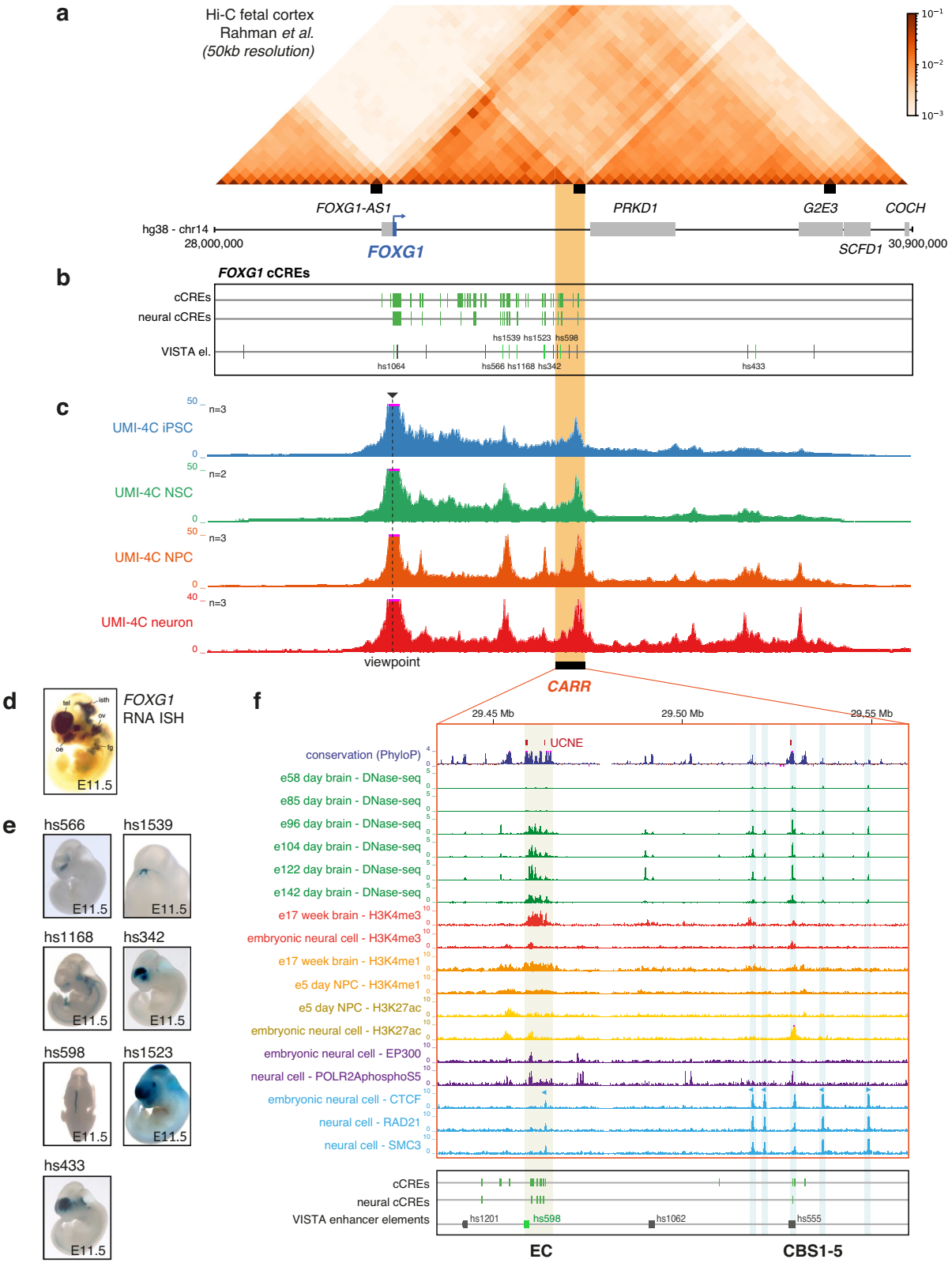
FOXG1 is dynamically regulated during the early phases of NPC differentiation

While analyses in LCLs revealed alterations in higher-order chromatin structure, *FOXG1* expression in these cells is low and therefore variable (even among controls), limiting their suitability for assessing transcriptional consequences. To investigate the functional impact of regulatory elements within the CARR on *FOXG1* expression, we established an in vitro model reflecting a biologically relevant developmental context. FOXG1 is a well-known marker of forebrain NPCs (fbNPCs)[36] and reanalysis of publicly available single-cell RNA-seq data from human brain development indeed mainly showed high *FOXG1* expression in ventral telencephalic NPCs and neuron[37] (Supplementary Fig. 10). Given the established role of *FOXG1* in NPC proliferation and differentiation, we sought to investigate the role of the EC and other regulatory elements in controlling *FOXG1* expression during NPC differentiation.

First, to evaluate the expression of *FOXG1* and other marker genes, we differentiated iPSCs into NPCs, using a general NPC induction protocol based on dual-SMAD inhibition[38]. Over a 12-day differentiation experiment, we noted variable induction of *FOXG1* expression between four wild-type iPSC lines. Although reproducible peak *FOXG1* expression was observed between days 6 and 8 of NPC differentiation, line-to-line variability increased after cell passaging on day 6 (Fig. 3a). To investigate whether this variability was attributable to a heterogeneous NPC population, we assessed the expression of various NPC markers for different brain regions: *FOXG1*, *DLX2*, *SIX3*, and *OTX2* (forebrain); *PAX6* (forebrain/midbrain); *LMX1A* (midbrain); *EN1* and *IRX3* (midbrain/hindbrain); and *HOXB4* (hindbrain/spinal cord) (Supplementary Fig. 11). Despite the lack of a clear correlation among most of these markers, variability in their expression levels suggests that differences in the NPC population may contribute to the observed variability in *FOXG1* expression.

To minimize the variability observed in *FOXG1* expression, we implemented a forebrain-directed differentiation protocol to generate a more homogeneous NPC population. This approach used dual-SMAD inhibition combined with Wnt/β-catenin inhibition to selectively

https://doi.org/10.1038/s41467-026-73770-1



promote fbNPC formation[39]. *FOXG1* was consistently induced during the 14-day fbNPC differentiation (Fig. 3b). Although this protocol more strongly induced *FOXG1* and reduced the spread of *FOXG1* induction between lines compared to the general NPC induction, some differences in *FOXG1* expression levels between wildtype lines remained. These residual differences likely reflect subtle, line-specific variation in the differentiation capacity of individual iPSC lines. Based on prior studies defining the timing of *FOXG1* activation in early forebrain NPCs, we selected day 9 (just prior to NPC passaging) and day 14 (differentiation endpoint) as key time points for downstream analyses.

Fig. 2 | *FOXG1* chromatin interactions and regulatory landscape. a TAD structure of the *FOXG1* locus as shown by the Hi-C interaction frequency matrix (fetal cortex, 50 kb resolution, Rahman et al.[32]). TAD boundaries are indicated by black boxes. **b** Candidate cis-regulatory elements (cCREs) within the *FOXG1* locus. Identified (neural) enhancer-like cCREs based on human DNase I hypersensitive sites (DHSs)[33] and ENCODE SCREEN[34] enhancer-like elements, as well as validated VISTA enhancer elements (positive: green, negative: gray)[25]. Commonly affected regulatory region (CARR) indicated in orange (chr14:29,435,515-29,559,423, hg38). **c** 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 ($n = 3$). **d** LacZ staining for Cre recombination in *Foxg1*-Cre mice illustrates *Foxg1* expression patterns in E11.5 mouse embryos (telencephalon (tel), mid-hind brain region (isthmus, isth), olfactory epithelium (oe), otic vesicle (ov), lens, retina, foregut (fg))[70]. **e** Activity patterns of validated enhancer elements from the VISTA enhancer browser in E11.5 mouse embryos (hs566: forebrain; hs1539: hindbrain; hs1168: cranial nerve, facial mesenchyme, hindbrain; hs342: forebrain; hs598: neural tube; hs1523: branchial arch, eye, facial mesenchyme, forebrain, midbrain, hindbrain; hs433: forebrain, hindbrain)[25]. **f** Closeup of CARR, showing PhyloP basewise conservation with ultra-conserved noncoding elements (UCNEs)[35], accessible chromatin through DNase-seq in embryonic brain, ChIP-seq for regulatory markers H3K4me3, H3K4me1, and H3K27ac, as well as active RNA polymerase II (POLR2A-phosphoS5) and the P300 cofactor in (embryonic) neural cells and NPCs, ChIP-seq for the architectural protein CTCF and cohesin subunits RAD21 and SMC3 in (embryonic) neural cells, (neural) cCREs, validated VISTA enhancer elements (positive: green, negative: gray)[25]. Highlights show regions of interest: enhancer cluster (EC) and CTCF-binding site (CBS) 1–5.

The enhancer cluster is activated during forebrain NPC differentiation and interacts with the *FOXG1* promoter

To understand the regulatory mechanisms underlying *FOXG1* induction during NPC differentiation, we characterized the epigenetic landscape of active enhancers and promoters in both NPCs and fbNPCs. CUT&RUN analysis of histone modifications associated with active promoters (H3K4me3, H3K27ac) and active enhancers (H3K4me1, H3K27ac) revealed a strong activation of the *FOXG1* promoter, VISTA enhancer elements, and the EC during both general NPC induction and directed fbNPC differentiation (Fig. 3c, d). Additionally, active promoter marks were observed at several NPC marker genes (Supplementary Fig. 12), indicating successful differentiation. During differentiation, activating epigenetic marks progressively accumulated at all elements (Fig. 3e). Yet, the strongest H3K27ac signal was consistently observed in fbNPCs, especially at VISTA forebrain enhancers (hs1523 and hs342) and at the EC. This is concordant with the much stronger induction of *FOXG1* transcription in fbNPCs.

In both NPC differentiations, we also observed dynamic enrichment of activating enhancer marks at an adjacent region, upstream of the EC (Fig. 3c, d). Notably, this region did not display enhancer-associated epigenomic marks in publicly available datasets from human embryonic brain tissue, suggesting that its activity may be specific to the developmental stage or cellular context represented by the in vitro NPC models. We therefore refer to this region as the progenitor element (PE) from here on (chr14:29451331-29455263, hg38). The identified PE showed the strongest H3K27ac signal enrichment in fbNPCs of all putative enhancer elements (Fig. 3e).

UMI-4C analysis further demonstrated increased chromatin interaction frequency between the *FOXG1* promoter and the EC, PE, and VISTA enhancer elements during NPC differentiation (Fig. 3c). The highest *FOXG1* contact frequency was again observed in fbNPCs for the forebrain VISTA enhancers (hs1523 and hs342), the PE, and the EC. Their strong epigenetic activation and frequent interaction with the *FOXG1* promoter suggest that these elements might control *FOXG1* transcription during fbNPC development.

