LETTERS
https://doi.org/10.1038/s41588-019-0433-8
nature genetics

Recessive gene disruptions in autism spectrum disorder

Ryan N. Doan[1,2], Elaine T. Lim[1,2], Silvia De Rubeis[3,4,5], Catalina Betancur[6], David J. Cutler[7], Andreas G. Chiocchetti[8], Lynne M. Overman[9], Aubrie Soucy[1], Susanne Goetze[1], Autism Sequencing Consortium[10], Christine M. Freitag[8], Mark J. Daly[2,11,12], Christopher A. Walsh[1,2,11,13], Joseph D. Buxbaum[3,4,5,14,15,16] and Timothy W. Yu[1,2,11*]

Autism spectrum disorder (ASD) affects up to 1 in 59 individuals[1]. Genome-wide association and large-scale sequencing studies strongly implicate both common variants[2–4] and rare de novo variants[5–10] in ASD. Recessive mutations have also been implicated[11–14] but their contribution remains less well defined. Here we demonstrate an excess of biallelic loss-of-function and damaging missense mutations in a large ASD cohort, corresponding to approximately 5% of total cases, including 10% of females, consistent with a female protective effect. We document biallelic disruption of known or emerging recessive neurodevelopmental genes (*CA2, DDHD1, NSUN2, PAH, RARB, ROGDI, SLC1A1, USH2A*) as well as other genes not previously implicated in ASD including *FEV* (FEV transcription factor, ETS family member), which encodes a key regulator of the serotonergic circuitry. Our data refine estimates of the contribution of recessive mutation to ASD and suggest new paths for illuminating previously unknown biological pathways responsible for this condition.

We performed a systematic analysis of exome data from the Autism Sequencing Consortium (ASC)[15], representing 2,343 affected and 5,852 unaffected individuals (Supplementary Table 1 and Supplementary Fig. 1). We cataloged a total of 696,143 autosomal loss-of-function (LOF) events (representing 28,685 unique variants in 11,745 unique genes) that introduce a stop codon or disrupt a canonical splice site (Table 1). After excluding common variants (allele frequency > 1%), there were 84,645 rare LOF events (27,648 unique alleles) for an average of approximately 10 LOF mutations per individual. After computational phasing, we found 298 events in the ASC (after filtering to exclude common polymorphisms), which is consistent with complete gene knockout (homozygous or compound heterozygous LOF mutations), affecting 266 individuals (Table 1). Affected individuals were disproportionately likely to harbor a gene knockout (62% more likely; 0.05 events per affected individual versus 0.031 per unaffected individual, $P = 0.0003$ by random permutation testing), with the bulk of the excess arising from the rarest alleles (Table 1, Supplementary Table 2 and Fig. 1a). To control for possible differences in population and family structure (for example, founder effects and/or consanguinity in the Finnish and Middle Eastern cohorts), we also normalized to the background burdens of biallelic synonymous variants (see Methods and Supplementary Tables 3–6). Individuals with autism spectrum disorder (ASD) continued to exhibit higher knockout rates after normalization (0.042 biallelic LOF events per affected individual versus 0.031 per unaffected individual, $P = 0.008$ by random permutation testing) (Supplementary Fig. 2a). Based on the observed ascertainment differentials between affected and unaffected individuals (0.050 versus 0.031, or 0.042 versus 0.031, after normalization), these burdens predict a contribution of biallelic LOF alleles to approximately 1–2% of ASD cases (see Methods).

We considered whether these findings might be extended to incorporate the impact of missense variation. Missense variation occurs much more frequently than truncating mutations, with a subset expected to be damaging. We evaluated potential missense contributions of two different types: all missense/non-synonymous variants (Nsyn), and damaging missense/non-synonymous variants (NsynD4, defined as missense changes classified as damaging by 4 or more predictive algorithms, representing 4% of all Nsyn events). Individuals with ASD exhibited an excess of biallelic events specifically involving a LOF allele in *trans* with an NsynD4 allele (Fig. 1b), an effect that persisted after normalization to synonymous rates (Supplementary Fig. 2b). Similarly, considering biallelic events involving either rare LOF or NsynD4 alleles, we found that individuals with ASD demonstrated an excess of biallelic variation ($P < 10^{-4}$) with or without normalization (Supplementary Fig. 2c and Fig. 1c). Biallelic NsynD4 alleles were also found in excess in cases (Fig. 1c), but this did not persist after normalization ($P = 0.07$, Supplementary Fig. 2c).

[1]Division of Genetics and Genomics, Department of Pediatrics, Boston Children's Hospital, Boston, MA, USA. [2]Harvard Medical School, Boston, MA, USA. [3]Seaver Autism Center for Research and Treatment, Icahn School of Medicine at Mount Sinai, New York, NY, USA. [4]Department of Psychiatry, Icahn School of Medicine at Mount Sinai, New York, NY, USA. [5]Mindich Child Health and Development Institute, Icahn School of Medicine at Mount Sinai, New York, NY, USA. [6]Neuroscience Paris Seine, Institut de Biologie Paris Seine, Sorbonne Université, INSERM, CNRS, Paris, France. [7]Department of Human Genetics, Emory University School of Medicine, Atlanta, GA, USA. [8]Department of Child and Adolescent Psychiatry, Psychosomatics and Psychotherapy, Autism Research and Intervention Center of Excellence, University Hospital Frankfurt, Goethe University, Frankfurt, Germany. [9]Human Developmental Biology Resource, Institute of Genetic Medicine, Newcastle University, International Centre for Life, Newcastle-upon-Tyne, UK. [10]A list of members and affiliations appears in the Supplementary Note. [11]The Broad Institute of MIT and Harvard, Cambridge, MA, USA. [12]Center for Human Genetic Research, Department of Medicine, Massachusetts General Hospital, Boston, MA, USA. [13]Howard Hughes Medical Institute, Boston Children's Hospital, Boston, MA, USA. [14]Department of Neuroscience, Icahn School of Medicine at Mount Sinai, New York, NY, USA. [15]Friedman Brain Institute, Icahn School of Medicine at Mount Sinai, New York, NY, USA. [16]Department of Genetics and Genomic Sciences, Icahn School of Medicine at Mount Sinai, New York, NY, USA.
*e-mail: timothy.yu@childrens.harvard.edu

Table 1 | Patterns of biallelic LOF mutation in the ASC

				Events per individual		
Allele frequency	**No. of biallelic LOF events**	**No. of unique genes involved**	**No. of individuals harboring a gene knockout**	**Unaffected**	**Affected**	***P***
All	237,936	636	8,195	28.997	29.128	0.2
≤10%	1,646	414	1,420	0.194	0.219	0.03
≤5%	621	324	559	0.068	0.096	0.0005
≤1%	298	237	266	0.031	0.050	0.003
≤0.5%	225	182	201	0.022	0.042	<0.0001
≤0.1%	170	138	146	0.016	0.032	<0.0001

