

Targeted sequencing identifies 91 neurodevelopmental-disorder risk genes with autism and developmental-disability biases

Holly A F Stessman^{1,29,30}, Bo Xiong^{1,2,30}, Bradley P Coe¹, Tianyun Wang³, Kendra Hoekzema¹, Michaela Fenckova^{4,5}, Malin Kvarnung^{6,7}, Jennifer Gerdts⁸, Sandy Trinh⁸, Nele Cosemans⁹, Laura Vives¹, Janice Lin¹, Tychelle N Turner¹, Gijs Santen¹⁰, Claudia Ruivenkamp¹⁰, Marjolein Kriek¹⁰, Arie van Haeringen¹⁰, Emmelien Aten¹⁰, Kathryn Friend^{11,12}, Jan Liebelt¹³, Christopher Barnett¹³, Eric Haan^{11,13}, Marie Shaw¹¹, Jozef Gecz^{11,12,14}, Britt-Marie Anderlid^{6,7}, Ann Nordgren^{6,7}, Anna Lindstrand^{6,7}, Charles Schwartz¹⁵, R Frank Kooy¹⁶, Geert Vandeweyer¹⁶, Celine Helmsmoortel¹⁶, Corrado Romano¹⁷, Antonino Alberti¹⁷, Mirella Vinci¹⁸, Emanuela Avola¹⁷, Stefania Giusto¹⁹, Eric Courchesne²⁰, Tiziano Prampero²⁰, Karen Pierce²⁰, Srinivasa Nalabolu²⁰, David G Amaral²¹, Ingrid E Scheffer^{22–24}, Martin B Delatycki^{22,25,26}, Paul J Lockhart^{22,26}, Fereydoun Hormozdiari²⁷, Benjamin Harich^{4,5}, Anna Castells-Nobau^{4,5}, Kun Xia³, Hilde Peeters⁹, Magnus Nordenskjöld^{6,7}, Annette Schenck^{4,5}, Raphael A Bernier⁸ & Evan E Eichler^{1,28}

Gene-disruptive mutations contribute to the biology of neurodevelopmental disorders (NDDs), but most of the related pathogenic genes are not known. We sequenced 208 candidate genes from >11,730 cases and >2,867 controls. We identified 91 genes, including 38 new NDD genes, with an excess of *de novo* mutations or private disruptive mutations in 5.7% of cases. *Drosophila* functional assays revealed a subset with increased involvement in NDDs. We identified 25 genes showing a bias for autism versus intellectual disability and highlighted a network associated with high-functioning autism (full-scale IQ >100). Clinical follow-up for *NAA15*, *KMT5B*, and *ASH1L* highlighted new syndromic and nonsyndromic forms of disease.

NDDs are a heterogeneous collection of psychiatric and clinical diagnoses that encompass autism spectrum disorder (ASD), intellectual disability/developmental delay (ID/DD), attention-deficit/hyperactivity, motor and tic disorders, and language communication disorders¹. Although each diagnosis is distinct in the Diagnostic and Statistical Manual of Mental Disorders, 5th edition (DSM-5)¹, NDDs often co-occur. Twin studies have shown that

NDDs have a heritable component². Beyond phenotypic overlaps, risk genotypes overlap among NDDs, as identified in copy-number variant (CNV) studies³. Although these data strongly suggest that common genetic etiologies underlie a subset of broadly defined NDDs, there has been criticism that gene-discovery efforts have failed to distinguish ID/DD genes from those contributing to ASD without ID⁴.

¹Department of Genome Sciences, University of Washington, Seattle, Washington, USA. ²Department of Forensic Medicine and Institute of Brain Research, Tongji Medical College, Huazhong University of Science and Technology, Wuhan, China. ³State Key Laboratory of Medical Genetics, School of Life Sciences, Central South University, Changsha, China. ⁴Department of Human Genetics, Radboud University Medical Center, Nijmegen, the Netherlands. ⁵Donders Institute for Brain, Cognition and Behaviour, Radboud University Medical Center, Nijmegen, the Netherlands. ⁶Department of Molecular Medicine and Surgery, Center for Molecular Medicine, Karolinska Institutet, Stockholm, Sweden. ⁷Department of Clinical Genetics, Karolinska University Hospital, Stockholm, Sweden. ⁸Department of Psychiatry and Behavioral Sciences, University of Washington, Seattle, Washington, USA. ⁹Centre for Human Genetics, KU Leuven and Leuven Autism Research (LAuRes), Leuven, Belgium. ¹⁰Department of Clinical Genetics, Leiden University Medical Center (LUMC), Leiden, the Netherlands. ¹¹School of Medicine and the Robinson Research Institute, the University of Adelaide at the Women's and Children's Hospital, Adelaide, South Australia, Australia. ¹²Genetics and Molecular Pathology, SA Pathology, Adelaide, South Australia, Australia. ¹³South Australian Clinical Genetics Service, SA Pathology (at the Women's and Children's Hospital), Adelaide, South Australia, Australia. ¹⁴South Australian Health and Medical Research Institute, Adelaide, South Australia, Australia. ¹⁵Center for Molecular Studies, J.C. Self Research Institute of Human Genetics, Greenwood Genetic Center, Greenwood, South Carolina, USA. ¹⁶Department of Medical Genetics, University of Antwerp, Antwerp, Belgium. ¹⁷Unit of Pediatrics & Medical Genetics, IRCCS Associazione Oasi Maria Santissima, Troina, Italy. ¹⁸Laboratory of Medical Genetics, IRCCS Associazione Oasi Maria Santissima, Troina, Italy. ¹⁹Unit of Neurology, IRCCS Associazione Oasi Maria Santissima, Troina, Italy. ²⁰Department of Neurosciences, UC San Diego Autism Center, School of Medicine, University of California San Diego, La Jolla, California, USA. ²¹MIND Institute and the University of California Davis School of Medicine, Sacramento, California, USA. ²²Department of Paediatrics, University of Melbourne, Royal Children's Hospital, Melbourne, Victoria, Australia. ²³Department of Medicine, University of Melbourne, Austin Health, Melbourne, Victoria, Australia. ²⁴Florey Institute of Neuroscience and Mental Health, Parkville, Victoria, Australia. ²⁵Victorian Clinical Genetics Services, Parkville, Victoria, Australia. ²⁶Bruce Lefroy Centre for Genetic Health Research, Murdoch Children's Research Institute, Parkville, Victoria, Australia. ²⁷Department of Biochemistry and Molecular Medicine, University of California, Davis, Davis, California, USA. ²⁸Howard Hughes Medical Institute, Seattle, Washington, USA. ²⁹Present address: Department of Pharmacology, Creighton University School of Medicine, Omaha, Nebraska, USA. ³⁰These authors contributed equally to this work. Correspondence should be addressed to E.E.E. (eee@gs.washington.edu).

Received 23 September 2016; accepted 22 January 2017; published online 13 February 2017; doi:10.1038/ng.3792

Large numbers of potentially pathogenic mutations have been identified through exome sequencing of NDD cohorts, but in most cases only a single occurrence of a *de novo* mutation in a particular gene has been discovered. Substantially larger numbers of cases and controls are required to demonstrate the statistical significance of individual genes. Leveraging samples and data from multiple comorbid conditions, in principle, can increase sensitivity in identifying risk genes. Phenotypic follow-up of cases broadly drawn from NDDs has previously allowed us to explore specific clinical phenotypes in a genotype-first manner⁵, as in the case of genes such as *CHD8* (ref. 6), *DYRK1A*⁷, and *ADNP*⁸.

Here, using single-molecule molecular inversion probes (smMIPs)^{9,10}, we sequenced the coding and splicing portions of 208 potential NDD risk genes in over 11,730 ASD, ID, and DD cases. smMIPs provide a highly sensitive, specific, and inexpensive approach to sequence the protein-coding portions of a moderate number of candidate genes in a large number of cases. Samples were collected as part of an international consortium termed the Autism Spectrum/Intellectual Disability (ASID) network, involving 15 centers across seven countries and four continents. The collection has the advantage over many others in that subsequent phenotypic follow-up is possible for a large fraction of the cases.

RESULTS

Mutation discovery

We selected 208 candidate NDD (ASD, ID, and DD) disease-risk genes, on the basis of published sequencing studies^{11–17} (Supplementary Tables 1–3), by using denovo-db¹⁸. Genes were selected and ranked according to the number of published *de novo* recurrences, overlap with a CNV morbidity map¹⁹, pathway connectivity²⁰, and absence of *de novo* variants in 1,909 published unaffected-sibling control exomes^{12,13}. We designed 12,016 smMIPs distributed across four smMIP pools to cover all annotated RefSeq coding exons as well as 5 bp of flanking intronic sequence (Supplementary Tables 4–7 and Online Methods). We targeted these genes for sequencing in 15 large cohorts of cases (some including unaffected siblings) with a primary ascertainment diagnosis of ASD, ID, or DD, for which exome sequencing had not previously been performed (Supplementary Tables 8 and 9). The set included 6,342 cases with a primary diagnosis of ASD and

7,065 cases with a diagnosis of ID/DD, from a large international collaboration between research and clinical investigators from the United States, Belgium, the Netherlands, Sweden, Italy, China, and Australia (Fig. 1).

After quality control (QC) (Online Methods, Supplementary Table 10 and Supplementary Figs. 1–4), we identified 61,315 QC-passing variants, excluding common dbSNP variants. Of these, 2,185 were private (i.e., found in only one family in the study; Supplementary Tables 10 and 11) and potentially deleterious (for example, nonsense, stop-gain, start-loss, frameshift, or disruptive splicing mutations) or missense events with a combined annotation-dependent depletion (CADD) score >30 (MIS30). The number of private, high-impact events identified in probands was significantly greater than that in unaffected siblings in the study (false discovery rate (FDR)-corrected $P = 1.44 \times 10^{-9}$; two-tailed Fisher's exact test; Supplementary Fig. 5a,b), as expected^{9,11}. This signal was primarily driven by likely gene-disruptive (LGD) events (corrected $P = 9.20 \times 10^{-15}$) and not MIS30 events (corrected $P = 0.83$). We validated 1,125 variants, including all private LGD events as well as 25% of the private MIS30 events, by Sanger sequencing (validation rate >97%; Supplementary Table 11).

Genes with an excess of severe *de novo* mutations

We assessed inheritance for 286 of the private variants, 35% of which were determined to be sporadic mutations (Supplementary Tables 12 and 13, and Supplementary Fig. 5c). The set represented 91 private *de novo* mutations—82 LGD and 9 MIS30—among cases and 9 *de novo* mutations—3 LGD and 6 MIS30—among unaffected siblings, and included 35 recently reported events^{11,12,21} (Supplementary Table 12 and Supplementary Fig. 5d). Allowing for an allele count (AC) ≤ 3 , we identified an additional 32 *de novo* LGD and 15 *de novo* MIS30 events in probands, for a total of 138 *de novo* proband events (114 LGD and 24 MIS30; Supplementary Table 12). Using a probabilistic model derived from human–chimpanzee divergence and an expected rate of 1.5 *de novo* mutations per exome^{9,22}, we calculated the overall probability of detecting 114 or more *de novo* LGD and 24 *de novo* MIS30 variants in our panel of 208 genes as $P = 1.6 \times 10^{-22}$ (one-tailed binomial test) with an odds ratio (OR) of 2.62 (95% confidence interval (CI) 2.2–3.09). By combining these results (Supplementary Table 12) with published exome data sets (Supplementary Tables 1 and 2),



Figure 1 ASID patient network. Probands ($n = 13,407$) with a primary diagnosis of ASD, ID, or DD, collected from 15 international groups, were screened on the basis of smMIPs. Circle size corresponds to the number of samples screened for each cohort. Cohort numbers 1–15 correspond to Supplementary Table 8.

