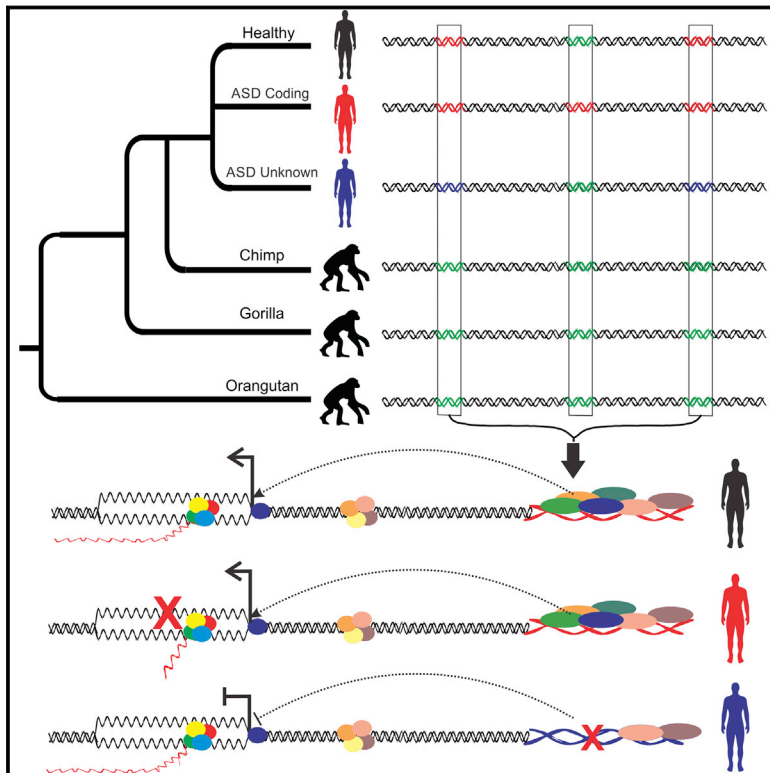# Cell

Mutations in Human Accelerated Regions Disrupt Cognition and Social Behavior

Graphical Abstract

Authors
Ryan N. Doan, Byoung-Il Bae, Beatriz Cubelos, ..., Generoso G. Gascon, Marta Nieto, Christopher A. Walsh

Correspondence
christopher.walsh@childrens.harvard.edu

In Brief
Mutations in genomic loci that have undergone accelerated evolution in humans are associated with increased risk for autism.

Highlights

- Human accelerated regions exhibit regulatory activity during neural development

- De novo CNVs impacting HARs are enriched in individuals with ASD

- Biallelic HAR mutations underlie up to 5% of consanguineous ASD cases

- Regulatory mutations reveal novel genetic architecture of ASD

CrossMark

CellPress

Mutations in Human Accelerated Regions Disrupt Cognition and Social Behavior

Ryan N. Doan,[1] Byoung-Il Bae,[1,12] Beatriz Cubelos,[2,3] Cindy Chang,[1] Amer A. Hossain,[1] Samira Al-Saad,[4] Nahit M. Mukaddes,[5] Ozgur Oner,[6] Muna Al-Saffar,[1,7] Soher Balkhy,[8] Generoso G. Gascon,[9] The Homozygosity Mapping Consortium for Autism, Marta Nieto,[3] and Christopher A. Walsh[1,10,11,13,*]

[1]Division of Genetics and Genomics, Manton Center for Orphan Disease Research, and Howard Hughes Medical Institute, Boston Children's Hospital, Boston, MA 02115, USA
[2]Department of Molecular Biology, Centro de Biología Molecular 'Severo Ochoa', Universidad Autonoma de Madrid, UAM-CSIC, Nicolas Cabrera 1, 28049 Madrid, Spain
[3]Department of Molecular and Cellular Biology, Centro Nacional de Biotecnología, CNB-CSIC, Darwin 3, Campus de Cantoblanco, 28049 Madrid, Spain
[4]Kuwait Center for Autism, 73455 Kuwait City, Kuwait
[5]Istanbul Institute of Child and Adolescent Psychiatry, 34365 Istanbul, Turkey
[6]Department of Child and Adolescent Psychiatry, Bahcesehir University School of Medicine, 34353 Istanbul, Turkey
[7]Department of Pediatrics, College of Medicine and Health Sciences, United Arab Emirates University, PO Box 17666, Al-Ain, United Arab Emirates
[8]Department of Pediatrics, King Faisal Specialist Hospital and Research Center, Jeddah 21499, Kingdom of Saudi Arabia
[9]Department of Neurology (Pediatric Neurology), Massachusetts General Hospital, Boston, MA 02114, USA
[10]Broad Institute of MIT and Harvard, Cambridge, MA 02142, USA
[11]Departments of Pediatrics and Neurology, Harvard Medical School, Boston, MA 02115, USA
[12]Present address: Department of Neurosurgery, School of Medicine, Yale University, New Haven, CT 06510, USA
[13]Lead Contact
*Correspondence: christopher.walsh@childrens.harvard.edu
http://dx.doi.org/10.1016/j.cell.2016.08.071

SUMMARY

Comparative analyses have identified genomic regions potentially involved in human evolution but do not directly assess function. Human accelerated regions (HARs) represent conserved genomic loci with elevated divergence in humans. If some HARs regulate human-specific social and behavioral traits, then mutations would likely impact cognitive and social disorders. Strikingly, rare biallelic point mutations—identified by whole-genome and targeted "HAR-ome" sequencing—showed a significant excess in individuals with ASD whose parents share common ancestry compared to familial controls, suggesting a contribution in 5% of consanguineous ASD cases. Using chromatin interaction sequencing, massively parallel reporter assays (MPRA), and transgenic mice, we identified disease-linked, biallelic HAR mutations in active enhancers for *CUX1*, *PTBP2*, *GPC4*, *CDKL5*, and other genes implicated in neural function, ASD, or both. Our data provide genetic evidence that specific HARs are essential for normal development, consistent with suggestions that their evolutionary changes may have altered social and/or cognitive behavior.

INTRODUCTION

The complex social and cognitive behaviors that characterize modern humans—including language, civilization, and society—reflect human-specific neurodevelopmental mechanisms that have both genetic and cultural roots (Miller et al., 2012; Somel et al., 2013). The ability of comparative genomics to identify loci that are highly different between humans and other species promises to allow their association to some of these human-specific traits (McLean et al., 2011). The accelerated divergence of human accelerated regions (HARs) between humans and other species has been suggested to reflect potential roles in the evolution of human-specific traits (Bird et al., 2007; Bush and Lahn, 2008; Capra et al., 2013; Lindblad-Toh et al., 2011; Pollard et al., 2006a; Prabhakar et al., 2008). The ultimate functional test of the importance of HARs comes in assessing the impact of deleterious mutations in HAR sequences, and virtually nothing is known about any potential functional effects of HAR mutations. Association of functional HAR mutations with disorders of cognition or social behavior could provide novel insights not only into the pathogenesis of important human disorders, but also into the mechanisms by which human-specific patterns of social and cognitive behavior originated.

The apparent loss of negative evolutionary pressure on HARs between humans and other primates could reflect any of three potential processes: neutral substitutions accumulating in sequences that have lost essential function, guanine-cytosine (GC)-biased gene-conversion events, or a switch from strictly negative selection to include positive selection. Variation in

healthy individuals and close hominin relatives suggests that rare mutations in HARs are generally deleterious, reflected in a paucity of recent alleles (i.e., non-ancestral alleles, 8.3%) and fixation of 96% of ancestral alleles (Burbano et al., 2012). Nonetheless, it appears that at least some HARs underwent a switch to positive selection in humans, possibly followed by a switch back to negative selection within stable human populations.

Several recent studies have elucidated potential functions of HARs using in silico predictions and animal models. Most HARs lie within 1 Mb of a gene and show enrichment for *CTCF* binding, suggesting that some HARs act as transcriptional enhancers through physical contact with promoters (Capra et al., 2013). Other HARs, such as HAR1, appear to encode RNAs (Pollard et al., 2006b). The enrichment of epigenetic signatures suggests that 29% of HARs function as enhancers during brain, heart, and limb development (Capra et al., 2013). Transgenic constructs encoding human or chimpanzee HAR sequences revealed species-specific functionality through expression levels or, in some cases, changes in anatomical regions of expression (Capra et al., 2013). Nonetheless, definitive proof of essential HARs requires the characterization of functional HAR mutations in humans, i.e., by the association with genetic disease.

In recent years, focus on regulatory mutations causing disease has increased, resulting in the identification of a handful of noncoding mutations. Unlike coding mutations, which are most often loss of function, regulatory mutations have the potential to be activating or inactivating of gene transcription in specific tissues (Bae et al., 2014; Weedon et al., 2014), typically due to gain or loss of essential transcription factor (TF) motifs. Regulatory mutations, including both biallelic and heterozygous, underlie a variety of human diseases (Weedon et al., 2014). The strong preservation of HARs between species identifies them as likely to be genetically essential and, hence, as favorable targets to identify undiscovered disease-associated mutations.

A subset of HARs shows evidence of neural function, with several within loci with significant associations to schizophrenia (Xu et al., 2015). Furthermore, GABAergic and glutamatergic genes were enriched among the HAR-associated schizophrenia genes. Interestingly, HAR-associated genes were enriched for processes involved in synaptic formation and exhibit a higher connectivity to regulatory networks in the prefrontal cortex (Xu et al., 2015). An intronic HAR with enhancer activity within *AUTS2*, a gene linked to Autism Spectrum Disorder (ASD) (Oksenberg et al., 2013), suggests that HAR variants could result in ASD, but no HAR mutations have yet been associated with any neurological diseases.

We utilized existing genomic data from healthy individuals and the Epigenomics Roadmap to demonstrate essential regulatory functions of HARs, particularly in neural tissues. We also performed 4C-sequencing, in combination with existing interaction data, to provide the first systematic map of target genes for more than 500 HARs. We investigated the mutational landscape of HARs by comparing de novo copy-number variations (CNVs) and biallelic point mutations in individuals with ASD and healthy controls, revealing a significant contribution of both types of mutations to ASD risk. We identify several specific HARs likely to have essential functions in the human brain that are potentially important targets of recent human brain evolution.

RESULTS

HARs Are Functionally Constrained

We investigated HARs for evidence of transcriptional regulatory functions by first defining their mutational landscape in healthy, diverse human populations. We identified 2,737 HARs from several recent publications (Bird et al., 2007; Bush and Lahn, 2008; Lindblad-Toh et al., 2011; Pollard et al., 2006a, 2006b; Prabhakar et al., 2008) (Table S1). Due in part to the identification methods, most HARs lie in intergenic and intronic noncoding regions (Figure S1), although several overlap promoters (48) or noncoding RNAs and pseudogenes (135). Within the Complete Genomics Diversity Panel (CG69) and low-coverage 1000 Genomes project (1000G), individual "HAR-omes" contain an average of 1,273 variants, with a 2-fold depletion of rare variants compared to coding, noncoding-conserved, and randomly selected loci (2.1% versus 4.5%, $p = 4.2 \times 10^{-44}$; 5.2%, $p = 4.7 \times 10^{-70}$; and 4.8%, $p = 2.0 \times 10^{-76}$; Table S2 and Figure 1A). While common HAR alleles exhibit a significant elevation in homozygosity (Figures 1B and S2), rare alleles are depleted for homozygosity, particularly for conserved nucleotides where they behave similar to coding mutations, suggesting possible damaging effects (Figure 1C).

To explore whether HAR variants occur at human-specific bases, we compared alleles from 1000G to their respective chimpanzee base, revealing that 11% occur at human-specific nucleotides within 1,401 HARs. While human-specific sites account for about 2% of HAR nucleotides, they are enriched for point mutations, including 690 rare alleles (2,255 of 12,973 sites; odds ratio [OR] = 6.6, 95% confidence interval [CI] 6.3–6.9, $p < 10^{-4}$), and display evidence of evolutionary constraint (i.e., fewer substitutions than expected by neutral models; average GERP-RS = 2.23) (Cooper et al., 2005). While enrichment of alleles at human-specific sites might represent recent mutational events or, more likely, polymorphic non-accelerated sites, 365 alleles do not revert to ancestral alleles, suggesting that a portion have been subject to recent mutations.

HARs Are Enriched for Neural Regulatory Elements

We expanded recent in silico predictions of developmental activity (Capra et al., 2013) by utilizing DNase I sensitivity, histone profiles, ChromHMM (Roadmap Epigenomics Consortium et al., 2015) segmentations, and in silico motif predictions. HARs show significant enrichment for fetal and adult brain H3K4me1 profiles ($p < 0.0001$), suggesting activity in the brain (Figure 1D). Furthermore, 81% of HARs overlapped marks of active transcription, promoters, or enhancers, including 45% in neural samples from ChromHMM (Figure 1E and Table S1), with many displaying tissue specificity (i.e., 168 in neural tissues).

We sought to define the TF motif landscape of HARs beyond the known enrichment for conserved TF sites (TFBS; Capra et al., 2013; Pollard et al., 2006a). HARs were enriched for TFBSs (78%, $p < 0.0001$, random sampling; Figure 1F) and high-quality predicted matrices (2.5-fold increase, 15,989 sites [Figure 1G]). Several enriched motifs are involved in neural developmental processes, including *MEF2A* ($p = 4.8 \times 10^{-5}$) and the SRY-related HMG-box gene 2 (*SOX2*) ($p = 2.9 \times 10^{-7}$) (Figure 1H), while others, like *ZNF333*, show broader expression. *SOX2* is
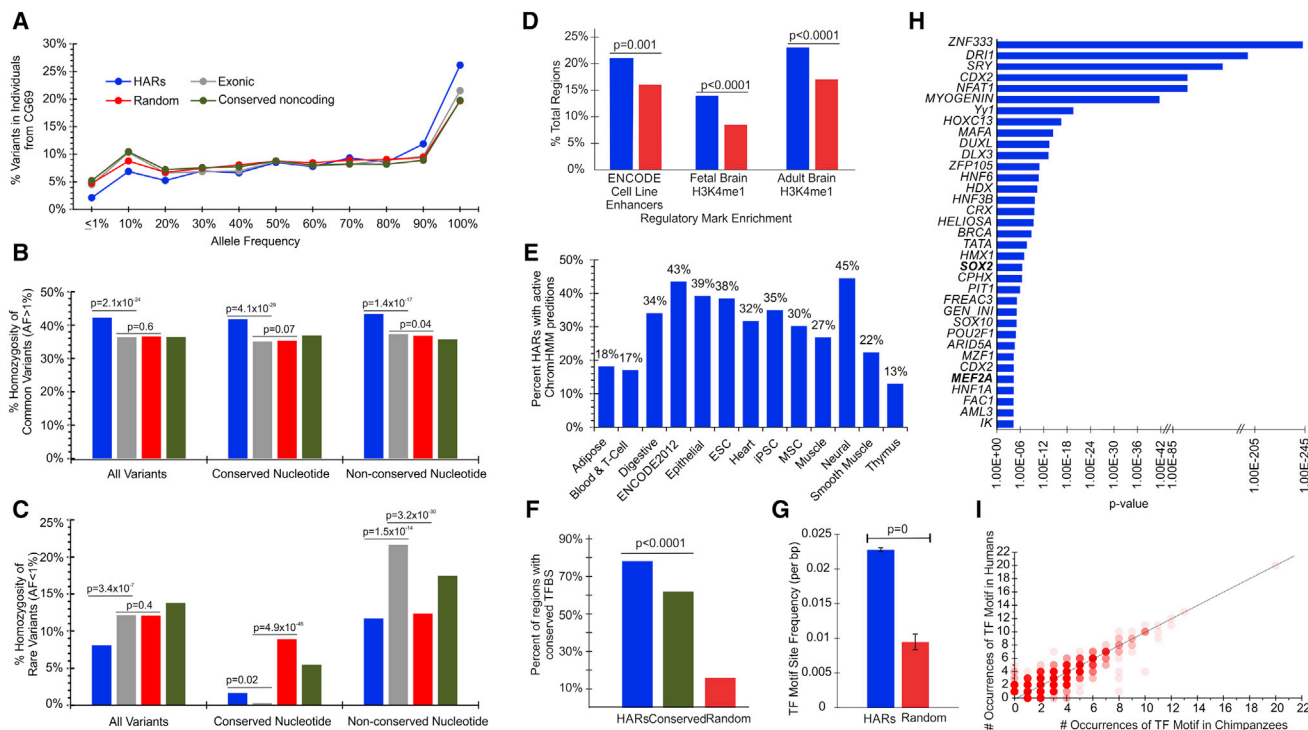
Figure 1. Evidence of Selective Pressures due to Regulatory Functionality of HARs

(A–C) Average distribution of variants by maximum AF in CG69. Average homozygosity in CG69 for all, conserved, and non-conserved nucleotides for (B) common and (C) rare variants.

