

Bridging between mouse and human enhancer-promoter long-range interactions in neural stem cells, to understand enhancer function in neurodevelopmental disease

Romina D'Aurizio  ^{1,✉}, Orazio Catona², Mattia Pitasi³, Yang Eric Li  ⁴, Bing Ren  ⁴, and Silvia K. Nicolis  ³

¹Institute of Informatics and Telematics (IIT), National Research Council (CNR), Pisa, Italy; ²Department of Medical Biotechnologies, University of Siena, Siena, Italy;
³Dipartimento di Biotecnologie e Bioscienze, University of Milano-Bicocca, Milano, Italy; ⁴University of California San Diego, La Jolla, CA, USA

Non-coding variation in complex human disease has been well established by genome-wide association studies, and it is thought to involve regulatory elements, such as enhancers, whose variation affects the expression of the gene responsible for the disease. The regulatory elements often lie far from the gene they regulate, or within introns of genes differing from the regulated gene, making it difficult to identify the gene whose function is affected by a given enhancer variation.

Enhancers are connected to their target gene promoters via long-range physical interactions (loops). In our study, we re-mapped, onto the human genome, more than 10,000 enhancers connected to promoters via long-range interactions, that we had previously identified in mouse brain-derived neural stem cells by RNApolII-ChIA-PET analysis, coupled to ChIP-seq mapping of DNA/chromatin regions carrying epigenetic enhancer marks. These interactions are thought to be functionally relevant. We discovered, in the human genome, thousands of DNA regions syntenic with the interacting mouse DNA regions (enhancers+connected promoters). We further annotated these human regions regarding their overlap with sequence variants (single nucleotide polymorphisms, SNPs; copy number variants, CNVs), that were previously associated with neurodevelopmental disease in human. We document various cases in which the genetic variant, associated in human to neurodevelopmental disease, affects an enhancer involved in long-range interactions: SNPs, previously identified by genome-wide association studies to be associated with schizophrenia, bipolar disorder, and intelligence, are located within our human syntenic enhancers, and alter transcription factors recognition sites. Similarly, CNVs associated to autism spectrum disease and other neurodevelopmental disorders overlap with our human syntenic enhancers. Some of these enhancers are connected (in mouse) to homologs of genes already associated to the human disease, strengthening the hypothesis that the gene is indeed involved in the disease. Other enhancers are connected to genes not previously as-

sociated with the disease, pointing to their possible pathogenetic involvement.

Our observations provide a resource for further exploration of neural disease, in parallel with the now widespread genome-wide identification of DNA variants in patients with neural disease.

enhancers | non-coding DNA | genome-wide association studies (GWAS) | copy number variants (CNV) | long-range interactions | neural stem-cells | neurodevelopmental diseases | single-nucleotide polymorphisms (SNP)
Correspondence: romina.daurizio@iit.cnr.it

Introduction

The study of the genetic basis of human inherited disease has so far centered mostly on the study of protein-coding genes, with the identification of disease-causing mutations involving gene sequences (exons), or sequences involved in the splicing of primary transcripts into mature mRNA. The potential of non-coding DNA (representing 98% of our genome) to contribute to genetic disease has remained comparatively unexplored. Further, it was estimated that only 42% of patients with severe developmental disorders carry pathogenic de novo mutations within coding sequences (Short et al., 2018), further encouraging to consider the noncoding genome as an important territory to chart for mutations. Genome-wide association studies (GWAS) identified thousands of sequence variants located in non-coding regions, possibly corresponding to gene regulatory elements, which might be contributing to disease by affecting the regulated gene (Hindorff et al., 2009; Maurano et al., 2012). The noncoding genome is known to contain transcriptional regulatory DNA elements, enhancers, and promoters, fundamental to orchestrate regulated gene expression during development and differentiation. Some “historical” examples well demonstrate the ability of enhancers and promoters to contribute to genetic disease (e.g. the beta-globin genes LCR and gamma-globin regulatory regions for hemoglobinopathies; enhancers of the *Sox9* and *SHH* genes for sex-determination and mesodermal pathologies (Kleinjan and Van Heyningen, 2005)). However, difficulties in identifying functional enhancers within the noncoding genome have until recently re-

resented a hurdle in searching for their role in disease pathogenesis in a more generalized way (Turner and Eichler, 2019). The advent of next-generation sequencing (NGS)-based functional genomics represented a major leap forward in the identification of enhancers, and in their annotation to the gene(s) they regulate. ChIPseq, and, more recently, CUT&RUN, allow to map epigenetic enhancer marks, such as histone modifications (H3K27Ac, H3K4me1), together with transcription factor (TF)-binding sites, in chromatin at the genome-wide level (<https://www.encodeproject.org/>) (Meers et al., 2019; Pagin et al., 2021a,b). Chromatin Interaction Analysis by Paired-End Tagging (ChIA-PET), and other related methods, identify long-range interactions between gene promoters and distant DNA sequences, carrying epigenetic enhancer marks, together with bound TF. In particular, RNApolIIChIA-PET identifies long-range interactions in chromatin, involving RNApolII, the enzyme that transcribes genes. Remarkably, RNApolII ChIA-PET showed that, at the genome-wide level, a given enhancer is very often not connected to the nearest promoter, but rather to a more distant one(s), skipping genes in between (Zhang et al., 2013). These observations point to the need for experimental determination of long-range interaction maps, to correctly annotate a given enhancer to the promoter it interacts with (often far away).

Recently, we performed RNApolIIChIA-PET analysis of mouse brain-derived neural stem cells (NSC) cultured from the mouse perinatal (postnatal day 0, P0) brain, coupled to ChIPseq profiling of epigenetic enhancer marks (H3K27Ac; H3K4me1), binding sites for the SOX2 and FOS transcription factor, and RNAseq analysis of gene expression in normal versus *Sox2*-deleted NSC (Bertolini et al., 2019; Pagin et al., 2021a). This identified thousands of interactions connecting gene promoters to distal enhancers, carrying epigenetic enhancer marks (and TF binding sites). Remarkably, 14 out of 15 enhancers tested in transgenesis assays directed gene expression to the developing brain; further, genes downregulated following *Sox2* deletion were highly enriched in long-range interactions (LRI) with enhancers carrying epigenetic enhancer marks, and *SOX2* binding sites. These observations pointed to the functional relevance of the identified enhancer-promoter interactions for the gene regulation (Bertolini et al., 2019).

In the present paper, we discover, in the human genome, thousands of DNA regions syntenic with the interacting regions (enhancers+connected promoters) previously identified, in mouse, by RNApolII-ChIA-PET (Bertolini et al., 2019). We term these regions human-mouse syntenic long-range interaction regions (hmsLRI). Within hmsLRI, we define regions homologous to mouse regions carrying epigenetic enhancer marks (termed hmsLRI-Enhancers, hmsLRI-E). We find that hmsLRI-E are highly enriched (within their mouse coun-

terpart) in candidate cis-regulatory elements (cCRE), previously defined on the basis of their chromatin accessibility, in single-nuclei analyses, in a variety of adult brain cell types (Li et al., 2021); in agreement with this, most (95.5%) hmsLRI-E carry epigenetic enhancer marks also in human neural cells.

We further relate these regions to sequence variants (single nucleotide polymorphisms, SNP; copy number variations, CNV), that were previously associated with neurodevelopmental disease (NDD) in human. We find that those sequence variants associated with NDD are highly represented within hmsLRI. We document various cases in which the genetic variant, associated in human to NDD, affects an enhancer, or a promoter, involved in LRI. These observations provide novel hypotheses regarding the pathogenesis of the NDD associated with those variants. Further, we propose that hmsLRI may provide new hypotheses as to the possible location of new pathogenic sequence variants in NDD, to be experimentally tested, in the future (e.g. by targeted sequencing of enhancers connected to genes whose mutation, in exons, is known to cause NDD).

Results

Identification of human DNA regions syntenic with mouse enhancer-promoter long-range interactions, and of putative human enhancer sequences therein

We sought to identify, within the human genome, DNA regions syntenic with the long-range interactions (LRI) previously identified by RNApolII-ChIA-PET in mouse NSC, connecting gene promoters with distal “epigenetic enhancers”, i.e. regions marked by H3K4me1 and H3K27Ac (Creyghton et al., 2010; Rada-Iglesias et al., 2011)(Methods)(Fig. 1A). We first identified, within mouse LRI, the “non-promoter” interaction anchors (Bertolini et al., 2019) carrying epigenetic enhancer marks. “Anchors” are defined as DNA regions, identified by RNApolII-ChIA-PET to interact with distant DNA regions; “Promoter anchors” contain a gene promoter, “non-promoter” anchors do not (Bertolini et al., 2019). We distinguished between regions carrying both H3K27Ac and H3K4me1 (“active enhancers”) and regions carrying H3K4me1 only (“poised enhancers”, representing enhancers that were proposed to become fully active following cell differentiation) (Creyghton et al., 2010; Rada-Iglesias et al., 2011) (Fig. 1B; Methods). We reciprocally mapped mouse enhancers and promoters of each LRI to their syntenic human regions by the Liftover tool (see Methods); we termed these human regions human-mouse syntenic long-range interaction regions (“hmsLRI”). We further distinguished within the hmsLRI the sequence homologous to the mouse epigenetic enhancer sequence (i.e. carrying active or poised enhancer marks); we term these sequences Enhancers within hmsLRI (hmsLRI-E) and hmsLRI-E-a (active) or hmsLRI-E-p (poised) according to the epigenetic marks (Fig. 1B; Table 1). Promot-

ers, as they are more conserved, showed a higher percentage of continuous overlap than enhancers, as the latter regions most often encompass intergenic and intragenic regions so they are less likely to be completely conserved in humans (Fig. S1). Overall, out of 10,994 mouse LRI originally identified in (Bertolini et al., 2019) (7554 involving active enhancers, and 3440 involving poised enhancers), 7698 LRI could be re-mapped onto the human genome regarding both the promoter and the enhancer (Fig. 1B, “P+E”), with more than 50% of bases “lifted over” to the human genome; these represent about 70% (72.7% for active and 64.1% for poised, respectively) of the LRI originally identified in mouse (Table 1, Table S1, Fig. 1C). These human regions (promoter+enhancer, P+E), corresponding to the mouse LRI via Liftover, are the hmsLRI (Table 1, Fig. 1B). Interestingly, for 33 mouse LRI (26 involving active and 7 involving poised enhancers), both enhancer and promoter could be re-mapped, but, in human, the corresponding sequences were located on two distinct chromosomes (interchromosomal LRI in Table 1, Fig. 1B, “split P and E”) implying a probable disruption of the mouse 3D chromatin interaction.

We also identified 493 enhancers (358 active, 135 poised), involved in LRI in mouse, which “lifted over” to the human genome, but for which the corresponding promoter anchor in human could not be identified as part of a syntenic region (Table 1, Fig. 1B, “E only”). On the other hand, for a total of 2308 mouse LRI, only the promoter sequences were conserved in a syntenic region of the human genome (Table 1, Fig. 1B, “P only”). These cases (E only, P only) may result from loss, in evolution, of the promoter or enhancer sequences present in the mouse (and presumably in a common ancestor), or from the gain of enhancers or promoters in the mouse, but not in human; they may, however, also result from the splitting, in human, of the single enhancer/promoter anchor sequence present in mouse, with the two split parts ending up on different chromosomes (or on distant parts of the same chromosome), thereby failing to reach the 50% overlap required to qualify as re-mapped by the chosen Liftover parameters. In the present work, we shall focus our attention on those 7698 LRI for which the mouse enhancer and interacting promoter could both be re-mapped in the human genome, and were part of a syntenic region, hence they were located on the same chromosome in both mouse and human (listed in Table S1). As expected, the number of LRI conserving the promoter regions exceeded those for which only the enhancer was conserved between the two species.

The hmsLRI-E showed significantly higher levels of sequence conservation among 100 vertebrate genomes as compared to random shuffled genomic regions (Fig. 2A; Methods). Each identified hmsLRI regions were annotated by comparison with the gene coordinates. To annotate hmsLRI-P, priority was given to check the distance from any transcription start site (TSS) within a

window of 5 kb centered in the TSS for every gene (Methods). Most of hmsLRI-E were intronic or intergenic (Fig. 2B). A total of 4635 hmsLRI-E were intragenic (i.e. intronic or exonic) and are reported in Table S1. Of note, around 50% (3797) of the syntenic enhancer-promoter interactions spanned multiple genes or the enhancers were in the intron of another gene. The full list of the identified hmsLRI and associated gene annotation is reported in Table S1, and a summary is reported in Table 1. We also created, for the WashU human genome browser, a track depicting the location of the identified hmsLRI, and hmsLRI-E, with hmsLRI represented by loops.

HmsLRI and hmsLRI-E overlap with DNA regions previously shown to be involved in long-range interactions, and to carry epigenetic enhancer marks, in human neural cells

Further, we asked if enhancer-promoter connections observed in mouse neural cells are also maintained within their hmsLRI counterparts. For this, we compared hmsLRI with 5 previously published human datasets of genome-wide long-range interactions, obtained mainly using HiC, and H3K4me3-mediated chromatin interactions to histone-marked active promoter regions, by PLAC-seq assays (Table 2). We found that, cumulatively, up to 51% of hmsLRI were represented within human experimentally detected long-range interactions. Notably, the number of interactions varied a lot between the different datasets, and a large proportion of hmsLRI identified in a given dataset were not identified in one or more of the other datasets, implying that the published HiC/PLAQ-seq data do not exhaustively represent the totality of interactions. This might reflect technical differences between different HiC experiments, as well as the fact that the source of the neural tissue/cells is quite diverse in the reported datasets, and some interactions may therefore be present in only a subset of the analyses reported. Other possible explanations are presented in Discussion. We conclude from the analysis in Table 2 that the cumulative representation of hmsLRI (51%) within published interactions datasets is likely a minimum estimate, that may increase when additional tissues/cell types are analyzed in more detail. Taking this result together with the extremely high conservation of epigenetic enhancer characteristics of hmsLRI-E in human neural cells, we are now confident that our present data may be useful for further studies.

Further NDD-related functional annotation of hmsLRI-E

To further complement the identification of hmsLRI-E with useful annotations and supporting evidence about the potential functional role of those regions specifically in the NDD context, we also checked hmsLRI-E overlaps with VISTA enhancers. The VISTA enhancer browser (<https://enhancer.lbl.gov/>) (Visel et al., 2009) collects experimentally validated noncoding

Long-Range Interaction (LRI)	Involving	Involving	Total
	Active enhancer	Poised enhancer	
Mouse Enhancer-Promoter LRI analyzed	7554 (769)	3440 (173)	10994
Human syntenic LRI identified in human (promoter+enhancer) ("P+E" in Fig. 1B)	5492 (501)	2206 (98)	7698
Only enhancer of mouse LRI identified in human ("E only" in Fig. 1B)	358	135	493
Only promoter of mouse LRI identified in human ("P only" in Fig. 1B)	1355	953	2308
Inter-chromosomal LRI identified in human ("split P and E" in Fig. 1B)	26	7	33
Mouse LRI (enhancer, promoter) not represented in human	323	139	462

Table 1. Human DNA regions identified by their synteny with mouse enhancer-promoter long-range interactions previously identified by RNApolIIChIA-PET

Total number of mouse LRI connecting promoter and enhancer anchors that were checked for syntenic conservation into the human genome (-minMatch=0.5 for LiftOver). "Involving Active enhancers" and "Involving Poised enhancers" report interactions in which the connected enhancer carries active and/or poised epigenetic enhancer marks (H3K27A, H3K4me1; see Fig. 1). Numbers in brackets report the subset of the indicated LRIs which are redundant, i.e. those sharing more than 50% of length for both promoter and enhancer anchors.

fragments with gene enhancer activity in human and mouse; in our previous work, we had found that VISTA enhancers are significantly represented within ChIA-PET-identified mouse enhancers (Bertolini et al., 2019). Here, we looked for evidence of confirmed gene enhancer activities also for hmsLRI-E regions. We found a total of 212 VISTA positive enhancers among the hmsLRI-E, of those 182 were active enhancers and 30 were poised enhancers (see Table S1).

Furthermore, we exploited the Gene2Phenotype databases (<http://www.ebi.ac.uk/gene2phenotype>) to check for the presence of high-confidence neurodevelopmental (brain) and eye disorder genes (DD panel and Eye panel in the dataset, respectively) among those involved in hmsLRI. We also annotated our hmsLRI for their overlap with the DDG2P dataset (<https://www.deciphergenomics.org/ddd/ddgenes>).

DDG2P includes genes with curated annotation of disease association level of certainty that the gene causes a developmental disease (e.g. definitive, strong, moderate...), known consequences of pathogenic mutations in the gene, as well as information about the tissues affected by gene mutations (Methods). In

Table S1, we annotated each human gene involved in hmsLRI, connected with the enhancer or the promoter anchor, if it was reported in the DDG2P database and specified the level of disease association. Regarding the Gene2Phenotype database, a total of 321 over 1474 genes involved in brain, cognition and eye disease, and 112 over 667 genes involved specifically in eye disease, were found involved in hmsLRI (see Table S1).

