# Supplemental Information

# Using Whole-Exome Sequencing to Identify Inherited Causes of Autism

Timothy W. Yu, Maria H. Chahrour, Michael E. Coulter, Sarn Jiralerspong, Kazuko Okamura-Ikeda, Bulent Ataman, Klaus Schmitz-Abe, David A. Harmin, Mazhar Adli, Athar N. Malik, Alissa M. D'Gama, Elaine T. Lim, Stephan J. Sanders, Ganesh H. Mochida, Jennifer N. Partlow, Christine M. Sunu, Jillian M. Felie, Jacqueline Rodriguez, Ramzi H. Nasir, Janice Ware, Robert M. Joseph, R. Sean Hill, Benjamin Y. Kwan, Muna Al-Saffar, Nahit M. Mukaddes, Asif Hashmi, Soher Balkhy, Generoso G. Gascon, Fuki M. Hisama, Elaine LeClair, Annapurna Poduri, Ozgur Oner, Samira Al-Saad, Sadika A. Al-Awadi, Laila Bastaki, Tawfeg Ben-Omran, Ahmad S. Teebi, Lihadh Al-Gazali, Valsamma Eapen, Christine R. Stevens, Leonard Rappaport, Stacey B. Gabriel, Kyriacos Markianos, Matthew W. State, Michael E. Greenberg, Hisaaki Taniguchi, Nancy E. Braverman, Eric M. Morrow, and Christopher A. Walsh

**Supplemental Information Inventory**

Supplemental Information includes five figures, five tables, Supplemental Experimental Procedures, Supplemental Text, and Supplemental References, and can be found with this article online.

Order according to how each item relates to the main figures and tables (in parentheses), including a short description of each:

Figure S1: Biochemical characterization of AMT alleles (Figure 1)

Figure S2: Biochemical characterization of AMT alleles (Figure 1)

Figure S5: Biochemical characterization of AMT alleles (Figure 1)

Figure S3: Quantification of RNAseq data presented in Figure 3E (Figure 3)

Figure S4: Pedigrees of families with candidate ASD mutations (Figure 4)

Table S1: Biochemical characterization of AMT alleles (Figure 1)

Table S2: Known disease genes used for whole exome sequencing analyses (Figure 4)

Table S3: Candidate variants in known disease genes (Table 1)

Table S4: Details of mutations reported in the study (Table 1)

Table S5: Resequencing for additional alleles of ASD genes (Figures 3 and 5)

**Supplemental Figures**

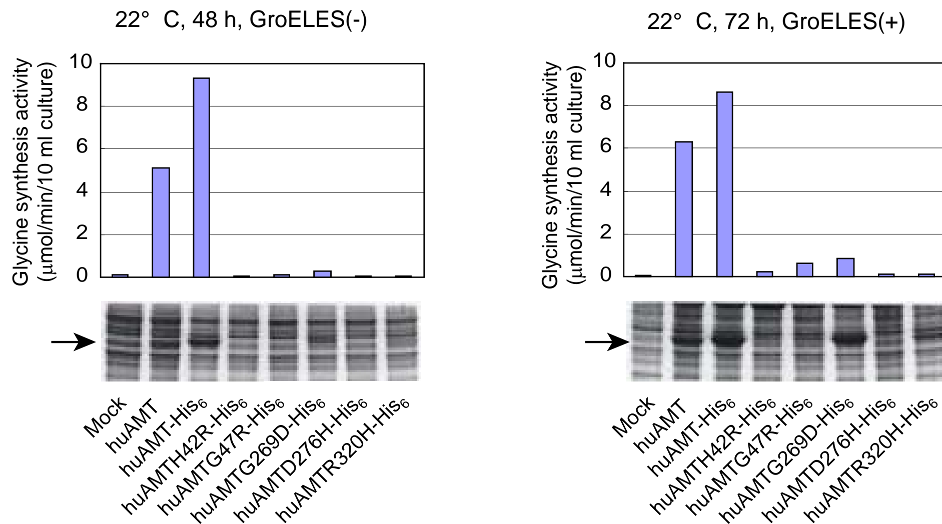
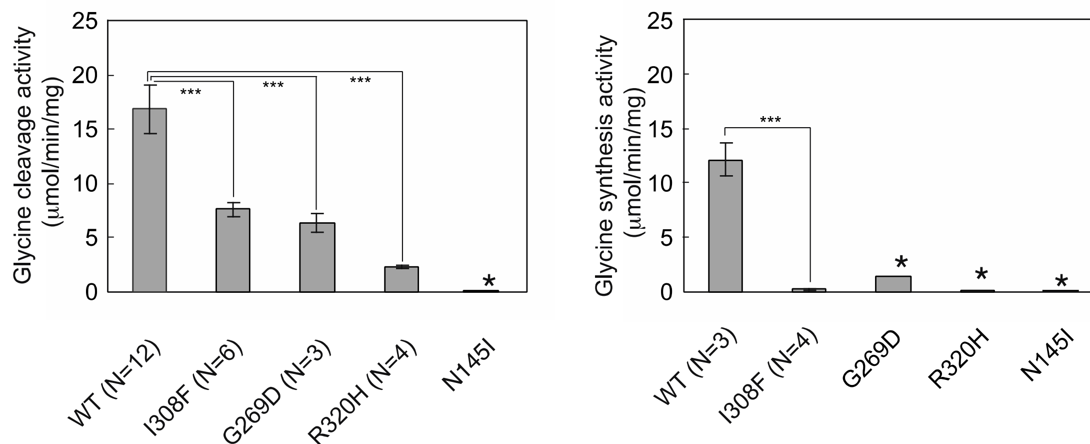



**Figure S1.** Protein solubility and enzymatic activity of five human AMT missense mutations associated with NKH. Wildtype and mutant 6xHis-human AMT proteins were expressed in *E. coli*. After lysis, soluble fractions were analyzed by SDS-PAGE and enzymatic assay of glycine synthesis activity. Untagged and 6xHis-tagged wildtype human AMT proteins are found in the supernatant as single bands (arrows). In the case of human AMT mutations associated with NKH, little to no soluble protein or enzymatic activity is detectable. Coexpression with GroES and GroEL heat shock chaperonin proteins restores solubility, but not enzymatic activity, of AMT p.G269D, but has little effect on the other AMT mutations associated with NKH.




* These values were cited from Okamura-Ikeda et al, JMB 2005

*** p<0.001

**Figure S2.** Quantification of enzymatic activity of human AMT I308F relative to wildtype and NKH-associated alleles. Wildtype and mutant human AMT proteins with a C-terminal His6-tag were expressed, purified, and assayed for enzymatic activity as previously described (Okamura-Ikeda et al., 2010; Okamura-Ikeda et al., 2005). AMT I308F retains 45% (SD 4.1%) of wildtype glycine cleavage specific activity and 1.8% (SD 0.5%) of glycine synthesis specific activity. Reported values for glycine cleavage activity of WT, I308F, G269D, and R320H AMT, and glycine synthesis activity of WT and I308F, represent direct comparisons in parallel experiments, whereas glycine cleavage activity of N145I, and glycine synthesis activities of G269D, R320H, and N145I, are cited from Okamura-Ikeda et al., 2005. Number of experimental replicates is shown in parentheses (N). Error bars represent standard deviation. For those samples run in parallel, statistically significant differences are indicated by brackets.

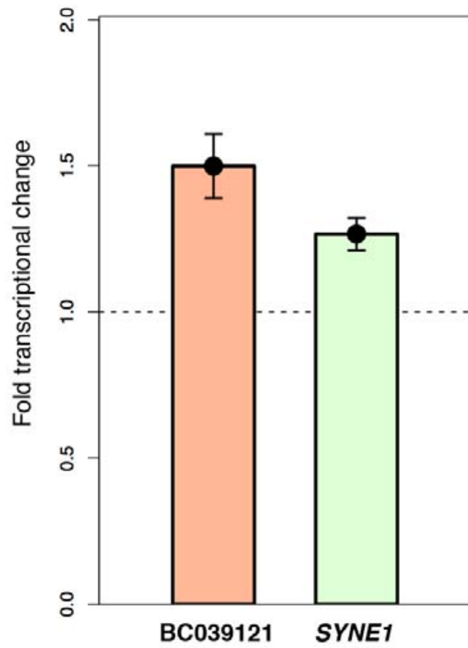



**Figure S3.** Expression of *SYNE1* and BC039121 is induced by neuronal activity. Quantification of RNAseq data presented in Figure 3E; error bars denote standard error of measurement across human neuronal cultures from five separate individuals. Transcription of *SYNE1* was induced 1.27-fold (n=5, S.E. 0.06, p=0.0203, *t* test, one-tailed) by neuronal activity, whereas transcription of BC039121 was induced by 1.50-fold (n=5, S.E. 0.11, p=0.0225, *t* test, one-tailed).