Candidate enhancers drive reporter activity in neural tissues during embryonic development

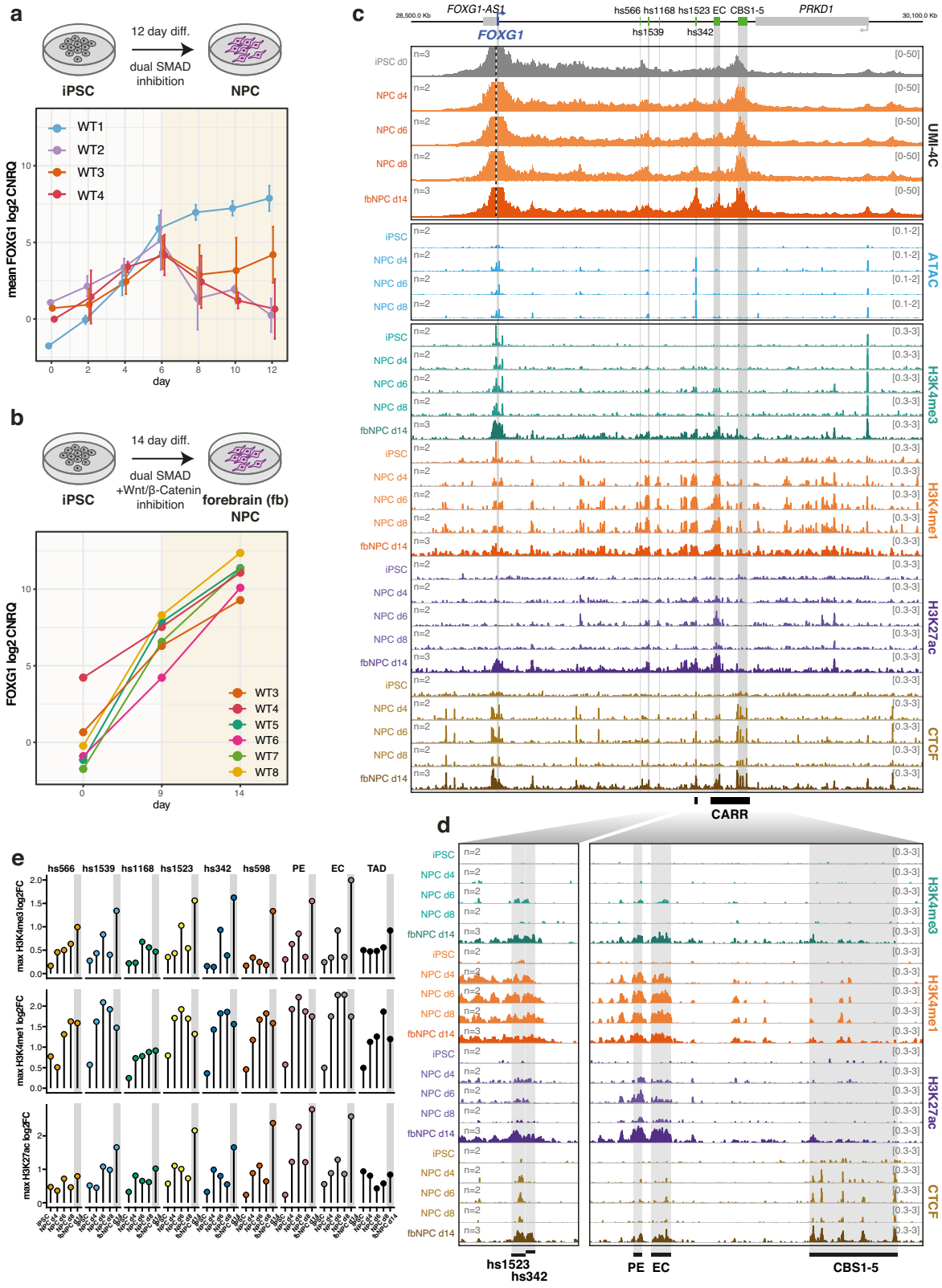
To functionally assess the regulatory potential of the EC and the identified progenitor element (PE), we performed in vivo enhancer assays in both zebrafish and mouse embryos, two complementary vertebrate systems that together capture the developmental window in which *Foxg1* becomes active. For the zebrafish assays, we delineated three candidate cis-regulatory elements (cCREs) within the PE (cCRE1–cCRE3) and five cCREs within the EC (cCREa–cCREe), based on conserved sequence stretches and epigenetic signatures in human brain samples (Fig. 4a and Supplementary Data 3). These sequences were subsequently cloned and tested through two independent in vivo zebrafish enhancer assays (i.e., E1b vector[40] and ZED vector[41]), and GFP expression was monitored at 24, 48, and 72 h post-fertilization (hpf). Seven out of the eight tested cCREs displayed (fore)brain activity in at least one of the assays (Fig. 4b, e, Supplementary Figs. 13, 14, and Supplementary Data 4). Only PE-cCRE1 failed to drive reporter gene expression in the developing brain. Both PE-cCRE2 and PE-cCRE3 showed forebrain activity during the early stages of development (24 and 48 hpf). All EC-cCREs displayed consistent (fore)brain activity across all stages, especially in the ZED vector assays. In addition, GFP expression was also observed in the neural tube (PE-cCRE2,3 and EC-cCREa,b,d), otic vesicle (EC-cCREa,d), olfactory pit (EC-cCREc,d), and heart (PE-cCRE1 and EC-cCREd,e). In sum, seven out of eight cCREs exhibited enhancer activity consistent with previously described Foxg1 expression patterns[36,42–46] and could therefore regulate tissue-specific *FOXG1* expression.

To evaluate enhancer activity in a mammalian context, we performed mouse transgenic enhancer assays using constructs spanning the full PE region and most of the EC (Supplementary Data 3). For the EC, we excluded the previously validated VISTA element hs598/EC-cCREa, as its neural tube activity had already been previously established (Fig. 2), and instead focused on the remaining cCREs (EC-cCREb-d). For the PE region (accessible in the VISTA enhancer browser as hs3168), nine transgenic embryos were obtained. Six embryos showed reporter expression, of which four carried a tandem and two a single construct. Robust forebrain LacZ expression was mainly observed in the tandem configuration, whereas single constructs showed weaker patterns (Fig. 4c, d). For the EC-cCREb-d construct (accessible in the VISTA enhancer browser as hs3169), two transgenic embryos were recovered, both of which showed reporter activity (one single and one tandem construct). LacZ staining was detected in the ear, abdominal region, and neural tube, with the single construct producing a stronger signal than the tandem configuration. Despite the consistent detection of a forebrain signal for the EC-cCREs in zebrafish embryos, no forebrain staining was observed for the EC-cCREb-d construct in mouse E11.5 embryos. However, epigenomic data from mouse embryonic development show that forebrain-specific H3K27ac is only markedly increased at the EC region from E13.5 onwards (Supplementary Fig. 15).

Deletion of functional elements in the *FOXG1* regulatory region results in decreased *FOXG1* expression during forebrain NPC differentiation

Next, we aimed to assess the impact of perturbing regulatory elements within the *FOXG1* locus, including the previously identified VISTA forebrain enhancers hs1523/hs342 (abbreviated: HS) as well as the identified functional elements within the CARR, i.e., the PE, EC, and TAD boundary, on *FOXG1* induction in the NPC models characterized above. We generated in vitro heterozygous knockout (KO) models targeting the HS, PE and EC elements, the TAD boundary and the entire CARR through CRISPR-Cas9 in iPSCs, using paired guide RNAs complementary to both sides of the target regions (Fig. 5a). Upon single-

https://doi.org/10.1038/s41467-026-73770-1



cell isolation and deletion screening, we selected edited clones and matched controls (i.e., wildtype clones that underwent the same experimental procedure) (Supplementary Data 5).

In an initial experiment, we differentiated EC and TAD boundary KO lines over eight days in triplicate using the general NPC induction protocol (control: 6 clones, EC DEL^{HET}: 3 clones, TAD DEL^{HET}: 6 clones). Quantitative RT-qPCR analysis at day 0, 4, 6, and 8 of differentiation showed the induction of *FOXG1* and other NPC markers (*PAX6*) and the loss of pluripotency markers (*OCT3/4*) for all genotypes, confirming successful differentiation of all lines (Supplementary Figs. 16 and 17). Despite the induction of *FOXG1* RNA expression across all genotypes, there was a notable decrease in *FOXG1* expression at day 8 in heterozygous EC and TAD boundary deletion lines compared to wildtype controls (Supplementary Fig. 16). Still, extreme variability in *FOXG1*

https://doi.org/10.1038/s41467-026-73770-1

Fig. 3 | *FOXG1* regulatory dynamics during neural progenitor cell (NPC) differentiation. a RNA expression of *FOXG1* during a 12-day general NPC induction for four wildtype (WT) iPSC lines. Data represent the mean ± SD of three technical replicates for four biological WT lines. Source data are provided as a Source data file. **b** RNA expression of *FOXG1* during a 14-day directed forebrain NPC (fbNPC) differentiation for six WT iPSC lines. Source data are provided as a Source data file. **c** Epigenomics data within the *FOXG1* locus in iPSCs; at day 4, day 6, and day 8 of general NPC induction; and at day 14 of fbNPC differentiation. Data includes *FOXG1* promoter interaction frequencies using UMI-4C; CUT&RUN (normalized log2 fold change signal/input) for H3K4me3 (active promoter mark), H3K4me1 (active enhancer mark), H3K27ac (active element mark), and CTCF (enriched at TAD boundaries). Validated VISTA enhancer elements, the progenitor element (PE), the enhancer cluster (EC), and TAD boundary (CBS1–5) are highlighted. **d** Close-up of epigenomics data at VISTA forebrain enhancers hs1523 and hs342, and at the commonly affected regulatory region (CARR). **e** Maximal (max) signal observed during (fb)NPC induction for H3K4me3, H3K4me1, and H3K27ac at validated VISTA enhancer elements, the PE, the EC, and the TAD boundary.

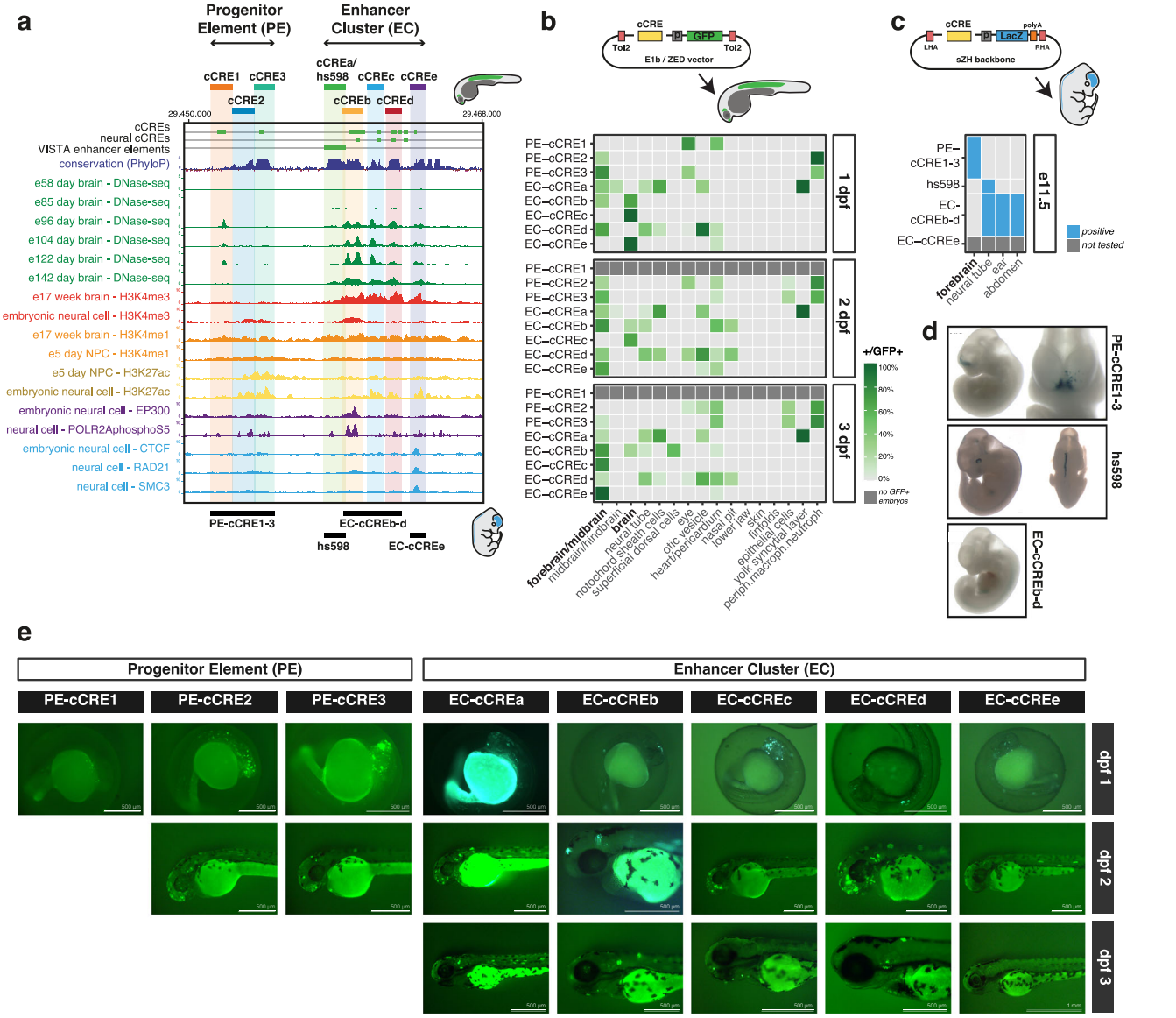


Fig. 4 | Candidate cis-regulatory elements (cCREs) show tissue-specific enhancer activity. a Delineation of candidate cis-regulatory elements (cCREs) within the progenitor element (PE) and enhancer cluster (EC) for in vivo enhancer assays in zebrafish and mouse embryos, based on conserved sequence stretches and publicly available epigenetic signatures in human fetal brain samples. **b** Frequency of transgenic F0 zebrafish embryos (ZED[41] vector) with tissue-specific GFP signal among total GFP-positive embryos at 1, 2, and 3 days post-fertilization (dpf). **c** Summary of tissue-specific signal observed in e11.5 transgenic mouse embryos for two tested constructs (sZH backbone) and the previously validated VISTA element hs598/EC-cCREa (hZR backbone). **d** Representative images of LacZ reporter expression in mouse embryos at e11.5. **e** Representative images of GFP reporter expression (ZED vector) in zebrafish embryos at 1, 2, or 3 dpf for each cCRE. The total number of injected embryos and the frequency of GFP-positive embryos are provided in Supplementary Data 4.

induction among wildtype (WT) lines, likely attributable to NPC population heterogeneity using the general NPC induction protocol, hampered the statistical interpretation of these results.