Mouse enhancers and hmsLRI-E are enriched in brain candidate cis-regulatory elements (cCRE)

In a recent study, Li et al. (Li et al., 2021) identified a set of candidate cis-regulatory elements (cCRE) in the mouse genome, by combining single-nucleus assays for transposase-accessible chromatin (sn-ATAC-seq) and single-cell RNA sequencing (scRNA-seq), from more than 800,000 cells from 45 different regions of the adult mouse brain. They identified 491,818 cCRE within 160 distinct cell types. We compared our mouse enhancers, involved in LRI, to mouse cCRE. We found that more than 95% of our mouse enhancers (10501 over 10994) overlap with cCRE (97.6% of the active enhancers, 7370 over 7554, and 91% of the poised ones, 3131

Figure 1

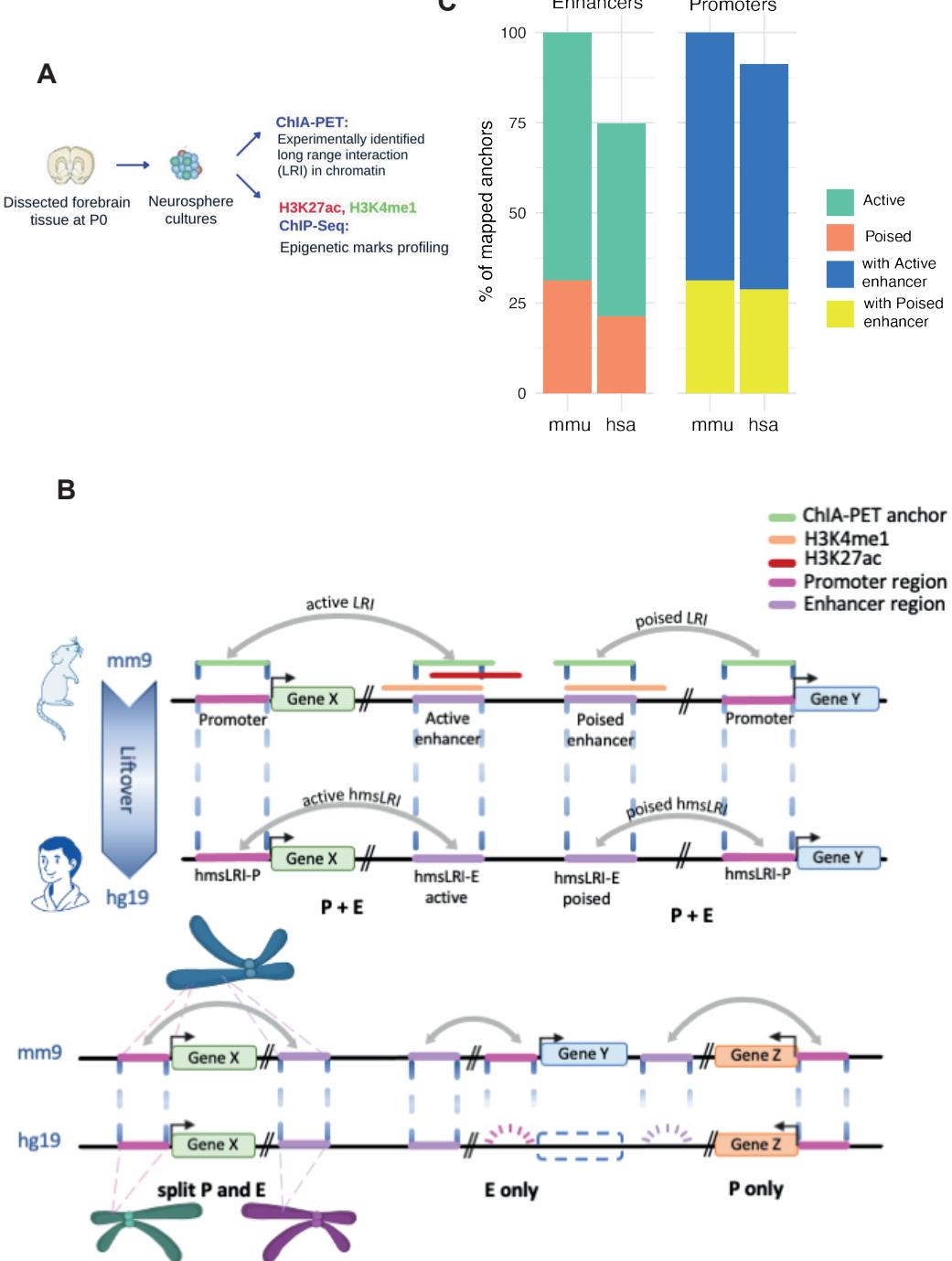


Figure 1. Identification of human DNA regions syntenic with mouse enhancer-promoter long-range interactions (LRI)

(A) Long-range interactions (LRI) identified by RNApolII ChIA-PET in neural stem cells (NSC) cultured in vitro from mouse neonatal forebrain, and regions carrying H3K4me1 and H3K27Ac histone modifications marks as identified by ChIP-Seq were obtained from (Bertolini et al. 2019).

(B) Graphical representation of the syntenic mapping between mouse (mm9) and human (hg19) reference genomes. Mouse LRI "anchors" that have syntenic correspondences in the human genome are indicated as hmsLRI-P (promoter) or hmsLRI-E (enhancer).

(C) The percentages of mouse promoter and enhancer anchors that "lifted over" to the human genome (imposing a 50% minimal overlap, see Methods).

Figure 2

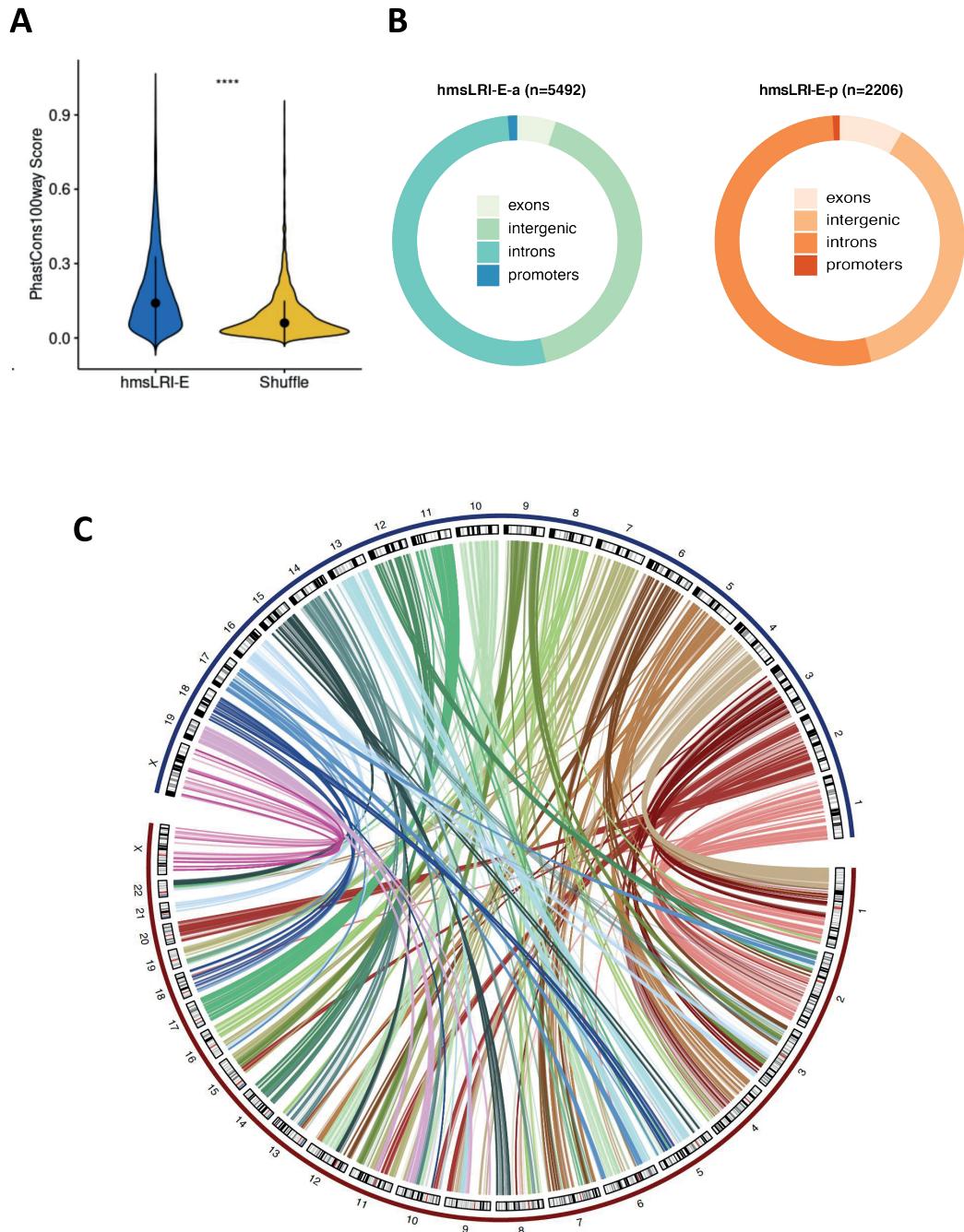


Figure 2. Conservation and synteny between mouse and human LRI-E and LRI

(A) - Violin plots showing sequence conservation of hmsLRI-E, versus random shuffled genomic regions with similar GC content (**** $P < 0.0001$, Wilcoxon rank-sum test). PhastCons score over 100 vertebrates is reported. Lines span the first to third quartiles (Q1 to Q3), dots denote the median.

(B) Ring chart with the fraction of hmsLRI-E (active on left and poised on right) overlapping exonic, intergenic, intronic or promoter regions in the human genome.

(C) - Circus plot showing the genomic correspondences of enhancers involved in LRI between mouse and human. Mouse chromosomes are represented in blue (chr1-19,X), human chromosomes are in red (chr1-22,X).

Figure 3

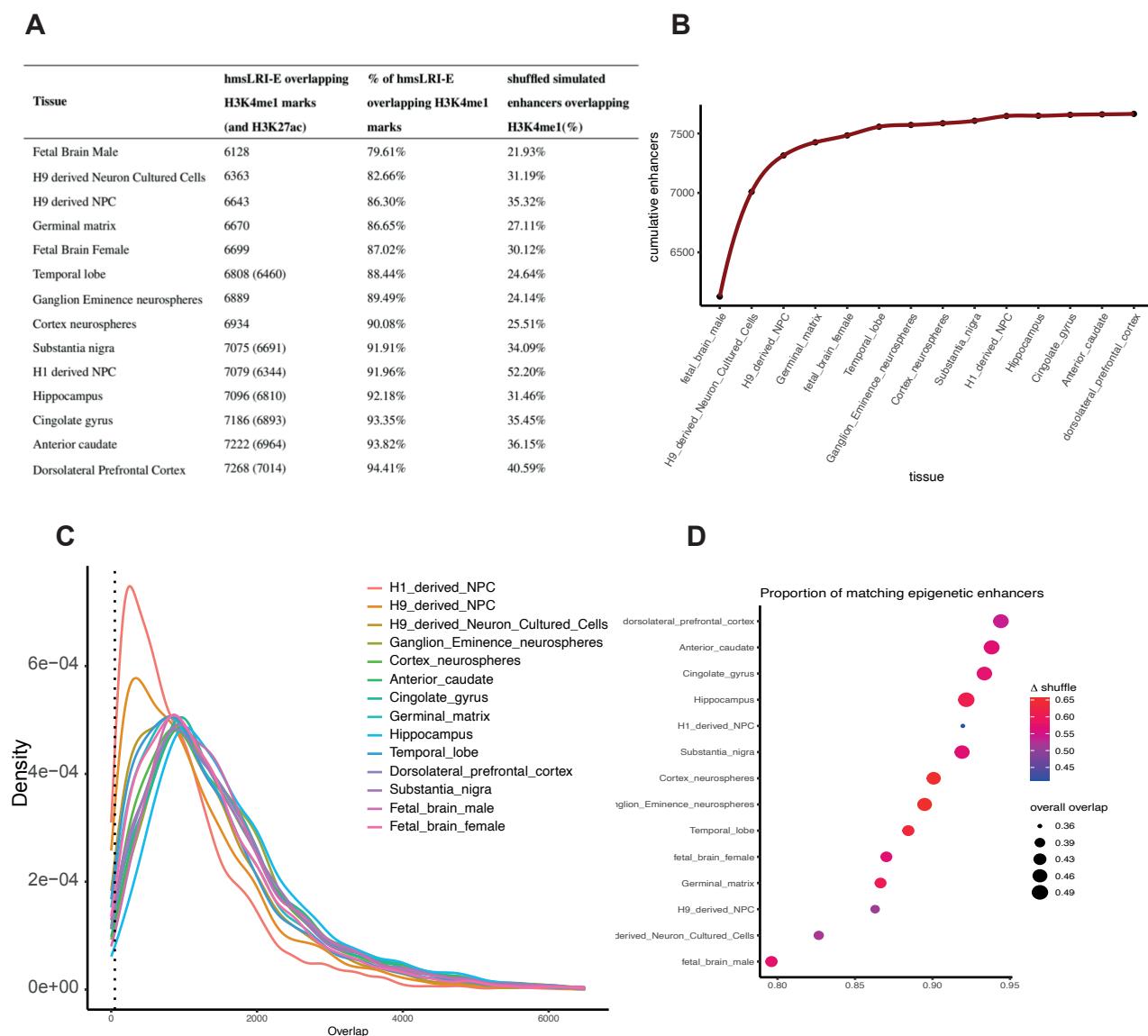


Figure 3. Concordance between hmsLRI-E and human DNA regions carrying epigenetic enhancer marks (H3K4me1; H3K27ac) from Roadmap Epigenomics (A) - For each human neural tissue, the number of hmsLRI-E encompassing enhancer marks (H3K4me1; H3K27ac where available) and their relative percentage are shown. The values are compared with the % of shuffled enhancer-like regions overlapping H3K4me1 marks.

(B) - Cumulative number of matched hmsLRI-E in all the tissues. On the x-axis, tissues are listed from the one with the lowest number of matching hmsLRI-E to the one with the highest number. The y-axis shows the cumulative number of unique hmsLRI-E that progressively match the epigenetic enhancer marks of the indicated tissue tissues (the percentage for each tissue is added to the ones already found for the previous ones).

(C) - Distribution of the extent of overlap (in base pairs, X axis) between hmsLRI-E and human regions carrying epigenetic enhancer marks in human cells according to the Roadmap Epigenomic database ("Density" in the Y axis). The dotted line at 50 bp represents the cutoff below which the hmsLRI-E and the enhancer marks were not considered matched.

(D) - Dot plot representing, on the X axis, the proportion of hmsLRI-E overlapping human DNA regions (min 50bp) carrying epigenetic enhancer marks in human neural cells ("matching epigenetic enhancers"), as previously defined in the Roadmap Epigenomic database, in the various tissues indicated on the Y axis. The width of the dots represents the proportion of bp covered by the methylation region out of the total length of the matched hmsLRI-E. Colour is dependent on the difference (delta) between the match of the hmsLRI-E with H3K4me1-marked regions and the match of the simulated shuffled enhancers, in terms of proportion.

over 3440, respectively). Overall, they spanned all the 160 cell types with different levels of occurrence ranging

from 12% to 78%. Accordingly, all 43 subclasses were present with variable fractions of enhancers overlapping

Cells Origin	N of hmsLRI represented	% of hmsLRI represented	Study
Hi-C derived CP and GZ of fetal brain	1531	20.67%	(Won et al., 2016)
Hi-C interactions for DPC, NPC, Hippocampus	1893	24.59%	(Jung et al., 2019)
HiC-derived enhancer-promoter linkages of PFC, TC, CB	627	8.14%	(Wang et al., 2018)
HiC loops associated with promoter regions of PFC, TC, CB	788	10.24%	(Wang et al., 2018)
PLAC-seq derived interactions of RG, IPC, EN, and IN from mgHC	2214	28.76%	(Song et al., 2020)
TOTAL	3950	51.31%	

Table 2. Human putative interactions in this study (hmsLRI) that are found within experimentally determined human interactions datasets in neural cells
The numbers of hmsLRI represented within the indicated datasets of interactions are indicated, together with the percentage of total hmsLRI (7698, Table 1) represented by these numbers. § interactions represented within more than one study were counted only once. CP, cortical plates; GZ, germinal zone; DPC, dorsolateral prefrontal cortex; NPC, neural progenitor cells; PFC, prefrontal cortex; TC, temporal cortex; CB, cerebellum; RG, radial glia; IPC, intermediate progenitor cell; EN, excitatory neurons; IN, interneurons; mgHC, mid-gestational human cortex.

with cCRE (12% to 76%, Fig. 4).

Remarkably, the most enriched cell types encompass Radial Glia-like Stem Cells (RGL, corresponding to the identity of our forebrain-derived cultured NSC), and other proliferating neural precursors (such as oligodendrocyte progenitor cells, OPC), but also many differentiated cell types within the in vivo brain (Fig. 4). Interestingly, the most represented cell types include various types of GABAergic neuroblasts and neurons, as well as astrocytes (ASC) and oligodendrocytes (OGC) (Fig. 4), which correspond to the cell types that our NSC generate upon in vitro differentiation ([Gritti et al., 1996](#); [Bertolini et al., 2019](#); [Pagin et al., 2021a,b](#)). Glutamatergic neurons, though present, are less represented (Fig. 4).