(D–G) HARs are significantly enriched for regulatory marks in fetal and adult brain (Dunham et al., 2012; Roadmap Epigenomics Consortium et al., 2015) with the greatest activity (E) in neural tissues. HARs are enriched for (F) conserved TF motifs (10,000 random samples) and (G) frequency of predicted motifs.

(H) Enrichment of TF motifs within HARs as determined against random sequences by TRANSFAC.

(I) Human-specific alleles in HARs alter TF motifs in comparison to chimpanzee.

See also Figures S1 and S2 and Tables S1 and S2.

expressed in stem cells, where it is essential for renewal of neural progenitors (Ferri et al., 2004; Kelberman et al., 2008). Upon decreased *SOX2* expression, neural progenitors differentiate into neurons. The enrichment of *SOX2* motifs supports a role in neural development, particularly during neurogenesis. Human-specific nucleotides altered the frequencies of some motifs (54 increased, 72 decreased, Figure 1I), including *REST*, *CTCF*, and *NFIA*. This divergence of TF motifs may reveal human-specific processes in cortical development.

Many HARs Act as Regulatory Elements for Dosage-Sensitive and Constrained Neural Genes

We combined existing ChIA-PET and HiC data (Goh et al., 2012; Jin et al., 2013; Li et al., 2010, 2012) to map 576 HARs to promoter regions of more than 700 target genes, including 123 interactions with large intergenic noncoding RNAs (lincRNAs) (Ma et al., 2015; Table S1). More than 42% of HARs interact with flanking gene promoters, suggesting that proximity is generally a good predictor of HAR target genes in lieu of cell-type-specific interaction data. Of the 132 target genes overlapping OMIM disorders, 28 associate with ASD/ID (p = 0.008, random sampling, Table S3 and Figure 2A) and 38 cause neural phenotypes in mice. Since regulatory elements alter gene-expression levels, but not protein structure, we hypothesized

that, for pathogenicity, target genes should be dosage sensitive, i.e., haploinsufficiency (HI). Indeed, the HI (Huang et al., 2010) of flanking and target genes are significantly elevated (i.e., dosage sensitive; HI = 0.35, p = 3.8×10^{-38}; HI = 0.34, p = 9.0×10^{-13}; *t* test; Figure 2B). The elevated dosage sensitivity is especially prominent for genes flanking neurally active HARs (HI = 0.42, p = 4.5×10^{-14}, *t* test; Figure 2B). In comparison, sampling of conserved noncoding and random genomic elements revealed an average HI of 0.33 and 0.31, respectively. Additionally, gene constraint scores (pLI) (Lek et al., 2015) of flanking and target genes are elevated (pLI = 0.42, p < 0.001; and 0.40, p < 0.001; random sampling). Together, HI and pLI scores suggest that HARs might provide a novel source of dosage regulation for highly constrained and haplosensitive genes.

The constraint of genes flanking (i.e., closest upstream and downstream) HARs suggests an important role and possibly an enrichment in specific essential biological functions. Consistent with prior studies, HAR genes were enriched for involvement in neural development, neuronal differentiation, and axonogenesis (Figure 2D and Table S4). Even more, HAR genes are highly enriched for mammalian phenotypes such as abnormal brain morphology (p = 2.3×10^{-27}, Figure 2C). Strikingly, genes near HARs were enriched for associations in the Genetic Association Database of Diseases with ASD (p = 0.03), schizophrenia
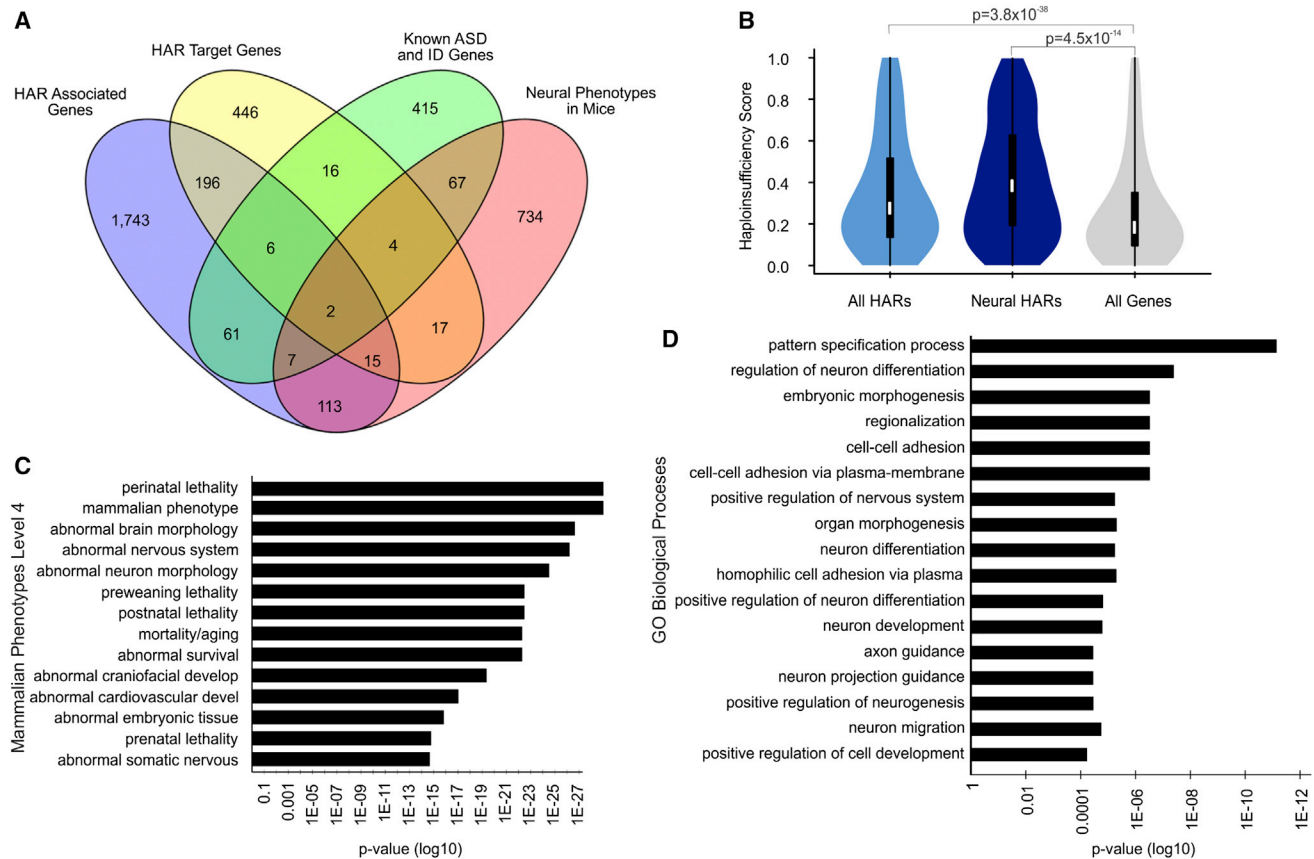
Figure 2. HARs Regulate Dosage-Sensitive Genes Involved in Neural Development

(A) Overlap of HAR-associated and target genes with genes linked to ASD/ID and neural mouse phenotypes.

(B) Comparison of haploinsufficiency scores (Huang et al., 2010) across all genes, HAR-associated genes, and HARs with predicted developmental brain regulatory activity (Capra et al., 2013).

(C) Mammalian phenotypes enriched within HAR genes.

(D) Enriched biological processes and mammalian phenotypes within HAR associated and target genes (Enrichr).

See also Tables S3 and S4.

(p = 0.001), and autonomic nervous system functions (p = 0.01), including *AUTS2*, *CDKL5*, *FOXP1*, *MEF2C*, *NRXN1*, and *SMARCA2*.

HARs in De Novo CNVs Associated with ASD

The functional characteristics of HARs suggest that, when mutated, they have the potential to negatively impact behavioral and cognitive functions. Our analysis of CNVs in 2,100 sibling-matched ASD probands from the Simons Simplex Collection (Sanders et al., 2015) identified rare, de novo CNVs involving HARs to be enriched 6.5-fold—a 2-fold greater excess than reported for all de novo CNVs—with 2.1% (45 out of 2,100) of probands versus 0.3% (7 out of 2,100) of siblings harboring a rare, de novo event (p < 0.0001, Figure 3A). In contrast, no significant difference was observed for inherited HAR-containing CNVs. Gender analyses revealed an excess of de novo CNVs affecting HARs in females (OR 2.48, p = 0.008) but no significant difference for unaffected individuals (p = 0.4). While many reported CNVs are large and impact exons, we identified two intergenic de novo HAR-containing CNVs in probands, but none in healthy sib-

lings. Interestingly, a de novo duplication is located 300 kb upstream of nuclear receptor subfamily 2, group F, member 2 (*NR2F2* or *COUP-TFII*), where existing data reveal an interaction between the *NR2F2* promoter and the duplicated HAR (Figure 3B). CNVs of *NR2F2* occur in individuals with ASD and ID, as well as heart defects (Nava et al., 2014). *NR2F2* is highly expressed in the caudal ganglionic eminence and migrating interneurons (Kanatani et al., 2015) and is downregulated as interneurons enter the cerebral cortex. Therefore, abnormal *NR2F2* expression may disrupt interneuron migration. Our data suggest that de novo CNVs affecting HARs or HAR-containing genes could be implicated in up to 1.8% of ASD cases in simplex families.

Enrichment of Biallelic Point Mutations in Neurally Active HARs

We assessed the impact of biallelic HAR mutations on ASD pathogenesis in 218 families enriched for consanguinity (mostly first-cousin marriages) with one or more children with ASD, as such offspring are enriched for recessive mutations (Morrow et al.,
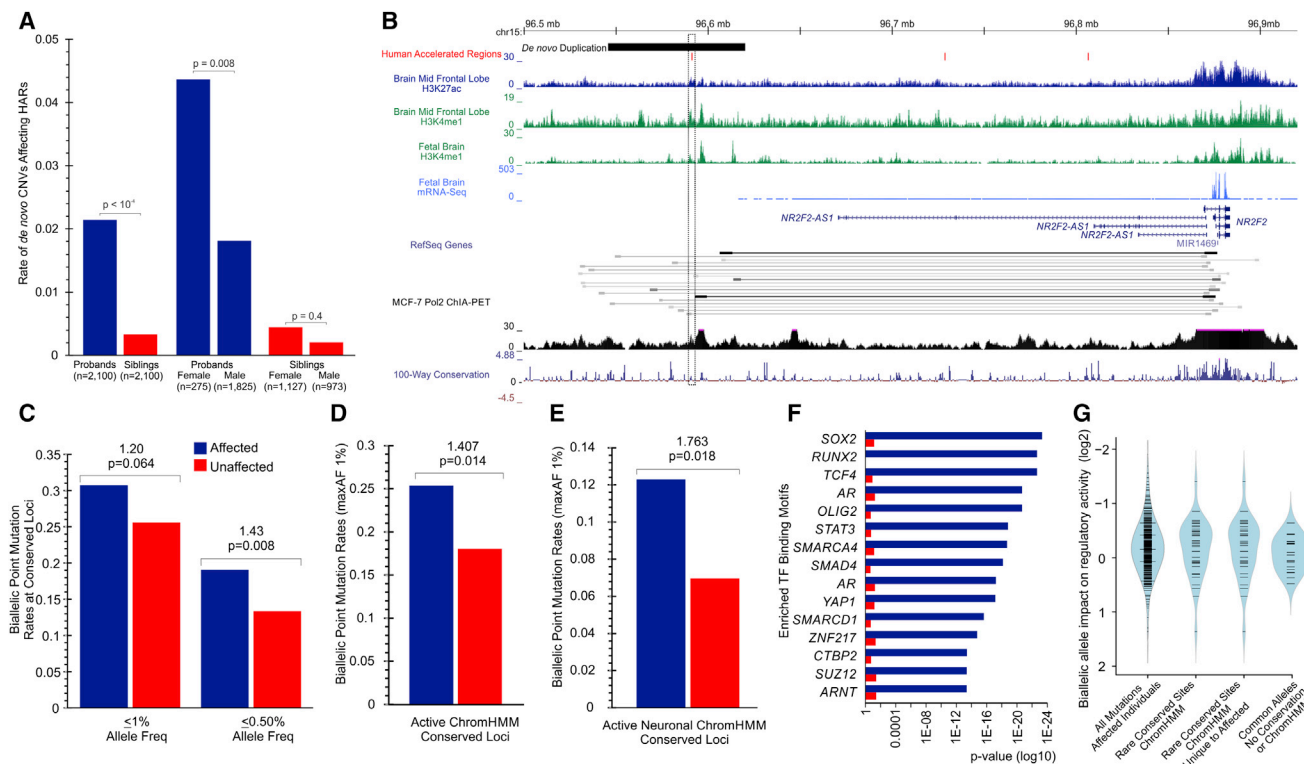
Figure 3. Enrichment of Rare De Novo CNVs and Biallelic Point Mutations in Individuals with ASD

(A) Excess de novo CNVs affecting HARs in affected individuals (blue) compared to normal siblings (red), with greater excess in affected females compared to affected males and no gender differences among healthy siblings.

(B) An intergenic de novo duplication (black bar) ~250 kb upstream of *NR2F2/COUP-TFII*, with existing ChIA-PET data indicating a direct interaction between the HAR and the *NR2F2* promoter.

(C–E) (C) Excess of rare biallelic mutations in probands (blue) versus unaffected (red) members of HMCA cohort. An excess of mutations in probands is also seen in conserved loci within active regulatory elements (maxAF < 1%) (D) and is further increased within active neural regulatory elements (E).

(F) Transcription factor binding enrichment for target and associated genes of rare biallelic mutations (maxAF < 1%) in affected (blue) and unaffected (red) individuals.

(G) Increased impact of the rare biallelic mutations seen in affected individuals at conserved sites using MPRA assay in primary mouse neurospheres.

See also Figure S3 and Tables S5–S7.

2008; Yu et al., 2013). We analyzed the whole-genome sequence (WGS) from 30 affected and 5 unaffected individuals, and we designed a custom "HAR-ome" capture array to sequence HARs in the others (188 affected and 172 unaffected, Table S6), covering 99% of targets with average read depth of 175×. Individuals possessed an average of 1,168 HAR variants (1,027 SNVs and 141 insertion deletions [INDELs]), of which at least 98% were common (MAF > 1%) within the population. The level of homozygosity (44%) was similar to constrained CG69 populations (e.g., Han Chinese, 43%), with a depletion of rare homozygous alleles.

While common HAR variants may contribute to ASD risk, we sought to quantitate an effect for rare biallelic mutations. Considering that biallelic mutation rates can be confounded by background homozygosity despite using healthy family members as controls, we used the bulk HAR mutation rate to correct for effects between affected and unaffected family members. Individuals with ASD exhibited an excess of rare (allele frequency [AF] < 0.5%) biallelic HAR alleles (43% excess versus unaffected, p = 0.008, random sampling; Figures 3C and S3). The mutational

excess in probands persists in HARs with active ChromHMM regulatory marks (Roadmap Epigenomics Consortium et al., 2015) (AF < 1%, ascertainment differential, AD = 0.07, 41% excess versus unaffected, p = 0.014, random sampling; Figure 3D). Strikingly, active neural elements harbored a 1.76-fold excess (AD = 0.05) of rare biallelic HAR mutations (p = 0.018, random sampling, Figure 3E), suggesting that up to 42% may contribute to ASD in ~5% of ASD cases. In comparison, non-neural ChromHMM conserved and non-conserved loci did not exhibit an excess of rare HAR mutations (0.13 versus 0.11, p = 0.246; 0.12 versus 0.11, p = 0.367), suggesting that the signal arising from neurally active HARs is not from unaccounted-for background population structure. Though our data do not distinguish highly penetrant mutations from risk alleles of more modest penetrance, these data expand, for the first time, the role of biallelic mutations in ASD by strongly implicating HARs.

