**Supplementary Information**

# Exome sequencing in sporadic autism spectrum disorders identifies severe *de novo* mutations

Brian J. O'Roak[1], Pelagia Deriziotis[2], Choli Lee[1], Laura Vives[1], Jerrod J. Schwartz[1], Santhosh Girirajan[1], Emre Karakoc[1], Alexandra P. MacKenzie[1], Sarah B. Ng[1], Carl Baker[1], Mark J. Rieder[1], Deborah A. Nickerson[1], Raphael Bernier[3], Simon E. Fisher[2,4], Jay Shendure[1*], and Evan E. Eichler[1,5*]

[1]Department of Genome Sciences, University of Washington School of Medicine, Seattle, WA, USA. [2]Wellcome Trust Centre for Human Genetics, University of Oxford, Oxford, United Kingdom. [3]Department of Psychiatry and Behavioral Sciences, University of Washington, Seattle, WA, USA. [4]Language and Genetics Department, Max Planck Institute for Psycholinguistics, Nijmegen, the Netherlands. [5]Howard Hughes Medical Institute, Seattle, WA, USA.

*Corresponding authors:

Evan E. Eichler, Ph.D.
University of Washington School of Medicine
Howard Hughes Medical Institute
Foege S413C, 3720 15th Ave NE
Box 355065
Seattle, WA 98195
Phone: (206) 543-9526
E-mail: eee@gs.washington.edu

Jay Shendure, M.D., Ph.D.
Department of Genome Sciences
University of Washington
Foege S210A, 3720 15th Ave NE
Box 355065
Seattle, WA 98195
(206) 685-8543
shendure@uw.edu

Nature Genetics: doi:10.1038/ng.835

**Supplementary Note**

**Clinical data.** The 20 ASD trios sequenced were selected from the Simons Simplex Collection[1](SSC) (19) and the Study of Autism Genetics Exploration (SAGE) (1). The SSC is a Simons Foundation funded project including 12 collaborating sites across North America (see sfari.org). Each of the 12 data collection sites independently recruited families with ASD that had not previously participated in a genetics research study following approved human subjects research guidelines at each university. Participation in the collection was restricted to ASD simplex families in which both parents were willing to participate. Families in which the child with ASD had a relative (up to 3rd degree) with ASD or who had a sibling who showed ASD related symptoms, such as social challenges necessitating an individualized education program, were excluded. Inclusion criteria required that children with meet ASD diagnosis standards on the Autism Diagnostic Observation Schedule[2] (ADOS) and the Collaborative Programs for Excellence in Autism (CPEA) criteria for the Autism Diagnostic Interview, Revised[3] (ADI-R). CPEA criteria require the child to score within 2 points of the cut-off on social or communication domains or within 1 point on both, with no requirement for the repetitive behavior or onset domains. Children also need to meet a nonverbal IQ estimate cut-off of 35. Additionally, children with significant hearing, vision or motor problems, significant birth complications (e.g. extended NICU stay), or having been diagnosed with ASD related disorders, such as Fragile X, were excluded.

Participation in the SSC for the children with ASD included a diagnostic evaluation, cognitive and adaptive assessment, comprehensive examination of medical and diagnostic history, and assessment of social and communicative abilities. Height, weight, head



Nature Genetics: doi:10.1038/ng.835

circumference, and DNA via blood sample were collected from all participating family members and social communicative abilities were also assessed in the parents and siblings of the children with ASD. Data collection, entry, and validation methods were standardized across sites to ensure reliability of sample collection. Each institution applied for and received approval from the institution's human subjects division. As required by each local institutional review board, all subjects provided consent to participate in the collection.

Recruited families participating in the SAGE study range in age from 24 months old to adulthood and meet diagnostic criteria for ASD with no other factors likely to contribute to the etiology (such as very low birth weight, other genetic conditions) or diagnostic criteria for non-ASD developmental delay. Families are recruited from clinic patients undergoing a diagnostic evaluation for autism at Seattle Children's Autism Center. Following recruitment and enrollment, DNA samples and family history information are collected and diagnoses are confirmed through record review. As reported by parent informant, ~75% are simplex families. For those families with identified pathogenic event, a comprehensive phenotypic workup is conducted that includes: medical and family history; treatment history; diagnostic workup; cognitive testing; language, adaptive, and motor skills assessment; 2D photos of face & hands; and completion of parent questionnaires re: broader autism phenotype, general psychological functioning, aberrant behaviors, social symptoms, and demographic information. The SAGE study received approval from the Seattle Children's Human Subjects Division and all subjects provided consent to participate in the collection.

DNA samples were de-identified prior to distribution. The use of non-identifiable biological materials in this study was deemed not human subjects research by the University of Washington Human Subjects Division.



Nature Genetics: doi:10.1038/ng.835

**Detailed Clinical Information.** NOTE: Standardized instruments included (mean 100, standard deviation is 15): cognitive: Differential Ability Scales-2nd edition, receptive language: Peabody Picture Vocabulary Test-4th edition, and adaptive skills: Vineland Adaptive Behavior Scales-2nd edition. Calibrated Severity Score is on 1-10 scale (>4 clinical; 10 most impaired). Abbreviations: Autism Diagnostic Interview (ADI); Autism Diagnostic Observation Schedule (ADOS); Intellectual Quotient (IQ).

**ID #: SSC 12817**

**Mutation: *FOXP1* frameshift**

**Demographics:** 9:5 male, White, non-Hispanic.

**Family and Developmental History:** No family history of ASD. Proband is 3rd of 3 siblings (1 older brother, 1 older sister). Paternal side: No significant psychiatric history. Maternal side: no significant psychiatric history. Pregnancy and birth: proband was 3rd pregnancy and delivery. Vaginal delivery at 41 weeks. Augmented labor. No anesthesia. No pregnancy or birth complications. No history of Rubella, Valproate use, Infections, Trauma, or use of Artificial Reproductive Technology.

**Medical History:** Likelihood of non-febrile seizures in proband (and paternal cousin).

**Paternal/Maternal age** in years at conception: 40/35.

**Cognitive:** FullScale IQ: 36 (<1st percentile); Nonverbal IQ: 34 (<1st percentile); Verbal IQ: 37 (<1st percentile); single word use.

**Receptive Language:** <1st percentile.

**Adaptive Behavior:** Daily Living Skills: 2nd percentile; Communication: 1st percentile; Social-Emotional: 2nd percentile; Overall: 1st percentile.



Nature Genetics: doi:10.1038/ng.835

**Diagnostic:** Meets research criteria for Autistic Disorder: (ADI +; ADOS +; clinical diagnosis +). Calibrated Severity Score = 8. Evidence of onset prior to 3 years of age. Regression (word loss). Language delay (phrases delayed).