For 7451 over 10501 mouse enhancers overlapping a cCRE, we could also associate by synteny a hmsLRI-E. In particular, we observed that almost all of our hmsLRI-E (97%, 7451 over 7698) overlaid with one or more cCREs, providing an additional evidence of their potential regulative role also in human. We annotated the hmsLRI-E within Table S1 by reporting the original cCRE ID (as in Table S8 of the Li et al. paper) of the cCREs overlapping the human LRI-E identified by

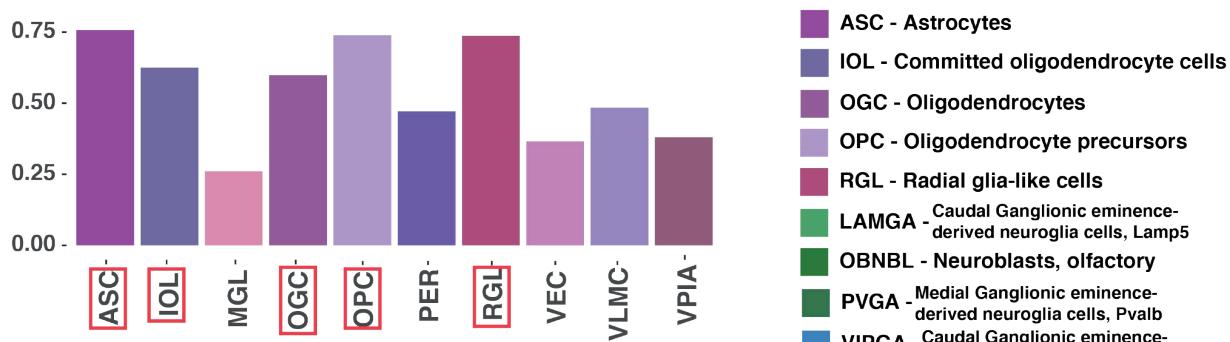
synteny with the mouse enhancers involved in LRI.

hmsLRI-E regions overlap with noncoding risk variants associated with NDD

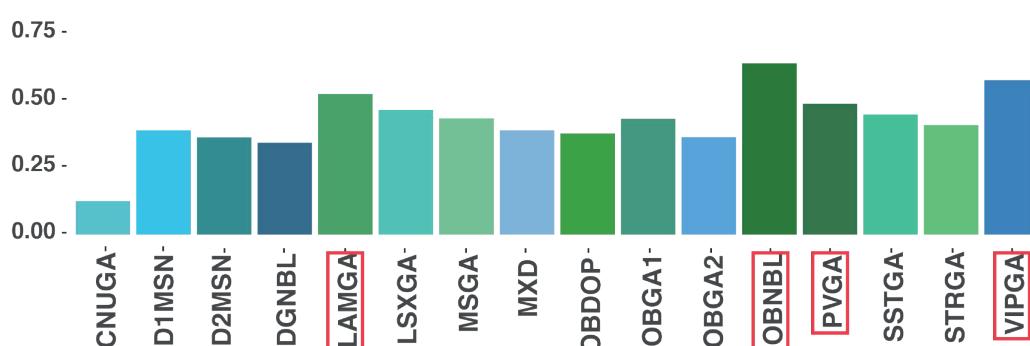
We next sought to relate the identified hmsLRI-E to sequence variants previously found in human by genome-wide association studies (GWAS) to be associated with neurological diseases and traits. Most such variants are located in the noncoding part of the genome, often lacking functional annotation, making it difficult to interpret their meaning ([Turner and Eichler, 2019](#)). To test whether our interaction maps could help the interpretation of some of these variants, we checked their presence within our hmsLRI-E to provide a way to functionally connect those variants to a specific gene promoter, via the long-range interaction. We thus searched our hmsLRI-E for the GWAS-derived variants with posterior probabilities of association (PPA) score larger than 1% with schizophrenia (SCZ), bipolar disorder (BP), and intelligence (I). Those potential causal variants were derived from the previous analyses by ([Li et al., 2021](#)). In all three datasets, we identified variants overlapping hmsLRI-E (Tables 3, S2 for details).

Figure 4

1.00 - Non-neurons, including neural stem/progenitor cells



GABAergic



Glutamatergic

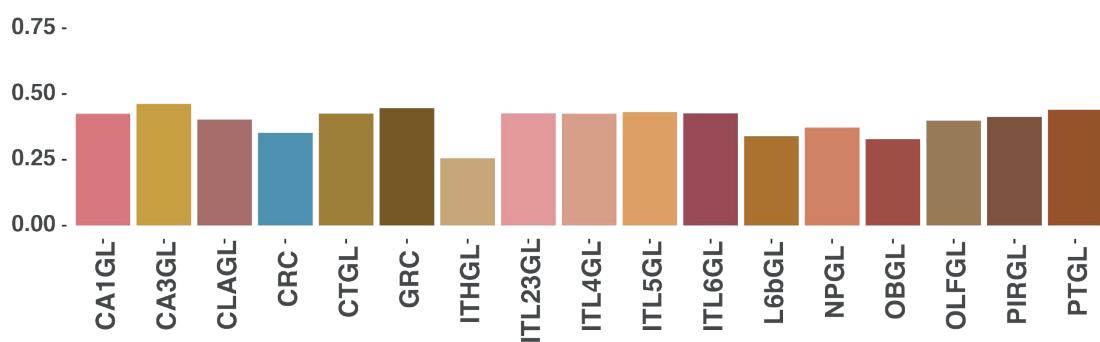


Figure 4. cCREs are highly represented within mouse enhancers involved in LRIs

Histograms represent the relative proportion of cCREs (Li et al. 2021) overlapping our mouse enhancers connected to promoters via LRI, for each of the 43 cell type subclasses (each of which is constituted by several different cell types, see Li et al. 2021). Red rectangles highlight subclasses with more than 50% of cCREs overlapping our mouse enhancers. Top, non-neuronal cell types, including neural stem/progenitor cells, grouped into 10 subclasses. Middle, GABAergic cell types grouped into 16 subclasses. Bottom, glutamatergic cell types grouped into 17 subclasses. For each subclass, mean values for all the different cell types constituting each subclass are reported. The cell type subclasses are defined in (Li et al. 2021, Table S8).

Significantly, 26 risk variants for schizophrenia were located in 31 hmsLRI-E regions and most of them were connected with different genes previously proposed to contribute to the pathology (*TRIM8*, *SREBF2*, *LINC00461*, *ETF1*, *CTNND1*, *PITPNM2*, *TCF4*, *FOXP1*, *NSD3*, Table 3, Table S2 for details). This suggests that the variant itself, or a closely associated mutation, may contribute to the disease by affecting

the transcription of the connected gene. In particular, a cluster of five different risk variants was within five different hmsLRI-Es (three of them were intergenic and two overlaid the 3'UTR of the *ARL3* gene) which were connected to the promoter of the *TRIM8* gene (Fig. 5). They may all be involved in the transcriptional deregulation of this gene, a proposed contributing factor to schizophrenia (Ding et al., 2021). Interestingly,

Gene	Risk variants overlapping hmsLRI-E	Number hmsLRI involved	Gene previously associated with the disease	Associated to SCZ in PsychENCODE dataset*	Gene product	Reference
SCHIZOPHRENIA						
TRIM8	5	5	Yes (by association studies)	No	Tripartite Motif Protein (E3 Ubiquitin-protein ligase?)	(Ding et al., 2021)
KIZ	1	1	Yes (by association with SNP)	No	Centrosomal protein	(Sheikh et al., 2018) (Consortium, 2018)
SREBF2	1	1	Yes (by association with intronic variant SNP)	No	Transcription factor	(Le Hellard et al., 2010)
LINC00461	1	1	Yes (by association studies)	No	Long noncoding RNA	(Liu et al., 2020)
ETF1	2	1	Yes (by association studies)	No	Translation termination factor	(Lam et al., 2019) (Zhao et al., 2021)
CTNND1	1	1	Yes (by association studies)	No	Catenin (cell adhesion, signal trasduction)	(Liu et al., 2021)
PITPNM2-AS1	1	1	Yes (PITPNM2) (by association studies)	Yes	Antisense RNA	(Liu et al., 2021)
CDK2AP1	2	1	Yes (by association studies)	Yes	Cyclin-dep kinase-assoc protein	(Liu et al., 2020)
EPN2	2	3	No	Yes	Eprin (clathrin interactor)	
MAPK7	1 (shared with EPN2)	1	No	No	Mitogen-activated protein kinase	
B9D1	1 (shared with EPN2)	1	No	No	B9-domain-cont protein (ciliogenesis)	
TCF4	4	4	Yes (by association with intronic variant SNP)	No	Transcription factor	(Gao et al., 2020)
FOXP1	1	1	Yes (by association studies)	Yes	Transcription factor	(Autism Spectrum, 2017)
LETM2	2	1	No	Yes	Transmembrane protein	
NSD3	2	1	Yes (by association studies)	Yes	Histone H3 methyltransferase, nuclear receptor binding protein	(Ma et al., 2013)
H2AC20	1	1	No	No	Replication-dependent histone	
BIPOLAR DISORDER						
NR2C2AP	2	1	No		Nuclear receptor associated protein	
MPP2	1	1	No		Membrane protein; synaptic function	
LSM12	2	1	No		Receptor channel	

Table 3. Sequence variants associated with Schizophrenia and Bipolar disorder overlap with human putative enhancers, involved in human-mouse syntenic long-range interactions (hmsLRI-E, see Fig. 1, Table 1)

"Gene" (first column) is the gene whose promoter is involved in the human-mouse syntenic long-range interaction (hmsLRI); the sequence variants are located within the connected hmsLRI-E.

We searched our hmsLRI-E for the GWAS-derived variants with posterior probabilities of association (PPA) score larger than 1% with schizophrenia and bipolar disorder; those potential causal variants were derived from the previous analyses by (Li et al. 2021, Ref. 10)(see their Table S25)

PsychENCODE association of genes to schizophrenia was by HiC connection between the variant-containing DNA region and the gene promoter, and by eQTL data (see Wang et al. 2018, Ref. 11, and (<http://resource.psychencode.org/>)

other variants were found in hmsLRI-E connected to genes that had not been previously associated with schizophrenia (*KIZ*, *CDKAP2*, *EPN2*, *MAPK7*, *B9D1*, Table 3). It is possible that these variants (or closely associated sequence variants) contribute to the disease by deregulation of these connected genes, pointing to their possible contribution to the pathology. Of note, among the hmsLRI-E that contain schizophrenia-associated risk variants (Table 3), 17 also overlapped with the human counterpart of cCREs (Tables 3, S2); this indicates that, in mouse, the enhancer involved in the LRI is in an open configuration in adult brain cell types. Interestingly, these enhancers were associated with 6 different genes by our LRI.

In the Bipolar disorder dataset, four risk variants were found to overlap with hmsLRI-E which were connected to the *NR2C2AP* gene (two variants), the *MPP2* gene (one variant), and the *LMS12* gene (two variants) (Table 3).

In the Intelligence dataset (collecting variants that segregate with different levels of intelligence), 38 risk variants were into 45 enhancer regions connected to 25 different genes (Table S2), interestingly including *FOXP1* (<https://www.deciphergenomics.org/gene/FOXP1/overview/clinical-info>) (already involved in language impairment and autism), *JARID2* (<https://www.deciphergenomics.org/gene/JARID2/overview/clinical-info>), *PTCH1* (<https://www.deciphergenomics.org/gene/PTCH1/overview/clinical-info>), and others (Table 3).

Overall, our data suggest that the identified variants are associated to enhancers, that may regulate the gene to which they are connected via LRI. Whether the sequence variant itself causes deregulation of the connected gene, or whether this is caused by a different mutation in close linkage with it, remains to be determined. To begin to investigate the possible pathogenic meaning of some of the variants overlapping our enhancers, we tested whether the variants overlapping the *TRIM8*-connected enhancers affect potential transcription factor binding sites, as reported in the ENCODE database of TF DNA-binding motifs. In Fig. 5B, we show that sequence variants overlapping with three different hmsLRI-E alter the recognition sequence for TF on DNA; in two cases the mutation is predicted to lead to loss or reduction of TF binding, in one case to lead to the creation of a TF binding site. These findings provide new, testable hypotheses regarding the functional effect of GWAS-derived sequence variants overlapping our enhancers.

CNVs associated with NDD involve hmsLRI and can disrupt enhancer-promoter interactions

We then considered clinically relevant Copy Number Variants (CNV) identified in NDD cohorts with the aim of bringing out a possible causal connection with the loss/gain of a copy of a DNA region involving hmsLRI.

Indeed, there are reports of neural disease caused by genomic rearrangements affecting the regulation of specific genes, without altering their coding regions; it is possible that these rearrangements disrupt gene connections to their enhancers, or the enhancers themselves (see (Boland et al., 2007; Allou et al., 2012; Ellaway et al., 2013; Takagi et al., 2013; Bidart et al., 2017) in relation to *FOXG1*, *Bcl11a*, *Arid1A*, *AKT3*). We analyzed two datasets of CNVs associated with NDD (Coe et al., 2014; Zarrei et al., 2019) for possible overlaps with our hmsLRI, and in particular with hmsLRI-E. The genome-wide catalogue of rare CNV generated by Zarrei et al. includes the genotypes of 2691 subjects affected by autism spectrum disorder (1838 ASD), attention deficit hyperactivity disorder (424 ADHD), schizophrenia (204 SCZ), and obsessive-compulsive disorder (222 OCD) encompassing 55256 copy number variants. Similarly, the dataset by Coe et al includes CNVs associated with neurocognitive disorders, developmental delay and/or ASD for a total of 29085 primarily pediatric cases in comparison with 19584 adult population controls. We asked about the representation of our hmsLRI within the CNVs harbored by affected cases in both datasets (Zarrei, Coe) and found a statistically significant enrichment ($p < 0.0001$, according to the hypergeometric test) for both, when compared to the control cohort in Coe's study, in agreement with hmsLRI containing genes relevant for the pathology, together with their relevant enhancers. The full list of hmsLRI showing altered copy number (loss or gain) in Zarrei and Coe's datasets is reported in Table S3. AnnotSV has been used to annotate the pathogenicity of every CNV (Table S3).

A particularly interesting class of CNVs is that in which the involved DNA region encompasses the enhancer (hmsLRI-E), but not the promoter connected to it (Table S3). Thus, we focused our attention on those alterations which were also present exclusively among the affected subjects compared with the controls. An example of such CNV (Fig. 6A) consists of deletions involving a noncoding DNA region 3' to the *FOXG1* gene, but not the *FOXG1* gene itself. These deletions were found in two different patients, affected by ASD forms with characteristics in common to the disease caused by mutations (CNV) affecting the *FOXG1* gene itself. Interestingly, we find that two such deletions overlap an hmsLRI-E, showing evolutionary conservation through multiple species as reflected by a high PhastCons score; further, the corresponding mouse enhancer overlaps cCREs (see above) active in various adult brain cell types (Fig. 6A). We note that other hmsLRI are identified for *FOXG1*, together with their associated hmsLRI-E (Fig. 6A), defining further candidates to investigate for the presence of pathogenic mutations in ASD in the future.

Further interesting cases include the ones spanning over hmsLRI-E connected to the promoter of the *TRIO* and *PHF21A* genes; both of these genes are

Figure 5

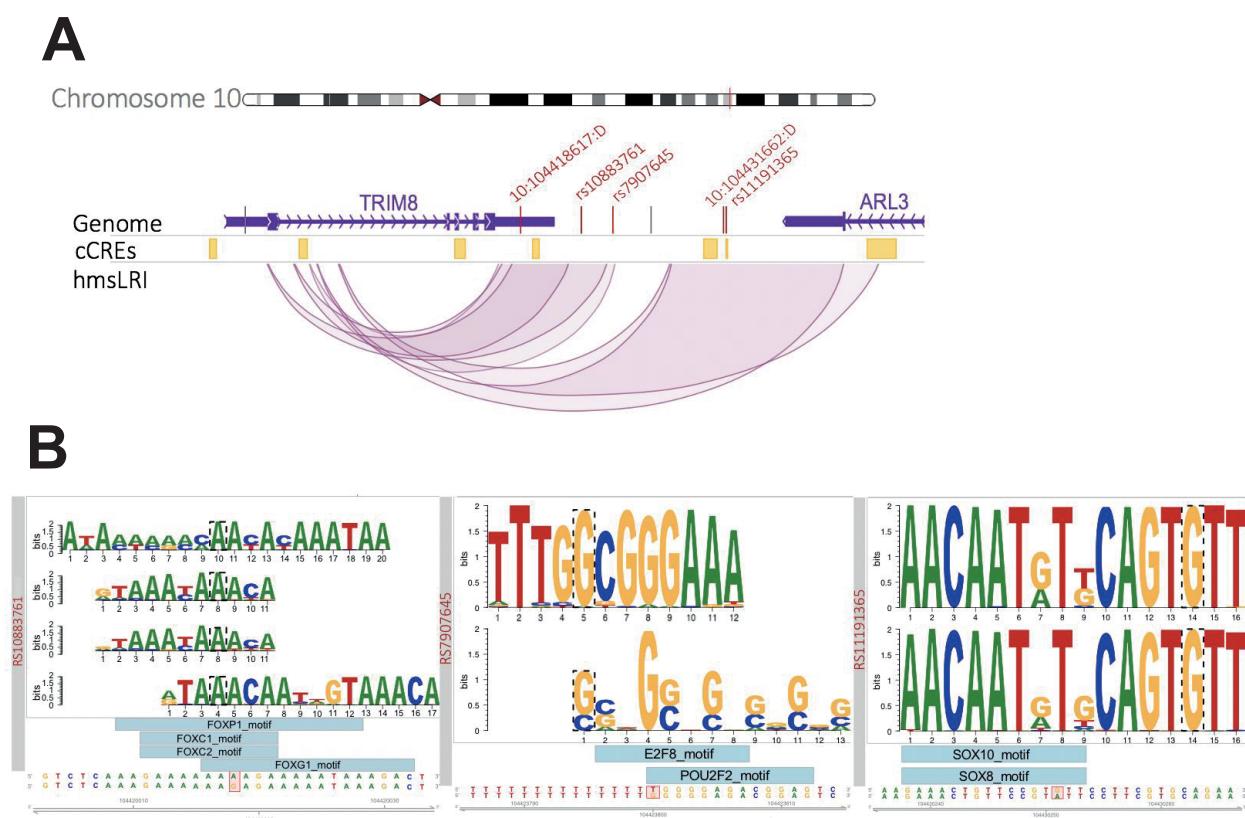


Figure 5. (A) GWAS-identified risk variants associated with schizophrenia overlap with five different hmsLRI-E connected with the TRIM8 gene. Visualization of the human TRIM8-ARL3 locus, located on 10q24.32. In the middle track, in yellow, cCREs identified by (Li et al. 2021) lifted over to the human genome. In the lower track, our hmsLRI: enhancers are represented in their full length, while promoter as their middle point. In the top genome track, vertical bars represent SNPs associated to schizophrenia according to (Li et al. 2021); the ones showed in red fall into one of our hmsLRI-E connected to the TRIM8 promoter.