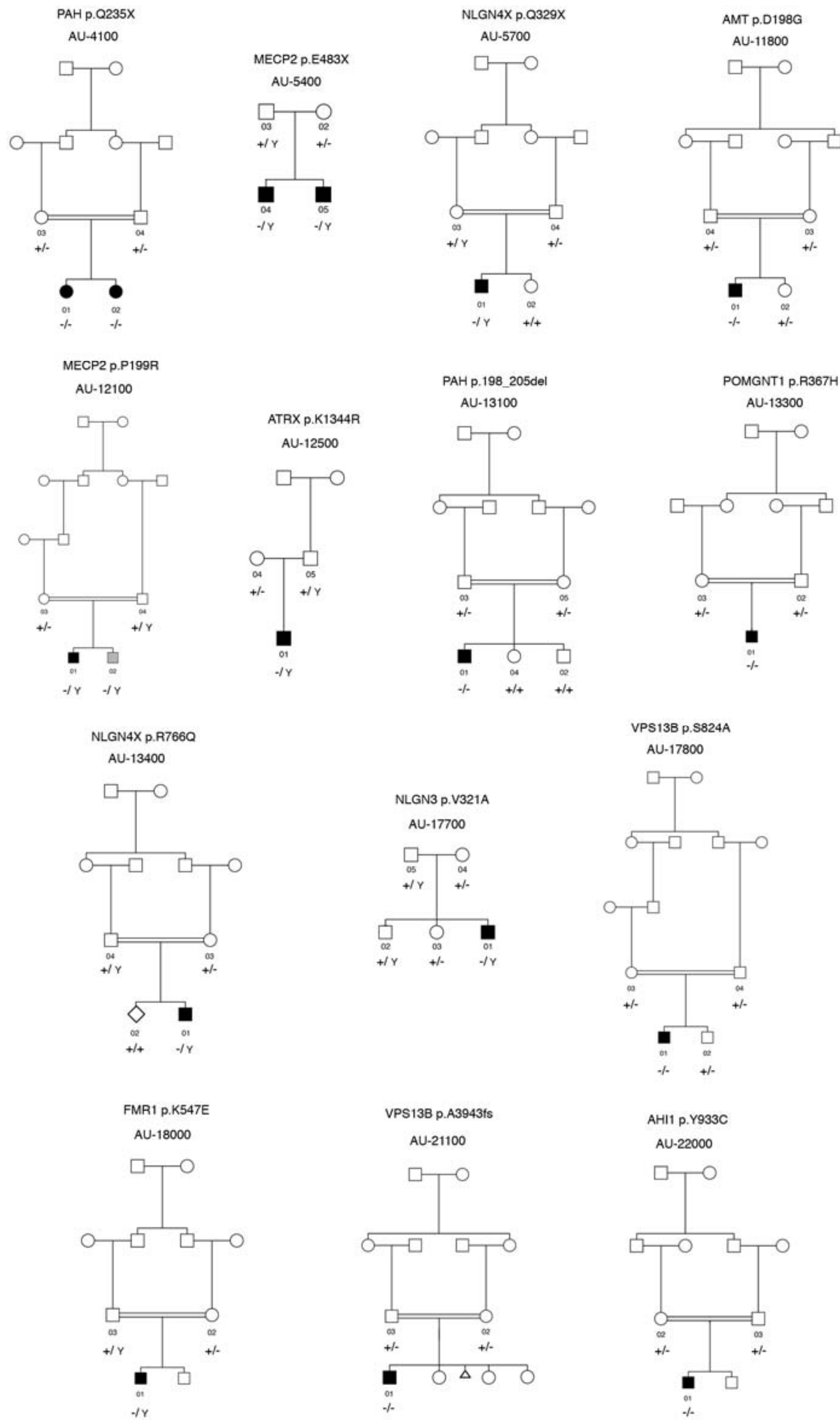
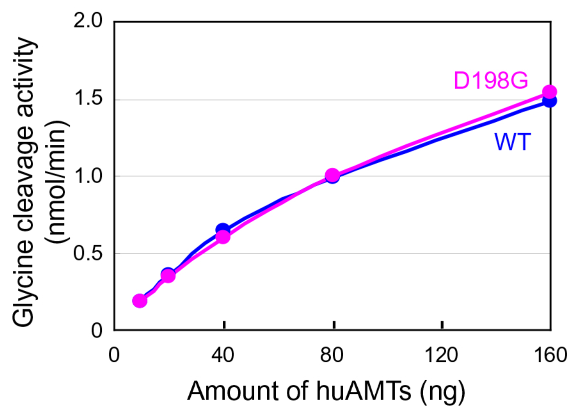
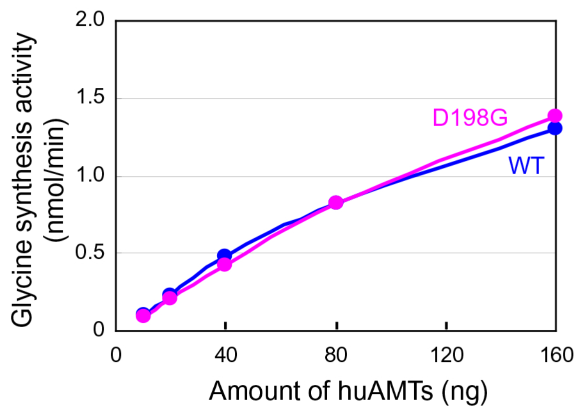



**Figure S4.** Pedigrees of the families in which segregating homozygous or hemizygous candidate ASD mutations were identified. WES in 163 families identified candidate ASD mutations in these pedigrees. Shaded symbols indicate affected individuals (black: autism, grey: ADHD). All mutations were confirmed by Sanger sequencing. Mutations and family numbers are indicated above each pedigree. Additional details about the mutations are listed in Tables S3 and S4.

**A**

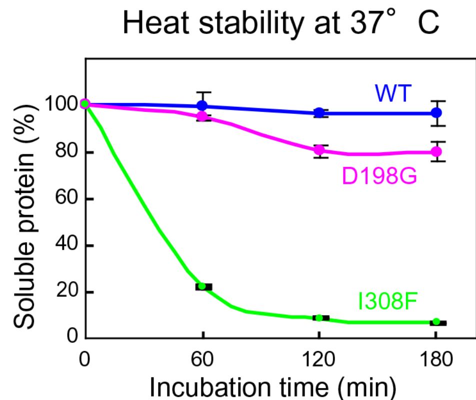


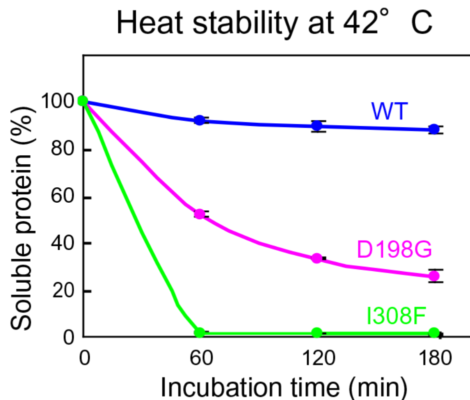




**B**

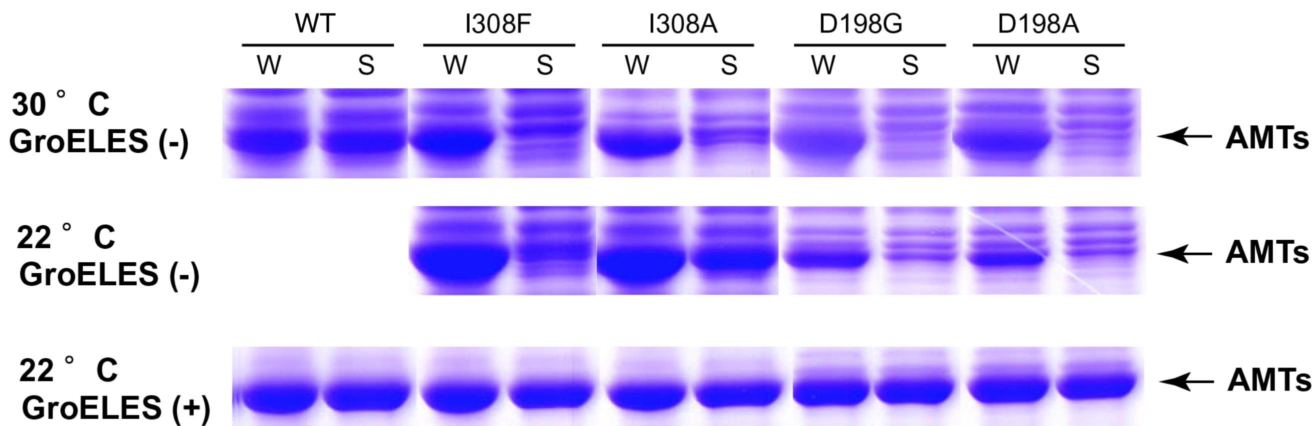
| huAMTs | Glycine cleavage activity | | Glycine synthesis activity | |
|---|---|---|---|---|
| | nmol/min/mg | % | nmol/min/mg | % |
| WT-$His_6$ | 15,503 ± 2,404 | 100 | 11,080 ± 960 | 100 |
| D198G-$His_6$ | 15,012 ±1,946 | 97.2 | 10,367 ±662 | 93.6 |

**C**




**D**




W : whole cell lysate, S : soluble fraction

**Figure S5.** AMT D198G results in protein folding and stability defects. (A, B) Wildtype and mutant 6xHis-human AMT was expressed in *E. coli*, purified, and assayed for glycine cleavage and glycine synthesis activity. AMT D198G retains enzymatic activity similar to wildtype as demonstrated by glycine cleavage (*left*) and glycine synthesis (*right*) assays. Quantification of enzymatic activity is reported plus or minus standard deviation across multiple experiments (n=4). (C) The AMT D198G recombinant protein demonstrates a temperature-sensitive stability defect. Error bars represent standard deviation across multiple experiments (n=4). (D) Unlike wildtype AMT, AMT D198G is insoluble under standard induction conditions (30°C), with little detectable protein in the supernatant despite similar overall expression levels. Reducing the induction temperature to 22°C and co-expression of GroEL and GroES are required to correct this protein folding defect.