Table 1 Genes reaching *de novo* significance

Gene	smMIP screening			Published exomes			Total		FDR-corrected <i>de novo</i> <i>P</i> value ^a		Study
	DN LGD	DN MIS30	Probands screened	DN LGD	DN MIS30	Probands screened	DN LGD	DN MIS30	LGD	MIS30	
SCN2A	10	1	13,407	11	6	5,237	21	7	8.45×10^{-45}	1.27×10^{-12}	Gold
ARID1B	6	0	13,407	16	0	5,237	22	0	9.84×10^{-42}	1	Gold
ADNP	7	0	13,407	9	0	6,158	16	0	9.57×10^{-34}	1	Gold
CHD8	5	0	13,407	12	2	6,158	17	2	4.26×10^{-33}	9.48×10^{-3}	Gold
SYNGAP1	1	0	13,407	16	1	6,158	17	1	2.52×10^{-27}	0.19	Gold
POGZ	4	1	13,407	5	1	5,237	9	2	2.06×10^{-22}	4.95×10^{-4}	Gold
DYRK1A	2	0	13,407	8	0	6,158	10	0	1.32×10^{-19}	1	Gold
CTNNB1	0	0	13,407	7	0	6,158	7	0	4.34×10^{-13}	1	Gold
ANKRD11	0	0	12,192	10	0	5,237	10	0	4.34×10^{-13}	1	ASD5
NAA15	4	0	12,192	2	0	5,237	6	0	1.52×10^{-12}	1	ASD4
MED13L	2	0	13,407	6	2	6,158	8	2	1.98×10^{-12}	1.40×10^{-2}	Gold
FOXP1	2	0	13,407	4	0	5,237	6	0	5.90×10^{-12}	1	Gold
TCF4	4	0	12,192	3	1	5,237	7	1	6.26×10^{-12}	0.22	ASD4
STXBP1	1	0	13,407	5	2	6,158	6	2	2.03×10^{-11}	7.73×10^{-3}	Gold
MECP2	2	0	11,731	3	1	5,237	5	1	4.12×10^{-11}	1.94×10^{-2}	ASD6
GRIN2B	2	1	13,407	4	1	6,158	6	2	6.39×10^{-10}	9.56×10^{-3}	Gold
WAC	1	0	13,407	4	0	6,158	5	0	9.63×10^{-10}	1	Gold
KMT5B	2	1	12,192	3	1	5,237	5	2	1.21×10^{-9}	1.93×10^{-3}	ASD4
TRIP12	2	0	13,407	4	2	6,158	6	2	1.84×10^{-9}	1.80×10^{-2}	Gold
CHD2	0	3	13,407	7	1	6,158	7	4	1.98×10^{-8}	7.66×10^{-4}	Gold
DSCAM	2	0	13,407	4	0	6,158	6	0	7.89×10^{-8}	1	Gold
KMT2A	0	0	11,731	6	0	5,237	6	0	8.57×10^{-8}	1	ASD6
CUL3	2	0	13,407	2	0	5,237	4	0	1.63×10^{-7}	1	Gold
ASH1L	2	2	13,407	3	0	5,237	5	2	2.26×10^{-7}	1.80×10^{-2}	Gold
TCF7L2	0	0	13,407	3	0	6,158	3	0	6.97×10^{-5}	1	Gold
ILF2	0	0	11,731	2	0	5,237	2	0	7.01×10^{-5}	1	ASD6
DDX3X	0	0	13,407	3	1	5,237	3	1	9.95×10^{-5}	0.15	Gold
RIMS1	1	0	12,192	2	0	5,237	3	0	9.95×10^{-5}	1	ASD4
KATNAL2	1	1	13,407	2	0	5,237	3	1	1.00×10^{-4}	0.19	Gold
NCKAP1	1	0	12,192	2	0	6,158	3	0	1.43×10^{-4}	1	ASD5
SETBP1	0	0	13,407	3	0	5,237	3	0	1.50×10^{-4}	1	Gold
SLC6A1	2	0	13,407	1	2	6,158	3	2	1.75×10^{-4}	2.49×10^{-2}	Gold
WDR45	0	0	11,731	2	0	5,237	2	0	1.75×10^{-4}	1	ASD6
SPAST	0	0	12,192	2	0	5,237	2	0	1.80×10^{-4}	1	ASD4
PTEN	0	0	13,407	2	0	6,158	2	0	2.46×10^{-4}	1	Gold
MYT1L	1	0	11,731	2	0	6,158	3	0	3.29×10^{-4}	1	ASD6
TNRC6B	1	0	13,407	2	0	5,237	3	0	3.49×10^{-4}	1	Gold
SETD5	0	1	13,407	3	0	6,158	3	1	3.61×10^{-4}	0.14	Gold
NRXN1	0	0	11,731	3	0	6,158	3	0	3.69×10^{-4}	1	ASD6
TBR1	0	0	13,407	2	0	6,158	2	0	5.21×10^{-4}	1	Gold
CASK	0	0	12,192	2	2	5,237	2	2	1.09×10^{-3}	2.98×10^{-3}	ASD4
PAX5	0	0	13,407	2	0	6,158	2	0	1.13×10^{-3}	1	Gold
PPM1D	0	0	13,407	2	0	6,158	2	0	2.73×10^{-3}	1	Gold
ANK2	0	1	12,192	4	1	5,237	4	2	2.93×10^{-3}	0.19	ASD5
HIVEP3	2	0	13,407	1	0	5,237	3	0	3.07×10^{-3}	1	Gold
CDC42BPB	0	0	12,192	2	0	5,237	2	0	3.51×10^{-3}	1	ASD4
CACNA2D3	0	0	11,731	2	0	5,237	2	0	3.54×10^{-3}	1	ASD6
GIGYF2	1	0	13,407	1	0	5,237	2	0	3.57×10^{-3}	1	Gold
DLG4	1	0	12,192	1	0	5,237	2	0	3.77×10^{-3}	1	ASD5
SMC3	1	0	12,192	1	1	5,237	2	1	4.51×10^{-3}	0.19	ASD4
KMT2E	0	0	12,192	2	0	5,237	2	0	7.16×10^{-3}	1	ASD4
PARD3B	1	0	12,192	1	0	5,237	2	0	7.79×10^{-3}	1	ASD5
PTK7	1	0	12,192	1	0	5,237	2	0	9.99×10^{-3}	1	ASD4
SRCAP	1	0	12,192	1	0	5,237	2	0	1.25×10^{-2}	1	ASD4
RELN	2	1	13,407	1	2	5,237	3	3	1.27×10^{-2}	1.40×10^{-2}	Gold
PHF2	0	0	12,192	2	0	5,237	2	0	1.60×10^{-2}	1	ASD5
ZC3H4	0	0	13,407	2	0	5,237	2	0	1.75×10^{-2}	1	Gold
SETD2	0	0	13,407	2	0	5,237	2	0	1.77×10^{-2}	1	Gold

(continued)

Table 1 Genes reaching *de novo* significance (continued)

Gene	smMIP screening			Published exomes			Total		FDR-corrected <i>de novo</i> <i>P</i> value ^a		Study
	DN LGD	DN MIS30	Probands screened	DN LGD	DN MIS30	Probands screened	DN LGD	DN MIS30	LGD	MIS30	
<i>DIP2A</i>	0	0	13,407	2	0	5,237	2	0	1.79×10^{-2}	1	Gold
<i>UNC80</i>	1	0	12,192	1	0	5,237	2	0	1.83×10^{-2}	1	ASD5
<i>ZNF292</i>	1	0	11,731	1	0	5,237	2	0	1.88×10^{-2}	1	ASD6
<i>PHIP</i>	1	0	13,407	1	0	5,237	2	0	1.88×10^{-2}	1	Gold
<i>WDFY3*</i>	0	0	12,192	2	0	5,237	2	0	1.93×10^{-2}	1	ASD5
<i>PLXNB1</i>	1	1	12,192	1	0	5,237	2	1	1.93×10^{-2}	0.4	ASD5
<i>ASXL3</i>	0	0	11,731	2	0	5,237	2	0	2.17×10^{-2}	1	ASD6
<i>LAMC3</i>	0	0	13,407	2	0	5,237	2	0	2.73×10^{-2}	1	Gold
<i>DOCK8</i>	1	0	11,731	1	0	5,237	2	0	5.40×10^{-2}	1	ASD6
<i>KMT2C</i>	0	0	11,731	2	0	5,237	2	0	7.15×10^{-2}	1	ASD6
<i>COL4A3BP</i>	0	0	13,407	0	4	5,237	0	4	1	1.75×10^{-7}	Gold
<i>PPP2R5D</i>	0	0	13,407	1	3	6,158	1	3	1.40×10^{-2}	1.04×10^{-6}	Gold
<i>TRIO</i>	0	1	13,407	1	3	5,237	1	4	0.18	1.44×10^{-4}	Gold
<i>TBL1XR1</i>	0	0	13,407	1	2	6,158	1	2	1.45×10^{-2}	6.37×10^{-4}	Gold
<i>PTPN11</i>	0	0	12,192	0	2	5,237	0	2	1	3.93×10^{-3}	ASD4
<i>DLGAP1</i>	0	1	12,192	0	1	5,237	0	2	1	5.10×10^{-3}	ASD4
<i>TANC2</i>	0	1	12,192	1	1	5,237	1	2	0.11	1.36×10^{-2}	ASD5
<i>SRGAP3</i>	0	1	12,192	0	1	5,237	0	2	1	1.47×10^{-2}	ASD5
<i>ITPR1</i>	0	1	11,731	0	2	5,237	0	3	1	1.80×10^{-2}	ASD6
<i>ADCY5</i>	0	0	12,192	0	2	5,237	0	2	1	1.94×10^{-2}	ASD5

DN LGD significant genes are shown first, followed by DN MIS30 significant genes, from most to least significant.

^aFDR corrections (on the one-tailed binomial test DN *P* value) were based on the number of samples for which parental DNA could be tested. *An LGD variant was identified in this gene by using previously published smMIPs; therefore, the LGD count differs compared with those in **Supplementary Table 11**, to avoid duplicate counting. DN, *de novo*.

we identified a total of 393 *de novo* LGD and 98 *de novo* MIS30 events in 208 screened genes, thus increasing the significance ($P = 1.28 \times 10^{-218}$; OR = 6.46 (95% CI 5.89–7.06)). Excluding known high-risk NDD genes (**Supplementary Table 3**), we recalculated the probability of identifying 136 or more *de novo* LGD and 13 *de novo* MIS30 variants among the 84 unknown genes (**Supplementary Table 3**) in which at least one *de novo* LGD mutation has been identified. The frequency of *de novo* mutations was significantly increased in probands ($P = 1.32 \times 10^{-55}$, OR = 5.12 (95% CI 4.33–6.01)), thus suggesting that many of these remaining genes contribute to NDD pathology.

Combining both smMIPs and exome sequence data, we identified 68 genes that reached *de novo* significance for LGD mutations and 23 genes that reached *de novo* significance for MIS30 mutations, at the level of the individual gene ($q < 0.1$ by binomial test and more than one LGD or MIS30 event in probands; **Table 1**, **Fig. 2a–c**, **Supplementary Fig. 6**, and **Supplementary Table 14**). Thirteen genes were significant for both *de novo* LGD and MIS30 genes; thus, 78 unique genes showed an excess of *de novo* mutations in cases (**Table 1**). Ten (13%) of these genes were unique to the MIS30 category for probands: *TANC2*, *TRIO*, *COL4A3BP*, *TBL1XR1*, *PPP2R5D*, *DLGAP1*, *SRGAP3*, *PTPN11*, *ADCY5*, and *ITPR1* (**Table 1** and **Supplementary Table 15**). Thirty-nine of the *de novo* LGD and seven of the *de novo* MIS30 significant genes have been previously linked to NDDs in the literature (**Table 1** and **Supplementary Table 15**). Of the 78 *de novo* significant genes, 32 have not been previously described as being associated with NDD phenotypes. The most significant of these genes were *TRIP12*, *KMT5B*, and *ASH1L* (**Fig. 3**), which were significant for both *de novo* LGD and MIS30 mutations, and *NAA15* and *DSCAM*¹¹, which were significant for only *de novo* LGD mutations (corrected $P < 1 \times 10^{-6}$). The most frequently *de novo*-mutated genes in this study were *SCN2A*, *ADNP*, *CHD8*, *DYRK1A*, and *POGZ* (**Supplementary Table 13**). *De novo* mutations in *NAA15* were also seen as frequently as those in *DYRK1A* and *POGZ* (**Supplementary Table 13**). No genes reached *de novo* LGD significance in unaffected-sibling controls, although

one gene, *TRRAP*, reached *de novo* MIS30 significance among the controls. Although it is possible that *de novo* mutation of this gene is protective, it is more likely that *TRRAP* represents a false positive, possibly because of a mutation rate higher than expected from our statistical model.

Inherited mutations and burden

The majority of validated LGD and MIS30 private variants were inherited (65%) from either mothers (33.2%) or fathers (31.8%) (**Supplementary Table 13** and **Supplementary Fig. 5e**). Among these variants, together with additional ultrarare ($AC \leq 3$) inherited events (**Supplementary Table 16**) and published private inherited counts from exome sequencing of the Simons Simplex Collection (SSC)¹³ (**Supplementary Table 17**), we observed a nominally significant maternal transmission bias for LGD (but not MIS30) events ($P = 0.037$, binomial test, with sex-chromosome events excluded). Although this analysis was underpowered to detect specific genes at the single-gene level, several genes showed an increased number of maternal transmissions ($\geq 3:1$ ratio; i.e., *AHNAK*, *DSCAM*, *NRXN1*, *NISCH*, *UIMC1*, *PLXNB1*, *PROX2*, *CHD1*, *TNRC18*, *PTK7*, and *MOV10*; **Supplementary Fig. 7**).

We also estimated the burden of private LGD and MIS30 variants at the single-gene level, regardless of inheritance status, by comparison with controls from the Exome Aggregation Consortium (ExAC) database, in which neuropsychiatric cases had been excluded ($n = 45,376$; **Supplementary Table 18**). Separate simulations of LGD and MIS30 events identified 30 and 13 genes with a significant LGD burden and MIS30 burden, respectively (1×10^6 simulations with Benjamini–Hochberg correction; **Table 2** and **Fig. 2d,e**). Four genes (*FOXPI*, *GRIN2B*, *SCN2A*, and *SETD5*) were significant for both LGD and MIS30 burden and are well-established NDD genes. Interestingly, 18 genes (eight LGD and ten MIS30) had a significant burden of private disruptive mutations in this study but did not reach *de novo* significance, probably because of our inability to test inheritance for

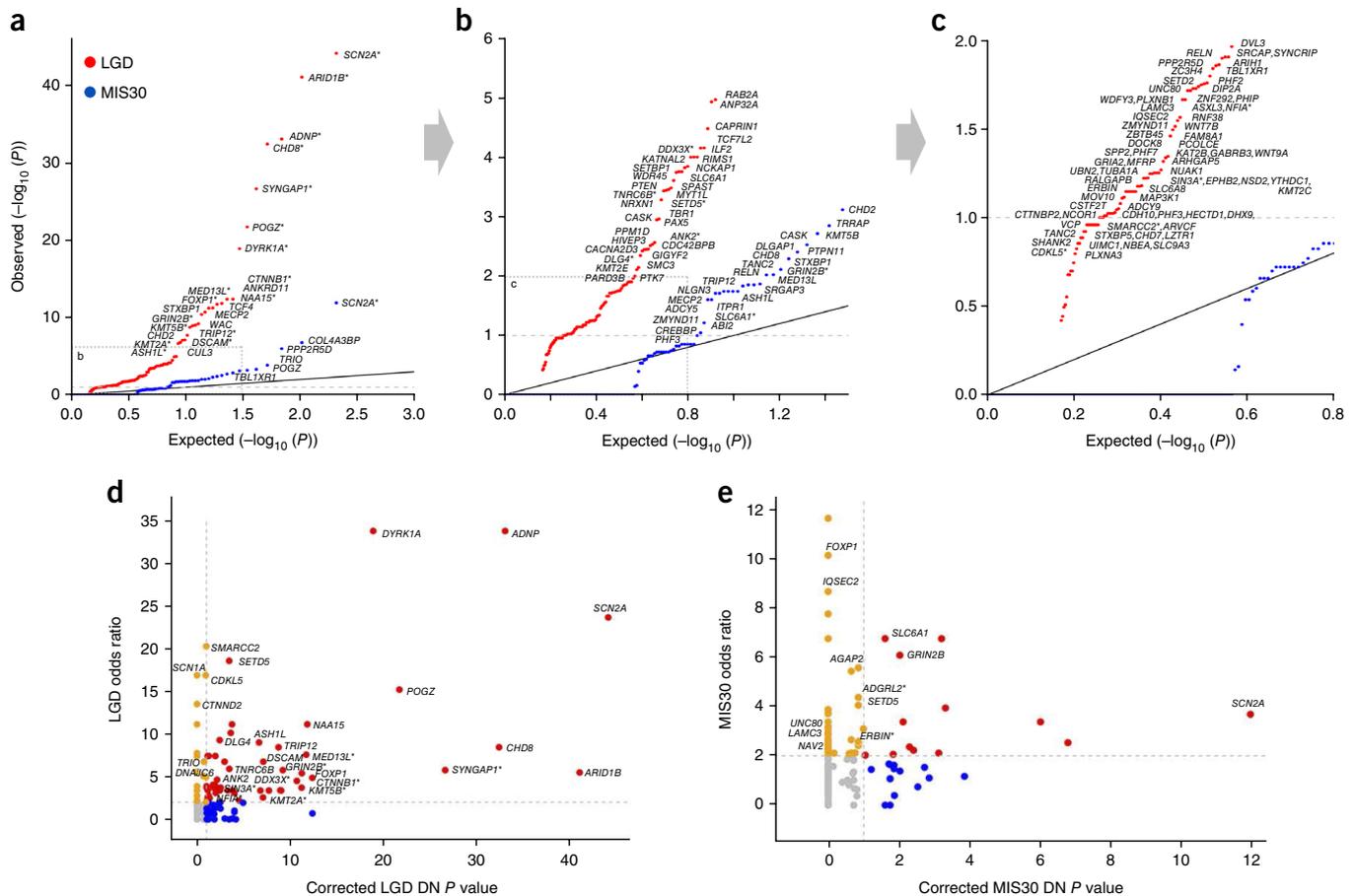


Figure 2 Targeted sequencing highlights genes reaching significance for *de novo* mutations and private disruptive variant burden. (a–c) Quantile-quantile plots comparing the probability (FDR corrected, inverse log transformed) of recurrent *de novo* mutation for individual genes among proband samples compared with a uniform distribution, given the number of genes tested (dashed gray line, significance threshold). Gray dashed boxes in **a** and **b** are shown in zoom view in **b** and **c**, respectively. Asterisk, genes that reached significance for mutation burden. (d,e) Scatter plots depicting the OR for private variants compared with unaffected controls from ExAC (y axis) versus the FDR-corrected *de novo* P value (one-tailed binomial test) (x axis; values are inverse log transformed for plotting) according to gene. Gray lines, significance threshold for the *de novo* P value (horizontal) and an OR of 2 (vertical). Genes are classified as *de novo* significant and OR >2 (red dots), OR >2 only (orange), and showing a significant *de novo* P value only (blue). Gene-name labels indicate a significant burden (FDR $q < 0.1$, simulation test) of either private LGD (d) or MIS30 (e) mutations in probands (Table 2 and Online Methods). Asterisk, genes in which no control counts were observed when the 95% lower confidence bound was used as the most conservative OR estimate. Underlying data are in Supplementary Table 14.

all events. Although *de novo* mutations in some of these genes have already been implicated in other NDD studies (for example, *FOXP1*, *TRIO*, *SCN1A*, *SIN3A*, and *IQSEC2*)^{23–27}, for others (*CTNND2*, *NAV2*, and *UNC80*), many of the severe mutations in pedigrees are inherited (Supplementary Tables 11 and 16). Given their involvement in neuronal function, axonal projection, dendrite spine formation and oligodendrocyte differentiation^{28–31}, these genes probably begin to define a class of inherited high-impact risk factors.