(B) Neural TF binding site alteration by the risk variants. Alteration by the risk variants of putative neural TF binding sites, as predicted by motifbreakR (Coetzee et al. 2015). From below upwards: reference DNA sequence and SNP alteration, highlighted in red; light blue boxes represent motifs of transcription factors whose binding is predicted to be strongly influenced by the SNP; detail of the binding motif, retrieved from ENCODE. The motif nucleotide affected by the SNP is surrounded by a dotted line; if the nucleotide is the same as the reference genome, the binding site is predicted to be lost, otherwise it is gained: rs10883761 and rs11191365 disrupt potential binding sites of the TF represented, while rs7907645 leads to the acquisition of the indicated potential novel TF binding sites.

linked to NDD according to the DDG2P database. In particular, one hmsLRI-E (active wTR2_9924 chr11:46315791-46318951, Table S3), connected via long-range interaction to *PHF21A*, overlaps two 50kb deletions characterized in SCZ patients reported in the Zarrei database (Fig. 6B, Table S3). *PHF21A* is part of a transcriptional complex regulating neuronal genes expression and its mutations or deletions have been associated with NDD (Hamanaka et al., 2019); intriguingly, the two deletions encompass part of *CREB3L1*, a gene which is, itself, not linked to NDD, and not highly expressed in neural cells according to the Human Protein Atlas Database (Thul and Lindskog,

2018). These observations suggest that the observed deletions within the region including *CREB3L1* might contribute to the NDD phenotype mainly by causing the loss of the *PHF21A*-connected enhancer, hence the transcriptional deregulation of *PHF21A*.

Regarding the *TRIO* gene, the two connected hmsLRI-E (Table S3) are intergenic and altered by distinct CNVs reported in the Zarrei dataset. Specifically, a 23kb deletion encompassing the first enhancer region (active wTR1_2401 chr5:14118146-14120275 and active wTR2_4867 chr5:14118458-14119337, Table S3), and not any coding region, is present in one

Figure 6

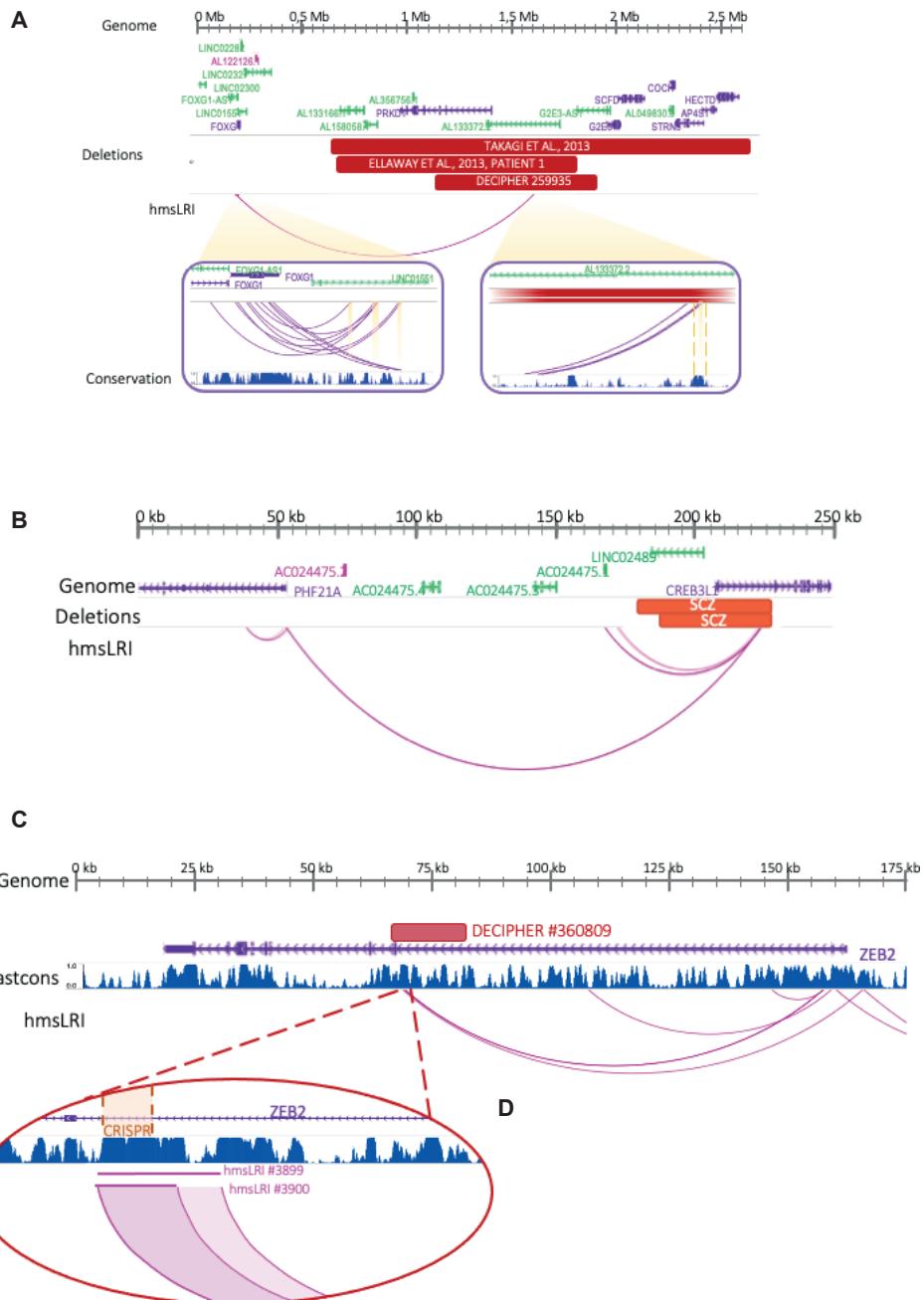


Figure 6. CNV involving hmsLRI-E, but not the interacting gene promoter, suggest that deregulation of the (intact) connected gene may be involved in pathogenesis

(A) CNV (microdeletions) identified in NDD patients involve an hmsLRI-E connected to the *FOGX1* promoter. The deletions track represents CNVs harbored by neurodevelopmental delay and Rett-like syndrome patients reported by (Ellaway et al. 2013; Takagi et al. 2013) and DECIPHER patient #259935. Magnification of distal and proximal hmsLRI is depicted in the purple boxes; the conservation track shown herein refers to the UCSC genome browser PhastCons score: distal hmsLRI fall into a highly conserved noncoding region. Highlighted hmsLRI overlap with cCREs.

(B) hmsLRI spanning the *PHF21A-CREB3L* locus. Active hmsLRI (wTR2_9924) connecting the promoter of *PHF21A* to an enhancer region located in an intron of *CREB3L*, overlaps two deletions reported in patients affected by schizophrenia (red horizontal rectangles marked SCZ).

(C) A CNV (14kb microdeletion, red horizontal box: DECIPHER #360809: chr2:145186591-145201317) affects two hmsLRI-E (wTR1_3899, chr2:145188059-145190658, and wTR1_3900, chr2:145188059-145189733). Phastcons score refers to the UCSC track.

(D) Magnification of the two hmsLRI. hmsLRI 3900 lies within an extremely conserved region; a notable portion of it (chr2:145188149-145189110), highlighted in orange, has been deleted via CRISPR by (Bar Yaacov et al. 2019), proving its role into *ZEB2* transcriptional regulation.

SCZ patient; the second *TRIO* enhancer (poised wTR3_2113 chr5:13976963-13977535, Table S3),

altered in two ASD siblings carrying a duplication (130kb), is located in an intergenic region (Fig. S3).

The *TRIO* gene encodes a GEF (Guanine Exchange Factor) that promotes axon guidance and neurite outgrowth (Deinhardt et al., 2011); being *Rho* and *Rac* its downstream effector, it plays a key role in p75^{NTR} complex, regulating and coordinating actin remodelling by its GEF activity in hippocampal neurons, within which it is localized in actin-rich protrusions (Barbosa et al., 2020). A recent study including individuals with pathogenic variants of *TRIO* gene highlighted a good genotype-phenotype correlation: missense mutations in the *RAC1*-activating GEF domain cause mild intellectual disability and microcephaly, missense mutations in the seventh spectrin repeat at the N-terminus underlie a phenotype characterized by severe intellectual disability and macrocephaly, while nonsense mutations in the *TRIO* sequence give rise to a more variable phenotype; interestingly, 31% of patients with pathogenic variants of the *TRIO* gene showed autistic traits (Sadybekov et al., 2017). A further study suggests that rare variants in the neurotrophin signaling pathway, in which p75 and *TRIO* are key components, are implicated in schizophrenia risk (Sadybekov et al., 2017; Prem et al., 2020). These pathways are important in cell migration, a key neurodevelopmental process altered in many NDD, including ASD and SCZ (Prem et al., 2020); as the disruption of the enhancer-promoter connectivity caused by CNVs presumably leads to a transcriptional deregulation of *TRIO*, the entire canonical cell migration pathway might be affected, hence providing a link between the CNVs and the NDD phenotype observed.

A further example of interesting hmsLRI-E overlapping microdeletions is represented by two intronic enhancers of *ZEB2* (i.e. hmsLRI-E wTR1_3899, including wTR1_3900, Fig. 6C), a gene whose mutations cause Mowat-Wilson Syndrome, characterized by intellectual disability and microcephaly (Mowat et al., 2003) and recently implicated in human-specific neurodevelopmental processes (Benito-Kwiecinski et al., 2021). An intronic microdeletion (14kb) overlapping these hmsLRI-E (chr2:145186591-145201317) is reported in the DECIPHER database, in an individual with intellectual disability (DECIPHER #360809). The deleted hmsLRI-Es overlaps a VISTA enhancer, active in the developing CNS (VISTA hs407, https://enhancer.lbl.gov/cgi-bin/imagedb3.pl?form=presentation&show=1&experiment_id=407&organism_id=1). The deletion, via CRISPR, of 961bp encompassing about two thirds of hmsLRI-E wTR1_3900 (Fig. 6D), was reported to significantly reduce *ZEB2* transcription in human HEK293 cells (Bar Yaacov et al., 2019), suggesting that loss of the hmsLRI-Es, overlapping the microdeletion, can contribute to the pathology by the downregulation of the connected *ZEB2* gene.

Mouse-human syntenic long-range interacting enhancers (hmsLRI-E) provide novel candidate sequences to search for mutations that may contribute to the pathogenesis of neurodevelopmental disease (NDD)

The degree to which specific NDDs can be explained by mutations within the exome vary greatly for different NDDs. In eye genetic disorders (Microphthalmia/Anophthalmia/Coloboma-Optic Nerve Hypoplasia, MAC-ONH), only between 40 and 50% of cases are explained by mutations within the exomes of several genes, including *SOX2*, *OTX2* and *PAX6* among the most frequently mutated (Williamson and FitzPatrick, 2014). Mutations or CNV/rearrangements involving enhancers may thus represent additional causes of disease.

Indeed, in pioneer studies on the genetic basis of eye disease, translocations were found in patients affected by aniridia (Kleinjan and Van Heyningen, 2005), involving a DNA region which is located, in the normal (wild type) genome, about 150.000 nucleotides 3' to the *PAX6* gene, whose mutation within the protein-coding exons was already known to cause aniridia. Molecular genetic analysis of aniridia patients bearing CNVs identified a 243.9kb critical region for *PAX6* transcriptional activation (Ansari et al., 2016), subsequently demonstrated to contain transcriptional enhancers, located within introns of the *Eip4* gene, but in fact controlling expression of *PAX6* (Kleinjan and Van Heyningen, 2005; Franzoni et al., 2017). One of our hmsLRI-E, (wTR2_9990; chr11:31579239-31580312) corresponding to a previously reported neural enhancer located in an intron of *ELP4* (EnhancerAtlas ID HS042-07410), completely lies within this critical region (Fig. 7A); other hmsLRI-E are connected to *Pax6*, located both 3' and 5' to *PAX6* (Fig. 7B). These sequences may be tested for the presence of mutations (CNV, structural variants or point mutations) in patients affected by NDD, involving the eye.

Our hmsLRI identified about 70 human genomic regions, harboring enhancers (hmsLRI-E), that are connected to eye disease-relevant genes in long-range interaction maps (Table S4-A). In particular, the *SOX2* (Fig. 8A), *CHD7* (Fig. 8B) and *PAX6* (Fig. 7B) genes display numerous long-range interactions connecting them to putative enhancers (Table S4-A). Further, this set of genes harbors several intronic epigenetic enhancers in mouse, that are conserved in humans; in turn, the enhancers are connected to other gene promoters (Table S4-B). The LRI perspective proposes these enhancers as candidates for targeted sequencing in those patients with genetic eye defects (MAC-ONH), that did not yet receive a molecular diagnosis based on exome sequencing.

Figure 7

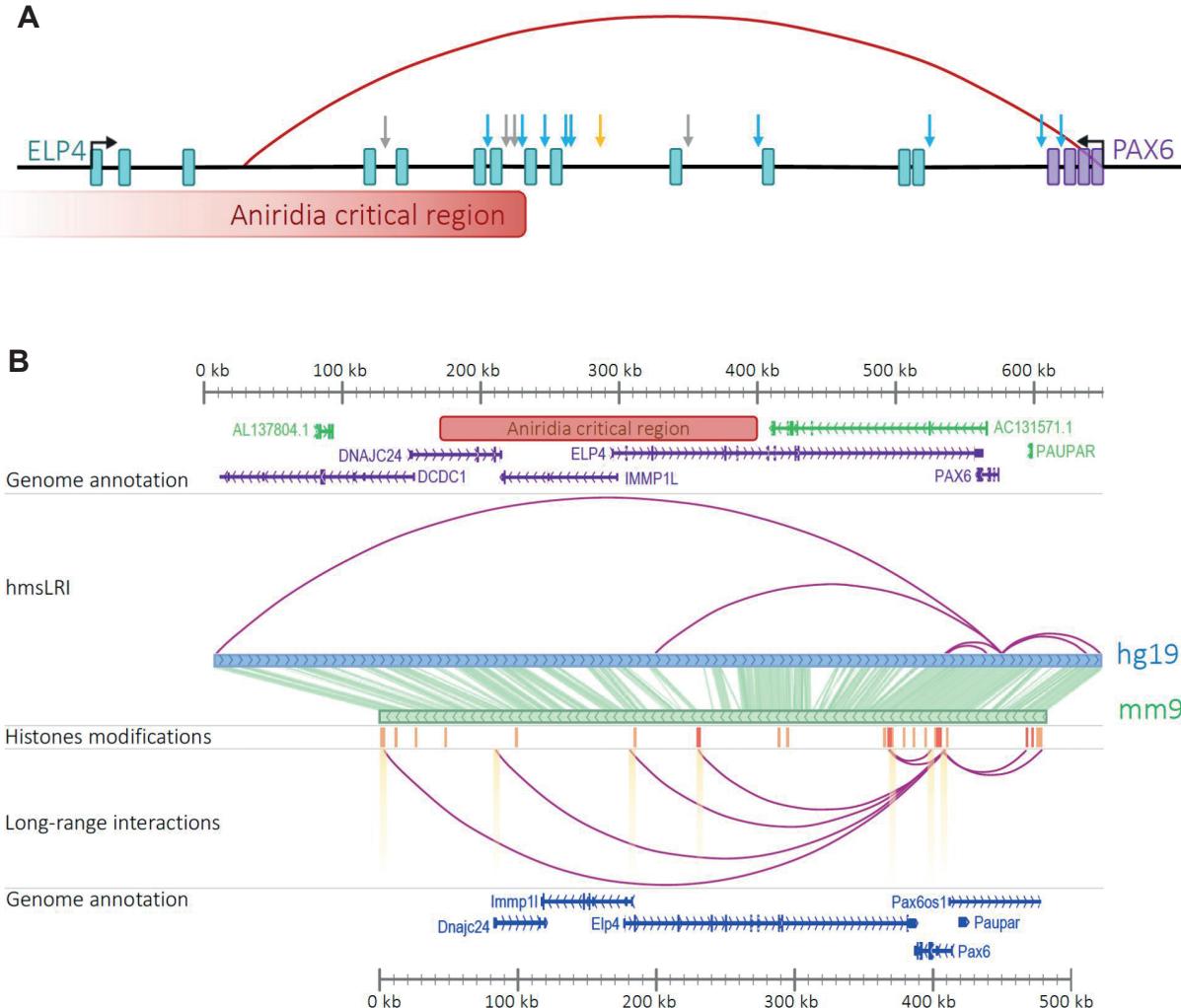


Figure 7. LRI connect PAX6 to multiple distal enhancers in the aniridia critical region (A) Blue arrows represent enhancers active in ocular structures, grey ones enhancers active in other brain structures; in orange, the SIMO element identified by (Kleinjan et al. 2001; Kleinjan and van Heyningen 2005b); the red arch connects PAX6 promoter to an hmsLRI-E. The aniridia critical region, chr11:31422424-31666340, encompasses our enhancer. (B) LRI involving the Pax6 promoter reported in (Bertolini et al. 2019); the track “histone modifications” shows regions carrying marks of poised (orange, H3K4me1) and active enhancers (red, both H3K4me1 and H3K27ac); highlighted LRI overlap with cCREs reported in (Li et al. 2021). In the upper part, long range interactions, involving active or poised enhancers, present among our hmsLRI-E. Highlighted enhancers and promoters overlap with cCREs reported (Li et al. 2021). The middle panel is a representation of corresponding regions between mouse (mm9) and human (hg19) genome.