## Supplemental Tables

**Table S1.** Interactions between protein solubility, specific activity, and total enzymatic activity in NKH-associated AMT mutations. (A) Protein soluble yield, enzyme specific activity and total enzymatic activity of wildtype AMT and AMT mutants. Wildtype and NKH-associated 6xHis-tagged AMT proteins were expressed in bacteria, purified (Ni-NTA fraction), and assayed for enzymatic activity (glycine synthesis, GS). NKH-associated mutations cause significant reductions in protein soluble yield, consistent with protein misfolding, and most are associated with reduced enzyme specific activity. An exception is AMT p.G47R, which retains significant specific GS activity (69.8% of wildtype), although overall GS activity (supernatant fraction) remains significantly reduced (3.9%). (B) Relationship of genotype and NKH phenotype for the mutations presented.

A

Protein souble yield and glycine synthesis activity in purified (Ni-NTA fraction) vs.total supernatant of wildtype and mutant huAMT

| | Ni-NTA fraction | | | | | Total supernatant | |
|---|---|---|---|---|---|---|---|
| | Protein soluble yield (mg) | (%) | GS activity (μmol/min) | Specific activity (nmol/min/mg) | (%) | GS activity (μmol/min) | (%) |
| huAMT-His$_6$ | 12.62 | 100 | 162.4 | 12867 | 100 | 190.4 | 100 |
| huAMTH42R-His$_6$ | 0.065 | 0.5 | 0.057 | 885 | 6.9 | 2.55 | 1.3 |
| huAMTG47R-His$_6$ | 0.108 | 0.9 | 0.969 | 8972 | 69.8 | 7.38 | 3.9 |
| huAMTG269D-His$_6$ | 2.62 | 20.8 | 3.21 | 1225 | 9.5 | 9.88 | 5.2 |
| huAMTD276H-His$_6$ | 0.026 | 0.2 | 0.004 | 160 | 1.2 | 1.88 | 0.99 |
| huAMTR320H-His$_6$ | 2.05 | 16.2 | 0.436 | 213 | 1.7 | 1.29 | 0.68 |

B

Relationship of AMT genotype and phenotype in NKH patients

| Genotype | Phenotype | References |
|---|---|---|
| H42R (+/+) | Typical NKH | 1 |
| G47R (+/-), R320H (+/-) | Atypical NKH | 2 |
| G269D (+/+) | Typical NKH | 2 |
| D276H (+/-), 183delC ( ➔ 94 aa peptide) (+/-) | Typical NKH | 3 |
| E211K (+/+), R320H (+/+) | Typical NKH | 4 |
| R320H (+/+) | Atypical NKH | 4 |
| N145I (+/-), R320H (+/-) | Typical NKH | 4 |


1. Kure, S. et al. (1998) *Hum. Genet.* 102, 430-434
2. Nanao, K. et al. (1994) *Hum Genet.* 93, 655-658
3. Kure, S. et al. (1998) *J. Hum. Genet.* 43, 135-137
4. Toone, J. R. et al. (2003) *Mol. Genet.* Metab. 79, 272-280

**Table S2.** Genes for syndromic recessive conditions reported to be associated with autism and/or intellectual disability that were screened in patient whole exome data. AR: autosomal recessive, AD: autosomal dominant, XL: X-linked.

| Gene symbol | Gene name | Location | Disorder | Inheritance | MIM number |
|---|---|---|---|---|---|
| *ACSL4* | Acyl-CoA synthetase long-chain family member 4 | Xq23 | Non-syndromic X-linked ID | XL | 300157 |
| *ADSL* | Adenylosuccinate lyase | 22q13.1 | Adenylosuccinate lyase deficiency | AR | 608222 |
| *AFF2* | Fragile X mental retardation 2 | Xq28 | Fragile X mental retardation 2 (FRAXE) | XL | 300806 |
| *AGTR2* | Angiotensin II receptor, type 2 | Xq23 | Non-syndromic X-linked ID | XL | 300034 |
| *AHI1* | Abelson helper integration site 1 | 6q23.3 | Joubert syndrome 3 | AR | 608894 |
| *ALDH5A1* | Aldehyde dehydrogenase 5 family, member A1 | 6p22.3 | Succinic semialdehyde dehydrogenase deficiency (gamma-hydroxybutyric aciduria) | AR | 610045 |
| *ALDH7A1* | Aldehyde dehydrogenase 7 family, member A1 | 5q23.2 | Pyridoxine-dependent epilepsy (antiquitin deficiency) | AR | 107323 |
| *AP1S2* | Adaptor-related protein complex 1, sigma 2 subunit | Xp22.2 | X-linked ID and autism syndrome characterized by hypotonia, speech delay, aggressive behavior, and brain calcifications | XL | 300629 |
| *ARHGEF6* | Rac/Cdc42 guanine nucleotide exchange factor 6 | Xq26.3 | Non-syndromic X-linked ID | XL | 300267 |
| *ARX* | Aristaless related homeobox | Xp21.3 | Epileptic encephalopathy, early infantile, 1 | XL | 300382 |
| *ATRX* | Transcriptional regulator ATRX | Xq21.1 | Large spectrum of phenotypes including ATRX syndrome (alpha thalassemia/mental retardation syndrome X-linked) and non-syndromic X-linked ID | XL | 300032 |
| *BTD* | Biotinidase | 3p25.1 | Biotinidase deficiency | AR | 609019 |
| *CACNA1F* | Calcium channel, voltage-dependent, L type, alpha 1F subunit | Xp11.23 | X-linked incomplete congenital stationary night blindness, severe form | XL | 300110 |
| *CASK* | Calcium/calmodulin-dependent serine protein kinase | Xp11.4 | Variable phenotypes, ranging from non-syndromic mild ID to severe ID with microcephaly, brain malformations, congenital nystagmus and dysmorphic facial features | XL | 300172 |
| *CDKL5* | Cyclin-dependent kinase-like 5 | Xp22.13 | Rett-like syndrome with infantile spasms and severe ID in female patients | XL | 300203 |
| *CEP290* | Centrosomal protein 290kDa | 12q21.32 | Joubert syndrome 5, Senior-Loken syndrome 6 | AR | 610142 |
| *CNTNAP2* | Contactin associated protein-like 2 | 7q35-q36 | Cortical dysplasia-focal epilepsy syndrome | AR, AD? | 604569 |
| *DCX* | Doublecortin | Xq23 | Type 1 lissencephaly | XL | 300121 |
| *DHCR7* | 7-Dehydrocholesterol reductase | 11q13.4 | Smith–Lemli–Opitz syndrome | AR | 602858 |
| *DMD* | Dystrophin | Xp21.2-p21.1 | Muscular dystrophy, Duchenne and Becker types | XL | 300377 |
| *FGD1* | Faciogenital dysplasia protein | Xp11.22 | Aarskog–Scott syndrome (faciogenital dysplasia), non-syndromic X-linked ID | XL | 300546 |
| *FMR1* | Fragile X mental retardation 1 | Xq27.3 | Fragile X syndrome | XL | 309550 |
| *FTSJ1* | FtsJ homolog 1 | Xp11.23 | Non-syndromic X-linked ID | XL | 300499 |
| *GAMT* | Guanidinoacetate N-methyltransferase | 19p13.3 | GAMT deficiency (brain creatine deficiency, synthesis defect) | AR | 601240 |
| *GATM* | Glycine amidinotransferase | 15q21.1 | AGAT deficiency | AR | 602360 |
| *GRIA3* | Glutamate receptor, ionotrophic, AMPA 3 | Xq25 | Non-syndromic X-linked ID | XL | 305915 |
| *GUCY2D* | Guanylate cyclase 2D, membrane | 17p13.1 | Leber congenital amaurosis 1, Cone-rod dystrophy 6 | AR | 600179 |

