Hotspots of missense mutation identify neurodevelopmental disorder genes and functional domains

Madeleine R Geisheker[1], Gabriel Heymann[2,3], Tianyun Wang[4], Bradley P Coe[1], Tychele N Turner[1], Holly A F Stessman[1,26], Kendra Hoekzema[1], Malin Kvarnung[5,6], Marie Shaw[7], Kathryn Friend[7,8], Jan Liebelt[9], Christopher Barnett[9,10], Elizabeth M Thompson[9,11], Eric Haan[9,11], Hui Guo[4], Britt-Marie Anderlid[5,6], Ann Nordgren[5,6], Anna Lindstrand[5,6], Geert Vandeweyer[12], Antonino Alberti[13], Emanuela Avola[13], Mirella Vinci[14], Stefania Giusto[15], Tiziano Pramparo[16], Karen Pierce[16], Srinivasa Nalabolu[16], Jacob J Michaelson[17], Zdenek Sedlacek[18], Gijs W E Santen[19], Hilde Peeters[20], Hakon Hakonarson[21–23], Eric Courchesne[16], Corrado Romano[13], R Frank Kooy[12], Raphael A Bernier[3], Magnus Nordenskjöld[5,6], Jozef Gecz[7,24], Kun Xia[4], Larry S Zweifel[2,3] & Evan E Eichler[1,25]

Although *de novo* missense mutations have been predicted to account for more cases of autism than gene-truncating mutations, most research has focused on the latter. We identified the properties of *de novo* missense mutations in patients with neurodevelopmental disorders (NDDs) and highlight 35 genes with excess missense mutations. Additionally, 40 amino acid sites were recurrently mutated in 36 genes, and targeted sequencing of 20 sites in 17,688 patients with NDD identified 21 new patients with identical missense mutations. One recurrent site substitution (p.A636T) occurs in a glutamate receptor subunit, GRIA1. This same amino acid substitution in the homologous but distinct mouse glutamate receptor subunit Grid2 is associated with Lurcher ataxia. Phenotypic follow-up in five individuals with *GRIA1* mutations shows evidence of specific learning disabilities and autism. Overall, we find significant clustering of *de novo* mutations in 200 genes, highlighting specific functional domains and synaptic candidate genes important in NDD pathology.

Multiple lines of evidence strongly support a genetic basis for autism spectrum disorders (ASD). *De novo* mutations, originating primarily in the parental germline, are individually rare but their collective risk is substantial and accounts for an estimated 30% of simplex ASD cases[1,2]. Most of the emphasis on identifying high-impact risk variants has focused on establishing burden for likely gene-disruptive (LGD) mutations (nonsense, frameshift or splice-site)[3–5]. High-impact risk genes with primarily *de novo* missense mutations have been understudied because a much smaller fraction (13%) are thought to be pathogenic when compared to *de novo* LGD mutations (42%)[1]. Moreover, *de novo* missense mutations are eight times as common, making it more challenging to prove their statistical relevance. Notwithstanding, a comparison of mutation rates in individuals with ASD and their unaffected siblings reveals that missense mutations contribute to disease risk in as many, if not more, cases than LGD mutations (12% versus 9%, respectively)[1].

The identification of genes with a substantial burden of missense mutations, then, is likely to highlight new classes of NDD risk genes.

