

Rare variation in noncoding regions with evolutionary signatures contributes to autism spectrum disorder risk

Taehwan Shin^{a,1}, Janet H.T. Song^{a,1}, Michael Kosicki^c, Connor Kenny^a, Samantha G. Beck^a, Lily Kelley^b, Xuyu Qian^a, Julieta Bonacina^b, Frances Papandile^a, Irene Antony^a, Dilenny Gonzalez^a, Julia Scotellaro^b, Evan M. Bushinsky^a, Rebecca E. Andersen^a, Eduardo Maury^a, Len A. Pennacchio^c, Ryan N. Doan^{b,*}, Christopher A. Walsh^{a,*}

^aDivision of Genetics and Genomics, Boston Children's Hospital; Departments of Pediatrics and Neurology, Harvard Medical School; Allen Discovery Center for Human Brain Evolution; Howard Hughes Medical Institute, Boston Children's Hospital, Boston, MA, 02115, USA

^bDivision of Genetics and Genomics, Boston Children's Hospital; Department of Pediatrics, Harvard Medical School; Allen Discovery Center for Human Brain Evolution, Boston, MA, 02115, USA

^cEnvironmental Genomics & Systems Biology Division, Lawrence Berkeley National Laboratory, Berkeley, CA, USA

Abstract

Little is known about the role of noncoding regions in the etiology of autism spectrum disorder (ASD). We examined three classes of noncoding regions: Human Accelerated Regions (HARs), which show signatures of positive selection in humans; experimentally validated neural Vista Enhancers (VEs); and conserved regions predicted to act as neural enhancers (CNEs). Targeted and whole genome analysis of >16,600 samples and >4900 ASD probands revealed that likely recessive, rare, inherited variants in HARs, VEs, and CNEs substantially contribute to ASD risk in probands whose parents share ancestry, which enriches for recessive contributions, but modestly, if at all, in simplex family structures. We identified multiple patient variants in HARs near *IL1RAPL1* and in a VE near *SIM1* and showed that they change enhancer activity. Our results implicate both human-evolved and evolutionarily conserved noncoding regions in ASD risk and suggest potential mechanisms of how changes in regulatory regions can modulate social behavior.

Keywords: autism spectrum disorder, noncoding regions, Human Accelerated Regions, Vista Enhancers, conserved neural enhancers, caMPRA, recessive variants, consanguineous families, *IL1RAPL1*, *SIM1*

Introduction

Autism spectrum disorder (ASD) is a highly heritable, phenotypically complex condition that affects 2-3% of children (Maenner et al., 2023) and shares comorbidity with many conditions including intellectual disability, attention-deficit/hyperactivity disorder, and obesity (Hyman et al., 2020). Over the last decade, immense progress has been made in understanding the genetic underpinnings of ASD. This has been largely driven by investigating *de novo* coding variants (De Rubeis et al., 2014; Iossifov et al., 2014; Zhou et al., 2022), and more recently, rare, recessive, inherited coding variants (Doan et al., 2019; Ruzzo et al., 2019; Zhou et al., 2022) of moderate to large effect size. Together, these efforts have identified more than 1000 candidate genes (Abrahams et al., 2013), with many identified ASD genes converging on similar gene programs, including synapse formation and maintenance, chromatin remodeling, and cytoskeletal pathways (De Rubeis et al., 2014; Ruzzo et al., 2019; Satterstrom et al., 2020).

Despite advances in understanding the role of coding variation in ASD, little is known about the role of noncoding varia-

tion. One major obstacle is that 98.5% of the genome is noncoding, and a systematic analysis of the entire noncoding genome requires a commensurately larger sample size to reach statistical significance. To address this issue, a number of studies have reduced the noncoding sequence search space to focus on noncoding regions that are likely to be functional and then queried whether specific classes of noncoding regions are enriched for patient variants. Evolutionary conservation has emerged as a strong marker of likely functional regions; many conserved noncoding regions are known to function as developmental enhancers, and disease-associated variants in these regions have been shown to disrupt gene regulation during development (Polychronopoulos et al., 2017). Indeed, recent studies found that *de novo* variants in conserved promoters are enriched in patients with ASD (An et al., 2018) and that *de novo* variants in conserved fetal brain enhancers are enriched in patients with severe neurodevelopmental disorders (Short et al., 2018). Consanguineous families, which are enriched for recessive contributions because of shared ancestry, have also proved powerful for identifying the contribution to ASD of noncoding regions, including inherited, homozygous deletions, which have not been detectable in nonconsanguineous families (Morrow et al., 2008; Schmitz-Abe et al., 2020).

Concurrently, multiple studies suggest that noncoding regions that show evolutionary signatures of selection in humans

*Corresponding authors

Email addresses: ryan.doan@childrens.harvard.edu (Ryan N. Doan), christopher.walsh@childrens.harvard.edu (Christopher A. Walsh)

¹These authors contributed equally. **NOTE: This preprint reports new research that has not been certified by peer review and should not be used to guide clinical practice.**

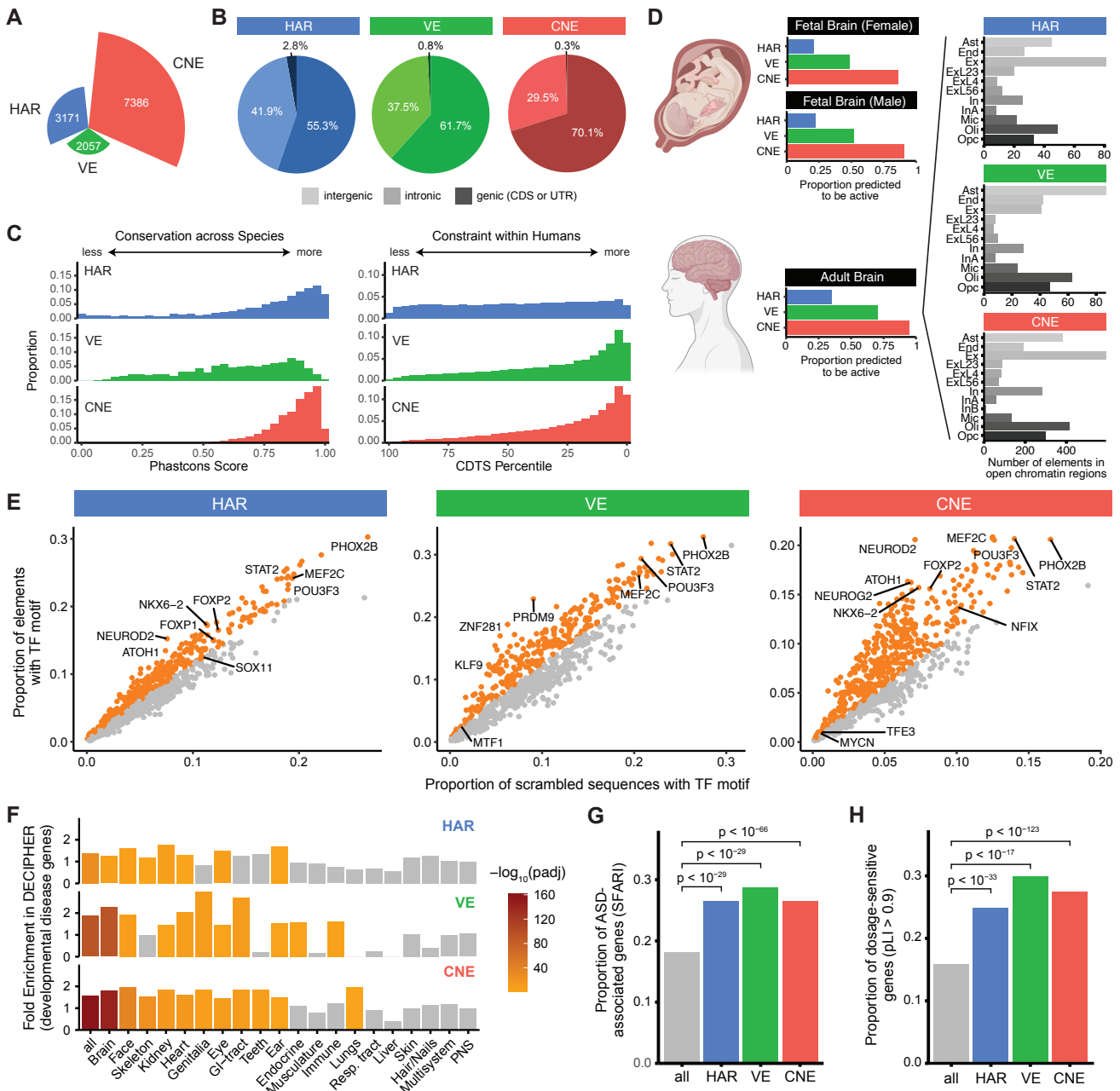


Figure 1: Genomic and epigenomic features of HARs, VEs, and CNEs. (A) Number of HARs, VEs, and CNEs. (B) Proportion of HARs, VEs, and CNEs in intergenic (light coloring), intronic (moderate coloring), and genic (dark coloring) regions. (C) Conservation across species (left) and constraint within humans (right) of HARs, VEs, and CNEs are represented by phastCons score (Siepel et al., 2005) and CDTS percentile (di Iulio et al., 2018), respectively. (D) Proportion of HARs, VEs, and CNEs predicted to be active by ChromHMM based on epigenomic data from a fetal male brain, a fetal female brain, and an adult brain (Kundaje et al., 2015) (left). Number of HARs, VEs, and CNEs that overlap open chromatin regions from scTHS-seq across cell types in the adult brain (Lake et al., 2018). Ast: astrocytes, End: endothelial cells, Ex: excitatory neurons, ExL23: layers 2-3 excitatory neurons, ExL4: layers 4 excitatory neurons, ExL56: layers 5-6 excitatory neurons, In: inhibitory neurons, InA: inhibitory neurons subtype A, InB: inhibitory neurons subtype B, Mic: microglia, Oli: oligodendrocytes, Opc: oligodendrocyte precursor cells. (E) Enrichment of transcription factor (TF) binding site motifs in HARs, VEs, and CNEs compared to nucleotide-matched scrambled sequences (Materials and Methods). Orange dots indicate significantly enriched elements at 5% FDR. (F) Enrichment of HARs, VEs, and CNEs near genes associated with developmental diseases in different body systems from the DECIPHER consortium (Firth et al., 2009). (G) HARs, VEs, and CNEs are enriched near ASD-associated genes annotated in the SFARI database (Abrahams et al., 2013). (H) Genes near HARs, VEs, or CNEs are enriched for genes with pLI > 0.9. Genes with pLI > 0.9 are considered loss-of-function intolerant (Lek et al., 2016).

may be preferentially vulnerable in human diseases (Oksenberg et al., 2013; Xu et al., 2015; Doan et al., 2016; Srinivasan et al., 2016; Song et al., 2018). For instance, human accel-

erated regions (HARs) are regions that are highly conserved across species, but show signals of positive selection in the human evolutionary lineage (Pollard et al., 2006; Bird et al., 2007;

Prabhakar et al., 2008; Lindblad-Toh et al., 2011; Gittelman et al., 2015; Girskis et al., 2021). HARs have been found to be enriched near genes associated with brain development (Capra et al., 2013; Kamm et al., 2013; Oksenberg et al., 2013; Boyd et al., 2015), and rare, recessive variants in HARs are enriched in patients with ASD in consanguineous families (Doan et al., 2016).

Here, we characterize three classes of noncoding regions that vary in evolutionary conservation and selection, examine their functional activity, and find that rare, inherited, likely recessive variants in all three classes are enriched, to varying degrees, in ASD cases compared to controls. We find that family structure strongly affects the contribution of rare, recessive variants across consanguineous, non-consanguineous multiplex, and non-consanguineous simplex cohorts. These ASD cohorts nominate a set of rare, recessive patient variants for further study, and we demonstrate that patient variants near the ASD-associated gene *ILIRAPL1* and the neurobehavioral gene *SIMI* regulate gene expression, suggesting that they may affect ASD risk.

Results

HARs, CNEs, and VEs may act as regulatory elements in the brain

Based on prior studies that suggest that regions that are highly conserved or under selection in humans may be selectively vulnerable in neurodevelopmental diseases (Doan et al., 2016; An et al., 2018; Short et al., 2018), we identified three classes of noncoding regions for characterization (Fig. 1, Table S1): (1) HARs, which are regions conserved through other mammals that are likely under positive selection in humans (Girskis et al., 2021) and which were previously shown to have elevated rates of rare, recessive variants in a consanguineous ASD cohort (Doan et al., 2016); (2) neural Vista Enhancers (VEs), which are conserved elements that have been experimentally tested to drive reporter activity in the brain in E11.5 transient transgenic reporter mice (Visel et al., 2007); and (3) conserved neural enhancers (CNEs). We defined CNEs as elements that are highly conserved across species, are highly constrained within humans, and are predicted to be enhancers in fetal brain and neurospheres or adult brain by ChromHMM from the Roadmap Epigenomics Project (Kundaje et al., 2015) (Materials and Methods).

Comparison of genomic and epigenomic features of HARs, VEs, and CNEs reveals similarities and differences in conservation, mutational constraint, and predicted functional activity. Most HARs and CNEs are highly conserved across species, whereas VEs exhibit variability in their level of conservation (Fig. 1C), likely because VEs often contain conserved segments flanked by stretches of less conserved sequences (Visel et al., 2007). In contrast, most VEs and CNEs are highly constrained within humans, whereas HARs, which could have either gained or lost functional activity in humans, exhibit variability in their levels of mutational constraint (Fig. 1C). As expected given our definition of CNEs and VEs, nearly all CNEs and the majority of VEs are predicted to be active in fetal or adult hu-

man brain by ChromHMM (Kundaje et al., 2015) (Fig. 1D). However, HARs, which were defined solely from genomic sequence changes, are less likely to be predicted to be active in fetal (~20%) or adult (~35%) brain. Substantial proportions of HARs, VEs, and CNEs are also predicted to be active in other tissues by ChromHMM (Fig. S1), and have differing cell type specificity in the adult brain (Lake et al., 2018) (Fig. 1D). HARs, VEs, and CNEs are all enriched for transcription factor (TF) binding sites of known neurodevelopmental TFs (Fig. 1E), including *FOXP2* in HARs and CNEs (den Hoed et al., 2021) and *ZNF281* in VEs (Pieraccioli et al., 2018), although only CNEs are enriched in aggregate for the motifs of TFs involved in neural functions (Table S1).

HARs, CNEs, and VEs are enriched near ASD-associated and dosage-sensitive genes

We might expect that if HARs, VEs, and CNEs modulate ASD risk, they would directly regulate the expression of genes previously implicated in ASD or other neurodevelopmental disorders. We find that HARs, VEs, and CNEs are enriched near genes implicated in severe developmental disorders that affect the brain, as annotated by the DECIPHER consortium (Firth et al., 2009) (Fig. 1F). We also observe a strong enrichment of HARs, VEs, and CNEs near ASD-associated genes, as annotated by SFARI (Abrahams et al., 2013) (adjusted $p < 10^{-29}$; Fig. 1G).

Given the restricted effect of a single regulatory element on gene expression (Cannavò et al., 2016; Osterwalder et al., 2018), noncoding regions that contribute to ASD risk might preferentially regulate genes that are dosage-sensitive, i.e., genes where a small change in expression can lead to a phenotypic outcome. As a measure of dosage sensitivity, we examined loss-of-function intolerance (Lek et al., 2016; Karczewski et al., 2020). ASD-associated genes (SFARI), which have been primarily identified from *de novo* heterozygous coding variants (Iossifov et al., 2014; Zhou et al., 2022), are strongly enriched for dosage-sensitive genes as expected (Fig. S3). Strikingly, HARs, VEs, and CNEs are all also significantly enriched near dosage-sensitive genes (adjusted $p < 10^{-17}$; Fig. 1H).

VEs and CNEs are more likely than HARs to act as enhancers in neural cells

To directly test whether HARs, VEs, and CNEs can act as enhancers, we used a capture-based Massively Parallel Reporter Assay (caMPRA) (Fig. 2A) (Girskis et al., 2021). Unlike conventional, oligonucleotide synthesis-based MPRA methods that can only test ~150-200bp sequences cost-effectively, caMPRA can test thousands of long (~500bp) noncoding sequences in parallel. This is critical because 60.6% of HARs, 34.7% of conserved cores of VEs, and 39.3% of CNEs are >200bp in length; conversely, 91.2% of HARs, 74.6% of conserved cores of VEs, and 92.9% of CNEs are <500bp in length (Materials and Methods; Fig. S4). Using this method, we tested HARs, VEs, and CNEs for regulatory activity in Neuro2A (N2A) cells, a neuroblastoma cell line that has been previously used to assess the neural function of noncoding regions (Doan et al., 2016; Mulvey and Dougherty, 2021; Arora et al., 2022; Mikl et al., 2022)