| *HOXA1* | Homeobox A1 | 7p15.2 | Bosley-Salih-Alorainy syndrome | AR | 142955 |
|---|---|---|---|---|---|
| *IL1RAPL1* | Interleukin 1 receptor accessory protein-like 1 | Xp21.3-p21.2 | Non-syndromic X-linked ID and/or ASD | XL | 300206 |
| *IQSEC2* | IQ motif and Sec7 domain 2 | Xp11.22 | Non-syndromic X-linked ID | XL | 300522 |
| *JARID1C* | Jumonji, AT rich interactive domain 1C | Xp11.22 | Large spectrum of phenotypes including ID with microcephaly, spasticity, short stature, epilepsy, and facial anomalies, as well as non-syndromic ID | XL | 314690 |
| *KIAA2022* | Hypothetical protein LOC340533 | Xq13.3 | X-linked ID, progressive quadriparesia, and autism | XL | 300524 |
| *L1CAM* | L1 cell adhesion molecule | Xq28 | Syndromic X-linked ID, MASA (mental retardation, aphasia, shuffling gait, and adducted thumbs) syndrome | XL | 308840 |
| *L2HGDH* | l-2-hydroxyglutarate dehydrogenase | 14q21.3 | L-2-hydroxyglutaric aciduria | AR | 609584 |
| *LAMP2* | Lysosomal-associated membrane protein 2 | Xq24 | Danon disease (X-linked vacuolar cardiomyopathy and myopathy) | XL | 309060 |
| *MAP2K1* | Mitogen-activated protein kinase kinase 1 | 15q22.31 | Cardiofaciocutaneous syndrome | AR | 176872 |
| *MECP2* | Methyl-CpG binding protein 2 | Xq28 | Rett syndrome | XL | 300005 |
| *MED12* | Mediator complex subunit 12 | Xq13.1 | Opitz-Kaveggia syndrome, Lujan-Fryns syndrome | XL | 300188 |
| *MID1* | Midline 1 | Xp22.2 | Opitz syndrome (Opitz/BBB syndrome) | XL | 300552 |
| *MKKS* | McKusick–Kaufman syndrome | 20p12.2 | McKusick-Kaufman syndrome, Bardet-Biedl syndrome 6 | AR | 604896 |
| *NDP* | Norrin | Xp11.3 | Norrie disease, Exudative vitreoretinopathy, X-linked | XL | 300658 |
| *NHS* | Nance–Horan syndrome | Xp22.13 | Nance–Horan syndrome (congenital cataracts and dental anomalies) | XL | 300457 |
| *NLGN3* | Neuroligin 3 | Xq13.1 | Non-syndromic X-linked ID and/or ASD | XL | 300336 |
| *NLGN4X* | Neuroligin 4, X-linked | Xp22.32-p22.31 | Non-syndromic X-linked ID and/or ASD | XL | 300427 |
| *NPHP1* | Nephrocystin-1 | 2q13 | Joubert syndrome 4 and nephronophthisis | AR | 607100 |
| *NRXN1* | Neurexin 1 | 2p16.3 | Disrupted in ASD, ID, and other neurodevelopmental and psychiatric disorders (AD?), Pitt–Hopkins-like syndrome-2 (AR) | AR, AD? | 600565 |
| *OCRL* | Phosphatidylinositol polyphosphate 5-phosphatase | Xq25-q26 | Lowe syndrome or oculo-cerebro-renal syndrome (ID, bilateral cataract and renal Fanconi syndrome) | XL | 300535 |
| *OPHN1* | Oligophrenin 1 | Xq12 | ID with cerebellar and vermis hypoplasia | XL | 300127 |
| *OTC* | Ornithine carbamoyltransferase | Xp11.4 | Ornithine transcarbamylase deficiency | XL | 300461 |
| *PAH* | Phenylalanine hydroxylase | 12q23.2 | Phenylketonuria | AR | 612349 |
| *PCDH19* | Protocadherin 19 | Xq22.1 | X-linked female-limited epilepsy and cognitive impairment | XL | 300460 |
| *PHF6* | PHD finger protein 6 | Xq26.2 | Borjeson–Forssman–Lehmann syndrome (ID, epilepsy, and hypogonadism) | XL | 300414 |
| *PHF8* | PHD finger protein 8 | Xp11.22 | Siderius–Hamel syndrome (ID with cleft lip or cleft palate) | XL | 300560 |
| *POMGNT1* | O-linked mannose | 1p34.1 | Muscle–eye–brain disease (congenital muscular dystrophy, structural eye abnormalities and lissencephaly) | AR | 606822 |
| *POMT1* | Protein-O-mannosyltransferase 1 | 9q34.13 | Limb-girdle muscular dystrophy with ID, Walker–Warburg syndrome | AR | 607423 |
| *PQBP1* | Polyglutamine binding protein 1 | Xp11.23 | Large spectrum of ID phenotypes, including Renpenning syndrome (microcephaly, short stature, small testes and dysmorphic features) and non-syndromic ID | XL | 300463 |
| *PRSS12* | Neurotrypsin precursor (protease, serine, 12) | 4q26 | Non-syndromic ID | AR | 606709 |
| *PTCHD1* | Patched domain containing 1 | Xp22.11 | X-linked ID and ASD | XL | 300828 |

| | | | | | |
|---|---|---|---|---|---|
| *RAB39B* | RAB39B, member RAS oncogene family | Xq28 | X-linked ID associated with autism, epilepsy, and macrocephaly | XL | 300774 |
| *RPE65* | Retinal pigment epithelium-specific protein | 1p31.3-p31.2 | Leber congenital amaurosis 2, Retinitis pigmentosa | AR | 180069 |
| *RPGRIP1L* | RPGRIP1-like (retinitis pigmentosa GTPase regulator-like) | 16q12.2 | Joubert syndrome 7, Meckel syndrome 5, COACH syndrome (Joubert syndrome with congenital hepatic fibrosis) | AR | 610937 |
| *SGSH* | N-sulfoglucosamine sulfohydrolase | 17q25.3 | Sanfilippo syndrome A (mucopolysaccharidosis III A) | AR | 605270 |
| *SLC6A8* | Solute carrier family 6 (neurotransmitter transporter, creatine), member 8 | Xq28 | Creatine deficiency syndrome, X-linked | XL | 300036 |
| *SLC9A6* | Solute carrier family 9 (sodium/hydrogen exchanger), member 6 | Xq26.3 | Syndromic X-linked ID, Christianson type (ID, microcephaly, epilepsy, and ataxia) | XL | 300231 |
| *SMC1A* | Structural maintenance of chromosomes 1A | Xp11.22 | Cornelia de Lange syndrome 2 | XL | 300040 |
| *SYN1* | Synapsin I | Xp11.23 | X-linked epilepsy with variable learning disabilities and behavior disorders | XL | 313440 |
| *UPF3B* | UPF3 regulator of nonsense transcripts homolog B | Xq24 | Non-syndromic X-linked ID | XL | 300298 |
| *VPS13B* | Vacuolar protein sorting 13 homolog B | 8q22.2 | Cohen syndrome | AR | 607817 |
| *ZNF674* | Zinc finger family member 674 | Xp11.3-p11.2 | Non-syndromic X-linked ID | XL | 300573 |
| *ZNF81* | Zinc finger protein 81 | Xp11.23 | Non-syndromic X-linked ID | XL | 314998 |

**Table S3.** Candidate missense variants in genes reported to be associated with autism and/or intellectual disability. M: male, F: female, ID: intellectual disability.

| Mutation | Known disease association | Family | Structure | Consanguinity | # Affected | # Unaffected | Linkage | Primary phenotype | Additional phenotypes | Investigations (Normal unless stated) |
|---|---|---|---|---|---|---|---|---|---|---|
| AFF2 p.R927H | Fragile X mental retardation 2 | AU-4400 | Simplex | No | 1 (M) | 1 | Yes | Autism | ID (IQ 75) | Array CGH |
| AHI1 p.Y933C | Joubert syndrome 3 | AU-22000 | Simplex | Yes | 1 (M) | 1 | Yes | Autism | ID | - |
| ATRX p.K1344R | ATRX syndrome (alpha thalassemia/mental retardation syndrome X-linked) and non-syndromic X-linked ID | AU-12500 | Simplex | No | 1 (M) | - | Yes | Autism | Developmental regression at 2y, seizures (following trauma at 8y, resolved), mitral valve prolapse, 3-5 café au lait spots, food and environmental allergies | 46, XY, Fragile X, 22q11 and 15q11 FISH, NF1 (no protein truncation segments 1-5), brain MRI (diminished white matter), EEG, audiogram, CBC, TFT, urine organic acid, 7-dehydrocholesterol, positive MBP antibody |
| FMR1 p.K547E | Fragile X mental retardation | AU-18000 | Simplex | Yes | 1 (M) | 1 | Yes | Autism | - | - |
| MECP2 p.P199R | Rett syndrome, ASD | AU-12100 | Simplex | Yes | 2 (M) | - | Yes | Autism | Brother with ADHD | Hearing exam, EEG |
| NLGN3 p.V321A | Non-syndromic X-linked ID and/or ASD | AU-17700 | Simplex | No | 1 (M) | 2 | Yes | Autism | ID, hyperactive, mild hearing loss, obesity, OFC at 65th percentile, dysmorphic features (low hairline, hair whorl, down-slanting and deep set eyes, narrow palpebral fissures) | Brain MRI |
| NLGN4X p.R766Q | Non-syndromic X-linked ID and/or ASD | AU-13400 | Simplex | Yes | 1 (M) | 1 | Yes | Autism | - | - |