[1]Department of Genome Sciences, University of Washington, Seattle, Washington, USA. [2]Department of Pharmacology, University of Washington, Seattle, Washington, USA. [3]Department of Psychiatry and Behavioral Sciences, University of Washington, Seattle, Washington, USA. [4]The State Key Laboratory of Medical Genetics, School of Life Sciences, Central South University, Changsha, Hunan, China. [5]Department of Molecular Medicine and Surgery, Center for Molecular Medicine, Karolinska Institutet, Stockholm, Sweden. [6]Department of Clinical Genetics, Karolinska University Hospital, Stockholm, Sweden. [7]Robinson Research Institute and the University of Adelaide at the Women's and Children's Hospital, North Adelaide, South Australia, Australia. [8]SA Pathology, Adelaide, South Australia, Australia. [9]South Australian Clinical Genetics Service, SA Pathology (at Women's and Children's Hospital), Adelaide, South Australia, Australia. [10]School of Paediatrics and Reproductive Health, University of Adelaide, Adelaide, South Australia, Australia. [11]School of Medicine, University of Adelaide, Adelaide, South Australia, Australia. [12]Department of Medical Genetics, University of Antwerp, Antwerp, Belgium. [13]Unit of Pediatrics & Medical Genetics, IRCCS Associazione Oasi Maria Santissima, Troina, Italy. [14]Laboratory of Medical Genetics, IRCCS Associazione Oasi Maria Santissima, Troina, Italy. [15]Unit of Neurology, IRCCS Associazione Oasi Maria Santissima, Troina, Italy. [16]University of California, San Diego, Autism Center of Excellence, La Jolla, California, USA. [17]Department of Psychiatry, The University of Iowa, Iowa City, Iowa, USA. [18]Department of Biology and Medical Genetics, Charles University 2nd Faculty of Medicine and University Hospital Motol, Prague, Czech Republic. [19]Department of Clinical Genetics, Leiden University Medical Center, Leiden, the Netherlands. [20]Centre for Human Genetics, KU Leuven and Leuven Autism Research, Leuven, Belgium. [21]Center for Applied Genomics, The Children's Hospital of Philadelphia, Philadelphia, Pennsylvania, USA. [22]Division of Genetics, The Children's Hospital of Philadelphia, Philadelphia, Pennsylvania, USA. [23]Department of Pediatrics, Perelman School of Medicine, University of Pennsylvania, Philadelphia, Pennsylvania, USA. [24]South Australian Health and Medical Research Institute, Adelaide, South Australia, Australia. [25]Howard Hughes Medical Institute, Seattle, Washington, USA. [26]Present address: Department of Pharmacology, Creighton University School of Medicine, Omaha, Nebraska, USA. Correspondence should be addressed to E.E.E. (eee@gs.washington.edu).

In some cases, this may reflect genes with such critical functions that LGD mutations are incompatible with life[1,6]. In other cases, the mutation's effect on the protein may differ. For example, missense mutations are more likely to have a gain-of-function effect[7] when compared to LGD mutations, which are predominantly loss-of-function. Clustering of missense mutations may highlight important and even previously unknown functional domains, providing insight into ASD pathogenesis and future downstream therapeutic targets. High-confidence ASD risk genes have been successfully identified by searching for mutation recurrence[3,4,8,9]. Given that missense mutations are more common and ~90% of them are thought to be incidental[1], a much larger sample size is required to prove pathogenicity. We took advantage of the substantial phenotypic and genotypic overlap between ASD, developmental delay and intellectual disability, epilepsy, congenital heart disease and schizophrenia[10] to study the pattern and distribution of *de novo* missense mutations more broadly. We focused on recurrent site and clustered mutations and tested a larger cohort of affected children to identify pathogenic events and implicate new missense 'hotspot' genes in NDD pathogenesis.

RESULTS

Properties of *de novo* missense mutations in NDD patients

We began by assessing the rates of *de novo* missense mutation in cases and controls. We identified a total of 5,807 *de novo* missense mutations in cases ($n = 8,477$) and 1,475 such events in controls ($n = 2,178$) (**Supplementary Table 1**). The fraction of probands with one or more event (50.7%) was significantly greater than the fraction of controls (47.8%; $P = 0.016$, odds ratio (OR) = 1.12 [1.02–1.24, 95% confidence interval (CI)], two-tailed Fisher's exact test) (**Fig. 1a**). As there were over three times as many cases as controls, we sought to limit the possibility that the signal was driven by rare outliers in cases and thus applied a secondary test, downsampling cases to match the number of controls. This further confirmed a significant increase in the rate of *de novo* missense mutations in cases (one-tailed empirical $P = 9.22 \times 10^{-4}$, OR = 1.12 [1.06–1.19, 95% CI, 1×10^6 permutations) (**Fig. 1a**). While the odds ratios for these two tests are nearly identical, the Fisher's exact test is considered more conservative and the hypergeometric distribution generates a wider confidence bound for the odds ratio when compared to that obtained by simulation.