A capture-based Massively Parallel Reporter Assay (caMPRA) workflow

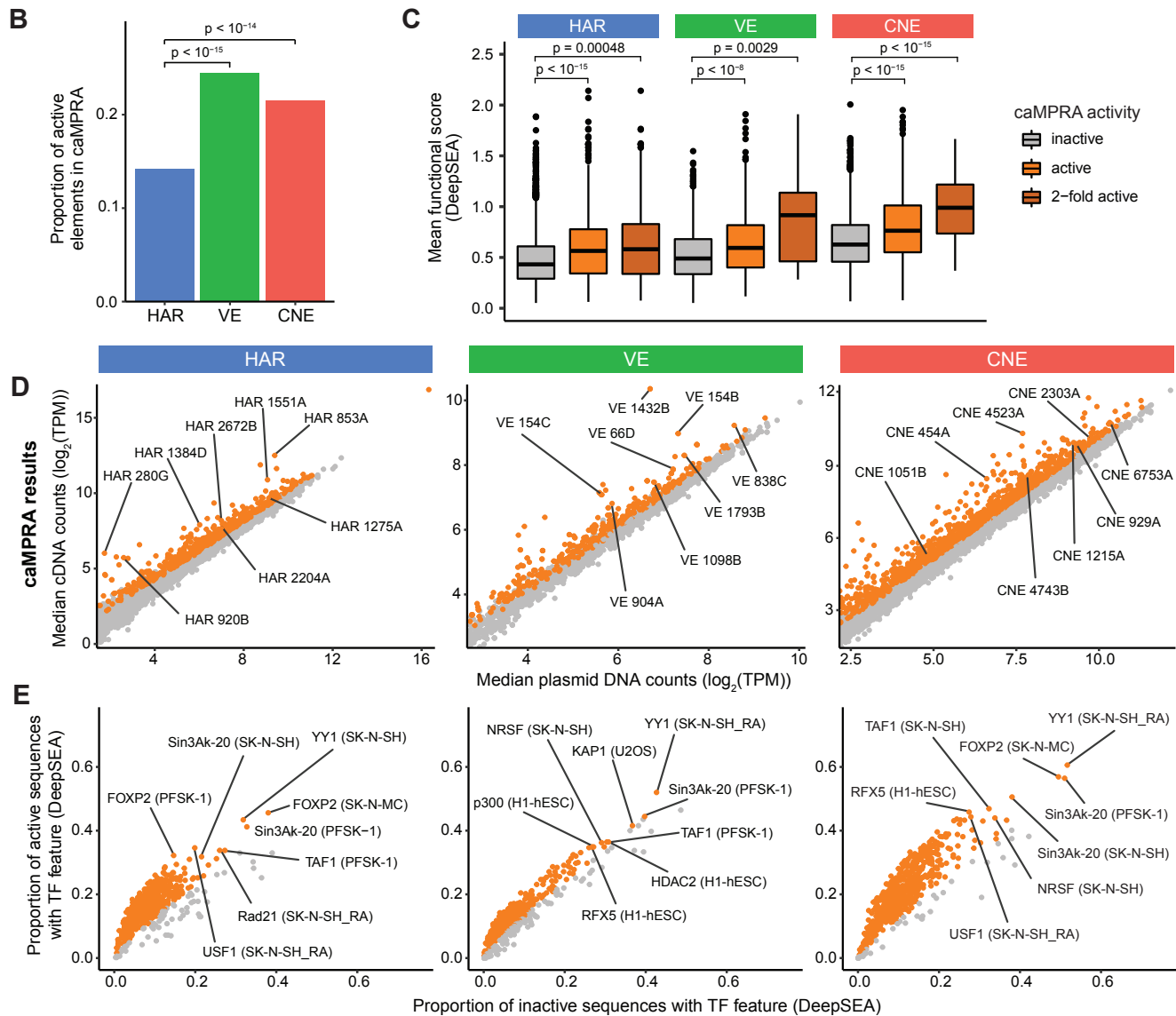
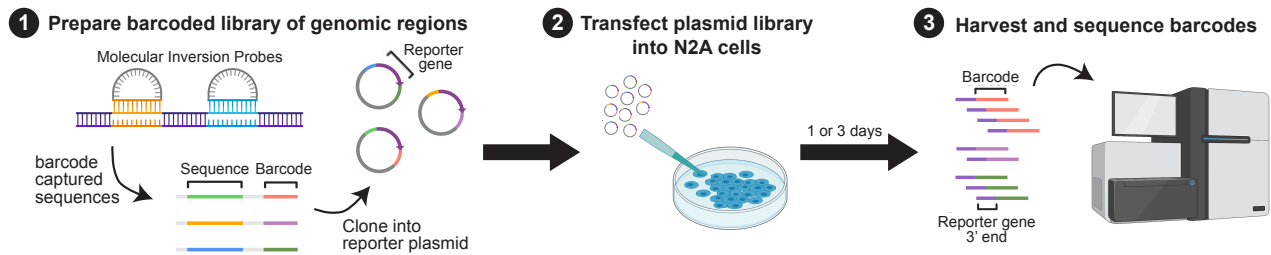


Figure 2: **HARs, VEs, and CNEs display enhancer activity in a capture-based Massively Parallel Reporter Assay (caMPRA).** (A) Schematic of caMPRA method. Sequences of interest are captured by molecular inversion probes (MIPs), barcoded, and cloned upstream of a minimal promoter driving luciferase expression. ~500bp sequences are captured by separate MIP probes per HAR, VE, or CNE element. The enhancer reporter plasmid library is then transfected into N2A cells and cells are harvested one (D1) or three (D3) days after transfection. Transcribed barcodes and barcodes from the original plasmid library are sequenced to examine enhancer activity. The results from the D3 caMPRA experiment are shown in this figure, and the results from the D1 caMPRA experiment are shown in Fig. S6. (B) Proportion of VEs or CNEs that have enhancer activity in at least one captured sequence is significantly higher than HARs by the chi-square test after FDR correction. (C) Sequences captured from HARs, VEs, and CNEs are classified as inactive, active, or 2-fold active and compared to their predicted mean functional score from DeepSEA (average of $-\log_{10}(e\text{-value})$ for every feature) (Zhou and Troyanskaya, 2015). P-values were determined with the hypergeometric test and adjusted by FDR correction. (D) Normalized cDNA counts vs normalized plasmid counts for sequences captured from HARs, VEs, and CNEs. Sequences with significant enhancer activity are in orange. (E) TF features were predicted by DeepSEA for each captured sequence. TF features significantly enriched in active sequences by caMPRA are shown in orange. Representative TF features are marked in the format: TF (cell type).

(Materials and Methods). Enhancer activity was highly correlated across replicates (Fig. S5, Fig. S6F).

Significantly more VEs (24.4%) and CNEs (21.5%) had enhancer activity compared to HARs (14.2%) in N2A cells ($p < 10^{-14}$; Fig. 2B). This finding is consistent with the definition of VEs and CNEs as experimentally validated or predicted enhancers respectively, whereas HARs are defined solely from genomic sequence changes in the human lineage. Many of the HARs, VEs, and CNEs that have enhancer activity are located near ASD-associated genes, including HAR3091 within *ILIRAPL1*, VE2019 near *ARX*, and CNE6536 near *ASXL3* (Abrahams et al., 2013). Active elements are also enriched for the motifs of neurodevelopmental TFs including FOXP2 and TAF1 (Fig. 2E, Fig. S6E) when assessed with DeepSEA, a deep learning model trained to predict thousands of features including TF binding (Zhou et al., 2018). Although N2A cells were not among the cell lines used for training and prediction in DeepSEA, the TF enrichment predictions from DeepSEA for active elements were specific for cell lines similar to N2A cells, including the neuroblastoma cell lines SK-H-SH and SK-N-MC and the neuroectodermal cell line PFSK-1, demonstrating the specificity of our assay for the TF milieu present in these cells. Further, we observed strong concordance between caMPRA-based activity and the predicted functional score from DeepSEA for HARs, VEs, and CNEs, particularly for sequences that exhibit a 2-fold increase in enhancer activity by caMPRA (Fig. 2C, Fig. S6C).

High-throughput mutagenesis of HARs causes gains as well as losses of enhancer activity

To determine whether single nucleotide variants in these non-coding regions could result in functional consequences in patients with ASD, we next sought to examine whether variants in these noncoding regions can affect regulatory activity. We focused on HARs and modified the caMPRA protocol to sparsely incorporate random variants into captured sequences using an error-prone PCR. Overall, we assessed 1,281 variants in 485 HARs across five replicate experiments (Materials and Methods; Fig. 3A, Fig. S8, Table S3).

Whereas most tested variants (81.5%) did not significantly alter regulatory activity (Fig. 3B, C), in general agreement with studies of other regulatory elements (Kircher et al., 2019; Snetkova et al., 2021), those that did were as likely to increase activity (10.8%) as to decrease it (7.6%). This is in contrast to prior mutagenesis studies where most variants that changed regulatory activity decreased expression (Kircher et al., 2019; Snetkova et al., 2021). It is possible that this difference reflects that prior studies mutagenized known enhancer elements, including VEs, whereas HARs likely comprise a mix of regulatory elements - including enhancers, repressors, and regulatory regions active in other species that have reduced or lost activity in humans. When we examine only sequences that contain a single introduced random variant, the proportion of variants that change expression and the distribution of fold changes were similar to sequences with more than 1 introduced variant (Fig. S7C, D), suggesting that single base pair changes in HARs

can have profound effects on both gains and losses of enhancer activity.

Rare, recessive variants in HARs, VEs, and CNEs are enriched in individuals with ASD in a consanguineous cohort

To examine whether HARs, VEs, and CNEs contribute to ASD risk, we examined whether there is an excess of rare, recessive variation in HARs, VEs, and CNEs in patients with ASD. Given the redundancy of regulatory networks even for highly conserved noncoding regions (Cannavò et al., 2016; Osterwalder et al., 2018), we reasoned that the bulk of our candidate enhancer sequences may act in a recessive manner, rather than via the *de novo* mode of contribution of highly constrained dominant genes. We first revisited a consanguineous cohort, the Homozygosity Mapping Collaborative for Autism (HMCA) (Morrow et al., 2008), where we had previously observed an enrichment of rare, recessive variants in HARs in ASD cases compared to controls using targeted sequencing (Doan et al., 2016). This enrichment was seen only when examining rare variants that were predicted to be damaging by conservation-based variant effect predictors (Doan et al., 2016).

When we now examine an expanded set of 3,171 HARs using whole-genome sequencing (WGS) on a larger number of families from HMCA (a total of 662 individuals including 193 probands), we continue to identify a strong enrichment of rare, recessive variants in HARs in cases compared to matched controls ($OR = 2.142$, adjusted $p = 0.001$; Fig. 4B). We defined recessive variants as variants that are homozygous, compound heterozygous, or hemizygous (specifically in male individuals for the X chromosome). Because hemizygous variants on the male X chromosome are much more likely to arise compared to homozygous variants on the female X chromosome, we only examined the autosomes when calculating combined rates for males and females, but included the X chromosome when analyzing males and females separately. As in the prior study, we observe an enrichment only when examining rare variants that are predicted to be damaging by conservation-based variant effect predictors (Materials and Methods; referred to hereafter as “conserved bases”), but not when examining non-conserved bases (adjusted $p > 0.05$).

VEs also had a large excess of rare, recessive variants in conserved bases when comparing cases to controls that was similar in magnitude to the excess seen in HARs ($OR = 2.074$, adjusted $p = 0.004$; Fig. 4B), whereas CNEs had a significant, but less pronounced excess of rare, recessive variants in cases compared to controls ($OR = 1.546$, adjusted $p = 0.041$; Fig. 4B). The enrichment of rare, recessive variants in HARs, VEs, and CNEs is stable across a range of low allele frequencies, suggesting that the signal we observe is not dependent on specific allele frequency cut-offs (Fig. S9A). Although we are underpowered to assess significance when each sex is analyzed separately, similar excesses in rare, recessive variants are observed in both males and females (Fig. S9B).

The observed rates of rare, recessive variants between cases and controls (0.239 versus 0.128 for HARs, 0.216 versus 0.117 for VEs, 0.414 versus 0.313 for CNEs) yield substantial estimated contributions to ASD of 9.9%, 11.1%, and 10.0% for

A capture-based Massively Parallel Reporter Assay (caMPRA) with mutagenesis workflow

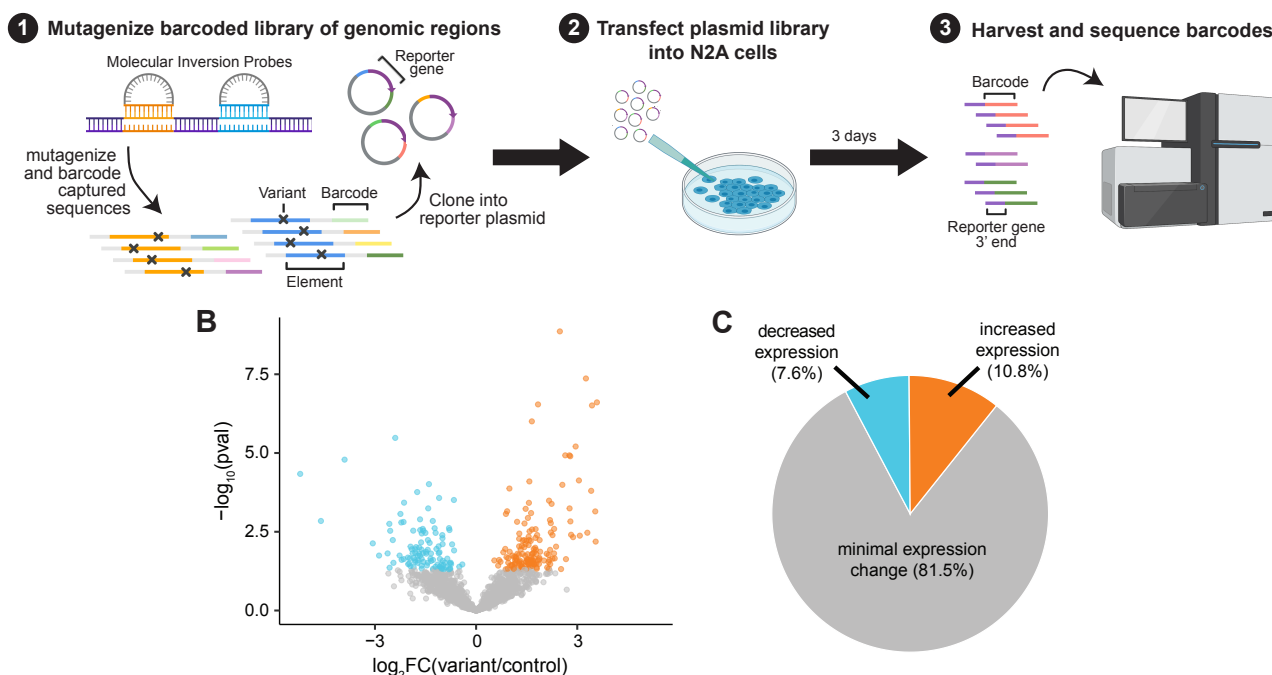


Figure 3: **Random variants in HARs can modulate enhancer activity.** (A) Schematic of caMPRA with random mutagenesis. (B) Volcano plot of fold change in expression and adjusted p-value for each mutagenized sequence. (C) Pie chart of percent of mutagenized sequences with decreased expression, increased expression, or no statistically significant change in expression.

recessive alleles in HARs, VEs, and CNEs, respectively (Materials and Methods). Together with a previous finding in this cohort of a $\sim 4x$ excess of rare, homozygous, inherited deletions in noncoding, but not in coding, genomic regions (Schmitz-Abe et al., 2020), our results suggest that homozygous noncoding variation in this cohort contributes significantly to ASD risk by several mechanisms, and is also consistent with a relatively modest contribution of recessive exonic mutations in this cohort (Yu et al., 2013).

Rare, recessive variants in HARs and VEs are enriched in individuals with ASD in non-consanguineous cohorts

We then examined whether the enrichment of rare, recessive variants in HARs, VEs, and CNEs is also observed in a larger, non-consanguineous cohort from the NIMH repository. We expect effect sizes for recessive variants to be considerably smaller in non-consanguineous cohorts compared to consanguineous cohorts where both direct consanguinity and endogamy make it more likely that the same rare variant is inherited from both parents (Bittles and Black, 2010). However, compared to the slightly less than 200 probands in the consanguineous cohort, the NIMH repository contains >2000 affected probands, offering greater resolution to detect small differences in recessive variants and to identify a larger set of patient variants for functional studies (Fig. 4A). We examined 660 probands from multiplex families, where inherited variants are more likely to play a role (Ruzzo et al., 2019; Ciriigliaro et al., 2023), and 1911 probands from families with only 1 affected child (either with or without an unaffected sibling). The

latter are likely to be simplex families, where recessive variants have a lower contribution to disease compared to *de novo* mechanisms (Sebat et al., 2007; Marshall et al., 2008).

Targeted sequencing of HARs, VEs, and CNEs with molecular inversion probes (Materials and Methods) showed a non-significant excess of rare, recessive variants in HARs and VEs at conserved bases when considering males and females jointly (HARs: $OR = 1.196$, adjusted $p = 0.186$; VEs: $OR = 1.193$, adjusted $p = 0.069$; Fig. S10A), while males considered alone, which captures hemizygous variants on chromosome X, revealed significant enrichment for rare, recessive variants in both HARs and VEs at conserved, but not at less conserved, bases (HARs: $OR = 1.262$, adjusted $p = 0.050$; VEs: $OR = 1.255$, adjusted $p = 0.030$; Fig. 4C) across allele frequency cut-offs (Fig. S11). In contrast, CNEs were not enriched for rare, recessive variants in cases versus controls for males (males: $OR = 0.909$, adjusted $p = 1$; Fig. 4C). Because females are much less likely than males to be diagnosed with ASD (Robinson et al., 2013), we have a much smaller number of female individuals in this cohort and were under-powered to examine females alone given the odds ratios observed in males (Fig. S10B). Notably, the odds ratios are similar when comparing males and females jointly or males separately. Consistent with the effect of family structure on the contribution of recessive variants to ASD risk, we also observe a larger effect size in multiplex families (HARs: $OR = 1.645$, VEs: $OR = 1.378$) compared with the likely simplex families (HARs: $OR = 1.189$, VEs: $OR = 1.180$) in males (Fig. S10C).

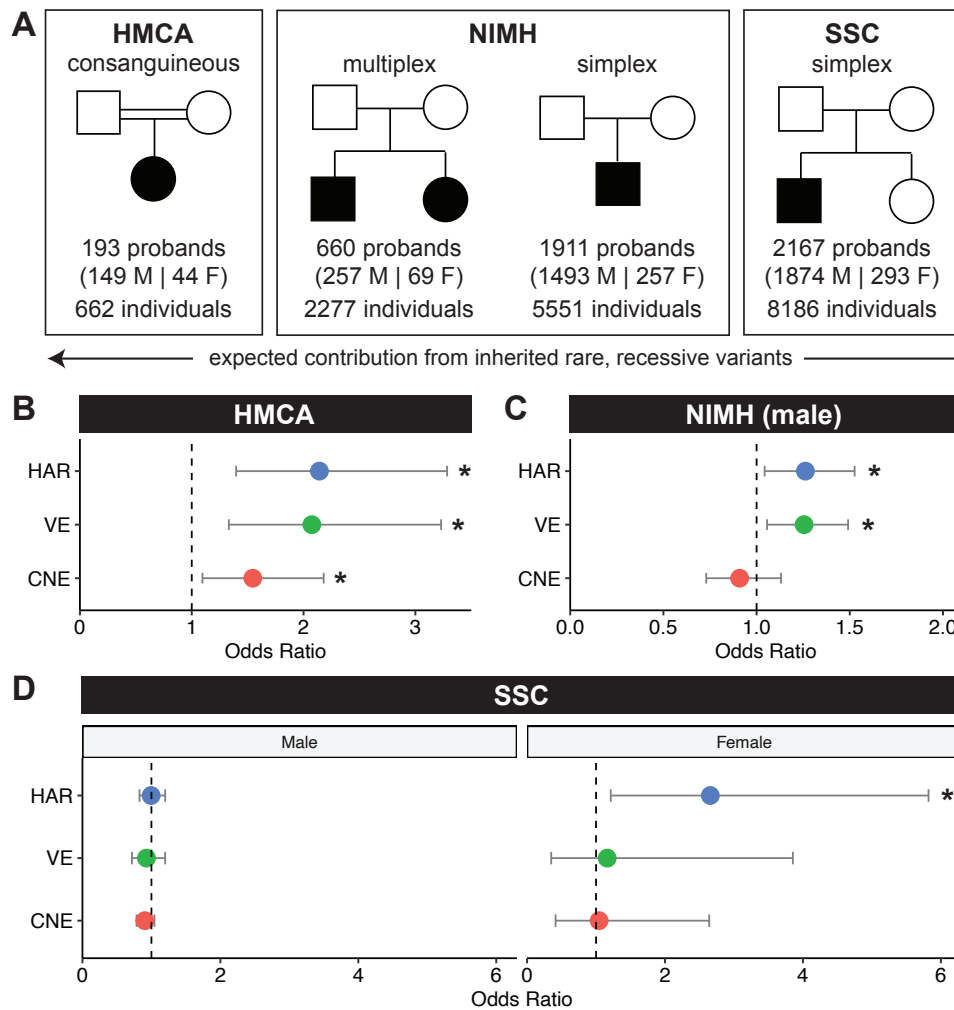


Figure 4: The contribution of rare, recessive variants in HARs, VEs, and CNEs to ASD varies across cohorts based on family structure. (A) We examined HARs, VEs, and CNEs in three cohorts: 1. HMCA, a consanguineous cohort; 2. NIMH, a cohort that includes multiplex and simplex families; 3. SSC, a cohort that only contains simplex families. (B) In the HMCA cohort, cases are enriched for rare, recessive variants in HARs (adjusted $p = 0.0014$), VEs (adjusted $p = 0.0038$), and CNEs (adjusted $p = 0.0412$) at allele frequency (AF) < 0.005 . (C) In the NIMH cohort, male cases are enriched for rare, recessive variants in HARs (adjusted $p = 0.0495$) and VEs (adjusted $p = 0.0297$) at $AF < 0.001$. Combined autosomal rates for males and females were not significant, and there was insufficient sample size to assess females separately (Fig. S10A, B). (D) In the SSC cohort, female cases are enriched for rare, recessive variants in HARs (adjusted $p = 0.0438$) at $AF < 0.005$. All analyses are done on conserved bases, and allele frequency cut-offs were chosen based on predefined heuristics (Materials and Methods). Odds ratios are consistent across different AF s (Fig. S9A, Fig. S11, Fig. S12). Statistical analyses are detailed in Materials and Methods.