**Table S4.** Mutations in genes reported to be associated with autism and/or intellectual disability. Genomic positions of identified variants are presented. Minor allele frequencies (MAF) were obtained from the 1000 Genomes Project (1000G), the NHLBI Exome Sequencing Project (EVS5000), and a dataset of 831 Middle Eastern exomes (831ME).

| Family | Gene | Transcript (RefSeq) | Position (hg19) | Mutation | Amino acid change | MAF (831ME) | MAF (1000G) | MAF (EVS5000) | SIFT | PolyPhen-2 |
|---|---|---|---|---|---|---|---|---|---|---|
| AU-4400 | *AFF2* | NM_002025 | chrX: 148,044,334G>A | c.G2780A | p.R927H | 0.24% | 0% | 0.09% | 0.02 | 0.999 |
| AU-22000 | *AHI1* | NM_017651 | chr6: 135,732,649T>C | c.A2798G | p.Y933C | 0.66 | 0 | 0.35 | 0.03 | 0 |
| AU-11800 | *AMT* | NM_000481 | chr3: 49,456,796T>C | c.A593G | p.D198G | 0 | 0 | 0 | 0.01 | 0.979 |
| AU-1700 | *AMT* | NM_000481 | chr3: 49,455,362T>A | c.A922T | p.I308F | 0 | 0 | 0 | 0 | 0.048 |
| AU-12500 | *ATRX* | NM_000489 | chrX: 76,918,960T>C | c.A4031G | p.K1344R | 0 | 0 | 0 | 0.02 | 0.877 |
| AU-3200 | *CEP152* | NM_001194998 | chr15: 49,054,766G>A | c.C2384T | p.T795I | 0 | 0 | 0 | 0.01 | 0.105 |
| AU-18000 | *FMR1* | NM_001185082 | chrX: 147,030,242A>G | c.A1639G | p.K547E | 0 | 0 | 0 | 0.01 | 0.019 |
| AU-12100 | *MECP2* | NM_004992 | chrX: 153,296,683G>C | c.C596G | p.P199R | 0 | 0 | 0 | 0.11 | 0.924 |
| AU-5400 | *MECP2* | NM_004992 | chrX: 153,295,832C>A | c.G1447T | p.E483X | 0 | 0 | 0 | 0 | 0.639 |
| AU-20300 | *MTRR* | NM_002454 | chr5: 7,892,937A>G | c.A1468G | p.T490A | 0.30 | 0.56 | 0.33 | 0.01 | 0.999 |
| AU-17700 | *NLGN3* | NM_001166660 | chrX: 70,387,029T>C | c.T962C | p.V321A | 0 | 0 | 0 | 0.14 | 0.988 |
| AU-5700 | *NLGN4X* | NM_020742 | chrX: 5,821,734G>A | c.C985T | p.Q329X | 0 | 0 | 0 | 0.21 | 0.737 |
| AU-13400 | *NLGN4X* | NM_020742 | chrX: 5,811,012C>T | c.G2297A | p.R766Q | 0.06 | 0 | 0 | 0.01 | 0.997 |
| AU-13100 | *PAH* | NM_000277 | chr12: 103,249,007_103,249,028 delCATAGCAAGCATGGGTTT TATA | c.592_613del | p.198_205del | 0 | 0 | 0 | 0 | 0 |
| AU-4100 | *PAH* | NM_000277 | chr12: 103,248,917G>A | c.C703T | p.Q235X | 0 | 0 | 0 | 1 | 0.735 |
| AU-3500 | *PEX7* | NM_000288 | chr6: 137,147,493G>C | c.G225C | p.W75C | 0 | 0 | 0 | 0 | 1 |
| AU-13300 | *POMGNT1* | NM_017739 | chr1: 46,658,987C>T | c.G1100A | p.R367H | 0 | 0 | 0 | 0 | 1 |
| AU-1600 | *SYNE1* | NM_182961 | chr6: 152,690,641G>T | c.C9616A | p.L3206M | 0 | 0 | 0 | 0.01 | 0.984 |
| AU-17800 | *VPS13B* | NM_152564 | chr8: 100,205,240T>G | c.T2470G | p.S824A | 0 | 0 | 0.04 | 0.29 | 0.002 |
| AU-21100 | *VPS13B* | NM_152564 | chr8: 100,887,727_100,887,728 insC | c.11827_11828insC | p.A3943fs | 0 | 0 | 0 | 0 | 0 |

**Table S5.** Analysis of biallelic (homozygous or compound heterozygous) mutations in the Simons Simplex Collection (SSC; whole exome sequence data from 612 families) in 6 genes: *AMT*, *PEX7*, *SYNE1*, *VPS13B*, *PAH*, and *POMGNT1*. Variants identified in affected probands (A) and unaffected siblings (B) are presented. For the purposes of this analysis, non protein-altering variants, and common variants with an allele frequency >1% (in the SSC whole exome cohort, in the 1000 Genomes project, or in the NHLBI Exome Sequencing Project) were excluded. MAF: minor allele frequency, 1000G: the 1000 Genomes Project, EVS5000: the NHLBI Exome Sequencing Project, SSC: allele frequency in the SSC whole exome cohort.

**A.** Affected probands

| Individual | Gene | Transcript (RefSeq) | Position (hg19) | Amino acid change | dbSNP135 | MAF (1000G) | MAF (EVS5000) | MAF (SSC) | Interpretation |
|---|---|---|---|---|---|---|---|---|---|
| 11540.p1 | *SYNE1* | NM_182961 | chr6: 152,776,571C>Y,<br>chr6: 152,690,150G>R/<br>chr6: 152,639,235C>Y | R961Q,<br>S3255L†/<br>R5518Q | rs76646638<br>rs114954026<br>rs150604289 | 0.140<br>0.410<br>– | 0.344<br>0.753<br>0.009 | 0.49<br>0.25<br>0.08 | Uncertain pathogenicity. Variants fall outside of BC039121 isoform, except S3255L which is homozygous in father (unaffected) |
| 12373.p1 | *SYNE1* | NM_182961 | chr6: 152,464,786G>R/<br>chr6: 152,470,773T>K | P8364L/<br>N8161H | rs148376885<br>rs36215251 | 0.090<br>0.320 | 0.214<br>– | 0.25<br>0.08 | Uncertain pathogenicity. Both fall outside of BC039121 isoform |
| 12424.p1 | *SYNE1* | NM_182961 | chr6: 152,647,218C>Y/<br>chr6: 152,605,116C>S | D5105N/<br>L6068F | rs35493783<br>rs138039375 | 0.460<br>0.230 | 0.679<br>0.047 | 0.25<br>0.16 | Uncertain pathogenicity. Both fall outside of BC039121 isoform |
| 12445.p1 | *SYNE1* | NM_182961 | chr6: 152,730,222G>R/<br>chr6: 152,771,967A>R | T2174I/<br>V1063A | rs141858284<br>rs141464488 | 0.050<br>0.090 | 0.093<br>0.418 | 0.49<br>0.65 | Uncertain pathogenicity. Both fall outside of BC039121 isoform |
| 13621.p1 | *SYNE1* | NM_182961 | chr6: 152,771,967A>R/<br>chr6: 152,792,791A>M | V1063A/<br>S525A | rs141464488<br>– | 0.090<br>– | 0.418<br>– | 0.65<br>0.08 | Uncertain pathogenicity. Both fall outside of BC039121 isoform |
| 11659.p1 | *VPS13B* | NM_152564 | chr8: 100,568,761A>R/<br>chr8: 100,729,575A>M | N1610S/<br>I2211L | – | – | –<br>0.009 | 0.08<br>0.08 | Uncertain pathogenicity. N1610S is highly conserved but I2211L is not. Proband reported to be nondysmorphic |
| 12100.p1 | *VPS13B* | NM_152564 | chr8: 100,847,933T>W/<br>chr8: 100,874,030G>R | S3303R/<br>A3691T | –<br>rs142476821 | –<br>0.230 | –<br>0.372 | 0.08<br>0.82 | Likely pathogenic. Both residues highly conserved, supportive SIFT and PP2 scores. Proband with dysmorphic features |
| 12651.p1 | *VPS13B* | NM_152564 | chr8: 100,396,500G>R/<br>chr8: 100,829,780G>R | W963X/<br>G2704R | –<br>– | –<br>– | –<br>– | 0.08<br>0.08 | Likely pathogenic. Both residues highly conserved, supportive SIFT and PP2 scores. Proband with dysmorphic features |