Out of 4,227 genes with rare *de novo* missense mutations in cases, 974 (23.0%) harbored mutations in two or more unrelated cases (**Supplementary Table 2**). In contrast, among controls, 101 out of 1,362 genes (7.4%) were mutated recurrently (**Supplementary Table 3**). Matching the number of cases and controls, we observed a significant increase in the number of genes among cases with two or more (one-tailed empirical $P = 0.011$, OR = 1.26 [1.10–1.42, 95% CI], 1×10^6 permutations) and three or more (one-tailed empirical $P = 3.10 \times 10^{-5}$, OR = 3.13 [2.22–4.03, 95% CI], 1×10^6 permutations) *de novo* missense mutations (**Fig. 1b**). The increased recurrence rate is not explained by increased mRNA or protein length, as genes with recurrent mutations in cases were significantly shorter than those with recurrent mutations in controls (mRNA, $P = 5.19 \times 10^{-3}$; protein, $P = 1.47 \times 10^{-3}$; two-tailed Wilcoxon rank-sum tests). Additionally, the total number of genes with mutations was smaller among cases (1,323 in downsampled cases versus 1,362 in controls), suggesting that mutations in cases are not randomly distributed but rather cluster within fewer genes.

We next compared the severity of *de novo* missense mutations between cases and controls by assessing the Combined Annotation–Dependent Depletion (CADD) score[11]. The CADD score distribution was significantly positively skewed in cases compared to controls, consistent with an increase in deleteriousness ($P = 2.2 \times 10^{-4}$,
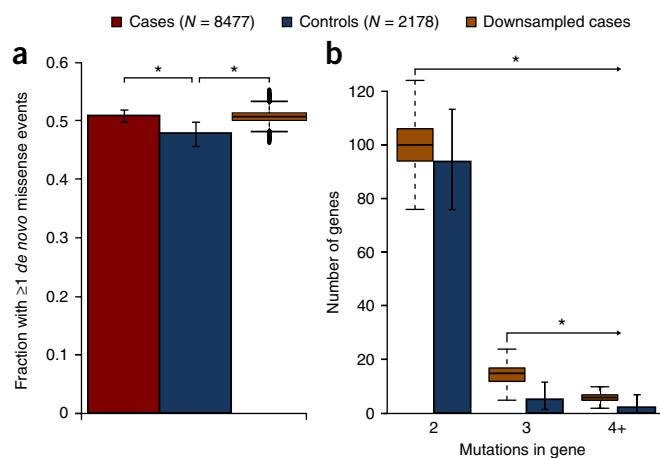

Figure 1 Burden and recurrence of *de novo* missense mutations. Bars for cases and controls represent observed data and error bars indicate the 95% CI for the observed proportions (Clopper-Pearson method). Box-and-whisker plots for downsampled cases represent the distribution of one million permutations. Boxes show interquartile range (IQR) with lines at the median, and whiskers are 1.5 times the IQR. *$P < 0.05$. (**a**) 4,301 out of 8,477 cases (50.7%) and 1,042 out of 2,178 controls (47.8%) have one or more *de novo* missense mutations (denovo-db v.0.9) that are rare in the general population (MAF < 0.1% in ESP). The fraction of individuals with one or more *de novo* missense mutation is significantly higher in cases ($P = 0.016$, OR = 1.12 [1.02–1.24, 95% CI], two-tailed Fisher's exact test) even after downsampling (empirical $P = 9.22 \times 10^{-4}$, OR = 1.12 [1.06–1.19, 95% CI]). (**b**) The number of genes with two or more mutations in downsampled cases is significantly greater than controls (empirical $P = 0.011$, OR = 1.26 [1.10–1.42, 95% CI]), as is the number of genes with three or more mutations (empirical $P = 3.1 \times 10^{-5}$, OR = 3.13 [2.22–4.03, 95% CI]).

two-tailed Wilcoxon rank-sum test). Further, at increasing minimum CADD score thresholds, the likelihood that an observed event can be attributed to a case increased (**Fig. 2a**). At a CADD threshold of 28, the likelihood rose dramatically (>1.2 positive likelihood ratio). Notably, mutations in genes with higher levels of recurrence in cases also showed significantly higher CADD scores ($P = 5.87 \times 10^{-29}$, $F = 45.12$, 3 d.f., one-way ANOVA), indicating that recurrence and severity are both valuable markers of missense pathogenicity and that they are highly correlated (**Fig. 2b**).