The rates of rare, recessive variants between male cases and controls are 0.141 versus 0.115 for HARs and 0.213 versus 0.181 for VEs, resulting in an estimated contribution of recessive alleles in HARs and VEs to 2.6% and 3.7% of ASD cases respectively. This contribution is similar to the 3-5% contribution of rare, recessive coding variants to ASD cases in a similar cohort (Doan et al., 2019), but smaller than the estimated ~10% contribution of rare, recessive alleles in HARs, VEs, and CNEs to ASD cases we observed in the consanguineous cohort. These results highlight the unique suitability of consanguineous cohorts to examine noncoding variation, but nevertheless confirm the contribution of HARs and VEs to ASD and provide a large set of rare, potentially damaging variants for further study (Table 1, Table S4).

We next examined the Simon Simplex Collection (SSC), which consists of 8,186 individuals with WGS data and is

specifically limited to simplex families with a single proband and unaffected siblings (Fischbach and Lord, 2010) (Fig. 4A). In such a cohort, recessive effects are expected to be minor, and potentially undetectable, compared to the consanguineous HMCA cohort and the NIMH cohort that contained multiplex families (Sebat et al., 2007; Yu et al., 2013). Indeed, we observed no excess of rare, recessive variants in HARs, VEs, or CNEs in almost all comparisons. When examining the cohort separated by sex, we do find a significant excess of rare, recessive variants in HARs in female ASD cases in conserved, but not at less conserved, bases ($OR = 2.657$, adjusted $p = 0.044$; rate of rare, recessive variants is 0.027 in cases and 0.010 in controls for an estimated 1.7% contribution; Fig. 4D) across allele frequency cut-offs (Fig. S12), with no similar enrichment in males, despite there being 1874 male probands and only 293 female probands in SSC. An enrichment in females, but not in

Element	Cohort	Variants found in cases and not controls (hg38)	Number of cases with variants	Number of controls with variants	Potential target genes	Disease and functional associations
HAR1362	NIMH	chr2:44493977 (G->A) chr2:44494211 (G->A)	2	0	<i>CAMKMT</i> , <i>SIX3*</i> , <i>PREPL</i>	Required for development of anterior neural structures (<i>SIX3</i>) (Lagutin et al., 2003)
HAR1479	NIMH	chr2:145221015 (G->A) chr2:145221025 (C->A)	2	0	<i>ZEB2*</i> , <i>GTDC1</i> , <i>ARHGAP15</i>	Mutations cause Mowat-Wilson syndrome (<i>ZEB2</i>) (Epifanova et al., 2019)
HAR3094	NIMH	chrX:30371544 (G->A) chrX:30371553 (A->G)	2	0	<i>NR0B1*</i> , <i>CXorf21</i> , <i>IL1RAPL1*</i> , <i>MAGEB1</i> , <i>MAGEB2</i> , <i>MAGEB3</i>	Mutations associated with ASD and ID (<i>IL1RAPL1</i>) (Bhat et al., 2008; Mikhail et al., 2011)
HAR3134	HMCA, NIMH	chrX:122662679 (T->C) chrX:122662692 (A->G)	3	0	<i>GRIA3*</i>	Mutations associated with ASD, X-linked syndromic ID, and schizophrenia (<i>GRIA3</i>) (Wu et al., 2007; Guilmatre et al., 2009; Singh et al., 2022)
HAR3162	NIMH	chrX:144625836 (G->C) chrX:144625865 (A->G) chrX:144625878 (G->A) chrX:144625958 (G->A)	4	1	<i>SLITRK2</i> , <i>SLITRK4</i>	Mutations associated with ID, DD, and neuropsychiatric symptoms (<i>SLITRK2</i>) (El Chehadeh et al., 2022)
VE15	NIMH	chr1:10737260 (T->C) chr1:10737343 (G->C)	2	0	<i>CASZ1*</i>	Mutations associated with ASD, ID, and DD (<i>CASZ1</i>) (Coe et al., 2019)
VE162	NIMH	chr1:213425381 (T->C) chr1:213425533 (A->C)	2	0	<i>PROX1*</i> , <i>RPS6K1</i> , <i>SMYD2</i>	Regulates interneuron differentiation (<i>PROX1</i>) (Miyoshi et al., 2015)
VE235	NIMH	chr2:63049151 (C->A) chr3:147847042 (T->C)	1	0	<i>OTX1*</i>	Mutations associated with ASD (Liu et al., 2011)
VE462	NIMH	chr3:147847133 (T->C) chr3:147847216 (T->C)	3	0	<i>ZIC1*</i> , <i>ZIC4</i>	Involved in medial telencephalon development (<i>ZIC1</i>) (Inoue et al., 2007)
VE644	NIMH	chr5:88396771 (G->T) chr5:88397035 (C->T)	2	0	<i>MEF2C*</i> , <i>TMEM161B</i>	Mutations associated with ASD (<i>MEF2C</i>) (Novara et al., 2010); Mutations associated with polymicrogyria (<i>TMEM161B</i>) (Akula et al., 2023; Wang et al., 2023)
CNE6445	HMCA	chr17:69607668 (C->T)	2	0	<i>KCNJ16</i> , <i>MAP2K6*</i> , <i>KCNJ2</i>	Member of MAP/ERK pathway, which has been linked to changes in social behavior (<i>MAP2K6</i>) (Albert-Gasco et al., 2020)
CNE7200	HMCA	chrX:18424091 (T->C)	2	0	<i>CDKL5*</i>	Mutations associated with Rett syndrome and epilepsy (<i>CDKL5</i>) (Weaving et al., 2004)

Table 1: **Examples of HARs, VEs, and CNEs that have more variants found in cases compared to controls.** Full list in Table S4. Asterisks indicate genes that are loss-of-function intolerant (pLI > 0.9) (Lek et al., 2016). Potential target genes were determined by gene proximity and by location within the same topologically associated domain (Dixon et al., 2012, 2015). For HAR3162, one variant was observed in both a case and a control individual in HMCA and was excluded from the table. Coordinates are in hg19. ID: intellectual disability. DD: developmental disorders.

males, may reflect the female protective effect, a phenomenon where female probands require a higher genetic burden (potentially including variants in HARs) than male probands to develop ASD (Gilman et al., 2011; Robinson et al., 2013; Jacquemont et al., 2014), and also parallels the larger contribution to ASD of recessive coding variants in females compared to males (Doan et al., 2019).

Our findings from multiple ASD cohorts indicate that rare, recessive variants in HARs, VEs, and CNEs contribute to ASD burden. As expected, the impact of these variants tracks the known contribution of recessive variants as a function of family structure, with by far the largest contribution seen in consanguineous families, a smaller contribution seen in male ASD cases in a cohort of non-consanguineous families that includes multiplex families, and a contribution that is only discernible in females among simplex families. Consanguineous families have been used to uncover noncoding variants underlying risk for other diseases (Schultz et al., 2009; Bae et al., 2014; Favaro et al., 2014; Tuncay et al., 2022), and our results suggest that further expanding consanguineous cohorts may be particularly suitable for identifying and analyzing noncoding contributions to disease.

Variants enriched in ASD patients implicate new genes in ASD risk

While we are underpowered to pinpoint specific HARs, VEs, or CNEs that are statistically enriched for patient variants, in-

dividual HARs, VEs, or CNEs with a numerical excess of rare, recessive variants in cases compared to controls represent potential candidates for further study, particularly since the number of controls far exceeds the number of cases in each cohort. We focused on rare, recessive variants enriched in ASD cases compared to controls in HARs, VEs, or CNEs from the HMCA cohort and in HARs or VEs from the NIMH cohort because there is a greater expected contribution of inherited variants from those cohorts compared to the simplex SSC cohort (Sebat et al., 2007; Bittles and Black, 2010; Yu et al., 2013).

HARs, VEs, and CNEs enriched for variants found in cases compared to controls are located near both ASD-associated genes and genes that have not been previously linked to ASD. Intriguingly, proteins encoded by many of the newly identified candidate genes are known to interact with proteins encoded by ASD-associated genes (Fig. S13A). This reflects recent studies that identify convergent effects on protein networks across multiple, distinct genetic models of ASD (Paulsen et al., 2022; Pintacuda et al., 2023; Li et al., 2023). Many newly identified candidate genes, as well as many genes previously associated with ASD, are also loss-of-function intolerant (blue circles indicate genes with pLI > 0.9 in Fig. S13A, B), suggesting that expression changes in specific cell types (such as through regulatory variants in HARs, VEs, or CNEs) may have significant phenotypic consequences.

We highlight a subset of HARs, VEs, and CNEs that are enriched for patient variants in Table 1 (full list in Table S4).

These include HAR3094 and VE644, which are near the ASD-associated genes *ILIRAPLI* (Bhat et al., 2008; Mikhail et al., 2011; Montani et al., 2019) and *MEF2C* (Novara et al., 2010) respectively. In contrast, HAR3162 (near *SLITRK2*), VE162 (near *PROXI*), and CNE6445 (near *MAP2K6*) are located near promising candidate genes that have not been previously associated with ASD. Mutations in *SLITRK2* result in moderate to severe intellectual disability with a range of behavioral and neuropsychiatric symptoms (El Chehadeh et al., 2022). In E11.5 embryonic mice, we find that HAR3162 has enhancer activity in the ventral telencephalon (Fig. S13C, D), where *SLITRK2* is expressed (Magdaleno et al., 2006; Diez-Roux et al., 2011), suggesting that HAR3162 may regulate *SLITRK2* expression. Similarly, VE162 has enhancer activity in the ventral telencephalon in E11.5 embryonic mice (Visel et al., 2007) and has been shown to physically interact with the promoter of *PROXI* (Song et al., 2020), a gene that regulates interneuron differentiation in the ventral telencephalon (Magdaleno et al., 2006; Miyoshi et al., 2015). Finally, CNE6445 is located near the loss-of-function intolerant MAP kinase gene *MAP2K6*. MAP/ERK signaling regulates many aspects of normal brain development, and dysregulation of this pathway has been associated with changes to social behaviors (Albert-Gascó et al., 2020).

HARs enriched for ASD patient variants regulate the neurodevelopmental gene ILIRAPLI

As an initial functional investigation into whether HARs, VEs, or CNEs that are enriched for patient variants might contribute to ASD risk, we characterized HAR3091 and HAR3094 (Fig. 5A). Both of these HARs reside near the gene *ILIRAPLI*, act as enhancers by caMPRA, and are enriched for patient variants. In the NIMH cohort, HAR3091 and HAR3094 each contained two variants in cases and none in controls at conserved bases (Table 1, Table S4).

ILIRAPLI is a gene important for synaptic density and dendrite formation at excitatory synapses (Montani et al., 2019) that is highly loss-of-function intolerant (Lek et al., 2016; Karczewski et al., 2020). Exonic point mutations, deletions, and duplications of *ILIRAPLI* have been associated with ASD and intellectual disability (Carrié et al., 1999; Tabolacci et al., 2006; Froyen et al., 2007; Piton et al., 2008; Firth et al., 2009; Behnecke et al., 2011; Franek et al., 2011; Du et al., 2018; Montani et al., 2019), suggesting that *ILIRAPLI* is dosage-sensitive to both gain and loss of expression. HAR3091 is located in the second intron of *ILIRAPLI*, and HAR3094 is located downstream of *ILIRAPLI* within the same topologically associated domain (Won et al., 2016).

Transgenic mouse analysis suggests that HAR3091 and HAR3094 have regionally restricted and species-specific enhancer activity in the developing brain. We generated transient transgenic mice with either the human or chimpanzee versions of HAR3091 or HAR3094 located upstream of a minimal promoter driving a *lacZ* reporter gene (Fig. 5B, Fig. S14). We harvested mouse embryos at E14.5, when *in situ* data show strong *ILIRAPLI* expression (Diez-Roux et al., 2011), and stained for *lacZ* expression (Materials and Methods). HAR3091 drives

lacZ expression predominantly in the telencephalon and olfactory bulb (arrowheads in Fig. 5B and Fig. S14), with more limited *lacZ* expression in the midbrain (asterisks in Fig. 5B and Fig. S14). In contrast, HAR3094 drives *lacZ* expression predominantly in the midbrain, with more limited expression in the telencephalon and the olfactory bulb. Interestingly, the chimpanzee version of HAR3091 drives much more robust activity in the telencephalon compared to the human version of HAR3091, whereas the human version of HAR3094 drives much more robust activity in the midbrain than the chimpanzee version of HAR3094. These results suggest that HAR3091 is primarily a telencephalon enhancer that has decreased activity in humans compared to chimpanzees, whereas HAR3094 is primarily a midbrain enhancer that has increased activity in humans compared to chimpanzees. Both HAR3091 and HAR3094 enhancer domains overlap with regions where *ILIRAPLI* is expressed at E14.5 (Diez-Roux et al., 2011) (Fig. 5B).

To directly test whether HAR3091 and HAR3094 might regulate *ILIRAPLI* expression, we used CRISPR inhibition (CRISPRi), which uses a nuclease-inactive Cas9 variant tethered to a KRAB domain (dCas9-KRAB) to heterochromatize and silence the target region (Gilbert et al., 2014). We induced *NGN2* expression (Materials and Methods) to differentiate human iPSCs into a heterogeneous mixture of excitatory neurons that resemble neurons derived from multiple brain regions, including the regions where HAR3091 and HAR3094 have enhancer activity (Lin et al., 2021). Whereas targeting the *ILIRAPLI* TSS significantly decreased *ILIRAPLI* expression compared to NTC gRNAs as expected (adjusted $p = 0.0001$; Fig. 5C), we also observed a significant decrease in *ILIRAPLI* expression when targeting HAR3094 (adjusted $p = 0.0002$), suggesting that HAR3094 acts as an *ILIRAPLI* enhancer. When targeting HAR3091, the median *ILIRAPLI* expression decreased nominally by 8.8% but did not reach statistical significance. Given that human HAR3091 acts as a weak enhancer in transgenic mice, it is possible that our CRISPRi assay lacked the required sensitivity to detect a significant decrease in expression, especially given the wide variability in gRNA efficacy (Fig. S15). These results suggest that HAR3094, and possibly also HAR3091, are *ILIRAPLI* enhancers in neurons, and that HAR3094 is a stronger enhancer than HAR3091.

Next, we asked whether patient variants may affect the enhancer activity of HAR3091 and HAR3094. Based on availability of patient DNA, we examined one of the two rare, recessive patient variants at conserved bases in HAR3091 and the two rare, recessive patient variants at conserved bases in HAR3094. In addition, we also examined an additional patient variant in HAR3091 and three additional patient variants in HAR3094 that are rare, recessive variants but at less conserved bases. HAR3091 or HAR3094 sequences containing these variants were cloned upstream of a minimal promoter driving luciferase expression, and luciferase activity was assessed in N2A cells (Materials and Methods). Strikingly, we find that patient variants for HAR3091 significantly increased luciferase activity compared to the control HAR3091 sequence, and that patient variants for HAR3094 significantly decreased