To overcome the difficulties associated with variability in the general NPC induction protocol and to improve the interpretability of the perturbation experiments, we decided to interrogate all targeted regulatory elements in a lineage-specific differentiation system, i.e., fbNPCs. All WT, HS DEL^{HET}, PE DEL^{HET}, EC DEL^{HET}, TAD DEL^{HET}, and CARR DEL^{HET} lines were differentiated to fbNPCs and analysed at day 9 and day 14. In this system, we assessed *FOXG1*, *PRKD1*, and *PAX6* expression by RT-qPCR and, additionally, performed flow cytometry for *FOXG1* and PAX6 at day 14 to determine whether transcriptional changes

https://doi.org/10.1038/s41467-026-73770-1

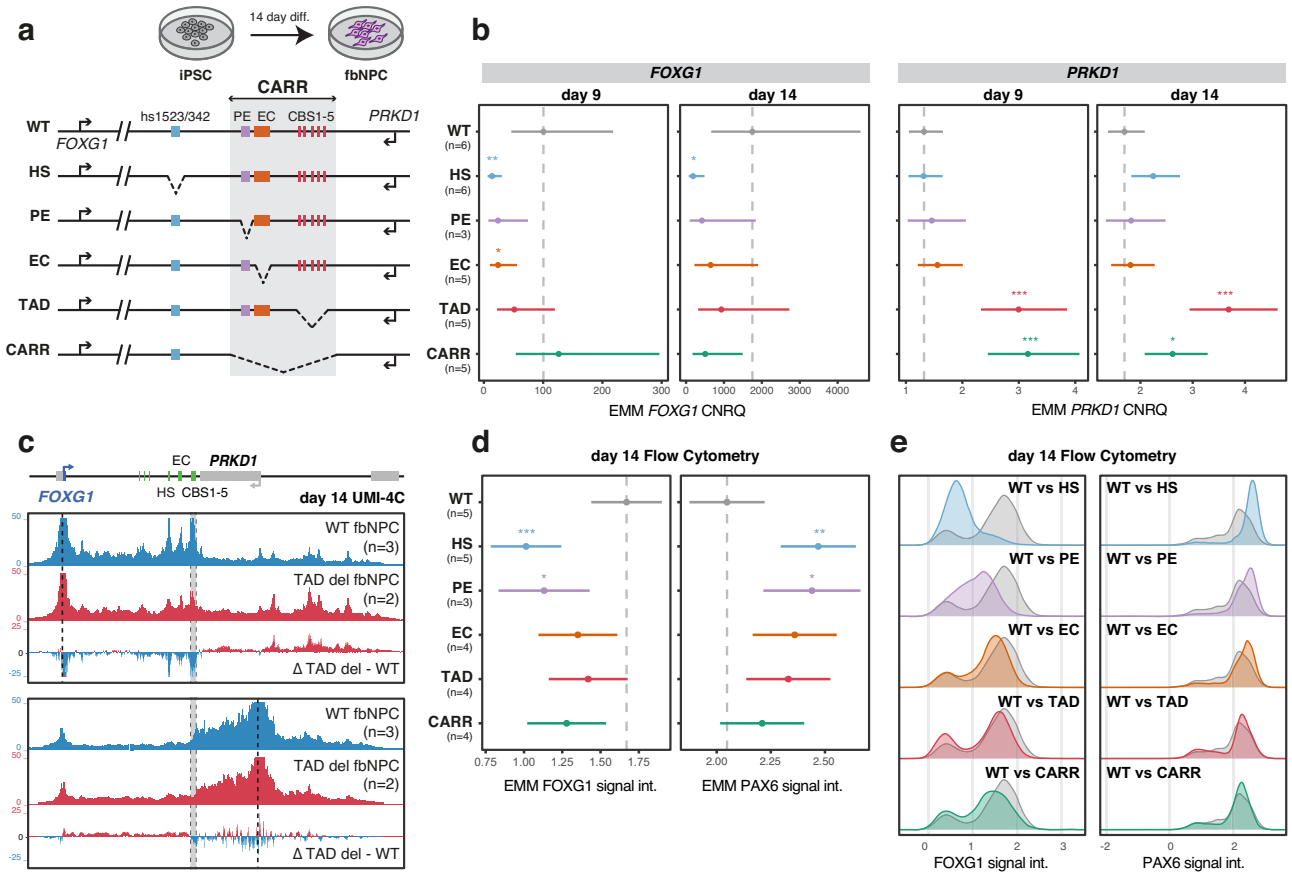


Fig. 5 | Deletion of functional elements in the *FOXG1* locus impacts *FOXG1* induction during forebrain neural progenitor cell (fbNPC) differentiation. **a** Schematic overview of heterozygous, iPSC-derived fbNPC knockout (KO) models, including targeted deletions of the hs1523/hs342 (HS) elements, progenitor 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* and *PRKD1* RNA expression at day 9 and day 14 of fbNPC differentiation for all KO models. Three to seven 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: ***($p < 0.001$), **($p < 0.01$), *($p < 0.05$). d9_HS DEL^{HET}_*FOXG1* ($p = 0.0034$), d9_EC DEL^{HET}_*FOXG1* ($p = 0.0438$), d14_HS DEL^{HET}_*FOXG1* ($p = 0.0034$), d9_TAD DEL^{HET}_*PRKD1* ($p = 0.0004$), d9_CARR DEL^{HET}_*PRKD1* ($p = 0.0002$), d14_TAD DEL^{HET}_*PRKD1* ($p = 0.0002$), d9_CARR DEL^{HET}_*PRKD1* ($p = 0.0291$). Source data are provided as a Source data file. **c** *FOXG1* and *PRKD1* UMI-4C interaction frequencies at day 14 of fbNPC differentiation in control versus TAD boundary (gray box) KO lines. The dotted line indicates the viewpoints. Subtraction ("delta") profiles highlight differential interactions. **d** Estimated marginal means (EMM) and 95% confidence intervals of *FOXG1* and PAX6 signal intensity as measured through flow cytometry at day 14 of fbNPC differentiation for all KO models. Three to six independent clones from each genotype were included. Generalized additive mixed model with two-sided Dunnett's adjusted *p*-values were calculated using the EMM contrasts and are indicated as follows: ***($p < 0.001$), **($p < 0.01$), *($p < 0.05$). d14_HS DEL^{HET}_PAX6 ($p = 0.0037$), d14_PE DEL^{HET}_PAX6 ($p = 0.0297$), d14_HS DEL^{HET}_*FOXG1* ($p = 0.0004$), d14_PE DEL^{HET}_*FOXG1* ($p = 0.0226$). Source data are provided as a Source data file. The gating strategy used for flow cytometry analysis is shown in Supplementary Fig. 21c. **e** Density plots of *FOXG1* and PAX6 signal intensity as measured through flow cytometry at day 14 of fbNPC differentiation for all KO models versus WT. Source data are provided as a Source data file. The gating strategy used for flow cytometry analysis is shown in Supplementary Fig. 21c.

corresponded to shifts in the proportions of *FOXG1*$^{+}$ and PAX6^{+} fbNPC populations.

Mean *FOXG1* RNA expression at day 9 was significantly reduced in HS DEL^{HET} (estimated marginal mean (EMM) = 13.9, 95% CI [6.4–30.3], $p = 0.0034$) and EC DEL^{HET} (EMM = 23.9, 95% CI [10.2–56.1], $p = 0.0438$) lines compared to WT lines (EMM = 100.5, 95% CI [46.3–218.0]) (Fig. 5b and Supplementary Fig. 18). By differentiation day 14, HS DEL^{HET} *FOXG1* RNA expression was 9.5 fold reduced (EMM = 183.5, 95% CI [69.4–485.5], $p = 0.0138$), while the other genotypes resulted in an observable (1.9- to 4.2-fold) but not significant decrease in *FOXG1* transcript levels compared to WT (EMM = 1750.6, 95% CI [665.8–4603.0]). In contrast, *PRKD1* expression was increased in TAD DEL^{HET} and CARR DEL^{HET} fbNPCs at both day 9 and day 14 compared to WT ($p < 0.05$) (Fig. 5b). We hypothesized that the 1.5–2.4 fold increase in *PRKD1* transcription is likely caused by the ectopic interaction between *FOXG1* enhancer elements and the *PRKD1* promoter due to the removal of the TAD boundary separating the *FOXG1* and *PRKD1* TADs. UMI-4C interaction mapping for *FOXG1* and *PRKD1* on day 14 TAD DEL^{HET} fbNPCs indeed showed increased interaction between *FOXG1* and the *PRKD1* TAD and vice versa (Fig. 5c and Supplementary Fig. 20). *PAX6* expression, a marker of early forebrain progenitors, was induced in all KO models (Supplementary Fig. 18). Interestingly, we noted a significant negative correlation between *FOXG1* and *PAX6* RNA expression (d14 corr coef = −0.5782, $p = 0.0008$) (Supplementary Fig. 18), consistent with the binding of the *FOXG1* repressor at the *PAX6* promoter in fbNPCs (Supplementary Fig. 19).

The trends observed at the transcriptional level were further supported by the *FOXG1* and PAX6 signal intensities measured through flow cytometry in day 14 fbNPCs (Fig. 5d and Supplementary Fig. 21), with the strongest reduction in *FOXG1* signal in HS DEL^{HET} fbNPCs (EMM = 1.0, 95% CI [0.8–1.2]) vs. WT (EMM = 1.7, 95% CI [1.4–1.9] $p = 0.0004$). We also observed a significant reduction in *FOXG1* signal in PE DEL^{HET} fbNPCs (PE EMM = 1.1, 95% CI [0.8–1.4]) vs. WT ($p = 0.0226$) and a trending decrease in the other models. Again, a negative

correlation between *FOXG1* and PAX6 signal intensities (corr coef = −0.622, $p = 0.0009$) was observed (Fig. 5d). Interestingly, *FOXG1* signal density plots revealed the presence and differential abundance of distinct fbNPC populations (Fig. 5e). In WT fbNPCs (gray), *FOXG1* intensities displayed a distinct bimodal distribution, indicating the presence of both low- and high-expressing *FOXG1* populations within the forebrain progenitor pool. In HS DEL^{HET} fbNPCs, the proportion of *FOXG1*-high cells was severely diminished, in agreement with RT-qPCR results at both time points. In EC DEL^{HET}, TAD DEL^{HET}, and CARR DEL^{HET} fbNPCs, the high-intensity *FOXG1* peak was subtly shifted leftwards toward lower intensities and accompanied by an increased low-intensity peak in TAD DEL^{HET} and CARR DEL^{HET}. PE DEL^{HET} showed a markedly broadened *FOXG1* distribution lacking clear peak structure, indicative of increased cell-to-cell and line-to-line variability and the presence of multiple subpopulations with differing *FOXG1* levels. PAX6 density plots revealed a mostly uniform PAX6 induction within the forebrain progenitor pool (Fig. 5e). In most KO models, the high-intensity peak median was slightly shifted upwards. This was again most pronounced in HS DEL^{HET} fbNPCs, consistent with the observations at the RNA level.

Together, the qPCR and flow-cytometry data demonstrate that regulatory deletions within the *FOXG1* locus not only modulate *FOXG1* transcription but also alter the cellular composition of the fbNPC population.

Discussion

The introduction of novel sequencing technologies in the clinic has led to improved SV detection and an increasing number of reported cases with putatively causal noncoding SVs[9]. Still, interpreting the functional impact of these SVs remains challenging, given the incomplete annotation of the noncoding human genome and their potential impact on 3D chromatin structure. The 14q12 *FOXG1* locus is recurrently affected by noncoding structural aberrations[16]. Through a comprehensive database and literature search, complemented by the identification of two additional cases (i.e., bca4 and del8), we assembled a cohort of 40 individuals with a *FOXG1* syndrome-like disorder, harboring SVs that disrupt the *FOXG1* regulatory domain. No clear association was observed between phenotype severity and either the type (CNV or BCA) or size of the structural variation. This may partially be attributed to the limited phenotypic information available for certain cases, potentially obscuring more nuanced associations.