Genes with recurrent missense mutations

To further assess gene-specific recurrent mutations, we applied a probabilistic model that calculates the expected number of mutations in a gene, based on locus- and base-specific relative substitution rates[12,13] (see Online Methods). We identified 35 genes that had significantly more *de novo* missense mutations in cases than expected (false discovery rate < 5%) (**Supplementary Table 2**). Only two genes, *YIF1A* and *PHKA2*, reached significance in controls (**Supplementary Table 3**). For 17 of the genes significant in cases, an excess of loss-of-function mutations has already been established by copy number variants (CNVs) and LGD mutation (for example, *GRIN2B*, *PTEN* and *SCN2A*)[14–16]. For 13 of the remaining significant genes, no LGD mutations have been identified in the 24 cohorts studied here or in individuals with NDD in the Online Mendelian Inheritance in Man (OMIM; http://omim.org/) or ClinVar (https://www.ncbi.nlm.nih.gov/clinvar/) databases. While six of these missense-only genes are well known and associated with specific phenotypes (for example, *PACS1* and Schuurs-Hoeijmakers syndrome[17]), the remaining seven warrant follow-up.
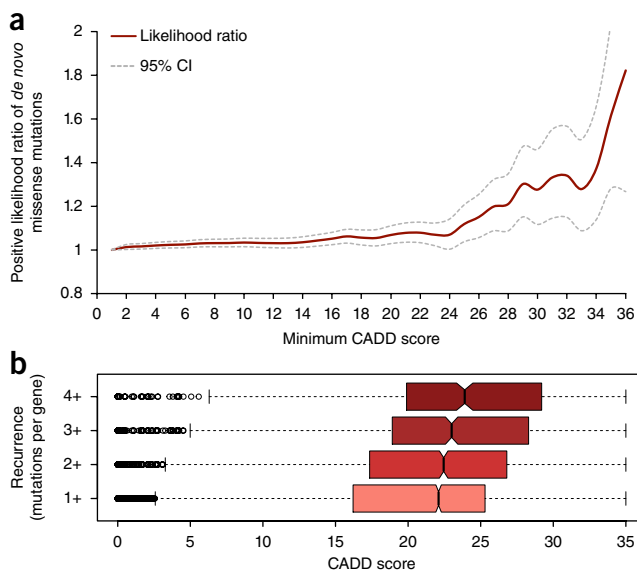
Figure 2 Severity of *de novo* missense mutations. (**a**) *De novo* missense mutations are more likely to be deleterious in cases ($n = 5,807$ mutations) than controls ($n = 1,475$ mutations), and the positive likelihood ratio increases as severity increases (as measured by CADD score). (**b**) The distribution of CADD scores skews significantly as the number of *de novo* missense mutations per gene in cases increases ($P = 5.87 \times 10^{-29}$, one-way ANOVA) indicating an enrichment for genes with pathogenic mutations. Boxes show IQR, with notches representing the 95% CI of the median; whiskers are 1.5 times the IQR. Circles are outliers.

As a set, the 35 genes with excess *de novo* missense mutations are enriched for aspects of neuronal communication such as post-synaptic membrane potential regulation (6 observed versus 0.17 expected, 35.3-fold enrichment, P_{adj} (adjusted *P*-value, Bonferroni corrected) $= 1.61 \times 10^{-4}$, two-tailed binomial test) and synaptic signaling (8 observed versus 0.7 expected, 11.4-fold enrichment, $P_{adj} = 3.30 \times 10^{-3}$, two-tailed binomial test), nervous system development (16 observed versus 3.7 expected, 4.4-fold enrichment, $P_{adj} = 1.05 \times 10^{-3}$, two-tailed binomial test) and gene expression regulation (3 observed versus 0.03 expected, >100-fold enrichment, $P_{adj} = 2.42 \times 10^{-2}$, two-tailed binomial test). There was also significant enrichment for genes involved in the presynapse (5 of 336 genes; $P = 3.74 \times 10^{-4}$, OR $= 9.25$ [3.40–∞, 95% CI], one-tailed Fisher's exact test) and postsynaptic density (11 of 1,755 genes; $P = 2.17 \times 10^{-4}$, OR $= 4.50$ [2.26–∞, 95% CI], one-tailed Fisher's exact test), and targets of FMRP (14 of 842 genes; $P = 1.20 \times 10^{-10}$, OR $= 14.37$ [7.57–∞, 95% CI], one-tailed Fisher's exact test).