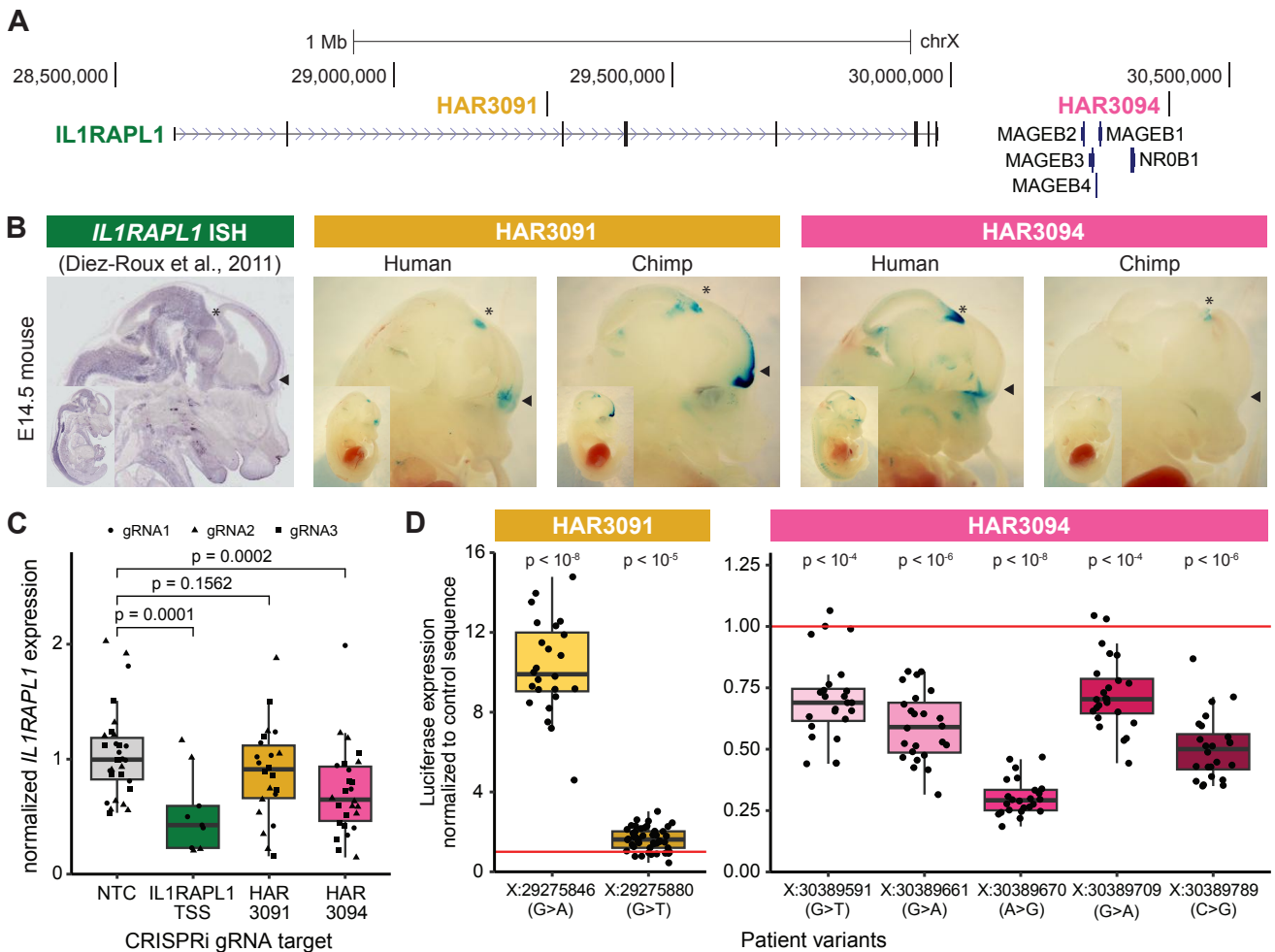


Figure 5: Patient variants in HAR3091 and HAR3094 likely regulate *ILIRAPL1* expression in multiple brain regions. (A) The genomic region containing *ILIRAPL1*, HAR3091, and HAR3094. (B) Constructs containing either the human or chimpanzee version of HAR3091 and HAR3094 cloned upstream of a minimal promoter driving lacZ expression were randomly integrated into mice and analyzed at E14.5. HAR3091 has enhancer activity predominantly in the telencephalon and olfactory bulb (filled arrowheads), and the human version of HAR3091 is a weaker enhancer than the chimpanzee version. In contrast, HAR3094 has enhancer activity predominantly in the midbrain (asterisks), and the human version of HAR3094 is a stronger enhancer than the chimpanzee version. Representative embryos are shown (all embryos are in Fig. S14). *In situ* hybridization of *ILIRAPL1* at E14.5 from the Eurexpress database (Diez-Roux et al., 2011) is shown for comparison. (C) CRISPRi targeting the *ILIRAPL1* TSS and HAR3094 significantly decrease *ILIRAPL1* expression compared to the non-targeting control (NTC) gRNAs in iPSC-derived neurons induced by *NGN2* expression. Multiple gRNAs were tested per target region. (D) Patient variants in HAR3091 and HAR3094 were tested for luciferase expression in N2A cells. HAR3091 patient variants significantly increased luciferase expression, whereas HAR3094 patient variants significantly decreased luciferase expression. Statistical analyses are detailed in Materials and Methods. Coordinates are in hg19.

luciferase activity compared to the control HAR3094 sequence (Fig. 5D, Fig. S16). The largest effect sizes were observed for the patient variants at conserved bases, consistent with the established link between conservation and functional activity and our finding that an excess of rare, recessive variants is observed in ASD cases compared to controls for conserved but not less conserved bases. These results indicate that patient variants modulate HAR3091 and HAR3094 enhancer activity and may result in changes to *ILIRAPL1* expression in specific brain regions.

ASD patient variants near SIM1, a human neurobehavioral gene, modulate in vivo enhancer activity

VE854, commonly referred to as hs576 (Visel et al., 2007), is an enhancer of the nearby obesity-associated gene *SIMI* (Ahi-

tuv et al., 2007; Bonnefond et al., 2013; Kim et al., 2014; Matharu et al., 2019) and contains two rare, recessive patient variants in the HMCA and NIMH cohorts (Fig. 6, Table S4). *SIMI* loss-of-function has been associated with obesity and neurobehavioral deficits; in one study that identified 13 obese individuals with rare, *de novo SIM1* mutations, 11 also presented with neurobehavioral abnormalities including ASD (Ramachandrapa et al., 2013). Genes downstream of *SIMI* have similarly been associated with both obesity and neurological phenotypes (Kasher et al., 2016), suggesting that modulating this pathway may contribute to obesity, ASD, and their comorbidity.

Hs576 has been previously found to drive enhancer activity in the developing brain, somites, and cranial nerves in transgenic E11.5 mice and in the forebrain and hippocam-

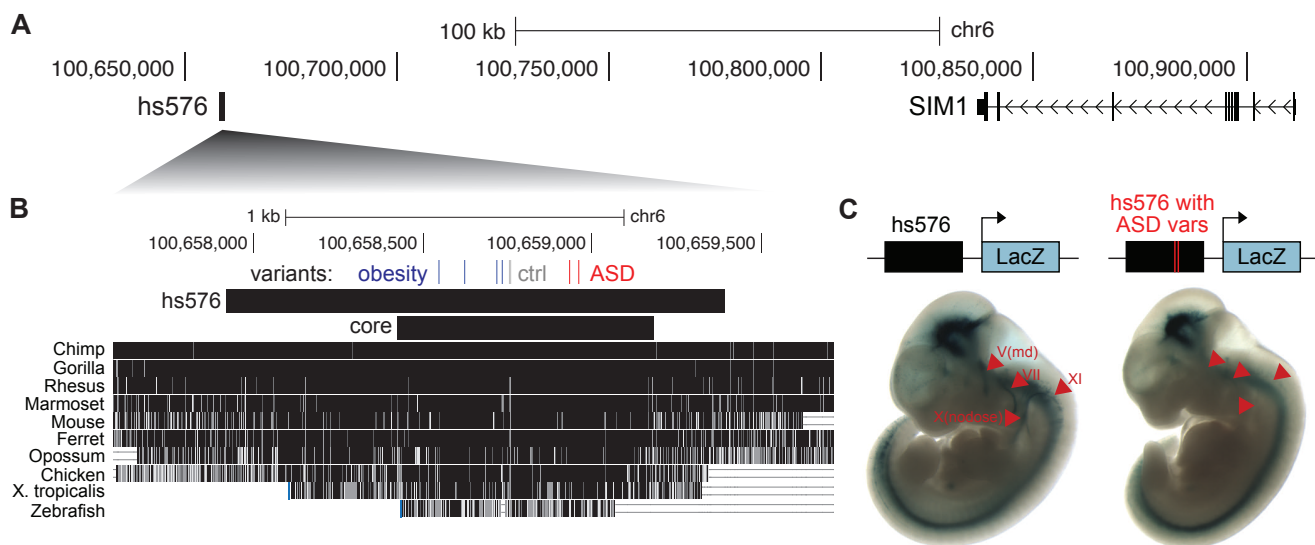


Figure 6: Patient variants in hs576 (VE854) reduce enhancer activity in cranial nerves. (A) The genomic region containing hs576 and *SIM1*. (B) The locations of two ASD patient variants from the HMCA and NIMH cohorts (ASD; red), two variants from one control individual (ctrl; gray), and four previously identified obesity-associated variants (blue) (Kim et al., 2014) are indicated. These variants are all located in the most conserved region of hs576 (core), which recapitulates most of the enhancer activity of the entire hs576 element (Kim et al., 2014). (C) Constructs containing hs576 without ($n = 4$) or with ($n = 6$) the two ASD patient variants upstream of a minimal promoter driving the lacZ gene were integrated into the safe-harbor H11 locus and analyzed for lacZ expression at E11.5 (Materials and Methods). Arrowheads indicate cranial nerves where the inclusion of the two ASD patient variants reduces enhancer activity. Representative embryos are shown (all embryos are in Fig. S17). Coordinates are in hg19.

pus in E14.5 mice, matching the expression pattern of *SIM1* (Visel et al., 2007; Kim et al., 2014). This enhancer activity is mainly derived from the most highly conserved portion of hs576 (“core” in Fig. 6B) (Kim et al., 2014). Intriguingly, rare variants associated with obesity (Kim et al., 2014) and our identified ASD patient variants are both located in this core region, albeit in separate clusters at the 5’ and 3’ ends respectively (Fig. 6B). There is also one control individual in the NIMH cohort that contains two neighboring variants in hs576 (Fig. 6B). To test whether the ASD patient variants affect the enhancer activity of hs576, we first confirmed its known expression pattern in E11.5 mice by integrating a construct containing hs576 upstream of a minimal promoter driving lacZ expression at the H11 safe-harbor locus (Materials and Methods; Fig. 6C). We then generated E11.5 transgenic mice where hs576 containing the two ASD patient variants was integrated at the H11 locus. Strikingly, we find that hs576 containing the two ASD patient variants had reduced or absent enhancer activity in multiple cranial nerves, in particular the mandibular portion of the trigeminal nerve (V), the facial nerve (VII), the nodose nerve / inferior part of vagus nerve (X), and the accessory nerve (XI) across multiple embryos (arrowheads in Fig. 6C; Fig. S17). These results suggest that ASD patient variants can alter *SIM1* expression and may contribute to ASD.

Discussion

In this study, we identify HARs, VEs, and CNEs as three classes of noncoding regions that are near ASD-associated and dosage-sensitive genes, and show contrasting degrees of evolutionary conservation, constraint within humans, and predicted

functional activity in the brain. Using caMPRA, we find that VEs and CNEs are significantly more likely than HARs to act as enhancers in N2A cells, and that single nucleotide variants in HARs can both increase and decrease enhancer activity. We identify a substantial contribution of rare, recessive variation in these noncoding elements to ASD risk but one that varies sharply with family structure: In a consanguineous cohort, where pathogenic *de novo* mutations are relatively less common compared to pathogenic inherited variants, recessive variation in HARs, VEs, and CNEs each contributes to ~10% of ASD cases. This is a substantial contribution when compared to the ~13% contribution of homozygous, noncoding copy number variants (Schmitz-Abe et al., 2020) and the ~30% contribution of homozygous, loss-of-function and missense variants (Yu et al., 2013; Doan et al., 2019) to ASD cases previously identified in this consanguineous cohort. In a non-consanguineous cohort of simplex and multiplex families, we observe a smaller (~3%) contribution of recessive variation in HARs and VEs to male ASD cases, comparable to the 3-5% contribution of recessive coding variants observed in a similar cohort (Doan et al., 2019). In contrast, the contribution of recessive variation in these noncoding elements to ASD risk is almost completely absent in an entirely simplex cohort, where *de novo*, rather than inherited, variants are more likely to contribute to ASD risk (Sebat et al., 2007; Yu et al., 2013; Cirmigliaro et al., 2023).

It is somewhat surprising that HARs consistently have the highest odds ratios for rare, recessive variants in ASD compared to controls across all three cohorts, followed by VEs and then by CNEs. Given that VEs are experimentally validated as neural enhancers and CNEs show strong predictions of neural enhancer activity, one might expect that VEs and CNEs would

have a stronger contribution to ASD risk than HARs, which are defined solely by brain-agnostic signatures of selection. The strong enrichment we observe in HARs may suggest that regions that are recently evolved in humans are more likely to contribute to disease risk than conserved regions. Intriguingly, two new sets of HARs were identified after the completion of this study (Keough et al., 2023; Bi et al., 2023). Although we are not powered to statistically assess these new, smaller HAR sets, both sets are also nominally enriched for rare, recessive variants in ASD cases compared to controls in the HMCA cohort (Fig. S18). Our results extend recent findings that common variation in noncoding, human-evolved regions may contribute to risk for neurological diseases (Xu et al., 2015; Srinivasan et al., 2016; Song et al., 2018), and suggests that rare variation in human-evolved regions may also preferentially contribute to human disease risk.

The proportion of HARs that showed enhancer activity in the caMPRA experiment was significantly lower than VEs and CNEs (Fig. 1D, Fig. 2B), which may reflect that HARs include regions with different regulatory functions. For instance, HARs include regions such as HAR3091 that were previously strong enhancers in other mammals, but where decreasing or silencing enhancer function appears to have been selected for in the human lineage. Variability in whether human-specific sequence variants in HARs increase or decrease enhancer activity has been observed in multiple MPRA comparing human and chimpanzee HAR sequences (Uebbing et al., 2021; Girskis et al., 2021; Whalen et al., 2023). We also clearly observe this heterogeneity in HAR function in the caMPRA mutagenesis experiment where a similar proportion of single nucleotide variants increase and decrease enhancer activity in HARs (Fig. 3C). In contrast, prior mutagenesis studies that examined known enhancers found that most functional variants decreased enhancer activity (Kircher et al., 2019; Snetkova et al., 2021). The strong enrichment for rare, recessive variants in HARs suggests the importance of examining noncoding regions that perform different regulatory functions for their contributions to ASD risk.

In contrast to HARs, CNEs had the lowest odds ratios for recessive ASD risk in all cohorts, despite being more highly conserved across species and more highly constrained within humans than HARs and VEs. In addition to a high level of conservation and constraint, CNEs are also predicted by the Roadmap Epigenomics Project to be active in a large number of body tissues outside of the brain (Fig. S1). This raises the possibility that variants in CNEs may be more likely to have large effect sizes or lead to pleiotropic effects that act in a dominant, *de novo* fashion rather than in a recessive, inherited fashion. Although we do not observe a case-specific enrichment of *de novo* variants in CNEs in the SSC cohort (Fig. S19), it is possible that an increased sample size may reveal a contribution of *de novo* variants in CNEs to ASD risk.

Prior work in non-consanguineous multiplex cohorts did not detect a significant contribution of noncoding, inherited variation when examining regions predicted to be functional (using similar heuristics as we use here to define CNEs) and suggested that sample sizes of 8000-9000 probands would be required for sufficient statistical power (Cirnigliaro et al., 2023). In contrast,

we find a significant enrichment for rare, recessive variants for HARs, VEs, and CNEs in a consanguineous cohort with only 193 probands and confirm this enrichment for HARs and VEs in a multiplex cohort. This suggests (1) that HARs and VEs, which are less likely than CNEs to be predicted to be active by epigenomic data (Fig. 1) but which were defined either by evolutionary signatures of selection in humans or by experimentally validated enhancer activity, are particularly impactful sets of noncoding regions, and that current predictors of functional activity require improvement; and (2) that consanguineous families may be especially suitable for analyzing noncoding contributions to disease risk, given that both direct consanguinity and endogamy enhance potential recessive contributions (Bittles and Black, 2010).

Although our cohorts are too small to ask whether individual noncoding elements are statistically enriched for rare, recessive variants, we posit that noncoding elements that show a numerical excess of rare, recessive variants in ASD cases compared to controls are nevertheless promising candidates for further study. We find that genes near noncoding elements with an excess of rare, recessive variants in cases compared to controls include both known ASD-associated genes and previously unassociated genes. Strikingly, proteins encoded by both ASD-associated and previously unassociated genes are known to interact, mirroring recent studies that identify convergent effects on protein networks across multiple, distinct genetic models of ASD (Paulsen et al., 2022; Pintacuda et al., 2023; Li et al., 2023). Many of these previously unassociated genes are dosage-sensitive, known to play critical roles in neurodevelopment, or mutated in severe developmental disorders (Table 1, Fig. S13). This suggests a model whereby coding variants in these genes lead to embryonic lethality or to multi-system developmental disorders, but noncoding variants in nearby regulatory sequences dysregulate gene expression in specific cell types or at specific developmental timepoints to contribute to ASD risk.

We closely investigated two HARs, HAR3091 and HAR3094, located near *ILIRAPLI* that contain rare, recessive variants in ASD cases but not in controls. We find that these HARs modulate *ILIRAPLI* expression in distinct brain regions; HAR3091 is a weak forebrain enhancer whereas HAR3094 is a strong midbrain enhancer in humans. We further find that ASD patient variants increase enhancer activity in HAR3091, but decrease enhancer activity in HAR3094. For both HAR3091 and HAR3094, patient variants reverted enhancer activity to more closely match the strength of the chimpanzee (ancestral) sequence. Together with studies showing that disease-associated variants in other human-evolved regions move functional activity in the derived direction (Huang et al., 2022), this suggests that disease variants in human-evolved regions may act either by reverting functional activity toward an ancestral state or by pushing functional activity toward an even more derived state. The opposing direction-of-effects of HAR3091 and HAR3094 patient variants further highlight that patient variants in noncoding regions may be more likely to act on dosage-sensitive genes like *ILIRAPLI*, where moderate increases or decreases in gene expression may result in phenotypic consequences.

ILIRAPLI encodes a post-synaptic organizer of excitatory synapses that has been repeatedly implicated in ASD and intellectual disability both by loss-of-function as well as by duplication (Carrié et al., 1999; Tabolacci et al., 2006; Froyen et al., 2007; Piton et al., 2008; Firth et al., 2009; Behnecke et al., 2011; Franek et al., 2011; Du et al., 2018; Montani et al., 2019). *ILIRAPLI* knock-out mice have deficits in learning and memory, reduced dendritic spine density, and a reduced number of excitatory synapses (Pavlovsky et al., 2010; Houbaert et al., 2013; Yasumura et al., 2014; Montani et al., 2017), whereas over-expression of *ILIRAPLI* in cultured neurons leads to an increased number of excitatory synapses (Pavlovsky et al., 2010; Valnegri et al., 2011). Changing excitatory synapses without affecting inhibitory synapses may result in an excitatory-inhibitory synaptic imbalance, which is one of the hallmark cellular phenotypes observed in ASD models (Nelson and Valakh, 2015). Additional studies will be needed to fully understand the role of *ILIRAPLI* and how modulation of its expression by HAR3091 and HAR3094 in specific brain regions may impact ASD risk. Nevertheless, our findings linking patient variants in HAR3091 and HAR3094 to *ILIRAPLI* expression suggest that noncoding regions that contain a numerical excess of rare, recessive variants are promising candidates for functional study.

We also investigated patient variants in hs576 (VE854), an enhancer of *SIMI*, which has been associated with both obesity and neurobehavioral deficits (Ramachandrapa et al., 2013), and found that patient variants reduced enhancer activity in multiple cranial nerves. Recent research has increasingly linked peripheral nervous system deficits, such as autonomic system or sensory neuron dysregulation, to common ASD symptoms, including decreased social engagement, flat facial expressions and intonation, touch and taste sensitivity, and gastrointestinal issues (Orefice et al., 2016; Jin and Kong, 2017; Huzard et al., 2022). These symptoms are likely driven, at least in part, by cranial nerves, including the trigeminal, facial, and vagus nerves affected by the patient variants in hs576. The vagus nerve, in particular, is also important in appetite regulation (de Lartigue, 2016), and its dysregulation in a subset of individuals with ASD may underlie the comorbidity of obesity and ASD (Phillips et al., 2014). Although both the previously identified obesity-associated variants and our newly identified ASD patient variants are located in the core enhancer region of hs576, detailed weight information for the ASD patients containing variants in hs576 is not available, and future research will be needed to determine whether these ASD patient variants affect *SIMI* expression in ways that solely contribute to neurobehavioral deficits or that may also contribute to obesity.