Autism versus intellectual disability and developmental delay

To determine whether individual genes showed a bias for clinical phenotype, we performed a separate burden analysis by using primary ascertainment diagnoses (i.e., ASD or ID (including DD diagnoses, per DSM-5 criteria)) combined with data from previous NDD studies¹³ (Supplementary Tables 2, 11, and 17). We identified 25 genes showing a bias for primary diagnosis (two one-tailed binomial tests, $P < 0.025$ for either ASD or ID/DD cases), considering both LGD and MIS30 (Fig. 4a). Eight genes had an ASD bias (*CHD2*, *CTTNBP2*, *CHD8*, *LAMC3*, *DIP2A*, *RELN*, *UNC80*, and *IQGAP3*). Of these, only *CHD8*, *CHD2*, and *DIP2A* have previously been implicated as

high-risk ASD loci³². Of the 17 ID/DD-biased genes, *NAA15*, *ZMYM2*, *PHIP*, and *STAG1* have not previously been linked to these phenotypes. We further separated the LGD and MIS30 events and identified additional significant genes for each mutation type, notably a bias for *CDH10* LGD and *NEMF* MIS30 mutations in ASD (Supplementary Fig. 8a) and *SCN1A* LGD and *NRXN1* MIS30 mutations in ID/DD (Supplementary Fig. 8b). Most genes, however, were mutated in both conditions, thus further highlighting the substantial genetic overlap between these comorbid conditions.

Phenotypic assessment of new risk genes

We recontacted individuals with mutations in *NAA15*, *KMT5B*, and *ASH1L* for further follow-up. We identified 12 LGD variants and one MIS30 private variant (Supplementary Table 11) in *NAA15* through our smMIP screening (Fig. 3a) and determined that four LGD mutations were sporadic, whereas two LGD variants, including a C-terminal mutation, were inherited (Supplementary Table 11). *NAA15* shows a burden of LGD events in cases (Table 2) as well as an excess of *de novo* LGD variants (Table 1). The gene *NAA15* encodes a protein that is a component of the NATA N-acetyltransferase complex, which includes *NAA10*—a

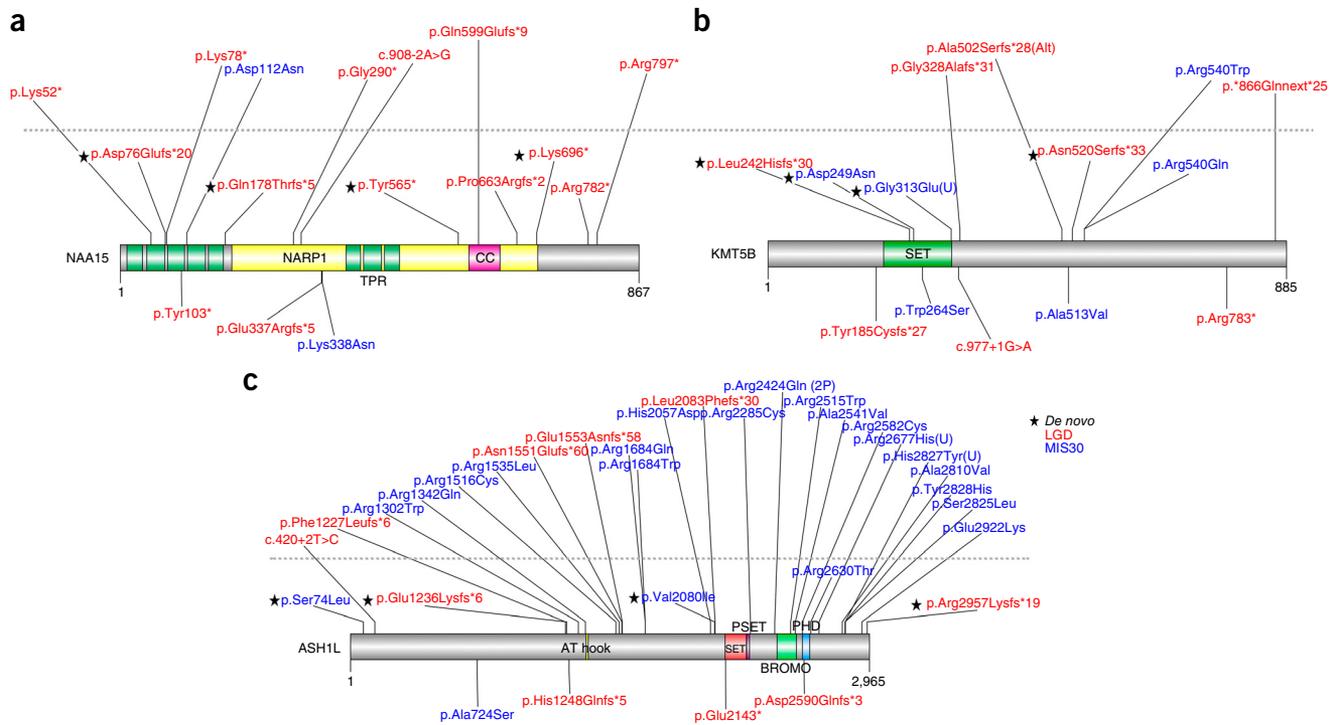


Figure 3 Protein locations of private disruptive variants in new candidate NDD risk genes. (a–c) Protein diagrams of NAA15 (a), KMT5B (b), and ASH1L (c), with novel private LGD and MIS30 mutations identified in this study and published *de novo* variants indicated in Human Genome Variation Society format. Annotated protein domains are shown (colored blocks) for the largest protein isoforms. Previously published *de novo* variants (annotated below the protein structure; **Supplementary Table 2**) are compared with new variants in this study (top). Variants above the dashed line are of unknown inheritance; variants below the line have been validated for inheritance. Domain abbreviations: NARP1, NMDA receptor–regulated protein 1; CC, coiled coil; TRP, tetratricopeptide-repeat region; PHD, plant homeodomain.

protein that is associated with Ogden syndrome as well as nonsyndromic DD³³ and is thought to tether the complex to the ribosome for post-translational modification of proteins as they exit the ribosome³⁴. To identify additional patients for clinical recontact, we relaxed our variant filter to allow for ultrarare ($AC \leq 3$) alleles. In total, we collected clinical information for ten probands with private variants and three probands with ultrarare variants in *NAA15* (**Supplementary Table 19**). Cases in our study with *NAA15* LGD and MIS30 mutations shared phenotypic features, including ID (10/11 cases; 91%), speech delay (5/6 cases; 83%), ASD diagnosis (formal diagnosis in 5/8 cases (63%) with ASD-like traits observed in two additional cases), and nonspecific growth abnormalities (for example, microcephaly, macrocephaly, and hypertelorism) (**Supplementary Table 19**). Given the incidence of DD (5.12%) in the general population^{35,36}, we estimated the penetrance of LGD *NAA15* mutations to be substantial at 35.3% (95% CI 15.7–63.6%).

Both *KMT5B* and *ASH1L* encode histone lysine *N*-methyltransferase proteins thought to be important in chromatin modification, occupancy and gene regulation. Although the roles of these genes in NDDs have not been established, a paralog of *ASH1L*, *SETBP1*, has been reported to be mutated in NDD and to be associated with ID and loss of expressive language¹⁹. We identified two *de novo* LGD mutations and two *de novo* MIS30 mutations in *KMT5B* in this study (**Supplementary Table 12**) in addition to three published *de novo* LGD variants and one *de novo* MIS30 variant^{12,14} (**Fig. 3b** and **Supplementary Table 2**). We were able to collect clinical information from three probands with private variants and four probands with ultrarare variants in *KMT5B* (**Supplementary Table 20**). Individuals with disruptive variation in *KMT5B* shared features such as ID/DD (7/7 cases; 100%), ASD diagnosis (5/6 cases; 83%), language delay (3/4

cases; 75%), motor delay (3/5 cases; 60%), and febrile seizures (3/5 cases; 60%). Attention deficits were also observed in three of these patients (**Supplementary Table 20**). For *ASH1L*, we identified two *de novo* LGD mutations and two *de novo* MIS30 mutations in addition to the three previously published *de novo* LGD mutations^{13,14}. We identified many additional LGD and MIS30 variants in *ASH1L* for which parental DNA was not available (**Fig. 3c**) and found that mutations clustered around the known annotated protein domains. We were able to obtain clinical information for two probands carrying private variants and three probands carrying ultrarare *ASH1L* variants. Individuals with *ASH1L* disruptive variation in this study had ID (5/5 cases; 100%), ASD (2/3 cases; 67%), and evidence of seizures (2/3 cases; 67%) (**Supplementary Table 21**).

Phenotypic comparisons and a high-functioning ASD network

We selected patients with *de novo* LGD mutations in 25 of our top-ranked genes in an effort to more broadly compare phenotypic features. Of the recontacted individuals, 70% (88/125) agreed to participate in a more comprehensive phenotypic evaluation. To increase our power to detect differences among patients, grouped by gene, we combined 215 case reports from the published literature with the findings collected as part of our recontact study. We assessed the general severity of each NDD by using a modified de Vries scale (**Supplementary Table 22**) and summarized phenotypic features collected during follow-up (**Table 3**), including rates of ASD, ID, seizures, macro- and microcephaly, and congenital abnormalities as well as mean IQ measures and ASD severity.

Several specific and global patterns emerged from this combined data set, in particular, an inverse relationship between ASD and ID

Table 2 Genes carrying a significant burden of private disruptive variation in cases

Gene	LGD				MIS30			
	Case count	Control count	Corrected burden <i>P</i> value	DN LGD significant	Case count	Control count	Corrected burden <i>P</i> value	DN MIS30 significant
<i>SCN2A</i>	14	2	8.80×10^{-5}	Yes	12	11	4.14×10^{-2}	Yes
<i>MED13L</i>	10	0	8.80×10^{-5}	Yes	7	16	0.63	Yes
<i>ADNP</i>	10	1	2.20×10^{-4}	Yes	1	3	0.92	No
<i>DYRK1A</i>	10	1	2.20×10^{-4}	Yes	3	8	0.83	No
<i>GRIN2B</i>	8	0	2.26×10^{-4}	Yes	9	5	3.66×10^{-2}	Yes
<i>SETD5</i>	11	2	2.26×10^{-4}	Yes	12	10	3.66×10^{-2}	No
<i>SYNGAP1</i>	8	0	2.26×10^{-4}	Yes	3	10	0.9	No
<i>NAA15</i>	12	4	3.08×10^{-4}	Yes	1	9	1	No
<i>CTNNB1</i>	7	0	6.06×10^{-4}	Yes	3	9	0.83	No
<i>POGZ</i>	9	2	1.32×10^{-3}	Yes	7	6	0.15	Yes
<i>CHD8</i>	10	4	2.62×10^{-3}	Yes	9	22	0.63	Yes
<i>ARID1B</i>	13	8	2.63×10^{-3}	Yes	14	34	0.56	No
<i>ASH1L</i>	8	3	7.04×10^{-3}	Yes	17	35	0.35	Yes
<i>DDX3X</i>	5	0	7.04×10^{-3}	Yes	2	0	0.3	No
<i>SIN3A</i>	5	0	7.04×10^{-3}	No	7	14	0.56	No
<i>KMT5B</i>	5	0	7.04×10^{-3}	Yes	3	0	0.13	Yes
<i>SMARCC2</i>	6	1	8.26×10^{-3}	No	2	4	0.76	No
<i>SCN1A</i>	5	1	2.68×10^{-2}	No	5	11	0.66	No
<i>KMT2A</i>	4	0	2.68×10^{-2}	Yes	3	9	0.83	No
<i>CDKL5</i>	5	1	2.68×10^{-2}	No	0	0	NA	No
<i>TNRC6B</i>	7	4	3.67×10^{-2}	Yes	7	18	0.7	No
<i>DSCAM</i>	6	3	4.76×10^{-2}	Yes	12	39	0.83	No
<i>TRIO</i>	6	3	4.76×10^{-2}	No	9	26	0.76	Yes
<i>FOXP1</i>	7	5	5.84×10^{-2}	Yes	6	2	4.34×10^{-2}	No
<i>TRIP12</i>	5	2	5.84×10^{-2}	Yes	6	19	0.83	Yes
<i>DLG4</i>	5	2	5.84×10^{-2}	Yes	4	9	0.7	No
<i>CTNND2</i>	4	1	7.04×10^{-2}	No	8	31	0.98	No
<i>ANK2</i>	9	9	7.04×10^{-2}	Yes	21	58	0.62	No
<i>NFIA</i>	3	0	7.15×10^{-2}	No	2	2	0.58	No
<i>DNAJC6</i>	6	4	7.28×10^{-2}	No	10	15	0.3	No
<i>UNC80</i>	5	13	0.68	Yes	30	38	1.21×10^{-2}	No
<i>ADGRL2</i>	1	0	0.48	No	6	0	1.21×10^{-2}	No
<i>SLC6A1</i>	1	0	0.48	Yes	10	5	2.30×10^{-2}	Yes
<i>NAV2</i>	1	9	1	No	42	74	3.63×10^{-2}	No
<i>IQSEC2</i>	2	0	0.2	No	7	3	4.14×10^{-2}	No
<i>AGAP2</i>	0	2	1	No	9	6	4.14×10^{-2}	No
<i>ERBIN</i>	1	0	0.48	No	4	0	4.34×10^{-2}	No
<i>LAMC3</i>	8	19	0.53	Yes	15	19	7.03×10^{-2}	No
<i>IQGAP3</i>	11	21	0.33	No	21	34	9.87×10^{-2}	No

P values were calculated by simulating the number of private LGD or MIS30 events found in the study compared with 45,375 ExAC controls (one-tailed binomial test) and were Benjamini–Hochberg corrected for the number of genes screened in which at least one private mutation was found in cases or controls ($n = 176$). Genes significant for LGD burden are shown first, followed by MIS30-burden-significant genes, from most to least significant. Genes in this table are labeled in **Figure 2d,e**. Corrected *P* values <0.1 were considered significant.

diagnoses by gene (Pearson's $R = -0.81$, $P = 9.84 \times 10^{-7}$; **Fig. 4b**). We partitioned cases into two categories: those most strongly associated with ASD (diagnostic rate >95%) and those more strongly associated with DD. Individuals with mutations in ASD genes showed significantly lower rates of seizures ($P = 1.20 \times 10^{-4}$), congenital abnormalities ($P = 1.88 \times 10^{-2}$), and microcephaly ($P = 1.79 \times 10^{-7}$), but higher rates of macrocephaly ($P = 5.25 \times 10^{-3}$), as compared with comorbid ASD and ID genes and strong ID genes (two-tailed Fisher's exact test; **Fig. 4c**). In addition, the ASD-dominated genes showed a significant difference with respect to sex. ASD genes were more likely ($P = 1.65 \times 10^{-4}$) to affect males, showing an overall ratio of 4:1 when compared with other genes (1.2:1 male/female ratio). Although the number of individual patients per gene was still low, interestingly, several genes showed an exclusive male bias (*KDM6B*, *DSCAM*, *WDFY3*, *CHD1*, and *WDR33*; **Fig. 4c** and **Table 3**). Pathway analysis of the 11 ASD genes indicated a functional enrichment in chromatin remodeling (corrected

$P = 4.72 \times 10^{-3}$; Enrichr tool)³⁷, thus implicating a functional network specifically associated with ASD individuals without ID.