Despite the diverse spectrum of SVs, with some CNVs affecting several megabases and encompassing multiple other protein-coding genes, in silico analysis using POSTRE[29] identified *FOXG1* as the only candidate causative gene for the observed neurodevelopmental phenotype, reinforcing the hypothesis that aberrant *FOXG1* expression is the underlying mechanism responsible for the *FOXG1* syndrome-like disorder in this cohort. Consistent with this, we observed a general decrease in *FOXG1* transcription in bca4, bca17, and del8 LCLs compared to control LCLs. Previous studies have also reported aberrant *FOXG1* expression in some of the included cases[15,19,20,23]. However, the effects were variable and assessed in sample types that may not adequately reflect *FOXG1* regulation and expression in a neurodevelopmental context, such as blood, fibroblasts, and, most often, LCLs. Consequently, these expression differences (or lack thereof) should be interpreted with care.

By identifying a ~416 kb de novo deletion in an individual with a *FOXG1* syndrome-like disorder, we refined the previously described smallest region of overlap to a 124 kb region downstream of *FOXG1*, disrupted across the entire *FOXG1*-like syndrome cohort, here annotated as the CARR. We show that the CARR contains two clusters of enhancer elements (PE and EC) and a TAD boundary consisting of five CBSs. In vivo enhancer assays validated neural enhancer activity for both PE and EC, each driving tissue-specific reporter expression patterns in zebrafish and mouse embryos, suggesting a role in regulating *FOXG1* expression within specific developmental contexts[36,42–46]. Moreover, the absence of detectable chromatin accessibility at the PE locus in publicly available human embryonic brain DNase-seq data further suggests that PE activity is likely restricted to specific cell types and developmental timepoints.

To further delineate the regulatory mechanisms underlying FOXG1 induction in a relevant in vitro model, we used iPSC-derived NPCs. FOXG1 is highly expressed in NPCs and orchestrates crucial processes like proliferation and differentiation[47]. Moreover, reduced FOXG1 expression impairs NPC expansion[47], ultimately resulting in microcephaly in individuals with *FOXG1* syndrome or *FOXG1* syndrome-like disorder. We performed epigenomic profiling during both general and forebrain-directed NPC differentiation protocols. This revealed the strongest activation and highest *FOXG1* contact frequencies for EC, PE, and the known forebrain VISTA enhancers hs1523/hs342 (HS), coinciding with transcriptional induction of *FOXG1*. While HS, PE, and EC exhibited dynamic activation and increased *FOXG1* promoter contacts under both differentiation conditions, the strongest activation was observed with the fbNPC protocol, indicating that their regulatory activity peaks in a forebrain-specified developmental context. For hs1523/hs342, this is further supported by the strong interactions with the *Foxg1* promoter found in mouse embryonic forebrain[48].

Next, we assessed the impact of perturbing the HS, PE, and EC enhancer elements, as well as the distal TAD boundary, on *FOXG1* expression. Perturbation assays during general NPC differentiation did not yield statistically interpretable results due to the variability among control *FOXG1* induction levels, associated with diverse cell populations obtained using dual SMAD inhibition protocols[36]. Deletion of both the HS and EC regions led to a strong decrease in *FOXG1* transcription early during fbNPC induction. At the endpoint of fbNPC differentiation, all models indicated reduced *FOXG1* RNA levels, with the strongest effect observed in HS DEL^{HET} fbNPCs. Even though forebrain-directed NPC differentiation increases overall cellular homogeneity, flow cytometry revealed the presence of distinct fbNPC populations with either high or low *FOXG1* expression. The balance between these two populations had most clearly shifted in HS DEL^{HET} fbNPCs, with the *FOXG1*-high population being almost completely absent. This indicates that in the absence of the hs1523/hs342 forebrain enhancers, NPCs fail to differentiate towards a *FOXG1*+ forebrain fate. In contrast, deletion of the EC and PE regions, although both exhibited equal or higher levels of H3K27ac in fbNPCs and forebrain activity in in vivo enhancer assays, only slightly shifted the *FOXG1*-high population.

Using Hi-C and UMI-4C on case LCLs (del8), we have shown that deletion of the entire CARR and CBS cluster removes the conserved *FOXG1* TAD boundary and results in increased contacts with the distal TAD. This was also confirmed in the in vitro TAD DEL^{HET} and CARR DEL^{HET} fbNPCs models. Furthermore, these deletions resulted in increased *PRKD1* expression, likely caused by ectopic interaction between *FOXG1* enhancer elements and the *PRKD1* promoter. Since all BCA cases involve distinct derivative chromosomes, and in all del cases (except for del8), *PRKD1* is also affected, ectopic expression of *PRKD1* is unlikely to contribute to the phenotype in these individuals. In addition, *FOXG1* transcription was slightly reduced in TAD DEL^{HET} fbNPCs. This coincided with decreased *FOXG1* contact frequency with HS, PE, and EC enhancer elements, possibly indicating a competition for interaction due to increased contact of the *FOXG1* promoter with loci in the distal TAD.

Previously, Ibn-Salem et al.[18] had proposed (fore)brain enhancer adoption by *FOXG1* as a potential pathomechanism in cases with deletions spanning the TAD boundary. However, our results support the case made by Mehrjouy et al.[10] against an enhancer adoption mechanism. We found a high phenotypic similarity within the assembled cohort of 40 cases, in many of whom either the TAD boundary

remained intact (translocation cases) or the deletion encompassed both the TAD boundary and enhancers (e.g., hs433) within the *PRKD1* TAD (most deletion cases). Also, our UMI-4C data for the *FOXG1* promoter in NPCs and neurons indicate that *FOXG1* interacts with elements in the *PRKD1* TAD, including hs433. While most frequent interactions are found within the *FOXG1* TAD (especially in NPCs), the insulation between both domains appears to weaken in more differentiated cell types (i.e., neurons). We also show that, while deletion of the TAD boundary increases *FOXG1* interaction frequencies with (fore) brain enhancers (including hs433, Fig. 5c) in the *PRKD1* TAD, mean *FOXG1* transcription in TAD boundary knockout fbNPCs was reduced (Fig. 5b). This indicates that reduced contact with its native forebrain enhancers outweighs the gain of interaction with elements, such as hs433.

While the CARR is affected in all *FOXG1* syndrome-like individuals with an SV downstream of *FOXG1*, in none of these cases is the aberration restricted to this region alone. Even though identifying the smallest region of overlap has been used as a strategy to identify functional noncoding regions in the past[10,49–53], it is important to consider that the phenotype in these individuals may arise from a cumulative effect caused by the deletion or translocation of multiple regulatory elements within the *FOXG1* TAD. In line with this, deletion of the entire CARR resulted in a lower mean *FOXG1* expression in day 14 fbNPCs than deletion of the individual PE and EC regions. Still, only deletion of the hs1523/hs342 forebrain enhancers resulted in an almost complete loss of the *FOXG1*-high fbNPC population. Given that these elements are removed in 36/40 individuals in the updated cohort, we propose that loss of the hs1523/hs342 forebrain enhancers is the principal driver of the *FOXG1* syndrome-like phenotype in these cases. The previous identification of two rare, de novo variants in the constrained region (i.e., Z-score of 4.02 in gnomAD v4.1.0, Supplementary Data 6) of the hs1523 element in individuals with neurodevelopmental disorders reinforces a critical neurodevelopmental role for the hs1523/hs342 enhancer region[48,54].

In conclusion, this study provides an in-depth functional characterization of the regulatory elements driving *FOXG1* expression in neurodevelopment, specifically during NPC differentiation. As such, our findings lay a foundation for the improved interpretation of (noncoding) variants at the 14q12 *FOXG1* locus.

Methods

Ethics

All human studies were conducted in accordance with the Declaration of Helsinki.

Ethical approval for the use of commercially available eukaryotic cell lines was obtained from the Ethics Committee of Ghent University Hospital under protocol number ONZ-2022-0100.

Ethical approval for the use of patient-derived cell lines and patient-derived material, involving the two probands described in the Supplementary Information, was obtained from the Ethics Committee of Ghent University Hospital under protocol number EC 2019/1430. Written informed consent was obtained from both participants and/or their legal guardians prior to sample collection and study participation, including consent for publication of potentially identifiable clinical information.

All mouse experiments were reviewed and approved by the Lawrence Berkeley National Laboratory Animal Welfare and Research Committee under protocol numbers 290003 and 290008.

All zebrafish experiments were performed prior to 5 days post-fertilization in accordance with applicable regulations and did not require ethical approval.

Cell culture

Lymphoblastoid cell lines (LCLs). LCLs were cultured in RPMI (Roswell Park Memorial Institute). RNA was extracted using the Maxwell® RSC instrument (AS4500, Promega) using the Maxwell® RSC Simply RNA Tissue kit (AS1340, Promega). Case lymphoblastoid cell lines (LCLs) were derived in-house from patient blood samples (del8, bca4, and bca17). Control LCLs were obtained from the Coriell Institute for Medical Research (ND02050, ND09729, ND11929, ND23969, ND23993, ND24284, GM07544, GM23712; GM23712).

Induced pluripotent stem cells (iPSCs). Human iPSC cells (Axol Bioscience, lot #30270222072019) were cultured in feeder-free conditions on Geltrex® LDEV-Free hESC-Qualified Reduced Growth Factor Basement Membrane Matrix (Fisher Scientific #12063569)-coated 6-well plates in mTeSR Plus medium (Stemcell Technologies #100-0276) that was changed every other day. Cells were dissociated with ReLeSR (Stemcell Technologies #100-0483) after reaching 80% confluency.

Neural stem cells (NSCs) and induced neurons (iNs). To minimize technical artifacts, hiPSCs were plated as single cells at 80% confluence on a Matrigel-coated 6-well plate with Y-27632 dihydrochloride (1 mg/mL). Polybrene was added at 8 mg/mL 1 h after re-plating. Cells were incubated with polybrene for 10–15 min prior to the addition of lentivirus. Lentiviral constructs for directed differentiation of hiPSCs into iNs were made as described previously[55]. hiPSC lines were differentiated into NSCs with the protocol detailed by Thermo Fisher (MAN0008031)[56].

Neural progenitor cells (NPCs). iPSCs were differentiated towards NPCs using the STEMdiff™ SMADi Neural Induction Kit following the monolayer protocol provided by the manufacturer (Catalog #08581), which directs differentiation by blocking TGF-β/BMP-dependent SMAD signaling. NPCs were seeded at a density of 4×10^5 cells per well in a Geltrex-coated 12-well plate in STEMdiff™ SMADi Neural Induction medium over 8 or 12 days, with media changes each day. For the 12-day protocol, the cells were split at day 6; for the 8-day protocol, no split was performed.

Forebrain neural progenitor cells (fbNPCs). iPSCs were differentiated towards forebrain-specific NPCs using a paired dual-SMAD and Wnt/β-catenin signaling inhibition protocol defined by Neaverson et al.[39]. iPSCs were dissociated into single cells using Accutase (Stemcell Technologies # 07920) and seeded at a density of 7.6×10^5 cells per well on Geltrex-coated 12-well plates in mTeSR Plus supplemented with 10 µM ROCK inhibitor (Y-27632). After 24 h, once a uniform monolayer had formed, the medium was replaced with neural induction medium (NIM) (TeSR™-E6, Stemcell Technologies # 05946) containing dual-SMAD inhibitors (10 µM SB431542, 1 µM Dorsomorphin) and a Wnt/β-catenin pathway inhibitor (2 µM XAV939). Cells were maintained in NIM for 9 days, with daily medium changes. Cells typically formed a continuous neuroepithelial sheet by day 7–9. At day 9, the sheet was dissociated to single cells using Accutase and replated at 7.6×10^5 cells per well on Geltrex-coated 12-well plates in neural maintenance medium (NMM). Cells were then cultured in NMM for the remainder of the differentiation, with medium changes every other day, until collection day 14 for downstream assays.

DNA extraction

Total DNA was extracted from iPSCs and iPSC-derived NPCs at day 8 of differentiation. The cells were rinsed twice with D-PBS and harvested with a cell scraper. The PBS-containing cells were centrifuged ($200 \times g$ for 5 min), the supernatant was discarded, and the pellet was resuspended in 200 µL D-PBS. DNA was obtained from this solution using the QiaAmp DNA Blood Mini kit (Qiagen #51106) on the Maxwell rapid Sample Concentrator (RSC), which allows automated extraction of DNA from different types of cells.