**B.** Unaffected siblings

| Individual | Gene | Transcript (RefSeq) | Position (hg19) | Amino acid change | dbSNP135 | MAF (1000G) | MAF (EVS5000) | MAF (SSC) | Interpretation |
|---|---|---|---|---|---|---|---|---|---|
| 13385.s1 | *SYNE1* | NM_182961 | chr6: 152,639,236G>R/<br>chr6: 152,776,571C>Y | R5518W/<br>R961Q | –<br>rs76646638 | –<br>0.14 | 0.009<br>0.344 | 0.16<br>0.49 | Both fall outside of BC039121 isoform |
| 13148.s1 | *VPS13B* | NM_152564 | chr8: 100,146,901G>T | Q416H | rs143024324 | 0.090 | 0.112 | 0.49 | Q416H alters a conserved residue.<br>Family not scored for dysmorphisms |

## Supplemental Experimental Procedures

### Patient recruitment

Diagnostic and Statistical Manual of Psychiatric Disease IV-Revised (DSMIV-R) criteria were confirmed in all individuals with an autism diagnosis. Site visits by a team of Boston Children's Hospital-affiliated clinicians (clinical psychology, genetics, developmental medicine, neurology) were conducted to perform additional phenotyping on select informative families, including additional quantitative measures (Autism Diagnostic Interview-Revised, Autism Diagnostic Observation Scale, Social Communication Questionnaire, Social Responsiveness Scale, Vineland Adaptive Behavior Scale) (Morrow et al., 2008).

### Genome-wide linkage and homozygosity scans

Genome-wide SNP screens were performed at the Broad Institute and Dana Farber Cancer Institute. Families were genotyped using Affymetrix 500K (NspI/Sty) or Affymetrix 6.0 microarrays. SNPs were filtered to remove loci genotyped at <95% of all samples, MAF>10%, and SNPs that deviate from Hardy-Weinberg equilibrium with a p-value of <0.00000001. Any markers missing from one or more individuals within a pedigree were removed from all members of the pedigree. Linkage disequilibrium-based SNP pruning was performed with PLINK, followed by filtering of loci homozygous in all samples and those with Mendelian inheritance errors. Multipoint LOD scores were calculated using MERLIN, assuming a recessive mode of disease inheritance, full penetrance, and a disease allele frequency of 0.0001. Runs of homozygosity were

calculated using custom Perl scripts, allowing for no more than 2 consecutive heterozygous SNPs in a run and 3 heterozygous calls in every 10 consecutive SNPs. Intervals homozygous for the same haplotype and shared by all affected individuals were used to narrow the locus in each family.

**Whole exome sequencing and data analysis**

DNA samples were sequenced at the Broad Institute. Whole blood DNA was subject to barcoded exome capture (SureSelect v2, targeting 98.2% of CCDS sequences as of November 2010; Agilent Technologies, Inc., Santa Clara, CA,) and paired-end sequencing (Illumina HiSeq, read length of 72 bp). Whole exome sequence was obtained on a total of 277 affected children and 409 parents, with a mean target coverage of 85.6% at >=20X and a mean read depth of 158X. For this study, families harboring known autism-associated CNVs were excluded (Supplemental Experimental Procedures).

Paired reads were aligned to NCBI human genome build v37 using BWA(Li and Durbin, 2009), Picard (http://picard.sourceforge.net/), and GATK(McKenna et al., 2010). Alignments underwent base quality score recalibration, PCR duplicate removal, and realignment around indels. Variants were called with GATK, subject to variant quality score recalibration and filtered to recover 99% of known sites. Variants were annotated with respect to gene model, pathogenicity and presence in public databases using ANNOVAR (Wang et al., 2010), and additional custom pipeline components to annotate all biallelic changes (homozygous and compound heterozygous variants). Under a rare-variant hypothesis, variants were filtered to exclude those with high minor allele frequency (MAF≥1%), using an internal dataset of 831 exomes from the Middle Eastern

population inclusive of the families under study and other families under investigation for non-autism neurological disease phenotypes. For the purposes of calculating each variant's MAF in this population, alleles contributed by the family under question were excluded in order to avoid artificially inflating the MAF. All reported variants were confirmed by Sanger sequencing.

**Sanger resequencing**

Sequencing by capillary electrophoresis was performed according to standard molecular biology practices at SeqWright, Inc. (Houston, Texas).

**Data visualization**

Protein modeling was performed using the UCSF Chimera package (Pettersen et al., 2004) from the Resource for Biocomputing, Visualization, and Informatics at the University of California, San Francisco (supported by NIH P41 RR001081). Genomic annotations and plots were produced as custom tracks on the UCSC Genome Browser (Kent et al., 2002).

**RNAseq of human neurons**

Primary human neuronal cells were purchased from Sciencell (Carlsbad, CA). For RNAseq experiments, neuronal cultures from 3 biological replicates were grown for around two weeks (DIV13-16). At the final day in culture, neurons in the experimental group were stimulated for 6 hours with 55mM KCl and were harvested along with the unstimulated control neurons using Trizol (Invitrogen). After total RNA was extracted,

ribosomal RNAs were depleted using the RiboMinus kit (Invitrogen). RNA quantity and quality were determined by Qubit (Invitrogen) and Bioanalayzer (Agilent), respectively. Next, strand-specific and paired-end cDNA libraries were generated using the PE RNAseq library kit (Illumina). Following quality control of the generated cDNA libraries, RNAseq was performed using HiSeq 2000 at the Broad Institute. 76-bp RNAseq reads were aligned to the human GRCh37/hg19 assembly using BWA. Reads that mapped uniquely with no more than 5 mismatches were further mapped to annotated (RefSeq) genic features. For quantification of *SYNE1* expression in response to depolarization, for each biological replicate, expression levels (normalized reads per bp) of individual exons were calculated, which allowed the calculation of the average expression level over any isoform of *SYNE1* comprising subsets of these exons. Then fold-change ratios (6 hours/unstimulated) of these levels were calculated for each replicate and isoform, and finally each isoform's mean and standard error over the replicates' fold changes.

**ChIPseq of human neurons**

The mini-ChIP assays were performed as previously described (Adli and Bernstein, 2011) on human neuronal cells that had been cultured for around two weeks. Briefly, cells were cross-linked, lysed, and sonicated for 3.5 minutes using a Branson 250. The fragmented chromatin was then immunoprecipitated overnight with 1µg of H3K27Ac (abcam Cat# ab4728) and H3K27me3 (Millipore Cat# 074490) antibodies. After samples were incubated with Protein A-Sepharose beads for 2 hours, they were collected and washed. The DNA was then eluted from the beads twice at 65°C for 10 minutes. The

eluted chromatin and the "input" samples were then reverse cross-linked at 65°C for 5 hours and were subsequently Proteinase K digested at 37°C for 2 hours. The ChIP DNA was recovered by standard phenol-chloroform-isoamyl alcohol extraction and was then precipitated overnight with ethanol. The ChIP DNA libraries were then constructed using ChIP-Seq DNA Sample Prep Kit (Illumina) and subsequently sequenced using HiSeq 2000 (Illumina) in Biopolymers facility at Harvard Medical School. ChIPseq reads were aligned to the human genome (GRCh37/hg19 assembly) using BWA with default parameters. Only uniquely mappable reads aligning to a single genomic position were used for findings reported in this manuscript. After ChIPseq reads were aligned, they were extended to 300bp to match the length of the DNA fragments that were sequenced. Finally, BigWig tracks were generated displaying the data normalized to 20 million reads per experimental condition.

**Electronic resources**

1000 Genomes Project: http://www.1000genomes.org/

dbSNP132: http://www.ncbi.nlm.nih.gov/projects/SNP/

NHLBI Exome Sequencing Project: https://esp.gs.washington.edu/drupal/

UCSC Genome Browser: http://genome.ucsc.edu

**Copy number variant analyses**

Genome-wide SNP screens were performed at the Broad Institute and Dana Farber Institute. Families were genotyped using Affymetrix 500K (NspI/Sty) or Affymetrix 6.0 microarrays. CNV analyses were performed using a custom analytic pipeline employing

four different algorithms: Birdsuite, PennCNV, Nexus and Affymetrix Genotyping Console. Full results of this pipeline will be reported elsewhere (Schmitz-Abe et al, in preparation). Criteria for CNV calling included the intersection of two or more algorithms, 10 or more probes (15 for Affymetrix 6.0 samples) and size >100 kb. Samples with a very large number of calls (120) were rejected, as they are typically the result of poor DNA hybridization. After CNV calling, results were compared to a reference catalog generated from 1,258 HapMap samples and filtered to remove common variants, and then annotated with respect to previously reported CNVs associated with autism and/or intellectual disability (>50% overlap with known boundaries) (Cooper et al., 2011; Sanders et al., 2011).