**Aberrant Behaviors**: Elevated for lethargy, hyperactivity, and inappropriate speech.

**ID #: SSC 12681**

**Mutation: *GRIN2B* 3' splice**

**Demographics:** 6:6 female, White, non-Hispanic.

**Family and Developmental History:** No family history of ASD. Proband is 3rd of 3 siblings (2 older sisters). Paternal side: No significant psychiatric history. Maternal side: no significant psychiatric history. Pregnancy and birth: proband was 3rd pregnancy and delivery. Upper respiratory infection and allergies reported during first trimester. Vaginal delivery at 41 weeks. Augmented labor. No anesthesia. Nuchal cord. No other pregnancy or birth complications. No history of Rubella, Valproate use, Trauma, or use of Artificial Reproductive Technology.

**Medical History:** Nothing of note.

**Paternal/Maternal age** in years at conception: 33/31.

**Cognitive:** FullScale IQ: 63 (1st percentile); Nonverbal IQ: 65 (1st percentile); Verbal IQ: 69 (2nd percentile); fluent language use.

**Receptive Language:** 21st percentile.

**Adaptive Behavior:** Daily Living Skills: 13th percentile; Communication: 5th percentile; Social-Emotional: 18th percentile; Motor: 7th percentile; Overall: 6th percentile.

**Fine Motor Ability:** Greater than 2 standard deviations below the mean.



Nature Genetics: doi:10.1038/ng.835

**Diagnostic:** Meets research criteria for Autistic Disorder: (ADI +; ADOS +; clinical diagnosis +). Calibrated Severity Score = 9. Evidence of onset prior to 3 years of age. Possible regression. Language delay (single word & phrases delayed).

**Aberrant Behaviors**: Elevated for hyperactivity.

**ID #: SSC 12499**

**Mutation: *SCN1A* PRO1894LEU**

**Demographics:** 6:11 male, White, non-Hispanic.

**Family and Developmental History:** No family history of ASD. Proband is 2nd of 2 siblings (1 older brother). Paternal side: cerebral palsy (pat cousin). Maternal side: migraines (mother, mat grandparent), lymphangioleicmyomatosis (lavi) (mat aunt/uncle). No significant psychiatric history. Pregnancy and birth: proband was 2nd pregnancy and delivery. Upper respiratory infection in trimester 2. Vaginal delivery at 40 weeks. No anesthesia. No pregnancy or birth complications. No history of Rubella, Valproate use, Trauma, or use of Artificial Reproductive Technology.

**Medical History:** Definite non-febrile seizures, dx of epilepsy.

**Paternal/Maternal age** in years at conception: 45/31.

**Cognitive:** FullScale IQ: 60 (<1st percentile); Nonverbal IQ: 57 (<1st percentile); Verbal IQ: 56 (<1st percentile); phrase speech use.

**Receptive Language:** 13th percentile.

**Adaptive Behavior:** Daily Living Skills: 8th percentile; Communication: 13th percentile; Social-Emotional: 6th percentile; Motor: 10th percentile; Overall: 76th percentile.



Nature Genetics: doi:10.1038/ng.835

**Diagnostic:** Meets research criteria for Autistic Disorder: (ADI +; ADOS +; clinical diagnosis +). Calibrated Severity Score = 8. Evidence of onset prior to 3 years of age. Possible regression. Language delay (single word & phrases delayed).

**Aberrant Behaviors**: No elevations.

**ID #: SSC 11666.p1**

**Mutation: *LAMC3* ASP339GLY**

**Demographics:** 7:9 male; White, non-Hispanic.

**Family and Developmental History:** no family history of ASD. Proband is 2nd of 2 children (1 older brother). Maternal side: migraines (mother); speech delay requiring therapy (maternal cousin); asthma (maternal cousin). Paternal side: eating disorder (paternal aunt, paternal grandmother), eczema (paternal grandparent). Pregnancy and birth: proband was 3rd pregnancy and 2nd delivery. Vaginal delivery at 40 weeks. Induced labor (failure to progress) via use of IV oxytocin. Anesthesia used. No pregnancy or birth complications. No history of Rubella, Valproate use, Infections, Trauma, or use of Artificial Reproductive Technology.

**Medical History:** Otitis Media at age 4 (treated). No other medical complications.

**Paternal/Maternal age** in years at conception: 32 years/30 years.

**Cognitive:** FullScale IQ: 49 (<1st percentile); Nonverbal IQ: 64 (1st percentile); Verbal IQ: 51 (<1st percentile); fluent language use.

**Receptive Language:** <1st percentile.

**Adaptive Behavior:** Daily Living Skills: 1st percentile; Communication: 2nd percentile; Social-Emotional: 2nd percentile; Overall: 1st percentile.

**Fine Motor Ability:** Greater than 2 standard deviations below the mean.



Nature Genetics: doi:10.1038/ng.835

**Diagnostic:** Meets research criteria for Autistic Disorder: (ADI +; ADOS +; clinical diagnosis +). Calibrated Severity Score = 10. Evidence of onset prior to 3 years of age. No regression.

**Aberrant Behaviors**: Elevated for Hyperactivity and Irritability; Elevated for Stereotyped and Restricted behaviors.

**Functional characterization of FOXP1 mutation.** To assess whether the *FOXP1* frameshift mutation is targeted by nonsense-mediated decay (NMD), we grew two parallel cultures of immortalized lymphoblasts derived from the family, one of which was chemically treated to inhibit NMD. In the untreated sample we see low expression of the *FOXP1* insertion allele, while in the treated sample we see roughly equal levels of the frameshift mutation and the normal allele, suggesting that most of the mutated transcript is degraded (**Supplementary Fig. 5a**). Residual mutated p.A339SfxX4 transcripts are predicted to yield a C-terminally truncated protein that lacks the characteristic forkhead-box DNA-binding domain. We analyzed the functional properties of this FOXP1 mutant protein (FOXP1mut), making comparisons to wild-type FOXP1, FOXP2, and previously studied FOXP2 variants, using transfected HEK293T cell lines (**Supplementary Fig. 5b-d**). FOXP1mut displays aberrant localization to the cytoplasm as opposed to the nucleus—similar to results obtained with FOXP2 mutations[4]. This disruption of nuclear targeting is most likely due to the loss of nuclear localization signals flanking the DNA-binding domain[4]. In contrast to a truncated version of FOXP2 previously associated with verbal dyspraxia[4,5], the FOXP1mut product appears stable.

HEK293T cell-based assays indicated that, as for FOXP2[6], increased levels of wild-type FOXP1 yield significantly reduced expression of *CNTNAP2* (p=0.0005) (**Supplementary Fig. 5d**). Intriguingly, in these experiments expression of the FOXP1mut protein was instead associated with a significant three-fold increase in *CNTNAP2* expression relative to control cells



Nature Genetics: doi:10.1038/ng.835

(p=0.0056). These data suggest that the aberrant FOXP1mut protein can interfere with the action of endogenous FOXP transcription factors present in the HEK293T cells[4], causing a further misregulation of *CNTNAP2* expression. Overall, we hypothesize that reduced dosage of wild-type FOXP1 transcripts (due to NMD of mutant transcripts), combined with dysfunction of FOXP1mut proteins that escape this process, may yield overexpression of CNTNAP2 proteins, amplifying any deleterious effects of H275A in the proband. Negligible expression of *CNTNAP2* in our only available tissue (immortalized lymphoblasts) precludes direct testing of this hypothesis[7].