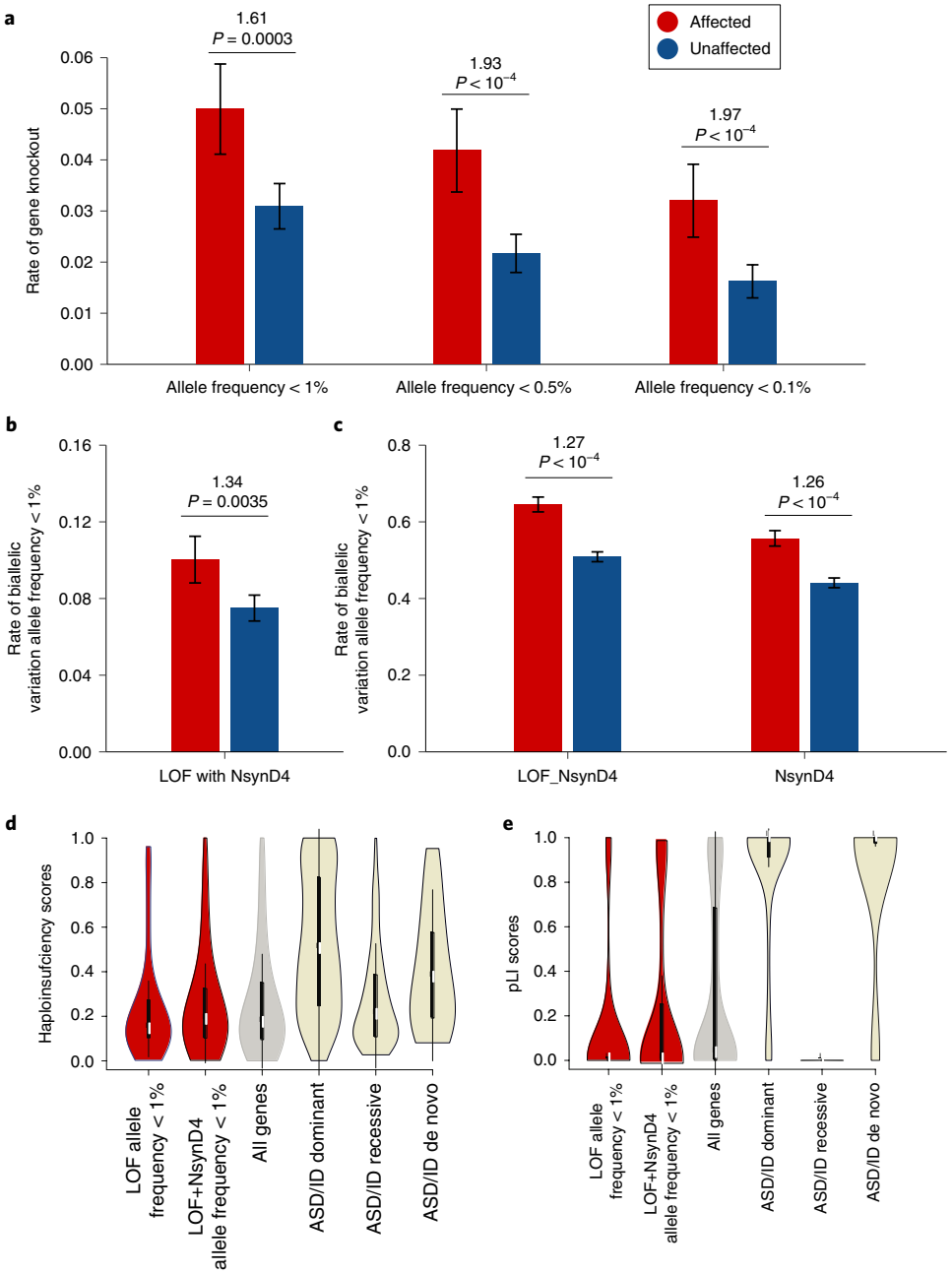

Fig. 1 | An excess of rare, damaging, biallelic mutation in ASD. a, Rates of biallelic gene knockout (strict LOF) in the ASC, stratified by diagnosis and allele frequency. **b**,**c**, Rates of biallelic variation, considering LOF variants paired with a damaging missense variant (NsynD4, predicted to be deleterious by at least four algorithms) (**b**), and LOF or NsynD4 variants, or NsynD4 variant alone (**c**). **d**,**e**, Genes impacted by biallelic variation in cases exhibit haploinsufficiency (**d**) and pLI score profiles consistent with recessive genes (**e**). Error bars represent 95% confidence intervals.

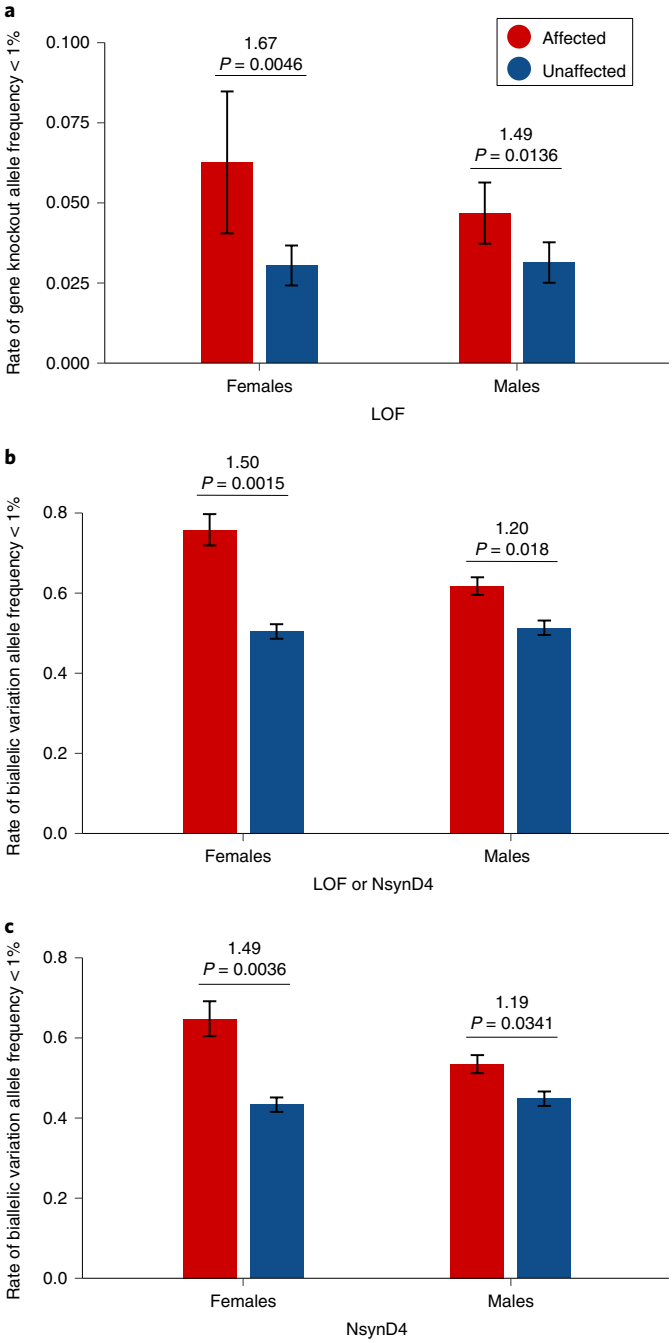

Fig. 2 | Biallelic mutations in ASD: the effects of sex. a, Rates of biallelic gene knockout (strict LOF) in the ASC, stratified by diagnosis and sex. **b**,**c**, Rates of biallelic variation stratified by sex, considering LOF or damaging missense (LOF or NsynD4) variants (**b**) or damaging missense variants alone (NsynD4) (**c**). Error bars represent 95% confidence intervals.