Collectively, these findings identify classes of noncoding regions that contribute to ASD disease risk and nominate specific noncoding elements and ASD patient variants for future study. We find that both regions that are highly conserved across species and likely act as neural enhancers (VEs and CNEs) and regions that are under positive selection in the human lineage and perform heterogeneous regulatory functions (HARs) contribute to ASD disease risk. This highlights the importance of examining a diverse set of noncoding regions for their contri-

bution to disease risk, including human-evolved elements and noncoding regions with diverse regulatory functions. Further, our data also demonstrate the importance of expanding cohort enrollment to diverse populations, and potentially focusing on populations with high rates of consanguinity and endogamy, since such families may be very powerful for elucidating the contribution of noncoding regions to ASD and other diseases.

Acknowledgements

We thank members of the Walsh and Doan labs for useful discussions. C.A.W. and R.N.D. are supported by the Allen Discovery Center program, a Paul G. Allen Frontiers Group advised program of the Paul G. Allen Family Foundation, and the Tan Yang Autism Center at Harvard University. T.S. is supported by the Stuart H.Q. & Victoria Quan Fellowship. J.H.T.S. is supported by the Y. Eva Tan Fellowship from the Tan Yang Autism Center at Harvard University and the Howard Hughes Medical Institute Fellowship from the Helen Hay Whitney Foundation. X.Q. is supported by the Howard Hughes Medical Institute Fellowship from the Helen Hay Whitney Foundation. I.A. is supported by grant no. T32GM144273 from NIGMS/NIH. R.E.A. is supported by Autism Speaks postdoctoral fellowship 13008. L.A.P. is supported by the grant no. R01HG003988 from NHGRI/NIH. C.A.W. is an Investigator of the Howard Hughes Medical Institute. The content is solely the responsibility of the authors and does not necessarily represent the official views of the National Institute of Health. This study makes use of data generated by the DECIPHER community. A full list of centers who contributed to the generation of the data is available from <https://deciphergenomics.org/about/stats> and via email from contact@deciphergenomics.org. Funding for the DECIPHER project was provided by Wellcome [grant number WT223718/Z/21/Z]. The clinicians who carried out the original analysis and collection of the data from the DECIPHER community bear no responsibility for the further analysis or interpretation of the data.

Author Contributions

Taehwan Shin: Conceptualization, Methodology, Formal analysis, Investigation, Writing - Original Draft, Writing - Review & Editing, Visualization. **Janet H.T. Song:** Conceptualization, Methodology, Formal Analysis, Investigation, Writing - Original Draft, Writing - Review & Editing, Visualization. **Michael Kosicki:** Investigation, Methodology, Writing - Original Draft, Writing - Review & Editing. **Connor Kenny:** Investigation. **Samantha G. Beck:** Investigation. **Lily Kelley:** Investigation. **Xuyu Qian:** Methodology, Investigation. **Julieta Bonacina:** Investigation. **Frances Papandile:** Investigation. **Irene Antony:** Investigation. **Dilenny Gonzalez:** Investigation. **Julia Scotellaro:** Investigation. **Evan Bushinsky:** Investigation. **Rebecca E. Andersen:** Methodology. **Eduardo Maury:** Formal Analysis. **Len A. Pennacchio:** Methodology, Resources, Writing - Review & Editing. **Ryan N. Doan:** Conceptualization, Methodology, Formal analysis, Investigation, Resources, Writing - Original Draft, Writing - Review &

Editing. **Christopher A. Walsh**: Conceptualization, Methodology, Resources, Writing - Original Draft, Writing - Review & Editing.

Declaration of Interests

The authors declare no competing interests.

Materials and Methods

Cell lines

Neuro2A (N2A) cells (ATCC, cat #CCL-131) were grown in 10% fetal bovine serum and 1% Penicillin-Streptomycin in Dulbecco's Modified Eagle Medium with L-Glutamine, 4.5g/L Glucose and Sodium Pyruvate (Fisher, cat #MT10013CV). We used a modified version of the male iPSC line WTC11 that contained stably integrated cassettes of a dox-inducible NGN2 and a degron-based inducible dCas9-KRAB (Tian et al., 2019). Both cell lines were maintained in a 5% CO₂ incubator at 37°C.

Mice

Transient transgenic mice were generated using plasmids containing either the human or chimpanzee versions of HAR3091 or HAR3094 located upstream of a minimal promoter driving a *lacZ* reporter gene. These constructs were generated by Vectorbuilder (VB210119-1206gb, VB210119-1208wrc, VB201008-1098whx, VB201020-1677ynn). Pronuclear injections of these constructs were performed in mice by the Mouse Engineering Core at Dana Farber / Harvard Cancer Center or by Cyagen (Santa Clara, CA). Mouse embryos were harvested at E14.5, bisected, and stained for *lacZ* expression. Embryos were cleared in 30% sucrose-PBS for imaging. Embryos with successful transgene insertions were determined by PCR for the *lacZ* gene. All animal experiments conformed to the guidelines approved by the Children's Hospital Animal Care and Use Committee.

Human subjects

From ASD families available through the NIMH repository, we processed 5551 samples (1911 probands) from likely simplex families, and 2277 samples (660 probands) from multiplex families in the Autism Genetic Resource Exchange (AGRE). Variant call format (VCF) files from whole-genome sequencing were obtained for the Homozygosity Mapping Collaborative for Autism (HMCA) from dbGaP phs001894.v1.p1. VCF files for SSC were obtained from SFARI. The number of male and female human subjects in each cohort is indicated in Fig. 4A. Research on human samples was conducted with approval of the Committee on Clinical Investigation at Boston Children's Hospital.

Selection of HARs, CNEs, VEs

The set of 3171 HARs examined in this study were selected from a number of different studies that identified HARs separately (Pollard et al., 2006; Bird et al., 2007; Prabhakar et al.,

2008; Bush and Lahn, 2008; Lindblad-Toh et al., 2011; Gittelman et al., 2015). Identified HARs that overlap were merged.

VEs were selected from active enhancers from the VISTA Enhancer browser (Visel et al., 2007). The enhancers were filtered for activity in the brain at E11.5, the time point evaluated by the VISTA enhancer group, and because many VEs are very long (greater than 1kb), VEs were subdivided. Only regions that contained at least mildly species-conserved bases (phastCons > 0.57) based on the 100-way vertebrate alignment from the UCSC Genome Browser (Navarro Gonzalez et al., 2021) were selected, with regions 50bp or closer merged to form a single element.

CNEs were selected based on epigenetic datasets, species conservation, and population constraint metrics. CNEs were filtered for conserved genomic regions, defined as having a >400 log-odds of being conserved using phastCons with the Viterbi setting (Siepel et al., 2005). Additionally, CNEs were required to have an enhancer-associated chromatin state (EnhG, Enh, or EnhBiv) based on ChromHMM (Kundaje et al., 2015) in neurospheres, fetal brain, or adult brain. Any elements that were annotated as exonic or splicing were filtered out. Furthermore, no more than 2% of bases in CNEs could have variants in Gnomad (Karczewski et al., 2020).

Capture-based Massively Parallel Reporter Assay (caMPRA) design, capture, and construction

Molecular inversion probes were designed to capture ~500bp regions using the MIPgen program (Boyle et al., 2014) to design targeting arms for the probes. Repetitive regions were masked prior to targeting. Flanking sequence was used to introduce AsiSI, PspXI, and SfiI restriction enzyme sites (NEB, cat #R0630L, R0656L, R0123L), along with a 10bp barcode for each probe. As targeted regions vary in length and many elements are longer than 500bp, probes were designed to double tile the bases of each element. MIPs were synthesized by Customarray, Inc (Redmond, WA).

The synthesized MIP oligos were amplified, amplification arms cleaved using MlyI (NEB, cat #R0610L), and purified using Ampure XP beads (Beckman Coulter) at 1.8x volume of the reaction mix. 15ng of the amplified MIP probes were then hybridized to 500ng of DNA from sample NA12878 (Coriell Institute) for 24 hours in 10x Ampligase buffer. The sequences in between the MIP targeting arms were captured by synthesis using Phusion DNA Polymerase (Thermo Fisher, cat #F530L) and circularized using Ampligase (VWR, cat #A3210K) at 60°C for 1 hour. Template DNA and uncaptured DNA were degraded using Exonuclease I (Thermo Fisher, cat #EN0582) and Exonuclease III (Thermo Fisher, cat #EN0191) at 37°C for 40 minutes and inactivated at 95°C for 5 minutes. The captured circles were then amplified using Phusion DNA Polymerase High Fidelity Master Mix (Thermo Fisher, cat #F531L), using primers that added SfiI restriction sites. Amplified, captured sequences were purified using Ampure XP beads at 0.65X to size select and remove unwanted shorter fragments.

The captured sequences and the pMPRA1 construct (Melnikov et al., 2014) (Addgene, cat #49349) were digested with SfiI and ligated using T4 DNA Ligase (Thermo Fisher, cat

#EL0014) at 16°C overnight. The ligated construct was purified and concentrated using the QIAquick Nucleotide Removal Kit (Qiagen, cat #28306). The ligated constructs were either transformed into 60 vials of Top10 chemically competent cells (Thermo Fisher, cat #C404006) and cultured overnight at 37°C in 200ml of LB/Ampicillin, or transformed into 1 vial of MegaX DH10B T1R Electrocompetent Cells (Thermo Fisher, cat #C640003) and plated on LB/Ampicillin agar plates (Molecular Devices X6023 BIOASSAY TRAYS; Fisher Scientific, cat #NC9372402) overnight at 37°C. Plasmid DNA was extracted the following day using the Qiagen Plasmid Maxi Kit (Qiagen, cat #12162). This plasmid library containing the captured sequences and a modified pMPRA donor2 (Melnikov et al., 2014) (Addgene, cat #49353) containing an AsiSI site were then digested using AsiSI and PspXI, and the fragment containing the minimal promoter and luciferase gene from pMPRA donor2 was cloned into the plasmid library containing the captured sequences between the captured element and the barcode. This final construct was then transformed, cultured, and harvested as above.

For the random mutagenesis of HARs, a 25bp barcode was used. We performed error-prone PCR using the GeneMorph II Random Mutagenesis Kit (Agilent, cat #200550) on the amplified, captured sequences. Based on the error rate of the Mutazyme II polymerase and PCR yield, we performed 7 cycles of error-prone PCR with 20ng of input captured sequence, with 25bp random barcode reverse primer. These mutagenized sequences were then cloned into the modified pMPRA1 construct as described above. In order to associate the mutagenized sequence with the random barcode, the cloned plasmid library containing the captured sequences was PCR amplified with primers containing sequencing adapters and sent out for 2x250bp sequencing on HiSeq Instruments at Psomagen.

Cell culture and transfection for caMPRA

N2A cells (ATCC, cat #CCL-131) were cultured in DMEM with L-Glutamine, 4.5g/L Glucose and Sodium Pyruvate (Fisher, cat #MT10013CV) with 10% fetal bovine serum and 1% penicillin/streptavidin at 37°C. Cells were maintained in 10cm TC-treated plates and split 1:5 every 4 days or when confluent with 0.25% (w/v) Trypsin – 0.53mM EDTA (Fisher, cat #25200-114). To minimize confounding due to passage number, we limited passage numbers to P3-6. For transfections, N2a cells were transfected at 70% confluency using Lipofectamine LTX with PLUS reagent (Thermo Fisher, cat #15338100) with 15ug of caMPRA plasmid, and cells were incubated with the transfection mix for 24 hours. After 1 day, media was changed. Cells were harvested either 1 day or 3 days after transfection. Cell pellets were washed with 1x PBS, and mRNA was extracted using the Dynabead mRNA Direct kit (Thermo Fisher, cat #61012), according to the manufacturer's instructions. mRNA was reverse transcribed using Superscript VILO Master Mix with EZ DNase (Thermo Fisher, cat #11766050). caMPRA barcodes were extracted using PCR amplification with primers containing illumina adapters for both the cDNA and plasmid pools and sent out for 150bp sequencing using HiSeq instruments at Psomagen (Rockville, MD).

Targeted sequencing of NIMH cohort

From the ASD families available through the NIMH repository, we processed 5551 samples (1911 probands) from likely simplex families, and 2277 samples (660 probands) from multiplex families in the Autism Genetic Resource Exchange (AGRE). The likely simplex families include families with one affected proband and no siblings and families with one affected proband and one or more unaffected siblings. Molecular inversion probes (MIPs) were designed, synthesized, and amplified as described above with the following changes: MIPs were designed to capture ~240 bp regions and were hybridized to a pool of DNA from NIMH samples. The purified library was quantified using a tapestation and sequenced at 2x150bp by Psomagen (Rockville, MD).

LacZ enhancer reporter assay in transgenic mice with random integration sites

We cloned either the human or chimpanzee versions of HAR3091 or HAR3094 located upstream of a minimal promoter driving a *lacZ* reporter gene with Vectorbuilder (VB210119-1206gb, VB210119-1208wrc, VB201008-1098whx, VB201020-1677ynn). Pronuclear injections of these constructs were performed in mice by the Mouse Engineering Core at Dana Farber / Harvard Cancer Center or by Cyagen (Santa Clara, CA). Mouse embryos were harvested at E14.5, bisected, and stained for *lacZ* expression. Embryos were cleared in 30% sucrose-PBS for imaging. Embryos with successful transgene insertions were determined by PCR for the *lacZ* gene. Because these mice are analyzed at F0, *lacZ* expression is dependent on the distribution and number of cells that integrate the reporter construct and the genomic location of the integration. Consequently, we expect that expression patterns driven by the sequence of interest rather than by the integration site will be consistently observed in multiple embryos and examined at least 10 PCR-positive embryos per construct to account for this variability.

LacZ enhancer reporter assay in transgenic mice with site-specific integration

Transgenic E11.5 mouse embryos were generated as described previously (Osterwalder et al., 2022). Briefly, superovulating female FVB mice were mated with FVB males and fertilized embryos were collected from the oviducts. Regulatory elements sequences were synthesized by Twist Biosciences. Inserts generated in this way were cloned into the donor plasmid containing minimal *Shh* promoter, *lacZ* reporter gene and H11 locus homology arms (Addgene, cat #139098) using NEBuilder HiFi DNA Assembly Mix (NEB, cat #E2621). The sequence identity of donor plasmids was verified using long-read sequencing (Primordium). Plasmids are available upon request. A mixture of Cas9 protein (Alt-R SpCas9 Nuclease V3, IDT, cat #1081058, final concentration 20 ng/ μ L), hybridized sgRNA against H11 locus (Alt-R CRISPR-Cas9 tracrRNA, IDT, cat #1072532 and Alt-R CRISPR-Cas9 locus targeting crRNA, gctgatggaacagtaacaa, total final concentration 50 ng/ μ L) and donor plasmid (12.5 ng/ μ L) was injected into

the pronucleus of donor FVB embryos. The efficiency of targeting and the gRNA selection process is described in detail in (Osterwalder et al., 2022). Embryos were cultured in M16 with amino acids at 37°C, 5% CO₂ for 2 hours and implanted into pseudopregnant CD-1 mice. Embryos were collected at E11.5 for lacZ staining as described previously (Osterwalder et al., 2022). Briefly, embryos were dissected from the uterine horns, washed in cold PBS, fixed in 4% PFA for 30 min and washed three times in embryo wash buffer (2mM MgCl₂, 0.02% NP-40, and 0.01% deoxycholate in PBS at pH 7.3). They were subsequently stained overnight at room temperature in X-gal stain (4mM potassium ferricyanide, 4mM potassium ferrocyanide, 1mg/mL X-gal and 20mM Tris pH 7.5 in embryo wash buffer). PCR using genomic DNA extracted from embryonic sacs digested with DirectPCR Lysis Reagent (Viagen, cat #301-C) containing Proteinase K (final concentration 6 U/mL) was used to confirm integration at the H11 locus and test for presence of tandem insertions (see (Osterwalder et al., 2022) for details). Only embryos with donor plasmid insertion at H11 were used. The stained transgenic embryos were washed three times in PBS and imaged from both sides using a Leica MZ16 microscope and Leica DFC420 digital camera.

CRISPR inhibition in iPSC-derived neurons

Guide RNAs (gRNAs) were designed to target the middle 200bp interval of HAR3091 and HAR3094 using GuideScan (Perez et al., 2017) with a specificity score cut-off > 0.2. Non-targeting control (NTC) gRNAs and gRNAs targeting the *IL1RAPL1* transcription start site (TSS) are from (Horlbeck et al., 2016). gRNA sequences can be found in Table S5. gRNAs were cloned into pBA904 (RRID: Addgene.122238), as previously described (Replogle et al., 2020). Lentivirus was made for each gRNA by ultracentrifugation.

We used a modified version of the iPSC line WTC11 that contained stably integrated cassettes of a dox-inducible NGN2 and a degenon-based inducible dCas9-KRAB (Tian et al., 2019). This modified WTC11 line was differentiated into neurons by inducing NGN2 expression as previously described (Chen et al., 2020). Twice the suggested number of cells (e.g. 2×10^5 cells per well in a 24-well plate) were plated at D0 to account for incomplete lentiviral infection, and immediately after plating, lentivirus was added at MOI 0.7 (so that ~50% of cells would be infected with lentivirus). dCas9-KRAB expression was induced by the addition of 20 μ M trimethoprim (Sigma Aldrich, cat #92131) from D0 until the neurons were collected at D18, and 1 μ g/ml puromycin was added from D2-D7 to select for infected neurons. RNA was extracted at D18 using the Direct-zol RNA Microprep Kit (Zymo, cat #R2062) per the manufacturer's instructions, and cDNA was synthesized from RNA with the SuperScript VILO cDNA Synthesis Kit (Thermo Fisher, cat #11756050). RT-qPCR was performed using Brilliant II SYBR Green Low ROX qPCR Master Mix (Agilent, cat #600830) on a CFX384 Touch Real-Time PCR Detection System (BioRad) for *IL1RAPL1* and *GAPDH* (primer sequences in Table S5) in triplicate.

Luciferase assays

The wild-type (WT) and mutant sequences of for HAR 3091 and 3094 were generated through PCR amplification from Promega Control Male human DNA (Promega, cat #G1471) and proband genomic DNA using primers containing unique restriction sites for directional cloning into a minimal promoter pGL4.25 luciferase. Two families harboring the same variant in HAR3091 (chrX:29275879, G>T) were amplified independently and cloned into separate plasmids containing the same variant. Plasmids were transformed into Top10 chemically competent cells (Thermo Fisher, cat #C404006). Genotypes and structures of the final plasmids were confirmed using Sanger sequencing. Plasmids (75ng) were co-transfected along with control Renilla (25ng) into mouse neuroblastoma Neuro-2a cells (N2a) (ATCC, cat #CCL-131) in 96-well plates using Lipofectamine 3000 (Thermo Fisher, cat #L3000015). Luciferase activities were assessed 48 hours post-transfection using standard procedures for the dual-luciferase reporter assay system (Promega, cat #E1960). The firefly luciferase activity was normalized using Renilla activity for each well of the 96-well plates. The luciferase activity measurements were performed with 8 replicates. Data from the two plasmids generated from different families with the same variant in HAR3091 yielded similar results, and are shown as one variant (Fig. 5D, Fig. S15).