CNV-seq and breakpoint fine-mapping

Molecular karyotyping was performed through low-pass whole-genome sequencing on genomic DNA (gDNA). Whole genome libraries were prepared with the Illumina DNA PCR-Free Prep kit (20041795, Illumina) according to the manufacturer's instructions. The libraries were subsequently paired-end sequenced (2 × 50 bp) on a NovaSeq6000 system (Illumina) to a depth of 75 M reads. The raw sequencing data were then aligned to the human reference genome GRCh38/hg38 with bowtie2[57]. Further processing and analyses were done using WisecondorX[58] and Vivar[59]. Breakpoints were then fine-mapped at base pair resolution through CNV-qPCR[60], followed by PCR amplification over the putative breakpoints and Sanger sequencing of the amplicon.

Long-read genome sequencing

Long-read genome sequencing (lrGS) of the bca4 proband was performed using the sequencing platform from Oxford Nanopore Technologies (ONT). Libraries of sizes ranging from 20 to 25 kb were generated and sequenced on a PromethION device using R10.4.1 flow cells. Reads were called with Guppy and aligned to the human genome reference hg38 with minimap2. Structural variants, including the exact breakpoints of the balanced translocation, were called with Sniffles2 and cuteSV, and single-nucleotide variants with Clair3 and Longshot. Variant files from lrGS were annotated with GenomeComb. Orthogonal validation of the identified breakpoints was done through PCR amplification.

RNA isolation, cDNA synthesis and quantitative PCR

Total RNA was extracted using QIAzol Lysis reagent (Qiagen #79306) and the Direct-zol RNA Miniprep Kit (Zymo Research #R2052). Complementary DNA (cDNA) was synthesized using the iScript cDNA synthesis kit (Bio-Rad #170-8891). Quantitative polymerase chain reaction (qPCR) was performed using 2× SYBR Green SsoAdvanced Supermix (Bio-Rad #172-5274) with 5 μM forward and reverse primers and 5 ng cDNA input. Expression values were analysed using qBase + software (Biogazelle, Ghent, Belgium) and normalized using two reference genes (UBC, SDHA). RNA expression data were analysed using linear mixed-effects models as described in the "Statistical analysis" section.

Flow cytometry

Flow cytometry was used to quantify PAX6- and *FOXG1*-positive cell populations and evaluate protein-level distributions in comparison to qPCR results. Day-14 fbNPCs were harvested by rinsing with D-PBS and incubating with 1 mL Accutase per well for 5.5 min. Detached cells were diluted in DMEM (2 mL per 1 mL Accutase), centrifuged (300 × *g*, 5 min), and resuspended in 1–2 mL NMM supplemented with Y-27632 (10 μM) for counting and allocation.

For staining, 5×10^5 cells per sample were washed with PBS (490 × *g*, 6 min, RT). Dead cells were labeled using LIVE/DEAD Fixable Aqua Dead Cell Stain Kit (0.2 μL dye in 48.9 μL PBS per sample, 405 nm excitation; Invitrogen L34957) for 20 min at RT in the dark, followed by two PBS washes. Cells were fixed and permeabilized using the FoxP3/Transcription Factor Staining Buffer Set (eBioscience 00-5523-00). A 1:4 dilution of fixation/permeabilization buffer (500 μL per sample) was applied for 30 min at RT, followed by one wash in 1× permeabilization buffer (PB) and two additional PBS washes. Samples were stored in PBS at 4 °C until intracellular staining.

Primary antibodies (Supplementary Data 7) were applied for 20 min at RT in 50 μL PB containing rabbit anti-*FOXG1* (1:60, Abcam ab196868) and/or Alexa Fluor 488 mouse anti-human PAX6 (1:20, BD Biosciences 561664). After two washes in PB, secondary staining was performed for 20 min at RT in the dark with Alexa Fluor 594 donkey anti-rabbit IgG (1:1000 in 150 μL PB, Invitrogen A21207). Cells were washed once and resuspended in 150 μL PBS for acquisition. Samples were run on a BD LSR Fortessa X-20 5-laser cytometer. Debris, doublets, and nonviable cells were excluded by sequential gating. Flow data were processed in FlowJo (v10.6) and analysed in R (v4.5.1) using generalized additive mixed models (GAMMs) as detailed in the "Statistical analysis" section.

Functional characterization of cCREs in zebrafish

The cCRE sequences were synthesized as gBlocks (IDT) with 30 bp flanking regions designed to overlap with the ends generated by AscI and PacI restriction digestion of the E1b vector[40]. The digested E1b vector and gBlocks were assembled using Gibson assembly (Gibson Assembly Master Mix, NEB # E2611L). In parallel, the same gBlocks were cloned into a TOPO vector via A-tailing. The TOPO vector, containing Gateway recombination sites, was then used for Gateway cloning into the ZED vector[41] using the Gateway LR Clonase II enzyme mix (Invitrogen #11791-020). E1b contains a minimal promoter followed by a GFP reporter gene. ZED, in addition, includes a transgenesis internal control cassette comprising the cardiac actin promoter driving the expression of red fluorescent protein and insulator sequences to mitigate position effect artifacts. The combination of these vectors enables a more accurate assessment of enhancer functionality, as each provides complementary insights into enhancer-driven gene expression. These vectors were microinjected into one-cell stage zebrafish embryos (Danio Rerio, AB strain), along with Tol2 transposase mRNA for genomic integration. As a readout, GFP fluorescence was observed, and its localization was annotated at 1, 2, and 3 dpf to evaluate enhancer activity. Zebrafish embryos used in enhancer reporter assays were imaged using a stereomicroscope (Leica M165 FC, Leica Microsystems) equipped with a PlanApo 1.0× objective. Images were captured using a Leica DFC450 C digital camera. Image acquisition was performed using Leica Application Suite (LAS v4.3).

Mouse transgenic enhancer assays

Transgenic E11.5 mouse embryos (Mus Musculus, FVB strain) were generated as described previously[61]. Briefly, super-ovulating female FVB mice were mated with FVB males and fertilized embryos were collected from the oviducts. Regulatory elements sequences were synthesized by Twist Biosciences. Inserts generated in this way were cloned into the donor plasmid containing a minimal Shh promoter, lacZ reporter gene and H11 locus homology arms (Addgene, 139098) using NEBuilder HiFi DNA Assembly Mix (NEB, E2621). The sequence identity of donor plasmids was verified using long-read sequencing (Primordium). Plasmids are available upon request. A mixture of Cas9 protein (Alt-R SpCas9 Nuclease V3, IDT, Cat#1081058, final concentration 20 ng/μL), hybridized sgRNA against H11 locus (Alt-R CRISPR-Cas9 tracrRNA, IDT, cat#1072532 and Alt-R CRISPR-Cas9 locus targeting crRNA, gctgatggaacaggtaacaa, total final concentration 50 ng/μL), and donor plasmid (12.5 ng/μL) was injected into the pronucleus of donor FVB embryos. The efficiency of targeting and the gRNA selection process is described in detail in Osterwalder 2022[61]. Embryos were cultured in M16 with amino acids at 37 °C, 5% CO_2 for 2 h and implanted into pseudopregnant CD-1 mice. Embryos were collected at E11.5 for lacZ staining as described previously[61]. Briefly, embryos were dissected from the uterine horns, washed in cold PBS, fixed in 4% PFA for 30 min and washed three times in embryo wash buffer (2 mM $MgCl_2$, 0.02% NP-40 and 0.01% deoxycholate in PBS at pH 7.3). They were subsequently stained overnight at room temperature in X-gal stain (4 mM potassium ferricyanide, 4 mM potassium ferrocyanide, 1 mg/mL X-gal and 20 mM Tris, pH 7.5 in embryo wash buffer). PCR using genomic DNA extracted from embryonic sacs digested with DirectPCR Lysis Reagent (Viagen, 301-C) containing Proteinase K (final concentration 6 U/mL) was used to confirm integration at the H11 locus and test for the presence of tandem insertions[61]. Only embryos with donor plasmid insertion at H11 were used. Embryos were not sexed. The stained transgenic embryos were washed three times in PBS and

imaged from both sides using a Leica MZ16 microscope and a Leica DFC420 digital camera. Animals were maintained under standard housing conditions (dark light cycle: 12 h on, 12 h off (light on 6 am–6 pm), temperature: 69–75 F, humidity: 30–70%).

Identification of candidate cis-regulatory elements (cCREs)

A set of cCREs with enhancer-like epigenomic properties was created by overlapping entries in the annotated regulatory index (DHS) by Meuleman et al. (meuleman.org)[32,33] and enhancer-annotated entries in the ENCODE SCREEN v3 dataset[34] using the GenomicRanges R package. To obtain a complete list of cCREs with accessibility in neuronal tissues, we selected peaks defined by Meuleman et al. as "tissue irrelevant" or "nervous," showing accessibility in 10% of relevant samples (Organ = "Neural"/"Brain"/"Germ"). Finally, enhancer-like cCREs located within the same TAD as *FOXG1* were selected, based on TAD boundaries determined using brain Hi-C data by Rahman et al.[32].

UMI-4C library preparation and analysis

UMI-4C was performed according to the protocol by Schwartzman et al.[30]. In brief, 1 (iNs), 5 (iPSCs, NSCs, and NPCs) or 10 (LCLs) million cells per sample were detached, resuspended in a PBS/10% FBS solution and crosslinked using formaldehyde (final concentration 2%) (Sigma-Aldrich, F1635-500ML) for 10 min. The reaction was quenched on ice using glycine (final concentration 0.125 M). Cells were then centrifuged, washed and resuspended in cold lysis buffer (50 mM Tris-HCl, 150 mM NaCl, 5 mM EDTA, 0.5% NP-40, 1% Triton X-100 and 1 tablet protease inhibitor per 10 mL (complete, Mini, Sigma-Aldrich, 11836153001)) and incubated for 10 min. After centrifugation, pellets were washed, snap-frozen in liquid nitrogen and stored at −80 °C until further processing. Frozen pellets were resuspended in 500 μL pre-diluted DpnII buffer and 15 μL pre-heated 10% SDS and incubated on a thermomixer for 1 h at 37 °C, shaking at 900 RPM. We halved these volumes and those mentioned below for smaller cell pellets (from 1 million cells). After adding 150 μL 10% Triton X-100, the solution was incubated again (1 h, 37 °C, 900 RPM). We digested the chromatin using 600 U DpnII (NEB, R0543L) in three stages (200 U for 2 h, 200 U overnight, 200 U for 2 h) at 37 °C and 900 RPM. The solution was incubated at 65 °C for 20 min to inactivate the restriction enzyme and put on ice. Next, the chromatin was ligated by adding 4000 U of T4 DNA ligase (NEB, M0202M) and 10× T4 DNA ligase buffer to a total volume of 1300 μL and incubating the solution overnight at 16 °C and 300 RPM. Ligated chromatin was de-crosslinked by incubation with 8 μL proteinase K (20 mg/mL, Qiagen, 19131) (overnight, 65 °C, 300 RPM). A 3C template was then purified using 1× Ampure XP beads (Beckman Coulter, A63881).

Up to 4 μg 3C template per sample was sheared using microTUBE snap-cap tubes (Covaris, 520045) in a Covaris M220 sonicator to an average fragment length of 300 bp. UMI-4C sequencing libraries were generated using the NEBNext Ultra II library prep kit (NEB, E7645L). For each sample, library prep was performed in four parallel reactions with a maximum input of 1000 ng per reaction, including an AmpureXP bead size selection targeting fragments with a length of 300–400 bp (unless input was <100 ng) and 4–8 cycles of PCR enrichment depending on the input. Next, two nested PCR reactions were performed to enrich for fragments captured by the viewpoint of interest, both using 2 μL 10 mM Illumina enrichment primer 2 and either a viewpoint-specific "upstream" (reaction 1, 2 μL 10 mM) or "downstream" (reaction 2, 2 μL 10 mM) primer (Supplementary Data 8). For each sample, we performed up to 8 nested PCR reactions in parallel with an input of 100–200 ng per reaction using the KAPA2G Robust ready mix (Sigma-Aldrich, KK5702) in a volume of 50 μL. PCR program: 3 min 95 °C, 20 cycles (18 cycles for reaction 2) of 15 s 95 °C, 15 s 55 °C, 60 s 72 °C and final elongation of 5 min 72 °C. Between PCR reactions, the product was cleaned up using 1× AmpureXP beads and eluted in 21 μL. The final PCR product was cleaned up using 0.7× AmpureXP beads and eluted in 25 μL. Reactions per sample were pooled, and library concentration was quantified via qPCR using the KAPA SYBR FAST qPCR Master Mix (Sigma-Aldrich, KK4602).