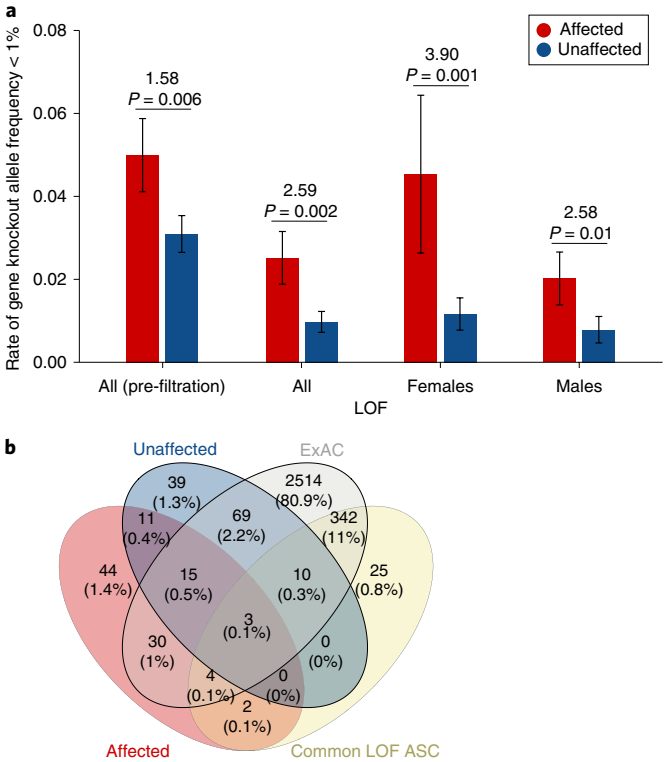

Fig. 3 | Biallelic mutations in ASD: ExAC filtration. a, Rates of biallelic gene knockout (strict LOF) after filtration of commonly inactivated genes in the ExAC. **b**, Breakdown of candidate ASD genes excluded by the presence of knockouts in the ExAC, or by knockouts that are common and/or in unaffected individuals of the ASC. Error bars represent 95% confidence intervals.

Alleles contributing to biallelic LOF knockout in cases were not in excess in the heterozygous state (Supplementary Fig. 3). Genes impacted by biallelic mutation also scored low with regard to haploinsufficiency[16] and gene constraint[17] (Fig. 1d,e and Supplementary Table 7), compared to known dominant and de novo ASD and ID genes[18]. These data support a recessive risk model, as opposed to the additive contributions of alleles acting codominantly.

ASD exhibits a striking male bias. Females with ASD exhibit a disproportionate burden of de novo and inherited single nucleotide variants (SNVs) and copy-number variants (CNVs), supporting a female protective/male susceptibility effect[19,20]. We asked whether this also applies to rare recessive mutations. Unaffected males and females did not differ in the burdens of LOF knockout (Fig. 2a and Supplementary Fig. 4a). However, females with ASD exhibited a larger excess of biallelic LOF mutations (Fig. 2a and Supplementary Fig. 4a). Similar patterns were observed when the analysis was extended to incorporate damaging missense variants, either alone (NsynD4; Fig. 2b and Supplementary Fig. 4b) or grouped with LOF mutations (LOF or NsynD4; Fig. 2c and Supplementary Fig. 4c). As observed before, LOF alleles were not in excess in the heterozygous state (Supplementary Fig. 3). From these patterns, we conclude that the female protective/male susceptibility effect also extends to recessive mutations.

The ascertainment differentials observed led us to estimate the total contribution of biallelic mutation to ASD to be 3–5% (2% from LOF mutations, 1–3% from missense mutations). Furthermore, based on the ascertainment differentials, identifiable biallelic LOF or damaging missense alleles appeared to contribute to approximately 10% of female ASD cases in the ASC (Supplementary Fig. 4b), suggesting that the category of biallelic mutation could be one of the largest genetic contributors to ASD in females.

To address the potential biological implications of these results, we examined the genes where biallelic mutations were found. We first focused on the category of strict biallelic LOF mutations, estimating from the observed ascertainment differentials that approximately 25–40% of those found in ASD patients (0.019 out of 0.05, or 0.011 out of 0.042, after normalization to synonymous burdens) contribute to their condition. A total of 109 unique genes were knocked out (biallelic LOF) in affected individuals in our dataset. We reasoned that the most relevant of these genes would be those

Table 2 | Clinically relevant gene knockouts in the ASC

Gene	Mutations	Disease relevance	Notes
CA2	hom c.232+1G>A (NM_000067)	Carbonic anhydrase deficiency	Intellectual disability, osteopetrosis, renal tubular acidosis
DDHD1	comp het c.1311-2A>T/c.2459-2A>T (NM_001160147)	Spastic paraplegia	Spastic paraplegia, mitochondrial abnormalities; only 3 cases reported
NSUN2	hom c.C1708T:p.Q570X (NM_001193455)	Autosomal recessive intellectual disability	Syndromic intellectual disability; only 7 mutations reported
PAH	hom c.C703T:p.Q235X and hom c.592_613del:p.Y198fs (NM_000277) (2 separate individuals)	Phenylketonuria	Approximately 20% of individuals with phenylketonuria have autism (finding previously reported[13])
SLC1A1	hom c.G142T:p.E48X (NM_004170)	Dicarboxylic aminoaciduria	Elevated urinary glutamate and aspartate, variable neuropsychiatric symptoms; only 3 cases reported
RARB	hom c.C78A:p.C26X (NM_000965)	Pulmonary hypoplasia-diaphragmatic hernia-anophthalmia-cardiac defect syndrome	Microphthalmia, pulmonary hypoplasia, diaphragmatic hernia, and cardiac defects; milder forms described
ROGDI	hom c.201-1G>T (NM_024589)	Kohlschütter–Tönz syndrome	Global developmental delay, epilepsy, spasticity, amelogenesis imperfecta
USH2A	comp het c.T12714G:p.Y4238X/ c.G6224A:p.W2075X (NM_206933)	Usher syndrome	Sensorineural hearing deficiencies, retinitis pigmentosa (finding previously reported[14])

that are selected against biallelic inactivation in the general population. Therefore, we developed a filter to remove any gene that was also knocked out in 1 or more of 60,706 individuals sequenced as a part of the Exome Aggregation Consortium (ExAC: see Methods)[17]. To test its effectiveness, we examined its impact on observed patterns and burdens of gene knockouts in ASD. Overall, 131 of the 227 (58%) gene knockouts observed in the ASC were also observed in the ExAC. Application of this filter reduced the overall burdens of observed biallelic knockout by half but left the ascertainment differentials intact (Fig. 3a and Supplementary Fig. 5). This substantially increased the observed proportional excess of biallelic LOF mutation: after filtering, on average, individuals with ASD were approximately 2.6 times as likely to have biallelic LOF mutations as individuals without ASD ($P=0.002$) (3.9× in females, $P=0.001$, and 2.6× in males, $P=0.01$) (Fig. 3a).

We applied a similar logic to biallelic missense mutations. We reasoned that missense alleles encountered in the homozygous state in the general population are not likely to contribute to a recessive condition. Approximately 50% of rare, biallelic damaging missense events in the ASC involved an allele encountered in the homozygous state in the ExAC (Supplementary Fig. 6a and Supplementary Table 8). Filtering these events allowed us to discern for the first time a significant excess of rare biallelic damaging missense variation (that is, NsynD4) in cases compared to controls (Supplementary Fig. 6a), with an ascertainment differential of approximately 4%. Biallelic Nsyn variants were also in slight excess ($P=0.0037$), which arose almost entirely from affected females (Supplementary Fig. 6b).