To identify additional genes that may be specifically associated with high-functioning ASD, we revisited the deep phenotyping data collected as part of the SSC and applied the MAGI network-building tool, which compares the spectrum of *de novo* mutations in probands and unaffected siblings to identify coexpression and protein–interaction networks enriched in patients^{12,13,20}. We specifically selected patients with a full-scale IQ (FSIQ) >100 ($n = 668$ SSC probands; male bias 9:1) to construct a protein–interaction network based on genes with *de novo* variants in this subset (Online Methods). One statistically significant model emerged ($P < 0.01$, simulation test; **Fig. 4d**), including 40 genes and *de novo* mutations in 31 individuals with FSIQ >100. Although it was primarily composed of *de novo* missense mutations, the network showed that *de novo* LGD mutations in *FBXW11*, *CHD1*, *CHD8*, *DOT1L*, *HDAC3*, *YTHDC1*, and *KLHDC10* may be

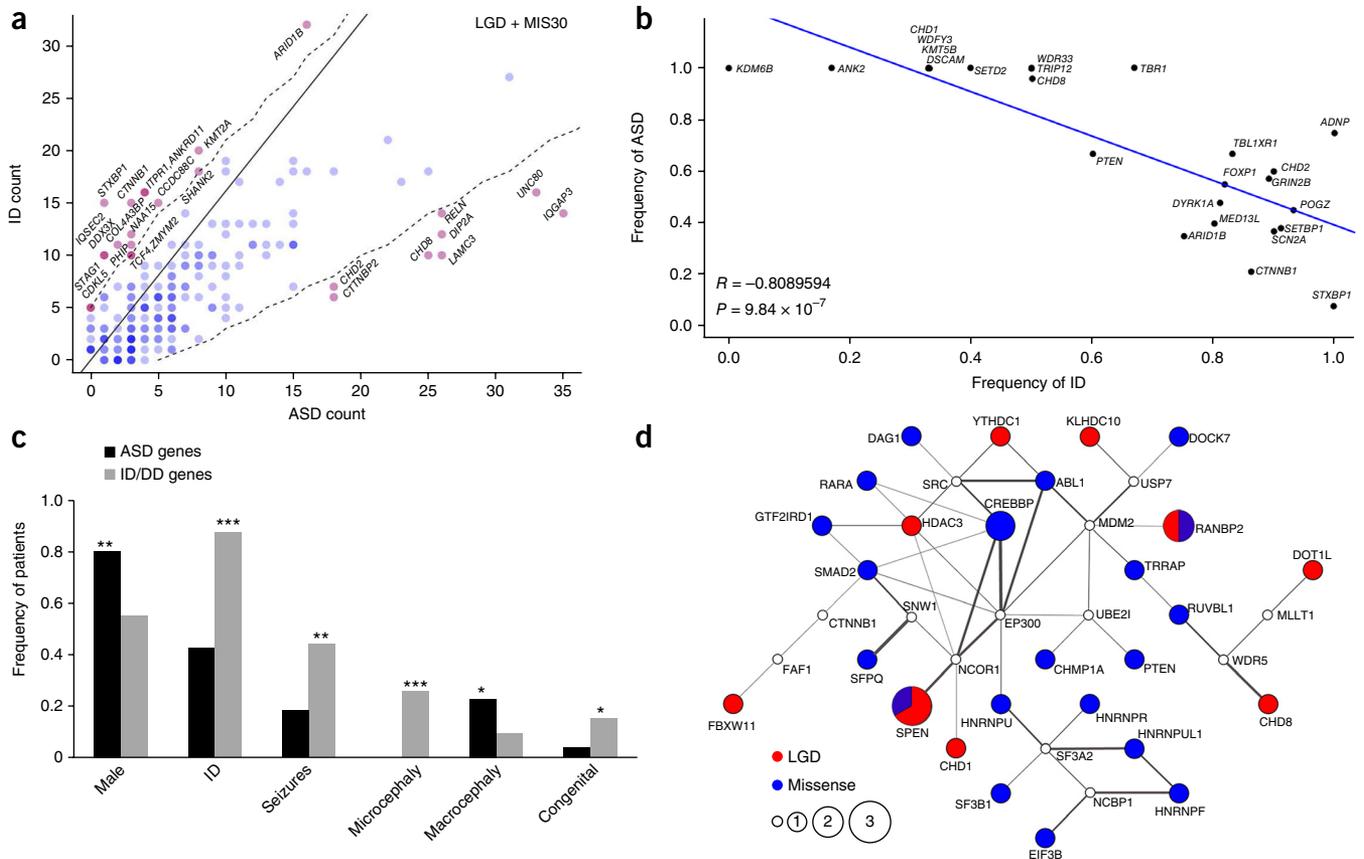


Figure 4 ASD versus ID/DD genes. (a) Probandes were categorized on the basis of primary ascertainment of either ASD or ID (including DD), and the combined numbers of LGD and MIS30 events per gene (previously published and from this study) are shown. Genes were tested for a bias toward one phenotype (ASD or ID) with two one-tailed binomial tests ($P < 0.025$ for either bias). The solid line indicates equal proportions of mutations corrected for the screened population size. Significantly biased genes (red) are indicated with respect to the threshold (dashed line) and insignificant genes (blue). Darker shades of red or blue indicate multiple genes. (b) Scatter plot showing a negative correlation (two-tailed Pearson's correlation) between ASD and ID diagnosis, according to gene (Table 3). (c) Bar graph comparing phenotypic features of patients in whom genes are primarily associated with ASD diagnosis (>95%, black bars) compared with all other genes (gray bars, representing ID/DD phenotypes) in Table 3. Significance was calculated by Fisher's two-tailed exact test, and P values were FDR corrected. Exact P values: males ($P = 1.65 \times 10^{-4}$), ID ($P = 1.20 \times 10^{-13}$), seizures ($P = 1.20 \times 10^{-4}$), microcephaly ($P = 1.79 \times 10^{-7}$), macrocephaly ($P = 5.25 \times 10^{-3}$), congenital abnormalities ($P = 1.88 \times 10^{-2}$). * $P < 0.05$; ** $P < 0.001$; *** $P < 0.0001$. (d) SSC probands with ASD and an FSIQ >100 were selected for pathway enrichment. Node size indicates the mutation score (calculated by MAGI on the basis of the number of *de novo* mutations), and the color of the node indicates the number of *de novo* LGD (red) and *de novo* missense (no CADD cutoff; blue) mutations observed in affected probands. For *SPEN*, two LGD and one missense mutation were observed, and for *RANBP2*, one LGD and one missense mutation were observed. White nodes indicate no *de novo* mutations observed. Gray lines connect genes with both protein-protein interactions and brain coexpression (Pearson's correlation coefficient $r^2 > 0.37$; Online Methods). Thicker lines correspond to more highly coexpressed gene pairs.

important in this patient subset. Both *CHD1* and *CHD8* individuals were included in our large-scale patient recontact and showed high specificity for ASD diagnosis (Table 3). A pathway analysis for this specific set of genes again implicated chromatin remodeling ($P = 0.0003$, Benjamini-Hochberg corrected Fisher's exact test (two-tailed)) as well as mRNA splicing ($P = 0.00026$) and Wnt signaling ($P = 0.03$) as potentially being important in ASD without ID.

Functional characterization of candidate genes in *Drosophila*

To provide additional functional evidence, especially with respect to nervous-system function and behavior, we performed a pilot study investigating 21 genes in *Drosophila melanogaster* (Supplementary Table 23). For 11 of these genes, *de novo* LGD mutations were significantly enriched in NDD patients. For the other ten genes, there were indications of possible association with ASD, such as a higher mutation rate in ASD cohorts or a central position in ASD gene interaction networks²⁰. Others, such as *NCKAP1* and *WDFY3*, were at the cusp of statistical significance. We used the UAS-Gal4 system and

inducible RNA interference (RNAi) lines to specifically knock down these genes in *Drosophila* neurons. When their locomotor function and overall vigor allowed (Supplementary Table 23), we subjected these knockdown flies to an ASD- and ID-relevant behavioral assay measuring light-off jump habituation, which has been shown to be affected in a number of ASD- or ID-related *Drosophila* models^{38–43}. In this assay, flies suppress their startle (jump) response to a repeated nonthreatening stimulus (light-off) as a result of experience. Their response thus gradually wanes (Fig. 5a,b). As the most fundamental evolutionarily conserved form of learning, habituation is thought to represent a prerequisite for higher cognitive functions⁴⁴. Beyond that, a number of studies have shown defective habituation of neural activity or behavior in ASD^{45–48}, and it has been proposed that disturbed habituation mechanisms may substantially contribute to defective filtering and other ASD features^{49,50}.

We first examined genes for which a significant excess of *de novo* LGD mutations (*NAA15*, *KMT5B*, *ASH1L*, and *TCF4*) was observed, and for which human phenotypic data strongly support a role in

Table 3 Key phenotypic traits across participants with gene-disrupting mutations

Gene	Total cases	Cases evaluated in depth	Mean testing age in months	Sex (% male)	Overall severity		ASD		ID		Seizures		Microcephaly		Macrocephaly		Congenital Abnormality		VIQ ^b		NVIQ ^c		ASD severity ^d	
					Modified de Vries ^a	n	Rate (%)	n	Rate (%)	n	Rate (%)	n	Rate (%)	n	Rate (%)	n	Rate (%)	Mean						
<i>KDM6B</i>	4	1	136	100	3.00	2	100	4	0	4	25	4	0	4	50	4	0	4	97.00	4	95.75	4	5.25	4
<i>ANK2</i>	6	1	141	83	5.00	2	100	6	17	6	0	6	0	6	17	6	0	6	80.50	6	81.17	6	6.83	6
<i>DSCAM</i>	3	2	133	100	6.50	2	100	3	33	3	0	3	0	3	33	3	0	3	71.33	3	55.33	3	9.00	3
<i>KMT5B</i>	3	1	115	67	7.00	1	100	3	33	3	33	3	0	3	33	3	0	3	47.33	3	54.67	3	8.33	3
<i>WDFY3</i>	3	2	181	100	6.00	3	100	3	33	3	33	3	0	3	67	3	0	3	57.33	3	77.00	3	8.00	3
<i>CHD1</i>	3	1	123	100	5.00	1	100	3	33	3	0	3	0	3	0	3	0	2	83.00	3	97.67	3	7.00	3
<i>SETD2</i>	5	2	190	60	4.25	3	100	4	40	5	40	5	0	5	40	5	0	5	100.67	3	94.67	3	7.67	3
<i>WDR33</i>	2	1	129	100	7.00	1	100	2	50	2	50	2	0	2	0	2	0	1	67.00	1	66.00	2	8.50	2
<i>TRIP12</i>	6	1	116	67	5.50	2	100	6	50	6	17	6	0	6	0	6	0	5	62.67	6	64.17	6	6.50	6
<i>TBR1</i>	6	1	102	50	5.50	2	100	6	67	6	17	6	0	6	0	6	0	4	62.25	4	60.40	5	7.20	5
<i>CHD8</i>	25	8	131	84	6.13	11	96	25	50	24	17	24	0	25	64	25	0	18	60.76	17	68.40	20	7.81	21
<i>ADNP</i>	20	4	81	65	7.40	14	75	20	100	20	25	20	10	19	10	19	10	18	33.25	4	36.00	5	6.71	7
<i>PTEN</i>	15	2	62	60	4.50	6	67	35	60	15	7	15	0	15	100	15	8	12	71.33	6	74.57	7	7.00	6
<i>TBL1XR1</i>	13	1	139	54	5.00	2	67	6	83	12	33	6	23	13	15	13	54	11	52.33	3	55.00	3	6.00	3
<i>CHD2</i>	12	4	134	50	5.75	8	60	10	90	10	83	12	10	10	10	10	25	8	71.33	6	63.33	6	8.17	6
<i>GRIN2B</i>	22	3	146	55	4.80	8	57	14	89	18	23	21	16	19	5	19	13	15	58.00	6	56.83	6	8.17	6
<i>FOXP1</i>	11	2	160	64	7.00	9	55	11	82	11	18	11	0	9	33	9	20	10	53.50	2	48.00	2	7.50	2
<i>DYRK1A</i>	21	6	195	57	7.76	17	48	21	81	21	57	21	90	21	0	21	0	21	44.29	7	51.57	7	7.43	7
<i>POGZ</i>	44	1	109	57	6.12	24	45	44	93	44	11	44	33	43	5	43	5	39	72.20	5	68.80	5	8.60	5
<i>MED13L</i>	15	1	112	53	5.44	9	40	10	80	15	20	15	20	15	7	15	36	14	57.33	3	73.67	3	6.67	3
<i>SETBP1</i>	24	1	98	58	6.20	5	38	8	91	23	71	24	9	11	0	11	18	11	60.33	3	69.67	3	6.00	3
<i>SCN2A</i>	55	2	75	53	5.14	13	37	35	90	52	75	55	25	44	2	44	21	23	52.38	8	57.25	8	6.50	8
<i>ARID1B</i>	28	3	137	43	6.45	22	35	26	75	28	32	28	7	28	14	28	21	28	67.14	7	67.00	7	7.00	7
<i>CTNWB1</i>	30	1	121	43	6.19	21	21	24	86	29	13	30	75	28	0	28	31	29	51.75	4	56.00	4	6.50	4
<i>STXBP1</i>	49	1	107	51	4.37	49	8	49	100	49	86	49	8	49	0	49	0	44	30.00	1	30.00	1	3.00	1

To maximize the number of cases for each assessment, the number of cases considered for calculated variables (*n*) differed. The *n* for each variable is listed. All *KDM6B* cases shown here have been published previously. Although originally considered for this study, *KDM6B* failed smMIP design, owing to high GC content.