Discussion

About 98% of our genome does not encode for proteins. Within this noncoding DNA, transcriptional regulatory regions (enhancers) are scattered, often at a great distance from the gene they regulate, and/or within introns of different genes. Indeed, this was the case for around half of the identified enhancer-promoter interactions. Furthermore, identifying the functional connections between enhancers and the genes they regulate will be of great help to understand genome function,

in normal development, genetic disease, and evolution. Our study, by re-mapping onto the human genome more than 10,000 enhancer-promoter LRI identified in mouse NSC by RNApolIIChIA-PET, provides a new resource and contribution to these investigations.

The vast majority of mouse enhancer-promoter LRI have corresponding syntenic regions in the human genome

In our work, we re-mapped 10,994 “epigenetic” enhancers (marked by H3K4me1/H3K27Ac enhancer

Figure 8

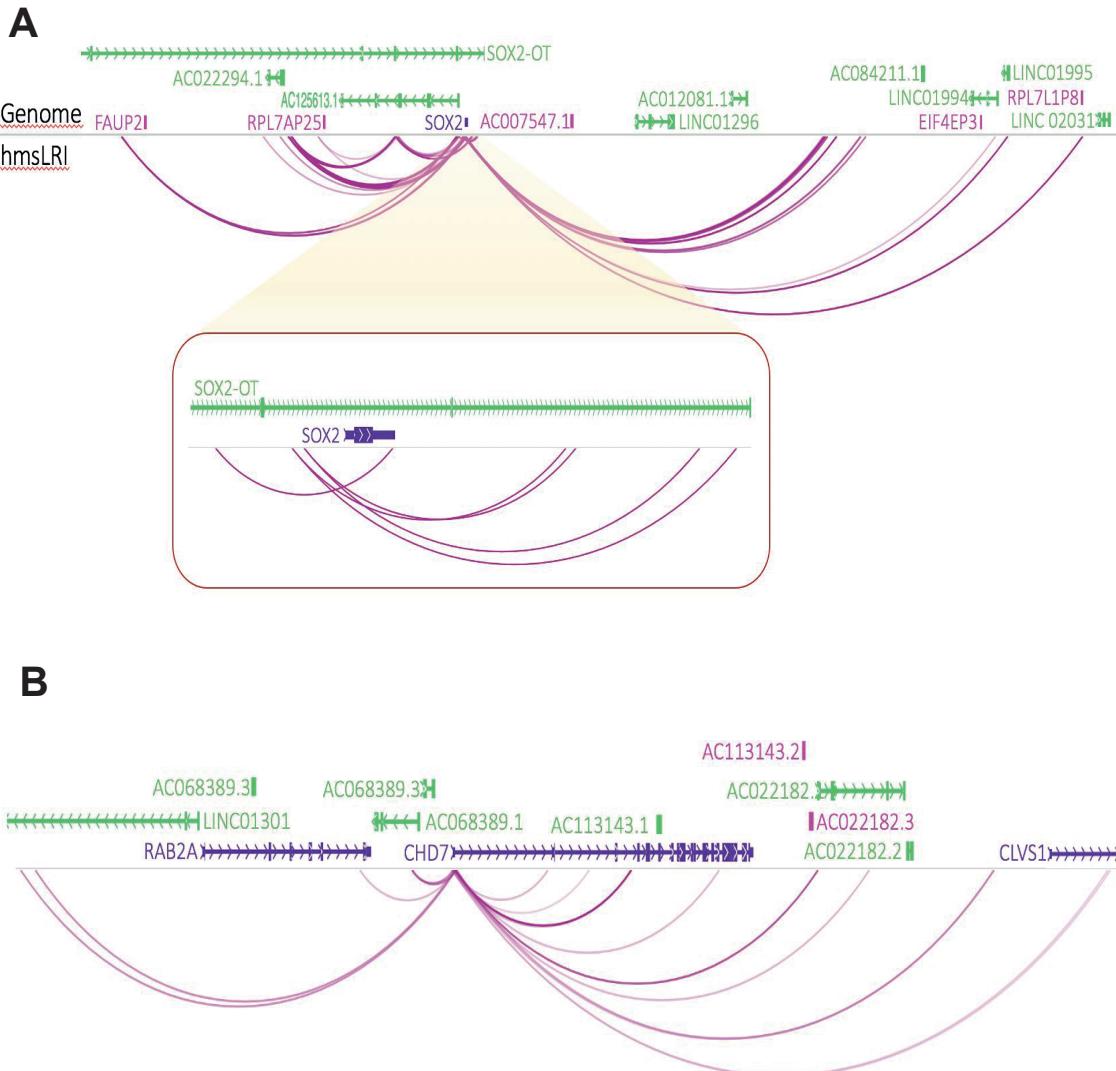


Figure 8. SOX2 and CHD7 promoters show intensive connectivity to distal enhancers The color intensity of arches is proportional to the number of interactions detected (see PET counts in Table S1) (A) SOX2 locus. The red box shows a magnification of the more proximal SOX2 region. (B) CHD7 locus.

marks), discovered in mouse NSC through their RNApolII-mediated interaction with gene promoters (Bertolini et al., 2019), onto the human genome. For most of them (7698), we found a syntenic region in the human genome, encompassing both the enhancer and the promoter connected to that enhancer in mouse (Fig. 1, Table 1 and S1, Fig. 2), despite the extensive “reshuffling” of genomic DNA sequences observed between mouse and human (Fig. 2B). Importantly, almost all hmsLRI-E carry epigenetic enhancer marks in human neural cells (Table 2), as they do in mouse. Moreover, a substantial proportion (over 50%) of hmsLRI are experimentally detected in HiC or PLAC-seq published datasets. As mentioned in Results, the failure to attain a higher representation may depend on the variety of cell types and tissues analyzed in those datasets, which would detect only a subset of the total population of interactions existing in the very diverse complex of human neural cells. An additional point is likely related to the technical differences between HiC

and RNApolII-ChIA-PET we originally used in our studies in mouse cells. HiC identifies all possible physical interactions within chromatin, giving origin to very large numbers of detected interactions which might play a great variety of functions, not all of which are necessarily related to gene transcription. In contrast, RNApolII-ChIA-PET selects those interactions that are mediated by active RNApolII, and thus identifies interactions that are related to transcription; these interactions represent a smaller subset than those obtained by HiC. It is possible that, in the presence of a very high number of relatively abundant HiC interactions, many of the RNApolII-ChIA-PET-identified interactions may escape detection. In this context, we analyzed the data reported in four HiC analyses (the same ones as those indicated in Table 2), asking which percentage of interactions detected in a given dataset is also identified in the sum of the other three datasets. The comparison (see Suppl. Fig. S4) clearly show that the total overlap of PsychENCODE long range interactions was less

than 53%, indicating that HiC data do not exhaustively represent the full repertoire of long-range interactions in a complex tissue, like brain, comprising hundreds of different cell types. This conclusion is further supported by data in Table 2: if we assume that the cumulative 51% of hmsLRI that find a counterpart in published HiC datasets are real, as they are confirmed by independent analyses, it is clear that each of the five HiC datasets detects a very small proportion of the cumulative real long range interactions, again confirming the conclusions about the incomplete representation of interactions within published HiC datasets discussed above. These observations highlight the importance of combining multiple independent studies with various techniques (HiC, PLAC-seq, ChIA-PET etc.), and of defined cell types, to approach a complete description of the repertoire of enhancer-promoter interactions in neural cells. On the other hand, for 33 LRI identified in mouse, the enhancer and the promoter remapped, in the human genome, onto different chromosomes, and for 493 enhancers, that interact with a promoter in mouse, a syntenic enhancer was remapped in the human genome, but no syntenic promoter was identified (Fig. 1, Table 1). We propose that these enhancers, though conserved in terms of sequence, may have changed their usage in evolution, possibly activating different genes in human and mouse. Vice versa, for 2308 gene promoters, connected to enhancers in mouse, we fail to detect a syntenic enhancer in human (Fig. 1, Table 1); also in this case, the loss of connected enhancers in human, or the acquisition of novel enhancers in mouse, may be the source of functionally relevant changes in gene activity. Loss or gain of enhancers have been proposed to be an important source of evolutionary change in vertebrates, a hypothesis validated by a few, though meaningful, functional studies (Indjeian et al., 2016; Long et al., 2016; Moriano and Boeckx, 2020), and our synteny maps may help develop novel hypotheses in this field.

Human-mouse syntenic enhancers involved in long-range interactions (hmsLRI-E) are highly enriched in candidate cis-regulatory elements accessible in the adult mouse brain

For the vast majority (about 97%) of the hmsLRI-E identified by our analysis, the mouse syntenic enhancers overlap with candidate cis-regulatory elements (cCRE), identified in mouse by their accessible chromatin state in single-nuclei assays performed on a variety of adult mouse brain cell types (Li et al., 2021) (Fig. 4). This is in a way surprising, since our LRI have been mapped in undifferentiated neonatal NSC. On the other hand, the differentiated cell types, whose cCRE are most enriched within our hmsLRI-E, correspond to the cell types that our NSC generate upon *in vitro* differentiation: astrocytes, GABAergic neurons, oligodendrocytes (Fig. 4). Our findings support the hypothesis that “epigenetic” enhancers, and their LRI with promoters, can mark, in

stem cells, sequences that will be later used for gene activation in differentiated neurons and glia (Creyghton et al., 2010; Rada-Iglesias et al., 2011; Cruz-Molina et al., 2017; Bertolini et al., 2019; Wei et al., 2019). Further, the annotation of our hmsLRI-E with the overlapping cCRE (Table S1), together with the knowledge of the specific cell types in which the cCRE are accessible, will represent a useful resource for future studies aimed at understanding the function of these enhancers in normal development and in disease (see below).

hmsLRI contain sequence and copy number variations associated to brain genetic disease, providing potential insight into disease mechanisms

We discovered that non-coding DNA sequence variants, associated by GWAS to brain diseases and traits (schizophrenia; bipolar disorder; intelligence), overlap with some of our hmsLRI-E (Table 3 and S2, Fig. 5B). This allows to hypothesize that the variant itself, or a closely linked mutation, acts by altering enhancer activity onto the connected gene. This is important, because, as already mentioned, LRI connect enhancers to promoters which can be hundreds of thousands, even millions of base pairs away, and can lie in introns of genes different from the gene promoter they actually control (see Introduction). Indeed, our interaction maps allow us to hypothesize the implication, in schizophrenia and bipolar disorder, of the regulation of genes, connected to variant-containing enhancers, that had not been previously involved in this disease (Table 3). For some other hmsLRI-E, the connected gene had already been hypothesized to contribute to the disease (schizophrenia, Table 3) by association studies between disease and polymorphic variants; in two instances, the variants were located directly within the putative candidate gene, whereas in other ones the candidate gene was lying within a broad locus implicated by GWAS (Table 3). In addition, a recent study (Rajarajan et al., 2018) using HiC data alone or in combination with additional criteria revealed some, but not all, of the putative schizophrenia genes identified by our study (Table 3). The fact that polymorphic variation is now detected both in a candidate disease gene indirectly implicated by GWAS, and in a putative enhancer connected by long-range interaction to that gene, strengthens the case for that particular gene being indeed, when mutated or dysregulated, a contributor to the disease. In conclusion, Table 3, also taken in connection with Table 2, indicates that previous HiC studies are not exhaustive in identifying candidate schizophrenia genes, and suggests that our work, based on RNApolII-ChIA-PET analysis, is a useful complement to HiC-based studies (possible reasons for this being discussed already in the previous paragraphs). Having said this, it is appropriate to mention that the functional effect of the targeted perturbation (by deletion or epigenetic repression) of a given enhancer on the expression of the connected

gene has been reported only in a few instances, on a small number of enhancers so far (e.g. two out of four connected genes were validated as targets of variant-containing enhancers, by functional enhancer activation/inactivation studies by Rajarajan et al. (Rajarajan et al., 2018); one enhancer was validated in Won et al. (Won et al., 2016). After this crucial step, the connected candidate gene will also have to be further validated as a schizophrenia-contributor gene by functional studies of the gene itself in developing neural brain cells.

As to molecular mechanisms, we hypothesize that the sequence variant in the enhancer acts by altering transcription factor binding to it, in turn leading to deregulation of the connected gene promoter. For enhancers connected to the TRIM8 gene, we show that some of the variants indeed fall within a transcription factor binding consensus; intriguingly, one variant alters a potential recognition site for *FOXG1*, a TF whose mutation causes a NDD of the Autism disorders spectrum (Fig. 5). These data will allow to study, in future experiments, whether transcription factor binding is indeed altered (by EMSA and chromatin immunoprecipitation assays), and whether this is relevant for the function of the connected gene (by targeted mutagenesis, or epigenetic modulation, of the enhancer via CRISPR-based approaches) (Doudna, 2020). For those variants that do not alter consensus binding sites, it will be possible to perform targeted sequencing of the whole hmsLRI-E in patients carrying the variant, to uncover possible pathogenic mutations in close linkage with the variant itself.

Approximately 97% of our hmsLRI-E overlap with cCRE (Fig. 4); it was thus somewhat unexpected to find that, among hmsLRI-E containing variants associated with schizophrenia and bipolar disorders, only 10 out of 32 overlapped with a cCRE (Tables 3, S2). As our hmsLRI-E were identified in neonatal NSC (Bertolini et al., 2019), it is possible that the disease-associated variants present in hmsLRI-E, but not in cCRE, affect enhancers active in cells more undifferentiated than those where cCRE are identified (adult brain). Indeed, our hmsLRI-E are enriched in VISTA enhancers, active in the mouse developing nervous system in early and mid-embryogenesis (Bertolini et al., 2019) (see also Table S1). This interpretation is in keeping with the likely early developmental origin of many instances of schizophrenia and bipolar disorders.

On the other hand, some of the GWAS-identified variants, overlapping our hmsLRI-E, do overlap the human counterpart (identified by synteny) of cCRE (Li et al., 2021) (Tables 3, S2). We know about the specific cell types of the adult (mouse) brain, where cCRE are in an open chromatin configuration (Li et al., 2021); this will be very useful to formulate hypotheses about the cellular environment where the variant enhancer functions, implying a possible involvement of that cell

type in the cellular mechanisms of pathogenesis. In this regard, it is useful to point out that the interactions that we detect in mouse NSC are largely dependent on the *SOX2* transcription factor, as *Sox2* is bound to a vast fraction of the interacting anchors, and *Sox2* deletion leads to the loss of a high proportion of the interactions. This implies that a *Sox2*-dependent transcriptional program might be related to the cell type where our hmsLRI-E are predominantly detected. Indeed, it is interesting to note that the cell types, displaying the cCRE that are most enriched within our mouse enhancers involved in LRI (Fig. 4) correspond to cell types that are most affected following *Sox2* conditional deletion in mouse (radial glia-like cells; GABAergic neurons and neuroblasts; specific astrocyte types (Ferri et al., 2004; Cavallaro et al., 2008; Favaro et al., 2009; Ferri et al., 2013; Cerrato et al., 2018; Mercurio et al., 2019, 2021)

We further discovered that copy number variants (CNV), associated with neurodevelopmental disease, affect our hmsLRI-E. For example, some microdeletions affect specific hmsLRI-E, though not the connected gene, with a resulting disease phenotype similar to the one resulting from mutations within the gene exons: an example is the deletions affecting an hmsLRI-E connected to *FOXG1* gene (associated to an autism-related syndrome) (Fig. 6A). These findings will direct future CRISPR-based experiments of targeted mutation, or epigenetic repression, of the identified enhancers, to test the effect on the expression of the connected gene, and the consequences on cell function (differentiation, survival etc.). More generally, our maps, and the annotation (Table S1) of each syntenic human LRI and LRI-E with CNVs in published datasets of CNVs associated with human NDD, will help understanding the specific contribution of enhancer CNVs to the disease.