UMI-4C libraries were pooled and sequenced on an Illumina HiSeq or NovaSeq (paired-end, 2 × 150 cycles). Reads were first filtered based on the presence of the downstream primer sequence (20% mismatch allowed). The resulting fastq files were used as input to the UMI-4C R package (https://github.com/tanaylab/umi4cpackage) to generate genomic interaction tracks representing UMI counts (i.e. unique interactions) per genomic restriction fragment. The package was then used to generate smoothed, viewpoint-specific interaction profiles for the region of interest. For each profile, interaction counts were normalized to the total UMI count within the profile.

Hi-C library preparation and analysis

Case LCLs were crosslinked, lysed and snapfrozen according to the protocol mentioned above under "UMI-4C." Crosslinked nuclei were then used to construct Hi-C libraries, following the in situ Hi-C protocol adopted by the 4D Nucleome consortium[31] with a few adaptations, as previously described by D'haene et al.[62]. Briefly, for each replicate, ~5 million pre-lysed, crosslinked nuclei were digested overnight using 250 U DpnII restriction enzyme (New England Biolabs, R0543L). DNA ends were marked by incorporating biotin-14-dATP (Life Technologies, 19524-016) and ligated for 4 h using 2000 U T4 DNA ligase (New England Biolabs, M0202L). Subsequently, crosslinks were reversed overnight using proteinase K (Qiagen, 19131), and Hi-C template DNA was purified using 1× AMPure XP beads (Beckman Coulter, A63881) and stored at 4 °C until library preparation. Hi-C template DNA was sheared to a size of 300–500 bp using microTUBE snap-caps (Covaris, 520045) in a Covaris M220 sonicator, and MyOne streptavidin T1 beads (Life Technologies, 65601) were used to pull down biotinylated ligation junctions. Next, samples were split into 5 μg aliquots for sequencing library preparation using the NEBNext Ultra II DNA Library Prep Kit (New England Biolabs, E7645L) and NEBNext Multiplex Oligos (New England Biolabs, E7335L). Amplified libraries were purified and size selected using 0.55× and 1.2× AMPure XP beads (Beckman Coulter, A63881). Pooled libraries were sequenced on an Illumina NovaSeq 6000 using 100 bp paired-end reads to a depth of ~500 million reads per sample.

FASTQ files containing raw sequencing data were processed into Hi-C contact matrices containing both raw and Knight–Ruiz (KR) normalized counts using the Juicer pipeline (v1.6)[63] with BWA-MEM mapping (v0.7.17)[64] to the hg38 reference genome. A processed and normalized Hi-C matrix for control LCLs was obtained from Melo et al.[13]. Insulating boundaries between self-interacting domains were identified based on diamond insulation score minima. We used cooltools[65] to calculate a genome-wide contact insulation score with 250 kb window size for KR normalized Hi-C contact matrices (MAPQ > 30) at 25 kb resolution. Insulating boundaries were determined by applying automated "Li" thresholding (from the scikit-image Python package) on boundary strength. We used FAN-C[66] to visualize KR normalized case vs. control Hi-C matrices and fold-change matrices for the region of interest.

Generating CRISPR Cas9-engineered iPSC lines

Design of gRNA's. The gRNAs were designed making use of the Custom Alt-R™ CRISPR-Cas9 guide RNA IDT design tool. For the KO of both the EC and TAD, two gRNAs (IDT) were designed within 1 kb upstream and downstream of the target region. The gRNAs with the highest on-target potential and lowest off-target risk were selected (Supplementary Data 9).

Formation of RNP complex. The crRNA and tracrRNA (IDT) were resuspended in IDTE buffer to final concentrations of 200 μM each. These two RNA oligonucleotides were mixed in equimolar

concentrations to a final duplex concentration of 100 μM. This mixture was heated at 95 °C for 5 min and afterwards cooled to room temperature for approximately 15 min and then placed on ice. For each well undergoing electroporation, the crRNA:tracrRNA duplex (1.01 μl per reaction) and Alt-R Cas9 enzyme (IDT) components (61 μM stock; 1.43 μl per reaction) were diluted in PBS (1.77 μl per reaction). This mixture was incubated at room temperature for 20 min and stored at 4 °C after incubation.

Electroporation of the RNP complex into iPSCs. Electroporation was performed using the Lonza® 4D-Nucleofector™ X Unit with a P3 primary cell 4D Nucleofector X kit.

Cells were collected during their exponential growth phase. The remaining medium was aspirated from the well, and 1 mL of Accutase (STEMCELL Technologies) was added to each well of a 6-well plate. The plate was incubated at 37 °C and 5% CO_2 for approximately 5 min, or until colonies appeared to be dissociated. Cells were gently washed from the surface of the plate with 2 ml mTeSR Plus (STEMCELL Technologies) using a pipettor fitted with a 1000 μl tip, pipetting the suspension up and down 2–3 times to break up small aggregates into single cells necessary for the electroporation. The cell suspension was transferred to a 15 mL conical tube, diluted in DMEM (4 mL per 1 mL Accutase), and centrifuged at 300 × *g* for 5 min. The supernatant was aspirated, being careful not to disturb the cell pellet. Cells were resuspended in at least 2 mL of Single-Cell Plating Medium (mTeSR Plus and Clone R (STEMCELL Technologies) in a 1:10 dilution, used to improve single-cell survival) and mixed by flicking the tube 2–3 times.

Cells were counted using an automated cell counter (Cellometer Auto T4). For each electroporation condition (i.e. EC, TAD, GFP, no nucleofection) 3×10^5 cells were added to a new 15 mL conical tube. The supernatant was removed without disturbing the pellet, and the cells were washed in 1 mL 1× PBS. Cells were centrifuged at 400 × *g* for 10 min and the supernatant was removed. Cells (300,000 cells/reaction) were resuspended in nucleofector solution with supplement (4.5:1 ratio solution:supplement). For each condition, 17 μl of the prepared solution was pipetted into a sterile microcentrifuge tube. For each condition (EC, TAD and GFP), 4.21 μl of each crRNA:tracrRNA (for the EC and TAD, both the upstream and downstream complexes are added, total 8.42 μl) and 0.84 μl (4 μM) of 100 μM Alt-R Cas9 Electroporation Enhancer (IDT) (total volume = ±26 μl) were added. The mixture was pipetted up and down, and 22 μl of the mixture was transferred to the wells of the nucleocuvette. The nucleocuvette was gently tapped to ensure no air bubbles were present. The nucleocuvette was placed in the device, and the electroporation program CA-137 (P3 solution) was started. After electroporation, the nucleocuvette was removed from the instrument, and 80 μl of pre-warmed culture media was added per well. Cells were resuspended by gently pipetting up and down before transferring to a Geltrex-coated 12-well plate, containing single-cell plating medium. A full medium change was performed after 24 h with mTeSR Plus. These edited iPSC were further expanded and kept in culture following standard iPSC culturing as described above.

Screening edited iPSCs in bulk. Successful editing was assessed in the bulk of the cells via DNA isolation and amplification of the target region, followed by targeted next-generation sequencing (NGS). For larger deletions, both primers outside and inside the deletion were designed (Supplementary Data 10). If successfully edited clones were present within the bulk, a PCR fragment was expected in all three PCR reactions. These were further kept in culture and expanded to perform serial limiting dilution to obtain monoclonal edited iPSC clones. Similar PCR reactions as described above were performed on the extracted DNA of these clones.

Limiting dilution. To obtain clonal cell lines, the transfected iPSCs were single-cell isolated through serial limiting dilutions. Once the bulk cells reached their exponential growth phase, they were dissociated into single cells using Accutase (STEMCELL Technologies). To prevent clumping of cells, a 40 μm cell strainer was used. The cells were diluted to a final concentration of 0.5 cells per 100 μl of Single-Cell plating medium, to reduce the likelihood of having multiple cells per well. One hundred microliters of diluted cells were added to each well of a 96-well plate, coated with Geltrex. After 4–8 days, individual colonies were observed and further used for expansion and screening.

Screening edited iPSC clones. To examine the successful deletion and assess monoclonality of the cloned hiPSCs, genomic DNA was extracted using the QuickExtract DNA isolation kit (QE09050, Epicentre). Cells were dissociated using ReLeSR (STEMCELL technologies) and resuspended into 100 μl of mTeSR Plus + 10 μM Rock Inhibitor. Half of the cells were transferred (~50 μL) to a 96-well plate containing mTesR Plus + 10 μM Rock Inhibitor for further cell expansion. The other half (~50 μL) was transferred to a 96-well PCR plate, while maintaining the location of each sample, for genomic DNA extraction. The PCR plate was spun down at 3800 rpm for 30 min at 4 °C, supernatants were removed, and 30 μl of the QuickExtract DNA solution was added to the pellet, and the protocol was performed further according to the manufacturer's instructions. Similar PCR reactions as described above were performed on the extracted DNA of these clones. Finally, monoclonality was confirmed via DNA isolation and amplification of the target region, followed by targeted NGS (indels) and CNV-seq (kb-sized deletions) (Supplementary Data 5).

CUT&RUN library preparation

On days 0, 4, 6, and 8 of differentiation, iPSCs/NPCs were washed twice with room temperature 1× PBS. Cells were lifted from the dish using a cell lifter and gently resuspended using a P1000 pipette. Cells were counted using an automated cell counter, and the appropriate number of cells was taken. For each condition, 5×10^5 cells were required, however it is recommended to batch process, and therefore cells for all conditions were pooled into a 15 ml Falcon tube. The lysate was transferred to a fresh 15 ml Falcon tube and spun down at 600 g at RT for 3 min. The supernatant was carefully discarded, and nuclei were isolated by addition of nuclei extraction buffer (NEB, 20 mM Hepes (KOH) pH 7.9, 10 mM KCl, 0.5 mM spermidine, 0.1% Triton X-100, 20% glycerol, 1 mM $MnCl_2$, 1 tablet cOmplete MINI EDTA-free Protease Inhibitor (Roche)) (100 μl per condition). The samples were incubated for 10 min on ice and spun down at 600 × *g* at 4 °C for 3 min. Supernatant was removed, and 100 μl NEB was added per condition.

For these first steps, it is recommended to work in batch and pool all samples of the same datapoint or cell type. Per reaction, 10 μl/sample concavalin-A conjugated beads were activated by washing twice in 100 μl/sample cold bead activation buffer (20 mM HEPES-KOH, pH 7.9, 10 mM KCl, 1 mM $CaCl_2$, 1 mM $MnCl_2$) on a magnetic stand and then resuspended in 11 μl/sample bead activation buffer. Freshly extracted nuclei (100 μl/sample) were added to the activated beads and incubated at 4 °C for 10 min with rotation. Beads-bound nuclei were placed on a magnetic stand and resuspended in Antibody buffer 50 μl/sample (2 mM EDTA, 1 tablet cOmplete MINI EDTA-free Protease Inhibitor (Roche) in wash buffer). Primary antibodies were added to the beads-bound nuclei and gently mixed, see Supplementary Data 7 (final dilution of 1:50 or 1:100 in 50 μl). Mixtures were incubated with rotation at 4 °C overnight on a nutator.

After washing twice on a magnetic stand in 250 μl/sample Washing Buffer (WB, 20 mM Hepes pH 7.5, 150 mM NaCl, 0.5 mM spermidine, 1 tablet cOmplete EDTA-free Protease Inhibitor), nuclei were resuspended in 50 μl/sample cold washing buffer. 2.5 μl of pAG-MNAse (15-1016-EPC Epicypher) was added to the mix and incubated with rotation for 10 min at 4 °C. Next, beads were washed twice in 250 μl/sample cold washing buffer on a magnetic stand and resuspended in 50 μl cold washing buffer. The samples were equilibrated to 0 °C in an

ice bath, and 1 µl/sample of 100 mM $CaCl_2$ was added to the reaction. After an incubation of 2 h at 4 °C on a nutator, 33 µl/sample 2X STOP buffer (340 mM NaCl, 20 mM EDTA, 4 mM EGTA, 50 µg/ml RNAseA, 50 µg/ml glycogen, 10 ng/ml spike-in DNA) was added, and beads were incubated for 10 min at 37 °C. The supernatant was extracted, and DNA was purified using the CUTANA™ Quick Cleanup DNA Purification Kit, following the manufacturer's instructions. DNA concentration was measured with Qubit (Thermo Fisher Scientific).