^aSupplementary Table 22 shows modified de Vries scoring criteria. ^bMean verbal IQ (VIQ) had a mean of 100 and an s.d. of 15. ^cMean nonverbal IQ (NVIQ) had a mean of 100 and an s.d. of 15. ^dMean ASD severity was derived from the ADOS-2 Calibrated Severity Score (CSS) and ranged from 1 to 10; scores between 4 and 10 represent symptoms of ASD, with 10 being the most severe.

NDDs. After knockdown of *Nat1* (ortholog of *NAA15*, Vienna Drosophila RNAi Center (VDRC) no. 110689), we observed erect wings, impaired locomotor activity, and adult early lethality (within 1 or 2 d after eclosion). After knockdown of *Nat1* in a second presumably weaker RNAi line (VDRC no. 17571), flies exhibited normal morphology and locomotion; however, when challenged in the light-off jump paradigm, the flies' initial response was impaired (19% frequency of initial jumping), thus precluding proper assessment of habituation (Supplementary Table 23). These results supported *Nat1*'s crucial role in nervous-system development. *Ash1* (ortholog of *ASH1L*) neuronal-knockdown flies also showed decreased fitness, and, like *Nat1* flies, could not be scored in the habituation paradigm. *Hmt4-20* (ortholog of *KMT5B*) and *da* (ortholog of *TCF4*) flies, in contrast, were healthy overall but showed specific and significant habituation deficits (Fig. 5a,c), thus suggesting that both genes play important roles in the molecular machinery that regulates habituation learning.

We observed habituation defects after knockdown of fly homologs for several other significant *de novo*-mutated genes, including *SYNGAP1*, *GRIN2B*, and *SRCAP* (Fig. 5c and Supplementary Table 23). In addition, *dom* (*SRCAP*) and *da* (*TCF4*) flies showed substantial morphological abnormalities at the neuromuscular junction (NMJ), a well-studied synaptic model system (Supplementary Fig. 9). *NCKAP1*, *WDFY3*, and *GIGYF2* were among the tested genes with borderline significance, on the basis of our human genetic data. Significant habituation defects were observed for flies with *hem* (ortholog of *NCKAP1*), *bchs* (ortholog of *WDFY3*), and *Gyf* (ortholog of *GIGYF2*) knocked down (Fig. 5c and Supplementary Table 23). Owing to the paucity of cases, little is known regarding the clinical phenotypes associated with loss-of-function mutations of these genes; however, these functional studies suggest that they have an important role in neuronal and cognitive function.

DISCUSSION

Targeted sequencing of candidate genes in a large NDD cohort identified three overlapping categories of high-risk genes. First, we identified 68 genes that reached *de novo* LGD mutation significance, 39 of which have previously been described. Owing to limited availability of parental samples, this estimate is probably conservative. Second, we highlighted 24 genes with a significant excess of *de novo* missense mutations in NDD patients; 63% (15/24) overlap genes with *de novo* LGD

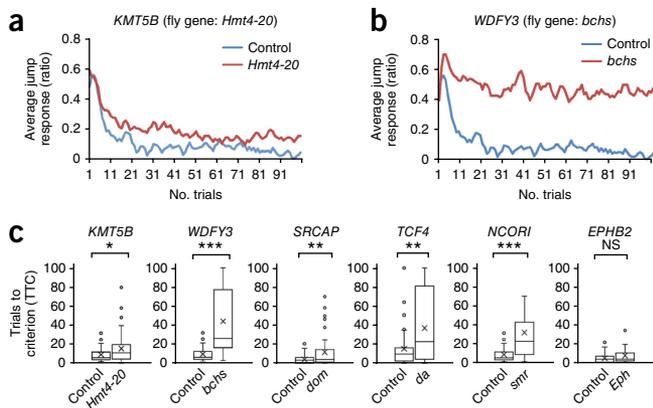


Figure 5 Habituation deficits in *Drosophila* knockdown models. (a,b) Representative jump-response curves for *Hmt4-20* (ortholog of *KMT5B*) (a) and *bchs* (ortholog of *WDFY3*) (b) pan-neuronal-knockdown flies. The ratios of flies that responded to light-off stimuli are plotted over 100 trials (in which 64 individual flies were tested for each genotype). Controls are plotted in blue, and knockdowns are plotted in red. (c) Distribution of trials to no-jump criterion (TTC, Online Methods) of knockdown versus corresponding control flies are plotted (cross, mean; center line, median; box boundaries, upper and lower quartiles; whiskers, maximum and minimum; dots, outliers). * $P < 0.05$; ** $P < 0.01$; *** $P < 0.001$; NS, not significant (linear-regression model; 64 flies tested for each genotype; exact P values in **Supplementary Table 23**).

significance (for example, *SCN2A*, *STXBP1*, *CHD2*, and *CASK*), whereas others were significant only on the basis of an excess of *de novo* MIS30 mutations (for example, *COL4A3BP*, *TRIO*, *TBL1XR1*, and *PPP2R5D*; **Table 1**), similarly to the Noonan-syndrome gene (*PTPN11*)⁵¹. Finally, 39 genes reached statistical significance on the basis of case-control burden testing (**Table 2**). Of interest were the 13 genes without *de novo*-mutation significance (*SMARCC2*, *CTNND2*, *SIN3A*, *SCN1A*, *NFIA*, *CDKL5*, *DNAJC6*, *IQSEC2*, *IQGAP3*, *ADGRL2*, *ERBIN*, *NAV2*, and *AGAP2*; **Table 2**), which may be potential inherited risk factors^{13,31,52}. In total, 44% (91/208) of our candidate genes reached locus-specific significance for disruptive mutations in 5.7% of patients, a result closely matching empirical expectations¹². However, mutation of these genes may not be necessary and sufficient to result in disease; for example, nine families (**Supplementary Fig. 10**) had disruptive mutations in two or more of the candidate genes.

Three genes without previous phenotypic information reached a high level of *de novo* significance (*NAA15*, *KMT5B*, and *ASH1L*). *NAA15* was originally identified as an *N*-methyl-D-aspartate (NMDA) glutamate-receptor-regulated gene through screens of *Nmdar1*-knockout mice⁵³. Knockdown of *Naa15* in *Drosophila* neurons caused severe locomotor defects and lethality. Missense mutations in the *NAA15*-interacting gene *NAA10* are known to cause Ogden syndrome, an X-linked disorder of infancy that can result in severe DD, craniofacial anomalies, hypotonia, cardiac arrhythmias, and, in some cases, death⁵⁴. This finding is consistent with the DD observed in our cases and the identification of *de novo* LGD mutations in people with congenital heart disease and NDDs⁵⁵. The identification of *KMT5B* and *ASH1L* highlighted the importance of histone methyltransferases, such as *EHMT56*, in ID and NDD. Mouse studies have shown that, in neurons, *Ash1l* protein represses *Nrxn1 α* protein, a known presynaptic adhesion molecule required for synaptic formation⁵⁷; mutations in *NRXN1* have been associated with ASD⁵⁸. Even less is known about the role of the *KMT5B* protein in the developing brain. However, studies have suggested that the histone H4 K20 trimethyl

mark established by the *KMT5B* protein may be involved in cell-cycle regulation in baboon neural stem progenitor cells⁵⁹. Our own analyses in *Drosophila* support roles of *NAA15* and *ASH1L* in neuronal development and of *KMT5B* and *TCF4* in habituation learning (**Supplementary Table 23**), in agreement with patient phenotypes (**Supplementary Tables 19–21**).

We designed the study such that approximately half of the cases were ascertained on the basis of a primary diagnosis of ASD, whereas the other half were diagnosed initially as ID/DD in an effort to test the diagnostic specificity of particular genes. Although most genes are clearly risk factors for NDD in a broad sense⁴, secondary analyses of both the genetic burden and subsequent patient follow-up for 25 genes in 303 cases did highlight genes with a statistical bias toward ASD versus ID/DD diagnosis (**Table 3** and **Fig. 4c**). We found that individuals with mutations in genes enriched in ASD showed significantly lower rates of seizures, congenital abnormalities, and microcephaly, but higher rates of macrocephaly compared with those for comorbid ASD and ID genes and strong ID genes (**Fig. 4c**). The latter finding is interesting in light of the observation of increased brain sizes and/or weights at early ages in subtypes of ASD, as compared with those associated with ID or found in typical toddlers^{60–63}. Although the number of exome-sequenced cases with a *de novo* mutation was small (4.6% or 31/668 patients), the data highlight a coexpression and protein-interaction network statistically enriched in high-functioning autism patients (FSIQ >100) compared with unaffected siblings. This network is biased for *de novo* missense compared with LGD mutations (2:1), thus indicating that less severe mutations may play a role in ASD without ID. The network highlights mRNA splicing as well as genes important in chromatin remodeling. The latter implicate early developmental programs that regulate cell proliferation, neural patterning and differentiation, axonal guidance consistent with cellular ASD models^{64–66}, and ASD postmortem^{63,67,68}, genomic^{69,70}, and developmental imaging^{60–62} studies.

URLs. Denovo-db, <http://denovo-db.gs.washington.edu/denovo-db/>; dbSNP, <http://www.ncbi.nlm.nih.gov/SNP/>; Exome Aggregation Consortium (ExAC), <http://exac.broadinstitute.org/>; Enrichr tool, <http://amp.pharm.mssm.edu/Enrichr/>; BrainSpan Atlas of the Developing Human Brain, <http://www.brainspan.org/>; GTEx, <http://www.gtexportal.org/>; Online Mendelian Inheritance in Man (OMIM), <http://www.omim.org/>; NHLBI GO Exome Sequencing Project (ESP), <http://evs.gs.washington.edu/EVS/>; MIPgen suite, <http://shendurelab.github.io/MIPGEN/>; Ensembl Variant Effect Predictor Tool for GRCh37, http://grch37.ensembl.org/Homo_sapiens/Tools/VEP/; Combined Annotation Dependent Depletion (CADD), <http://cadd.gs.washington.edu/>; NCBI Exome Sequencing Project (ESP), <http://evs.gs.washington.edu/EVS/>; Vienna *Drosophila* Resource Center (VDRC), <http://stockcenter.vdrc.at/control/main/>; UCSC Genome Browser, <http://genome.ucsc.edu/>; NCBI Gene, <http://www.ncbi.nlm.nih.gov/gene/>.

METHODS

Methods, including statements of data availability and any associated accession codes and references, are available in the **online version of the paper**.

Note: Any Supplementary Information and Source Data files are available in the online version of the paper.

ACKNOWLEDGMENTS

We thank the individuals and their families for participation in this study. We acknowledge the Vienna *Drosophila* Resource Center and Bloomington

Drosophila Stock Center (NIH P40D018537). This research was supported in part by the following: the Simons Foundation Autism Research Initiative (SFARI 303241) and NIH (R01MH101221) to E.E.E.; VIDI and TOP grants (917-96-346, 912-12-109) from the Netherlands Organization for Scientific Research and Horizon 2020 Marie Skłodowska–Curie European Training Network (MiND, 643051) to A.S.; an NHGRI Interdisciplinary Training in Genome Science grant (T32HG00035) to H.A.F.S. and T.N.T.; Australian NHMRC grants 1091593 and 1041920 and Channel 7 Children's Research Foundation support to J.G.; the National Basic Research Program of China (2012CB517900) and the National Natural Science Foundation of China (81330027, 81525007 and 31400919) to K.X.; the China Scholarship Council (201406370028) and the Fundamental Research Funds for the Central Universities (2012zzts110) to T.W.; National Health and Medical Research Council of Australia Project grants (556759 and 1044175) to I.E.S., P.J.L., and M.B.D., and a Practitioner Fellowship (1006110) to I.E.S.; grants from the Jack Brockhoff Foundation and Perpetual Trustees, the Victorian State Government Operational Infrastructure Support and Australian Government NHMRC IRIISS, the Swedish Brain Foundation, the Swedish Research Council, and the Stockholm County Council; the University of California, San Diego Clinical and Translational Research Institute (KL2TR00099 and 1KL2TR001444) to T.P.; and the Research Fund–Flanders (FWO) to R.F.K. and G.V.D.W. We are grateful to all of the families at the participating SSC sites, as well as the principal investigators (A. Beaudet, R. Bernier, J. Constantino, E. Cook, E. Fombonne, D. Geschwind, R. Goin-Kochel, E. Hanson, D. Grice, A. Klin, D. Ledbetter, C. Lord, C. Martin, D. Martin, R. Maxim, J. Miles, O. Ousley, K. Pelphrey, B. Peterson, J. Piggot, C. Saulnier, M. State, W. Stone, J. Sutcliffe, C. Walsh, Z. Warren, and E. Wijsman). We appreciate access to phenotypic data on SFARI Base. We gratefully acknowledge the resources provided by the Autism Genetic Resource Exchange (AGRE) Consortium and the participating AGRE families. AGRE is a program of Autism Speaks and is supported in part by grant 1U24MH081810 from the National Institute of Mental Health to C.M. Lajonchere. We thank N. Brown, K. Pereira, T. Vick, T. Desai, C. Green, A.L. Doebley, and L. Grillo for their valuable contributions as well as T. Brown for assistance in editing this manuscript. H.P. is supported as a Senior Clinical Investigator of FWO. E.E.E. is supported as an investigator of the Howard Hughes Medical Institute.

AUTHOR CONTRIBUTIONS

E.E.E., H.A.F.S., B.X., and B.P.C. designed the study. H.A.F.S., B.X., T.W., K.H., L.V., and J. Lin performed the experiments. B.P.C. assisted with smMIP design and data analysis. F.H. performed the gene network analysis. R.A.B., J. Gerdt, and S.T. analyzed the patient data. B.X., M.F., B.H., and A.C.-N. performed and analyzed the *Drosophila* experiments. Other authors participated in the sample collection and DNA extraction and/or preparation. E.E.E., H.A.F.S., B.P.C., B.X., A.S., M.F., and R.A.B. wrote the manuscript with input from all authors. B.P.C. and T.W. contributed equally to this effort and should be regarded as joint second authors.

COMPETING FINANCIAL INTERESTS

The authors declare competing financial interests: details are available in the [online version of the paper](#).

Reprints and permissions information is available online at <http://www.nature.com/reprints/index.html>.