These results indicate that the observed excess in ASD is driven by genes and alleles that are biallelically constrained in the control population, strongly supporting the biological relevance of these findings.

De novo gene-disrupting mutations contribute strongly to ASD[6]. To examine possible interactions between de novo and biallelic gene-disrupting events, we compared the burden of biallelic LOF events in cases with and without LOF de novo mutations (Supplementary Fig. 10) (ref. [6]). An excess of rare biallelic LOF events is strongly evident in ASD cases lacking de novo LOF variants (Supplementary Fig. 10, $P=0.0006$). Their contribution in ASD cases with de novo LOF variants is harder to distinguish, since the observed difference in burdens of biallelic LOF events did not achieve statistical significance (Supplementary Fig. 7). These may represent orthogonal risk categories, although larger sample sizes will be required to clarify this in the future.

After ExAC filtering, a total of 57 biallelically constrained genes harbored biallelic LOF mutations in at least one ASD case. Of these, we can estimate that approximately 60% of these (based on the observed burdens, see Methods), and 80% of the knockouts observed in females, probably confer an ASD risk. To further enrich for relevant genes, we also set aside 13 genes knocked out in one or more unaffected individuals in our ASC dataset (Fig. 3b and Supplementary Table 7) and three that failed Sanger confirmation, leaving a final set of 41 genes that were specifically knocked out in individuals diagnosed with ASD, but never in the 66,558 controls (60,706 from the ExAC and 5,852 unaffected individuals from the ASC) (Supplementary Tables 9 and 10).

Several of these represent genes already implicated in disease. Eight (*CA2, DDHD1, NSUN2, PAH, RARB, ROGDI, SLC1A1, USH2A*) were associated with recessive genetic conditions involving neurodevelopmental delay (see Table 2 and Supplementary Information for clinical details). In at least six of these cases (*CA2*, *NSUN2*, *PAH*, *RARB*, *ROGDI*, *USH2A*), available medical and phenotypic records were concordant with the clinical features predicted from the genetic finding (Supplementary Information). Phenotypic data were either unavailable or incomplete, but not discordant, for other cases. In two cases, knockouts were encountered in genes that cause well-characterized autosomal recessive conditions without clear connection to autism (*RFX5* and *DNAI2*, associated with autosomal recessive immunodeficiency and primary ciliary dyskinesia, respectively). In two other cases, knockouts were observed in genes associated with autosomal dominant disorders via recurrent gain-of-function mutations (*IFITM5* and *KIF22*, causing osteogenesis imperfecta type V and spondyloepimetaphyseal dysplasia with joint laxity type 2, respectively). The phenotypic consequence of recessive LOF mutations for these two genes is unknown. Interestingly, *KIF22* is one of 25 genes in the 593 kb core interval affected by 16p11.2 deletions[21]. *KIF22* was proposed as one of two genes in the interval that demonstrates dosage sensitivity, with knockdown causing axon tract damage and behavioral defects in zebrafish[22]; disruption of a *KIF22* ortholog in *Drosophila* causes defects in synapse development[23].

Several of the remaining 33 genes are of strong neurobiological interest. One affected individual bore a homozygous LOF mutation (p.E48X) in *SLC1A1*, where mutations have been linked to

Table 3 | Biallelic mutations identified in cases and controls previously reported as pathogenic or likely pathogenic in ClinVar

Gene	Condition	Sample identifier	Sex	Mutation	ClinVar (HGMD)	Variant pathogenicity	Damaging predictions
Found in 13 cases							
ASS1	Citrullinemia type I	SKUSE5080161	M	hom c.1168G>A:p.G390R (NM_000050.4)	6329 (CM900037)	Pathogenic	10/10
				het c.3207C>A:p.H1069Q (NM_000053.3)	3848 (CM930059)	Pathogenic	11/11
ATP7B	Wilson disease	80001103644	M	het c.4087T>C:p.S1363P (NM_000053.3)	Same site as p.S1363F (ClinVar 188908, HGMD CM992829)	(p.S1363F = pathogenic)	12/12
B3GALNT2	Muscular dystrophy with brain malformations	80001103921	F	het c.1368+1G>A (NM_152490.4)		(Likely pathogenic)	N/A
				het c.979G>A:p.D327N (NM_152490.4)	424766 (CM144415)	Likely pathogenic	10/12
CA2	Osteopetrosis autosomal recessive 3	AU-10501	F	hom c.232+1G>A (NM_000067.2)	288909	Pathogenic	N/A
COQ8A	Coenzyme Q10 deficiency	AU-13901	M	hom c.1805C>G:p.P602R (NM_020247.4)	214050 (CM141297)	Pathogenic	11/12
GBE1	Glycogen storage disease type IV and adult polyglucosan body disease	AU-24601	F	hom c.986A>G:p.Y329C (NM_000158.3)	371439 (CM128776)	Pathogenic	11/11
GLRA1	Hereditary hyperekplexia	AU-12001	M	hom c.277C>T:p.R93W (NM_000171.3)	225379 (CM104312)	Different missense, likely pathogenic in ClinVar (225379)	10/11
HADH	Deficiency of 3-hydroxyacyl-CoA dehydrogenase	09C86928	M	hom c.676T>C:p.Y226H (NM_005327.4)	212734 (CM064046)	Likely pathogenic	11/11
MYO15A	Deafness, autosomal recessive 3	SAGA-96	F	het c.5515C>T:p.Q1839X (NM_016239.3)	N/A	(Likely pathogenic)	N/A
				het c.8183G>A:p.R2728H (NM_016239.3)	228276 (CM117958)	Pathogenic	10/10
NAGLU	Sanfilippo syndrome B	AU-16201	F	homc.934G>A:p.D312N (NM_000263.3)	437446 (CM113463)	Pathogenic	12/12
PAH	Phenylketonuria	AU-16201	F	het c.782G>A:p.R261Q (NM_000277.1)	582 (CM910287)	Pathogenic	12/12
				het c.842C>T:p.P281L (NM_000277.1)	589 (CM910292)	Pathogenic	11/11
POLG	Alpers–Huttenlocher syndrome, childhood myocerebrohepatopathy spectrum disorder	AU-16201	F	hom c.3151G>A:p.G1051R (NM_002693.2)	13501 (CM040472)	Pathogenic	11/11
USH2A	Usher syndrome type 2A	09C96107	M	het c.12714T>G:p.Y4238X (NM_206933.2)	48405 (CM134383)	Pathogenic	N/A
				het c.6224G>A:p.W2075X (NM_206933.2)	48554 (CM134384)	Pathogenic	
Found in 3 controls							
MCCC2	3-methylcrotonyl-CoA carboxylase deficiency	AU-13303	F	hom c.295G>C:p.E99Q(NM_022132.4)	1920 (CM010910)	Pathogenic	10/10
PAH	Phenylketonuria	AU-16503	F	hom c.898G>T:p.A300S(NM_000277.1)	92751 (CM920555)	Pathogenic	12/12
		AU-16502	M	hom c.898G>T:p.A300S(NM_000277.1)	92751 (CM920555)	Pathogenic	

dicarboxylic aminoaciduria[24], an extremely rare disorder with prior associations with intellectual disability[25,26] and obsessive compulsive disorder[24]. *SLC1A1* encodes a brain-expressed, activity-regulated glutamate transporter[27,28] that modulates glutamatergic neurotransmission via a variety of mechanisms[29–34] *SLC1A1* knockout in mice results in altered locomotor behavior[35] and learning and memory defects[36].