**Comparison to Vissers *et al.*** Vissers and colleagues reported a similar study on sporadic moderate to severe ID using Agilent/SOLiD technologies[8]. In contrast to our study, they reported putatively causative events (one of which was inherited) in 6/10 families suggesting that coding *de novo* events of major effect may be more prevalent in ID than ASD. Using an analogous approach, they reported that the predicted disease-associated events occurred at highly conserved sites and had disruptive Grantham scores. While the two approaches are similar, we did observed subtle differences in the mutation rates and characteristics. For example, although we both report ~1 coding *de novo* event per trio and sufficient coverage of ~90% of target, their set was larger (37 Mb versus 26 Mb). Based on our observed rate we would expect ~1.3 events per trio from a 37 Mb exome capture (versus the ~1.0 events per trio reported in their study). Based on their reported rate, we would expect ~0.7 events per trio from a 26 Mb capture (versus the ~0.9 events per trio observed in our study). In addition, 9/16 (56%) of the *de novo* coding substitutions we observed were nonsynonymous, while they reported 8/9 (89%) were nonsynonymous. We also observed markedly different transition/transversion ratios: 18:2 in this study versus 6:3 in the Vissers *et al.* study. However, none of these differences are statistically significant, and given the



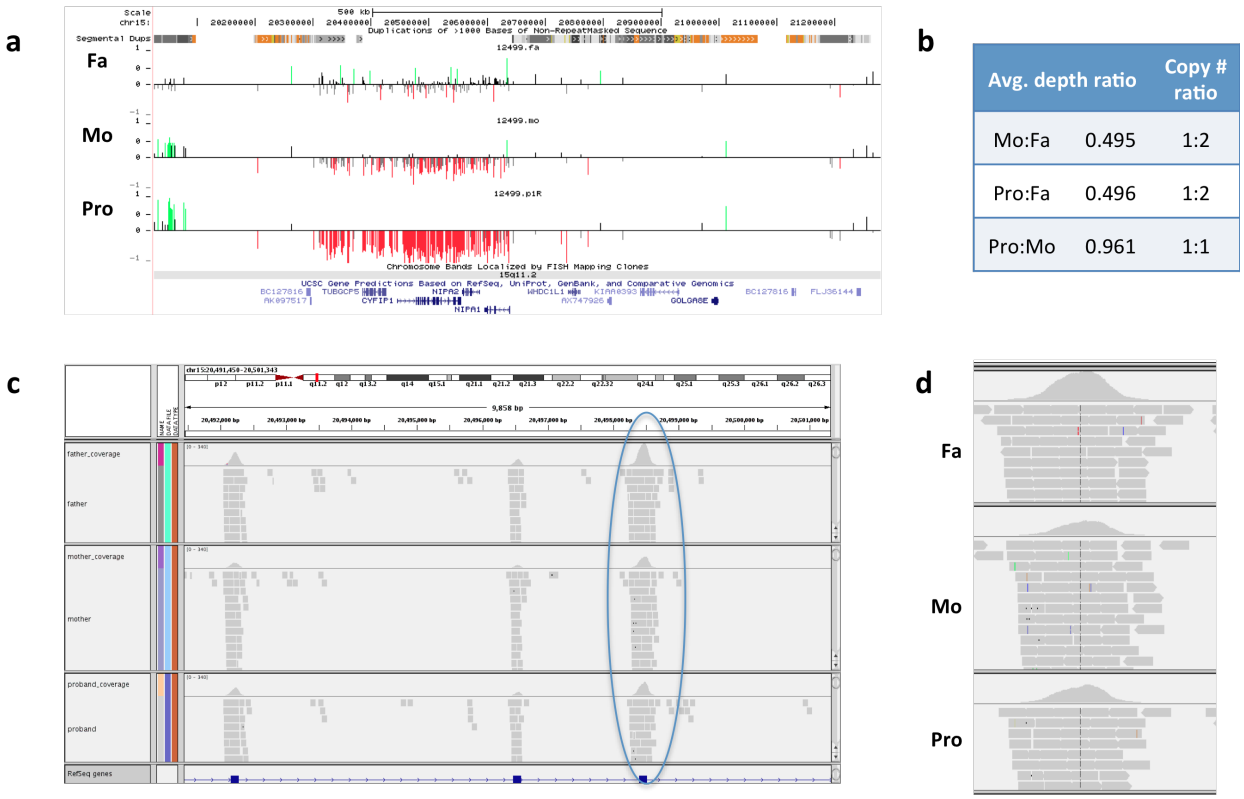
Nature Genetics: doi:10.1038/ng.835

relatively small number of observations at this time, it is unknown whether they reflect any true differences in either capture/sequencing/analysis methodology or in ID versus ASD patient populations.

**Description of other conserved/damaging variants.** The remaining *de novo* mutation sites were not considered to be strong candidates for involvement in ASD, although it is entirely possible that sequencing of additional ASD cases may further implicate one or more of them. Missense mutations affecting conserved nucleotides and also predicted to be damaging based on Grantham scores occurred in *SYNE1*, *SLC30A5*, *TLK2*, and *RBM3*. SYNE1 (p.Y282C, CCDS5236.1) is a spectrin repeat containing protein expressed in skeletal and smooth muscle, an extremely large transcript, and previously associated with spinocerebellar ataxia, autosomal recessive 8, and emery-dreifruss muscular dystrophy 4. SLC30A5 (p.S561R, CCDS3996.1) is a zinc solute carrier that functions in the pancreas. TLK2 (p.S595L, CCDS11633.1) is a serine/threonine kinases potentially involved in the regulation of chromatin assembly. RBMS3 (p.T383M, CCDS33724.1) is an RNA-binding protein and potential regulator of hepatic stellate cells. The three remaining missense variants affect sites that are neither highly conserved nor predicted to be disruptive from the Grantham score. TGM3 (p.V144I, CCDS33435.1) is a transglutaminase involved in hair follicle development. GPR139 (p.S151G, CCDS32398.1) is a G-coupled receptor, which is expressed in the brain. XIRP1 (p.V483M, CCDS2683.1) is the Xin actin-binding repeat-containing protein 1, associated with cardiomyopathy, and likely functions to protect actin filaments from depolymerization. The only synonymous and untranslated region (UTR) sites that occur at conserved nucleotides are at ARHGAP15 and MYO1A, respectively. ARHGAP15 is an RHO GTPase-activating protein, weakly expressed in brain. *MYO1A* is a myosin superfamily gene, associated with autosomal dominant deafness.