In addition to recurrent mutations within the protein-coding portion of genes, we also assessed amino acids in which two or more *de novo* missense mutations in unrelated individuals with NDDs have been identified, hereafter referred to as sites. We identified 40 sites in 36 genes, 10 of which have a significant burden of *de novo* missense mutation, after excluding mutations observed in population controls (minor allele frequency (MAF) > 0.01% in the Exome Sequencing Project (ESP; NHLBI GO ESP Exome Variant Server, Seattle, WA; http://evs.gs.washington.edu/EVS/, August 2016) ($n = 6,503$) or the Exome Aggregation Consortium[18] (ExAC) database v.0.3 without neuropsychiatric disorders ($n = 45,376$)) (**Supplementary Table 4**). None of these mutations were observed in unaffected controls in denovo-db v.0.9. Seven sites had more than two recurrent mutations (for example, *PACS1* with six mutations at residue 203) and some genes (for example, *SCN2A*) had more than one recurrently mutated amino acid residue. Sixteen of the amino acid sites involved adjacent
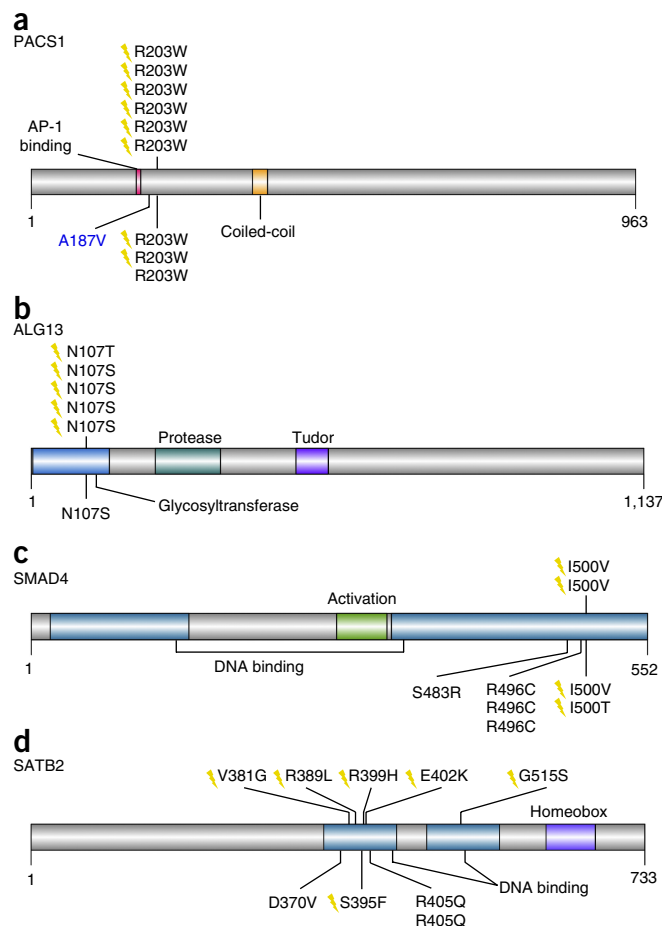
Figure 3 Recurrent mutations fall in or near functional domains. Published mutations in NDD patients are above the protein sequence and new mutations identified by targeted sequencing are below it. *De novo* mutations (lightning bolt) and paternally inherited mutations (blue) are indicated. Inheritance is unknown for the remaining mutations. Protein domains are from UniProt. (**a**) PACS1, NP_060496.2. (**b**) ALG13, NP_060936.1. (**c**) SMAD4, NP_005350.1. (**d**) SATB2, NP_001165988.1.

mutations in the same codon. Twenty-eight of the 40 sites (36 of 56 mutations) involved CpG dinucleotides, consistent with their known association with hotspots of single-nucleotide variation[19]. Thirty-four sites had average mutation CADD scores of 20 or greater and 17 had a score over 30, indicating that they are in the top 1% of deleterious mutations in the human genome. This observation stands in contrast to the pattern of *de novo* recurrent missense in controls, where only one of the three sites had a CADD score greater than 20, although the number of events compared is few.