Two brothers with ASD bore homozygous stop-gain mutations in *FEV* (FEV transcription factor, ETS family member), which encodes a transcription factor that is required for both the development and function of serotonergic neurons (Supplementary Fig. 8a). These neurons send projections that ramify broadly throughout the central nervous system (Supplementary Fig. 8b), and have wide-ranging neuromodulatory effects on many physiological states and behaviors, including mood, aggression, anxiety, sleep and movement[37]. Serotonergic dysfunction has been strongly implicated in neuropsychiatric conditions including ASD[38], schizophrenia and mood disorders[37]. In mice, *Fev/Pet-1* has been shown to be highly expressed in the serotonergic raphe nuclei[39], and inactivation of *Fev* during development[40] or adulthood[41] results in the depletion of serotonergic neurons and neurobehavioral changes including anxiety and aggression[40,41]. Human *FEV* is also expressed in the serotonergic raphe neurons in the adult brain[42,43]; we found similar results via in situ hybridization analysis of human fetal brain (Supplementary Fig. 8c–f). The older affected brother was diagnosed with autism and had a measured IQ of 69. He exhibited severe, stereotyped aggressive and self-injurious behaviors. He had a body mass index of 25.6 (99%) and mildly dysmorphic facial features. The electroencephalogram showed bilateral Rolandic focus with generalization. The younger affected brother was diagnosed with pervasive developmental disorder-not otherwise specified and intellectual disability (IQ not measured), had a body mass index of 17.7 (91%) and no facial dysmorphism (further details can be found in the Supplementary Information). We are not aware of any previous report of a human LOF phenotype for *FEV*.

Gene knockouts were also identified in *FCHSD2*, encoding a cytoskeletal adapter protein that is highly expressed in human fetal brain and is the human homolog of *Drosophila* Nervous Wreck[44], *NCOA7*, a brain-specific transcriptional coactivator for steroid and nuclear hormone receptors[45] and *MOB1B*, a downstream effector in the Hippo signaling pathway, a pathway recently implicated in the regulation of brain size[46]. Many additional genes in the candidate set have abundant brain expression (*BRMS1L*, *ELOF1*, *GCN1L1*, *NUPR1*, *SLC22A6*) but less clear neurobiological activity.

We also expanded our scope to incorporate biallelic missense mutations. There were 409 distinct genes that harbored biallelic, damaging missense mutations in cases but not in controls (Supplementary Fig. 9 and Supplementary Table 11). Only one gene was independently hit in multiple families: aminomethyltransferase (*AMT*), an established cause of non-ketotic hyperglycinemia. Hypomorphic mutations in *AMT* have previously been reported to cause ASD without the non-ketotic hyperglycinemia's characteristic metabolic abnormalities. No other gene was hit more than once, so larger sample sizes are necessary to establish any of these genes as ASD genes on this basis alone (Supplementary Table 11 and Supplementary Figs. 10 and 11). However, there were 13 cases where the affected individuals carried biallelic mutations in alleles of established medical importance (in ClinVar as pathogenic or likely pathogenic; Table 3, Supplementary Table 12 and Supplementary Information). All were diagnostic of genetic disorders with known neurodevelopmental consequences, including classic metabolic diseases, mitochondrial depletion syndrome and other syndromic conditions. When combined with the previously described 8 LOF cases, biallelic mutations in known genes constituted approximately 1% of our cohort (21 out of 2,343 affected individuals), underscoring the importance of clinical screening for monogenic recessive conditions in this patient population.

Our study extends prior estimates of the contribution of recessive mutations to ASD[13,14] and proves that the female protective/male susceptibility effect in ASD also applies to biallelic mutations. Expanding these efforts to larger ASD cohorts should offer a path forward to delineate previously unknown neurobiological mechanisms in autism.

Online content

Any methods, additional references, Nature Research reporting summaries, source data, statements of code and data availability and associated accession codes are available at https://doi.org/10.1038/s41588-019-0433-8.

Received: 25 October 2018; Accepted: 2 May 2019;
Published online: 17 June 2019

References

1. Baio, J. et al. Prevalence of autism spectrum disorder among children aged 8 years. Autism and developmental disabilities monitoring network, 11 sites, United States, 2014. *MMWR Surveill. Summ.* **63**, 1–22 (2014).
2. Robinson, E. B. et al. Genetic risk for autism spectrum disorders and neuropsychiatric variation in the general population. *Nat. Genet.* **48**, 552–555 (2016).
3. Gaugler, T. et al. Most genetic risk for autism resides with common variation. *Nat. Genet.* **46**, 881–885 (2014).
4. Weiner, D. J. et al. Polygenic transmission disequilibrium confirms that common and rare variation act additively to create risk for autism spectrum disorders. *Nat. Genet.* **49**, 978–985 (2017).
5. Krumm, N. et al. Excess of rare, inherited truncating mutations in autism. *Nat. Genet.* **47**, 582–588 (2015).
6. De Rubeis, S. et al. Synaptic, transcriptional and chromatin genes disrupted in autism. *Nature* **515**, 209–215 (2014).
7. Levy, D. et al. Rare de novo and transmitted copy-number variation in autistic spectrum disorders. *Neuron* **70**, 886–897 (2011).
8. Sanders, S. J. et al. De novo mutations revealed by whole-exome sequencing are strongly associated with autism. *Nature* **485**, 237–241 (2012).
9. Iossifov, I. et al. De novo gene disruptions in children on the autistic spectrum. *Neuron* **74**, 285–299 (2012).
10. Ronemus, M., Iossifov, I., Levy, D. & Wigler, M. The role of de novo mutations in the genetics of autism spectrum disorders. *Nat. Rev. Genet.* **15**, 133–141 (2014).
11. Morrow, E. M. et al. Identifying autism loci and genes by tracing recent shared ancestry. *Science* **321**, 218–223 (2008).
12. Chahrour, M. H. et al. Whole-exome sequencing and homozygosity analysis implicate depolarization-regulated neuronal genes in autism. *PLoS Genet.* **8**, e1002635 (2012).
13. Yu, T. W. et al. Using whole-exome sequencing to identify inherited causes of autism. *Neuron* **77**, 259–273 (2013).
14. Lim, E. T. et al. Rare complete knockouts in humans: population distribution and significant role in autism spectrum disorders. *Neuron* **77**, 235–242 (2013).
15. Buxbaum, J. D. et al. The autism sequencing consortium: large-scale, high-throughput sequencing in autism spectrum disorders. *Neuron* **76**, 1052–1056 (2012).
16. Huang, N., Lee, I., Marcotte, E. M. & Hurles, M. E. Characterising and predicting haploinsufficiency in the human genome. *PLoS Genet.* **6**, e1001154 (2010).
17. Lek, M. et al. Analysis of protein-coding genetic variation in 60,706 humans. *Nature* **536**, 285–291 (2016).
18. Betancur, C. Etiological heterogeneity in autism spectrum disorders: more than 100 genetic and genomic disorders and still counting. *Brain Res.* **1380**, 42–77 (2011).
19. Jacquemont, S. et al. A higher mutational burden in females supports a "female protective model" in neurodevelopmental disorders. *Am. J. Hum. Genet.* **94**, 415–425 (2014).
20. Robinson, E. B., Lichtenstein, P., Anckarsäter, H., Happé, F. & Ronald, A. Examining and interpreting the female protective effect against autistic behavior. *Proc. Natl Acad. Sci. USA* **110**, 5258–5262 (2013).
21. Weiss, L. A. et al. Association between microdeletion and microduplication at 16p11.2 and autism. *N. Engl. J. Med.* **358**, 667–675 (2008).
22. Blaker-Lee, A., Gupta, S., McCammon, J. M., De Rienzo, G. & Sive, H. Zebrafish homologs of genes within 16p11.2, a genomic region associated with brain disorders, are active during brain development, and include two deletion dosage sensor genes. *Dis. Model. Mech.* **5**, 834–851 (2012).
23. Park, S. M., Littleton, J. T., Park, H. R. & Lee, J. H. *Drosophila* homolog of human KIF22 at the autism-linked 16p11.2 loci influences synaptic connectivity at larval neuromuscular junctions. *Exp. Neurobiol.* **25**, 33–39 (2016).