Functional characterization of HARs with rare homozygous alleles revealed a significant enrichment for 13 TFBS motifs, including *HMX1* (p = 9.7×10^{-6}), *BBX* (p = 5.8×10^{-3}), and

Table 1. Identification of Candidate HAR Mutations

Location (hg19)	Reference Allele	Alternate Allele	Gene Location	Interacting Gene[a] (Brain Expression)	Potential Target Genes (Normalized Brain Expression Level)	Family ID	Predicted Regulatory Activity[b]	Phenotype	Diseases and Function
7:101249641	G	A	intergenic	*CUX1* (26.5)	*LINC01007,MYL10* (0.3)	AU-20400 AU-13200	brain enhancer	ASD with ID, (Epilepsy in 1 child)	regulates synaptic spines and dendritic complexity in upper cortical layers (*CUX1*) (Cubelos et al., 2014; Nieto et al., 2004)
5:87776690	T	G	intergenic	*MEF2C* (323.5), *TMEM161B* (2.9)	*LINC00461* (88.0)	AU-9200	brain enhancer	ASD with ID, dysmorphic features (small pointed chin, large ears)	ASD, ID, dysmorphic features (e.g., small chin, prominent forehead, large ears) (Bienvenu et al., 2013; Morrow et al., 2008)
1:97463563	TGGGTAC	TA	intergenic	*PTBP2* (62.5)	*DPYD* (4.0)	AU-4700	brain enhancer	ASD with ID	brain-specific splicing regulator; regulates neurogenesis; locus associated with ID (*PTBP2*)
X:132496780	T	C	intronic	*GPC4* (89.9)		AU-16100	brain enhancer	ASD with ID	regulates synapses (Kalscheuer et al., 2003);
X:132496834	GT	G	intronic	*GPC4* (89.9)		AU-23200	brain enhancer	ASD with ID, Epilepsy, ACC	Simpson-Golabi-Behmel syndrome (ID in 47%) (Hughes-Benzie et al., 1996)
X:18445494	G	GAGCTGTAG	promoter	*CDKL5* (28.7)		AU-13500 AU022203	brain enhancer	ASD with ID	autism, Rett syndrome, epilepsy and Angelman syndrome (Avino and Hutsler, 2010; Hutsler and Casanova, 2015)
4:145793760	C	T	intergenic		*HHIP* (13.8),*ANAPC10* (5.0)	AU-13700	enhancer (stem cell, fetal lung)	ASD with ID	deletion associated with ID (*ANAPC10*)
7:39073195	T	TC	intronic		*POU6F2* (4.4),*VPS41* (27.6)	AU-19000	brain enhancer	ASD	strong association to ASD in GWAS and Wilms tumor (*POU6F2*) (Mizuno et al., 2006)
17:55675858	C	G	intronic	*USP32* (22.3)	*MSI2* (21.1)	AU-24600	brain enhancer	ASD with ID	Regulation of neural precursors during CNS development (*MSI2*) (Gusev et al., 2014)
X:136372832	C	T	intergenic		*GPR101* (11),*ZIC3* (57.8)	AU-21500	repressed Polycomb (brain)	PDD-NOS, ADHD	involved in adrenergic receptor activity in brain (*GPR101*); X-linked visceral heterotaxy and cerebellar dysgenesis (*ZIC3*)

(*Continued on next page*)

Table 1. *Continued*

Location (hg19)	Reference Allele	Alternate Allele	Gene Location	Interacting Gene[a] (Brain Expression)	Potential Target Genes (Normalized Brain Expression Level)	Family ID	Phenotype	Predicted Regulatory Activity[b]	Diseases and Function
4:30653187	T	G	intergenic		MIR4275,PCDH7 (29.7)	AU-5000	ASD	low H3K9ac in derived neural progenitor	possible ASD; high expression in developing brain; regulated by MECP2 during establishment of synaptic connections (PCDH7)
5:39527315	G	A	intergenic	DAB2 (7.9)	PTGER4 (4.3)	AU-26800	ASD with ID	enhancer: HESC, mesendoderm, trophoblast	essential for embryonic development (DAB2)

The location of the HAR mutations with target genes and location relative to refSeq gene annotations, with numbers in parentheses indicating brain expression. Expression profiles were obtained from BrainSpan's normalized LCM microarray values of the developing and adult human brain (BrainSpan, 2011). See also Figure S5. Phenotypes of patients and potential association of the genes to diseases are also provided.

[a]Interacting genes determined from 4C-sequencing, ChIA-Pet (Fullwood et al., 2010; Li et al., 2010), and HiC data (Jin et al., 2013). The interaction distances are limited to 1 Mb. Potential interacting genes represent those containing intronic HARs and, for intergenic HARs, the closest flanking upstream and downstream genes without a clear link to HAR in existing chromatin interaction data.

[b]Predicted regulatory activity was determined using Capra et al. (2013) and ChromHMM predictions from Epigenomics Roadmap.

CDX2 (p = 3.3 × 10^{-5}). Strikingly, genes flanking candidate mutations are highly enriched for interactions with essential TFs involved in cerebral and hippocampal development, such as *SOX2* (p = 4.8 × 10^{-24} versus 0.07 in controls), *OLIG2* (p = 2.1 × 10^{-21} versus 0.2 in controls), and *SMARCA4* (p = 2.0 × 10^{-19} versus 0.07 in controls, Figure 3F). Even more, 70% of genes associated with candidate HAR mutations are expressed in the brain and enriched for ASD (19 of 140, p = 0.02), abnormal brain morphology phenotypes (p = 0.001), and perinatal lethality (p = 0.001), suggesting an impact on cerebral cortical neurogenesis.

We assessed the functional impact of 343 biallelic mutations in affected individuals (85 with AF < 0.01) in primary mouse neurospheres using a custom MPRA assay (Melnikov et al., 2012). A greater portion of rare alleles at conserved loci with predicted regulatory function altered activity than those at nonconserved nucleotides that lacked ChommHMM predictions (35% versus 10%, p = 0.02, Fisher's exact, two-sided, Figure 3G and Table S7). While burden analyses implicate 31% of rare conserved biallelic mutations in neurally active HARs in ASD, remarkably, 29% of such mutations altered regulatory activity by MPRA (19% decreased, 10% increased activity). The enrichment of regulation-altering mutations in HARs with predicted activity suggests that many may contribute to the pathogenesis and diversity of ASD.

We also identified 335 ultra-conserved HARs, completely devoid of mutations within our ASD cohort, which were enriched for neural activity (143/335, p = 0.02) and neurodevelopmental processes (Table S5). Of these HARs, 75 lacked mutation in 1000G samples. We find that TF motifs for neural developmental TFs *POU6F1* (5-fold increase, p = 2.7×10^{-6}) and *POU2F1* (11-fold increase, p = 3.2×10^{-3}) are highly enriched. Even more, 79% (59 of the 75) exhibit regulatory activity in at least one cell type, including 40 in neural tissues (53%). This intriguing subset of HARs could reveal essential functions, potentially relating to more severe brain abnormalities when mutated.

HAR Mutation Affecting *CUX1*

We identified several rare homozygous mutations with active regulatory marks that were proximal to known neurodevelopmental and disease-associated genes (Table 1) in patients who lacked plausibly causative coding mutations. One such HAR mutation is a rare G>A mutation within HAR426 (Prabhakar et al., 2008), for which existing ChIA-Pet data show an interaction with the dosage-sensitive (pLI = 1.0) *CUX1* promoter 200 kb distal (Figures 4A–4C). The mutation was homozygous in three affected individuals from two unrelated consanguineous families of first-cousin parents, one with two sons with ASD and IQs below 40 (AU-20400) and the other family having one daughter with ASD and an IQ below 40 (AU-13200). Linkage analysis revealed a linked region encompassing *CUX1* in both families (AU-20400, LOD = 1.81; AU-13200, LOD = 1.33) but no shared candidate exonic mutations. The HAR mutation occurs only rarely in healthy individuals (1000G, MAF 9.8 × 10^{-4}; middle eastern controls, MAF 0.5%) and only heterozygously and is located within a methylated CpG dinucleotide. Interestingly, the mutation is predicted to create several TF motifs (Figures 4D and 4E). *CUX1* encodes a vertebrate homolog of *Drosophila*
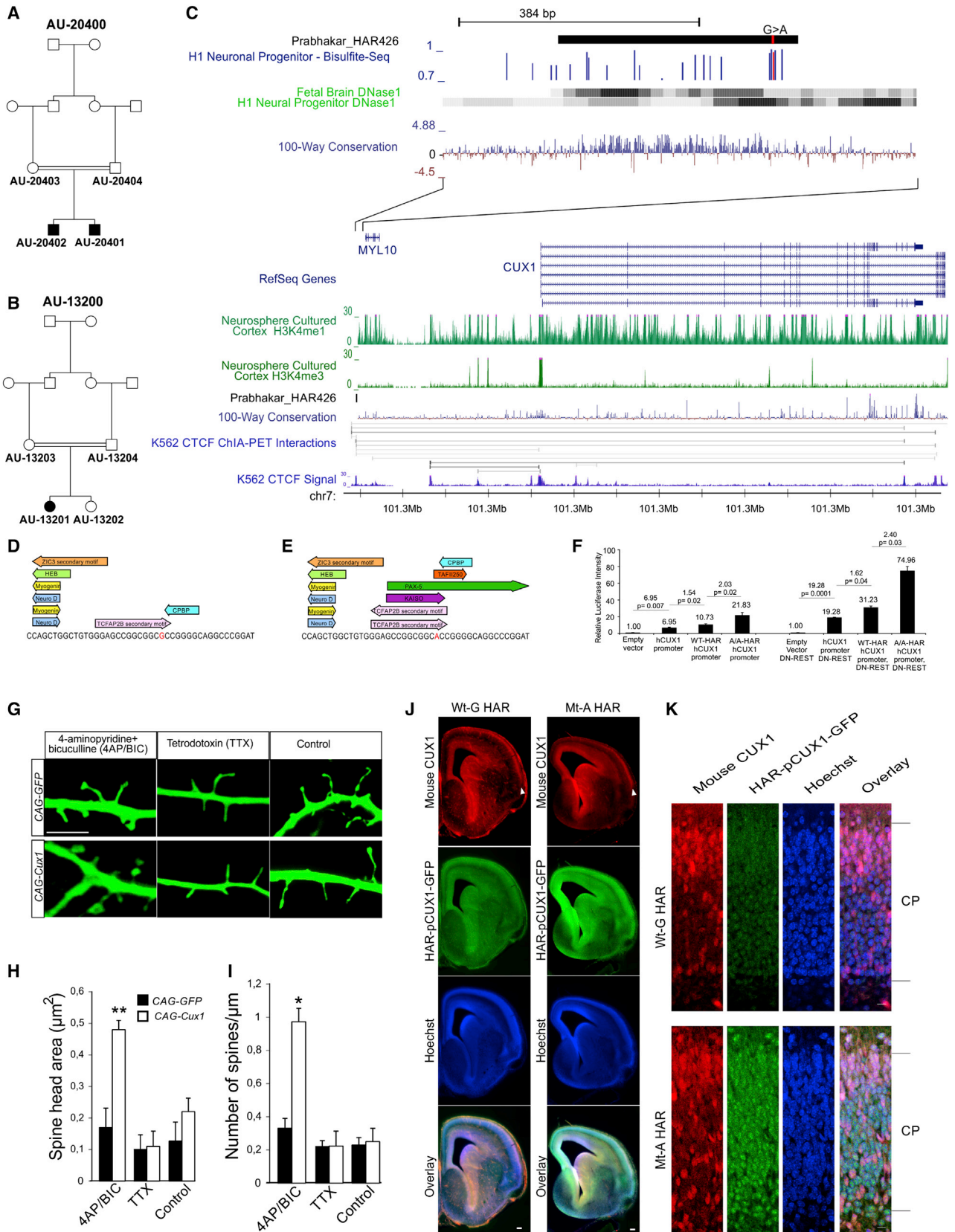
(legend on next page)

Cut, a classical gene with both gain- and loss-of-function effects on neuronal morphology and regulation of synaptic spine density in flies, as well as mammalian cortical neurons (Cubelos et al., 2010, 2014; Grueber et al., 2003).

HAR426 joined to the human *CUX1* promoter shows strong enhancer activity in a luciferase reporter assay, and the G>A mutation boosted this activity >3-fold (Figure 4F) compared to only a 1.5-fold increase from the wild-type allele, suggesting that the HAR426 G>A variant strongly increases neuronal expression of *CUX1*. This increase is among the 95th percentile of enhancer mutations, where the average change due to mutation is 5.5% (95% CI of 1%) (Patwardhan et al., 2012). Overexpression of *Cux1* in cultured differentiating cortical neurons increased synaptic spine density in an activity-dependent manner (Figures 4G–4I), as well as increasing spine head area, indicative of strengthened and more stable synapses, indicating that *Cux1* modifies the intrinsic response of neurons to depolarization, possibly inhibiting dendritic spine pruning or promoting excessive synaptogenesis. The defects in spine morphology induced by *CUX1* expression suggest that aberrant overexpression of *CUX1* caused by the HAR mutation may interfere with normal spine refinement.

The in vivo functional effect of the *CUX1* HAR mutation was modeled in transient transgenic mice (E16.5) expressing a GFP reporter controlled by the HAR (wild-type G [Wt-G] and mutant A [Mt-A] alleles) and the human *CUX1* promoter. Transgenic mice with the Mt-A and *CUX1* promoter showed GFP expression in cortex, consistent with the short *CUX1* HAR isoform in both humans and mice (Nonaka-Kinoshita et al., 2013; Saito et al., 2011), while mice carrying the Mt-A allele showed elevated expression compared to the Wt-G HAR (Figure 4J and 4K). Comparison of the HAR and *CUX1* promoter to predicted sequences of the last common placental ancestor (105MyA) and phylogenetic reconstruction showed surprisingly divergent sequences of the *CUX1* HAR and promoter in rodents compared to other mammals (Figure S4), suggesting rodent-specific evolution (as well as human-specific evolution), whereas *CUX1* exons are highly conserved. These data suggest that allelic and evolutionary changes in the *CUX1*-associated HAR could affect expression levels that, in turn, can regulate spine density or other aspects of neuronal morphology.

HAR Mutation Affecting PTBP2, an Essential Neural Splicing Regulator

In a family of two brothers with ASD and ID, we identified a previously unreported 5-bp INDEL located between the *DPYD* and *PTBP2* genes within an active brain enhancer element (Table 1 and Figure 5A). Our circular chromosome conformation capture sequence [4C-seq] in human SH-SY5Y cells suggests an interaction of the HAR and the promoter of the dosage-sensitive (pLI = 0.99) *PTBP2* gene, encoding a brain-specific splicing regulator essential for neuronal differentiation (Li et al., 2014; Licatalosi et al., 2012). Heterozygous deletions of the locus (*PTBP2* and *DPYD*) are associated with ASD and ID (Carter et al., 2011; Willemsen et al., 2011). TF motif analysis suggests that the INDEL alters TF binding through the gain and loss of motifs (Figures 5B and 5C). Luciferase analysis confirmed HAR enhancer activity with a 50% reduction in activity in the mutated HAR when co-transfected with DN-REST, thus in a neuronal-like state (Figures 5D and 5E). MPRA analysis revealed a similar loss (40% decrease) of enhancer activity caused by the mutation in primary mouse neurospheres (Table S7). No effects of the mutation on enhancer activity were seen in the progenitor-like state, suggesting an effect relatively specific to neurons, where *PTBP2* also shows the most striking phenotypes in mice (Li et al., 2014).

Two Intronic *GPC4* HAR Mutations Affecting Regulatory Activity

Two unrelated families harbored previously unreported homozygous mutations in the same HAR within an intron of *GPC4*, encoding Glypican-4 (Figure 6A). The occurrence of two different rare biallelic mutations in the same HAR is extremely unlikely due to random chance, given the low mutation rate. Our interaction data revealed interactions between the HAR and the promoter of a dosage-sensitive (HI = 0.8) *GPC4* gene in primary adult human brain tissue (Figure 6A). *GPC4* is essential for excitatory synapse development in mice (Allen et al., 2012). Human *GPC3* and *GPC4* are implicated in Simpson-Golabi-Behmel syndrome, which includes ID (Hughes-Benzie et al., 1996; Veugelers et al., 2000). The HAR exhibits activity in neural, muscle, and embryonic stem cells, and the two mutations (Table 1) were predicted to remove TF motifs (Figure 6B). Both mutations reduce regulatory activity by 20%–25% in mouse N2A cells (Figure 6C), with a diminished effect in a neuronal-like state using co-transfection of DN-REST (10%–15% decrease, Figure 6D). Our

---

Figure 4. Autism-Linked HAR Variant Increases Human *CUX1* Promoter Activity in All Cortical Layers

(A–E) Two unrelated consanguineous families, (A) AU-20400 and (B) AU-13200 show a homozygous mutation within long distance regulatory element of *CUX1* within a methylated CpG marked by the presence of H3K4me1 in neuronal cell types (C). ChIA-PET (Fullwood et al., 2010; Li et al., 2010) data demonstrate interaction of HAR426 with promoter of *CUX1*. The mutation alters TF motifs in the reference genome (D) by adding additional motifs (E).

(F) The G>A *CUX1*-interacting mutation results in a 2-fold increase in *CUX1* promoter activity in N2A cells (neural precursor-like condition) and a 2.5-fold increase in the presence of dominant-negative (DN) REST (a neuronal-like condition).

(G–I) Comparison of dendritic filopodia and spines of cortical cultured neurons overexpressing Cux1 (CAG-Cux1) co-transfected with CAG-GFP plasmid to neurons expressing CAG-GFP alone under TTX, 4AP/BIC, and standard (control) conditions. Bar represents 2 μm. Cux1 overexpression with 4AP/BIC treatment results in significant increase in (H) spine head surface area and (I) spine density as compared to CAG-GFP transfected neurons, with no significant differences observed when overexpressing Cux1 in neurons under standard growing conditions (control) and TTX treatment. Student's *t* test; p * ≤ 0.01; ** ≤ 0.0001 compared to CAG-GFP expressing neurons with 4AP/BIC.