Nature Genetics: doi:10.1038/ng.835

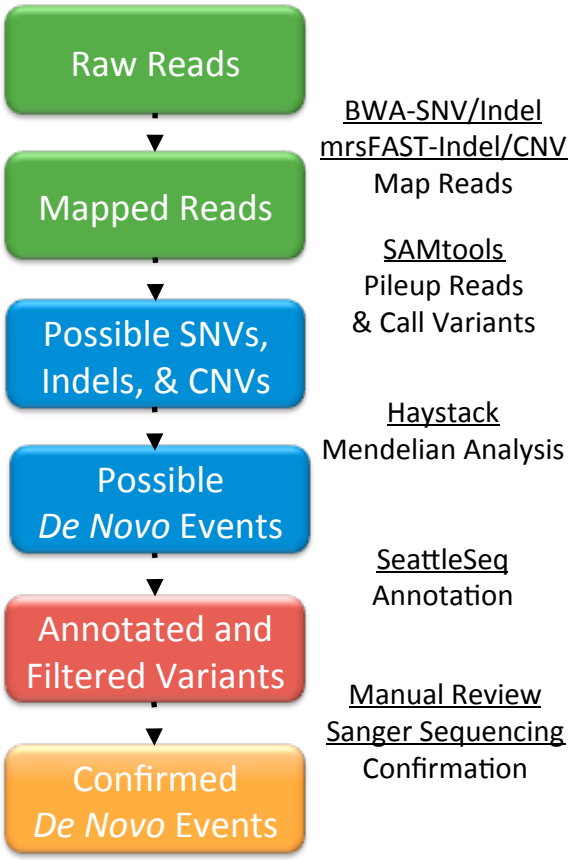



| Avg. depth ratio | | Copy # ratio |
|---|---|---|
| Mo:Fa | 0.495 | 1:2 |
| Pro:Fa | 0.496 | 1:2 |
| Pro:Mo | 0.961 | 1:1 |

**Supplementary Figure 1** Maternally inherited large CNV at 15q11.2 in family 12499. **a,** Genome browser diagram showing raw array CGH results indicating that the mother and proband are heterozygous deletion carriers. The deletion spans from approximately chr15:20,300,028-20,647,960, including *TUBGCP5*, *CYFIP1*, *NIPA2*, and *NIPA1*. Red indicates significantly deviated probes. **b,** Average exome read depth across the interval shows the expected relative copy numbers of the trio. **c,** IGV browser[9] view shows drop in read depth across multiple exons. **d,** Expanded view of exon highlighted in **c**.



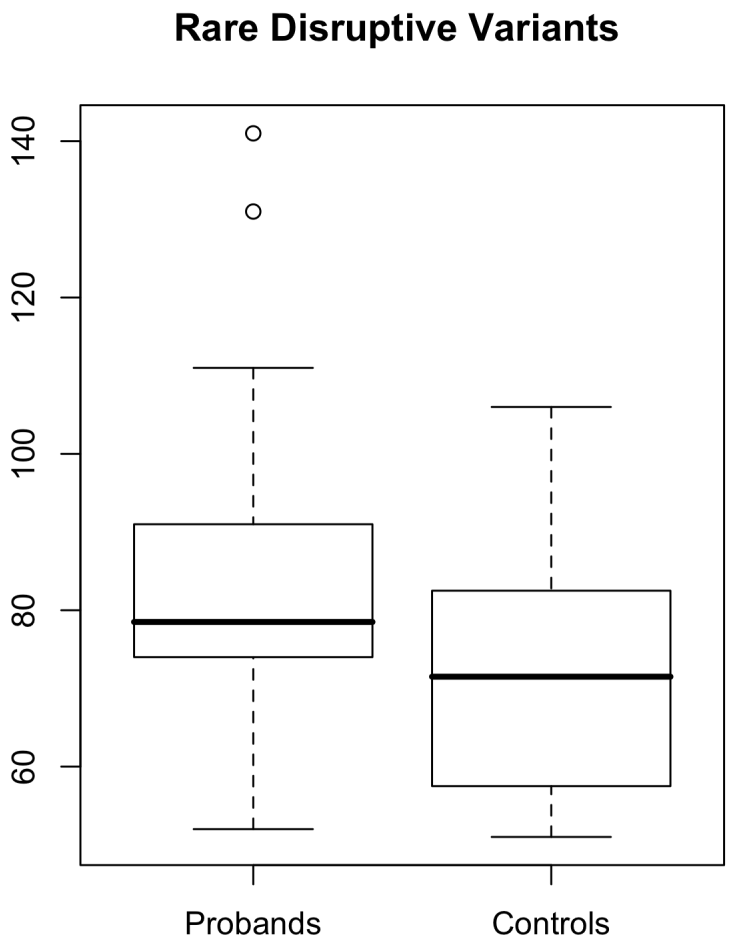
Nature Genetics: doi:10.1038/ng.835




**Supplementary Figure 2** Flow diagram of sequence analysis pipeline. First, raw reads are mapped to the reference genome using with *BWA* for SNV/small indel detection and *mrsFAST* for indel/CNV detection. Second, discordant read pairs and duplicates are removed, followed by genotype calling using *SAMtools*. Third, high quality variants are then run through a custom pipeline, *Haystack*, to evaluate their inheritance. Fourth, variant positions flagged as potentially *de novo* are then filtered against other sequenced exomes to remove systematic artifacts and annotated using the SeattleSeq server. Lastly, novel variants (e.g. not called in dbSNP, the 1000 Genomes Project Pilot, or 1490 other exomes) are then visually inspected to remove variants with >10% variant alleles in one or both parents and the remainder subjected to bi-directional Sanger sequencing.



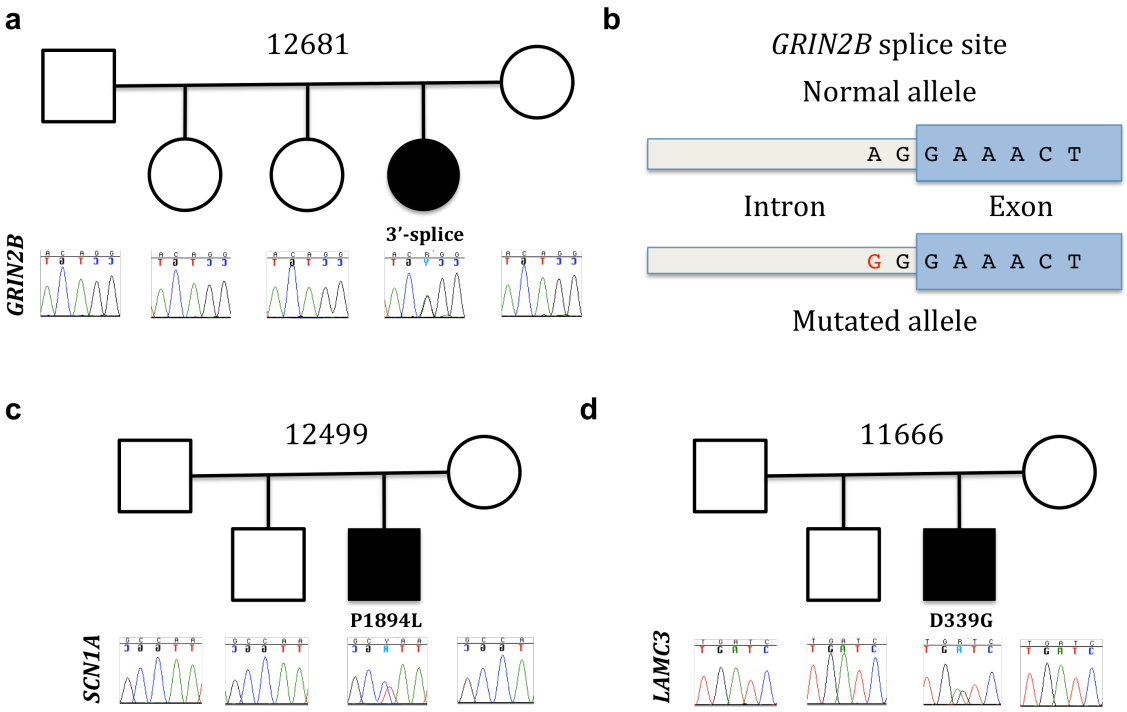
Nature Genetics: doi:10.1038/ng.835




**Supplementary Figure 3** Boxplots showing number of "private" rare protein disruptive variants identified in the ASD proband and HapMap controls.