24. Bailey, C. G. et al. Loss-of-function mutations in the glutamate transporter SLC1A1 cause human dicarboxylic aminoaciduria. *J. Clin. Invest.* **121**, 446–453 (2011).
25. Teijema, H. L., van Gelderen, H. H., Giesberts, M. A. & Laurent de Angulo, M. S. Dicarboxylic aminoaciduria: an inborn error of glutamate and aspartate transport with metabolic implications, in combination with a hyperprolinemia. *Metab. Clin. Exp.* **23**, 115–123 (1974).
26. Swarna, M., Rao, D. N. & Reddy, P. P. Dicarboxylic aminoaciduria associated with mental retardation. *Hum. Genet.* **82**, 299–300 (1989).
27. Rothstein, J. D. et al. Localization of neuronal and glial glutamate transporters. *Neuron* **13**, 713–725 (1994).
28. Ross, J. R., Porter, B. E., Buckley, P. T., Eberwine, J. H. & Robinson, M. B. mRNA for the EAAC1 subtype of glutamate transporter is present in neuronal dendrites in vitro and dramatically increases in vivo after a seizure. *Neurochem. Int.* **58**, 366–375 (2011).
29. Nieoullon, A. et al. The neuronal excitatory amino acid transporter EAAC1/EAAT3: does it represent a major actor at the brain excitatory synapse? *J. Neurochem.* **98**, 1007–1018 (2006).
30. Diamond, J. S. Neuronal glutamate transporters limit activation of NMDA receptors by neurotransmitter spillover on CA1 pyramidal cells. *J. Neurosci.* **21**, 8328–8338 (2001).
31. Bianchi, M. G., Bardelli, D., Chiu, M. & Bussolati, O. Changes in the expression of the glutamate transporter EAAT3/EAAC1 in health and disease. *Cell. Mol. Life Sci.* **71**, 2001–2015 (2014).
32. Stafford, M. M., Brown, M. N., Mishra, P., Stanwood, G. D. & Mathews, G. C. Glutamate spillover augments GABA synthesis and release from axodendritic synapses in rat hippocampus. *Hippocampus* **20**, 134–144 (2010).
33. Mathews, G. C. & Diamond, J. S. Neuronal glutamate uptake contributes to GABA synthesis and inhibitory synaptic strength. *J. Neurosci.* **23**, 2040–2048 (2003).
34. Scimemi, A., Tian, H. & Diamond, J. S. Neuronal transporters regulate glutamate clearance, NMDA receptor activation, and synaptic plasticity in the hippocampus. *J. Neurosci.* **29**, 14581–14595 (2009).
35. Peghini, P., Janzen, J. & Stoffel, W. Glutamate transporter EAAC-1-deficient mice develop dicarboxylic aminoaciduria and behavioral abnormalities but no neurodegeneration. *EMBO J.* **16**, 3822–3832 (1997).
36. Lee, S., Park, S. H. & Zuo, Z. Effects of isoflurane on learning and memory functions of wild-type and glutamate transporter type 3 knockout mice. *J. Pharm. Pharmacol.* **64**, 302–307 (2012).
37. Lucki, I. The spectrum of behaviors influenced by serotonin. *Biol. Psychiatry* **44**, 151–162 (1998).
38. Muller, C. L., Anacker, A. M. J. & Veenstra-VanderWeele, J. The serotonin system in autism spectrum disorder: from biomarker to animal models. *Neuroscience* **321**, 24–41 (2016).
39. Hendricks, T., Francis, N., Fyodorov, D. & Deneris, E. S. The ETS domain factor Pet-1 is an early and precise marker of central serotonin neurons and interacts with a conserved element in serotonergic genes. *J. Neurosci.* **19**, 10348–10356 (1999).
40. Hendricks, T. J. et al. *Pet-1* ETS gene plays a critical role in 5-HT neuron development and is required for normal anxiety-like and aggressive behavior. *Neuron* **37**, 233–247 (2003).
41. Liu, C. et al. *Pet-1* is required across different stages of life to regulate serotonergic function. *Nat. Neurosci.* **13**, 1190–1198 (2010).
42. Iyo, A. H., Porter, B., Deneris, E. S. & Austin, M. C. Regional distribution and cellular localization of the ETS-domain transcription factor, FEV, mRNA in the human postmortem brain. *Synapse* **57**, 223–228 (2005).
43. Maurer, P. et al. The Ets transcription factor Fev is specifically expressed in the human central serotonergic neurons. *Neurosci. Lett.* **357**, 215–218 (2004).
44. Cao, H. et al. FCHSD1 and FCHSD2 are expressed in hair cell stereocilia and cuticular plate and regulate actin polymerization in vitro. *PLoS ONE* **8**, e56516 (2013).
45. Shao, W., Halachmi, S. & Brown, M. ERAP140, a conserved tissue-specific nuclear receptor coactivator. *Mol. Cell. Biol.* **22**, 3358–3372 (2002).
46. Poon, C. L. C., Mitchell, K. A., Kondo, S., Cheng, L. Y. & Harvey, K. F. The Hippo pathway regulates neuroblasts and brain size in *Drosophila melanogaster*. *Curr. Biol.* **26**, 1034–1042 (2016).

Acknowledgements

We thank A. Hossain and N. Hatem for their help with sample preparation, and J. Kerwin for her help with analysis of in situ hybridization results. R.N.D. was supported by an NIH T32 fellowship from the Fundamental Neurobiology Training Grant (no. 5 T32 NS007484-14) and the Nancy Lurie Marks Clinical and Research Fellowship Program in Autism. The ASC is supported by the National Institute of Mental Health (NIMH; grant nos. MH100233, MH100229, MH100209, MH100239, MH111661, MH111660, MH111662 and MH111658). Collection of the PAGES cohort is supported by the NIMH (grant no. MH097849). This work was supported in part through the computational resources provided by the Department of Scientific Computing at the Icahn School of Medicine at Mount Sinai, and the Research Information Technology Group at Harvard Medical School, which is partially supported by National Institutes of Health grant no. NCRR 1S10RR028832-01. Human embryonic and fetal material was provided jointly by the MRC/Wellcome Trust (grant no. MR/R006237/1) Human Developmental Biology Resource (www.hdbr.org). C.A.W. is an Investigator of the Howard Hughes Medical Institute. C.A.W. and T.W.Y. were supported by NIMH grant nos. RC2MH089952 and RO1MH083565. T.W.Y. was supported by grant nos. NIH/NIMH R01MH113761, NICHD/NHGRI/NIH U19HD077671 and NIH/NICHD U24HD0938487, and by a SFARI Pilot Research Award. S.D.R. and J.D.B. are supported by the Beatrice and Samuel A. Seaver Foundation.