(J) Mutant A allele (Mt-A) HAR increases transcriptional activity of human *CUX1* promoter linked to a GFP reporter, compared to wild-type G allele (Wt-G) HAR, in E16.5 transgenic mice.

(K) Both Wt-G and Mt-A mutations drive expression of CUX1 across all layers of the cortical plate, with increased expression due to the mutation.
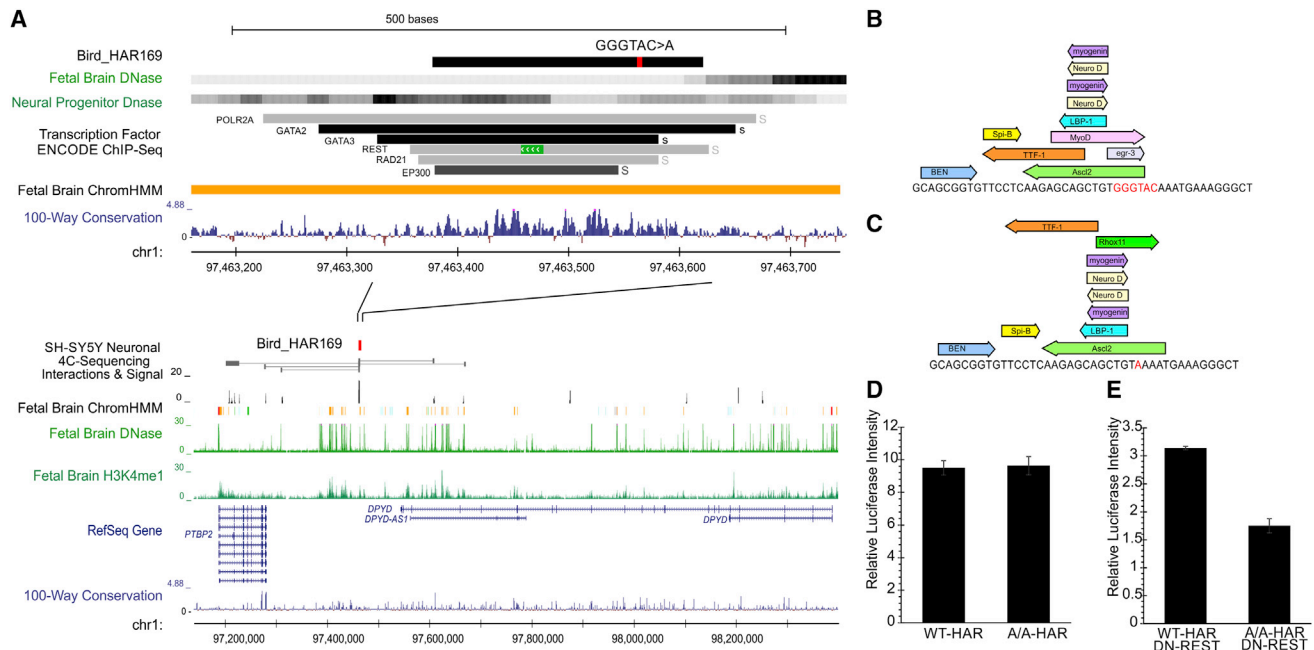
See also Figure S4.

Figure 5. Homozygous INDEL within a HAR that Interacts with Promoter Region *PTBP2*, which Encodes an Essential Splicing Regulator

(A–E) Homozygous GGGTAC>A mutation between *PTBP2* and *DPYD* within a HAR showing enhancer marks in fetal brain (ChromHMM), several TF motifs from TF-chip, and chromatin interactions with the promoter region of *PTBP2* by 4C-seq. The INDEL is located within predicted TF motifs in the reference genome (B), and disrupts some of them (C). Luciferase analysis of the mutant (Mt) and reference (Wt) HAR using a minimal promoter showed equivalent activity in N2A cells in a neural progenitor-like state (D), while co-transfection with DN-REST (E) resulted in a neuronal-like state where the Mt sequence shows a 50% decrease in activity compared to the normal sequence.

MPRA data in primary neural cells from neurospheres also revealed a 30% decrease in activity for the T>C allele (Table S7). Together, these data suggest potential involvement of this HAR in synapse formation.

In addition to these functionally tested mutations, we find several additional properly segregating HAR mutations within affected individuals, many of which exhibit regulatory functions and are located near genes that are important for neurodevelopment (Table 1). Despite tissue specificity of many HARs, we find that several of these candidate HAR mutations affect regulatory potential, including variants within the promoter region of *CDKL5* (Table 1, 25% loss in activity) and 400 kb downstream of *MEF2C* (Table 1 and Figure S5, 50% loss in activity), both of which are known genes underlying ASD and ID. Together, these rare biallelic HAR mutations highlight a previously undiscovered set of noncoding genomic loci, with possible implications on human brain development and the manifestation of ASD.

DISCUSSION

We provide the first direct evidence that rare de novo CNVs involving HARs can contribute to simplex ASD and that rare biallelic mutations in neurally active HARs can confer risk to ASD in as many as 5% of individuals from a consanguineous population. Epigenomic profiling and in vitro analyses showed functional effects of candidate mutations in several HARs that interact with promoters of dosage-sensitive neurodevelopmental genes, including *CUX1*, *PTBP2*, *GPC4*, and *MEF2C*. Abnormal expres-

sion of these genes elicits severe defects in synaptogenesis or other developmental processes, suggesting that HAR mutations in ASD may confer risk through such processes and identifying a set of HARs with potentially essential functions in cognition and behavior.

The striking enrichment of SOX2 interactions among genes associated with HARs with rare biallelic ASD mutations, combined with the overrepresentation of *SOX2* and other developmental dosage-sensitive TF binding motifs (e.g., *OLIG2* and *SMARCA4*) within these HARs, suggests potential roles in neurogenesis. *SOX2* is essential for maintenance of neural progenitors and neural differentiation, and loss of *SOX2* causes hippocampal and cerebral malformations (Kelberman et al., 2008). Therefore, the alteration of TF binding within a HAR could affect the precise timing and cell types involved in neurogenesis or other neurodevelopmental processes.

A fundamental question about HARs is whether their characteristics in humans indicate loss of activity (i.e., assuming a neutral mutation rate) or indicate that they underwent positive selection with evolved functions before switching to negative selection among humans (i.e., functionally constrained in humans). We provide the largest population-level analysis of HARs using WGS data from CG69 and 1000G to reveal evidence of directional selection, which previous studies have suggested to include a combination of positive and negative pressures. If some HARs are still undergoing positive selection, they would thus represent an interesting class of evolving loci that may contribute to differences observed among diverse human populations.
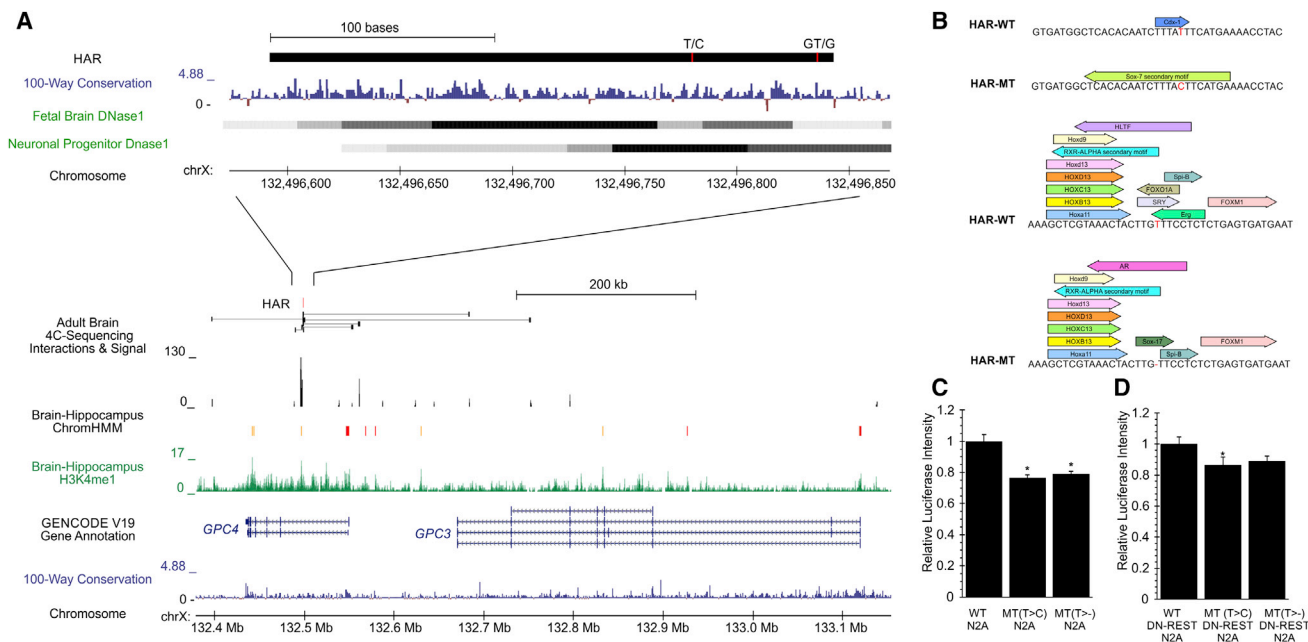
Figure 6. Two Independent Mutations Reduce Regulatory Activity of Intronic HAR within *GPC4*

(A) Homozygous mutations in two unrelated families with ASD affect a HAR within the intron of *GPC4* with enhancer activity in brain tissue (ChromHMM); both proximity and interaction data suggest interaction with the *GPC4* promoter.

(B–D) (B) Predicted TF motifs in the HAR are predicted to be disrupted by both mutations. Luciferase analysis of mutant (Mt) and reference (Wt) HAR sequence using the *GPC4* promoter revealed (C) a 20%–25% decrease in regulatory activity in N2A cells, while co-transfection with DN-REST resulted in a slightly diminished effect (D). *t* test; *p ≤ 0.05 compared to WT.

The conservation of most human-specific HAR alleles within human populations suggests that they modified important functions that might be deleterious if lost. Analyses of human and chimpanzee alleles in TF motifs revealed enrichment for the creation and loss of essential developmental transcription factors such as *CTCF* and *REST*, suggesting that many human-specific nucleotides alter transcriptional regulatory potential, which is further supported by previous transgenic mice (Capra et al., 2013). Therefore, the comparative profiling of HARs for essential TF motifs can be used to identify candidate HARs and alleles regulating human-specific neural developmental processes for social and cognitive functions, as well as synaptic complexity and brain size.

The divergence of humans compared to chimpanzees is apparent in a wide range of phenotypic differences: not only brain size, but also recently acquired or expanded social and cognitive traits. The complexity and spectrum of such traits strongly suggest roles for diverse, recently evolved, noncoding regulatory elements. The overlap of HAR- associated and target genes with diverse neurodevelopmental phenotypes and OMIM disorders (ASD, microcephaly, and seizures) suggests that further mutational and functional characterization of HARs may further dissect those involved in regulation of brain size, structure, and/or other cognitive or social traits. Even more, the intriguing conservation of HARs within human populations— but divergence from chimpanzees—highlights their importance in human-specific neural evolution and diversification.

While our study focused on a cohort specifically selected to be enriched for biallelic mutations, other unconstrained co-horts, as well as other neurodevelopmental disorders, may harbor excesses in biallelic, inherited heterozygous, and de novo risk alleles in HARs, as well as other regulatory elements (Lim et al., 2013; Yu et al., 2013). HAR alleles will likely vary in risk from mild to highly penetrant or pathogenic. Future studies will likely expand our understanding of noncoding mutations in developmental disorders by implicating additional HARs and other conserved regions through targeted sequencing and WGS. Our rapid and cost-effective "HAR-ome" sequence capture strategy can be productively applied to probe the roles of HARs in other conditions and cohorts or in diverse or ancient human populations. Our data help to explain part of the phenotypic variability and missing heritability of autism by implicating rare biallelic mutations within noncoding regulatory HARs.

STAR★METHODS

Detailed methods are provided in the online version of this paper and include the following:

- KEY RESOURCES TABLE
- CONTACT FOR REAGENT AND RESOURCE SHARING
- EXPERIMENTAL MODEL AND SUBJECT DETAILS
  - Human subjects
  - Mice
  - Cell Lines
  - Primary Neuronal Culture

- **METHOD DETAILS**
  - HAR-ome Sequence Capture
  - Massively Parallel Reporter Assay
  - Enhancer Activity Assay
  - Neuronal Spine Analysis
  - 4C-sequencing
  - Functional Validation of Mutation Affecting Regulation of Myocyte Enhancer Factor-2c
- **QUANTIFICATION AND STATISTICAL ANALYSIS**
  - Population Analysis of HARs
  - Functional Predictions of HARs
  - Transcription Factor Binding Motif Analyses
  - HAR-ome Sequence Capture
  - Burden Analysis
  - Functional Prediction Analyses
  - ASD CNV Analysis
  - Evolutionary Analysis of HAR and *CUX1* Promoter
- **DATA AND SOFTWARE AVAILABILITY**

SUPPLEMENTAL INFORMATION

Supplemental Information includes five figures and seven tables and can be found with this article online at http://dx.doi.org/10.1016/j.cell.2016.08.071.

An audio PaperClip is available at http://dx.doi.org/10.1016/j.cell.2016.08.071#mmc6.

CONSORTIA

The contributing members of The Homozygosity Mapping Consortium for Autism are Mahmoud M. Abeidah, Mazhar Adli, Al Noor Centre for Children with Special Needs, Sadika Al-Awadi, Lihadh Al-Gazali, Zeinab I. Alloub, Samira Al-Saad, Muna Al-Saffar, Bulent Ataman, Soher Balkhy, A. James Barkovich, Brenda J. Barry, Laila Bastaki, Margaret Bauman, Tawfeg Ben-Omran, Nancy E. Braverman, Maria H. Chahrour, Bernard S. Chang, Haroon R. Chaudhry, Michael Coulter, Alissa M. D'Gama, Azhar Daoud, Dubai Autism Center, Valsamma Eapen, Jillian M. Felie, Stacey B. Gabriel, Generoso G. Gascon, Micheal E. Greenberg, Ellen Hanson, David A. Harmin, Asif Hashmi, Sabri Herguner, R. Sean Hill, Fuki M. Hisama, Sarn Jiralerspong, Robert M. Joseph, Samir Khalil, Najwa Khuri-Bulos, Omar Kwaja, Benjamin Y. Kwan, Elaine LeClair, Elaine T. Lim, Manzil Centre for Challenged Individuals, Kyriakos Markianos, Madelena Martin, Amira Masri, Brian Meyer, David Miller, Ganeshwaran H. Mochida, Eric M. Morrow, Nahit M. Mukaddes, Ramzi H. Nasir, Zafar Nawaz, Saima Niaz, Kazuko Okamura-Ikeda, Ozgur Oner, Jennifer N. Partlow, Annapurna Poduri, Anna Rajab, Leonard Rappaport, Jacqueline Rodriguez, Klaus Schmitz-Abe, Sharjah Autism Centre, Yiping Shen, Christine R. Stevens, Joan M. Stoler, Christine M. Sunu, Wen-Hann Tan, Hisaaki Taniguchi, Ahmad Teebi, Christopher A. Walsh, Janice Ware, Bai-Lin Wu, Seung-Yun Yoo, and Timothy W. Yu.

AUTHOR CONTRIBUTIONS

R.N.D. and C.A.W. designed the experiments. R.N.D. performed sequencing experiments and analyses. B.-I.B., B.C., C.C., A.A.H., M.N., and R.N.D. performed Sanger and functional validation. S.A.-S., N.M.M., O.O., M.A.-S., S.B., G.G.G., and HMCA provided samples and clinical details for individuals involved in the study. R.N.D. and C.A.W. wrote the manuscript and all authors contributed to the final version of the manuscript.

REFERENCES

Adzhubei, I.A., Schmidt, S., Peshkin, L., Ramensky, V.E., Gerasimova, A., Bork, P., Kondrashov, A.S., and Sunyaev, S.R. (2010). A method and server for predicting damaging missense mutations. Nat. Methods 7, 248–249.

Allen, N.J., Bennett, M.L., Foo, L.C., Wang, G.X., Chakraborty, C., Smith, S.J., and Barres, B.A. (2012). Astrocyte glypicans 4 and 6 promote formation of excitatory synapses via GluA1 AMPA receptors. Nature 486, 410–414.

1000 Genomes Project Consortium, Auton, A., Brooks, L.D., Durbin, R.M., Garrison, E.P., Kang, H.M., Korbel, J.O., Marchini, J.L., McCarthy, S., McVean, G.A., and Abecasis, G.R. (2015). A global reference for human genetic variation. Nature 526, 68–74.

Avino, T.A., and Hutsler, J.J. (2010). Abnormal cell patterning at the cortical gray-white matter boundary in autism spectrum disorders. Brain Res. 1360, 138–146.