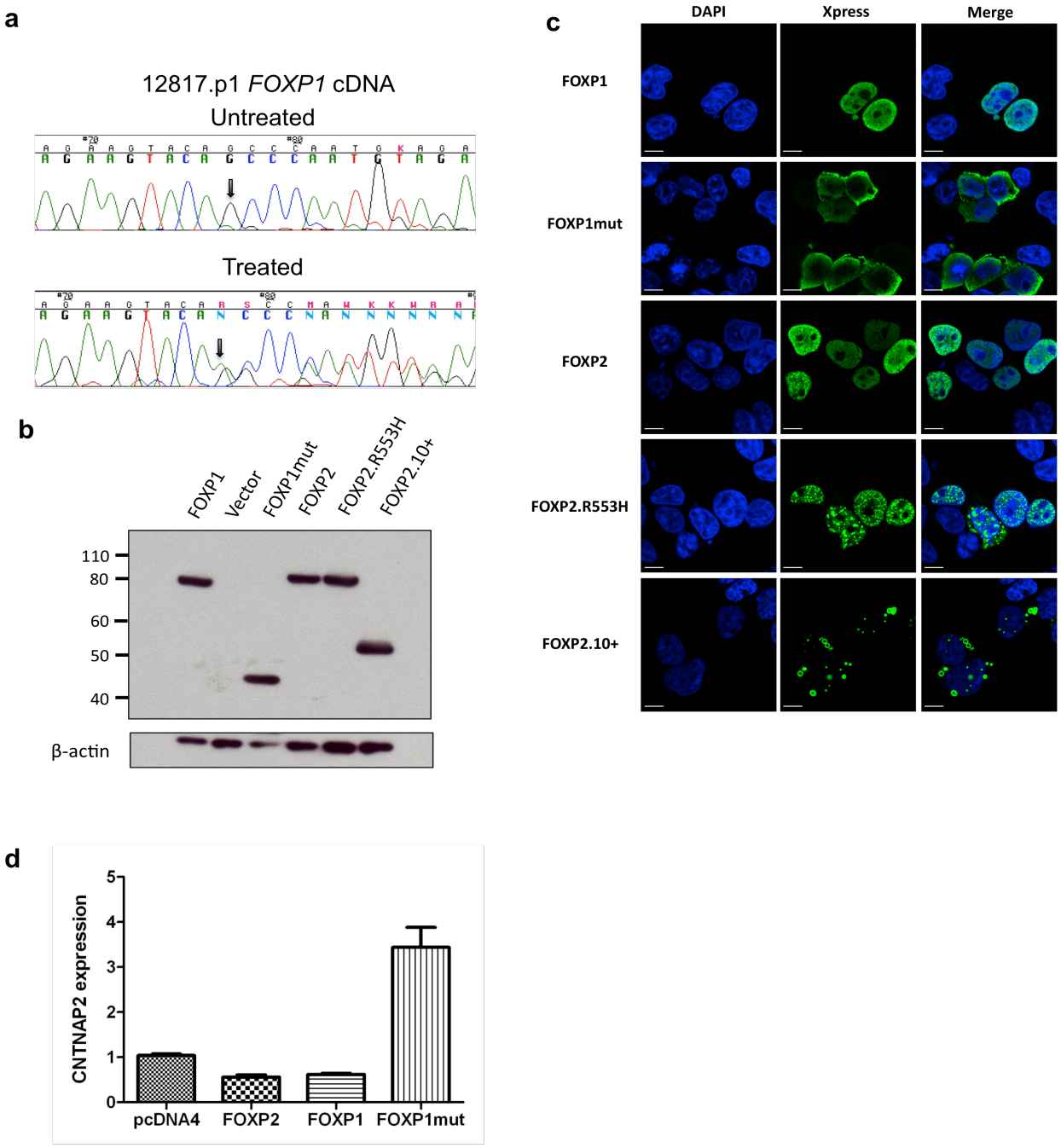
Nature Genetics: doi:10.1038/ng.835




**Supplementary Figure 4** SNV *de novo* mutation events that are potentially causative. **a,** Pedigree showing chromatogram traces surrounding *GRIN2B* 3' splice mutation. **b,** Diagram showing the affected *GRIN2B* intron-exon boundary. The A->G mutation results in loss of the canonical 3' AG splice site. **c,** Pedigree showing chromatogram traces surrounding *SCN1A* p.P1894L missense mutation. **d,** Pedigree showing chromatogram traces surrounding *LAMC3* p.D339G missense mutation.



Nature Genetics: doi:10.1038/ng.835




**Supplementary Figure 5** Functional characterization of *FOXP1 de novo* mutation. **a,** Sanger traces of PCR amplified cDNA from 12817.p1 lymphoblasts. Untreated cells show low levels of the mutant allele p.A339SfsX4. Cells treated to inhibit NMD show approximately equal levels of



Nature Genetics: doi:10.1038/ng.835

the normal and mutation allele. Arrow indicates insertion site. **b-d,** HEK293T cell-based expression of FOXP1, FOXP1mut-p.A339SfsX4, FOXP2, FOXP2-p.R553H, and the alternatively spliced FOXP2-10+ isoform. All transcripts were cloned in a pcDNA4/HisMax expression vector, in frame with its N-terminal Xpress™ tag. **b,** Western blotting of HEK293T transfected whole-cell extracts using an antibody against the N-terminal Xpress™ tag demonstrated the presence of recombinant proteins around the predicted molecular weight for all constructs. The FOXP1mut construct yielded a truncated product of ~45 kDa, approximately half the size of the wild-type protein (FOXP1, ~75 kDa). Levels of the truncated product were similar to those of wild-type FOXP1. The p.R553H variant of FOXP2, an etiological mutation causing severe speech and language deficits in a large multigenerational family[10], was of similar molecular weight to wild-type FOXP2. The 10+ variant represents an alternatively spliced version of FOXP2 encoded by a putative mRNA transcript that contains a polyadenylation site in the intron following exon 10 and thus excludes exons 11-17; as expected this yielded a product of ~50 kDa[4]. Equivalent loading across samples was confirmed using a beta-actin internal loading control. **c,** Confocal images of HEK293T cells transiently transfected with expression constructs followed by immunofluorescence with an antibody to the Xpress™ tag (green). DAPI counterstain (blue) depicts the location of nuclei. Both wild-type FOXP1 and FOXP2 localize predominantly to the nucleus and, in line with previously reported data[4], are generally excluded from nucleoli. In contrast, the FOXP1mut protein displays predominantly cytoplasmic localization. The FOXP2.R553H protein shows both cytoplasmic and nuclear localization. In some cells it appears to form small aggregates. The FOXP2.10+ isoform is predominantly localized to the cytoplasm and forms cytoplasmic aggregations, which have been suggested to represent aggresomes[4]. Scale bar, 10 μm. **e,** Quantitative RT-PCR results for cDNA prepared