Libraries were prepared using the NEBNext Ultra II Library prep Kit (E7645S, New England Biolabs) according to the manufacturer's protocol using half of the volumes. Adapters were used at a working concentration of 15, 1.5, 0.6 µM based on the DNA input. A cleanup of the adapter-ligated DNA without size selection was performed, making use of AMPure XP beads (1.8X). Libraries (13 µL) were PCR amplified with unique indexes (15 µL NEBNext Ultra II Q5 Master Mix and 1 µL i7 Primer) (New England Biolabs) for 8–15 cycles depending on the start DNA input.

A cleanup of the PCR reaction with size selection was performed using 0.5× and 1.2× AMPure XP beads. After the cleanup, the fragment size and concentration was analyzed by Fragment Analyzer. CUT&RUN library concentrations were measured with the Illumina Kapa Library quantification kit (Roche #07960140001). All pre-made libraries were sequenced on the Illumina NextSeq 2000 platform (150 bp paired-end sequencing), at a sequencing depth of 10 million reads per sample. The resulting fastq files were used as input to the nf-core/cutandrun package (v3.2.2) with Bowtie2 (v2.4.4) mapping.

Statistical analysis

All statistical analyses were performed in R (v4.5.1) using the "tidyverse," "lme4," "lmerTest," and "emmeans" packages, unless otherwise specified.

Replicate design. All analyses accounted for the hierarchical structure of the data. Independent differentiations and clonal iPSC lines were modeled as random effects to account for non-independence and batch-to-batch variation. Technical replicates (e.g., qPCR triplicates) were averaged prior to statistical testing.

qPCR analysis. Normalized expression values were analysed using linear mixed-effects models (LMMs) in the form:

$$\text{Expression}_{ij} \sim \text{Genotype} + (1|\,\text{Clone}) + (1|\,\text{Experiment})$$

Residuals were inspected to confirm model assumptions. Pairwise comparisons were performed using estimated marginal means (EMMs) with multiple-testing correction applied using Dunnett's contrasts.

Flow cytometry analysis. Fluorescence intensity distributions for PAX6 and *FOXG1* were analysed using GAMMs using the "mgcv" package, following logicle transformation. The relationship between signal intensity and experimental condition was modeled using a tensor-product smooth over centered and scaled scanning order (time_z) and forward-scatter width (fsc_w_z), with a random-effect smooth from clonal line identity:

$$\text{Intensity}_{ij} \sim \text{Genotype}_i + f_1\left(\text{time}_{z,ij}, \text{fsc\_w}_{z,ij}\right) + u_{\text{clone(j)}}$$

Here, $f_1(\cdot)$ denotes a tensor-product smooth with dimensions $k = \{10, 10\}$, and $u_{\text{clone}(j)}$ represents the random-effect smooth. Smoothing parameters were estimated using REML. Model diagnostics included inspection of residuals, concurvity assessment, and evaluation of smoothness selection. Pairwise comparisons were performed using EMMs, with multiple testing correction performed using Dunnett's contrasts.

CUT&RUN analysis

Read processing and peak calling was performed using a dedicated nextflow CUT&RUN pipeline available through nf-core[67] (https://doi.org/10.5281/zenodo.5653535). Briefly, raw reads were trimmed, QC filtered and aligned to the hg38 reference genome using Bowtie2 (v2.4.4)[57]. Upon Q-score filtering, duplicate removal and CPM normalization, peaks in target samples were called with SEACR (v1.3)[68] using corresponding IgG samples as controls. A consensus peak list per sample was obtained by requiring the presence of peaks in both replicates.

Reanalysis of single-cell RNA-seq data

Single-cell RNA-seq data of the developing human brain (Braun et al.[69]) was accessed through GitHub (https://github.com/linnarsson-lab/developing-human-brain) in h5ad format with CELLxGENE annotations (human_dev.h5ad). Single-cell RNA-seq data from the human neural organoid cell atlas HNOCA[37] were obtained from CELLxGENE (https://cellxgene.cziscience.com/collections/de379e5f-52d0-498c-9801-0f850823c847). Datasets were subset and processed in Python, and heatmaps were generated using the "seaborn" package heatmap function.

Reporting summary

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

Data availability

Some patient data included in Fig. 1 were collected from www.deciphergenomics.org (v11.27); details of the included cases are provided in Supplementary Data 1. Public epigenomics data included in Fig. 2 were collected from https://genome.ucsc.edu (ENCODE trackhub-DNAse-seq/Histone ChIP-seq (by target)/TF ChIP-seq (by target); Comparative Genomic – UCSC 100 Vertebrates), and VISTA enhancer data were collected from https://vista-enhancer.lbl.gov. Public data supporting the identification of candidate cis-regulatory elements were collected from the Annotated Regulatory Index [https://www.meuleman.org/research/dhsindex/] and ENCODE SCREEN v3 [https://screen.wenglab.org]. Single-cell RNA-seq data of the developing human brain[69] were accessed through GitHub [https://github.com/linnarsson-lab/developing-human-brain]. The single-cell RNA-seq data from the human neural organoid cell atlas HNOCA was sourced from CELLxGENE [https://cellxgene.cziscience.com/collections/de379e5f-52d0-498c-9801-0f850823c847]. All datasets generated in this study are available through the Gene Expression Omnibus (GEO) repository under accession numbers: GSE292880 (UMI-4C) and GSE292881 (CUT&RUN). The Hi-C datasets have been deposited in the European Genome-phenome Archive (EGA) under study accession number EGAS50000001768. The individual sets are available under accession numbers EGAD50000002535 (del8) and EGAD50000002536 (bca4). These data are under controlled access. Requests for access will be handled by the Center for Medical Genetics Ghent. Access will be granted upon request, provided that: a relevant scientific motivation for the use of the data is supplied; the proposed use is disease-specific research; the use is nonprofit and non-commercial; appropriate citation of our study and dataset is ensured in any resulting publication; a collaboration agreement is established where appropriate; and a data transfer agreement is signed prior to data sharing. Access can be requested through the EGA website. Source data are provided as a Source data file. Source data are provided with this paper.

References

1. Hou, P.-S., hAilín, D. Ó., Vogel, T. & Hanashima, C. Transcription and beyond: delineating *FOXG1* function in cortical development and disorders. *Front. Cell. Neurosci.* **14**, 35 (2020).

2. Chen, D. et al. Loss of Foxg1 impairs the development of cortical SST-interneurons leading to abnormal emotional and social behaviors. *Cereb. Cortex* **29**, 3666–3682 (2019). 1991.
3. Zhu, W. et al. Precisely controlling endogenous protein dosage in hPSCs and derivatives to model *FOXG1* syndrome. *Nat. Commun.* **10**, 928 (2019).
4. Mariani, J. et al. *FOXG1*-dependent dysregulation of GABA/glutamate neuron differentiation in autism spectrum disorders. *Cell* **162**, 375–390 (2015).
5. Won, H. et al. Chromosome conformation elucidates regulatory relationships in developing human brain. *Nature* **538**, 523–527 (2016).
6. Mitter, D. et al. *FOXG1* syndrome: genotype phenotype association in 84 patients with *FOXG1* variants. *Neuropediatrics* **48**, S1–S45 (2017).
7. Gold, W. A., Krishnarajy, R., Ellaway, C. & Christodoulou, J. Rett syndrome: a genetic update and clinical review focusing on comorbidities. *ACS Chem. Neurosci.* **9**, 167–176 (2018).
8. Kortum, F. et al. The core *FOXG1* syndrome phenotype consists of postnatal microcephaly, severe mental retardation, absent language, dyskinesia, and corpus callosum hypogenesis. *J. Med. Genet.* **48**, 396–406 (2011).
9. D'haene, E. & Vergult, S. Interpreting the impact of noncoding structural variation in neurodevelopmental disorders. *Genet. Med. Off. J. Am. Coll. Med. Genet.* **23**, 34–46 (2021).
10. Mehrjouy, M. M. et al. Regulatory variants of *FOXG1* in the context of its topological domain organisation. *Eur. J. Hum. Genet.* **26**, 186–196 (2018).
11. Poszewiecka, B. et al. TADeus2: a web server facilitating the clinical diagnosis by pathogenicity assessment of structural variations disarranging 3D chromatin structure. *Nucleic Acids Res.* **50**, W744–W752 (2022).
12. Lu, G. et al. Identification of a de novo mutation of the *FOXG1* gene and comprehensive analysis for molecular factors in Chinese *FOXG1*-related encephalopathies. *Front. Mol. Neurosci.* **15**, 1039990 (2022).
13. 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).
14. Pienkowski, V. M. et al. Mapping of breakpoints in balanced chromosomal translocations by shallow whole-genome sequencing points to EFNA5, BAHD1 and PPP2R5E as novel candidates for genes causing human Mendelian disorders. *J. Med. Genet.* **56**, 104–112 (2019).
15. Allou, L. et al. 14q12 and severe Rett-like phenotypes: new clinical insights and physical mapping of *FOXG1*-regulatory elements. *Eur. J. Hum. Genet.* **20**, 1216–1223 (2012).
16. Redin, C. et al. The genomic landscape of balanced cytogenetic abnormalities associated with human congenital anomalies. *Nat. Genet.* **49**, 36–45 (2016). 2016 491.
17. Craig, C. P. et al. Diagnosis of *FOXG1* syndrome caused by recurrent balanced chromosomal rearrangements: case study and literature review. *Mol. Cytogenet.* **13**, 40 (2020).
18. Ibn-Salem, J. et al. Deletions of chromosomal regulatory boundaries are associated with congenital disease. *Genome Biol.* **15**, 423 (2014).
19. Ellaway, C. J. et al. 14q12 microdeletions excluding *FOXG1* give rise to a congenital variant Rett syndrome-like phenotype. *Eur. J. Hum. Genet.* **21**, 522–527 (2013).
20. Takagi, M. et al. A 2.0 Mb microdeletion in proximal chromosome 14q12, involving regulatory elements of *FOXG1*, with the coding region of *FOXG1* being unaffected, results in severe developmental delay, microcephaly, and hypoplasia of the corpus callosum. *Eur. J. Med. Genet.* **56**, 526–528 (2013).
21. Shoichet, S. A. et al. Haploinsufficiency of novel FOXG1B variants in a patient with severe mental retardation, brain malformations and microcephaly. *Hum. Genet.* **117**, 536–544 (2005).
22. Alosi, D. et al. Dysregulation of *FOXG1* by ring chromosome 14. *Mol. Cytogenet.* **8**, 24 (2015).
23. Goubau, C. et al. Platelet defects in congenital variant of Rett syndrome patients with *FOXG1* mutations or reduced expression due to a position effect at 14q12. *Eur. J. Hum. Genet.* **21**, 1349–1355 (2013).
24. Firth, H. V. et al. DECIPHER: database of chromosomal imbalance and phenotype in humans using Ensembl resources. *Am. J. Hum. Genet.* **84**, 524–533 (2009).
25. Visel, A., Minovitsky, S., Dubchak, I. & Pennacchio, L. A. VISTA enhancer browser—a database of tissue-specific human enhancers. *Nucleic Acids Res.* **35**, D88–D92 (2007).
26. Spielmann, M., Lupiáñez, D. G. & Mundlos, S. Structural variation in the 3D genome. *Nat. Rev. Genet.* **19**, 453–467 (2018).
27. Sifrim, A. et al. Distinct genetic architectures for syndromic and nonsyndromic congenital heart defects identified by exome sequencing. *Nat. Genet.* **48**, 1060–1065 (2016).
28. Alter, S. et al. Telangiectasia-ectodermal dysplasia-brachydactyly-cardiac anomaly syndrome is caused by de novo mutations in protein kinase D1. *J. Med. Genet.* **58**, 415–421 (2021).
29. Sánchez-Gaya, V. & Rada-Iglesias, A. POSTRE: a tool to predict the pathological effects of human structural variants. *Nucleic Acids Res.* **51**, e54 (2023).
30. Schwartzman, O. et al. UMI-4C for quantitative and targeted chromosomal contact profiling. *Nat. Methods* **13**, 685–691 (2016).
31. Rao, S. S. P. et al. A 3D map of the human genome at kilobase resolution reveals principles of chromatin looping. *Cell* **159**, 1665–1680 (2014).
32. 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).
33. Meuleman, W. et al. Index and biological spectrum of human DNase I hypersensitive sites. *Nature* **584**, 244–251 (2020).
34. Moore, J. E. et al. Expanded encyclopaedias of DNA elements in the human and mouse genomes. *Nature* **583**, 699–710 (2020).
35. Dimitrieva, S. & Bucher, P. UCNEbase—a database of ultraconserved non-coding elements and genomic regulatory blocks. *Nucleic Acids Res.* **41**, D101–D109 (2013).
36. Moya, N., Cutts, J., Gaasterland, T., Willert, K. & Brafman, D. A. Endogenous WNT signaling regulates hPSC-derived neural progenitor cell heterogeneity and specifies their regional identity. stem. *Cell Rep.* **3**, 1015–1028 (2014).
37. He, Z. et al. An integrated transcriptomic cell atlas of human neural organoids. *Nature* **635**, 690–698 (2024).
38. Chambers, S. M. et al. Highly efficient neural conversion of human ES and iPS cells by dual inhibition of SMAD signaling. *Nat. Biotechnol.* **27**, 275–280 (2009).
39. Neaverson, A. et al. Differentiation of human induced pluripotent stem cells into cortical neural stem cells. *Front. Cell Dev. Biol.* **10**, 1023340 (2023).
40. Birnbaum, R. Y. et al. Coding exons function as tissue-specific enhancers of nearby genes. *Genome Res.* **22**, 1059–1068 (2012).
41. 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).
42. Manuel, M. et al. The transcription factor *Foxg1* regulates the competence of telencephalic cells to adopt subpallial fates in mice. *Dev. Camb. Engl.* **137**, 487–497 (2010).
43. Salehi, H. et al. Differentiation of human ES cell-derived neural progenitors to neuronal cells with regional specific identity by co-culturing of notochord and somite. *Stem Cell Res.* **8**, 120–133 (2012).