Conclusion

The observations reported in Table 3 and Figs. 4,5 show that a small, but useful, number of disease-associated sequence variants, detected in hmsLRI-E, allows the identification of connected genes already identified as candidates for a role in brain disease, when mutated; as well as a substantial number of new candidate genes, indicating the usefulness of adding our approach to existing and future HiC-based studies.. This implies that, in principle, our strategy may also identify novel disease genes which had not been previously identified as disease genes (see cases in Table 3). In view of the ongoing large effort in genome-wide sequencing of patients with neural disease, it is likely that many new mutations in enhancers will be discovered, allowing to extend our present strategy to the identification of new candidate disease genes. Likewise, a deeper understanding of functional effects of deletions or duplications (CNV) can be obtained by the progressive addition, and analysis,

of new cases in which enhancers are involved.

Methods

Mapping mouse LRI-syntenic regions within the human genome

We used mouse LRIs previously identified by (Bertolini et al., 2019) in NSC cultured from mouse neonatal forebrain cultured in vitro. Starting from ChIP-Seq experiments for H3K4me1 and H3K27Ac histone modifications (Table S4 of (Bertolini et al., 2019) and ChIA-PET experiments for 3-D chromatin interactions (Table S1 of (Bertolini et al., 2019), we obtained the list of coordinates of anchors (L, left anchor; R, right anchor) defining mouse LRIs for wTR1, wTR2 and wTR3 replicate experiments (“anchors” are defined as DNA regions associated via long-range interactions in ChIA-PET experiments, see (Bertolini et al., 2019)(Fig. 1). As a close correspondence had been found between these three replicates, we considered all the interactions previously identified for wTR1, wTR2 and wTR3. From this list, we selected the interactions between promoter (P) anchors (defined as $\pm 2.5\text{kb}$ of TSS) and non-promoter, genic anchors (G, i.e. overlapping with a gene region) or intergenic anchors (I, i.e. not overlapping with any gene or promoter region). The “TxDb.Mmusculus.UCSC.mm9KNOWNGENEMOUSE” Bioconductor package was used for updating the anchors annotation (Table S1). Non-promoter anchors embedding both H3K27Ac and H3K4me1 (300 bp min length) epigenetic marks with 10 bp min reciprocal overlap were defined “active enhancers” while those carrying only H3K4me1 marked region (300 bp min length) were considered “poised enhancers”. For downstream analysis, we kept only mouse LRIs with enhancer anchors. Both anchor coordinates were separately mapped to human genome by means of LiftOver command line tool (Kent et al., 2002) using <http://hgdownload.soe.ucsc.edu/goldenPath/mm9/liftOver/mm9ToHg19.over.chain.gz> chain file and specifying -minMatch 0.5 which means that more than 50% of bases lifted over to the human genome. Subsequently, mouse LRIs for which both anchors were conserved in a syntenic region of the human genome were selected and called “hmsLRI” (Table 1). The identified human syntenic regions were annotated using the UCSC KnownGene track “TxDb.Hsapiens.UCSC.hg19KNOWNGENE” (Carlson, Marc and Maintainer, 2015) distinguishing among exonic, intergenic, intronic or promoter (Fig. 2B). Human syntenic enhancer regions were classified intragenic if they overlapped (1 bp min.) any gene coordinates (5'UTR, exon, intron, 3' UTR) while syntenic promoter regions were first checked for overlap with any promoter region (defined as $\pm 2.5\text{kb}$ of TSS). Conservation level for hmsLRI-E (enhancers involved in hmsLRI) was measured using UCSC phastCons conservation scores for the human genome (hg19)

calculated from multiple alignments with other 99 vertebrate species and retrieved in R using Bioconductor package phastCons100way.UCSC.hg19 (Siepel et al., 2005). A set of random shuffled genomic regions with similar length and chromosomal distribution of hmsLRI were simulated using randomizeRegions() function of regioneR package (GelB. et al. (2015) regioneR: an R/Bioconductor package for the association analysis of genomic regions based on permutation tests. Bioinformatics, 32, 289–291).

Comparing hmsLRI-E and hmsLRI with published datasets of epigenetic enhancer marks and long-range interactions previously determined in human neural cells

For TAD regions we dowloaded the DER-18-TAD_adultbrain resource from <http://resource.psychencode.org/> and CP/GZ_TAD.bed files from GEO (GSE77565 dataset).

We retrieved histone modification signals (broad inferred peaks) from the Roadmap Epigenomics database https://egg2.wustl.edu/roadmap/web_portal/meta.html and checked for an overlap of at least 50 bp with the hmsLRI-E. H3K4me1 peaks corresponding to potential enhancers were available for each neural tissue while H3K27ac only for a subset. Human interactions were from <http://resource.psychencode.org/> (HiC-derived Enhancer-Gene and Promoter linkages, INT-16_HiC_EP_linkages file)

Annotation of hmsLRI and hmsLRI-E by their overlap with VISTA enhancers, cCREs, and association with NDD

VISTA enhancers:

All fasta sequences of human 998 “positive” and 944 “negative” enhancers with evidence of in vivo activity in developing mouse embryos were downloaded on 29 september 2021 from the VISTA enhancer browser (https://enhancer.lbl.gov/frnt_page_n.shtml, (Visel et al., 2007, 2009) and compared with syntenic human enhancer anchors. Human VISTA enhancers had been tested in transgenic mouse embryos (Visel et al., 2007, 2009); each enhancer was defined “positive” if it showed reproducible expression in the same CNS area in at least three independent transgenic embryos, otherwise it is considered “negative”.

Comparison with brain candidate cis-regulatory elements (cCRE):

The full list of the 491818 mouse cCRE along with their genomic coordinates and cell type assignments were retrieved from Supplementary Table 8 of (Li et al., 2021). We checked for any overlap with our mouse enhancer anchors using findOverlaps function of GenomicRanges.

Gene2Phenotype annotation:

Developmental Disorders Genotype-to-Phenotype Database (DDG2P, v.2021.11.12) and EyeG2P (2021.10.5) gene lists were downloaded from the

Gene2Phenotype (G2P) website

(<https://www.ebi.ac.uk/gene2phenotype/>). In DDG2P we selected only genes associated to eye and brain/cognition ("organ specificity list"), the full list was instead kept for EyeG2P. After matching with hmsLRI associated genes we reported their gene-diseases confidence level of association (Table S1,S2) as reported by the Gene Curation Coalition (<https://thegencc.org/>). Gene-disease connection is classified by this database as definitive, strong, moderate, supportive, limited, disputed, refuted, animal or unknown.

Mapping DDD risk variants with respect to hmsLRI-E

We used the subset of GWAS variants with a posterior probability of association (PPA) larger than 1% in schizophrenia (Consortium, 2014), bipolar disorder (Hou et al., 2016) and intelligence (Savage et al., 2018) previously identified as potential causal variants by (Li et al., 2021), which had been overlapped with cCREs. The list of sequence variants was crossed with the hmsLRI-E (Table S1), and the variants showing overlap are reported in Table S2, together with the coordinates of the overlapping hmsLRI-E.

The full lists of variant regions of Zarrei's and Coe's datasets were downloaded from dbVar study accession number nstd173 and nstd100, respectively. We filtered variant regions with copy number gain/loss (hg19 coordinates) and compared them to hmsLRI using findOverlaps function of GenomicRanges Bioconductor package. AnnotSV v 2.2 (Geoffroy et al., 2021) was used to classify CNV according to their potential clinical impact. Three classes were taken into account: benign (encompassing AnnotSV Class 1 and 2), variant of unknown significance (equivalent to AnnotSV class 3) and pathogenic (encompassing AnnotSV Class 4 and 5) variants. The study from Coe et al. 2014 is available from dbVar nstd100; it was designed to identify CNVs associated with Developmental delay. In this dataset we have data from individuals with intellectual disability, developmental delay and/or ASD (although the precise phenotype is not specified among the dataset) for a total of 29,085 primarily pediatric cases in comparison to 19,584 adult population controls. The variant calls from Coe et al. were mapped with respect to hmsLRI, to possibly identify rare CNVs that can disrupt the interaction

Bibliography

Allou, L., Lambert, L., Amsallem, D., Bieth, E., Edery, P., Destrée, A., Rivier, F., Amor, D., Thompson, E., Nicholl, J., Harbord, M., Nemos, C., Saunier, A., Moustaine, A., Vigouroux, A., Jonveaux, P., and Philippe, C. 14q12 and severe Rett-like phenotypes: New clinical insights and physical mapping of FOXG1-regulatory elements. *European Journal of Human Genetics*, 20(12):1216–1223, 2012. doi: 10.1038/ejhg.2012.127.

Ansari, M., Rainger, J., Hanson, I. M., Williamson, K. A., Sharkey, F., Harewood, L., Sandilands, A., Clayton-Smith, J., Dolfus, H., Bitoun, P., Meire, F., Fantes, J., Franco, B., Lorenz, B., Taylor, D. S., Stewart, F., Willoughby, C. E., McEntagart, M., Khaw, P. T., Clericuzio, C., Van Maldergem, L., Williams, D., Newbury-Ecob, R., Traboulsi, E. I., Silva, E. D., Madlom, M. M., Goudie, D. R., Fleck, B. W., Wieczorek, D., Kohlhase, J., McTrusty, A. D., Gardiner, C., Yale, C., Moore, A. T., Russell-Eggett, I., Islam, L., Lees, M., Beales, P. L., Tuft, S. J., Solano, J. B., Spill, M., Hertz, J. M., Prescott, T. E., Shears, D. J., Nischal, K. K., Doco-Fenzy, M., Prieur, F., Temple, I. K., Lachlan, K. L., Damante, G., Morrison, D. A., Van Heyningen, V., and Fitzpatrick, D. R. Genetic analysis of 'PAX6-negative' individuals with aniridia or Gillespie Syndrome. *PLoS ONE*, 11(4):1–15, 2016. doi: 10.1371/journal.pone.0153757.

Bar Yaacov, R., Eshel, R., Farhi, E., Shemeluvich, F., Kaplan, T., and Birnbaum, R. Y. Functional characterization of the ZEB2 regulatory landscape. *Human molecular genetics*, 28 (9):1487–1497, 2019. doi: 10.1093/hmg/ddy440.

Barbosa, S., Greville-Heygate, S., Bonnet, M., Godwin, A., Fagotto-Kaufmann, C., Kajava, A. V., Laoutouet, D., Mawby, R., Wai, H. A., Dingemans, A. J., Hehir-Kwa, J., Willems, M., Capri, Y., Mehta, S. G., Cox, H., Goudie, D., Vansenne, F., Turnpenny, P., Vincent, M., Cogné, B., Lescas, G., Hertecant, J., Rodriguez, D., Keren, B., Burglen, L., Gérard, M., Putoux, A., Cantagrel, V., Siquier-Pernet, K., Rio, M., Banka, S., Sarkar, A., Steeves, M., Parker, M., Clement, E., Moutton, S., Tran Mau-Them, F., Piton, A., de Vries, B. B., Guille, M., Debant, A., Schmidt, S., and Baralle, D. Opposite Modulation of RAC1 by Mutations in TRIO Is Associated with Distinct, Domain-Specific Neurodevelopmental Disorders. *The American Journal of Human Genetics*, 106(3):338–355, mar 2020. doi: 10.1016/j.ajhg.2020.01.018.

Benito-Kwiecinski, S., Giandomenico, S. L., Sutcliffe, M., Riis, E. S., Freire-Pritchett, P., Kelava, I., Wunderlich, S., Martin, U., Wray, G. A., McDole, K., and Lancaster, M. A. An early cell shape transition drives evolutionary expansion of the human forebrain. *Cell*, 184(8):2084–2102.e19, apr 2021. doi: 10.1016/j.cell.2021.02.050.

Bertolo, A. J., Favaro, R., Zhu, Y., Pagin, M., Ngan, C. Y., Wong, C. H., Tjong, H., Vermunt, M. W., Martynoga, B., Barone, C., Mariani, J., Cardozo, M. J., Tabanera, N., Zambelli, F., Mercurio, S., Ottolenghi, S., Robson, P., Creyghton, M. P., Bovolenta, P., Pavesi, G., Guillemot, F., Nicolis, S. K., and Wei, C. L. Mapping the Global Chromatin Connectivity Network for Sox2 Function in Neural Stem Cell Maintenance. *Cell Stem Cell*, 24(3): 462–476.e6, 2019. doi: 10.1016/j.stem.2019.02.004.

Bidart, M., El Atifi, M., Miladi, S., Rendu, J., Satre, V., Ray, P. F., Bosson, C., Devillard, F., Lehalle, D., Malan, V., Amiel, J., Mencarelli, M. A., Baldassari, M., Renieri, A., Clayton-Smith, J., Vieville, G., Thevenon, J., Amlard, F., Berger, F., Jouk, P. S., and Coutton, C. Microduplication of the ARID1A gene causes intellectual disability with recognizable syndromic features. *Genetics in Medicine*, 19(6):701–710, 2017. doi: 10.1038/gim.2016.180.

Boland, E., Clayton-Smith, J., Woo, V. G., McKee, S., Manson, F. D., Medne, L., Zackai, E., Swanson, E. A., Fitzpatrick, D., Millen, K. J., Sherr, E. H., Dobyns, W. B., and Black, G. C. Mapping of deletion and translocation breakpoints in 1q44 implicates the serine/threonine kinase AKT3 in postnatal microcephaly and agenesis of the corpus callosum. *American Journal of Human Genetics*, 81(2):292–303, 2007. doi: 10.1086/519999.

Carlson, Marc and Maintainer, B. P. TxDb.Hsapiens. UCSC. hg19. knownGene. 2015.

Cavallaro, M., Mariani, J., Lancini, C., Latorre, E., Caccia, R., Gullo, F., Valotta, M., DeBiasi, S., Spinardi, L., Ronchi, A., Wanke, E., Brunelli, S., Favaro, R., Ottolenghi, S., and Nicolis, S. K. Impaired generation of mature neurons by neural stem cells from hypomorphic Sox2 mutants. *Development*, 135(3):541–557, 2008. doi: 10.1242/dev.010801.

Cerrato, V., Mercurio, S., Leto, K., Fucà, E., Hoxha, E., Bottes, S., Pagin, M., Milanese, M., Ngan, C. Y., Concina, G., Ottolenghi, S., Wei, C. L., Bonanno, G., Pavesi, G., Tempia, F., Buffo, A., and Nicolis, S. K. Sox2 conditional mutation in mouse causes ataxic symptoms, cerebellar vermis hypoplasia, and postnatal defects of Bergmann glia. *Glia*, 66(9):1929–1946, 2018. doi: 10.1002/glia.23448.

Coe, B. P., Witherspoon, K., Rosenfeld, J. A., Van Bon, B. W., Vulto-Van Silfhout, A. T., Bosco, P., Friend, K. L., Baker, C., Buono, S., Vissers, L. E., Schuurs-Hoeijmakers, J. H., Hoischen, A., Pfundt, R., Krumm, N., Carvill, G. L., Li, D., Amaral, D., Brown, N., Lockhart, P. J., Scheffer, I. E., Alberti, A., Shaw, M., Pettinato, R., Tervo, R., De Leeuw, N., Reijnders, M. R., Torchia, B. S., Peeters, H., Thompson, E., O'Roak, B. J., Fichera, M., Hehir-Kwa, J. Y., Shendure, J., Mefford, H. C., Haan, E., Gécz, J., De Vries, B. B., Roman, C., and Eichler, E. E. Refining analyses of copy number variation identifies specific genes associated with developmental delay. *Nature Genetics*, 46(10):1063–1071, 2014. doi: 10.1038/ng.3092.

Consortium, S. W. G. o. t. P. G. Biological insights from 108 schizophrenia-associated genetic loci. *Nature*, 511(7510):421–427, jul 2014. doi: 10.1038/nature13595.

Consortium, T. B. Analysis of shared heritability in common disorders of the brain. *Science*, 360(6395), 2018. doi: 10.1126/science.aap8757.

Creyghton, M. P., Cheng, A. W., Welstead, G. G., Kooistra, T., Carey, B. W., Steine, E. J., Hanna, J., Lodato, M. A., Frampton, G. M., Sharp, P. A., Boyer, L. A., Young, R. A., and Jaenisch, R. Histone H3K27ac separates active from poised enhancers and predicts developmental state. *Proceedings of the National Academy of Sciences of the United States of America*, 107(20):21931–21936, 2010. doi: 10.1073/pnas.1016071107.

Cruz-Molina, S., Respuela, P., Tebar, C., Kolovos, P., Nikolic, M., Fueyo, R., van IJcken, W. F., Grosfeld, V., Frommolt, P., Bazzi, H., and Rada-Iglesias, A. PRC2 Facilitates the Regulatory Topology Required for Poised Enhancer Function during Pluripotent Stem Cell Differentiation. *Cell Stem Cell*, 20(5):689–705.e9, 2017. doi: 10.1016/j.stem.2017.02.004.

Deinhardt, K., Kim, T., Spellman, D. S., Mains, R. E., Eipper, B. A., Neubert, T. A., Chao, M. V., and Hempstead, B. L. Neuronal Growth Cone Retraction Relies on Proneurotrophin Receptor Signaling Through Rac. *Science Signaling*, 4(202):1–9, dec 2011. doi: 10.1126/scisignal.2002060.

Ding, C., Zhang, C., Kopp, R., Kuney, L., Meng, Q., Wang, L., Xia, Y., Jiang, Y., Dai, R., Min, S., Yao, W. D., Wong, M. L., Ruan, H., Liu, C., and Chen, C. Transcription factor POU3F2 regulates TRIM8 expression contributing to cellular functions implicated in schizophrenia. *Molecular Psychiatry*, 26(7):3444–3460, 2021. doi: 10.1038/s41380-020-00877-2.

Doudna, J. A. The promise and challenge of therapeutic genome editing. *Nature*, 578(7794): 229–236, 2020. doi: 10.1038/s41586-020-1978-5.