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, 764–768.

Bienvenu, T., Diebold, B., Chelly, J., and Isidor, B. (2013). Refining the phenotype associated with MEF2C point mutations. Neurogenetics 14, 71–75.

Bird, C.P., Stranger, B.E., Liu, M., Thomas, D.J., Ingle, C.E., Beazley, C., Miller, W., Hurles, M.E., and Dermitzakis, E.T. (2007). Fast-evolving noncoding sequences in the human genome. Genome Biol. 8, R118.

BrainSpan (2011). BrainSpan: Atlas of the Developing Human Brain. http://www.brainspan.org/.

Burbano, H.A., Green, R.E., Maricic, T., Lalueza-Fox, C., de la Rasilla, M., Rosas, A., Kelso, J., Pollard, K.S., Lachmann, M., and Pääbo, S. (2012). Analysis of human accelerated DNA regions using archaic hominin genomes. PLoS ONE 7, e32877.

Bush, E.C., and Lahn, B.T. (2008). A genome-wide screen for noncoding elements important in primate evolution. BMC Evol. Biol. 8, 17.

Capra, J.A., Erwin, G.D., McKinsey, G., Rubenstein, J.L., and Pollard, K.S. (2013). Many human accelerated regions are developmental enhancers. Philos. Trans. R. Soc. Lond. B Biol. Sci. 368, 20130025.

Carter, M.T., Nikkel, S.M., Fernandez, B.A., Marshall, C.R., Noor, A., Lionel, A.C., Prasad, A., Pinto, D., Joseph-George, A.M., Noakes, C., et al. (2011). Hemizygous deletions on chromosome 1p21.3 involving the DPYD gene in individuals with autism spectrum disorder. Clin. Genet. 80, 435–443.

Chen, E.Y., Tan, C.M., Kou, Y., Duan, Q., Wang, Z., Meirelles, G.V., Clark, N.R., and Ma'ayan, A. (2013). Enrichr: interactive and collaborative HTML5 gene list enrichment analysis tool. BMC Bioinformatics 14, 128.

Chong, J.A., Tapia-Ramírez, J., Kim, S., Toledo-Aral, J.J., Zheng, Y., Boutros, M.C., Altshuller, Y.M., Frohman, M.A., Kraner, S.D., and Mandel, G. (1995). REST: a mammalian silencer protein that restricts sodium channel gene expression to neurons. Cell 80, 949–957.

Cooper, G.M., Stone, E.A., Asimenos, G., Green, E.D., Batzoglou, S., and Sidow, A.; NISC Comparative Sequencing Program (2005). Distribution and intensity of constraint in mammalian genomic sequence. Genome Res. 15, 901–913.

Cubelos, B., Sebastián-Serrano, A., Beccari, L., Calcagnotto, M.E., Cisneros, E., Kim, S., Dopazo, A., Alvarez-Dolado, M., Redondo, J.M., Bovolenta, P., et al. (2010). Cux1 and Cux2 regulate dendritic branching, spine morphology, and synapses of the upper layer neurons of the cortex. Neuron 66, 523–535.

Cubelos, B., Briz, C.G., Esteban-Ortega, G.M., and Nieto, M. (2014). Cux1 and Cux2 selectively target basal and apical dendritic compartments of layer II-III cortical neurons. Dev. Neurobiol. 75, 163–172.

Drmanac, R., Sparks, A.B., Callow, M.J., Halpern, A.L., Burns, N.L., Kermani, B.G., Carnevali, P., Nazarenko, I., Nilsen, G.B., Yeung, G., et al. (2010). Human genome sequencing using unchained base reads on self-assembling DNA nanoarrays. Science 327, 78–81.

Dunham, I., Kundaje, A., Aldred, S.F., Collins, P.J., Davis, C.A., Doyle, F., Epstein, C.B., Frietze, S., Harrow, J., Kaul, R., et al.; ENCODE Project Consortium (2012). An integrated encyclopedia of DNA elements in the human genome. Nature 489, 57–74.

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

Fullwood, M.J., Han, Y., Wei, C.L., Ruan, X., and Ruan, Y. (2010). Chromatin interaction analysis using paired-end tag sequencing. Curr. Protoc. Mol. Biol. Chapter 21, Unit 21 15 21–25.

Gao, F., Wei, Z., Lu, W., and Wang, K. (2013). Comparative analysis of 4C-Seq data generated from enzyme-based and sonication-based methods. BMC Genomics 14, 345.

Gaugler, T., Klei, L., Sanders, S.J., Bodea, C.A., Goldberg, A.P., Lee, A.B., Mahajan, M., Manaa, D., Pawitan, Y., Reichert, J., et al. (2014). Most genetic risk for autism resides with common variation. Nat. Genet. 46, 881–885.

Goh, Y., Fullwood, M.J., Poh, H.M., Peh, S.Q., Ong, C.T., Zhang, J., Ruan, X., and Ruan, Y. (2012). Chromatin Interaction Analysis with Paired-End Tag Sequencing (ChIA-PET) for mapping chromatin interactions and understanding transcription regulation. J. Vis. Exp. Published online April 30, 2012. http://dx.doi.org/10.3791/3770.

Grueber, W.B., Jan, L.Y., and Jan, Y.N. (2003). Different levels of the homeodomain protein cut regulate distinct dendrite branching patterns of Drosophila multidendritic neurons. Cell 112, 805–818.

Gusev, A., Lee, S.H., Trynka, G., Finucane, H., Vilhjalmsson, B.J., Xu, H., Zang, C., Ripke, S., Bulik-Sullivan, B., Stahl, E., et al. (2014). Partitioning heritability of regulatory and cell-type-specific variants across 11 common diseases. Am. J. Hum. Genet. 95, 535–552.

Hiller, M., Schaar, B.T., Indjeian, V.B., Kingsley, D.M., Hagey, L.R., and Bejerano, G. (2012). A "forward genomics" approach links genotype to phenotype using independent phenotypic losses among related species. Cell Rep. 2, 817–823.

Huang, W., Sherman, B.T., and Lempicki, R.A. (2009). Systematic and integrative analysis of large gene lists using DAVID bioinformatics resources. Nat. Protoc. 4, 44–57.

Huang, N., Lee, I., Marcotte, E.M., and Hurles, M.E. (2010). Characterising and predicting haploinsufficiency in the human genome. PLoS Genet. 6, e1001154.

Hughes-Benzie, R.M., Pilia, G., Xuan, J.Y., Hunter, A.G., Chen, E., Golabi, M., Hurst, J.A., Kobori, J., Marymee, K., Pagon, R.A., et al. (1996). Simpson-Golabi-Behmel syndrome: genotype/phenotype analysis of 18 affected males from 7 unrelated families. Am. J. Med. Genet. 66, 227–234.

Hutsler, J.J., and Casanova, M.F. (2015). Cortical construction in autism spectrum disorder: columns, connectivity and the subplate. Neuropathol. Appl. Neurobiol. 42, 115–134.

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, 216–221.

Jin, F., Li, Y., Dixon, J.R., Selvaraj, S., Ye, Z., Lee, A.Y., Yen, C.A., Schmitt, A.D., Espinoza, C.A., and Ren, B. (2013). A high-resolution map of the three-dimensional chromatin interactome in human cells. Nature 503, 290–294.

Kalscheuer, V.M., Tao, J., Donnelly, A., Hollway, G., Schwinger, E., Kubart, S., Menzel, C., Hoeltzenbein, M., Tommerup, N., Eyre, H., et al. (2003). Disruption of the serine/threonine kinase 9 gene causes severe X-linked infantile spasms and mental retardation. Am. J. Hum. Genet. 72, 1401–1411.

Kanatani, S., Honda, T., Aramaki, M., Hayashi, K., Kubo, K., Ishida, M., Tanaka, D.H., Kawauchi, T., Sekine, K., Kusuzawa, S., et al. (2015). The COUP-TFII/Neuropilin-2 is a molecular switch steering diencephalon-derived GABAergic neurons in the developing mouse brain. Proc. Natl. Acad. Sci. USA 112, E4985–E4994.

Kelberman, D., de Castro, S.C., Huang, S., Crolla, J.A., Palmer, R., Gregory, J.W., Taylor, D., Cavallo, L., Faienza, M.F., Fischetto, R., et al. (2008). SOX2 plays a critical role in the pituitary, forebrain, and eye during human embryonic development. J. Clin. Endocrinol. Metab. 93, 1865–1873.

Kent, W.J., Sugnet, C.W., Furey, T.S., Roskin, K.M., Pringle, T.H., Zahler, A.M., and Haussler, D. (2002). The human genome browser at UCSC. Genome Res. 12, 996–1006.

Kircher, M., Witten, D.M., Jain, P., O'Roak, B.J., Cooper, G.M., and Shendure, J. (2014). A general framework for estimating the relative pathogenicity of human genetic variants. Nat. Genet. 46, 310–315.

Roadmap Epigenomics Consortium, 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, 317–330.

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. (2015). Analysis of protein-coding genetic variation in 60,706 humans. Nature 536, 285–291.

Li, H., and Durbin, R. (2009). Fast and accurate short read alignment with Burrows-Wheeler transform. Bioinformatics 25, 1754–1760.

Li, H., Handsaker, B., Wysoker, A., Fennell, T., Ruan, J., Homer, N., Marth, G., Abecasis, G., and Durbin, R.; 1000 Genome Project Data Processing Subgroup (2009). The Sequence Alignment/Map format and SAMtools. Bioinformatics 25, 2078–2079.

Li, G., Fullwood, M.J., Xu, H., Mulawadi, F.H., Velkov, S., Vega, V., Ariyaratne, P.N., Mohamed, Y.B., Ooi, H.S., Tennakoon, C., et al. (2010). ChIA-PET tool for comprehensive chromatin interaction analysis with paired-end tag sequencing. Genome Biol. 11, R22.

Li, G., Ruan, X., Auerbach, R.K., Sandhu, K.S., Zheng, M., Wang, P., Poh, H.M., Goh, Y., Lim, J., Zhang, J., et al. (2012). Extensive promoter-centered chromatin interactions provide a topological basis for transcription regulation. Cell 148, 84–98.

Li, Q., Zheng, S., Han, A., Lin, C.H., Stoilov, P., Fu, X.D., and Black, D.L. (2014). The splicing regulator PTBP2 controls a program of embryonic splicing required for neuronal maturation. eLife 3, e01201.

Licatalosi, D.D., Yano, M., Fak, J.J., Mele, A., Grabinski, S.E., Zhang, C., and Darnell, R.B. (2012). Ptbp2 represses adult-specific splicing to regulate the generation of neuronal precursors in the embryonic brain. Genes Dev. 26, 1626–1642.

Lim, E.T., Raychaudhuri, S., Sanders, S.J., Stevens, C., Sabo, A., MacArthur, D.G., Neale, B.M., Kirby, A., Ruderfer, D.M., Fromer, M., et al.; NHLBI Exome Sequencing Project (2013). Rare complete knockouts in humans: population distribution and significant role in autism spectrum disorders. Neuron 77, 235–242.

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.; Broad Institute Sequencing Platform and Whole Genome Assembly Team; Baylor College of Medicine Human Genome Sequencing Center Sequencing Team; Genome

Institute at Washington University (2011). A high-resolution map of human evolutionary constraint using 29 mammals. Nature *478*, 476–482.

Ma, W., Ay, F., Lee, C., Gulsoy, G., Deng, X., Cook, S., Hesson, J., Cavanaugh, C., Ware, C.B., Krumm, A., et al. (2015). Fine-scale chromatin interaction maps reveal the cis-regulatory landscape of human lincRNA genes. Nat. Methods *12*, 71–78.

McLean, C.Y., Reno, P.L., Pollen, A.A., Bassan, A.I., Capellini, T.D., Guenther, C., Indjeian, V.B., Lim, X., Menke, D.B., Schaar, B.T., et al. (2011). Human-specific loss of regulatory DNA and the evolution of human-specific traits. Nature *471*, 216–219.

Melnikov, A., Murugan, A., Zhang, X., Tesileanu, T., Wang, L., Rogov, P., Feizi, S., Gnirke, A., Callan, C.G., Jr., Kinney, J.B., et al. (2012). Systematic dissection and optimization of inducible enhancers in human cells using a massively parallel reporter assay. Nat. Biotechnol. *30*, 271–277.

Miller, D.J., Duka, T., Stimpson, C.D., Schapiro, S.J., Baze, W.B., McArthur, M.J., Fobbs, A.J., Sousa, A.M., Sestan, N., Wildman, D.E., et al. (2012). Prolonged myelination in human neocortical evolution. Proc. Natl. Acad. Sci. USA *109*, 16480–16485.

Mizuno, A., Villalobos, M.E., Davies, M.M., Dahl, B.C., and Muller, R.A. (2006). Partially enhanced thalamocortical functional connectivity in autism. Brain Res. *1104*, 160–174.

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*, 218–223.

Nava, C., Keren, B., Mignot, C., Rastetter, A., Chantot-Bastaraud, S., Faudet, A., Fonteneau, E., Amiet, C., Laurent, C., Jacquette, A., et al. (2014). Prospective diagnostic analysis of copy number variants using SNP microarrays in individuals with autism spectrum disorders. Eur. J. Hum. Genet. *22*, 71–78.

Ng, P.C., and Henikoff, S. (2003). SIFT: Predicting amino acid changes that affect protein function. Nucleic Acids Res. *31*, 3812–3814.

Nieto, M., Monuki, E.S., Tang, H., Imitola, J., Haubst, N., Khoury, S.J., Cunningham, J., Gotz, M., and Walsh, C.A. (2004). Expression of Cux-1 and Cux-2 in the subventricular zone and upper layers II-IV of the cerebral cortex. J. Comp. Neurol. *479*, 168–180.

Nonaka-Kinoshita, M., Reillo, I., Artegiani, B., Martínez-Martínez, M.A., Nelson, M., Borrell, V., and Calegari, F. (2013). Regulation of cerebral cortex size and folding by expansion of basal progenitors. EMBO J. *32*, 1817–1828.

Novara, F., Rizzo, A., Bedini, G., Girgenti, V., Esposito, S., Pantaleoni, C., Ciccone, R., Sciacca, F.L., Achille, V., Della Mina, E., et al. (2013). MEF2C deletions and mutations versus duplications: a clinical comparison. Eur. J. Med. Genet. *56*, 260–265.

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 Genet. *9*, e1003221.

Patwardhan, R.P., Hiatt, J.B., Witten, D.M., Kim, M.J., Smith, R.P., May, D., Lee, C., Andrie, J.M., Lee, S.I., Cooper, G.M., et al. (2012). Massively parallel functional dissection of mammalian enhancers in vivo. Nat. Biotechnol. *30*, 265–270.

Pennacchio, L.A., Ahituv, N., Moses, A.M., Prabhakar, S., Nobrega, M.A., Shoukry, M., Minovitsky, S., Dubchak, I., Holt, A., Lewis, K.D., et al. (2006). In vivo enhancer analysis of human conserved non-coding sequences. Nature *444*, 499–502.

Pollard, K.S., Salama, S.R., King, B., Kern, A.D., Dreszer, T., Katzman, S., Siepel, A., Pedersen, J.S., Bejerano, G., Baertsch, R., et al. (2006a). Forces shaping the fastest evolving regions in the human genome. PLoS Genet. *2*, e168.

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. (2006b). An RNA gene expressed during cortical development evolved rapidly in humans. Nature *443*, 167–172.

Pollard, K.S., Hubisz, M.J., Rosenbloom, K.R., and Siepel, A. (2010). Detection of nonneutral substitution rates on mammalian phylogenies. Genome Res. *20*, 110–121.

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*, 1346–1350.

Quinlan, A.R., and Hall, I.M. (2010). BEDTools: a flexible suite of utilities for comparing genomic features. Bioinformatics *26*, 841–842.

Rice, P., Longden, I., and Bleasby, A. (2000). EMBOSS: the European Molecular Biology Open Software Suite. Trends Genet. *16*, 276–277.

Rodríguez-Tornos, F.M., San Aniceto, I., Cubelos, B., and Nieto, M. (2013). Enrichment of conserved synaptic activity-responsive element in neuronal genes predicts a coordinated response of MEF2, CREB and SRF. PLoS ONE *8*, e53848.

Saito, T., Hanai, S., Takashima, S., Nakagawa, E., Okazaki, S., Inoue, T., Miyata, R., Hoshino, K., Akashi, T., Sasaki, M., et al. (2011). Neocortical layer formation of human developing brains and lissencephalies: consideration of layer-specific marker expression. Cereb. Cortex *21*, 588–596.

Sanders, S.J., He, X., Willsey, A.J., Ercan-Sencicek, A.G., Samocha, K.E., Cicek, A.E., Murtha, M.T., Bal, V.H., Bishop, S.L., Dong, S., et al.; Autism Sequencing Consortium (2015). Insights into Autism Spectrum Disorder Genomic Architecture and Biology from 71 Risk Loci. Neuron *87*, 1215–1233.