Targeted sequencing of missense mutations

Using single-molecule molecular inversion probes (smMIPs), we targeted 20 of these recurrent sites for sequencing in a large cohort of 17,688 patients with a primary diagnosis of ASD or developmental delay (including intellectual disability; **Supplementary Tables 5** and **6**). The set included primarily patients with idiopathic NDDs not yet tested by exome sequencing. We also included a set of unaffected siblings and other unaffected individuals as an additional control ($n = 3,023$). We identified and validated 21 recurrent missense variants at 12 sites in 11 genes among cases (**Fig. 3a–c** and **Table 1**). No variants were observed at any of the 20 sites in controls. The inheritance status for only eight of the variants identified in cases could be

Table 1 New recurrent mutations at targeted missense sites

Gene	Site	Alternative amino acid(s)	Protein identifier	Mutations in denovo-db v.0.9 (N = 8,477)	Codon de novo P	Codon de novo P, genome-wide correction[a]	Mutations identified with smMIPs (N = 17,688)			Total de novo	Codon de novo P	Codon de novo P, genome-wide correction[a]	ExAC v.0.3 allele count (N = 45,376)
							Inherited*	De novo*	Unknown*				
PACS1	p.Arg203	Trp	NP_060496.2	6	1.03×10^{-24}	1.13×10^{-17}		2 (St, Tr)	1 (Ad)	8	1.37×10^{-29}	1.51×10^{-22}	1
PPP2R5D	p.Glu198	Lys	NP_006236.1	4	3.84×10^{-18}	4.22×10^{-11}			1 (Tr)	4	3.48×10^{-16}	3.83×10^{-9}	0
ALG13	p.Asn107	Ser, Thr	NP_060936.1	5	5.26×10^{-17}	5.79×10^{-10}			1 (Tr)	5	1.47×10^{-14}	1.62×10^{-7}	0
SCN2A	p.Arg937	Cys, His	NP_001035232.1	3	9.47×10^{-12}	1.04×10^{-04}		1 (ACGC)	1 (ACGC)	4	8.26×10^{-14}	9.09×10^{-7}	0
SMAD4	p.Ile500	Val, Thr	NP_005350.1	2	1.12×10^{-7}	1		2 (An, Le)		4	1.88×10^{-13}	2.07×10^{-6}	0
PTPN11	p.Gly503	Arg, Glu	NP_002825.3	3	1.66×10^{-11}	1.83×10^{-04}	1[b] (AGRE)			3	5.10×10^{-10}	5.62×10^{-3}	0
GRIA1	p.Ala636	Thr	NP_000818.2	2	1.11×10^{-7}	1		1 (St)		3	4.89×10^{-10}	5.39×10^{-3}	0
SCN2A	p.Arg379	His	NP_001035232.1	2	7.39×10^{-8}	8.14×10^{-01}			2 (St)	2	7.04×10^{-7}	1	0
CLCN4	p.Arg718	Trp, Gln	NP_001821.2	2	9.57×10^{-8}	1	1 (St)		1 (Ad)	2	9.11×10^{-7}	1	0/2[c]
KCNQ3	p.Arg230	Cys	NP_004510.1	2	3.36×10^{-7}	1			1 (Tr)	2	3.20×10^{-6}	1	0
ZNF215	p.Arg473	Gln	NP_037382.2	2	6.54×10^{-7}	1			3 (Ad)	2	3.49×10^{-6}	1	5
CUX2	p.Glu590	Lys	NP_056082.2	2	5.38×10^{-7}	1			1 (Tr)	2	5.12×10^{-6}	1	0

aBonferroni family-wise error rate correction based on 1.1×10^7 codons in genome. bIn two affected siblings. cAllele in denovo-db v.0.9 has 0 occurrences in ExAC; allele found with smMIPs has been seen twice. ACGC, Autism Clinical and Genetic Resources in China; Ad, Adelaide; AGRE, Autism Genetic Resource Exchange; An, Antwerp; Le, Leuven; St, Stockholm; Tr, Troina.