Assessing accessibility and epigenetic marks in human tissue

ChromHMM annotations from the Epigenomics Roadmap Project (Kundaje et al., 2015) were overlapped with HARs, CNEs, and VEs using bedtools intersect (Quinlan and Hall, 2010) to identify the number of elements in each class that were annotated as active (TssA, TssAFlnk, TxFlnk, Tx, TxWk, EnhG, Enh, TssBiv, BivFlnk, or EnhBiv in the 15-state model) in each assessed tissue. For cell type-specific annotations in the adult brain, scTHS-seq data was used (Lake et al., 2018) to determine accessibility in different brain cell types for HARs, VEs, and CNEs.

TF binding analysis

We examined HARs, VEs, and CNEs, as well as matched background sequences generated using BiasAway (Khan et al., 2021), for transcription factor binding sites. PWMSCAN (Ambrosini et al., 2018) was used to scan sequences for motifs from the JASPAR motif database (Castro-Mondragon et al., 2022) to identify potential transcription factor binding sites. To control for false positives, a p-value cutoff of 10^{-4} was used. The presence of motifs was then aggregated and the enrichment of specific motifs in HARs, VEs, or CNEs compared to matched background sequences was determined. P-values for enrichment were generated using the hypergeometric test and were adjusted for multiple hypothesis testing with the Benjamini-Hochberg correction. Gene set enrichment analysis for TFs with motifs in HARs, VEs, and CNEs was performed using clusterProfiler (Yu et al., 2012) and adjusted for multiple hypothesis testing with the Benjamini-Hochberg correction.

In addition to motif-based matching, HARs, VEs, and CNEs were also annotated using DeepSEA (Zhou et al., 2018) using

the online DeepSEA server. We used the Beluga model that was trained on 2,002 chromatin features. TF ChIP-seq features were selected for, and only those with an e-value (defined as the expected proportion of common SNPs with a larger predicted effect) of less than 0.05 were interpreted as associated with the element.

Analysis of genes near HARs, VEs, and CNEs

Gene ontology analysis was performed with GREAT (McLean et al., 2010) with the binomial test at 5% FDR. The binomial test at 5% FDR were also used to assess whether HARs, VEs, and CNEs are enriched near disease-associated genes. HARs, VEs, and CNEs were assigned to nearby genes as previously described (McLean et al., 2010). We separated genes implicated in severe, developmental disorders from the DECIPHER consortium (v. 13_7_2022) (Firth et al., 2009) based on the phenotypes of the affected individuals. If affected individuals had phenotypes in multiple body systems, affected genes were assigned to all affected body systems. We also examined autism-associated genes from the SFARI database (Abrahams et al., 2013).

To examine whether autism-associated genes and genes near HARs, VEs, or CNEs were enriched for dosage-sensitive genes, we examined pLI and LOEUF scores (Lek et al., 2016; Karczewski et al., 2020). pLI > 0.9 and low LOEUF scores indicate loss-of-function intolerance. The hypergeometric test was used to test whether ASD-associated genes from the SFARI database and genes near HARs, VEs, or CNEs were enriched for genes with pLI > 0.9 at 5% FDR. The Wilcoxon rank-sum test was used to test whether the LOEUF scores of ASD-associated genes or genes near HARs, VEs, or CNEs differed from LOEUF scores for all genes at 5% FDR.

caMPRA analysis

To count barcodes to assess regulatory activity, sequencing data from the plasmid DNA and cDNA libraries described above were processed with cutadapt to remove adapters (Martin, 2011). Barcodes were extracted using UMI-tools (Smith et al., 2017) and reads were mapped using bwa mem (Li and Durbin, 2009). Reads were assigned to caMPRA probes using featureCounts (Liao et al., 2014). The 10bp barcode were clustered to recover barcodes with small sequencing errors using the multiplexed version of the UMI-tools directional method.

Barcodes for plasmid and cDNA samples were normalized to a barcode per million format to remove bias due to sequencing coverage. Each cDNA barcode was normalized to the barcode count in the plasmid pool, and \log_2 transformed. Only barcodes that were found in the plasmid pool and all 5 cDNA replicate pools were used in the analysis. After filtering, we were able to examine 2932 HARs, 1702 VEs, and 5155 CNEs. Elements were considered active if at least one probe overlapping that element was active in the assay (Fig. 2B, Fig. S6B). The proportion of active probes (sequences) is shown in Fig. S6A, and full results are detailed in Table S2.

Analysis of caMPRA data from random mutagenesis

Analysis of the plasmid and cDNA barcode pools was performed as described above. Variants from each caMPRA probe were called using bcftools mpileup (Li, 2011), and associated with the appropriate barcode. Sequences captured using MIPs may include regions that flank the sequence of interest. Mutagenized sequences that only included variants in the flanking sequence were excluded. Variants found in NA12878 compared to the reference genome were filtered out. The correlation between replicate experiments (Fig. S8) was assessed prior to removing mutagenized sequences that only included variants in the flanking sequence.

MIP sequencing pre-processing and variant calling

Analysis of targeted sequencing was performed using a custom pre-processing pipeline combined with GATK-based variant calling. Briefly, sequenced reads were trimmed for adapters using cutadapt (Martin, 2011). UMI-tools (Smith et al., 2017) was used to extract the 5bp unique molecular index (UMI), and reads were mapped to the human genome (hg19) using bwa mem (Li and Durbin, 2009). Reads that mapped off-target compared to the intended target were filtered out. The extension and ligation arms (the targeting arms) were clipped off the mapped reads using bamclipper (Au et al., 2017). Samtools was used to remove multimapping reads, unpaired/broken read pairs, and unmapped reads (Li et al., 2009). UMI-tools was used to collapse sequences based on UMI sequence. Finally, sequences were base recalibrated using GATK base recalibrator, and variants were joint-called using the GATK Haplotype caller and suggested GATK best practices (Van der Auwera et al., 2013). After targeted sequencing and processing, we were able to resolve HARs in 6464 individuals, VEs in 5273 individuals, and CNEs in 5983 individuals.

Variant filtering, classification, and analysis

The HMCA consanguineous cohort was filtered for variants of AD > 2, DP > 10, GQ > 20. For the targeted sequencing of the NIMH cohort, variants were required to have a minimum of 10x coverage and GQ > 20 and AD > 4. Only variants produced from collapsed reads were used for accuracy. The Simon Simplex Collection (SSC) cohort of 8,186 individuals was filtered to remove alleles with AD < 3, DP < 5, and GQ < 20.

For recessive variant analysis, our definition included homozygous, compound heterozygous, and hemizygous variants (specifically in male individuals for the X chromosome). Because hemizygous variants are much more likely to appear, we performed our analysis with each sex separately when examining genome-wide rates.

In order to enrich for functionality, we created a classification that uses an ensemble of different conservation-based variant effect predictors – GERP++ (Davydov et al., 2010), CADD (Rentzsch et al., 2019), DANN (Quang et al., 2015), FATHMMnc (Shihab et al., 2015) - to annotate variants and base positions. Variants were filtered to exclude those within exonic regions of protein-coding genes (based on RefSeq and Gencode v28). For variants that fall within either the UTRs,

within 1 kb upstream of a transcriptional start site, or within a predicted promoter element from the Eukaryotic Promoter Database (Dreos et al., 2015), these variants must overlap a conserved element from the 100-way phastCons from the UCSC genome browser. All variants were filtered for GERP > 2 and (CADD > 15 or DANN > 0.85 or FATHMMnc > 0.85).

The variant contributions of rare germline events were assessed for rare, recessive and *de novo* predicted damaging variants identified in individuals with ASD and healthy familial controls. Statistical testing of variant contributions was performed as follows. First, the odds ratio (OR), standard error, and 95% confidence intervals were calculated using the approach described by (Altman, 1991): $OR = \frac{a/b}{c/d}$, where a = number of cases with variants, b = number of cases without variants, c = number of controls with variants, and d = number of controls without variants. The standard errors of the log odds ratio were calculated using the following formula: $SE\{\ln(OR)\} = \sqrt{\frac{1}{a} + \frac{1}{b} + \frac{1}{c} + \frac{1}{d}}$. The 95% confidence intervals were determined using $95\% CI = \exp(\ln(OR) \pm 1.96 \times SE\{\ln(OR)\})$. P-values of the ORs were calculated under the assumption of the deviation from a normal distribution, using: $z\text{-value} = \frac{\ln(OR)}{SE\{\ln(OR)\}}$. The allele frequency (AF) cut-off for statistical analysis was set at either $AF < 0.005$ (HMCA and SSC) or $AF < 0.001$ (NIMH), using the lowest AF where there were >5 predicted damaging variants for cases and controls for each sex. The contribution of variants to ASD risk was estimated as the difference between the rate of rare, recessive variants in cases compared to controls as previously described (Doan et al., 2019).

In cohorts with known elevations of homozygosity that could impact the recessive contribution (e.g., HMCA), we determined the rates of likely benign events at non-conserved sites within gene promoters that have no predicted functional impact under the assumption that these rates should be equivalent in cases and controls due to the lack of selection bias on the sites. Next, the rates of predicted damaging events in cases were reduced proportionally to the excess detected in the non-conserved sites, as was done previously using synonymous rates for recessive protein coding variation (Doan et al., 2019). Following correction for elevated consanguinity, variation contributions and significance were determined, using the above described approach.

Protein-protein interaction networks

Variants found in HARs, VEs, or CNEs in HMCA and variants found in HARs or VEs in NIMH were aggregated to the level of individual HARs, VEs, and CNEs. HARs, VEs, and CNEs with a numerical excess of variants found in patients compared to controls were associated with nearby genes using GREAT (McLean et al., 2010). A protein-protein interaction network for these nearby genes was constructed using STRING version 11.5 using the online interface with default parameters (Szklarczyk et al., 2019).

Analysis of CRISPR inhibition in iPSC-derived neurons and luciferase assays

To analyze the CRISPR inhibition data, the quantity of *ILIRAPL1* for each sample was calculated by comparing the Ct value for *ILIRAPL1* to a standard curve of pooled samples and then normalized to the expression of the housekeeping gene *GAPDH* in the same sample. This normalized value was then divided by the normalized quantity of *ILIRAPL1* in samples infected with NTC gRNAs. Each point represented in Fig. 5 and Fig. S14 is from a separate well; each condition was tested in at least 3 different differentiations. The Wilcoxon rank-sum test was used to compare each individual gRNA to the NTC gRNAs, and p-values for gRNAs targeting the same region were combined with Fisher's method. P-values were adjusted with the Benjamini-Hochberg correction and considered significant at 5% FDR.

For the luciferase assays, the Wilcoxon rank-sum test was used to compare each test sequence to the control sequence, and p-values for each replicate were combined with Fisher's method. P-values were adjusted with the Benjamini-Hochberg correction and considered significant at 5% FDR.

References

- Abrahams, B. S., Arking, D. E., Campbell, D. B., Mefford, H. C., Morrow, E. M., Weiss, L. A., Menashe, I., Wadkins, T., Banerjee-Basu, S., and Packer, A., et al., 2013. SFARI Gene 2.0: A community-driven knowledgebase for the autism spectrum disorders (ASDs). *Molecular Autism*, **4**:36.
- Ahituv, N., Kavaslar, N., Schackwitz, W., Ustaszewska, A., Martin, J., Hébert, S., Doelle, H., Ersoy, B., Kryukov, G., Schmidt, S., et al., 2007. Medical sequencing at the extremes of human body mass. *American Journal of Human Genetics*, **80**(4):779–791.
- Akula, S. K., Marciano, J. H., Lim, Y., Exposito-Alonso, D., Hylton, N. K., Hwang, G. H., Neil, J. E., Dominado, N., Bunton-Stasyshyn, R. K., Song, J. H. T., et al., 2023. TMEM161B regulates cerebral cortical gyration, Sonic Hedgehog signaling, and ciliary structure in the developing central nervous system. *Proceedings of the National Academy of Sciences*, **120**(4):e2209964120.
- Albert-Gascó, H., Ros-Bernal, F., Castillo-Gómez, E., and Olucha-Bordonau, F. E., 2020. MAP/ERK signaling in developing cognitive and emotional function and its effect on pathological and neurodegenerative processes. *International Journal of Molecular Sciences*, **21**(12):4471.
- Altman, D. G., 1991. *Practical Statistics for Medical Research*. Chapman and Hall, London.
- Ambrosini, G., Groux, R., and Bucher, P., 2018. PWMScan: a fast tool for scanning entire genomes with a position-specific weight matrix. *Bioinformatics*, **34**(14):2483–2484.
- An, J.-Y., Lin, K., Zhu, L., Werling, D. M., Dong, S., Brand, H., Wang, H. Z., Zhao, X., Schwartz, G. B., Collins, R. L., et al., 2018. Genome-wide *de novo* risk score implicates promoter variation in autism spectrum disorder. *Science*, **362**(6420):eaat6576.
- Arora, A., Castro-Gutierrez, R., Moffatt, C., Eletto, D., Becker, R., Brown, M., Moor, A. E., Russ, H. A., and Taliaferro, J. M., 2022. High-throughput identification of RNA localization elements in neuronal cells. *Nucleic Acids Research*, **50**(18):10626–10642.
- Au, C. H., Ho, D. N., Kwong, A., Chan, T. L., and Ma, E. S. K., 2017. BAM-Clipper: Removing primers from alignments to minimize false-negative mutations in amplicon next-generation sequencing. *Scientific Reports*, **7**(1):1567.
- Bae, B. I., Tietjen, I., Atabay, K. D., Evrony, G. D., Johnson, M. B., Asare, E., Wang, P. P., Murayama, A. Y., Im, K., Lisgo, S. N., et al., 2014. Evolutionarily dynamic alternative splicing of GPR56 regulates regional cerebral cortical patterning. *Science*, **343**(6172):764–768.
- Behnecke, A., Hinderhofer, K., Bartsch, O., Nümann, A., Ipach, M.-L., Damatova, N., Haaf, T., Dufke, A., Riess, O., and Moog, U., et al., 2011. In-