Nature Genetics: doi:10.1038/ng.835

from transiently transfected HEK293T cells. *CNTNAP2* expression levels (y-axis) are the means of three independent cDNA experiments and are normalized with an internal control, *GAPDH*. FOXP1 significantly repressed *CNTNAP2* expression relative to the empty vector (pcDNA4), two-way p=0.0005 (**Supplementary Table 8**). Overexpression of FOXP1mut resulted in an approximately three-fold increase in *CNTNAP2* expression relative to the empty vector, two-way p=0.0056, suggesting that expression of the mutant FOXP1 protein may lead to amplification of any potentially damaging effects of the CNTNAP2 H275A allele. Error bars indicate ±SEM.

**Supplementary Table 1 Core descriptive clinical values on ASD probands**
Attached file: oroak_supplementary_table1.xlsx


Nature Genetics: doi:10.1038/ng.835

**Supplementary Table 2 Exome genotype and Illumina 1M/1MDuo concordance and heterozygous detection rate**

| | All Overlapping SNP Sites | | | Heterozygous Overlapping SNP Sites | | |
|---|---|---|---|---|---|---|
| **Individual** | **All Concordant** | **All Sites** | **# of All Sites** | **All Detected Het** | **All Het Sites** | **# of Het Sites** |
| 11048.fa | 34209 | 34293 | 0.997550521 | 5951 | 5971 | 0.996650477 |
| 11048.mo | 34070 | 34167 | 0.997161003 | 5913 | 5932 | 0.996797033 |
| 11048.p1 | 32761 | 32854 | 0.997169294 | 5556 | 5575 | 0.996591928 |
| 11307.fa | 33528 | 33606 | 0.997678986 | 5701 | 5713 | 0.997899527 |
| 11307.mo | 33411 | 33498 | 0.99740283 | 5909 | 5925 | 0.997299578 |
| 11307.p1 | 33425 | 33521 | 0.997136124 | 5647 | 5669 | 0.996119245 |
| 11580.fa | 32976 | 33095 | 0.996404291 | 5745 | 5782 | 0.99360083 |
| 11580.mo | 33488 | 33590 | 0.996963382 | 5776 | 5794 | 0.996893338 |
| 11666.fa | 33389 | 33491 | 0.996954406 | 5647 | 5677 | 0.994715519 |
| 11666.mo | 34267 | 34363 | 0.997206297 | 6018 | 6041 | 0.996192683 |
| 11666.p1 | 34488 | 34562 | 0.99785892 | 5901 | 5922 | 0.996453901 |
| 12325.fa | 33994 | 34074 | 0.997652169 | 5726 | 5747 | 0.99634592 |
| 12325.mo | 33900 | 33987 | 0.997440198 | 5897 | 5916 | 0.996788371 |
| 12325.p1 | 33914 | 34013 | 0.997089348 | 5684 | 5710 | 0.995446585 |
| 12499.fa | 34742 | 34821 | 0.997731254 | 6090 | 6114 | 0.996074583 |
| 12499.mo | 34744 | 34820 | 0.997817346 | 5933 | 5947 | 0.997645872 |
| 12499.p1 | 34650 | 34724 | 0.997868909 | 5862 | 5875 | 0.997787234 |
| 12575.fa | 34021 | 34096 | 0.997800328 | 5892 | 5909 | 0.997123033 |
| 12575.mo | 34771 | 34838 | 0.998076813 | 6051 | 6066 | 0.997527201 |
| 12575.p1 | 34504 | 34577 | 0.99788877 | 5923 | 5932 | 0.998482805 |
| 12647.fa | 34619 | 34684 | 0.998125937 | 5916 | 5922 | 0.998986829 |
| 12647.mo | 34751 | 34836 | 0.997559995 | 6161 | 6170 | 0.998541329 |
| 12647.p1 | 34019 | 34122 | 0.99698142 | 5926 | 5946 | 0.996636394 |
| 12681.fa | 34540 | 34623 | 0.99760275 | 5961 | 5984 | 0.996156417 |
| 12681.mo | 34744 | 34816 | 0.997931985 | 6279 | 6289 | 0.998409922 |
| 12681.p1 | 34115 | 34192 | 0.997748011 | 6092 | 6108 | 0.997380485 |
| 12817.fa | 34532 | 34600 | 0.998034682 | 5979 | 5984 | 0.999164439 |
| 12817.mo | 34772 | 34831 | 0.998306107 | 6172 | 6180 | 0.998705502 |
| 12817.p1 | 34313 | 34391 | 0.997731965 | 5944 | 5959 | 0.997482799 |
| 13095.fa | 34392 | 34473 | 0.997650335 | 6095 | 6117 | 0.996403466 |
| 13095.mo | 34123 | 34203 | 0.997661024 | 5958 | 5973 | 0.997488699 |
| 13095.p1 | 34452 | 34529 | 0.99776999 | 6109 | 6123 | 0.997713539 |
| | | **AVG** | 0.997561106 | | **AVG** | 0.997047046 |



Nature Genetics: doi:10.1038/ng.835

**Supplementary Table 3 Proband sites covered in each trio**

| Proband | Trio Covered | Total Sites | Total % |
|---|---|---|---|
| 11048.p1 | 13626 | 14095 | 0.966725789 |
| 11307.p1 | 12966 | 13509 | 0.959804575 |
| 11580.p1 | 13079 | 13912 | 0.940123634 |
| 11666.p1 | 13008 | 14306 | 0.909268838 |
| 12325.p1 | 13182 | 13866 | 0.950670705 |
| 12499.p1 | 14089 | 14479 | 0.973064438 |
| 12575.p1 | 13703 | 14568 | 0.940623284 |
| 12647.p1 | 13827 | 14144 | 0.97758767 |
| 12680.p1 | 13681 | 14124 | 0.968634948 |
| 12681.p1 | 14290 | 14750 | 0.968813559 |
| 12817.p1 | 13931 | 14364 | 0.969855194 |
| 12974.p1 | 13466 | 13990 | 0.962544675 |
| 13095.p1 | 13995 | 14605 | 0.958233482 |
| 13253.p1 | 13158 | 13775 | 0.955208711 |
| 13284.p1* | 16855 | 17806 | 0.946591037 |
| 13466.p1 | 13542 | 14023 | 0.965699208 |
| 13683.p1 | 13314 | 14419 | 0.923365005 |
| 13708.p1 | 13475 | 13997 | 0.962706294 |
| 13970.p1 | 13678 | 14293 | 0.956971944 |
| SAGE4022.p1 | 13772 | 14538 | 0.947310497 |