1. *Diagnostic and Statistical Manual of Mental Disorders* 5th edn. (American Psychiatric Association, 2013).
2. Posthuma, D. & Polderman, T.J. What have we learned from recent twin studies about the etiology of neurodevelopmental disorders? *Curr. Opin. Neurol.* **26**, 111–121 (2013).
3. Torres, F., Barbosa, M. & Maciel, P. Recurrent copy number variations as risk factors for neurodevelopmental disorders: critical overview and analysis of clinical implications. *J. Med. Genet.* **53**, 73–90 (2016).
4. Matson, J.L. & Shoemaker, M. Intellectual disability and its relationship to autism spectrum disorders. *Res. Dev. Disabil.* **30**, 1107–1114 (2009).
5. Stessman, H.A., Bernier, R. & Eichler, E.E. A genotype-first approach to defining the subtypes of a complex disease. *Cell* **156**, 872–877 (2014).
6. Bernier, R. *et al.* Disruptive *CHD8* mutations define a subtype of autism early in development. *Cell* **158**, 263–276 (2014).
7. van Bon, B.W. *et al.* Disruptive *de novo* mutations of *DYRK1A* lead to a syndromic form of autism and ID. *Mol. Psychiatry* **21**, 126–132 (2016).
8. Helsmoortel, C. *et al.* A SWI/SNF-related autism syndrome caused by *de novo* mutations in *ADNP*. *Nat. Genet.* **46**, 380–384 (2014).
9. O'Roak, B.J. *et al.* Multiplex targeted sequencing identifies recurrently mutated genes in autism spectrum disorders. *Science* **338**, 1619–1622 (2012).
10. Hiatt, J.B., Pritchard, C.C., Salipante, S.J., O'Roak, B.J. & Shendure, J. Single molecule molecular inversion probes for targeted, high-accuracy detection of low-frequency variation. *Genome Res.* **23**, 843–854 (2013).
11. O'Roak, B.J. *et al.* Recurrent *de novo* mutations implicate novel genes underlying simplex autism risk. *Nat. Commun.* **5**, 5595 (2014).
12. Iossifov, I. *et al.* The contribution of *de novo* coding mutations to autism spectrum disorder. *Nature* **515**, 216–221 (2014).
13. Krumm, N. *et al.* Excess of rare, inherited truncating mutations in autism. *Nat. Genet.* **47**, 582–588 (2015).
14. De Rubeis, S. *et al.* Synaptic, transcriptional and chromatin genes disrupted in autism. *Nature* **515**, 209–215 (2014).
15. de Ligt, J. *et al.* Diagnostic exome sequencing in persons with severe intellectual disability. *N. Engl. J. Med.* **367**, 1921–1929 (2012).
16. Rauch, A. *et al.* Range of genetic mutations associated with severe non-syndromic sporadic intellectual disability: an exome sequencing study. *Lancet* **380**, 1674–1682 (2012).
17. Deciphering Developmental Disorders Study. Large-scale discovery of novel genetic causes of developmental disorders. *Nature* **519**, 223–228 (2015).
18. Turner, T.N. *et al.* *de novo*-db: a compendium of human *de novo* variants. *Nucleic Acids Res.* **45**, D804–D811 (2017).
19. Coe, B.P. *et al.* Refining analyses of copy number variation identifies specific genes associated with developmental delay. *Nat. Genet.* **46**, 1063–1071 (2014).
20. Hormozdiari, F., Penn, O., Borenstein, E. & Eichler, E.E. The discovery of integrated gene networks for autism and related disorders. *Genome Res.* **25**, 142–154 (2015).
21. Wang, T. *et al.* *De novo* genetic mutations among a Chinese autism spectrum disorder cohort. *Nat. Commun.* **7**, 13316 (2016).
22. Turner, T.N. *et al.* Genome sequencing of autism-affected families reveals disruption of putative noncoding regulatory DNA. *Am. J. Hum. Genet.* **98**, 58–74 (2016).
23. Hamdan, F.F. *et al.* *De novo* mutations in *FOXP1* in cases with intellectual disability, autism, and language impairment. *Am. J. Hum. Genet.* **87**, 671–678 (2010).
24. Ba, W. *et al.* TRIO loss of function is associated with mild intellectual disability and affects dendritic branching and synapse function. *Hum. Mol. Genet.* **25**, 892–902 (2016).
25. Han, S. *et al.* Autistic-like behaviour in *Scn1a*^{+/−} mice and rescue by enhanced GABA-mediated neurotransmission. *Nature* **489**, 385–390 (2012).
26. Witteveen, J.S. *et al.* Haploinsufficiency of MeCP2-interacting transcriptional co-repressor *SIN3A* causes mild intellectual disability by affecting the development of cortical integrity. *Nat. Genet.* **48**, 877–887 (2016).
27. Shoubridge, C. *et al.* Mutations in the guanine nucleotide exchange factor gene *IQSEC2* cause nonsyndromic intellectual disability. *Nat. Genet.* **42**, 486–488 (2010).
28. Chan, C.B. *et al.* PIKE is essential for oligodendroglia development and CNS myelination. *Proc. Natl. Acad. Sci. USA* **111**, 1993–1998 (2014).
29. McNeill, E.M. *et al.* Nav2 hypomorphic mutant mice are ataxic and exhibit abnormalities in cerebellar development. *Dev. Biol.* **353**, 331–343 (2011).
30. Stray-Pedersen, A. *et al.* Biallelic mutations in *UNC80* cause persistent hypotonia, encephalopathy, growth retardation, and severe intellectual disability. *Am. J. Hum. Genet.* **98**, 202–209 (2016).
31. Turner, T.N. *et al.* Loss of δ -catenin function in severe autism. *Nature* **520**, 51–56 (2015).
32. Sanders, S.J. *et al.* Insights into autism spectrum disorder genomic architecture and biology from 71 risk loci. *Neuron* **87**, 1215–1233 (2015).
33. Rope, A.F. *et al.* Using VAAST to identify an X-linked disorder resulting in lethality in male infants due to N-terminal acetyltransferase deficiency. *Am. J. Hum. Genet.* **89**, 28–43 (2011).
34. Liszczak, G. *et al.* Molecular basis for N-terminal acetylation by the heterodimeric NatA complex. *Nat. Struct. Mol. Biol.* **20**, 1098–1105 (2013).
35. Baird, P.A., Anderson, T.W., Newcombe, H.B. & Lowry, R.B. Genetic disorders in children and young adults: a population study. *Am. J. Hum. Genet.* **42**, 677–693 (1988).
36. Rosenfeld, J.A., Coe, B.P., Eichler, E.E., Cuckle, H. & Shaffer, L.G. Estimates of penetrance for recurrent pathogenic copy-number variations. *Genet. Med.* **15**, 478–481 (2013).
37. Chen, E.Y. *et al.* Enrichr: interactive and collaborative HTML5 gene list enrichment analysis tool. *BMC Bioinformatics* **14**, 128 (2013).
38. Stessman, H.A. *et al.* Disruption of *POGZ* is associated with intellectual disability and autism spectrum disorders. *Am. J. Hum. Genet.* **98**, 541–552 (2016).
39. Esmaeli-Nieh, S. *et al.* BOD1 is required for cognitive function in humans and *Drosophila*. *PLoS Genet.* **12**, e1006022 (2016).
40. Lugtenberg, D. *et al.* *De novo* loss-of-function mutations in *WAC* cause a recognizable intellectual disability syndrome and learning deficits in *Drosophila*. *Eur. J. Hum. Genet.* **24**, 1145–1153 (2016).
41. Kleefstra, T. *et al.* Disruption of an EHMT1-associated chromatin-modification module causes intellectual disability. *Am. J. Hum. Genet.* **91**, 73–82 (2012).
42. van Bon, B.W. *et al.* CEP99 is required for mitochondrial metabolism and neuronal function in man and fly. *Hum. Mol. Genet.* **22**, 3138–3151 (2013).
43. Willemsen, M.H. *et al.* *GATAD2B* loss-of-function mutations cause a recognisable syndrome with intellectual disability and are associated with learning deficits and synaptic undergrowth in *Drosophila*. *J. Med. Genet.* **50**, 507–514 (2013).
44. Schmid, S., Wilson, D.A. & Rankin, C.H. Habituation mechanisms and their importance for cognitive function. *Front. Integr. Neurosci.* **8**, 97 (2015).
45. Kleinhans, N.M. *et al.* Reduced neural habituation in the amygdala and social impairments in autism spectrum disorders. *Am. J. Psychiatry* **166**, 467–475 (2009).

46. Dinstein, I. *et al.* Unreliable evoked responses in autism. *Neuron* **75**, 981–991 (2012).
47. Pellicano, E., Rhodes, G. & Calder, A.J. Reduced gaze aftereffects are related to difficulties categorising gaze direction in children with autism. *Neuropsychologia* **51**, 1504–1509 (2013).
48. Ethridge, L.E. *et al.* Reduced habituation of auditory evoked potentials indicate cortical hyper-excitability in Fragile X Syndrome. *Transl. Psychiatry* **6**, e787 (2016).
49. Cascio, C.J., Woynaroski, T., Baranek, G.T. & Wallace, M.T. Toward an interdisciplinary approach to understanding sensory function in autism spectrum disorder. *Autism Res.* **9**, 920–925 (2016).
50. Ramaswami, M. Network plasticity in adaptive filtering and behavioral habituation. *Neuron* **82**, 1216–1229 (2014).
51. Tartaglia, M. *et al.* Mutations in *PTPN11*, encoding the protein tyrosine phosphatase SHP-2, cause Noonan syndrome. *Nat. Genet.* **29**, 465–468 (2001).
52. Iossifov, I. *et al.* Low load for disruptive mutations in autism genes and their biased transmission. *Proc. Natl. Acad. Sci. USA* **112**, E5600–E5607 (2015).
53. Sugiura, N., Patel, R.G. & Corriveau, R.A. N-methyl-D-aspartate receptors regulate a group of transiently expressed genes in the developing brain. *J. Biol. Chem.* **276**, 14257–14263 (2001).
54. Myklebust, L.M. *et al.* Biochemical and cellular analysis of Ogden syndrome reveals downstream Nt-acetylation defects. *Hum. Mol. Genet.* **24**, 1956–1976 (2015).
55. Homsy, J. *et al.* *De novo* mutations in congenital heart disease with neurodevelopmental and other congenital anomalies. *Science* **350**, 1262–1266 (2015).
56. van Bokhoven, H. Genetic and epigenetic networks in intellectual disabilities. *Annu. Rev. Genet.* **45**, 81–104 (2011).
57. Zhu, T. *et al.* Histone methyltransferase Ash1L mediates activity-dependent repression of neurexin-1 α . *Sci. Rep.* **6**, 26597 (2016).
58. Griswold, A.J. *et al.* Targeted massively parallel sequencing of autism spectrum disorder-associated genes in a case control cohort reveals rare loss-of-function risk variants. *Mol. Autism* **6**, 43 (2015).
59. Rhodes, C.T. *et al.* Cross-species analyses unravel the complexity of H3K27me3 and H4K20me3 in the context of neural stem progenitor cells. *Neuroepigenetics* **6**, 10–25 (2016).
60. Courchesne, E. *et al.* Unusual brain growth patterns in early life in patients with autistic disorder: an MRI study. *Neurology* **57**, 245–254 (2001).
61. Shen, M.D. *et al.* Early brain enlargement and elevated extra-axial fluid in infants who develop autism spectrum disorder. *Brain* **136**, 2825–2835 (2013).
62. Schumann, C.M. *et al.* Longitudinal magnetic resonance imaging study of cortical development through early childhood in autism. *J. Neurosci.* **30**, 4419–4427 (2010).
63. Redcay, E. & Courchesne, E. When is the brain enlarged in autism? A meta-analysis of all brain size reports. *Biol. Psychiatry* **58**, 1–9 (2005).
64. Marchetto, M.C. *et al.* Altered proliferation and networks in neural cells derived from idiopathic autistic individuals. *Mol. Psychiatry* <http://dx.doi.org/10.1038/mp.2016.95> (2016).
65. Sugathan, A. *et al.* CHD8 regulates neurodevelopmental pathways associated with autism spectrum disorder in neural progenitors. *Proc. Natl. Acad. Sci. USA* **111**, E4468–E4477 (2014).
66. Cotney, J. *et al.* The autism-associated chromatin modifier CHD8 regulates other autism risk genes during human neurodevelopment. *Nat. Commun.* **6**, 6404 (2015).
67. Courchesne, E. *et al.* Neuron number and size in prefrontal cortex of children with autism. *J. Am. Med. Assoc.* **306**, 2001–2010 (2011).
68. Stoner, R. *et al.* Patches of disorganization in the neocortex of children with autism. *N. Engl. J. Med.* **370**, 1209–1219 (2014).
69. Chow, M.L. *et al.* Age-dependent brain gene expression and copy number anomalies in autism suggest distinct pathological processes at young versus mature ages. *PLoS Genet.* **8**, e1002592 (2012).
70. Prampero, T. *et al.* Cell cycle networks link gene expression dysregulation, mutation, and brain maldevelopment in autistic toddlers. *Mol. Syst. Biol.* **11**, 841 (2015).

ONLINE METHODS

Patient samples. Whole-blood or cell-line DNA from patients with ASD, ID, or DD diagnoses were collected from 15 international clinical and research cohorts (Fig. 1). Only DNA samples from The Autism Simplex Collection (TASC) and Autism Genetic Resource Exchange (AGRE) cohorts were derived from cell lines. Clinical workup, including diagnostic evaluation, medical examination, and neuropsychological assessment was made available for many patients upon request, specifically for patients with mutations in *NAA15*, *KMT5B*, and *ASH1L* (case reports in **Supplementary Note**). Best-estimate clinical DSM-5 diagnoses were made by experienced, licensed clinicians on the basis of all available information collected during the evaluation. Descriptions, the number of individuals represented, and the primary ascertainment criteria for each cohort in this study are shown in **Supplementary Table 8**. In addition, 2,867 unaffected-sibling control individuals were analyzed for genetic and phenotypic comparison (**Supplementary Table 9**). All experiments carried out on these individuals were in accordance with the ethical standards of the responsible institutional and national committees on human experimentation (the University of Washington Institutional Review Board approved the human research performed in this study), and proper informed consent was obtained for sequencing, recontact for inheritance testing, and phenotypic workup. All sequencing of patient samples was performed at the University of Washington.

Detailed descriptions of clinical cohorts. *Autism Clinical and Genetic Resources in China (ACGC).* This cohort has been described previously²¹.

Autism Genetic Resource Exchange (AGRE). This cohort has been described previously⁷¹.

Leuven. The Leuven cohort consists of patients with ASD, as diagnosed by the multidisciplinary team in the Expert Centre for Autism Leuven according to DSM-IV-TR (American Psychiatric Association, 2000) criteria. All patients were examined by a clinical geneticist. Patients with known monogenic conditions were excluded after a routine genetics workup.