Ellaway, C. J., Ho, G., Bettella, E., Knapman, A., Collins, F., Hackett, A., McKenzie, F., Darmanian, A., Peters, G. B., Fagan, K., and Christodoulou, J. 14q12 microdeletions excluding FOXG1 give rise to a congenital variant Rett syndrome-like phenotype. *European Journal*

of Human Genetics, 21(5):522–527, 2013. doi: 10.1038/ejhg.2012.208.

Favarro, R., Valotta, M., Ferri, A. L., Latorre, E., Mariani, J., Giachino, C., Lancini, C., Tosetti, V., Ottolenghi, S., Taylor, V., and Nicolis, S. K. Hippocampal development and neural stem cell maintenance require Sox2-dependent regulation of Shh. *Nature Neuroscience*, 12(10):1248–1256, 2009. doi: 10.1038/nn.2397.

Ferri, A. L., Cavallaro, M., Braida, D., Di Cristofano, A., Canta, A., Vezzani, A., Ottolenghi, S., Pandolfi, P. P., Sala, M., DeBiasi, S., and Nicolis, S. K. Sox2 deficiency causes neurodegeneration and impaired neurogenesis in the adult mouse brain. *Development*, 131(15):3805–3819, 2004. doi: 10.1242/dev.01204.

Ferri, A., Favarro, R., Beccari, L., Bertolini, J., Mercurio, S., Nieto-Lopez, F., Verzneroli, C., La Regina, F., De Pietri Tonelli, D., Ottolenghi, S., Bovolenta, P., and Nicolis, S. K. Sox2 is required for embryonic development of the ventral telencephalon through the activation of the ventral determinants Nkx2.1 and Shh. *Development (Cambridge)*, 140(6):1250–1261, 2013. doi: 10.1242/dev.073411.

Franzoni, A., Russo, P. D., Baldan, F., D'Elia, A. V., Puppin, C., Penco, S., and Damante, G. A CGH array procedure to detect PAX6 gene structural defects. *Molecular and Cellular Probes*, 32:65–68, 2017. doi: 10.1016/j.mcp.2016.12.001.

Gao, J. Y., Ma, P., Li, Y., Yan, C. X., Zhang, Q., Yang, X. X., Shi, Q., Zhang, B., and Wen, X. P. Association between a TCF4 Polymorphism and Susceptibility to Schizophrenia. *BioMed Research International*, 2020, 2020. doi: 10.1155/2020/1216303.

Geoffroy, V., Guignard, T., Kress, A., Gaillard, J. B., Solli-Nowlan, T., Schalk, A., Gatinois, V., Dolfus, H., Scheidecker, S., and Muller, J. AnnotSV and knotAnnotSV: A web server for human structural variations annotations, ranking and analysis. *Nucleic Acids Research*, 49(W1):W21–W28, 2021. doi: 10.1093/nar/gka402.

Gritti, A., Parati, E., Cova, L., Frolichthal, P., Galli, R., Wanke, E., Faravelli, L., Morassutti, D., Roisen, F., Nickel, D., and Vescovi, A. Multipotent stem cells from the adult mouse brain proliferate and self-renew in response to basic fibroblast growth factor. *The Journal of Neuroscience*, 16(3):1091–1100, feb 1996. doi: 10.1523/JNEUROSCI.16-03-01091.1996.

Hamanaka, K., Sugawara, Y., Shimoji, T., Nordtveit, T. I., Kato, M., Nakashima, M., Saito, H., Suzuki, T., Yamakawa, K., Aukrust, I., Houge, G., Mitsuhashi, S., Takata, A., Iwama, K., Aikanaq, A., Fujita, A., Imagawa, E., Mizuguchi, T., Miyake, N., Miyatake, S., and Matsumoto, N. De novo truncating variants in PHF21A cause intellectual disability and craniofacial anomalies. *European Journal of Human Genetics*, 27(3):378–383, 2019. doi: 10.1038/s41431-018-0289-x.

Hindorf, L. A., Sethupathy, P., Junkins, H. A., Ramos, E. M., Mehta, J. P., Collins, F. S., and Manolio, T. A. Potential etiologic and functional implications of genome-wide association loci for human diseases and traits. *Proceedings of the National Academy of Sciences*, 106(23):9362–9367, jun 2009. doi: 10.1073/pnas.0903103106.

Hou, L., Bergen, S. E., Akula, N., Song, J., Hultman, C. M., Landén, M., Adli, M., Alda, M., Ardua, R., Arias, B., Aubry, J.-M., Backlund, L., Badner, J. A., Barrett, B. B., Bauer, M., Baune, T. B., Bellivier, F., Benaberre, A., Bengesser, S., Berrettini, W. H., Bhattacharjee, A. K., Biernacka, J. M., Birner, A., Bloss, C. S., Brichant-Petitjean, C., Bui, E. T., Byerley, W., Cervantes, P., Chilotti, C., Cichon, S., Colom, F., Coryell, W., Craig, D. W., Cruceanu, C., Czerski, P. M., Davis, T., Dwyer, A., Degenhardt, F., Del Zompo, M., DePaulo, J. R., Edenberg, H. J., Étain, B., Falkai, P., Foroud, T., Forstner, A. J., Frisén, L., Frye, M. A., Fullerton, J. M., Gard, S., Garnham, J. S., Gershon, E. S., Goes, F. S., Greenwood, T. A., Grigoriou-Serbanescu, M., Hauser, J., Heilbronner, U., Heilmann-Heimbach, S., Herms, S., Hippolito, M., Hittlingappaa, S., Hoffmann, P., Hoffman, A., Jamain, S., Jiménez, E., Kahn, J.-P., Kassem, L., Kelsoe, J. R., Kittel-Schneider, S., Kliwicki, S., Koller, D. L., König, B., Lackner, N., Laje, G., Lang, M., Lavebratt, C., Lawson, W. B., Leboyer, M., Leckband, S. G., Liu, C., Maaser, A., Mahon, P. B., Maier, W., Maj, M., Manchia, M., Martinsson, L., McCarthy, M. J., McElroy, S. L., McInnis, M. G., McKinney, R., Mitchell, P. B., Mitjans, M., Mondimore, F. M., Monteleone, P., Mühlleisen, T. W., Nievergelt, C. M., Nöthen, M. M., Novák, T., Nurnberger, J. J., Nwulia, E. A., Ösby, U., Pfennig, A., Potash, J. B., Propping, P., Reif, A., Reininghaus, E., Rice, J., Rietschel, M., Rouleau, G. A., Rybakowski, J. K., Schalling, M., Scheftner, W. A., Schofield, P. R., Schork, N. J., Schulze, T. G., Schumacher, J., Schweizer, B. W., Severini, G., Shekhtman, T., Shilling, P. D., Simhandl, C., Slaney, C. M., Smith, E. N., Squassina, A., Stamm, T., Stopkova, P., Streit, F., Strohmaier, J., Szelinger, S., Tighe, S. K., Tortorella, A., Turecki, G., Vieta, E., Volkert, J., Witt, S. H., Wright, A., Zandi, P. P., Zhang, P., Zolnier, S., and McMahon, F. J. Genome-wide association study of 40,000 individuals identifies two novel loci associated with bipolar disorder. *Human Molecular Genetics*, 25(15):3383–3394, aug 2016. doi: 10.1093/hmg/ddw181.

Indjeian, V. B., Kingman, G. A., Jones, F. C., Guenther, C. A., Grimwood, J., Schmutz, J., Myers, R. M., and Kingsley, D. M. Evolving New Skeletal Traits by cis-Regulatory Changes in Bona Morphogenetic Proteins. *Cell*, 164(1-2):45–56, 2016. doi: 10.1016/j.cell.2015.12.007.

Jung, I., Schmitt, A., Diao, Y., Lee, A. J., Liu, T., Yang, D., Tan, C., Eom, J., Chan, M., Chee, S., Chiang, Z., Kim, C., Masliyah, E., Barr, C. L., Li, B., Kuan, S., Kim, D., and Ren, B. A compendium of promoter-centered long-range chromatin interactions in the human genome. *Nature Genetics*, 51(10):1442–1449, oct 2019. doi: 10.1038/s41588-019-0494-8.

Kent, W. J., Sugnet, C. W., Furey, T. S., Roskin, K. M., Pringle, T. H., Zahler, A. M., Haussler, and David. The Human Genome Browser at UCSC. *Genome Research*, 12(6):996–1006, 2002. doi: 10.1101/022910.

Kleinjan, D. A. and Van Heyningen, V. Long-range control of gene expression: Emerging mechanisms and disruption in disease. *American Journal of Human Genetics*, 76(1):8–32, 2005. doi: 10.1086/426833.

Lam, M., Chen, C.-Y., Li, Z., Martin, A. R., Bryois, J., Ma, X., Gaspar, H., Ikeda, M., Benyamin, B., Brown, B. C., Liu, R., Zhou, W., Guan, L., Kamatani, Y., Kim, S.-W., Kubo, M., Kusumawardhani, A. A. A. A., Liu, C.-M., Ma, H., Periyasamy, S., Takahashi, A., Xu, Z., Yu, H., Zhu, F., Chen, W. J., Faraoone, S., Glatt, S. J., He, L., Hyman, S. E., Hwu, H.-G., McCarroll, S. A., Neale, B. M., Sklar, P., Wildenauer, D. B., Yu, X., Zhang, D., Mowry, B. J., Lee, J., Holmans, P., Xu, S., Sullivan, P. F., Ripke, S., O'Donovan, M. C., Daly, M. J., Qin, S., Sham, P., Iwata, N., Hong, K. S., Schwab, S. G., Yue, W., Tsuang, M., Liu, J., Ma, X., Kahn, R. S., Shi, Y., and Huang, H. Comparative genetic architectures of schizophrenia in East Asian and European populations. *Nature Genetics*, 51(12):1670–1678, dec 2019. doi: 10.1038/s41588-019-0512-x.

Le Hellard, S., Mühlleisen, T. W., Djurovic, S., Fernø, J., Ouriaghli, Z., Mattheisen, M., Vasilescu, C., Raeder, M. B., Hansen, T., Strohmaier, J., Georgi, A., Brockschmidt, F. F., Melle, I., Nenadic, I., Sauer, H., Rietschel, M., Nöthen, M. M., Werge, T., Andreassen, O. A., Cichon, S., and Steen, V. M. Polymorphisms in SREBF1 and SREBF2, two antipsychotic-activated transcription factors controlling cellular lipogenesis, are associated with schizophrenia in German and Scandinavian samples. *Molecular Psychiatry*, 15(5):463–472, 2010. doi: 10.1038/mp.2008.110.

Li, Y. E., Preissl, S., Hou, X., Zhang, Z., Zhang, K., Qiu, Y., Poirion, O. B., Li, B., Chiou, J., Liu, H., Pinto-Duarte, A., Kubo, N., Yang, X., Fang, R., Wang, X., Han, J. Y., Lucero, J., Yan, Y., Miller, M., Kuan, S., Gorkin, D., Gaulton, K. J., Shen, Y., Nunn, M., Mukamel, E. A., Behrens, M. M., Ecker, J. R., and Ren, B. An atlas of gene regulatory elements in adult mouse cerebrum. *Nature*, 598(7879):129–136, 2021. doi: 10.1038/s41586-021-03604-1.

Li, H., Sun, Y., Zhang, X., Li, S., Hu, D., Xiao, L., Chen, Y., He, L., and Wang, D. W. Integrated Analysis of Summary Statistics to Identify Pleiotropic Genes and Pathways for the Comorbidity of Schizophrenia and Cardiometabolic Disease. *Frontiers in Psychiatry*, 11(April):1–12, 2020. doi: 10.3389/fpsyg.2020.00256.

Li, S., Rao, S., Xu, Y., Li, J., Huang, H., Zhang, X., Fu, H., Wang, Q., Cao, H., Baranova, A., Jin, C., and Zhang, F. Identifying common genome-wide risk genes for major psychiatric traits. *Human Genetics*, 139(2):185–198, 2020. doi: 10.1007/s00439-019-02096-4.

Li, J., Li, X., and Luo, X. J. Proteome-wide Association Study Provides Insights Into the Genetic Component of Protein Abundance in Psychiatric Disorders. *Biological Psychiatry*, 90(11):781–789, 2021. doi: 10.1016/j.biopsych.2021.06.022.

Long, H. K., Prescott, S. L., and Wysocka, J. Ever-Changing Landscapes: Transcriptional Enhancers in Development and Evolution. *Cell*, 167(5):1170–1187, 2016. doi: 10.1016/j.cell.2016.09.018.

Ma, L., Tang, J., Wang, D., Zhang, W., Liu, W., Wang, D., Liu, X. H., Gong, W., Yao, Y. G., and Chen, X. Evaluating risk loci for schizophrenia distilled from genome-wide association studies in Han Chinese from central China. *Molecular Psychiatry*, 18(6):638–639, 2013. doi: 10.1038/mp.2012.63.

Maurano, M. T., Humbert, R., Rynes, E., Thurman, R. E., Haugen, E., Wang, H., Reynolds, A. P., Sandstrom, R., Qu, H., Brody, J., Shaffer, A., Neri, F., Lee, K., Kutyavin, T., Stehling-Sun, S., Johnson, A. K., Canfield, T. K., Giste, E., Diegel, M., Bates, D., Hansen, R. S., Neph, S., Sabo, P. J., Heimfeld, S., Raubitschek, A., Ziegler, S., Cotsapas, C., Sotoodehnia, N., Glass, I., Sunyaev, S. R., Kaul, R., and Stamatoyannopoulos, J. A. Systematic Localization of Common Disease-Associated Variation in Regulatory DNA. *Science*, 337 (6099):1190–1195, sep 2012. doi: 10.1126/science.1222794.

Meers, M. P., Bryson, T. D., Henikoff, J. G., and Henikoff, S. Improved cutrun chromatin profiling tools. *eLife*, 8:1–16, 2019. doi: 10.7554/eLife.46314.001.

Mercurio, S., Serra, L., Motta, A., Gesuita, L., Sanchez-Arrones, L., Inverardi, F., Foglio, B., Barone, C., Kaimakis, P., Martynoga, B., Ottolenghi, S., Studer, M., Guillermot, F., Frassoni, C., Bovolenta, P., and Nicolis, S. K. Sox2 Acts in Thalamus Neurons to Control the Development of Retina-Thalamus-Cortex Connectivity. *iScience*, 15:257–273, 2019. doi: 10.1016/j.isci.2019.04.030.

Mercurio, S., Alberti, C., Serra, L., Meneghini, S., Berico, P., Bertolini, J., Becchetti, A., and Nicolis, S. K. An early Sox2-dependent gene expression programme required for hippocampal dentate gyrus development. *Open Biology*, 11(2), 2021. doi: 10.1098/rsob.200339.

Moriano, J. and Boeckx, C. Modern human changes in regulatory regions implicated in cortical development. *BMC Genomics*, 21(1):1–10, 2020. doi: 10.1186/s12864-020-6706-x.

Mowat, D. R., Wilson, M. J., and Goossens, M. Mowat-Wilson syndrome. 40(5):305–310, may 2003. doi: 10.1136/jmg.40.5.305.

Pagin, M., Pernebrink, M., Giubellini, S., Barone, C., Sambruni, G., Zhu, Y., Chiara, M., Ottolenghi, S., Pavesi, G., Wei, C. L., Cantù, C., and Nicolis, S. K. Sox2 controls neural stem cell self-renewal through a Fos-centered gene regulatory network. *Stem Cells*, 39 (8):1107–1119, 2021a. doi: 10.1002/stem.3373.

Pagin, M., Pernebrink, M., Pitasi, M., Malignetti, F., Ngan, C.-Y., Ottolenghi, S., Pavesi, G., Cantù, C., and Nicolis, S. K. FOS Rescues Neuronal Differentiation of Sox2-Deleted Neural Stem Cells by Genome-Wide Regulation of Common SOX2 and AP1(FOS-JUN) Target Genes. *Cells*, 10(7):1757, jul 2021b. doi: 10.3390/cells10071757.

Prem, S., Millonig, J. H., and DiCicco-Bloom, E. Dysregulation of Neurite Outgrowth and Cell Migration in Autism and Other Neurodevelopmental Disorders. In Schousboe, A., editor. *Neurodevelopmental Disorders*, pages 109–153. Springer US, 2020. ISBN 9783030454920. doi: 10.1007/978-3-030-45493-7_5. URL http://link.springer.com/10.1007/978-3-030-45493-7_5.

Rada-Iglesias, A., Bajpai, R., Swigut, T., Brugmann, S. A., Flynn, R. A., and Wysocka, J. A. Unique chromatin signature uncovers early developmental enhancers in humans. *Nature*, 470(7333):279–285, 2011. doi: 10.1038/nature09692.

Rajaratnam, P., Borrman, T., Liao, W., Schrode, N., Flaherty, E., Casiño, C., Powell, S., Yashaswini, C., LaMarca, E. A., Kassim, B., Javidfar, B., Espeso-Gil, S., Li, A., Won, H., Geschwind, D. H., Ho, S.-M., MacDonald, M., Hoffman, G. E., Rousos, P., Zhang, B., Hahn, C.-G., Weng, Z., Brennan, K. J., and Akbarian, S. Neuron-specific signatures in the chromosomal connectome associated with schizophrenia risk. *Science*, 362(6420), dec 2018. doi: 10.1126/science.aat4311.