determined due to missing parental DNA. Six of these were *de novo* missense mutations (**Table 1**): PACS1 p.Arg203 (two mutations), GRIA1 p.Ala636, SCN2A p.Arg937, and SMAD4 p.Ile500 (two mutations). Of note, one of the inherited mutations (PTPN11 p.Gly503) is adjacent to the well-known Noonan syndrome recurrent mutation[20] (PTPN11 p.Ser502) and was transmitted paternally to two children both affected with ASD and intellectual disability. No information on the father's phenotype is currently available. Five genes corresponding to six sites were identified with two or more recurrent missense mutations in the NDD cohort, namely GRIA1 p.Ala636, PACS1 p.Arg203, SCN2A p.Arg379, SCN2A p.Arg937, SMAD4 p.Ile500, and ZNF215 p.Arg473. Phenotypic similarities are present in patients with shared mutations, such as those in ALG13 (**Fig. 3b**), where all six individuals with a mutation at residue 107 have both epilepsy and developmental delay even though they were recruited from cohorts with different primary diagnostic criteria. Both individuals with newly found mutations at SMAD4 p.Ile500 have features consistent with Myhre syndrome, including intellectual disability, short stature, facial dysmorphisms and hearing loss[21,22] (**Supplementary Clinical Case Reports**).

We also observed rare, potentially disruptive, missense variants in close proximity to the original recurrent site mutations, such as that in *SMAD4* (**Fig. 3c**), and therefore reexamined our database for regions where multiple *de novo* missense mutations mapped within 10 amino acids. We designed smMIPs for 17 clustered regions as well as the 20 recurrent sites (in 30 total genes) and sequenced this extended set (~5 kbp of coding sequence) in a subset of the NDD cohort (**Supplementary Tables 5** and **6**). Combined with targeted sites, we discovered a total of 139 recurrent or clustered missense variants in 137 cases compared to 7 variants in 6 unaffected controls, representing a significant enrichment ($P = 1.11 \times 10^{-4}$, OR = 3.93 [1.76–10.89, 95% CI], two-tailed Fisher's exact test) (**Table 2** and **Supplementary Table 7**). Twelve of the clustered missense mutations in cases were confirmed to be *de novo*, including events in *SATB2* (**Fig. 3d**), *GRIA1* (**Fig. 4a**), *SCN2A*, *KCNQ3*, *SCN8A*, *DEAF1* and *PPR2R1A* (**Supplementary Table 8**).

In addition to new variants at sites in denovo-db v.0.9, targeted sequencing established 14 new sites, although inheritance status for most variants remains unknown. The specific variants at SCN8A p.Arg1617 and STXBP1 p.Arg551 have been seen previously in NDD. While Myhre syndrome has been associated[22] only with residue 500 of SMAD4, *in silico* predictions suggest that the p.Arg496Cys mutations we identified are also likely to be pathogenic as the residue is highly conserved across species and the amino acid substitution is nonconservative[23]. Detailed phenotypic information on one patient with this mutation indicates characteristics of the syndrome, including intellectual disability, short stature and dysmorphic facial features, suggesting that Myhre syndrome is not only limited to one amino acid[22]. Phenotypic commonalities are also present among individuals with clustered mutations, indicating the functional relevance of protein domains. For example, seven out of eight patients with a mutation in the first DNA binding domain of SATB2 (**Fig. 3d**) have facial dysmorphisms and seven out of eight have developmental delay.

De novo missense mutations in *GRIA1*

We identified a recurrently mutated amino acid in *GRIA1* (**Fig. 4a**), encoding GluA1, a subunit of AMPA glutamate receptors, which was originally reported in one patient with intellectual disability[24] and another with ASD[14]. Both patients share a *de novo* G>A mutation resulting in an alanine-to-threonine amino acid replacement at residue 636 (NP_000818.2). Resequencing identified the same variant in three more patients with a primary diagnosis of ASD. One newly