Melbourne & Murdoch. All participants had a DSM-IV or DSM-5 diagnosis of ASD. Diagnoses were community based and were performed by a multidisciplinary team (pediatrician, psychologist, and speech pathologist). Diagnoses were confirmed for research purposes through ascertainment of previous ASD and cognitive assessments and a telephone interview with the parents. Information about pregnancy and birth, developmental milestones, comorbidities, medications, and general health was collected during the interviews and from medical records. Current ASD symptomatology was measured via the Social Responsiveness Scale (SRS) or Social Responsiveness Scale, 2nd edition (SRS-2). For each family, pedigrees were constructed detailing the family's history of medical conditions, mental health disorders, intellectual impairment, ASD diagnoses and ASD traits. DNA was collected from blood or saliva from probands and their parents. Most probands underwent molecular karyotyping for CNVs and single-nucleotide polymorphisms and fragile-X DNA testing; older participants had routine karyotyping.

The Autism Simplex Collection (TASC). This cohort has been described previously⁷².

Adelaide. Individuals with intellectual disability or developmental delay were recruited who were referred but were negative in molecular testing for fragile X and exhibited large CNVs in array comparative genomic hybridization (CGH). The majority were also singletons and were recently clinically diagnosed/ascertained patients, for recontacting purposes.

Leiden. The cohort consists of patients with developmental delay with or without autistic features. Clinical microarrays to detect CNVs were run on all index patients, and identification of a likely causal CNV was an exclusion criterion. No formal DSM criteria were used in the diagnoses. All patients were seen by experienced clinical geneticists and, if indicated, specific gene tests were requested. Parents of the patients provided verbal consent for inclusion in this study.

Stockholm. For all cases, array CGH was performed on an Agilent platform with a 180K genome-wide design. The cases were referred for genetic investigation after receiving a diagnosis of ID/ASD, but the DSM-5 guidelines were not used systematically.

Candidate-gene selection. Candidate genes with *de novo* mutations were identified from whole-exome and targeted-sequencing studies of ASD, ID,

and DD, on the basis of previously published studies, and included 4,874 probands with ASD^{9,11–14}, 151 probands with ID^{15,16}, and 1,133 probands with DD¹⁷ (**Supplementary Table 1**). Genes were ranked on the basis of the following criteria: (i) presence of two or more LGD mutations; (ii) presence of multiple missense mutations and at least one LGD mutation; (iii) presence of at least one LGD mutation overlapping a region of interest in our published DD CNV morbidity map¹⁹; and (iv) presence of at least one LGD mutation with network connectivity to either chromatin remodeling/transcription or long-term potentiation, as described previously²⁰. Genes with expression in the brain (on the basis of the BrainSpan Atlas of the Developing Human Brain) and GTEx databases⁷³) were prioritized. We eliminated genes associated with likely unrelated disorders in OMIM, and genes that were deemed highly mutable (on the basis of data from 6,503 control individuals in the NHLBI GO Exome Sequencing Project). Finally, we filtered genes according to the number of *de novo* mutations by whole-exome sequencing among unaffected siblings from the SSC^{11–13}. More details on selection criteria can be found in **Supplementary Table 3**.

smMIP sequencing and variant validation. smMIP sequencing was performed as previously described^{9,10}. We targeted the coding portions of all RefSeq annotated transcripts for these 208 genes as well as 5 bp into each exon-adjacent intron to capture variation at splice-donor/acceptor sites, thus resulting in the design of 12,016 smMIPs. smMIPs were split into four pools (Gold, ASD4, ASD5, and ASD6; gene breakdown in **Supplementary Tables 4–7**), and each pool was rebalanced so that poorer-performing smMIPs were spiked in at a concentration of 10× or 50×. Approximately 192 samples per lane were barcoded and sequenced with an Illumina HiSeq 2000, as previously described¹¹, and data analysis was performed with the MIPgen suite of tools. Variant calling of smMIP data was performed on each sequencing lane with FreeBayes v0.9.14 with default settings and the hg19 reference. For each of the four pools, all FreeBayes output was combined with GATK. Allele counts per genotype (AC) and the total number of alleles per genotype (AN) were recalibrated on the combined variant set with VCFtools. Multiallelic sites were split into separate entries with vcflib (vcfbreakmulti), and we removed sequencing-error repeats and common single-nucleotide polymorphisms (all of dbSNP v129 and variants found in dbSNP v141 at a minor-allele frequency ≥ 0.01 in at least one major population with at least two unrelated individuals having the minor allele). From the individual genotypes with sequencing depth (DP) of $>8\times$ and a quality score (QUAL) of >20 , a private filter (found in only one family in the study) was applied to each pooled data set (i.e., ASD4, ASD5, ASD6, or Gold). These variants were annotated with the Ensembl Variant Effect Predictor tool for GRCh37 and with CADD scores⁷⁴. All private LGD variants and a portion of MIS30 variants were validated by Sanger sequencing. Specifically, a CADD >30 was chosen for validation, because these events are very rare ($<0.1\%$ of all missense events in control genomes⁷⁵) and are more likely to be pathogenic²¹. When available, parents were also Sanger sequenced to determine the inheritance status of these variants. In total, we targeted $>16,000$ unique samples, including 13,407 probands and 2,867 unaffected siblings, by using each of the four smMIP pools. 1,744 of the DNA samples did not have sufficient DNA for all four pools; gene sets were prioritized on the basis of potential disease significance (Gold $>$ ASD4 $>$ ASD5 $>$ ASD6; **Supplementary Table 10**), with 13,407 probands (Gold), 12,192 probands (ASD4 and ASD5), and 11,731 probands (ASD6). In addition, we sequenced 2,867 unaffected-sibling samples with each of the four pools. To determine the performance of each smMIP pool, ten plates of unaffected siblings (960 samples) were compared in each pool by plotting the frequency of these 960 samples that reached at least $8\times$ sequencing coverage for each individual smMIP in the study (**Supplementary Table 3**). Each of these data points was plotted by gene within each pool (i.e., Gold, ASD4, ASD5, and ASD6, shown in **Supplementary Figs. 1–4**, respectively). 165 genes passed all QC metrics (75% of smMIPs by gene reached at least $8\times$ coverage in $\geq 80\%$ of controls), but some exons, owing to their size or GC composition, did not pass these thresholds. For those regions that did not pass QC, we considered variant genotypes identified in samples if they were of high quality (read depth (DP) >8 , phred scaled QUAL >20); however, these variants were not considered for assessments of mutation burden.

Clinical recontact and phenotyping. To systematically compare the effects of particular LGD mutations of targeted genes on phenotype, we recontacted individuals with identified LGD mutations and conducted a comprehensive phenotypic workup assessing function across multiple domains. No statistical method was used to predetermine sample size. The experiments were not randomized and were not performed with blinding. Per our human-subject approval, we recontacted only individuals who had consented to be approached about future studies during their original assessment. Families were invited to participate in a comprehensive clinical workup that included diagnostic evaluation, medical examination, and neuropsychological assessment (test battery and procedures in **Supplementary Note**). Importantly, all assessments were conducted by examiners naive to the individual's genetic event, thereby decreasing clinician bias in rendering diagnostic dispositions. To make comparisons across groups with similar LGD mutations, each participant was scored according to a modified version of the de Vries scale as a proxy for the overall severity of the phenotype^{76–78}. The modified de Vries scale included the presence of facial dysmorphisms, congenital abnormalities, postnatal head growth abnormalities, ID/DD, and the number of DSM-5 diagnoses and medical diagnoses conferred, thus yielding a score ranging from 0 to 12 (**Supplementary Table 22**). Data collected from these patients were combined with published case reports to increase our power to detect enrichments among patients sharing *de novo* LGD mutations in the same gene or pathway. A total of 323 case reports of individuals with *de novo* LGD mutations of interest and relevant data were included. The relevant phenotype data extracted from cases in the published literature were combined with information collected from individuals that were able to complete the in-person comprehensive evaluation. The modified de Vries scale scores for individuals with the same disrupted gene were averaged and then rank-ordered to estimate the effect of the gene mutation on phenotype. Only genes with six or more study participants and published case reports were included in the analysis. Patients that were considered had LGD mutations in one of the following genes: *ADNP*, *ARID1B*, *CHD2*, *CHD8*, *CTNNB1*, *DYRK1A*, *FOXP1*, *GRIN2B*, *MED13L*, *POGZ*, *PTEN*, *SCN2A*, *SETBP1*, *STXBP1*, or *TBL1XR1*.

Patient workups. Comprehensive clinical workups included diagnostic evaluation, medical examination, and neuropsychological assessment. Best-estimate clinical DSM-5 diagnoses were made by experienced, licensed clinicians using all available information collected during the research evaluations. The battery included autism-specific diagnostic measures, the Autism Diagnostic Observation Schedule⁷⁹ and the Autism Diagnostic Interview–Revised⁸⁰, both administered by research-reliable clinicians. The battery also included assessment of cognitive ability (Differential Ability Scales, DAS⁸¹), language ability (Peabody Picture Vocabulary Test, 4th edition (PPVT) and Expressive Vocabulary Test, 2nd edition (EVT)), adaptive functioning (Vineland Adaptive Behavior Scales, 2nd edition), motor ability (Movement ABC; Purdue Pegboard), and behavioral and psychiatric disorders (Child and Adolescent Symptom Inventory, 5th edition (CASI-5), Child Behavior Checklist (CBCL), and Aberrant Behavior Checklist (ABC)). Medical diagnoses were assessed with the SSC medical history interview⁸² and by physical examination by a developmental pediatrician conducting a standardized medical examination.

Participants undergoing comprehensive phenotypic assessment and published case reports in the literature were scored with an adapted de Vries scale as a proxy for the overall severity of the phenotype. Modifications included the removal of stature and prenatal-onset growth retardation, the inclusion of medical and psychiatric diagnoses, revision of weighting of intellectual disability into three points, and an increase to a total score of 12. Borderline intellectual disability or general delays were rated with one point, mild to moderate intellectual disability was scored with two points, and severe-profound intellectual disability was scored with three points. Psychiatric and medical diagnoses were tallied and scored as one if an individual had one diagnosis in these domains and was scored as two if the child had two or more diagnoses in these domains.

DSM-5 diagnoses included: ASD (299.00), attention-deficit/hyperactivity disorders (314.01, 314.00), language disorder (315.39), speech sound disorder (315.39), developmental coordination disorder (315.4), anxiety disorders (309.21, 300.29, 300.01, 300.02, 300.09), behavior disorders (313.81, 312.34, 312.81, 312.9), mood disorders (311.0, 296.99, 300.4), and elimination disorders

(307.6, 307.7). Intellectual disability (319, 315.8) was not tallied in the DSM-5 diagnosis domain. Medical diagnoses were tallied according to system: cardiac, gastrointestinal, genital, neurological, pulmonary, renal, and visual and auditory. To not double-code diagnoses, microcephaly, macrocephaly and congenital abnormalities were not tallied under the medical diagnoses domain.

Relevant phenotypic data were extracted from published case reports of individuals with *de novo* LGD mutations to increase the power to detect enrichments among patients sharing *de novo* LGD mutations in the same gene. A total of 323 case reports of individuals with *de novo* LGD mutations of interest and relevant data were identified, and 215 case reports had sufficient data to incorporate in the de Vries scale. LGD mutations included: *ADNP*^{8,17,83}, *ARID1B*^{17,23,84}, *CHD2* (refs. 17,23,85), *CHD8* (refs. 6,86), *CTNNB1* (refs. 15,17,87–89), *DYRK1A*^{7,17}, *FOXP1* (refs. 17,90,91), *GRIN2B*^{15,17,23,85,92–95}, *MED13L*^{17,23,96,97}, *POGZ*^{17,38,98,99}, *PTEN*^{85,100,101}, *SCN2A*^{15,17,85,102–106}, *SETBP1* (refs. 17,19,23,107–109), *STXBP1* (refs. 17,85,110), and *TBL1XR1* (refs. 17,111).

Network analysis. We investigated modules significantly disrupted in high-functioning-autism samples (full-scale IQ >100). We applied MAGI²⁰ on all of the samples from the ASD probands in SSC with FSIQ above 100. This subset of samples covered over 500 total *de novo* missense mutations and 100 LGD mutations. We applied the MAGI tool for module discovery on these variants by protein–interaction networks, gene coexpression networks and severe mutations reported in a control population from ESP ($n = 6,500$ individuals). The protein–interaction network used was a combination of networks from HPRD¹¹² and STRING¹¹³ databases, and the coexpression network was built with the BrainSpan Atlas resource. The exact same training networks were used in our previous analysis for autism-module discovery²⁰. The parameters were that the pairwise gene coexpression inside modules be, on average, at least 0.415 and the average protein–interaction density be 0.085. MAGI found one module of 40 genes significantly enriched in *de novo* mutations ($P < 0.01$ on the basis of 100 random-mutation permutation tests).

***Drosophila* knockdown models.** *Drosophila* orthologs of the genes of interest were determined with Ensembl, Unigene, and FlyBase databases^{114,115}. Their expression was knocked down with the UAS–Gal4 system¹¹⁶ to induce conditional RNAi. The pan-neuronal promoter line *w1118*; *2xGMR-wIR*; *elav-Gal4*, *UAS-Dicer-2* (ref. 42) and two independent RNAi constructs per gene were used whenever available (VDRC)¹¹⁷, thereby fulfilling stringent specificity criteria ($s19$ value ≥ 0.98)¹¹⁸. Strains containing identical genetic backgrounds to the RNAi constructs (no. 60000 and no. 60100) were crossed to the driver line and used as controls. No effects in our assays were observed when crossing the ‘40DUAS’ line¹¹⁹ (containing UAS repeats but no functional short hairpin RNA, a potential source of dominant phenotypes due to an integration locus of the VDRC KK library^{119,120}) to our pan-neuronal Gal4 driver. Flies were cultured according to standard procedures. Experimental randomization was not applicable to the *Drosophila* experiments, and power calculations were not performed in this study.

***Drosophila* light-off jump-reflex habituation assay.** The *Drosophila* light-off jump habituation assay was performed as previously described¹²¹. Briefly, flies were reared at 25 °C and 70% humidity, in a 12-h:12-h light/dark cycle. For all healthy lines, at least 64 3- to 7-d-old male flies were tested per genotype, in at least two independent experiments. Flies were transferred into individual vials of two 16-unit habituation systems (Aktogen) and, after a 5-min adaptation, were exposed to 100 light-off pulses (15 ms each) with a 1-s inter-pulse interval. Jump responses were recorded by two sensitive microphones placed in each vial. A carefully chosen threshold was applied to distinguish the jump responses from the background noise. Data from 64 individual flies per genotype were collected (two independent experiments) and analyzed with custom Labview Software (National Instruments). Investigators were blinded to genotypes during the experiments, and data were automatically analyzed. Flies that jumped at least once in the first five trials were evaluated for habituation (preestablished criterion). Initial jumping responses to light-off pulse decreased with the number of pulses, and flies were considered to be habituated when they stopped jumping for five consecutive trials (no-jump criterion). Habituation was scored as the number of trials to the

no-jump criterion (TTC). The main effect of genotype, corrected for testing day and 16-unit system on TTC values, was determined through linear-model regression analysis with R statistical software (v.3.0.0).