Sadybekov, A., Tian, C., Arnesano, C., Katritch, V., and Herring, B. E. An autism spectrum disorder-related de novo mutation hotspot discovered in the GEF1 domain of Trio. *Nature Communications*, 8(1):1–12, 2017. doi: 10.1038/s41467-017-00472-0.

Savage, J. E., Jansen, P. R., Stringer, S., Watanabe, K., Bryois, J., de Leeuw, C. A., Nagel, M., Awasthi, S., Barr, P. B., Coleman, J. R. I., Grasby, K. L., Hammerschlag, A. R., Kaminski, J. A., Karlsson, R., Krapohl, E., Lam, M., Nygaard, M., Reynolds, C. A., Trampush, J. W., Young, H., Zabaneh, D., Hägg, S., Hansell, N. K., Karlsson, I. K., Linnarsson, S., Montgomery, G. W., Muñoz-Manchado, A. B., Quinlan, E. B., Schumann, G., Skene, N. G., Webb, B. T., White, T., Arking, D. E., Avramopoulos, D., Bilder, R. M., Bitsios, P., Burdick, K. E., Cannon, T. D., Chiba-Falek, O., Christoforou, A., Cirulli, E. T., Congdon, E., Corvin,

A., Davies, G., Deary, I. J., DeRosse, P., Dickinson, D., Djurovic, S., Donohoe, G., Conley, E. D., Eriksson, J. G., Espeseth, T., Freimer, N. A., Giakoumaki, S., Giegling, I., Gill, M., Glahn, D. C., Hariri, A. R., Hatzimanolis, A., Keller, M. C., Knowles, E., Koltai, D., Konte, B., Lahti, J., Le Hellard, S., Lencz, T., Liewald, D. C., London, E., Lundervold, A. J., Malhotra, A. K., Melle, I., Morris, D., Need, A. C., Oliier, W., Palotie, A., Payton, A., Pendleton, N., Poldrack, R. A., Räikkönen, K., Reinvang, I., Roussos, P., Rujescu, D., Sabb, F. W., Scult, M. A., Smeland, O. B., Smyrnis, N., Starr, J. M., Steen, V. M., Stefanis, N. C., Straub, R. E., Sundet, K., Tiemeier, H., Voineskos, A. N., Weinberger, D. R., Widen, E., Yu, J., Abecasis, G., Andreassen, O. A., Breen, G., Christiansen, L., Debrabant, B., Dick, D. M., Heinz, A., Hjerling-Leffler, J., Ikram, M. A., Kendler, K. S., Martin, N. G., Medland, S. E., Pedersen, N. L., Plomin, R., Polderman, T. J. C., Ripke, S., van der Sluis, S., Sullivan, P. F., Vrieze, S. I., Wright, M. J., and Posthuma, D. Genome-wide association meta-analysis in 269,867 individuals identifies new genetic and functional links to intelligence. *Nature Genetics*, 50(7):912–919, jul 2018. doi: 10.1038/s41588-018-0152-6.

Sheikh, T. I., Harrapul, R., Ayub, M., and Vincent, J. B. MeCP2 AT-Hook1 mutations in patients with intellectual disability and/or schizophrenia disrupt DNA binding and chromatin compaction in vitro. *Human Mutation*, 39(5):717–728, may 2018. doi: 10.1002/humu.23409.

Short, P. J., McRae, J. F., Gallone, G., Sifrim, A., Won, H., Geschwind, D. H., Wright, C. F., Firth, H. V., Fitzpatrick, D. R., Barrett, J. C., and Hurles, M. E. De novo mutations in regulatory elements in neurodevelopmental disorders. *Nature*, 555(7698):611–616, 2018. doi: 10.1038/nature25983.

Siepel, A., Bejerano, G., Pedersen, J. S., Hinrichs, A. S., Hou, M., Rosenbloom, K., Clawson, H., Spieth, J., Hillier, L. D. W., Richards, S., Weinstock, G. M., Wilson, R. K., Gibbs, R. A., Kent, W. J., Miller, W., and Haussler, D. Evolutionarily conserved elements in vertebrate, insect, worm, and yeast genomes. *Genome Research*, 15(8):1034–1050, 2005. doi: 10.1101/gr.371505.

Song, M., Pebworth, M.-P., Yang, X., Abnousi, A., Fan, C., Wen, J., Rosen, J. D., Choudhary, M. N. K., Cui, X., Jones, I. R., Bergenholz, S., Eze, U. C., Juric, I., Li, B., Maliskova, L., Lee, J., Liu, W., Pollen, A. A., Li, Y., Wang, T., Hu, M., Kriegstein, A. R., and Shen, Y. Cell-type-specific 3D epigenomes in the developing human cortex. *Nature*, 587(7835):644–649, nov 2020. doi: 10.1038/s41586-020-2825-4.

Takagi, M., Sasaki, G., Mitsui, T., Honda, M., Tanaka, Y., and Hasegawa, T. A 2.0Mb 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. *European Journal of Medical Genetics*, 56(9):526–528, 2013. doi: 10.1016/j.ejmg.2013.05.012.

Thul, P. J. and Lindskog, C. The human protein atlas: A spatial map of the human proteome. *Protein Science*, 27(1):233–244, 2018. doi: 10.1002/pro.3307.

Turner, T. N. and Eichler, E. E. The Role of De Novo Noncoding Regulatory Mutations in Neurodevelopmental Disorders. *Trends in Neurosciences*, 42(2):115–127, 2019. doi: 10.1016/j.tins.2018.11.002.

Visel, A., Minovitsky, S., Dubchak, I., and Pennacchio, L. A. VISTA Enhancer Browser—a database of tissue-specific human enhancers. *Nucleic Acids Research*, 35(Database):D88–D92, jan 2007. doi: 10.1093/nar/gkl822.

Visel, A., Blow, M. J., Li, Z., Zhang, T., Akiyama, J. A., Holt, A., Plajzer-Frick, I., Shoukry, M., Wright, C., Chen, F., Afzal, V., Ren, B., Rubin, E. M., and Pennacchio, L. A. ChIP-seq accurately predicts tissue-specific activity of enhancers. *Nature*, 457(7231):854–858, 2009. doi: 10.1038/nature07730.

Visel, A., Rubin, E. M., and Pennacchio, L. A. Genomic views of distant-acting enhancers. *Nature*, 461(7261):199–205, 2009. doi: 10.1038/nature08451.

Wang, D., Liu, S., Warrell, J., Won, H., Shi, X., Navarro, F. C., Clarke, D., Gu, M., Emani, P., Yang, Y. T., Min, Y., Gandal, M. J., Lou, S., Zhang, J., Park, J. J., Yan, C., KyongRhee, S., Manakongtreeep, K., Zhou, H., Aparna Natha, A., Peters, M., Mattei, E., Fitzgerald, D., Brunetti, T., Moore, J., Jiang, Y., Girdhar, K., Hoffman, G. E., Kalayci, S., Gümuş, Z. H., Crawford, G. E., Roussos, P., Akbarian, S., Jaffe, A. E., White, K. P., Weng, Z., Sestan, N., Geschwind, D. H., Knowles, J. A., and Gerstein, M. B. Comprehensive functional genomic resource and integrative model for the human brain. *Science*, 362(6420), 2018. doi: 10.1126/science.aat8464.

Wei, C. L., Nicolis, S. K., Zhu, Y., and Pagin, M. Sox2-Dependent 3D Chromatin Interactomes in Transcription, Neural Stem Cell Proliferation and Neurodevelopmental Diseases. *Journal of Experimental Neuroscience*, 13, 2019. doi: 10.1177/1179069519868224.

Williamson, K. A. and FitzPatrick, D. R. The genetic architecture of microphthalmia, anophthalmia and coloboma. *European Journal of Medical Genetics*, 57(8):369–380, 2014. doi: 10.1016/j.ejmg.2014.05.002.

Won, H., de la Torre-Ubieta, L., Stein, J. L., Parkishak, N. N., Huang, J., Opland, C. K., Gandal, M. J., Sutton, G. J., Hormozdiari, F., Lu, D., Lee, C., Eskin, E., Voineagu, I., Ernst, J., and Geschwind, D. H. Chromosome conformation elucidates regulatory relationships in developing human brain. *Nature*, 538(7626):523–527, 2016. doi: 10.1038/nature19847.

Zarrei, M., Burton, C. L., Engchuan, W., Young, E. J., Higginbotham, E. J., MacDonald, J. R., Trost, B., Chan, A. J., Walker, S., Lamoureux, S., Heung, T., Mojarrad, B. A., Kellam, B., Paton, T., Faheem, M., Miron, K., Lu, C., Wang, T., Samler, K., Wang, X., Costain, G., Hoang, N., Pellecchia, G., Wei, J., Patel, R. V., Thiruvahindrapuram, B., Roifman, M., Merico, D., Goodale, T., Drmic, I., Speevak, M., Howe, J. L., Yuen, R. K., Buchanan, J. A., Vorstman, J. A., Marshall, C. R., Wintle, R. F., Rosenberg, D. R., Hanna, G. L., Woodbury-Smith, M., Cytrynbaum, C., Zwaigenbaum, L., Elsabbagh, M., Flanagan, J., Fernandez, B. A., Carter, M. T., Szatmari, P., Roberts, W., Lerch, J., Liu, X., Nicolson, R., Georgiades, S., Weksberg, R., Arnold, P. D., Bassett, A. S., Crosbie, J., Schachar, R., Stavropoulos, D. J., Anagnostou, E., and Scherer, S. W. A large data resource of genomic copy number variation across neurodevelopmental disorders. *npj Genomic Medicine*, 4(1), 2019. doi: 10.1038/s41525-019-0098-3.

Zhang, Y., Wong, C. H., Birnbaum, R. Y., Li, G., Favaro, R., Ngan, C. Y., Lim, J., Tai, E., Poh, H. M., Wong, E., Mulawadi, F. H., Sung, W. K., Nicolis, S., Ahituv, N., Ruan, Y., and Wei, C. L. Chromatin connectivity maps reveal dynamic promoter-enhancer long-range associations. *Nature*, 504(7479):306–310, 2013. doi: 10.1038/nature12716.

Zhao, X., Chen, J., Xiao, P., Feng, J., Nie, Q., and Zhao, X. M. Identifying age-specific gene signatures of the human cerebral cortex with joint analysis of transcriptomes and functional connectomes. *Briefings in Bioinformatics*, 22(4):1–12, 2021. doi: 10.1093/bib/bbaa388.

Supplementary Information

Figure S1

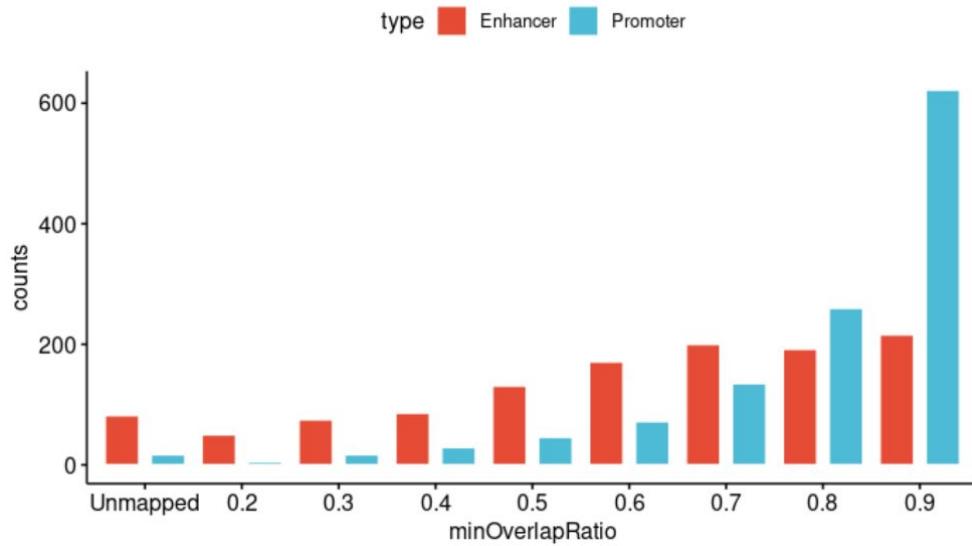


Figure S1. Number of mouse enhancer/promoter anchors of wTR1 experiment which were mapped to human genome by liftOver, foreach percentage of overlap (0.2 to 0.9). Unmapped anchors were also shown. Promoter anchors generally mapped at higher overlap level than enhancer

Figure S2

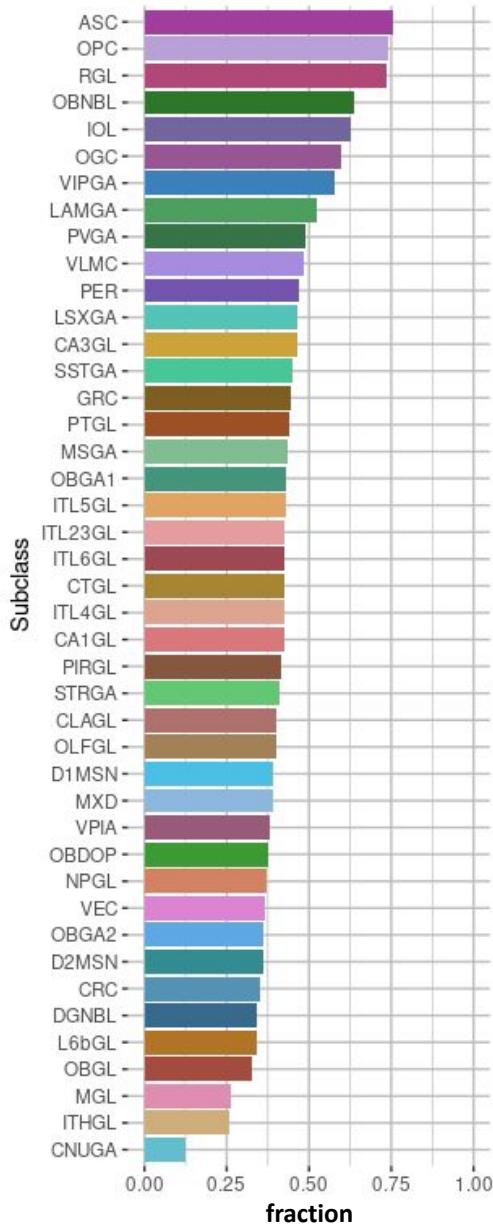


Figure S2

Sorted histograms showing the relative proportion of cCREs (Li et al. 2021) overlapping our mouse enhancers connected to promoters via LRI, for each of the 43 cell type subclasses. For abbreviations of cell subclasses names, see (Li et al. 2021).

Figure S3

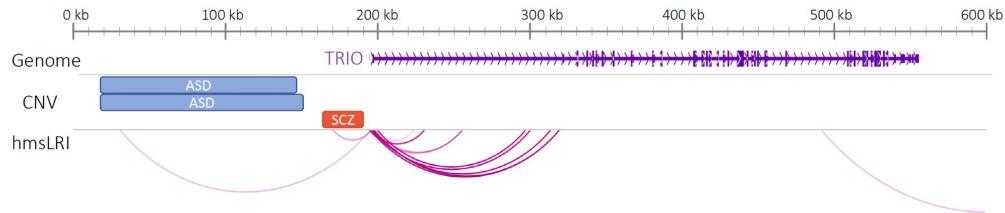


Figure S3

CNV associated to autism spectrum disorder (ASD; copy number gain, blue rectangle) or schizophrenia (SCZ; copy number loss, red rectangle) overlap with human putative enhancers connected to the TRIO gene promoter by the indicated hmsLRI.

Figure S4

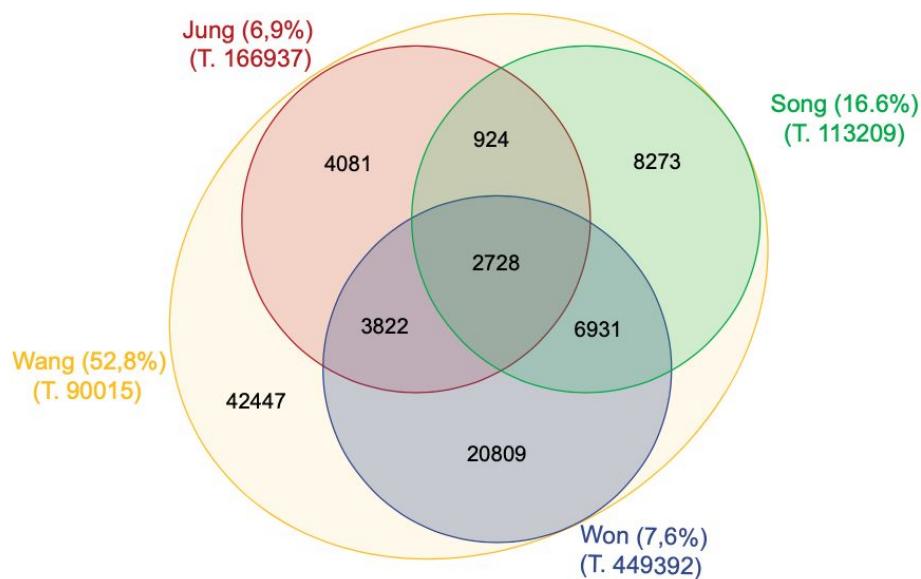


Figure S4

The Venn diagram shows how many human HiC-derived enhancer-promoter interactions from PsychENCODE Consortium (Wang et al 2018) were also identified by three independent studies (Won et al 2016, Jung et al 2019 and Song et al 2020). For each study, the total number of interactions and the percentage of those in common with the rest of the studies are shown in parentheses.