## Supplemental Text

### AMT p.I308F family (AU-1700) Clinical Information

This family originates from the Kingdom of Saudi Arabia and exhibits complex consanguinity. A son and daughter have autism spectrum disorder with intellectual disability, mild dysmorphology, and seizures. A younger son has autism spectrum disorder. Another son and daughter are unaffected and otherwise in good health.

The eldest affected son (AU-1702) was evaluated at 12 years of age. He exhibited poor eye contact, lack of social and emotional reciprocity, and absence of joint attention. He also exhibited self-stimulatory behaviors, hyperactivity, and inappropriate laughing. He scored a 24 on the Social Communication Questionnaire (SCQ) (cutoff >15). His birth history was notable for an unremarkable pregnancy, nuchal cord at delivery, and birth

weight of 3 kilograms (15$^{th}$ percentile). He developed brief apneic episodes on day of life three, and head ultrasound revealed a small subependymal hemorrhage in the right caudate nucleus. He was treated with phenytoin for presumptive seizures and was observed in the ICU until discharge on day of life 12. Seizures recurred at 10 months, requiring extended hospitalization for two months. EEGs showed synchronous sharp wave slow wave complexes, high amplitude delta activity and a burst suppression pattern. A karyotype was normal (46, XY). A plasma amino acid profile was normal. At 22 months of age, he was readmitted to the hospital with frequent seizures having lasted almost one week and accompanied by fever and flu-like illness. Investigations including CSF culture showed no evidence of infection. Magnetic resonance imaging (MRI) of the brain was done at three years of age and revealed a prominent cisterna magna, mild dilation of the ventricular system and mild frontotemporal atrophy. Neuropsychology evaluation at three years of age described global developmental and severe language delay. Expressive language consisted of only a few words. He did not follow commands and had poor self care skills. He walked and ran but was unsteady and fell much of the time. He displayed some tremor and difficulty in manipulating objects. He did not play with toys much and would be content to tear paper for long periods of time. On the Bailey Scale of Infant Development, motor development was estimated at 15 months and his cognitive development age to 10 months. At age 12 years, physical exam revealed trigonencephaly, large ears, a long tubular nose, a high arched palate and dental crowding, inverted and hypoplastic nipples, and unusual palmar creases. On neurologic exam, he had mild to moderate hypotonia, 1+ reflexes in upper limbs and knees but trace reflexes in the ankles. His knees locked and feet everted upon walking.

The second affected child (AU-1701) was a 9 year old girl. She exhibited poor eye contact, lack of social reciprocity, and self-stimulatory behaviors including hand flapping and body rocking. She would laugh and vocalize, but could not communicate. She scored a 25 on the SCQ (cutoff, 15). Her birth history was unremarkable. Medical history was notable for epilepsy only intermittently requiring antiepileptic medication, normal brain MRI, and normal plasma amino acid profiling by tandem mass spectroscopy. On examination, facial features were mildly dysmorphic and she had a low posterior hairline, hypertrichosis of the arms and back, long eyelashes and nose, a high arched palate, and everted feet. Neurologicaly she had mild alternating esotropia, hypotonia, normal reflexes, down-going Babinski reflexes bilaterally, and a wide-based gait.

The youngest affected child (AU-1707) was a two years boy at the time of evaluation. He demonstrated language and motor delays (did not sit until 1 year old, not yet walking by age 2) and mild to moderate hypotonia. He had a history of a single febrile seizure. MR imaging of the brain was reportedly normal as was a karyotype (46, XY). He was diagnosed with PDD but his age precluded a definitive autism diagnosis.

**PEX7 p.W75C family (AU-3500) Clinical Information**

This family originates from Pakistan and the parents are first cousins who are in general in good health. They have six children, including two sons and one daughter affected by ASD; and two other sons and a daughter who are otherwise healthy. At the time of initial investigation, the eldest affected son (AU-3501) was 24 years old, the affected daughter (AU-3504) was 20 years old and the youngest affected son (AU-3506) was 13 years old.

All were diagnosed with ASD and intellectual disability. They understand simple commands of daily functioning and can communicate with a limited vocabulary.

AU-3501 exhibited poor eye contact and no social interests. He could understand simple commands and was described as being able to "engage in very limited conversation." He attended a special education school for several years before being withdrawn to live at home by his parents for perceived lack of progress. His medical history was notable for epilepsy (generalized tonic clonic seizures, on three antiepileptic medications) and bilateral cataracts.

AU-3506 also exhibited social deficits. He had more expressive language than his brother. He was echolalic. He had a fascination with automobiles. He had stereotyped movements of the upper limbs and lips. Self-injurious behavior was not documented. He never attended school. He also had cataracts, but no epilepsy.
AU-3504 had social impairments and intellectual disability but was the most mildly affected. She had the highest level of functional skills. Like her brother AU-3506, she was enrolled in special education for 2-3 years before being withdrawn due to lack of progress. All three siblings were independent with toileting, but required assistance with other activities of daily living such as bathing (all three) and dressing (AU-3501 and AU-3506). AU-3504 was able to participate in some household tasks in the kitchen with her family. AU-3504 did not have cataracts. She developed seizures during young adulthood but at the time of evaluation was seizure-free without medication.

On physical examination, all three siblings were not dysmorphic. There was no microcephaly and there were no skeletal or dermatological abnormalities. Elemental neurologic examinations revealed no focal deficits. All three siblings walked normally, but running was awkward, and none were able to ride a bicycle.

**Previously reported PEX7 Family (PBD075) Clinical Information**

Initial clinical and biochemical descriptions of individuals in this family were previously published (Braverman et al., 2002; Moser et al., 1995) (family PBD075). We recontacted this family and obtained additional interval history via direct clinical interview and examination.

This family had three children, born to nonconsanguineous, neurotypical, Caucasian parents. Two daughters were affected by atypical RCDP (see below) and a third child was a neurotypical male.

The eldest daughter in this family had unremarkable prenatal and neonatal histories. She first came to medical attention at the age of two when she was noted to have corneal clouding at a routine pediatric visit. She was evaluated by ophthalmology and diagnosed with bilateral cataracts. She described at this time as an “active girl” without behavioral or developmental concerns. Her head circumference was 47cm (34%ile). Very long chain fatty acid (VLCFA) levels were normal, but serum phytanic acid was measured at 22 (normal <3). She exhibited intermediate reductions in erythrocyte plasmalogen levels. At age 4 years 9 months, repeat developmental assessment showed the ability to repeat word sequences at a 4-5 year old level. She exhibited some gross motor clumsiness but

had otherwise a normal physical exam. Phytanic acid levels remained elevated (12.8, normal <3). Cataract surgery was done at 6 years of age. At age 7 while in the first grade, she was noted to have minor learning difficulties and attentional problems and was put on an IEP. Examination at that time made note of a slightly simple ear shape but no other dysmorphic features. Weight was 22.9kg ($50^{th}$%ile), height was 123cm ($75^{th}$%ile) and head circumference was 50.5cm (25%ile). Her neurological exam was unremarkable. Currently, this young woman is in her 20s. She continues to carry an ADHD diagnosis, but attended college, lives independently, drives a car, and works as a teaching assistant with children with developmental difficulties.

Like her sister, the younger girl in the family had an unremarkable prenatal and neonatal course. She was noted to have bilateral diffuse cataracts at 9 months of age. By age 10 months her gross motor development was noted to be delayed. Her head measured at 42.7 (5%ile). At age 2 she was diagnosed at Children's Hospital Boston with delays in receptive and expressive language, communication, pretend play, and social/adaptive behavior. Hearing problems were ruled out. She was diagnosed with pervasive developmental delay. Subsequent evaluations noted autistic features including repetitive, self-stimulatory and stereotypic behaviors. By age 4y9m she had 2 seizures, one in the setting of a low fever. EEG and MRI were reportedly normal. She was noted to be otherwise healthy and had no dysmorphic features and normal tone and strength. Phytanic acid levels at age 2 were measured on three serial exams at 8, 4.5, and 6.1 (normal <3). VLCFA levels were normal in fibroblasts and serum, plasmalogen synthesis was reduced (0.71, normal 1.65 +/- 0.66), and phytanic acid oxidation was

deficient at 0.121 (normal >3). She had cataract surgery at age 9. At age 11 poor weight gain was noted and endoscopy revealed chronic *H. Pylori gastritis* and colonic lymphoid hyperplasia. As a young adult she carries the diagnosis of autism, is on seizure medication and lives in an adult residential facility.