Author contributions

T.W.Y., R.N.D., E.T.L. and M.J.D. designed the study, with important additional contributions from C.B., D.J.C, C.A.W. and J.D.B. R.N.D., E.T.L. and A.S. performed the data analyses. S.D.R. and S.G. performed the Sanger validation. A.G.C. and C.M.F. characterized the *FEV* family. L.O., S.G. and T.W.Y. designed and performed the in situ expression analyses. R.N.D. and T.W.Y. wrote the manuscript. All authors reviewed and approved the manuscript.

Competing interests

The authors declare no competing interests.

Additional information

Supplementary information is available for this paper at https://doi.org/10.1038/s41588-019-0433-8.

Reprints and permissions information is available at www.nature.com/reprints.

Correspondence and requests for materials should be addressed to T.W.Y.

Publisher's note: Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Methods

Whole-exome data analyses. Whole-exome sequencing data were gathered as part of the ASC[47]. We focused our analysis on nine datasets generated using the Illumina sequencing platform (Supplementary Table 1). Alignment and variant calling were performed as described previously[6]. Low-quality variants were filtered with the following criteria: (1) SNPs with genotyping quality < 20; (2) SNPs not passing the standard PASS filter; and (3) all indels (due to excessive false positives leading to spurious frameshift calls). Variant annotation was performed using ANNOVAR[48] v.2017June01 to add allele frequencies, functional predictions, conservation and gene annotations including Online Mendelian Inheritance in Man and Human Gene Mutation Database (HGMD)[49] disease associations. Allele frequencies were also calculated for each contributing cohort and across the entire collection of ASC samples analyzed. Variants were categorized by their predicted impact into synonymous, non-synonymous (missense) or LOF (altering a canonical splice site or resulting in a stop-gain). Additional classification of damaging missense variants (NsynD4) was performed for variants predicted to be deleterious by at least 4 out of 6 algorithms (SIFT[50], PolyPhen2_HDIV[51], PolyPhen2_HVAR, MutationTaster[52], MutationAssessor[53] and LRT[54]).

Phasing was performed using Beagle v.4 (ref. [55]) on the ASC dataset. Variant call formats were subdivided by chromosome and cohort before processing with Beagle using default parameters.

Biallelic burden calculations. Allele frequency filtration was performed as described in the text using available public databases (1000 Genomes, Exome Variant Server, ExAC). For rare variant calculations at the maximum allele frequencies of 0.1, 0.5, 1, 5 and 10%, additional cohort-specific filters (cohort-specific allele frequencies of 5 or 10%) were applied to account for allelic variants that may be prevalent in a population but are not well represented in existing control databases. The burdens of homozygous and compound heterozygous variants were then counted for each cohort and for the entire ASC, stratifying by affected/unaffected status, sex and mutational category (LOF, Nsyn, NsynD4, Syn).

To control for potential differences in underlying family and population structure of some of the constituent cohorts, we also measured rates of synonymous biallelic variation. In two cohorts (from the Middle East and Finland, respectively), rates of biallelic synonymous variation in affected individuals were higher than rates of biallelic synonymous variation in unaffected individuals (Supplementary Table 3). Therefore, for these cohorts, we normalized the LOF, NsynD4 and Nsyn burden calculations to the background rate of biallelic synonymous variation, to isolate the impact of biallelic LOF variation independent of background homozygosity in these cohorts.

The contribution of mutations to ASD risk was estimated using the following approach. First, the ascertainment differential is calculated as the difference between the burden of biallelic mutations in cases and in controls and represents the estimated fraction of mutations that contribute to ASD. For example, an ascertainment differential of 0.01 suggests that 1% of cases have a biallelic mutation contributing to ASD risk. The fraction of observed biallelic mutations contributing to disease is then calculated as the ascertainment differential divided by the total mutation burden in cases. For example, given a burden of 0.025 biallelic LOF events in cases and 0.01 biallelic LOF events in controls, the ascertainment differential is 0.015 and the percentage of mutations identified that contribute to disease risk is 0.015 divided by 0.025, or 60%.

For burden comparisons, significance was assessed by permutation testing using 10,000 random assignments of diagnostic categorization to individuals in the study. A significance threshold of $P < 0.05$ (that is, without correction for multiple testing) was used since the allele frequency cutoffs and variant annotation classes (LOF, LOF + NsynD4, LOF or NsynD4, NsynD4, Nsyn) analyzed are not discrete. We note that all ExAC-filtered biallelic burden results (that is, excess biallelic LOF, NsynD4 and Nsyn mutations, summarized in Fig. 3 and Supplementary Figs. 5 and 6) were significant even with the application of maximally conservative Bonferroni correction (for example, $P < 0.01$, assuming five fully independent hypotheses).

ExAC filtration. To generate a list of genes commonly inactivated by biallelic LOF mutation, we analyzed whole-exome data from 60,706 individuals compiled by ExAC[17]. We applied ANNOVAR to systematically re-annotate all variants from ExAC against RefSeq complete protein-coding genes (v.2017June01). Since phasing was unavailable, we focused on genes bearing putative homozygous LOF variants (any variant in stop-gain, frameshift or canonical splicing change). Homopolymer expansions/contractions in canonical splice regions were ignored, since expansions and contractions in these regions are common and have little impact on splicing. With these parameters, a total of 2,986 genes were found to be inactivated by homozygous stop-gain, frameshift or canonical splice site mutation in ExAC. We also used this re-annotated dataset to generate a list of ExAC alleles encountered in the homozygous state, to identify missense and other mutations that are more likely to be clinically benign. For the final LOF candidate genes presented in Supplementary Tables 9 and 10, contributory alleles were also confirmed to be absent in the homozygous state in the Genome Aggregation Database dataset[17].

Variant confirmation. Sanger sequencing was used to confirm the accuracy of exome variant calls for a subset of LOF and NsynD4 variants with an empirical validation rate of 96% (107 of 114 alleles tested). These included all LOF alleles reported in this study for which DNA samples were available. If DNA samples were not available, we noted the presence of allelic variants in known databases and/or whether predicted genotypes followed Mendelian segregation patterns in parental exome data (where available) as additional support.

Human tissue collection and in situ hybridization for *FEV*. Human embryonic and fetal tissues were obtained from the MRC/Wellcome Trust Human Developmental Biology Resource (www.hdbr.org)[56] following appropriate consent from the women involved and adhering to the relevant Human Tissue Authority guidelines with approval from Research Ethics Committee, NRES Committee North East, Newcastle & North Tyneside 1.

RNA probes were produced by amplifying human embryonic complementary DNA using a forward primer to *FEV* (NM_017521, nucleotides 1286–1305) tagged with a T7 RNA promoter sequence and a reverse primer (NM_017521, nucleotides 1862–1843) tagged with an SP6 RNA promoter sequence (Supplementary Table 13). The resulting fragment was transcribed in vitro using the DIG RNA Labeling Kit (Roche).