Schwarz, J.M., Rödelsperger, C., Schuelke, M., and Seelow, D. (2010). MutationTaster evaluates disease-causing potential of sequence alterations. Nat. Methods *7*, 575–576.

Shihab, H.A., Gough, J., Cooper, D.N., Stenson, P.D., Barker, G.L., Edwards, K.J., Day, I.N., and Gaunt, T.R. (2013). Predicting the functional, molecular, and phenotypic consequences of amino acid substitutions using hidden Markov models. Hum. Mutat. *34*, 57–65.

Somel, M., Liu, X., and Khaitovich, P. (2013). Human brain evolution: transcripts, metabolites and their regulators. Nat. Rev. Neurosci. *14*, 112–127.

Stamatakis, A. (2014). RAxML version 8: a tool for phylogenetic analysis and post-analysis of large phylogenies. Bioinformatics *30*, 1312–1313.

Untergasser, A., Nijveen, H., Rao, X., Bisseling, T., Geurts, R., and Leunissen, J.A. (2007). Primer3Plus, an enhanced web interface to Primer3. Nucleic Acids Res. Published online May 7, 2007. http://dx.doi.com/10.1093/nar/gkm306.

Veugelers, M., Cat, B.D., Muyldermans, S.Y., Reekmans, G., Delande, N., Frints, S., Legius, E., Fryns, J.P., Schrander-Stumpel, C., Weidle, B., et al. (2000). Mutational analysis of the GPC3/GPC4 glypican gene cluster on Xq26 in patients with Simpson-Golabi-Behmel syndrome: identification of loss-of-function mutations in the GPC3 gene. Hum. Mol. Genet. *9*, 1321–1328.

Wang, K., Li, M., and Hakonarson, H. (2010). ANNOVAR: functional annotation of genetic variants from high-throughput sequencing data. Nucleic Acids Res. *38*, e164.

Weedon, M.N., Cebola, I., Patch, A.M., Flanagan, S.E., De Franco, E., Caswell, R., Rodríguez-Seguí, S.A., Shaw-Smith, C., Cho, C.H., Lango Allen, H., et al.; International Pancreatic Agenesis Consortium (2014). Recessive mutations in a distal PTF1A enhancer cause isolated pancreatic agenesis. Nat. Genet. *46*, 61–64.

Willemsen, M.H., Vallès, A., Kirkels, L.A., Mastebroek, M., Olde Loohuis, N., Kos, A., Wissink-Lindhout, W.M., de Brouwer, A.P., Nillesen, W.M., Pfundt, R., et al. (2011). Chromosome 1p21.3 microdeletions comprising DPYD and MIR137 are associated with intellectual disability. J. Med. Genet. *48*, 810–818.

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. Mol. Biol. Evol. *32*, 1148–1160.

Yamada, T., Yang, Y., Huang, J., Coppola, G., Geschwind, D.H., and Bonni, A. (2013). Sumoylated MEF2A coordinately eliminates orphan presynaptic sites and promotes maturation of presynaptic boutons. J. Neurosci. *33*, 4726–4740.

Yang, Z. (1995). A space-time process model for the evolution of DNA sequences. Genetics *139*, 993–1005.

Yu, T.W., Chahrour, 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*, 259–273.

Zang, C., Schones, D.E., Zeng, C., Cui, K., Zhao, K., and Peng, W. (2009). A clustering approach for identification of enriched domains from histone modification ChIP-Seq data. Bioinformatics *25*, 1952–1958.

STAR★METHODS

KEY RESOURCES TABLE

REAGENT or RESOURCE	SOURCE	IDENTIFIER
Antibodies		
Anti-GFP antibody (chicken IgY)	Abcam	Cat#AB_300798; RRID: AB_300798
Anti-CDP(CUX1) antibody (rabbit IgG)	Santa Cruz	Cat#AB_2261231; RRID: AB_2261231
Rabbit anti-GFP	Life Technologies	Cat#A11122; RRID: AB_10073917
Goat anti-rabbit-Alexa488	Life Technologies	Cat#A11034; RRID: AB_2576217
Hoechst 33342	ThermoFisher Scientific	Cat#62249
Chemicals, Peptides, and Recombinant Proteins		
Lipofectamine 2000 Transfection Reagent	ThermoFisher Scientific	Cat#11668-019
IGF1 Recombinant Human Protein	ThermoFisher Scientific	Cat#PHG0071
Critical Commercial Assays		
Neural Tissue Dissociation Kit-T	Miltenyl Biotech	Cat#130-093-231
Neon Transfection System Kit	ThermoFisher Scientific	Cat#MPK10025
Dual-Luciferase Reporter Assay System	Promega	Cat#E1910
PolyFect Transfection Reagent	QIAGEN	Cat#301105
Agilent Haloplex Custom Capture Kit 501kb-2.5Mb	Agilent	Cat#G9911B
Phusion Polymerase HotStart Flex	NEB	Cat#M535S
Dynabeads mRNA DIRECT Purification Kit	ThermoFisher Scientific	Cat#61012
Deposited Data		
Whole Genome Sequence Data	dgGAP	phs000639
Complete Genomics Diversity Panel	(Drmanac et al., 2010)	ftp://ftp2.completegenomics.com/
1000 Genomes Data	(1000 Genomes Project Consortium et al., 2015)	ftp://ftp.1000genomes.ebi.ac.uk/vol1/ftp/
Epigenomics Roadmap data (including ChromHMM)	(Roadmap Epigenomics Consortium et al., 2015)	http://www.roadmapepigenomics.org/
SIFT	(Ng and Henikoff 2003)	http://annovar.openbioinformatics.org/en/latest/
Polyphen 2	(Adzhubei et al., 2010)	http://annovar.openbioinformatics.org/en/latest/
Mutation Taster	(Schwarz et al., 2010)	http://annovar.openbioinformatics.org/en/latest/
Fathmm	(Shihab et al., 2013)	http://annovar.openbioinformatics.org/en/latest/
CADD	(Kircher et al., 2014)	http://annovar.openbioinformatics.org/en/latest/
ENCODE	(Dunham et al., 2012)	https://www.encodeproject.org
ChIA-PET Data	(Li et al., 2010)	http://genome.ucsc.edu/
HiC Data	(Jin et al., 2013)	Supplementary Data 2 from Jin et al. (2013)
DNase HiC Data	(Ma et al., 2015)	Supplemental Tables 8, 9, 16, and 17 from Ma et al. (2015)
HAR functional predictions	(Capra et al., 2013)	Supplemental Table 5 from Capra et al. (2013)
100-way Conservation	UCSC Genome Browser	http://genome.ucsc.edu/
VISTA enhancers	(Pennacchio et al., 2006)	http://genome.ucsc.edu/
Brain Span LCM and RNA-seq Expression profiles	(BrainSpan, 2011)	http://www.developinghumanbrain.org/
CNVs in the SSC ASD collection	(Sanders et al., 2015)	Tables S2 and S3
ExAC pLI Scores	(Lek et al., 2015)	ftp://ftp.broadinstitute.org/pub/ExAC_release/release0.3/functional_gene_constraint/
Experimental Models: Cell Lines		
Mouse: Neuro2a	ATCC	Cat# CCL-131
Human: SH-SY5Y	ATCC	Cat# CRL 2266

(Continued on next page)

Continued

REAGENT or RESOURCE	SOURCE	IDENTIFIER
Mouse: Primary neuronal culture	This paper	N/A
Experimental Models: Organisms/Strains		
Healthy Adult BA9 brain tissue	Maryland Brain Bank	Cat#UMB1455
Mus musculus/FVB/NJ	The Jackson Laboratory	Cat#001800
Recombinant DNA		
pCAG-CUX1	(Cubelos et al., 2010)	N/A
pMPRA1	Addgene	Plasmid#49349
pMPRADonor2	Addgene	Plasmid#49353
BAC containing human CUX1 gene	Children's Hospital Oakland Research Institute	Cat#RP11-962B9
pCAG-GFP	Gift from C. Cepko (Addgene)	Plasmid#11150
Firefly Luciferase Vectors pGL4.12	Promega	Plasmid#E6671
pGL4.12 human CUX1 (hCUX1) promoter	This paper	N/A
pCDNA1 dominant negative (DN) REST	(Chong et al., 1995)	Gift from G. Mandel
Software and Algorithms		
Agilent SureCall v3	Agilent	http://www.genomics.agilent.com/en/home.jsp
CLC Genomics Workbench v7	QIAGEN	Cat#832000
Bedtools v2.23	(Quinlan and Hall, 2010)	http://bedtools.readthedocs.io/en/latest/
BWA v0.7.8	(Li and Durbin, 2009)	http://bio-bwa.sourceforge.net/
ANNOVAR annotation tool	(Wang et al., 2010)	http://annovar.openbioinformatics.org/en/latest/
FastX toolkit v0.0.13	Hannon Lab	http://hannonlab.cshl.edu/fastx_toolkit/
TransFac	BioBase	http://www.biobase-international.com/product/transcription-factor-binding-sites
Geneious	Biomatters	http://www.geneious.com/
Sicer	(Zang et al., 2009)	http://home.gwu.edu/~wpeng/Software.htm
Samtools	(Li et al., 2009)	http://www.htslib.org/
SIFT	(Ng and Henikoff, 2003)	http://sift.jcvi.org/
Polyphen-2	(Adzhubei et al., 2010)	http://genetics.bwh.harvard.edu/pph2/
Mutation Taster	(Schwarz et al., 2010)	http://www.mutationtaster.org/
Fathmm	(Shihab et al., 2013)	http://fathmm.biocompute.org.uk/
CADD	(Kircher et al., 2014)	http://cadd.gs.washington.edu/
Enrichr	(Chen et al., 2013)	http://amp.pharm.mssm.edu/Enrichr/
DAVID functional annotation tool	(Huang et al., 2009)	https://david.ncifcrf.gov/
UCSC Genome Browser	(Kent et al., 2002)	http://genome.ucsc.edu/
EMBOSS tools	(Rice et al., 2000)	http://www.ebi.ac.uk/Tools/emboss/
RAxML	(Stamatakis 2014)	http://sco.h-its.org/exelixis/web/software/raxml/
Primer3Plus	(Untergasser et al., 2007)	http://www.bioinformatics.nl/cgi-bin/primer3plus/primer3plus.cgi/
PHAST tools	(Yang, 1995)	http://compgen.cshl.edu/phast/

CONTACT FOR REAGENT AND RESOURCE SHARING

Please contact C.A.W. (christopher.walsh@childrens.harvard.edu) for reagents and resources generated in this study.

EXPERIMENTAL MODEL AND SUBJECT DETAILS

Human subjects
Research on human samples was conducted following written informed consent and with approval of the Committees on Clinical Investigation at Boston Children's Hospital and Beth Israel Deaconess Medical Center, and participating local institutions.

We selected simplex and multiplex families of Middle Eastern descent with known consanguinity (predominately 1st-cousin parents) with a diagnosis of ASD where WES did not reveal clear causative mutations (Table S6). Additionally, 15 families from the AGRE collection were included for WGS and targeted sequencing. A single affected and unaffected family member was chosen from each family in order to mitigate the effects of population stratification in the HAR burden analyses (Table S5). Beyond ASD, individuals also presented with ID and/or other comorbidities. Genomic DNA was isolated from blood samples.

Mice

Transient transgenic mice were generated using plasmids containing *CUX1*-interacting HAR426 (Wt-G or Mt-A), a 2.2-kb human *CUX1* promoter and GFP. The 2.2-kb human *CUX1* promoter, chr7:101,457,129-101,459,367 (hg19), was retrieved from bacterial artificial chromosome RP11-962B9 (Children's Hospital Oakland Research Institute) by recombineering. The plasmids were linearized, and pronuclearly injected into the pronucleus of one-cell FVB/N mouse embryos (The Jackson Laboratory, n = 50-80 embryos/plasmid) at the Brigham and Women's Hospital Transgenic Core. Wt-G HAR-hCUX1-GFP mice (n = 7) and Mt-A HAR-hCUX1-GFP mice (n = 15) were analyzed at E16.5. Brains were perfused and fixed in 4% paraformaldehyde/PBS, vibratome-sectioned at 75 μm, and stained for GFP and endogenous Cux1 expression using chicken anti-GFP antibody (Abcam) and rabbit anti-CDP/CUX1 antibody (Santa Cruz). All animal experiments conformed to the guidelines approved by the Children's Hospital Animal Care and Use Committee.

Cell Lines

N2A cells (ATCC) were grown in 10% fetal bovine serum, Dulbecco's Modified Eagle Medium, and 1X Penicillin-Streptomycin. SH-SY5Y cell (ATCC) medium was 10% fetal bovine serum, Dulbecco's Modified Eagle Medium/F12, and 1X Penicillin-Streptomycin. Both cell lines were maintained in a 5% CO_2 incubator at 37°C.

Primary Neuronal Culture

For neuronal culture preparations E18 embryo cortex were trypsinized (0.25 μg/ml trypsin) (SIGMA-Aldrich) in EBSS (GIBCO, Invitrogen, Carlsbad, CA) 3.8% MgSO4 (Sigma-Aldrich, St. Louis, MO), and penicillin/streptomycin (GIBCO, Invitrogen, Carlsbad, CA). The reaction was stopped and cells were mechanically dissociated in EBSS media complemented with 0.26 mg/ml Trypsin inhibitor, 0.08 mg/ml DNase, and 3.8% MgSO4 heptahydrate (all from Sigma-Aldrich, St. Louis, MO). Dissociated cells were seeded onto 24 well Poly-D-Lys (Sigma-Aldrich, St. Louis, MO) coated plates in neurobasal media supplemented with B27 complement 1x, glutamax 1x and penicillin/streptomycin (GIBCO, Invitrogen, Carlsbad, CA). 500 μl media were replaced every 2 days.

METHOD DETAILS

HAR-ome Sequence Capture

Our sequence capture panel utilizes the Agilent Haloplex technology to capture 99% of targeted HARs (808.7kb of 815.6kb, 2,733 of 2737 HARs). Deep sequencing of all samples was performed using the Illumina Hi-Seq 2500 with 150-bp paired-end reads with pooled 96-indexed samples per lane. All raw HAR-ome sequences were processed through our existing CLC genomics pipeline. Briefly, all FASTQ files were trimmed to remove low quality reads and bases. All sequences were then aligned with three rounds of local realignments. Variants were identified using the fixed ploidy variant caller incorporating base quality scores (20), read depth (10X), and uniquely mapped reads. In addition, VCF formatted genetic variants from WGS of ASD families, along with the Complete Genomics Diversity panel and 1000 Genomes project, were converted into an Annovar input format using the convert2annovar command.

Massively Parallel Reporter Assay

Rare biallelic HAR mutations in affected and unaffected individuals were used to design both wild-type and mutant MPRA probes. Each mutation was tiled using three unique 115bp ssDNA oligonucleotides overlapping by 90bp. Oligos were created as a single ssDNA pool by CustomArray (Bothell, WA). The MPRA assay was conducted as previously described (Melnikov et al., 2012). Briefly, the custom ssDNA oligo pool was amplified by emulsion PCR and cloned into pMPRA1 vector (Addgene). The luciferase gene and minimal promoter (pMPRA_donor2, Addgene, 49353) was cloned into the pMPRA1_Oligo construct pool. Neurospheres were created using the Miltenyi Biotec Neural Tissue Dissociation Kit. Briefly, the cortex from E14.5 embryonic mice was minced and dissociated with using the Miltenyi Biotec Neural Tissue Dissociation Kit. The cells were grown in suspended culture for 24 hr. Then, 2ug of MPRA constructs were transfected into batches of 5 million harvested cells resuspended in Neon resuspension buffer using the Neon System with electroporation settings: 1,400 V, 40 ms, 1 pulse. The cells were harvested 48 hr post-transfection and mRNA was isolated using the Dynabeads mRNA DIRECT Purification Kit. Resulting Tag-seq amplified cDNA was sequenced using 75bp sequencing with Illumina's MiSeq technology. Reads were filtered for quality and perfect match to oligo barcodes.

Enhancer Activity Assay

We selected HAR mutations for functional testing based on proximity to important neurodevelopmental genes, established regulatory activity in Epigenomics Roadmap data, and existing chromatin interaction data. HARs were cloned from control and patient

lymphocyte-derived DNA by PCR amplification. HARs and either a minimal *Hsp68* or the target promoter were subcloned into lucif-erase vector pGL4.12 (Promega). The human *CUX1* promoter used in the enhancer activity assay is identical to the one used for the generation of transgenic mice. Luciferase plasmids were transfected into N2A cells, along with an internal control plasmid (phRL-TK(Int-), Promega) and dominant-negative REST (DN-REST) (gift from Dr. Gail Mandel) or GFP expression plasmids using Polyfect (QIAGEN)(Chong et al., 1995). Luciferase activities were measured 72 hr later.