- tragic deletions of IL1RAPL1: Report of two cases and review of the literature. *American Journal of Medical Genetics Part A*, **155**(2):372–379.
- Bhat, S. S., Ladd, S., Grass, F., Spence, J. E., Brasington, C. K., Simensen, R. J., Schwartz, C. E., Dupont, B. R., Stevenson, R. E., and Srivastava, A. K., et al., 2008. Disruption of the IL1RAPL1 gene associated with a pericentromeric inversion of the X chromosome in a patient with mental retardation and autism. *Clinical Genetics*, **73**(1):94–96.
- Bi, X., Zhou, L., Zhang, J.-J., Feng, S., Hu, M., Cooper, D. N., Lin, J., Li, J., Wu, D.-D., and Zhang, G., et al., 2023. Lineage-specific accelerated sequences underlying primate evolution. *Science Advances*, **9**(22):eac9507.
- Bird, C. P., Stranger, B. E., Liu, M., Thomas, D. J., Ingle, C. E., Beazley, C., Miller, W., Hurles, M. E., and Dermizakis, E. T., 2007. Fast-evolving noncoding sequences in the human genome. *Genome Biology*, **8**(66):1–12.
- Bittles, A. H. and Black, M. L., 2010. Consanguinity, human evolution, and complex diseases. *Proceedings of the National Academy of Sciences*, **107**:1779–1786.
- Bonnefond, A., Raimondo, A., Stutzmann, F., Ghossaini, M., Ramachandrapra, S., Bersten, D. C., Durand, E., Vatin, V., Balkau, B., Lantieri, O., et al., 2013. Loss-of-function mutations in SIM1 contribute to obesity and Prader-Willi-like features. *Journal of Clinical Investigation*, **123**(7):3037–3041.
- Boyd, J. L., Skove, S. L., Rouanet, J. P., Pilaz, L. J., Bepler, T., Gordan, R., Wray, G. A., and Silver, D. L., 2015. Human-chimpanzee differences in a FZD8 enhancer alter cell-cycle dynamics in the developing neocortex. *Current Biology*, **25**(6):772–779.
- Boyle, E. A., O’Roak, B. J., Martin, B. K., Kumar, A., and Shendure, J., 2014. MIPgen: Optimized modeling and design of molecular inversion probes for targeted resequencing. *Bioinformatics*, **30**(18):2670–2672.
- Bush, E. C. and Lahn, B. T., 2008. A genome-wide screen for noncoding elements important in primate evolution. *BMC Evolutionary Biology*, **8**:17.
- Cannavò, E., Khouchi, P., Garfield, D. A., Geeleher, P., Zichner, T., Gustafson, E. H., Ciglar, L., Korb, J. O., and Furlong, E. E. M., 2016. Shadow enhancers are pervasive features of developmental regulatory networks. *Current Biology*, **26**(1):38–51.
- Capra, J. A., Erwin, G. D., McKinsey, G., Rubenstein, J. L., and Pollard, K. S., 2013. Many human accelerated regions are developmental enhancers. *Philosophical Transactions of the Royal Society B*, **368**(1632):20130025.
- Carrié, A., Jun, L., Bienvenu, T., Vinet, M.-C., McDonnell, N., Couvert, P., Zenni, R., Cardona, A., Van Buggenhout, G., Frints, S., et al., 1999. A new member of the IL-1 receptor family highly expressed in hippocampus and involved in X-linked mental retardation. *Nature Genetics*, **23**(11):25–31.
- Castro-Mondragon, J. A., Riudavets-Puig, R., Rauluseviciute, I., Lemma, R. B., Turchi, L., Blanc-Mathieu, R., Lucas, J., Boddie, P., Khan, A., Manosalva Pérez, N., et al., 2022. JASPAR 2022: The 9th release of the open-access database of transcription factor binding profiles. *Nucleic Acids Research*, **50**(D1):D165–D173.
- Chen, M., Draeger, N., Kampmann, M., Leng, K., Li, E., Ludwig, C., Mohl, G., Samelson, A., Sattler, S., and Tian, R., et al., 2020. iNeuron pre-differentiation & differentiation protocol. <https://www.protocols.io/view/ineuron-pre-differentiation-and-differentiation-pr-bcrjv4n>.
- Cirmigliaro, M., Chang, T. S., Arteaga, S. A., Pérez-Cano, L., Ruzzo, E. K., Gordon, A., Bicks, L. K., Jung, J.-Y., Lowe, J. K., Wall, D. P., et al., 2023. The contributions of rare inherited and polygenic risk to ASD in multiplex families. *Proceedings of the National Academy of Sciences*, **120**(31):e2215632120.
- Coe, B. P., Stessman, H. A. F., Sulovari, A., Geisheker, M. R., Bakken, T. E., Lake, A. M., Dougherty, J. D., Lein, E. S., Hormozdiari, F., Bernier, R. A., et al., 2019. Neurodevelopmental disease genes implicated by de novo mutation and copy number variation morbidity. *Nature Genetics*, **51**(1):106–116.
- Davydov, E. V., Goode, D. L., Sirota, M., Cooper, G. M., Sidow, A., and Batzoglou, S., 2010. Identifying a high fraction of the human genome to be under selective constraint using GERP++. *PLoS Computational Biology*, **6**(12):e1001025.
- de Lartigue, G., 2016. Role of the vagus nerve in the development and treatment of diet-induced obesity. *The Journal of Physiology*, **594**(20):5791–5815.
- De Rubeis, S., He, X., Goldberg, A. P., Poultney, C. S., Samocha, K., Ercument Cicek, A., Kou, Y., Liu, L., Fromer, M., Walker, S., et al., 2014. Synaptic, transcriptional and chromatin genes disrupted in autism. *Nature*, **515**(75267526):209–215.
- den Hoed, J., Devaraju, K., and Fisher, S. E., 2021. Molecular networks of the FOXP2 transcription factor in the brain. *EMBO Reports*, **22**(8):e52803.
- di Iulio, J., Bartha, I., Wong, E. H. M., Yu, H.-C., Lavrenko, V., Yang, D., Jung, I., Hicks, M. A., Shah, N., Kirkness, E. F., et al., 2018. The human noncoding genome defined by genetic diversity. *Nature Genetics*, **50**(33):333–337.
- Diez-Roux, G., Banfi, S., Sultan, M., Geffers, L., Anand, S., Rozado, D., Magen, A., Canidio, E., Pagani, M., Peluso, I., et al., 2011. A high-resolution anatomical atlas of the transcriptome in the mouse embryo. *PLoS Biology*, **9**(1):e1000582.
- Dixon, J. R., Jung, I., Selvaraj, S., Shen, Y., Antosiewicz-Bourget, J. E., Lee, A. Y., Ye, Z., Kim, A., Rajagopal, N., Xie, W., et al., 2015. Chromatin architecture reorganization during stem cell differentiation. *Nature*, **518**(7539):331–336.
- Dixon, J. R., Selvaraj, S., Yue, F., Kim, A., Li, Y., Shen, Y., Hu, M., Liu, J. S., and Ren, B., 2012. Topological domains in mammalian genomes identified by analysis of chromatin interactions. *Nature*, **485**(7398):376–380.
- Doan, R. N., Bae, B. I., Cubelos, B., Chang, C., Hossain, A. A., Al-Saad, S., Mukaddes, N. M., Oner, O., Al-Saffar, M., Balkhy, S., et al., 2016. Mutations in human accelerated regions disrupt cognition and social behavior. *Cell*, **167**(2):341–354.
- Doan, R. N., Lim, E. T., De Rubeis, S., Betancur, C., Cutler, D. J., Chiochetti, A. G., Overman, L. M., Soucy, A., Goetze, S., Freitag, C. M., et al., 2019. Recessive gene disruptions in autism spectrum disorder. *Nature Genetics*, **51**(7):1092–1098.
- Dreos, R., Ambrosini, G., Périer, R. C., and Bucher, P., 2015. The Eukaryotic Promoter Database: Expansion of EPDnew and new promoter analysis tools. *Nucleic Acids Research*, **43**(D1):D92–D96.
- Du, X., Gao, X., Liu, X., Shen, L., Wang, K., Fan, Y., Sun, Y., Luo, X., Liu, H., Wang, L., et al., 2018. Genetic diagnostic evaluation of trio-based whole exome sequencing among children with diagnosed or suspected autism spectrum disorder. *Frontiers in Genetics*, **9**.
- El Chehadeh, S., Han, K. A., Kim, D., Jang, G., Bakhtiari, S., Lim, D., Kim, H. Y., Kim, J., Kim, H., Wynn, J., et al., 2022. SLITRK2 variants associated with neurodevelopmental disorders impair excitatory synaptic function and cognition in mice. *Nature Communications*, **13**(1):4112.
- Epifanova, E., Babaev, A., Newman, A. G., and Tarabykin, V., 2019. Role of Zeb2/Sip1 in neuronal development. *Brain Research*, **1705**:24–31.
- Favaro, F. P., Alvizi, L., Zechi-Ceide, R. M., Bertola, D., Felix, T. M., de Souza, J., Raskin, S., Twigg, S. R., Weiner, A. M., Armas, P., et al., 2014. A noncoding expansion in EIF4A3 causes Richieri-Costa-Pereira syndrome, a craniofacial disorder associated with limb defects. *American Journal of Human Genetics*, **94**(1):120–128.
- Firth, H. V., Richards, S. M., Bevan, A. P., Clayton, S., Corpas, M., Rajan, D., Vooren, S. V., Moreau, Y., Pettett, R. M., and Carter, N. P., et al., 2009. DECIPHER: Database of chromosomal imbalance and phenotype in humans using Ensembl resources. *American Journal of Human Genetics*, **84**(4):524–533.
- Fischbach, G. D. and Lord, C., 2010. The Simons Simplex Collection: A resource for identification of autism genetic risk factors. *Neuron*, **68**(2):192–195.
- Franeck, K. J., Butler, J., Johnson, J., Simensen, R., Friez, M. J., Bartel, F., Moss, T., DuPont, B., Berry, K., Bauman, M., et al., 2011. Deletion of the immunoglobulin domain of IL1RAPL1 results in nonsyndromic X-linked intellectual disability associated with behavioral problems and mild dysmorphism. *American Journal of Medical Genetics Part A*, **155**(5):1109–1114.
- Froyen, G., Van Esch, H., Bauters, M., Hollanders, K., Frints, S. G. M., Vermeesch, J. R., Devriendt, K., Frys, J.-P., and Marynen, P., 2007. Detection of genomic copy number changes in patients with idiopathic mental retardation by high-resolution X-array-CGH: Important role for increased gene dosage of XLMR genes. *Human Mutation*, **28**(10):1034–1042.
- Gilbert, L. A., Horlbeck, M. A., Adamson, B., Villalta, J. E., Chen, Y., Whitehead, E. H., Guimaraes, C., Panning, B., Ploegh, H. L., Bassik, M. C., et al., 2014. Genome-scale CRISPR-mediated control of gene repression and activation. *Cell*, **159**(3):647–661.
- Gilman, S. R., Iossifov, I., Levy, D., Ronemus, M., Wigler, M., and Vitkup, D., 2011. Rare de novo variants associated with autism implicate a large functional network of genes involved in formation and function of synapses. *Neuron*, **70**(5):898–907.
- Girskis, K. M., Stergachis, A. B., DeGennaro, E. M., Doan, R. N., Qian, X., Johnson, M. B., Wang, P. P., Sejourne, G. M., Nagy, M. A., Pollina, E. A., et al., 2021. Rewiring of human neurodevelopmental gene regulatory programs by human accelerated regions. *Neuron*, :109, 1–13.

- Gittelman, R. M., Hun, E., Ay, F., Madeoy, J., Pennacchio, L., Noble, W. S., Hawkins, R. D., and Akey, J. M., 2015. Comprehensive identification and analysis of human accelerated regulatory DNA. *Genome Research*, **25**(9):1245–1255.
- Guilmatre, A., Dubourg, C., Mosca, A.-L., Legallie, S., Goldenberg, A., Drouin-Garraud, V., Layet, V., Rosier, A., Briault, S., Bonnet-Brilhault, F., et al., 2009. Recurrent rearrangements in synaptic and neurodevelopmental genes and shared biologic pathways in schizophrenia, autism, and mental retardation. *Archives of General Psychiatry*, **66**(9):947.
- Horlbeck, M. A., Gilbert, L. A., Villalta, J. E., Adamson, B., Pak, R. A., Chen, Y., Fields, A. P., Park, C. Y., Corn, J. E., Kampmann, M., et al., 2016. Compact and highly active next-generation libraries for CRISPR-mediated gene repression and activation. *eLife*, **5**:e19760.
- Houbaert, X., Zhang, C.-L., Gambino, F., Lepleux, M., Deshors, M., Normand, E., Levet, F., Ramos, M., Billuart, P., Chelly, J., et al., 2013. Target-specific vulnerability of excitatory synapses leads to deficits in associative memory in a model of intellectual disorder. *Journal of Neuroscience*, **33**(34):13805–13819.
- Huang, J., Zhong, Y., Makohon-Moore, A. P., White, T., Jasin, M., Norell, M. A., Wheeler, W. C., and Iacobuzio-Donahue, C. A., 2022. Evidence for reduced BRCA2 functional activity in Homo sapiens after divergence from the chimpanzee-human last common ancestor. *Cell Reports*, **39**(5):110771.
- Huzard, D., Martin, M., Maingret, F., Chemin, J., Jeanneteau, F., Mery, P.-F., Fossat, P., Bourinet, E., and François, A., 2022. The impact of C-tactile low-threshold mechanoreceptors on affective touch and social interactions in mice. *Science Advances*, **8**(26):eabo7566.
- Hyman, S. L., Levy, S. E., Myers, S. M., COUNCIL ON CHILDREN WITH DISABILITIES, S. O. D., and PEDIATRICS, B., 2020. Identification, evaluation, and management of children with autism spectrum disorder. *Pediatrics*, **145**(1):e20193447.
- Inoue, T., Ota, M., Ogawa, M., Mikoshiba, K., and Aruga, J., 2007. Zic1 and Zic3 regulate medial forebrain development through expansion of neuronal progenitors. *Journal of Neuroscience*, **27**(20):5461–5473.
- Iossifov, I., O’Roak, B. J., Sanders, S. J., Ronemus, M., Krumm, N., Levy, D., Stessman, H. A., Witherspoon, K. T., Vives, L., Patterson, K. E., et al., 2014. The contribution of de novo coding mutations to autism spectrum disorder. *Nature*, **515**(7526):216–221.
- Jacquemont, S., Coe, B. P., Hersch, M., Duyzend, M. H., Krumm, N., Bergmann, S., Beckmann, J. S., Rosenfeld, J. A., and Eichler, E. E., 2014. A higher mutational burden in females supports a “female protective model” in neurodevelopmental disorders. *American Journal of Human Genetics*, **94**(3):415–425.
- Jin, Y. and Kong, J., 2017. Transcutaneous vagus nerve stimulation: A promising method for treatment of autism spectrum disorders. *Frontiers in Neuroscience*, **10**:609.
- Kamm, G. B., Pisciotto, F., Kliger, R., and Franchini, L. F., 2013. The developmental brain gene NPAS3 contains the largest number of accelerated regulatory sequences in the human genome. *Molecular Biology and Evolution*, **30**(5):1088–1102.
- Karczewski, K. J., Francioli, L. C., Tiao, G., Cummings, B. B., Alföldi, J., Wang, Q., Collins, R. L., Laricchia, K. M., Ganna, A., Birnbaum, D. P., et al., 2020. The mutational constraint spectrum quantified from variation in 141,456 humans. *Nature*, **581**(78097809):434–443.
- Kasher, P. R., Schertz, K. E., Thomas, M., Jackson, A., Annunziata, S., Ballesta-Martinez, M. J., Campeau, P. M., Clayton, P. E., Eaton, J. L., Granata, T., et al., 2016. Small 6q16.1 deletions encompassing POU3F2 cause susceptibility to obesity and variable developmental delay with intellectual disability. *American Journal of Human Genetics*, **98**(2):363–372.
- Keough, K. C., Whalen, S., Inoue, F., Przytycki, P. F., Fair, T., Deng, C., Steyert, M., Ryu, H., Lindblad-Toh, K., Karlsson, E., et al., 2023. Three-dimensional genome rewiring in loci with human accelerated regions. *Science*, **380**(6643):eabm1696.
- Khan, A., Riudavets Puig, R., Boddie, P., and Mathelier, A., 2021. BiasAway: Command-line and web server to generate nucleotide composition-matched DNA background sequences. *Bioinformatics*, **37**(11):1607–1609.
- Kim, M. J., Oksenberg, N., Hoffmann, T. J., Vaisse, C., and Ahituv, N., 2014. Functional characterization of SIM1-associated enhancers. *Human Molecular Genetics*, **23**(7):1700–1708.
- Kircher, M., Xiong, C., Martin, B., Schubach, M., Inoue, F., Bell, R. J. A., Costello, J. F., Shendure, J., and Ahituv, N., 2019. Saturation mutagenesis of twenty disease-associated regulatory elements at single base-pair resolution. *Nature Communications*, **10**(11):3583.
- Kundaje, A., Meuleman, W., Ernst, J., Bilenky, M., Yen, A., Heravi-Moussavi, A., Kheradpour, P., Zhang, Z., Wang, J., Ziller, M. J., et al., 2015. Integrative analysis of 111 reference human epigenomes. *Nature*, **518**(7539):317–330.
- Kvon, E. Z., Zhu, Y., Kelman, G., Novak, C. S., Plajzer-Frick, I., Kato, M., Garvin, T. H., Pham, Q., Harrington, A. N., Hunter, R. D., et al., 2020. Vivo Interrogation Reveals Phenotypic Impact of Human Enhancer Variants. *Cell*, **180**(6):1262–1271.
- Lake, B. B., Chen, S., Sos, B. C., Fan, J., Kaeser, G. E., Yung, Y. C., Duong, T. E., Gao, D., Chun, J., Kharchenko, P. V., et al., 2018. Integrative single-cell analysis of transcriptional and epigenetic states in the human adult brain. *Nature Biotechnology*, **36**(1):70–80.
- Lek, M., Karczewski, K. J., Minikel, E. V., Samocha, K. E., Banks, E., Fennell, T., O’Donnell-Luria, A. H., Ware, J. S., Hill, A. J., Cummings, B. B., et al., 2016. Analysis of protein-coding genetic variation in 60,706 humans. *Nature*, **536**(76167616):285–291.
- Li, C., Fleck, J. S., Martins-Costa, C., Burkard, T. R., Themann, J., Stuempflen, M., Peer, A. M., Vertesy, A., Littleboy, J. B., Esk, C., et al., 2023. Single-cell brain organoid screening identifies developmental defects in autism. *Nature*, **621**(7978):373–380.
- Li, H., 2011. A statistical framework for SNP calling, mutation discovery, association mapping and population genetic parameter estimation from sequencing data. *Bioinformatics*, **27**(21):2987–2993.
- Li, H. and Durbin, R., 2009. Fast and accurate short read alignment with Burrows-Wheeler transform. *Bioinformatics*, **25**(14):1754–1760.
- Li, H., Handsaker, B., Wysoker, A., Fennell, T., Ruan, J., Homer, N., Marth, G., Abecasis, G., and Durbin, R., 2009. The Sequence Alignment/Map format and SAMtools. *Bioinformatics*, **25**(16):2078–2079.
- Liao, Y., Smyth, G. K., and Shi, W., 2014. featureCounts: An efficient general purpose program for assigning sequence reads to genomic features. *Bioinformatics*, **30**(7):923–930.
- Lin, H.-C., He, Z., Ebert, S., Schörnig, M., Santel, M., Nikolova, M. T., Weigert, A., Hevers, W., Kasri, N. N., Taverna, E., et al., 2021. NGN2 induces diverse neuron types from human pluripotency. *Stem Cell Reports*, **16**(9):2118–2127.
- Lindblad-Toh, K., Garber, M., Zuk, O., Lin, M. F., Parker, B. J., Washietl, S., Kheradpour, P., Ernst, J., Jordan, G., Mauceli, E., et al., 2011. A high-resolution map of human evolutionary constraint using 29 mammals. *Nature*, **478**(73707370):476–482.
- Liu, X., Malenfant, P., Reesor, C., Lee, A., Hudson, M. L., Harvard, C., Qiao, Y., Persico, A. M., Cohen, I. L., Chudley, A. E., et al., 2011. 2p15-p16.1 microdeletion syndrome: Molecular characterization and association of the OTX1 and XPO1 genes with autism spectrum disorders. *European Journal of Human Genetics*, **19**(12):1264–1270.
- Maenner, M. J., Warren, Z., Williams, A. R., Amoakohene, E., Bakian, A. V., Bilder, D. A., Durkin, M. S., Fitzgerald, R. T., Fournier, S. M., Hughes, M. M., et al., 2023. Prevalence and characteristics of autism spectrum disorder among children aged 8 years - autism and developmental disabilities monitoring network, 11 sites, united states, 2020. *Morbidity and Mortality Weekly Report. Surveillance Summaries*, **72**(2):1–14.
- Magdaleno, S., Jensen, P., Brumwell, C. L., Seal, A., Lehman, K., Asbury, A., Cheung, T., Cornelius, T., Batten, D. M., Eden, C., et al., 2006. BGEM: An *In Situ* hybridization database of gene expression in the embryonic and adult mouse nervous system. *PLoS Biology*, **4**(4):e86.
- Marshall, C. R., Noor, A., Vincent, J. B., Lionel, A. C., Feuk, L., Skaug, J., Shago, M., Moessner, R., Pinto, D., Ren, Y., et al., 2008. Structural variation of chromosomes in autism spectrum disorder. *American Journal of Human Genetics*, **82**(2):477–488.
- Martin, M., 2011. Cutadapt removes adapter sequences from high-throughput sequencing reads. *EMBnet journal*, **17**(1):10–12.
- Matharu, N., Rattanasopha, S., Tamura, S., Maliskova, L., Wang, Y., Bernard, A., Hardin, A., Eckalbar, W. L., Vaisse, C., and Ahituv, N., et al., 2019. CRISPR-mediated activation of a promoter or enhancer rescues obesity caused by haploinsufficiency. *Science*, **363**(6424):eaau0629.
- McLean, C. Y., Bristol, D., Hiller, M., Clarke, S. L., Schaar, B. T., Lowe, C. B., Wenger, A. M., and Bejerano, G., 2010. GREAT improves functional interpretation of cis-regulatory regions. *Nature Biotechnology*, **28**(5):495–501.
- Melnikov, A., Zhang, X., Rogov, P., Wang, L., and Mikkelsen, T. S., 2014. Massively parallel reporter assays in cultured mammalian cells. *Journal of Visualized Experiments*, (90):e51719.
- Mikhail, F. M., Lose, E. J., Robin, N. H., Descartes, M. D., Rutledge, K. D.,