The in situ hybridizations were performed as described previously[47], except that a 60 ng RNA probe in 100 µl ULTRAhyb (Ambion) per slide was used for the hybridization. The post-hybridization washes were modified as follows: slides were washed for 10 min in 5× SSC buffer, twice in 2× SSC, 50% formamide/1× SSC, 1× SSC and finally in 0.1× SSC. The hybridization incubation and all post-hybridization steps were performed at 63.5 °C. Three human fetal brain samples were used: a Carnegie Stage 23 brain (12959), a Carnegie Stage 23 head (12011) and a 10 weeks post-conception hindbrain (sample 12721). The findings described were confirmed in at least two specimens. Expression patterns were analyzed relative to spatial models provided by the Human Developmental Studies Network human gene expression spatial database project (http://www.hudsen.org)[57].

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

Data availability

Data included in this manuscript have been deposited at the database of Genotypes and Phenotypes, merged with published data under accession number phs000298.v4.p3. Correspondence and requests for materials should be addressed to Timothy.yu@childrens.harvard.edu.

References

47. Cheng, Y. Z. et al. Investigating embryonic expression patterns and evolution of *AHI1* and *CEP290* genes, implicated in Joubert syndrome. *PLoS ONE* **7**, e44975 (2012).
48. Wang, K., Li, M. & Hakonarson, H. ANNOVAR: functional annotation of genetic variants from high-throughput sequencing data. *Nucleic Acids Res.* **38**, e164 (2010).
49. Stenson, P. D. et al. The Human Gene Mutation Database: towards a comprehensive repository of inherited mutation data for medical research, genetic diagnosis and next-generation sequencing studies. *Hum. Genet.* **136**, 665–677 (2017).
50. Ng, P. C. & Henikoff, S. SIFT: predicting amino acid changes that affect protein function. *Nucleic Acids Res.* **31**, 3812–3814 (2003).
51. Adzhubei, I. A. et al. A method and server for predicting damaging missense mutations. *Nat. Methods* **7**, 248–249 (2010).
52. Schwarz, J. M., Rödelsperger, C., Schuelke, M. & Seelow, D. MutationTaster evaluates disease-causing potential of sequence alterations. *Nat. Methods* **7**, 575–576 (2010).
53. Reva, B., Antipin, Y. & Sander, C. Predicting the functional impact of protein mutations: application to cancer genomics. *Nucleic Acids Res.* **39**, e118 (2011).
54. Chun, S. & Fay, J. C. Identification of deleterious mutations within three human genomes. *Genome Res.* **19**, 1553–1561 (2009).
55. Browning, S. R. & Browning, B. L. Rapid and accurate haplotype phasing and missing-data inference for whole-genome association studies by use of localized haplotype clustering. *Am. J. Hum. Genet.* **81**, 1084–1097 (2007).
56. Gerrelli, D., Lisgo, S., Copp, A. J. & Lindsay, S. Enabling research with human embryonic and fetal tissue resources. *Development* **142**, 3073–3076 (2015).
57. Kerwin, J. et al. The HUDSEN Atlas: a three-dimensional (3D) spatial framework for studying gene expression in the developing human brain. *J. Anat.* **217**, 289–299 (2010).

nature research

Corresponding author(s): Timothy W. Yu

Last updated by author(s): May 2, 2019

Reporting Summary

Nature Research wishes to improve the reproducibility of the work that we publish. This form provides structure for consistency and transparency in reporting. For further information on Nature Research policies, see Authors & Referees and the Editorial Policy Checklist.

Statistics

For all statistical analyses, confirm that the following items are present in the figure legend, table legend, main text, or Methods section.

n/a	Confirmed	
☐	☒	The exact sample size (*n*) for each experimental group/condition, given as a discrete number and unit of measurement
☐	☒	A statement on whether measurements were taken from distinct samples or whether the same sample was measured repeatedly
☐	☒	The statistical test(s) used AND whether they are one- or two-sided *Only common tests should be described solely by name; describe more complex techniques in the Methods section.*
☐	☒	A description of all covariates tested
☐	☒	A description of any assumptions or corrections, such as tests of normality and adjustment for multiple comparisons
☐	☒	A full description of the statistical parameters including central tendency (e.g. means) or other basic estimates (e.g. regression coefficient) AND variation (e.g. standard deviation) or associated estimates of uncertainty (e.g. confidence intervals)
☐	☒	For null hypothesis testing, the test statistic (e.g. *F*, *t*, *r*) with confidence intervals, effect sizes, degrees of freedom and *P* value noted *Give P values as exact values whenever suitable.*
☒	☐	For Bayesian analysis, information on the choice of priors and Markov chain Monte Carlo settings
☐	☒	For hierarchical and complex designs, identification of the appropriate level for tests and full reporting of outcomes
☐	☒	Estimates of effect sizes (e.g. Cohen's *d*, Pearson's *r*), indicating how they were calculated

Our web collection on statistics for biologists contains articles on many of the points above.

Software and code

Policy information about availability of computer code

Data collection	Sequencing data was provided by the Autism Sequencing Consortium and variants were called using GATK by the individuals labs who submitted data to the ASC.
Data analysis	Beagle4.0 was used for computational phasing. VCFs were processed and annotated using VCFtools and Annovar.

For manuscripts utilizing custom algorithms or software that are central to the research but not yet described in published literature, software must be made available to editors/reviewers. We strongly encourage code deposition in a community repository (e.g. GitHub). See the Nature Research guidelines for submitting code & software for further information.

Data

Policy information about availability of data

All manuscripts must include a data availability statement. This statement should provide the following information, where applicable:

- Accession codes, unique identifiers, or web links for publicly available datasets
- A list of figures that have associated raw data
- A description of any restrictions on data availability

Data included in this manuscript is deposited at dbGAP merged with published data under accession number phs000298.v4.p3.

Field-specific reporting

Please select the one below that is the best fit for your research. If you are not sure, read the appropriate sections before making your selection.

☒ Life sciences ☐ Behavioural & social sciences ☐ Ecological, evolutionary & environmental sciences

Life sciences study design

All studies must disclose on these points even when the disclosure is negative.

Sample size	The Autism Sequencing Consortium (ASC), represents 2,343 affected and 5,852 unaffected individuals. these were collected across several independent exome sequencing studies.
Data exclusions	Data was initially passed through a standard QC filtering to remove low-quality calls prior to starting any analyses. This included removing alleles which did not have the PASS filter, those with a DP less than 10, or a GQ less than 20.
Replication	Mutational excesses observed in the entire ASC dataset were also present in each of the independent cohorts making up the ASC data collection.
Randomization	The affected status of the patients and controls was randomly assigned to confirm significance of identified mutational excesses.
Blinding	The study design was not blinded to affected status of patients.

Reporting for specific materials, systems and methods

We require information from authors about some types of materials, experimental systems and methods used in many studies. Here, indicate whether each material, system or method listed is relevant to your study. If you are not sure if a list item applies to your research, read the appropriate section before selecting a response.

Materials & experimental systems

n/a	Involved in the study
☒	☐ Antibodies
☒	☐ Eukaryotic cell lines
☒	☐ Palaeontology
☒	☐ Animals and other organisms
☒	☐ Human research participants
☒	☐ Clinical data

Methods

n/a	Involved in the study
☒	☐ ChIP-seq
☒	☐ Flow cytometry
☒	☐ MRI-based neuroimaging