Neuronal Spine Analysis

For spine analysis, primary neurons were transfected with CAG-GFP or CAG-Cux1; CAG-GFP using lipofectamine 2000 as previ-ously described (Rodríguez-Tornos et al., 2013). Next, 12hr prior to inducing neuronal activity, cells were incubated with 2 mM tetro-dotoxin (TTX) (Alomone-labs, Jerusalem, Israel). Then media was replaced with media containing 4-aminopyridine (4AP) 100 μM, strychnine 1 μM, glycine 100 μM and bicuculline (BIC) 30 μM (Sigma-Aldrich, St. Louis, MO). After stimulation, cells were fixed and stained with mouse anti-GFP (A11122) and goat anti-rabbit-Alexa488 (Life Technologies) as described (Cubelos et al., 2010). Confocal microscopy was performed with a TCS-SP5 (Leica) Laser Scanning System on a Zeiss Axiovert 200 microscope. Dendritic spine and filopodia processes of individual neurons were measured as previously described (Cubelos et al., 2010). Confocal micro-scopy was performed with a TCS-SP5 (Leica) Laser Scanning System on a Zeiss Axiovert 200 microscope and 50 μm sections were analyzed by taking 0.2 μm serial optical sections with Lasaf v1.8 software (Leica). Images were acquired using a 1024x1024 scan format with a 63x objective and analyzed using Fiji.

4C-sequencing

We used a modified version of a previously reported 4C-sequencing method to determine chromatin interactions between genomic loci and selected HARs. Briefly, we cultured human neuroblastoma SH-SY5Y cells with and without the addition of IGF-1 for 72 hr to differentiate the cells into mature neurons. Approximately 30M cells from each condition were collected for 4C-sequencing. In addi-tion, fresh-frozen bulk adult human cortical brain tissue was homogenized on ice in 1X PBS containing protease inhibitors. The homogenized tissue was processed as previously reported including brief fixation in 1% formaldehyde, nuclear extraction and chro-matin sonication using a Covaris sonicator (Gao et al., 2013). Indexed paired-end libraries were created following PCR enrichment of selected HARs using Phusion polymerase (NEB), followed by sequencing using Illumina MiSeq. Reads were mapped using BWA (Li and Durbin, 2009), allowing for chimeric reads representing the fusion of distal interaction regions created through the ligation step of 4C. All data were visually inspected for mapped reads spanning distinct loci within 1Mb. Mapped BAM files were converted to bed-graphs using BEDtools (Quinlan and Hall, 2010) and significant peaks were identified using SICER (Zang et al., 2009). Next, reads mapping within or spanning peaks located within 1Mb of the targeted HAR were extracted from mapped BAM files using Samtools (Li et al., 2009). We defined the known interacting genes for several HARs containing candidate mutations as either having a previously reported physical interactions or by the presence of mapped paired-reads from 4C-sequencing.

Functional Validation of Mutation Affecting Regulation of Myocyte Enhancer Factor-2c

Another candidate HAR mutation is a homozygous T>G SNV in a developmental brain enhancer (family AU-9200) (Table 1 and Fig-ure S5A). The family comprises healthy, first cousin parents with three healthy sons and two sons with ASD, ID, and mild dysmor-phisms including small pointed chin and large ears, a phenotype commonly observed with *MEF2C* mutations (Bienvenu et al., 2013; Novara et al., 2013). The point mutation is present in low frequencies in the 1000G (0.7%), though more common in South Asia, and absent in healthy Middle Eastern controls. One person of South Asian decent is reported to be homozygous for the mutation in 1000G, though no clinical data are available. This single homozygous allele might have arisen as an artifact of the low coverage sequencing in 1000G. Analysis of the phased 1000G data suggest that the frequency of rare alleles is much lower than expected based on the CG69 WGS, our targeted sequencing, and public datasets. This depletion of rare alleles persists across all genomic categories (*i.e.*, coding, conserved noncoding, random regions, and HARs), suggesting at least two possible contributing factors arising from the low (< 1X) read coverage in the 1000G data. First, 1000G was designed to use low coverage sequencing in a large population in order to identify common alleles in diverse populations. Such coverage will inherently limit the identification of rare al-leles. Second, the imputation of missing genotypes in populations, especially smaller populations (e.g., South Asian) could lead to the artificial inflation of some allele frequencies, particularly in rare or recurrent mutations. Therefore, while the HAR mutation is estimated to be at a higher frequency in the South Asian population, the exact frequency would require higher-depth sequencing of a larger population. Regardless of the population frequency limitations, this HAR mutation is likely to represent a phenotypic modifier that may act with other coding or noncoding mutations to account for the phenotypic spectrum in the family in a similar way to that proposed by common coding mutations (Gaugler et al., 2014).

We assessed the potential impact of rare coding mutations in this family using exome sequencing data of both affected individuals. Our analysis of rare biallelic coding mutations including stop-gain, INDELs (frameshift and non-frameshift), damaging missense (pre-dicted pathogenic by SIFT, POLYPHEN, MUTATION TASTER, and FATHMM)(Adzhubei et al., 2010; Ng and Henikoff, 2003; Schwarz et al., 2010; Shihab et al., 2013) revealed only a single missense mutation in *ARHGAP36* (p.Ile432Val) which was predicted to be damaging by 2 of the 4 programs. *ARHGAP36* is a dosage sensitive gene lacking association with human disease. The combination of the gene's likely dominant acting mechanism (pLI = 0.99), presence of other homozygous mutations in the flanking amino acid (p.His431Arg) in healthy individuals, and inconsistent damaging predictions suggest that the identified mutation is likely benign.

The lack of plausible coding mutations raises the possibility of a role for the HAR mutation in at least a portion of the individuals' phenotypes, especially those matching typical *MEF2C* deletions, but we document this mutation here because we regard its pathogenicity as less definitive than others.

Given the potential functional role of this mutation, we sought to identify the target gene of the HAR and assess the mutation's impact on gene regulation. Using our own 4C-seq of human SH-SY5Y neuroblastoma cells in a neuronal state, we identified interaction with the promoter regions of *TMEM161B* and *MEF2C* (HI = 0.97), an interaction that is not observed in existing ChIA-PET data from cancer cell lines suggesting a neural-specific interaction (Figure S5B). The mutation is predicted to create a myocyte enhancer factor-2a (*MEF2A*) motif; which is highly expressed in differentiated neurons during synaptogenesis and controls postsynaptic dendritic differentiation as well as presynaptic differentiation in the brain (Yamada et al., 2013) (Figure S5C and S5D). Analysis of the regulatory activity of the reference and mutated HAR sequence using the predicted promoter region of *MEF2C* in N2A cells revealed a 30% decrease in activity in neural progenitor like cells (Figure S5E). Since the mutation is predicted to create a motif for *MEF2A*, we would expect overexpression of *MEF2A* to amplify the mutation's effect, which it did, resulting in a 53% decrease in activity (Figure S5E). Next, given the complexity of the promoter region of *MEF2C* (Figure S5B), we selected a second locus supported by epigenetic data as being a promoter and assessed its activity under the same conditions. While the second promoter had less activity than the first, we find that the mutation decreased the HARs activity by 16% in progenitor-like cells and 30% in the presence of *MEF2A* (Figure S5E). The 30%–50% loss of *MEF2C* due to the HAR mutation would likely affect tissues where the HAR is active and be similar to the heterozygous loss of *MEF2C* due to coding mutations. The phenotypic similarity between these patients and others with heterozygous *MEF2C* deletions—who often show severe ID, seizures, verbal and motor developmental delay, large ears, small chin, prominent forehead, and either hypo or hypertonia (Bienvenu et al., 2013; Novara et al., 2013)—suggests a possible role for this SNV in the individual's phenotype.

QUANTIFICATION AND STATISTICAL ANALYSIS

Population Analysis of HARs

HARs were obtained from recent publications and converted into the human (hg19) genome assembly using the UCSC genome browser liftover tool. All HARs located within 30 base-pairs of another HAR were merged into a single region prior to annotating with Annovar (Wang et al., 2010) using databases obtained from UCSC Genome Browser (Kent et al., 2002) and Annovar. All mutations overlapping HARs, conserved noncoding elements, refSeq coding exons, or randomly selected genomic loci of equal number and size to HARs were extracted from the 1000 Genomes and CG69 VCF files (Drmanac et al., 2010; 1000 Genomes Project Consortium et al., 2015). The rates of mutations were assessed for each population and subpopulation and across the entire collection of individuals for each genomic category (HARs, conserved noncoding elements, refSeq coding exons, and the average of 1000 randomly selected genomic loci equivalent to the size distribution of HARs). The total mutational rates, as well as the fraction of rare (AF < 1%) mutations were compared between the HARs and the other categories using the 2-tailed p value from a T-Test. In addition, groups were compared using the 95% confidence intervals. All statistical parameters and values are located in the text and figure legends.

Functional Predictions of HARs

Human and fetal brain H3K4me1 bed files, ChromHMM, and DNase hypersensitivity data were obtained and analyzed from the Epigenomics Roadmap server (Roadmap Epigenomics Consortium et al., 2015). The significance of enrichments was computed using 10,000 samplings of random and conserved genomic regions. The DAVID functional annotation tool (Huang et al., 2009) was used to test genes for enrichment in the Genetic Association Database of Diseases, while Enrichr was used to determine enrichment of biological processes, KEGG pathways, mammalian phenotypes, and human phenotypes (Chen et al., 2013; Huang et al., 2009). Results of statistical tests are located in the figure legends and results section.

Using existing ChIA-Pet, we overlaid all 2,737 HARs with existing datasets, revealing HAR-promoter interactions among 180 HARs and 207 genes. Additionally, despite the limited resolution of HiC compared to ChIA-Pet, 349 HARs were mapped to their target promoters in IMR90 human fibroblasts.

While interactions between HARs and promoters likely indicate regulatory activity, they could also result from random contacts in the nuclei. Therefore, we associated predicted regulatory activity of these loci in the ENCODE and Epigenomics Roadmap data revealing that 93% of HARs with identified chromatin interactions exhibit transcriptional or regulatory activity in at least one cell type, including 66% in ENCODE or IMR90 cell lines. Interestingly, these HARs were more frequently active in embryonic stem cells (75% overlap), followed by ENCODE and epithelial samples. Even more, we found several HARs that interacted with different genes depending on the cell line assayed, suggesting that some fraction might regulate different genes based on their developmental time point. The tissue specific activity suggests that primary tissues will be required to fully understand the promoter interaction among most HARs.

Transcription Factor Binding Motif Analyses

Transcription factor binding motifs for all HARs using the human and chimpanzee sequences were performed using the FMatch tool with TransFac (BioBase) with the settings: vertebrate_non_redundant_miFP, database version 2015.3, significance p < 0.01. Enrichment of motifs was determined using options: randomly generated sequences and 1000bp shift of input regions.

HAR-ome Sequence Capture

The genetic variants from all individuals were annotated using Annovar against public variant databases, regulatory databases (i.e., ENCODE, Epigenomics Roadmap, Ensembl HAR enhancer predictions (Capra et al., 2013), and VISTA enhancers (Pennacchio et al., 2006)), CADD(Kircher et al., 2014), 100-way Vertebrate sequence conservation, GERP, gene databases (Refseq, Ensembl, and UCSC Known genes), and existing chromatin interaction maps (ChIA-PET and HiC) (Fullwood et al., 2010; Jin et al., 2013; Li et al., 2010). Simulations on 10,000 sets of randomly selected genomic and conserved regions within each individual whole genome sequence were analyzed and compared to HARs. The level, zygosity, and burden of HAR variation within our Middle Eastern cohort were compared to CG69. Furthermore, the identification of candidate mutations for ASD was performed by filtering all variants, removing all common variants (> 1% allele frequency), heterozygous variants, and those that were homozygous in the control samples, as well as those common within our cohort (AF > 2%).

Candidate mutations were screened for allelic segregation within each family. Additionally, mutations were screened in 101 healthy individuals from the United Arab Emirates. PCR primers were designed to span candidate mutations using Primer3Plus (Untergasser et al., 2007). All PCRs were conducted using a standard touchdown protocol. PCR products were purified and sequenced using Beckman Genomics' Sanger sequencing service. Trace data from all family members and 101 controls were trimmed and multi-aligned using default settings in Geneious.

Burden Analysis

The rates of high-quality homozygous variants were compared between affected and unaffected individuals at 5%, 1%, and 0.5% maximum allele frequencies in public databases as well as within our cohort. Additionally, all variants within 10bp of another variant were excluded in order to identify highly conserved loci. Furthermore, alleles with low quality calls at greater than 10% of the covering reads were excluded as these often represented poorly mapped and difficult regions for sequencing. The ratio of mutation rates between affected and unaffected individuals for each allele frequency was used to normalize the affected rates in affected individuals, thereby reducing our final burden to more accurately reflect an elevation due to ASD instead of population structure (Figure S2). All variants were required to have either GERP scores > 2 or CADD scores > 0. We overlaid ChromHMM predictions from the Epigenomics Roadmap project to identify alleles within loci with active enhancer, promoter, or transcriptional properties. We further screened these data to create a list of neuronal loci which included all brain tissues as well as neurosphere derived cells. Significance of observed excesses were tested using 10,000 random assignments of affected status.

Estimation of contribution of HAR mutations to ASD was determined using the population corrected mutational rates. First, the ascertainment differential was calculated as the difference between the corrected rates in affected individuals and the rate in unaffected individuals (Iossifov et al., 2014). This value represents the rate of alleles contributing to ASD and, when divided by the total rate in affected individuals, can be used to estimate the proportion of identified alleles that contribute to the diagnoses.

Functional Prediction Analyses

In order to investigate enriched functional categories in HARs, as well as in our mutation datasets, we analyzed their target and associated genes. Associated genes include those where HARs are within the introns, within or near (less than 1kb) 5′ and 3′ UTRs, or are the closest flanking gene, as annotated by Annovar. All closest flanking genes (both upstream and downstream) were less than 2.1Mb away, with 70% being less than 500kb away. All gene sets were analyzed using Enrichr (Chen et al., 2013) and DAVID functional annotation tools. Using our highly annotated datasets, candidate variants in promoter, intronic, and intergenic regions were identified after considering overlap with base conservation, histone data, TFBS, chromatin interactions, and predicted enhancers. The predicted functional consequences of each variant were determined first by overlapping DNA and histone modifications with potential gene targets. Next, all candidates were analyzed using the vertebrate database and the default settings including filters to minimize false positives in TRANSFAC (Biobase) in order to determine the variants' impact on TFBS. Finally, fetal and adult human brain expression profiles of potential target genes of HARs were determined using existing microarray analyses of LCM-acquired tissues available from BrainSpan. All normalized microarray expression values represent the maximum observed across all ages and tissues.

ASD CNV Analysis

Rare de novo CNVs in ASD cases and sibling-matched controls were obtained from recently published de novo analyses of the SSC ASD (Sanders et al., 2015). All CNVs were annotated with gene annotations and HARs using Annovar (Wang et al., 2010). Statistical comparisons of cases to the controls were conducted using Fisher's Exact test.

Evolutionary Analysis of HAR and *CUX1* Promoter

Comparative evolutionary analysis of HAR426 and *CUX1* was performed using a modified version of the recently published "Forward Genomics" approach (Hiller et al., 2012). A multi-fasta file was extracted from the existing 100-way vertebrate multiple alignments from the UCSC genome browser (Kent et al., 2002; Pollard et al., 2010). The sequence of the last common placental ancestor was predicted from the phylip formatted alignment using the prequel algorithm (–keep-gaps–no-probs–msa-format PHYLIP), part of the PHAST tools (Yang, 1995). The percent identities were calculated through pairwise alignments of all species against the predicted common ancestral sequence using Needleall, part of the EMBOSS tools (Rice et al., 2000). Species with low-quality or assembly gaps were excluded from the analysis. To further validate these findings, phylogenetic reconstruction was performed using the

alignment files for the loci using RAxML(Stamatakis, 2014). Using default settings, 1,000 fast bootstraps and the GRT+gamma model, a maximum likelihood phylogram was created for the HAR and *CUX1* promoter. These phylograms were compared to the existing phylogram available from the 100-way vertebrate alignment, obtained from UCSC genome browser.

DATA AND SOFTWARE AVAILABILITY

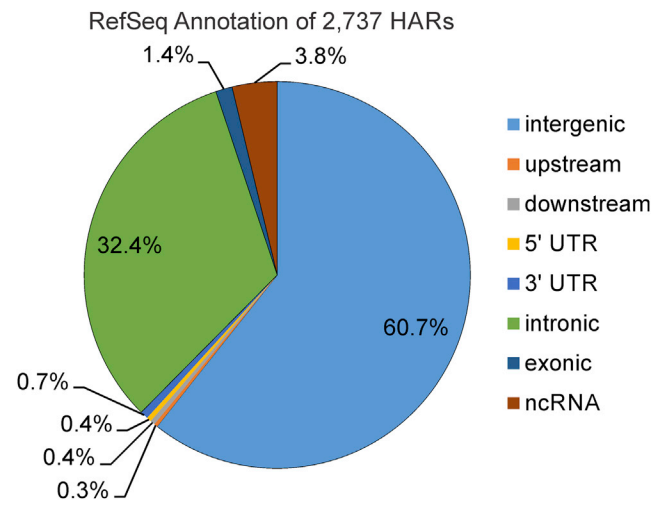The accession number for the whole genome sequence data reported in this paper is dbGAP: phs000639.