44. Salehi, H. et al. Neuronal induction and regional identity by co-culture of adherent human embryonic stem cells with chicken notochords and somites. *Int. J. Dev. Biol.* **55**, 321–326 (2011).
45. Kawauchi, S. et al. *Foxg1* promotes olfactory neurogenesis by antagonizing Gdf11. *Dev. Camb. Engl.* **136**, 1453–1464 (2009).
46. Yang, H. et al. SOXE group transcription factors regulates the expression of *FoxG1* during inner ear development. *Biochem. Biophys. Res. Commun.* **623**, 96–103 (2022).
47. Hettige, N. C. et al. *FOXG1* dose tunes cell proliferation dynamics in human forebrain progenitor cells. *Stem Cell Rep.* **17**, 475–488 (2022).
48. Chen, Z. et al. Increased enhancer–promoter interactions during developmental enhancer activation in mammals. *Nat. Genet.* **56**, 675–685 (2024).
49. Zhang, F. & Lupski, J. R. Non-coding genetic variants in human disease. *Hum. Mol. Genet.* **24**, R102–R110 (2015).
50. Baetens, D., Mendonça, B. B., Verdin, H., Cools, M. & De Baere, E. Non-coding variation in disorders of sex development. *Clin. Genet.* **91**, 163–172 (2017).
51. Liao, J. et al. Deletion of conserved non-coding sequences downstream from NKX2-1: A novel disease-causing mechanism for benign hereditary chorea. *Mol. Genet. Genom. Med.* **9**, e1647 (2021).
52. Benko, S. et al. Disruption of a long distance regulatory region upstream of SOX9 in isolated disorders of sex development. *J. Med. Genet.* **48**, 825–830 (2011).
53. Verdin, H. et al. Microhomology-mediated mechanisms underlie non-recurrent disease-causing microdeletions of the FOXL2 gene or its regulatory domain. *PLoS Genet.* **9**, e1003358 (2013).
54. Short, P. J. et al. De novo mutations in regulatory elements in neurodevelopmental disorders. *Nature* **555**, 611–616 (2018).
55. Zhang, Y. et al. Rapid single-step induction of functional neurons from human pluripotent stem cells. *Neuron* **78**, 785–798 (2013).
56. Mohajeri, K. et al. Transcriptional and functional consequences of alterations to MEF2C and its topological organization in neuronal models. *Am. J. Hum. Genet.* **109**, 2049 (2022).
57. Langmead, B. & Salzberg, S. L. Fast gapped-read alignment with Bowtie 2. *Nat. Methods* **9**, 357–359 (2012).
58. Raman, L., Dheedene, A., De Smet, M., Van Dorpe, J. & Menten, B. WisecondorX: improved copy number detection for routine shallow whole-genome sequencing. *Nucleic Acids Res.* **47**, 1605–1614 (2019).
59. Sante, T. et al. ViVar: a comprehensive platform for the analysis and visualization of structural genomic variation. *PLoS ONE* **9**, e113800 (2014).
60. D'haene, B., Vandesompele, J. & Hellemans, J. Accurate and objective copy number profiling using real-time quantitative PCR. *Methods* **50**, 262–270 (2010).
61. Osterwalder, M. et al. Characterization of mammalian in vivo enhancers using mouse transgenesis and CRISPR genome editing. in (ed. Dworkin, S.) *Craniofacial Development: Methods and Protocols* (Springer, 2022).
62. D'haene, E. et al. Comparative 3D genome analysis between neural retina and retinal pigment epithelium reveals differential cis-regulatory interactions at retinal disease loci. *Genome Biol.* **25**, 123 (2024).
63. Durand, N. C. et al. Juicer provides a one-click system for analyzing loop-resolution Hi-C experiments. *Cell Syst.* **3**, 95–98 (2016).
64. Li, H. & Durbin, R. Fast and accurate short read alignment with Burrows-Wheeler transform. *Bioinformatics* **25**, 1754–1760 (2009).
65. Open2C et al. Cooltools: enabling high-resolution Hi-C analysis in Python. *PLoS Comput. Biol.* **20**, e1012067 (2024).
66. Kruse, K., Hug, C. B. & Vaquerizas, J. M. FAN-C: a feature-rich framework for the analysis and visualisation of chromosome conformation capture data. *Genome Biol.* **21**, 303 (2020).
67. Ewels, P. A. et al. The nf-core framework for community-curated bioinformatics pipelines. *Nat. Biotechnol.* **38**, 276–278 (2020).
68. Meers, M. P., Tenenbaum, D. & Henikoff, S. Peak calling by sparse enrichment analysis for CUT&RUN chromatin profiling. *Epigenetics Chromatin* **12**, 42 (2019).
69. Braun, E. et al. Comprehensive cell atlas of the first-trimester developing human brain. *Science* **382**, eadf1226 (2023).
70. Hébert, J. M. & McConnell, S. K. Targeting of cre to the *Foxg1* (BF-1) locus mediates loxP recombination in the telencephalon and other developing head structures. *Dev. Biol.* **222**, 296–306 (2000).

Acknowledgements

We extend our gratitude to the participating individuals and their families for their contribution in this study. Crosslinked cell pellets from human NSCs and neurons were generously provided by Dr. Kiana Mohajeri-Stickels and Prof. Michael Talkowski. Hi-C interaction matrices from control LCLs were kindly provided by Dr. Uira Melo and Prof. Stefan Mundlos. We thank the Core Zebrafish Facility Ghent (ZFG) BOF/COR/2022/009, and Dr. Andy Willaert for their expert technical assistance with in vivo zebrafish enhancer assays. The E1b vector and the ZED vector were generously provided by Prof. Ramon Y. Birnbaum and Dr. Jose Bessa, respectively. We thank the technical team at the Lawrence Berkeley National Laboratory for generating transgenic mouse embryos. We also thank Maarten Verdonckt and Sophie Vermaut of the CORE Flow Cytometry BOF/COR/2022/001, at Ghent University, for their support and for access to the BD LSR Fortessa X-20 5-laser cytometer used for all flow cytometry experiments; and Eline Kees, M.Sc., for her contribution to the findings in patient bca4. This study makes use of data generated by the DECIPHER community. A full list of centers who contributed to the generation of the data is available from https://deciphergenomics.org/about/stats and via email from contact@deciphergenomics.org. DECIPHER is hosted by EMBL-EBI, and funding for the DECIPHER project was provided by the Wellcome Trust [grant number WT223718/Z/21/Z]. We would specifically like to thank Prof. Alessandra Renieri, Dr. Margherita Baldassarri, Dr. Chiara Fallerini, Dr. Edward Blair, Dr. Deborah Shears and Dr. Pernille M. Torring for their kind cooperation and for providing information on cases included in the DECIPHER database. Financial support has been provided by grants 1520518 N, G055422N and G063423N to S.V. from the Research Foundation-Flanders (FWO) and BOF.STG.2019.0103.01, BOF.BAF.2024.0263.01, and BOF.BAF.2025.0095.01 to S.V. from the Special Research Fund (BOF) from Ghent University. E.D. is supported by a postdoctoral grant from the Research Foundation Flanders (FWO 12D8523N) and was previously supported by Marguerite-Marie Delacroix and FWO doctoral fellowships. L.H., N.V.L. are supported by a doctoral fellowship from the Research Foundation Flanders (L.H.: FWO 11I2525N and N.V.L.: FWO 11A4C25N) and were previously supported by a Marguerite-Marie Delacroix doctoral fellowship. M.d.R.P.B. was, and L.C. is supported by a Marguerite-Marie Delacroix doctoral fellowship. A.V., M.K. were supported by U.S. National Institutes of Health (NIH) grant R01HG003988. Research conducted at the E.O. Lawrence Berkeley National Laboratory and performed under U.S. Department of Energy Contract DE-AC02-05CH11231, University of California (UC). L.L.P.R. received a PhD fellowship from the European Union's Horizon 2020 research and innovation programme under the Marie Sklodowska-Curie grant agreement number 101034290 (EMERALD International PhD Programme for Medical Doctors). This work was supported in part by funding from the Fund for Scientific Research-Flanders (FWO) (research grants G048220N and G0A2122N to A.J.), the Association Belge contre les Maladies neuro-Musculaires ASBL (ABMM) grants to A.J., the French Muscular Dystrophy Association (AFM-Telethon) grants 16179 and 23708 to A.J.; the Bulgarian National Recovery and Resilience Plan financed by the National Science Fund of Bulgaria (BNSF) BG-RRP-2.004-0004-C01 to A.J., D.M.A.-T.

Author contributions

E.D., B.M. and S.V. conceived the overall project. E.D., L.H., M.B.V. and S.V. designed the experiments. B.C., A.J., L.L.P.R., D.M.A.-T. and K.D.

examined the patients, provided biological samples and gathered clinical information. H.G. contributed additional clinical information. A.D., B.M. provided cytogenetic analysis and molecular diagnosis. E.D., L.H., M.B.V., N.V.L., M.d.R.P.B., S.L., L.G., E.D. and L.V. performed functional experimental assays. E.D., M.B.V., S.L. and L.C. performed data analyses. M.K., A.V. designed, performed and analysed mouse enhancer assays. E.D., L.H., M.B.V. and S.V. interpreted the results and wrote the manuscript. All authors reviewed and approved the final version of the manuscript.

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/s41467-026-73770-1.

Correspondence and requests for materials should be addressed to Sarah Vergult.

Peer review information *Nature Communications* thanks Niels Tommerup, Malte Spielmann and the other anonymous reviewer(s) for their contribution to the peer review of this work. [A peer review file is available].

[1]Center for Medical Genetics Ghent, Ghent University Hospital, Ghent, Belgium. [2]Department of Biomolecular Medicine, Ghent University, Ghent, Belgium. [3]Department of Human Structure and Repair, Ghent University, Ghent, Belgium. [4]Environmental Genomics & System Biology Division, Lawrence Berkeley National Laboratory, Berkeley, CA, USA. [5]Molecular Neurogenomics Group, VIB Center for Molecular Neurology, VIB, Antwerp, Belgium. [6]Department of Biomedical Sciences, University of Antwerp, Antwerp, Belgium. [7]Department of Clinical Genetics, University Pediatric Hospital, Medical University-Sofia, Sofia, Bulgaria. [8]Hunter Genetics, Newcastle, NSW, Australia. [9]Newcastle-School of Medicine and Public Health, Faculty of Health, University of Newcastle, Newcastle, NSW, Australia. [10]Center for Human Genetics, University Hospitals Leuven, Leuven, Belgium. [11]Molecular Medicine Center, Department of Medical Chemistry and Biochemistry, Medical University-Sofia, Sofia, Bulgaria. [12]School of Natural Sciences, University of California, Merced, CA, USA. [13]US Department of Energy Joint Genome Institute, Berkeley, CA, USA. [14]These authors contributed equally: Lisa Hamerlinck, Eva D'haene.
✉e-mail: Sarah.Vergult@UGent.be