***13284 Included Additional RefSeq Targets**

**Supplementary Table 4 Observed and expected *de novo* coding mutation rates**

| | Total | Avg | Expected* |
|---|---|---|---|
| All Coding | 18 | 0.9 | 0.591 |
| Synon | 7 | 0.35 | 0.14 |
| Missense | 9 | 0.45 | 0.41 |
| Nonsense | 0 | NA | 0.022 |
| Splice | 1 | 0.05 | 0.013 |
| Indels | 1 | 0.05 | 0.035 |

***Based on a haploid mutation rate of substitution rate of 1.28E-08, deletion rate of 5.80E-10, and insertion rate of 2.00E-10, 22,546,796 coding and 523,843 splice bases covered in an average trio**



Nature Genetics: doi:10.1038/ng.835

**Supplementary Table 5 Variant positions of 21 genes with identified *de novo* events from 1000 genomes[9] pilot data, 20 HapMap, and 20 ASD probands**

Attached file: oroak_supplementary_table5.xlsx



Nature Genetics: doi:10.1038/ng.835

## Supplementary Table 6 Rare protein disruptive variants intersecting with SFARI gene list

**Variants Intersecting Genes Unique to Cases (*de novo* events excluded)**

| ID | Type | Chromosome:Position | Gene Symbol | Variant | AA Change | GERP Score | Grantham Score |
|---|---|---|---|---|---|---|---|
| 11048.p1 | missense | chr16:87874950 | ANKRD11 | R | P1834L | 2.75 | 98 |
| 12680.p1 | missense | chr16:87878688 | ANKRD11 | R | S588L | 5.16 | 145 |
| 12325.p1 | frameshift | chr3:53817801 | CACNA1D | +C | | | |
| 11307.p1 | missense | chr20:44303060 | CDH22 | R | R167C | 3.85 | 180 |
| 11580.p1 | missense | chr5:26938483 | CDH9 | K | K371T | 5.19 | 78 |
| 12817.p1 | missense | chr7:146449073 | CNTNAP2 | R | H275R | 5.47 | 29 |
| 13253.p1 | missense | chr11:522746 | HRAS | Y | D154N | 3.81 | 23 |
| 13284.p1 | missense | chr6:114485886 | HS3ST5 | W | L90H | 5.57 | 99 |
| 13683.p1 | missense | chr2:100908186 | NPAS2 | R | N60S | 5.1 | 46 |
| 12325.p1 | missense | chr7:107608049 | NRCAM | K | K886T | 5.08 | 78 |
| 12499.p1 | missense | chr15:86481659 | NTRK3 | Y | R201H | 4.23 | 29 |
| 13095.p1 | frameshift | chr4:134291098 | PCDH10 | -C | | | |
| 12647.p1 | missense | chr1:40867593 | RIMS3 | S | L171V | 4.14 | 32 |
| 11580.p1 | missense | chr6:166782242 | RPS6KA2 | Y | R439Q | 3.93 | 43 |
| 13683.p1 | missense | chr4:53468386 | SCFD2 | Y | M613V | 5.36 | 21 |
| 13466.p1 | missense | chr11:70011071 | SHANK2 | R | V397A | 4.52 | 64 |
| 12680.p1 | missense | chr15:23201720 | UBE3A | Y | D18N* | 3.49 | 23 |
| 13970.p1 | missense | chr15:23201720 | UBE3A | Y | D18N* | 3.49 | 23 |

***Variant observed in two probands**

**Variants Intersecting Genes Unique to Controls**

| ID | Type | Chromosome:Position | Gene Symbol | Variant | AA Change | GERP Score | Grantham Score |
|---|---|---|---|---|---|---|---|
| 1842 | missense | chr5:112202615 | APC | R | N1142S | 5.41 | 46 |
| 1846 | missense | chr7:117219988 | CTTNBP2 | S | K166N | 2.45 | 94 |
| 1798 | missense | chrX:32276529 | DMD | R | S1788L | 5.14 | 145 |
| 1843 | missense | chrX:12638530 | FRMPD4 | Y | S521L | 5.12 | 145 |
| 1805 | missense | chr15:24568644 | GABRB3 | Y | M80V | 4.22 | 21 |
| 1798 | missense | chr2:154705212 | GALNT13 | R | A87T | 5.78 | 58 |
| 1804 | missense | chr6:102240920 | GRIK2 | Y | T317M | 5.49 | 81 |
| 1838 | frameshift | chr6:30018712 | HLA-A | +C | | | |
| 1796 | missense | chr7:107389291 | LAMB1 | S | R642G | 4.8 | 125 |
| 1838 | missense | chr10:102756781 | LZTS2 | S | R626G | 3.59 | 125 |
| 1804 | missense | chr1:160593539 | NOS1AP | K | A310S | 5.14 | 99 |
| 1804 | missense | chr17:7985215 | PER1 | S | G1257R | 3.71 | 125 |
| 1798 | missense | chrX:107217677 | PSMD10 | R | T141I | 4.86 | 89 |
| 1840 | missense | chr2:166609996 | SCN1A | M | L489V | 2.84 | 32 |
| 1804 | missense | chr16:19102430 | SYT17 | S | E137D | 2.92 | 45 |

**Variants Intersecting Genes in Both Cases & Controls**

| ID | Type | Chromosome:Position | Gene Symbol | Variant | AA Change | GERP Score | Grantham Score |
|---|---|---|---|---|---|---|---|
| 13708.p1 | missense | chr8:143955163 | CYP11B1 | R | R24C | 1.58 | 180 |
| 1847 | missense | chr8:143958147 | CYP11B1 | Y | A29T | -5.38 | 58 |
| 12575.p1 | missense | chr1:97811940 | DPYD | Y | I435V | 5.34 | 29 |
| 1841 | missense | chr1:97753928 | DPYD | M | R561L | 4.95 | 102 |
| 12575.p1 | missense | chr17:17641687 | RAI1 | Y | P1567L | 4.79 | 98 |
| 1796 | missense | chr17:17639511 | RAI1 | Y | R842W | 3.6 | 101 |
| 1803 | missense | chr17:17639587 | RAI1 | Y | A867V | 4.72 | 64 |
| 12647.p1 | missense | chr7:102970438 | RELN | Y | R2216Q | 5.47 | 43 |
| 1800 | missense | chr7:102949739 | RELN | R | A2545V | 5.37 | 64 |
| 13970.p1 | missense | chr16:2078053 | TSC2 | Y | M1691T | 4.3 | 81 |
| 1804 | missense | chr16:2078136 | TSC2 | R | A1719T | 4.16 | 58 |