- Rutledge, S. L., Korf, B. R., and Carroll, A. J., 2011. Clinically relevant single gene or intragenic deletions encompassing critical neurodevelopmental genes in patients with developmental delay, mental retardation, and/or autism spectrum disorders. *American Journal of Medical Genetics. Part A*, **155A**(10):2386–2396.
- Mikl, M., Eletto, D., Nijim, M., Lee, M., Lafzi, A., Mhamedi, F., David, O., Sain, S. B., Handler, K., and Moor, A. E., et al., 2022. A massively parallel reporter assay reveals focused and broadly encoded RNA localization signals in neurons. *Nucleic Acids Research*, **50**(18):10643–10664.
- Miyoshi, G., Young, A., Petros, T., Karayannis, T., McKenzie Chang, M., Lavado, A., Iwano, T., Nakajima, M., Taniguchi, H., Huang, Z. J., et al., 2015. Prox1 regulates the subtype-specific development of caudal ganglionic eminence-derived gabaergic cortical interneurons. *Journal of Neuroscience*, **35**(37):12869–12889.
- Montani, C., Gritti, L., Beretta, S., Verpelli, C., and Sala, C., 2019. The synaptic and neuronal functions of the X-linked intellectual disability protein interleukin-1 receptor accessory protein like 1 (IL1RAPL1). *Developmental Neurobiology*, **79**(1):85–95.
- Montani, C., Ramos-Brossier, M., Ponzoni, L., Gritti, L., Cwetsch, A. W., Braida, D., Saillour, Y., Terragni, B., Mantegazza, M., Sala, M., et al., 2017. The X-linked intellectual disability protein IL1RAPL1 regulates dendrite complexity. *Journal of Neuroscience*, **37**(28):6606–6627.
- Morrow, E. M., Yoo, S.-Y., Flavell, S. W., Kim, T.-K., Lin, Y., Hill, R. S., Mukaddes, N. M., Balkhy, S., Gascon, G., Hashmi, A., et al., 2008. Identifying autism loci and genes by tracing recent shared ancestry. *Science*, **321**(5886):218–223.
- Mulvey, B. and Dougherty, J. D., 2021. Transcriptional-regulatory convergence across functional MDD risk variants identified by massively parallel reporter assays. *Translational Psychiatry*, **11**(11):1–13.
- Navarro Gonzalez, J., Zweig, A. S., Speir, M. L., Schmelter, D., Rosenbloom, K. R., Raney, B. J., Powell, C. C., Nassar, L. R., Maulding, N. D., Lee, C. M., et al., 2021. The UCSC Genome Browser database: 2021 update. *Nucleic Acids Research*, **49**(D1):D1046–D1057.
- Nelson, S. B. and Valakh, V., 2015. Excitatory/inhibitory balance and circuit homeostasis in autism spectrum disorders. *Neuron*, **87**(4):684–698.
- Novara, F., Beri, S., Giorda, R., Ortibus, E., Nageshappa, S., Darra, F., Dalla Bernardina, B., Zuffardi, O., and Van Esch, H., 2010. Refining the phenotype associated with MEF2C haploinsufficiency. *Clinical Genetics*, **78**(5):471–477.
- Oksenberg, N., Stevison, L., Wall, J. D., and Ahituv, N., 2013. Function and regulation of AUTS2, a gene implicated in autism and human evolution. *PLoS Genetics*, **9**(1):e1003221.
- Orefice, L. L., Zimmerman, A. L., Chirila, A. M., Sleboda, S. J., Head, J. P., and Ginty, D. D., 2016. Peripheral mechanosensory neuron dysfunction underlies tactile and behavioral deficits in mouse models of ASDs. *Cell*, **166**(2):299–313.
- Osterwalder, M., Barozzi, I., Tissieres, V., Fukuda-Yuzawa, Y., Mannion, B. J., Afzal, S. Y., Lee, E. A., Zhu, Y., Plajzer-Frick, I., Pickle, C. S., et al., 2018. Enhancer redundancy provides phenotypic robustness in mammalian development. *Nature*, **554**(7691):239–243.
- Osterwalder, M., Tran, S., Hunter, R. D., Meko, E. M., von Maydell, K., Harrington, A. N., Godoy, J., Novak, C. S., Plajzer-Frick, I., Zhu, Y., et al., 2022. Characterization of mammalian in vivo enhancers using mouse transgenesis and CRISPR genome editing. *Methods in Molecular Biology*, **2403**:147–186.
- Paulsen, B., Velasco, S., Kedaigle, A. J., Pigoni, M., Quadrato, G., Deo, A. J., Adiconis, X., Uzquiano, A., Sartore, R., Yang, S. M., et al., 2022. Autism genes converge on asynchronous development of shared neuron classes. *Nature*, **602**(78967896):268–273.
- Pavlovsky, A., Gianfelice, A., Pallotto, M., Zanchi, A., Vara, H., Khelfaoui, M., Valnegri, P., Rezai, X., Bassani, S., Brambilla, D., et al., 2010. A post-synaptic signaling pathway that may account for the cognitive defect due to IL1RAPL1 mutation. *Current Biology*, **20**(2):103–115.
- Perez, A. R., Pritykin, Y., Vidigal, J. A., Chhangawala, S., Zamparo, L., Leslie, C. S., and Ventura, A., 2017. GuideScan software for improved single and paired CRISPR guide RNA design. *Nature Biotechnology*, **35**(44):347–349.
- Phillips, K. L., Schieve, L. A., Visser, S., Boulet, S., Sharma, A. J., Kogan, M. D., Boyle, C. A., and Yeargin-Allsopp, M., 2014. Prevalence and impact of unhealthy weight in a national sample of us adolescents with autism and other learning and behavioral disabilities. *Maternal and Child Health Journal*, **18**(8):1964–1975.
- Pieraccioli, M., Nicolai, S., Pitolli, C., Agostini, M., Antonov, A., Malewicz, M., Knight, R. A., Raschella, G., and Melino, G., 2018. ZNF281 inhibits neuronal differentiation and is a prognostic marker for neuroblastoma. *Proceedings of the National Academy of Sciences*, **115**(28):7356–7361.
- Pintacuda, G., Hsu, Y.-H. H., Tsaou, K., Li, K. W., Martín, J. M., Riseman, J., Biagini, J. C., Ching, J. K. T., Mena, D., Gonzalez-Lozano, M. A., et al., 2023. Protein interaction studies in human induced neurons indicate convergent biology underlying autism spectrum disorders. *Cell Genomics*, **3**(3).
- Piton, A., Michaud, J. L., Peng, H., Aradhya, S., Gauthier, J., Mottron, L., Champagne, N., Lafrenière, R. G., Hamdan, F. F., team, S., et al., 2008. Mutations in the calcium-related gene IL1RAPL1 are associated with autism. *Human Molecular Genetics*, **17**(24):3965–3974.
- Pollard, K. S., Salama, S. R., Lambert, N., Lambot, M. A., Coppens, S., Pedersen, J. S., Katzman, S., King, B., Onodera, C., Siepel, A., et al., 2006. An RNA gene expressed during cortical development evolved rapidly in humans. *Nature*, **443**(7108):167–172.
- Polychronopoulos, D., King, J. W. D., Nash, A. J., Tan, G., and Lenhard, B., 2017. Conserved non-coding elements: Developmental gene regulation meets genome organization. *Nucleic Acids Research*, **45**(22):12611–12624.
- Prabhakar, S., Visel, A., Akiyama, J. A., Shoukry, M., Lewis, K. D., Holt, A., Plajzer-Frick, I., Morrison, H., Fitzpatrick, D. R., Afzal, V., et al., 2008. Human-specific gain of function in a developmental enhancer. *Science*, **321**(5894):1346–1350.
- Quang, D., Chen, Y., and Xie, X., 2015. DANN: A deep learning approach for annotating the pathogenicity of genetic variants. *Bioinformatics*, **31**(5):761–763.
- Quinlan, A. R. and Hall, I. M., 2010. BEDTools: A flexible suite of utilities for comparing genomic features. *Bioinformatics*, **26**(6):841–842.
- Ramachandrapa, S., Raimondo, A., Cali, A. M., Keogh, J. M., Henning, E., Saeed, S., Thompson, A., Garg, S., Bochukova, E. G., Brage, S., et al., 2013. Rare variants in single-minded 1 (SIM1) are associated with severe obesity. *Journal of Clinical Investigation*, **123**(7):3042–3050.
- Rentzsch, P., Witten, D., Cooper, G. M., Shendure, J., and Kircher, M., 2019. CADD: Predicting the deleteriousness of variants throughout the human genome. *Nucleic Acids Research*, **47**(D1):D886–D894.
- Replogle, J. M., Norman, T. M., Xu, A., Hussmann, J. A., Chen, J., Cogan, J. Z., Meer, E. J., Terry, J. M., Riordan, D. P., Srinivas, N., et al., 2020. Combinatorial single-cell CRISPR screens by direct guide RNA capture and targeted sequencing. *Nature Biotechnology*, **38**(8):954–961.
- Robinson, E. B., Lichtenstein, P., Anckarsäter, H., Happé, F., and Ronald, A., 2013. Examining and interpreting the female protective effect against autistic behavior. *Proceedings of the National Academy of Sciences*, **110**(13):5258–5262.
- Ruzzo, E. K., Pérez-Cano, L., Jung, J.-Y., Wang, L.-k., Kashef-Haghighi, D., Hartl, C., Singh, C., Xu, J., Hoekstra, J. N., Leventhal, O., et al., 2019. Inherited and de novo genetic risk for autism impacts shared networks. *Cell*, **178**(4):850–866.e26.
- Satterstrom, F. K., Kosmicki, J. A., Wang, J., Breen, M. S., De Rubeis, S., An, J.-Y., Peng, M., Collins, R., Grove, J., Klei, L., et al., 2020. Large-scale exome sequencing study implicates both developmental and functional changes in the neurobiology of autism. *Cell*, **180**(3):568–584.e23.
- Schmitz-Abe, K., Sanchez-Schmitz, G., Doan, R. N., Hill, R. S., Chahrour, M. H., Mehta, B. K., Servattalab, S., Ataman, B., Lam, A.-T. N., Morrow, E. M., et al., 2020. Homozygous deletions implicate non-coding epigenetic marks in autism spectrum disorder. *Scientific Reports*, **10**(11):14045.
- Schultz, J. M., Khan, S. N., Ahmed, Z. M., Riazuddin, S., Waryah, A. M., Chhatre, D., Starost, M. F., Ploplis, B., Buckley, S., Velásquez, D., et al., 2009. Noncoding mutations of HGF are associated with nonsyndromic hearing loss, DFNB39. *American Journal of Human Genetics*, **85**(1):25–39.
- Sebat, J., Lakshmi, B., Malhotra, D., Troge, J., Lese-Martin, C., Walsh, T., Yamrom, B., Yoon, S., Krasnitz, A., Kendall, J., et al., 2007. Strong association of de novo copy number mutations with autism. *Science*, **316**(5823):445–449.
- Shihab, H. A., Rogers, M. F., Gough, J., Mort, M., Cooper, D. N., Day, I. N. M., Gaunt, T. R., and Campbell, C., 2015. An integrative approach to predicting the functional effects of non-coding and coding sequence variation. *Bioinformatics*, **31**(10):1536–1543.
- 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., et al., 2018. De novo mutations in regulatory elements in neurodevelopmental disorders. *Nature*, **555**(76987698):611–616.

- Siepel, A., Bejerano, G., Pedersen, J. S., Hinrichs, A. S., Hou, M., Rosenbloom, K., Clawson, H., Spieth, J., Hillier, L. W., Richards, S., et al., 2005. Evolutionarily conserved elements in vertebrate, insect, worm, and yeast genomes. *Genome Research*, **15**(8):1034–1050.
- Singh, T., Poterba, T., Curtis, D., Akil, H., Al Eissa, M., Barchas, J. D., Bass, N., Bigdeli, T. B., Breen, G., Bromet, E. J., et al., 2022. Rare coding variants in ten genes confer substantial risk for schizophrenia. *Nature*, **604**(79067906):509–516.
- Smith, T., Heger, A., and Sudbery, I., 2017. UMI-tools: Modeling sequencing errors in unique molecular identifiers to improve quantification accuracy. *Genome Research*, **27**(3):491–499.
- Snetkova, V., Ypsilanti, A. R., Akiyama, J. A., Mannion, B. J., Plajzer-Frick, I., Novak, C. S., Harrington, A. N., Pham, Q. T., Kato, M., Zhu, Y., et al., 2021. Ultraconserved enhancer function does not require perfect sequence conservation. *Nature Genetics*, **53**(44):521–528.
- Song, J. H. T., Lowe, C. B., and Kingsley, D. M., 2018. Characterization of a human-specific tandem repeat associated with bipolar disorder and schizophrenia. *American Journal of Human Genetics*, **103**(3):421–430.
- Song, M., Pebworth, M.-P., Yang, X., Abnoui, A., Fan, C., Wen, J., Rosen, J. D., Choudhary, M. N. K., Cui, X., Jones, I. R., et al., 2020. Cell-type-specific 3D epigenomes in the developing human cortex. *Nature*, **587**(7835):644–649.
- Srinivasan, S., Bettella, F., Mattingsdal, M., Wang, Y., Witoelar, A., Schork, A. J., Thompson, W. K., Zuber, V., Winsvold, B. S., Zwart, J. A., et al., 2016. Genetic markers of human evolution are enriched in schizophrenia. *Biological Psychiatry*, **80**(4):284–292.
- Szklarczyk, D., Gable, A. L., Lyon, D., Junge, A., Wyder, S., Huerta-Cepas, J., Simonovic, M., Doncheva, N. T., Morris, J. H., Bork, P., et al., 2019. STRING v11: Protein–protein association networks with increased coverage, supporting functional discovery in genome-wide experimental datasets. *Nucleic Acids Research*, **47**(Database issue):D607–D613.
- Tabolacci, E., Pomponi, M. G., Pietrobono, R., Terracciano, A., Chiurazzi, P., and Neri, G., 2006. A truncating mutation in the IL1RAPL1 gene is responsible for X-linked mental retardation in the MRX21 family. *American Journal of Medical Genetics Part A*, **140A**(5):482–487.
- Tian, R., Gachechiladze, M. A., Ludwig, C. H., Laurie, M. T., Hong, J. Y., Nathaniel, D., Prabhu, A. V., Fernandopulle, M. S., Patel, R., Abshari, M., et al., 2019. CRISPR interference-based platform for multimodal genetic screens in human iPSC-derived neurons. *Neuron*, **104**(2):239–255.e12.
- Tuncay, I. O., Parmalee, N. L., Khalil, R., Kaur, K., Kumar, A., Jimale, M., Howe, J. L., Goodspeed, K., Evans, P., Alzghoul, L., et al., 2022. Analysis of recent shared ancestry in a familial cohort identifies coding and noncoding autism spectrum disorder variants. *npj Genomic Medicine*, **7**(11):1–13.
- Uebbing, S., Gockley, J., Reilly, S. K., Kocher, A. A., Geller, E., Gandotra, N., Scharfe, C., Cotney, J., and Noonan, J. P., 2021. Massively parallel discovery of human-specific substitutions that alter enhancer activity. *Proceedings of the National Academy of Sciences*, **118**(2).
- Valnegri, P., Montrasio, C., Brambilla, D., Ko, J., Passafaro, M., and Sala, C., 2011. The X-linked intellectual disability protein IL1RAPL1 regulates excitatory synapse formation by binding PTP δ and RhoGAP2. *Human Molecular Genetics*, **20**(24):4797–4809.
- Van der Auwera, G. A., Carneiro, M. O., Hartl, C., Poplin, R., Del Angel, G., Levy-Moonshine, A., Jordan, T., Shakir, K., Roazen, D., Thibault, J., et al., 2013. From FastQ data to high confidence variant calls: The Genome Analysis Toolkit best practices pipeline. *Current Protocols in Bioinformatics*, **43**:1–11.
- Visel, A., Minovitsky, S., Dubchak, I., and Pennacchio, L. A., 2007. VISTA Enhancer Browser—a database of tissue-specific human enhancers. *Nucleic Acids Research*, **35**(Database issue):D88–92.
- Wang, L., Heffner, C., Vong, K. I., Barrows, C., Ha, Y.-J., Lee, S., Lara-Gonzalez, P., Jhamb, I., Van Der Meer, D., Loughnan, R., et al., 2023. TMEM161B modulates radial glial scaffolding in neocortical development. *Proceedings of the National Academy of Sciences*, **120**(4):e2209983120.
- Weaving, L. S., Christodoulou, J., Williamson, S. L., Friend, K. L., McKenzie, O. L. D., Archer, H., Evans, J., Clarke, A., Pelka, G. J., Tam, P. P. L., et al., 2004. Mutations of CDKL5 cause a severe neurodevelopmental disorder with infantile spasms and mental retardation. *American Journal of Human Genetics*, **75**(6):1079–1093.
- Whalen, S., Inoue, F., Ryu, H., Fair, T., Markenscoff-Papadimitriou, E., Keough, K., Kircher, M., Martin, B., Alvarado, B., Elor, O., et al., 2023. Machine learning dissection of human accelerated regions in primate neurodevelopment. *Neuron*, **111**(6):857–873.e8.
- Won, H., de la Torre-Ubieta, L., Stein, J. L., Parikshak, N. N., Huang, J., Opland, C. K., Gandal, M. J., Sutton, G. J., Hormozdiari, F., Lu, D., et al., 2016. Chromosome conformation elucidates regulatory relationships in developing human brain. *Nature*, **538**(7626):523–527.
- Wu, Y., Arai, A. C., Rumbaugh, G., Srivastava, A. K., Turner, G., Hayashi, T., Suzuki, E., Jiang, Y., Zhang, L., Rodriguez, J., et al., 2007. Mutations in ionotropic AMPA receptor 3 alter channel properties and are associated with moderate cognitive impairment in humans. *Proceedings of the National Academy of Sciences*, **104**(46):18163–18168.
- Xu, K., Schadt, E. E., Pollard, K. S., Roussos, P., and Dudley, J. T., 2015. Genomic and network patterns of schizophrenia genetic variation in human evolutionary accelerated regions. *Molecular Biology and Evolution*, **32**(5):1148–1160.
- Yasumura, M., Yoshida, T., Yamazaki, M., Abe, M., Natsume, R., Kanno, K., Uemura, T., Takao, K., Sakimura, K., Kikusui, T., et al., 2014. IL1RAPL1 knockout mice show spine density decrease, learning deficiency, hyperactivity and reduced anxiety-like behaviours. *Scientific Reports*, **4**:6613.
- Yu, G., Wang, L. G., Han, Y., and He, Q. Y., 2012. clusterProfiler: An R package for comparing biological themes among gene clusters. *OMICS*, **16**(5):284–287.
- Yu, T. W., Chahrouh, M. H., Coulter, M. E., Jiralerspong, S., Okamura-Ikeda, K., Ataman, B., Schmitz-Abe, K., Harmin, D. A., Adli, M., Malik, A. N., et al., 2013. Using whole-exome sequencing to identify inherited causes of autism. *Neuron*, **77**(2):259–273.
- Zhou, J., Theesfeld, C. L., Yao, K., Chen, K. M., Wong, A. K., and Troyanskaya, O. G., 2018. Deep learning sequence-based ab initio prediction of variant effects on expression and disease risk. *Nature Genetics*, **50**(8):1171–1179.
- Zhou, J. and Troyanskaya, O. G., 2015. Predicting effects of noncoding variants with deep learning–based sequence model. *Nature Methods*, **12**(10):931–934.
- Zhou, X., Feliciano, P., Shu, C., Wang, T., Astrovskaya, I., Hall, J. B., Obiajulu, J. U., Wright, J. R., Murali, S. C., Xu, S. X., et al., 2022. Integrating de novo and inherited variants in 42,607 autism cases identifies mutations in new moderate-risk genes. *Nature Genetics*, :1–15.