***Drosophila* neuromuscular junction (NMJ) experiments.** Flies were reared at 28 °C, 60% humidity and a 12-h:12-h light/dark cycle. Type 1b NMJs of muscle 4 were analyzed. Wandering third-instar larvae were collected, dissected and fixed in 3.7% paraformaldehyde for 30 min. Preparations were rinsed with PBS and permeabilized with 0.3% Triton X-100 in PBS for 2 h at room temperature. Discs large protein (Dlg1) was visualized with the primary antibody anti-Dlg1 (1:25) (Dlg1-4F3, Developmental Studies Hybridoma Bank) conjugated with the Zenon Alexa Fluor 568 Mouse IgG1 labeling kit (Invitrogen), according to the manufacturer's protocol. Antibody validation is provided on the vendors' websites. The preparations were incubated for 1.5 h at room temperature, extensively washed, and mounted in ProLong Gold Antifade Mountant (Thermo Fisher Scientific). Investigators were blinded to genotypes during the experiments, and data were automatically analyzed. On the basis of previous experience^{38–40}, fluorescence images were acquired for each genotype with an automated Leica DMI6000B high-content microscope. Morphometric analysis was performed in FIJI¹²² with the *Drosophila* NMJ Morphometrics macro¹²³. The resulting images were visually inspected for accurate image segmentation. Inaccurately segmented parameters were excluded, as previously described¹²³. The NMJ bouton number in *Hmt4-20*-knockdown flies was manually assessed by two independent investigators blinded to genotype, and data were averaged to obtain the final counts. Statistical analysis was performed with GraphPad PRISM. The area, perimeter, length, longest branch length, and number of boutons were analyzed with Student's *t*-tests. The number of branches, branching points, and islands were analyzed with Mann–Whitney *U* tests.

Statistical analyses. To calculate the significance and penetrance of LGD and MIS30 mutations in ASD/ID, we compared our smMIP data against two control data sets: the first included the 2,867 unaffected-sibling controls that were sequenced through the same smMIP pipeline (described above), and the second included mutation data from the ExAC database⁷⁵, in which neuropsychiatric cases were removed (ExAC v0.3), representing 45,376 samples. The .vcf file for these 45,376 ExAC samples was annotated with VEP and CADD (as described above). The ExAC data set was filtered through the same pipeline as the smMIP data (described above) including only 'PASS' variants. Private variants were filtered as an AC = 1 for burden statistics. To compare the ExAC control counts and the smMIP data, only LGD and MIS30 events were considered. LGD and MIS30 counts by gene for ExAC can be found in **Supplementary Table 18**. To calculate the significance of private LGD or MIS30 mutation burden in our smMIP data set compared with unaffected (ExAC) controls, we performed a simulation by shuffling the labels of private case and control observations 1×10^6 times and calculated the probability of observing at least the number of LGD or MIS30 events seen among our cases. These *P* values were corrected (Benjamini–Hochberg) for the number of genes in the study ($n = 208$). Penetrance and its confidence bounds were calculated with a previously described model³⁶:

$$P(D|G) = \frac{P(G|D)P(D)}{P(G|D)P(D) + P(G|\bar{D})P(\bar{D})}$$

where *D* is disease, *G* is genotype (the presence of the specific type of event in the gene), and \bar{D} is the absence of disease. The general population incidence of ID/DD in our cohort was assumed to match that described in Rosenfeld *et al.*^{35,36} ($P(D) = 5.12\%$), because our cohort composition has a similar representation of youth-onset diseases with an important genetic component with broad exclusion of chromosomal disorders. *De novo* significance was calculated as previously described⁹, using a statistical framework that considers the length of the gene and divergence between chimpanzees and humans. To calculate *de novo* significance for MIS30 variants, we modified the model to separately enumerate prior probabilities for CADD <30 and CADD ≥30 missense sites with CADD v1.3.

To compare clinical phenotypes, a Pearson's correlation coefficient and *P* value were used to compare overall ASD versus ID diagnosis on the basis of genetic event. Phenotypic rates were compared among individuals carrying variants in ASD versus DD genes with two-tailed Fisher's exact tests.

A linear-model regression analysis was performed with 64 flies per genotype, collected in two independent experiments, for the habituation data calculations. For the NMJ experiments, the area, perimeter, length, longest branch length, and number of boutons were analyzed with two-tailed Student's *t*-tests (degrees of freedom for *SRCAP* (*dom*) experiments: area = 60, length = 65, boutons = 73, perimeter = 60; degrees of freedom for *TCF4* (*da*) experiments: area = 63, length = 68, branches = 68, branching = 68). Branches, branching points, and islands were analyzed with Mann–Whitney *U* tests.

Data availability. The smMIP sequencing data for this study can be downloaded from the NIMH data repository National Database for Autism Research (NDAR) at <http://dx.doi.org/10.15154/1340671> and are available to all qualified researchers after data-use certification. Approved researchers can obtain the SSC population data set described in this study (<http://sfari.org/resources/simons-simplex-collection/>) by applying at <https://base.sfari.org/>.

71. Geschwind, D.H. *et al.* The autism genetic resource exchange: a resource for the study of autism and related neuropsychiatric conditions. *Am. J. Hum. Genet.* **69**, 463–466 (2001).
72. Buxbaum, J.D. *et al.* The Autism Simplex Collection: an international, expertly phenotyped autism sample for genetic and phenotypic analyses. *Mol. Autism* **5**, 34 (2014).
73. Ardlie, K.G. *et al.*; GTEx Consortium. Human genomics. The Genotype-Tissue Expression (GTEx) pilot analysis: multitissue gene regulation in humans. *Science* **348**, 648–660 (2015).
74. Kircher, M. *et al.* A general framework for estimating the relative pathogenicity of human genetic variants. *Nat. Genet.* **46**, 310–315 (2014).
75. Lek, M. *et al.* Analysis of protein-coding genetic variation in 60,706 humans. *Nature* **536**, 285–291 (2016).
76. Feenstra, I. *et al.* Balanced into array: genome-wide array analysis in 54 patients with an apparently balanced *de novo* chromosome rearrangement and a meta-analysis. *Eur. J. Hum. Genet.* **19**, 1152–1160 (2011).
77. Vulto-van Silfhout, A.T. *et al.* Clinical significance of *de novo* and inherited copy-number variation. *Hum. Mutat.* **34**, 1679–1687 (2013).
78. de Vries, B.B. *et al.* Clinical studies on submicroscopic subtelomeric rearrangements: a checklist. *J. Med. Genet.* **38**, 145–150 (2001).
79. Lord, C., Rutter, M., DiLavore, P.C. & Risi, S. *Autism Diagnostic Observation Schedule* (Western Psychological Services, 2001).
80. 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**, 659–685 (1994).
81. Elliott, C.D. *Differential Ability Scales: Introductory and Technical Manual* 2nd edn. (Harcourt Assessment, 2007).
82. Fischbach, G.D. & Lord, C. The Simons Simplex Collection: a resource for identification of autism genetic risk factors. *Neuron* **68**, 192–195 (2010).
83. Pescosolido, M.F. *et al.* Expansion of the clinical phenotype associated with mutations in activity-dependent neuroprotective protein. *J. Med. Genet.* **51**, 587–589 (2014).
84. Hoyer, J. *et al.* Haploinsufficiency of ARID1B, a member of the SWI/SNF-a chromatin-remodeling complex, is a frequent cause of intellectual disability. *Am. J. Hum. Genet.* **90**, 565–572 (2012).
85. Epi4K Consortium. *et al.* *De novo* mutations in epileptic encephalopathies. *Nature* **501**, 217–221 (2013).
86. Merner, N. *et al.* A *de novo* frameshift mutation in chromodomain helicase DNA-binding domain 8 (CHD8): A case report and literature review. *Am. J. Med. Genet. A* **170A**, 1225–1235 (2016).
87. Kuechler, A. *et al.* *De novo* mutations in beta-catenin (*CTNNB1*) appear to be a frequent cause of intellectual disability: expanding the mutational and clinical spectrum. *Hum. Genet.* **134**, 97–109 (2015).
88. Tucci, V. *et al.* Dominant β -catenin mutations cause intellectual disability with recognizable syndromic features. *J. Clin. Invest.* **124**, 1468–1482 (2014).
89. Winczewska-Wiktor, A. *et al.* A *de novo* *CTNNB1* nonsense mutation associated with syndromic atypical hyperekplexia, microcephaly and intellectual disability: a case report. *BMC Neurol.* **16**, 35 (2016).
90. Lozano, R., Vano, A., Lozano, C., Fisher, S.E. & Deriziotis, P. A *de novo* *FOXP1* variant in a patient with autism, intellectual disability and severe speech and language impairment. *Eur. J. Hum. Genet.* **23**, 1702–1707 (2015).
91. Sollis, E. *et al.* Identification and functional characterization of *de novo* *FOXP1* variants provides novel insights into the etiology of neurodevelopmental disorder. *Hum. Mol. Genet.* **25**, 546–557 (2016).
92. Adams, D.R. *et al.* Three rare diseases in one Sib pair: *RAI1*, *PCK1*, *GRIN2B* mutations associated with Smith-Magenis Syndrome, cytosolic PEPCCK deficiency and NMDA receptor glutamate insensitivity. *Mol. Genet. Metab.* **113**, 161–170 (2014).
93. Endeley, S. *et al.* Mutations in *GRIN2A* and *GRIN2B* encoding regulatory subunits of NMDA receptors cause variable neurodevelopmental phenotypes. *Nat. Genet.* **42**, 1021–1026 (2010).

94. Freunschtl, I. *et al.* Behavioral phenotype in five individuals with *de novo* mutations within the *GRIN2B* gene. *Behav. Brain Funct.* **9**, 20 (2013).
95. Lemke, J.R. *et al.* *GRIN2B* mutations in West syndrome and intellectual disability with focal epilepsy. *Ann. Neurol.* **75**, 147–154 (2014).
96. Cafiero, C. *et al.* Novel *de novo* heterozygous loss-of-function variants in *MED13L* and further delineation of the *MED13L* haploinsufficiency syndrome. *Eur. J. Hum. Genet.* **23**, 1499–1504 (2015).
97. van Haelst, M.M. *et al.* Further confirmation of the *MED13L* haploinsufficiency syndrome. *Eur. J. Hum. Genet.* **23**, 135–138 (2015).
98. Fukai, R. *et al.* A case of autism spectrum disorder arising from a *de novo* missense mutation in *POGZ*. *J. Hum. Genet.* **60**, 277–279 (2015).
99. White, J. *et al.* *POGZ* truncating alleles cause syndromic intellectual disability. *Genome Med.* **8**, 3 (2016).
100. Busa, T. *et al.* Clinical presentation of *PTEN* mutations in childhood in the absence of family history of Cowden syndrome. *Eur. J. Paediatr. Neurol.* **19**, 188–192 (2015).
101. Buxbaum, J.D. *et al.* Mutation screening of the *PTEN* gene in patients with autism spectrum disorders and macrocephaly. *Am. J. Med. Genet. B. Neuropsychiatr. Genet.* **144B**, 484–491 (2007).
102. Baasch, A.L. *et al.* Exome sequencing identifies a *de novo* *SCN2A* mutation in a patient with intractable seizures, severe intellectual disability, optic atrophy, muscular hypotonia, and brain abnormalities. *Epilepsia* **55**, e25–e29 (2014).
103. Dhamija, R., Wirrell, E., Falcao, G., Kirmani, S. & Wong-Kissel, L.C. Novel *de novo* *SCN2A* mutation in a child with migrating focal seizures of infancy. *Pediatr. Neurol.* **49**, 486–488 (2013).
104. Dimassi, S. *et al.* Whole-exome sequencing improves the diagnosis yield in sporadic infantile spasm syndrome. *Clin. Genet.* **89**, 198–204 (2016).
105. Nakamura, K. *et al.* Clinical spectrum of *SCN2A* mutations expanding to Ohtahara syndrome. *Neurology* **81**, 992–998 (2013).
106. Tavassoli, T. *et al.* *De novo* *SCN2A* splice site mutation in a boy with Autism spectrum disorder. *BMC Med. Genet.* **15**, 35 (2014).
107. Herenger, Y. *et al.* Long term follow up of two independent patients with Schinzel-Giedion carrying *SETBP1* mutations. *Eur. J. Med. Genet.* **58**, 479–487 (2015).
108. Miyake, F. *et al.* West syndrome in a patient with Schinzel-Giedion syndrome. *J. Child Neurol.* **30**, 932–936 (2015).
109. Takeuchi, A. *et al.* Progressive brain atrophy in Schinzel-Giedion syndrome with a *SETBP1* mutation. *Eur. J. Med. Genet.* **58**, 369–371 (2015).
110. Stamberger, H. *et al.* *STXBP1* encephalopathy: a neurodevelopmental disorder including epilepsy. *Neurology* **86**, 954–962 (2016).
111. Heinen, C.A. *et al.* A specific mutation in *TBL1XR1* causes Pierpont syndrome. *J. Med. Genet.* **53**, 330–337 (2016).
112. Keshava Prasad, T.S. *et al.* Human Protein Reference Database: 2009 update. *Nucleic Acids Res.* **37**, D767–D772 (2009).
113. Szklarczyk, D. *et al.* The STRING database in 2011: functional interaction networks of proteins, globally integrated and scored. *Nucleic Acids Res.* **39**, D561–D568 (2011).
114. Wheeler, D.L. *et al.* Database resources of the National Center for Biotechnology. *Nucleic Acids Res.* **31**, 28–33 (2003).
115. Attrill, H. *et al.* FlyBase: establishing a Gene Group resource for *Drosophila melanogaster*. *Nucleic Acids Res.* **44**, D786–D792 (2016).
116. Brand, A.H. & Perrimon, N. Targeted gene expression as a means of altering cell fates and generating dominant phenotypes. *Development* **118**, 401–415 (1993).
117. Dietzl, G. *et al.* A genome-wide transgenic RNAi library for conditional gene inactivation in *Drosophila*. *Nature* **448**, 151–156 (2007).
118. Oortveld, M.A. *et al.* Human intellectual disability genes form conserved functional modules in *Drosophila*. *PLoS Genet.* **9**, e1003911 (2013).
119. Green, E.W., Fedele, G., Giorgini, F. & Kyriacou, C.P. A *Drosophila* RNAi collection is subject to dominant phenotypic effects. *Nat. Methods* **11**, 222–223 (2014).
120. Vissers, J.H., Manning, S.A., Kulkarni, A. & Harvey, K.F. A *Drosophila* RNAi library modulates Hippo pathway-dependent tissue growth. *Nat. Commun.* **7**, 10368 (2016).
121. Kramer, J.M. *et al.* Epigenetic regulation of learning and memory by *Drosophila* EHMT/G9a. *PLoS Biol.* **9**, e1000569 (2011).
122. Schindelin, J. *et al.* Fiji: an open-source platform for biological-image analysis. *Nat. Methods* **9**, 676–682 (2012).
123. Nijhof, B. *et al.* A new Fiji-based algorithm that systematically quantifies nine synaptic parameters provides insights into *Drosophila* NMJ morphometry. *PLoS Comput. Biol.* **12**, e1004823 (2016).