Nature Genetics: doi:10.1038/ng.835

## Supplementary Table 7 Multiple mutations affecting a single proband

| Proband | Type* | Source | Chromosome:Position | Gene Symbol | Variant | AA Change | GERP Score | Grantham Score |
|---|---|---|---|---|---|---|---|---|
| 11580.p1 | missense | inherited | chr5:26938483 | CDH9 | K | K371T | 5.19 | 78 |
| 11580.p1 | missense | inherited | chr6:166782242 | RPS6KA2 | Y | R439Q | 3.93 | 43 |
| 12325.p1 | frameshift | inherited | chr3:53817801 | CACNA1D | +C | | | |
| 12325.p1 | missense | inherited | chr7:107608049 | NRCAM | K | K886T | 5.08 | 78 |
| 12499.p1 | missense | inherited | chr15:86481659 | NTRK3 | Y | R201H | 4.23 | 29 |
| 12499.p1 | missense | *de novo* | chr2:166556317 | SCN1A | R | P1894L | 5.55 | 98 |
| 12499.p1 | missense | *de novo* | chr6:152865504 | SYNE1 | Y | Y282C | 4.48 | 194 |
| 12499.p1 | CNV | inherited | chr15:20300028-20647960 | TUBGCP5, CYFIP1, NIPA2, NIPA1 | DEL | | | |
| 12680.p1 | missense | inherited | chr15:23201720 | UBE3A | Y | D18N | 3.49 | 23 |
| 12680.p1 | missense | inherited | chr16:87878688 | ANKRD11 | R | S588L | 5.16 | 145 |
| 12817.p1 | frameshift | *de novo* | chr3:71132860 | FOXP1 | +T | A339SfsX4 | | |
| 12817.p1 | missense | inherited | chr7:146449073 | CNTNAP2 | R | H275R | 5.47 | 29 |
| 13683.p1 | missense | inherited | chr2:100908186 | NPAS2 | R | N60S | 5.1 | 46 |
| 13683.p1 | missense | inherited | chr4:53468386 | SCFD2 | Y | M613V | 5.36 | 21 |

***Includes disruptive d*e novo*, SFARI, and copy number rare variants**

## Supplementary Table 8 *CNTNAP2* expression in HEK293T cell assays

| | Mean | SEM |
|---|---|---|
| pcDNA4 | 1.04 | ±0.03198 |
| FOXP2 | 0.5592 | ±0.05054 |
| FOXP1 | 0.6172 | ±0.02632 |
| FOXP1mut | 3.44 | ±0.4409 |

| comparison p-value |
|---|
| pcDNA4 vs FOXP2=0.0033 |
| pcDNA4 vs FOXP1=0.0005 |
| pcDNA4 vs FOXP1mut=0.0056 |
| FOXP2 vs FOXP1=ns |
| FOXP1 vs FOXP1mut=0.0031 |

Calculated using a two-tailed unpaired *t* test



Nature Genetics: doi:10.1038/ng.835

## Supplementary Table 9 Primer sequences for exome capture and qPCR

| | |
|---|---|
| **Adapters** | |
| Adapter_PE_Hi | ACA CTC TTT CCC TAC ACG ACG CTC TTC CGA TC*T |
| Adapter_PE_Lo | /5Phos/GAT CGG AAG AGC GGT TCA GCA GGA ATG CCG AG |
| | |
| **Library Primers** | |
| **Standard (pre-cap/pre-seq/blocking)** | |
| SLXA_Pair_For_Amp | AAT GAT ACG GCG ACC ACC GAG ATC TAC ACT CTT TCC CTA CAC GAC GCT CTT CCG ATC *T |
| SLXA_Pair_Rev_Amp | CAA GCA GAA GAC GGC ATA CGA GAT CGG TCT CGG CAT TCC TGC TGA ACC GCT CTT CCG ATC *T |
| | |
| **Barcode** | |
| **pre-cap** | |
| PreCapIndex_Fwd_Amp_Common | AAT GAT ACG GCG ACC ACC GAG ATC TAC ACT CTT TCC CTA CAC GAC GC |
| Rev_Amp_w/Barcode | CAA GCA GAA GAC GGC ATA CGA GAT CAA GGT CAC GGT CTC GGC ATT CCT GCT GAA CCG |
| 1/96 barcodes: barcode sequence | CAA GGT CA |
| **blocking** | |
| ET-Nbgn-Index-BlockFwd | AAT GAT ACG GCG ACC ACC GAG ATC TAC ACT CTT TCC CTA CAC GAC GCT CTT CCG ATC T |
| ET-Nbgn-Index-BlockRev1 | CAA GCA GAA GAC GGC ATA CGA GAT |
| ET-Nbgn-Index-BlockRev1 | CAA GCA GAA GAC GGC ATA CGA GAT |
| **pre-seq** | |
| Post_Cap_Short_Fwd_Amp | AAT GAT ACG GCG ACC ACC GAG ATC T |
| Post_Cap_Short_Rev_Amp | CAA GCA GAA GAC GGC ATA CGA GAT |

"*" refers to a phosphorothioate bond.

| | |
|---|---|
| **QPCR** | |
| CNTNAP2_F | TCC CTC CAC GTC CCA AAA ATG |
| CNTNAP2_R | TCT TGG CAT AGC CGG GAG AA |
| GAPDH_F | CAG TCC ATG CCA TCA CTG C |
| GAPDH_R | TTC GTT GTC ATA CCA GGA AAT G |

## Supplementary References


1. Fischbach, G.D. & Lord, C. The Simons Simplex Collection: a resource for identification of autism genetic risk factors. *Neuron* **68**, 192-5 (2010).
2. Lord, C. et al. The autism diagnostic observation schedule-generic: a standard measure of social and communication deficits associated with the spectrum of autism. *J Autism Dev Disord* **30**, 205-23 (2000).
3. Lord, C., Rutter, M. & Le Couteur, A. Autism Diagnostic Interview-Revised: a revised version of a diagnostic interview for caregivers of individuals with possible pervasive developmental disorders. *J Autism Dev Disord* **24**, 659-85 (1994).
4. Vernes, S.C. et al. Functional genetic analysis of mutations implicated in a human speech and language disorder. *Hum Mol Genet* **15**, 3154-67 (2006).




Nature Genetics: doi:10.1038/ng.835

5. MacDermot, K.D. et al. Identification of FOXP2 truncation as a novel cause of developmental speech and language deficits. *Am J Hum Genet* **76**, 1074-80 (2005).

6. Vernes, S.C. et al. A functional genetic link between distinct developmental language disorders. *N Engl J Med* **359**, 2337-45 (2008).

7. Bakkaloglu, B. et al. Molecular cytogenetic analysis and resequencing of contactin associated protein-like 2 in autism spectrum disorders. *Am J Hum Genet* **82**, 165-73 (2008).

8. Vissers, L.E. et al. A de novo paradigm for mental retardation. *Nat Genet* (2010).

9. Durbin, R.M. et al. A map of human genome variation from population-scale sequencing. *Nature* **467**, 1061-73 (2010).

10. Lai, C.S., Fisher, S.E., Hurst, J.A., Vargha-Khadem, F. & Monaco, A.P. A forkhead-domain gene is mutated in a severe speech and language disorder. *Nature* **413**, 519-23 (2001).



Nature Genetics: doi:10.1038/ng.835