**SYNE1 Family (AU-1600) Clinical Information**

This is a Kuwaiti family in which parents were double first cousins. Four of five children were autistic. The proband, AU-1601, was born at full term by normal vaginal delivery after an unremarkable pregnancy. Labor lasted more than 21 hours. Her birth weight was 2.8 kg (13$^{th}$ percentile) and she was reported to have neonatal jaundice. Her medical history was remarkable for an inguinal hernia repair at 4 months age, an unspecified adverse reaction to a BCG vaccine at 4 years of age, an infection of a “cyst in her left shoulder” at 4 years of age for which antibiotics were given for 1.5 years, and chicken pox at 8 years of age. There was no history of seizures. Her motor developmental history was reported to have been normal, but she suffered from delays in language, communication, and socialization. She was reported to have reasonably good self-care skills, and enjoyed sewing. Evaluation by Boston Children’s clinicians at age sixteen, including an ADOS assessment and consideration of DSM_IV-TR criteria, confirmed a diagnosis of autism. She had markedly reduced eye contact with very brief, infrequent gaze sharing. She used gestures and a minimal amount of oral language to communicate her basic needs and wishes. She showed high interest in single word labeling of objects, asking the examiner to name the object and then repeating the word with delayed echolalia. Sensory seeking behaviors, such as chewing her cheek, were frequently

observed, as were other stereotypical, repetitive mouth movements. Throughout the evaluation she appeared content and interested in the activities and responded well to praise by clapping with delight at her own successes. On examination, she was non-dysmorphic and had a head circumference of 53 cm (11$^{th}$ percentile). She used single words and occasional two-word phrases; word order was often incorrect. She was able to say her name and the school she attended when asked. Pupils were 5 mm bilaterally and reacted to 3 mm. Extraocular movements were full without nystagmus. There was no facial weakness. She had normal muscle tone and gait. Deep tendon reflexes were 2+ and symmetrical, and her toes were down-going bilaterally. Past laboratory investigations include complete blood count, thyroid function test, amino acid and organic acid chromatography, Fragile X DNA testing (FRAXA and FRAXE) and karyotype (46, XX), all of which were reported to be within normal limits.

AU-1601 has four full siblings, three of whom were autistic. An older brother was reported to have developmental delay and unspecified communication problems. A younger brother was also reported to have communication and language problems as well as unspecified renal abnormalities. A younger sister was reported to have communication and language problems and decreased comprehension. Another younger brother was reported to be neurotypical. Intellectual disability, without autism, was reported in a maternal uncle. Detailed medical and developmental histories were not available on these other affected family members.

**VPS13B/COH1 p.A3943fs Family (AU-21100) Clinical Information**

At the time of evaluation by the Boston Children's Hospital team, the proband was a 9 years 8 month old boy with a history of autism and developmental delay, born to consanguineous parents. He had some receptive language but expressively was limited to single words. He had three unaffected siblings (sisters). The BCH evaluation team concurred with the diagnosis of autism. Prior investigations had made note of normal karyotyping, SignatureChip CGH microarray, and brain MRI studies. On examination he was at ~3rd percentile for length, weight, and head circumference was 49.0 cm (<<3rd percentile). His head was noted to be dysmorphic with a long head, possible bitemporal narrowing, a left preauricular tag, a prominent nasal tip with nasal septum extending below alae nasi, a protruding maxilla, short philtrum (~0.5 cm, <3rd percentile), slightly high arched palate, and poor dentition. He also had a Marfanoid appearance and hyperextensible joints. He was scoliotic. He had pes planus with medial rotation of feet. He had a beighton score of 4 or 5 (thumb x 2, knees x 2, hips unknown). His elemental neurologic exam was notable for hypotonia. DTRs were 2+ in upper and lower limbs.

**VPS13B/COH1 SSC Family Clinical Information**

Screening a cohort of 612 families from the Simons Simplex Collection who underwent exome sequencing (Sanders et al, 2012), we identified two individuals with inherited, compound heterozygous mutations in *VPS13B/COH1*. The first (12651.p1) was an affected male child carrying the heterozygous mutations p.W963X and p.G2704R (NP_689777.3), inherited one from each parent. G2740R alters a conserved residue, and was scored as probably damaging by SIFT and PolyPhen2, whereas W963X is an allele

previously reported in Cohen Syndrome (Kolehmainen et al., 2004). Review of available phenotypic information in SFARI Base indicated that this child had multiple dysmorphic features, with specific abnormalities noted in hair growth pattern, nose size, mouth and lip, philtrum, teeth, fingers, thumbs, and feet. The child did not previously carry a diagnosis of Cohen Syndrome. In a second, unrelated family, the affected male child (12100.p1) was found to carry the heterozygous mutations p.S3303R and p.A3691T, each inherited from a separate parent. Both alter highly conserved residues and are predicted to be deleterious or pathogenic by SIFT and Polyphen2. This child was not microcephalic (head circumference 54.5cm (50%), height 156.5 cm (97%) at age 11.3 years), but was noted to have dysmorphisms of the face and extremities and an abnormal hair growth pattern.

**Supplemental References**

Adli, M., and Bernstein, B.E. (2011). Whole-genome chromatin profiling from limited numbers of cells using nano-ChIP-seq. Nat Protoc *6*, 1656-1668.

Braverman, N., Chen, L., Lin, P., Obie, C., Steel, G., Douglas, P., Chakraborty, P.K., Clarke, J.T.R., Boneh, A., Moser, A.*, et al.* (2002). Mutation analysis of PEX7 in 60 probands with rhizomelic chondrodysplasia punctata and functional correlations of genotype with phenotype. Human mutation *20*, 284-297.

Cooper, G.M., Coe, B.P., Girirajan, S., Rosenfeld, J.A., Vu, T.H., Baker, C., Williams, C., Stalker, H., Hamid, R., Hannig, V.*, et al.* (2011). A copy number variation morbidity map of developmental delay. Nature Genetics *43*, 838-846.

Kent, W.J., Sugnet, C.W., Furey, T.S., Roskin, K.M., Pringle, T.H., Zahler, A.M., and Haussler, A.D. (2002). The Human Genome Browser at UCSC. Genome Research *12*, 996-1006.

Kolehmainen, J., Wilkinson, R., Lehesjoki, A.-E., Chandler, K., Kivitie-Kallio, S., Clayton-Smith, J., Träskelin, A.-L., Waris, L., Saarinen, A., Khan, J.*, et al.* (2004). Delineation of Cohen Syndrome Following a Large-Scale Genotype-Phenotype Screen. The American Journal of Human Genetics *75*, 122-127.

Li, H., and Durbin, R. (2009). Fast and Accurate Short Read Alignment with Burrows-Wheeler Transform. Bioinformatics (Oxford, England).

McKenna, A., Hanna, M., Banks, E., Sivachenko, A., Cibulskis, K., Kernytsky, A., Garimella, K., Altshuler, D., Gabriel, S., Daly, M.*, et al.* (2010). The Genome Analysis Toolkit: a MapReduce framework for analyzing next-generation DNA sequencing data. Genome Research *20*, 1297-1303.

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 (New York, NY) *321*, 218-223.

Moser, A.B., Rasmussen, M., Naidu, S., Watkins, P.A., McGuinness, M., Hajra, A.K., Chen, G., Raymond, G., Liu, A., and Gordon, D. (1995). Phenotype of patients with peroxisomal disorders subdivided into sixteen complementation groups. The Journal of pediatrics *127*, 13-22.

Okamura-Ikeda, K., Hosaka, H., Maita, N., Fujiwara, K., Yoshizawa, A.C., Nakagawa, A., and Taniguchi, H. (2010). Crystal structure of aminomethyltransferase in complex with dihydrolipoyl-H-protein of the glycine cleavage system: implications for recognition

of lipoyl protein substrate, disease-related mutations, and reaction mechanism. The Journal of biological chemistry *285*, 18684-18692.

Okamura-Ikeda, K., Hosaka, H., Yoshimura, M., Yamashita, E., Toma, S., Nakagawa, A., Fujiwara, K., Motokawa, Y., and Taniguchi, H. (2005). Crystal Structure of Human T-protein of Glycine Cleavage System at 2.0  Resolution and its Implication for Understanding Non-ketotic Hyperglycinemia. Journal of Molecular Biology *351*, 1146-1159.

Pettersen, E.F., Goddard, T.D., Huang, C.C., Couch, G.S., Greenblatt, D.M., Meng, E.C., and Ferrin, T.E. (2004). UCSF Chimera--a visualization system for exploratory research and analysis. Journal of Computational Chemistry *25*, 1605-1612.

Sanders, S.J., Ercan-Sencicek, A.G., Hus, V., Luo, R., Murtha, M.T., Moreno-De-Luca, D., Chu, S.H., Moreau, M.P., Gupta, A.R., Thomson, S.A., *et al.* (2011). Multiple Recurrent De Novo CNVs, Including Duplications of the 7q11.23 Williams Syndrome Region, Are Strongly Associated with Autism. Neuron *70*, 863-885.

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