Supplemental Figures

Cell

Figure S1. Annotation of 2,737 HARs Using the refGene Database, Related to Figure 1
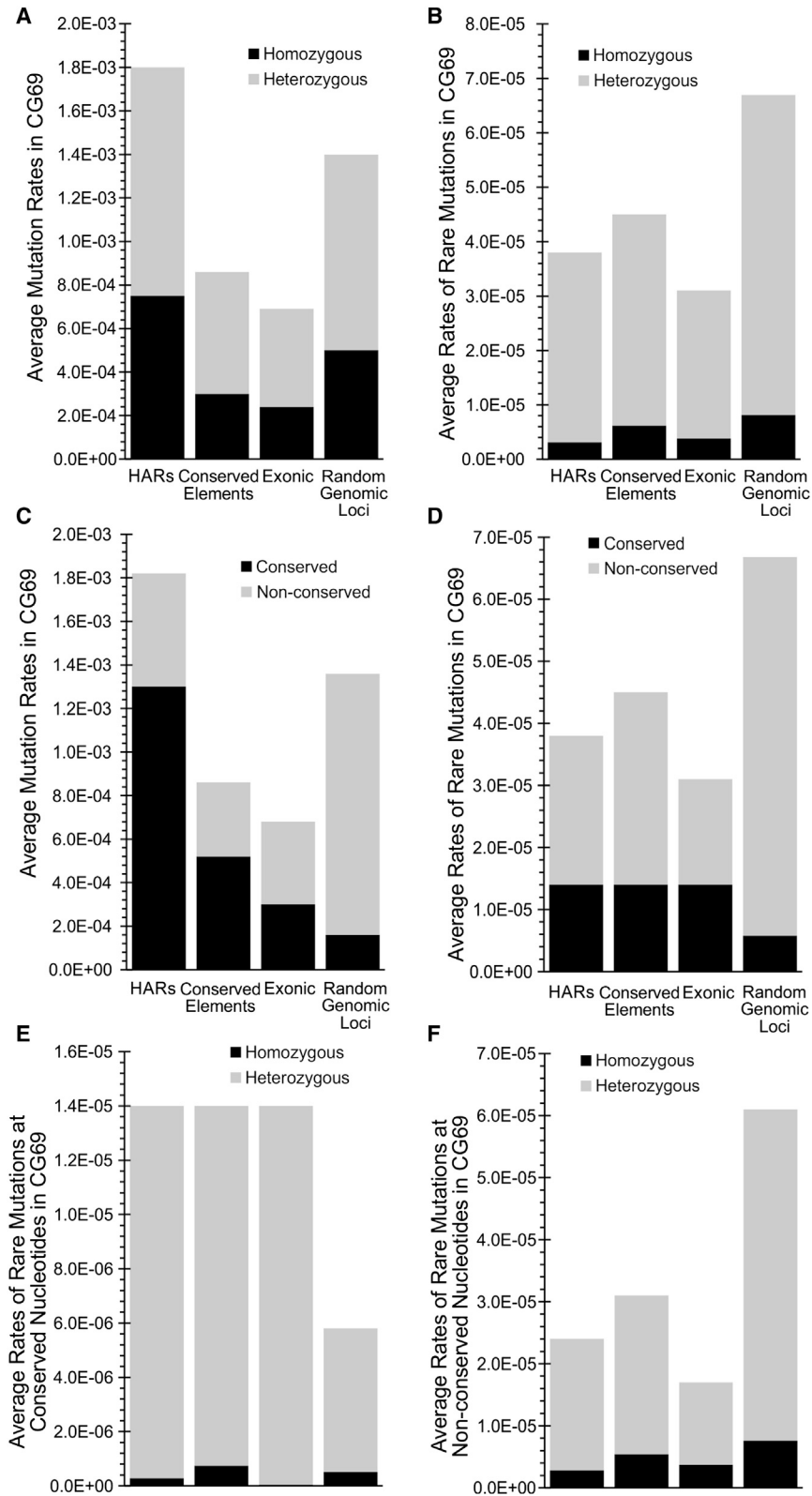
Figure S2. Differences in Mutation Rates among HARs and Other Genomic Loci in CG69, Related to Figure 1
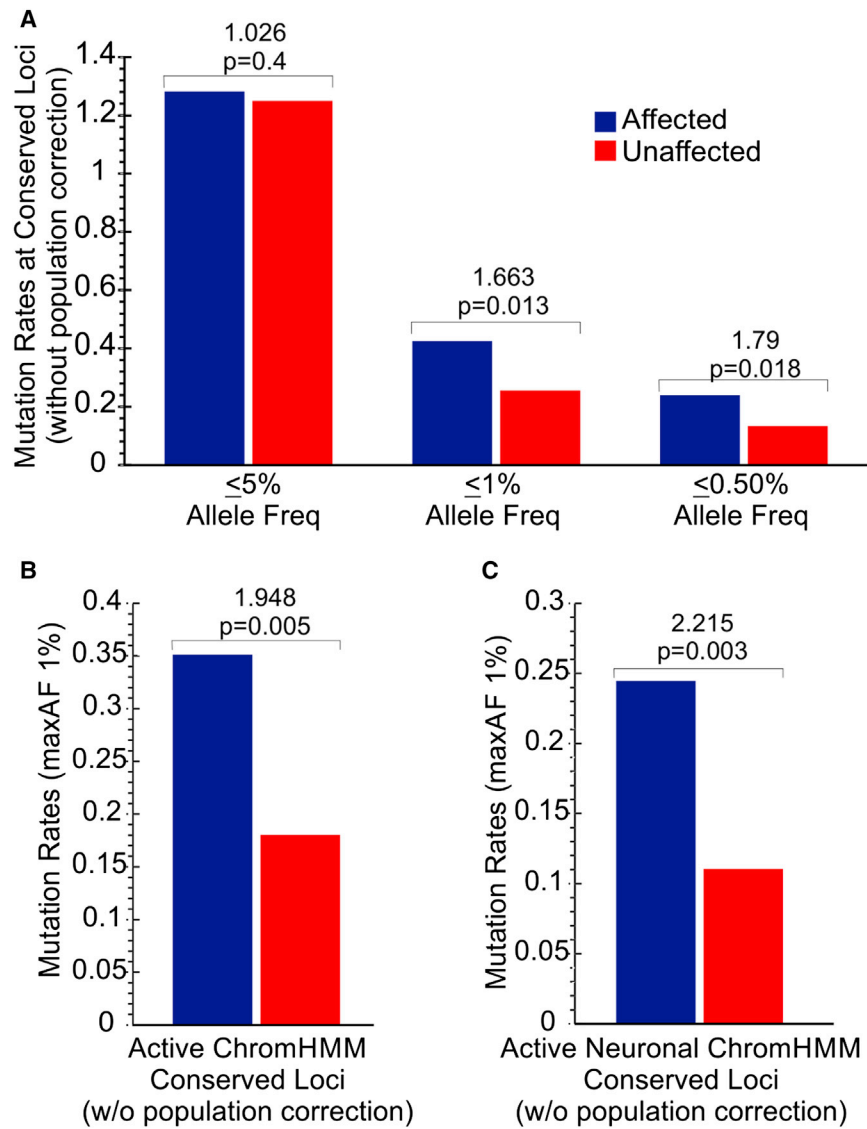
Figure S3. Enrichment of Biallelic Point Mutations in Individuals with ASD Prior to Population Structure Correction, Related to Figure 3
(A) Rate of biallelic mutations without correction for elevated homozygosity reveals the greatest excess with the rarest alleles with maxAF < 1%.
(B and C) (B) Excess of rare biallelic mutations (maxAF < 1%) at conserved loci within active regulatory elements, which is elevated within (C) active neuronal regulatory elements.
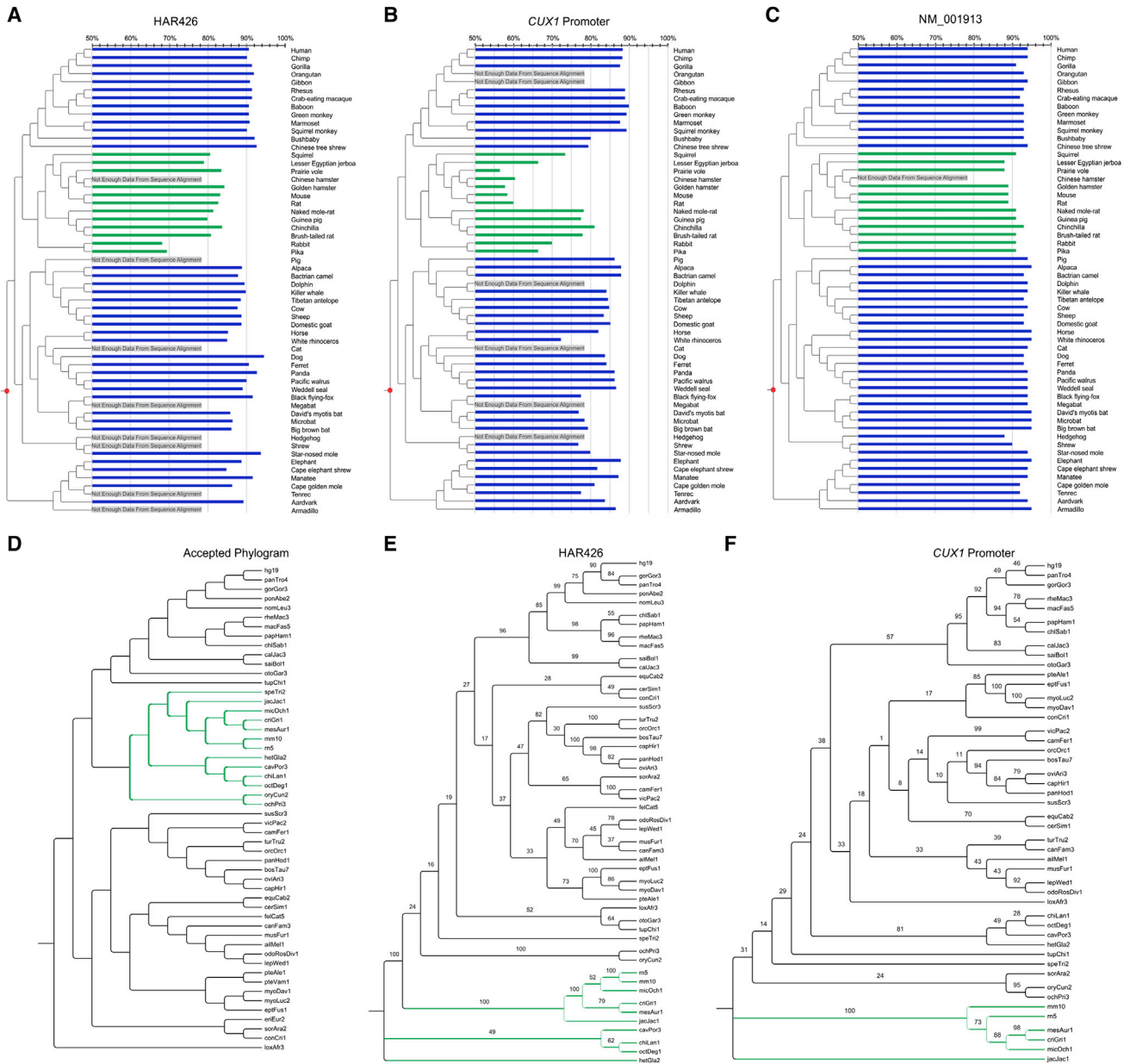
Figure S4. Phylogenetic Analysis of HAR426 and *CUX1* Promoter across Placental Mammals, Related to Table 1

(A–C) Comparison of placental mammals against the predicted common ancestor where species with assembly gaps were excluded, as indicated, for the (A) HAR426, (B) *CUX1* promoter, and (C) the coding sequence of *CUX1* transcript NM_001913.

(D–F) Comparison of (D) established phylogenetic tree against phylogenetic reconstruction of placental mammals using RaxML for (E) HAR426 and (F) the *CUX1* promoter.
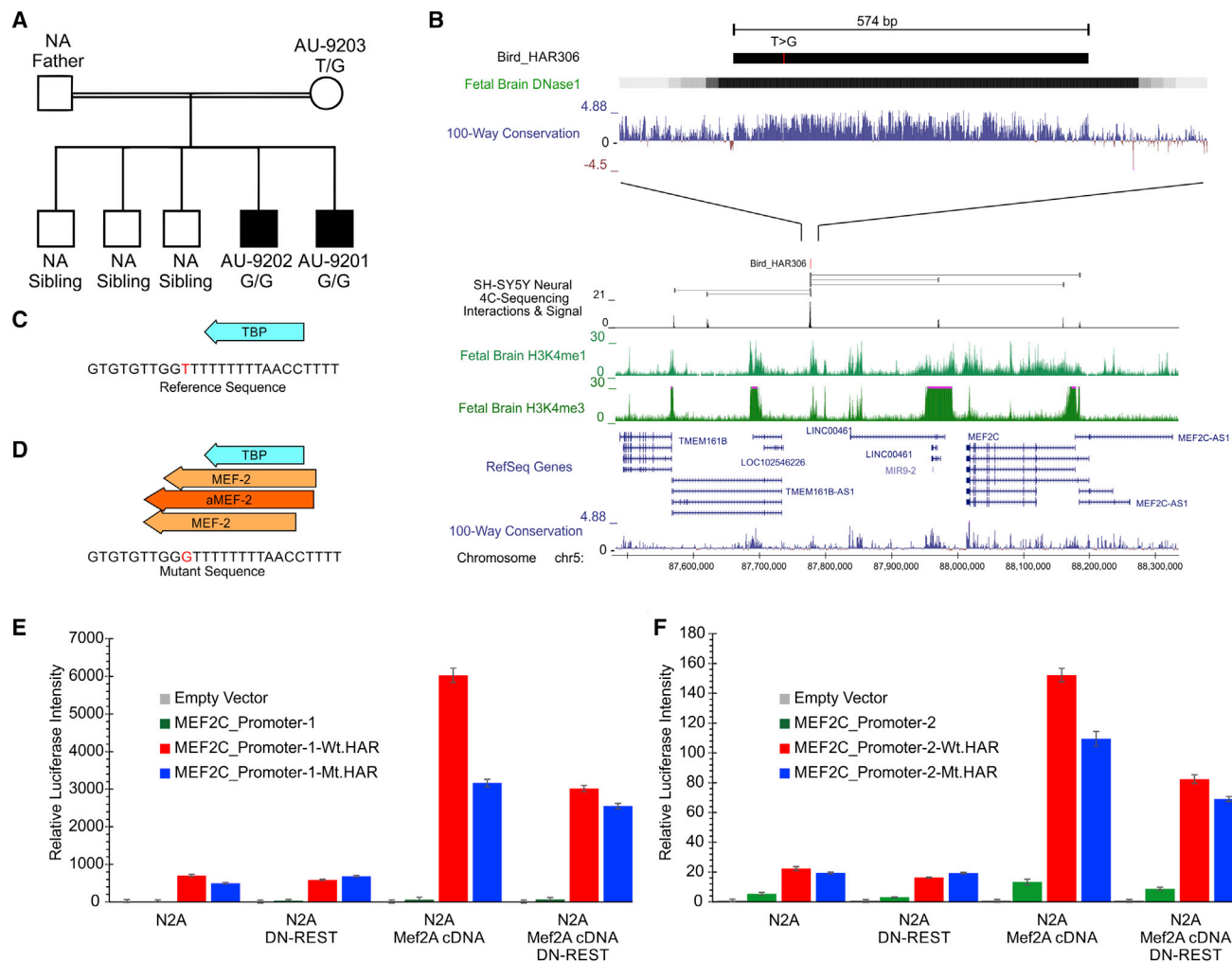
Figure S5. Homozygous Mutation Alters Regulatory Activity of HAR Interacting with Promoter Region of Known ASD/ID Gene *MEF2C*, Related to Table 1

(A and B) Consanguineous families (A) AU-9200 with a (B) homozygous T>G HAR mutation located between *TMEM161B* and *MEF2C*, where 4C-sequencing in neural tissue revealed interactions with the promoter region of *MEF2C*.

(C and D) The mutation is predicted to affect the (C) TF motifs in the reference sequence by (D) adding a *MEF2A* motif.

(E and F) Luciferase analysis of the mutant (Mt) and reference (Wt) HAR sequence using 2 flanking promoter regions (E and F) with co-transfections of DN-REST and *MEF2A* cDNA expression constructs reveal a significant decrease in enhancer activity that is exacerbated by the addition of *MEF2A*.