Table 2 Rare[a] clustered missense mutations identified by targeted sequencing (CADD > 20)

| Gene	Site or cluster	Protein identifier	denovo-db v.0.9	Controls N	Controls Missense variants	Cases N	Cases Missense variants	N	All cases[b] Missense variants	All cases[b] Known de novo
PACS1	p.Arg203	NP_060496.2	6	3,023	0	17,688	4	26,165	10	8
PPP2R5D	p.Glu198	NP_006236.1	4	3,023	1	17,688	3	26,165	7	4
SCN2A	p.Arg937	NP_001035232.1	3	3,023	0	17,688	5	26,165	8	5
DEAF1	p.Gln264	NP_066288.2	2	3,023	0	17,688	7	26,165	9	5
ALG13	p.Asn107	NP_060936.1	5	3,023	0	17,688	1	26,165	6	5
GRIA1	p.Ala636	NP_000818.2	3	3,023	0	17,688	5	26,165	7	4
COL4A3BP	p.Ser260	NP_001123577.1	3	3,023	0	17,688	2	26,165	5	3
SCN8A	p.Gly214–p.Asn215	NP_055006.1	3	169	0	11,631	1	20,108	4	3
DEAF1	p.Leu219–p.Gly220	NP_066288.2	2	169	0	11,631	2	20,108	4	3
SATB2	p.Arg399–p.Glu402	NP_001165988.1	2	169	0	11,631	3	20,108	5	3
PPP2R1A	p.Arg182	NP_055040.2	2	3,023	0	17,688	2	26,165	4	3
SCN8A	p.Arg1617–p.Gly1625	NP_055006.1	2	169	0	11,631	2	20,108	4	3
PTPN11	p.Gly503	NP_002825.3	3	3,023	0	17,688	6	26,165	9	3
STXBP1	p.Gly544	NP_001027392.1	2	3,023	0	17,688	5	26,165	7	3
BCL11A	p.Thr47–p.Cys48	NP_075044.2	2	169	0	11,631	4	20,108	6	2
KCNQ3	p.Arg230	NP_004510.1	2	3,023	0	17,688	3	26,165	5	3
TRRAP	p.Trp1848	NP_003487.1	2	3,023	2	17,688	9	26,165	11	3
CLTCL1	p.Val641–p.Asn647	NP_009029.3	2	169	0	11,631	2	20,108	4	2
TRPM7	p.Thr379–p.Glu387	NP_060142.3	2	169	0	11,631	2	20,108	4	2
SCN2A	p.Arg853	NP_001035232.1	2	3,023	1	17,688	2	26,165	4	2
DUSP15	p.Thr4	NP_001012662.1	2	3,023	0	11,112	3	19,589	5	2
SCN2A	p.Arg379	NP_001035232.1	2	3,023	0	17,688	6	26,165	8	2
SATB2	p.Val381–p.Arg389	NP_001165988.1	2	169	0	11,631	1	20,108	3	2
SMAD4	p.Ile500	NP_005350.1	2	3,023	0	17,688	6	26,165	8	4
SLC35C2	p.Ser173–p.Gly176	NP_057029.8	2	169	0	11,631	3	20,108	5	2
CLCN4	p.Arg718	NP_001821.2	2	3,023	1	17,688	4	26,165	6	2
PCGF2	p.Pro65	NP_009075.1	2	3,023	0	17,688	5	26,165	7	2
KAT6B	p.Ser1380–p.Glu1389	NP_001243398.1	2	169	0	11,631	4	20,108	6	2
TRIO	p.Pro1461	NP_009049.2	2	3,023	0	17,688	1	26,165	3	2
NCAN	p.Pro1219–p.Val1221	NP_004377.2	2	169	0	11,631	8	20,108	10	2
ITPR1	p.Thr267–p.Arg269	NP_001161744.1	2	169	0	11,631	1	20,108	3	2
CUX2	p.Glu590	NP_056082.2	2	3,023	1	17,688	3	26,165	5	2
ZNF215	p.Arg473	NP_037382.2	2	3,023	1	11,112	10	19,589	12	2
SMARCA2	p.Arg525	NP_003061.3	2	3,023	0	17,688	3	26,165	5	2
TBR1	p.Trp271	NP_006584.1	1	3,023	0	17,688	3	26,165	4	1
PDCD11	p.Arg964	NP_055791.1	1	3,023	0	17,688	8	26,165	9	1
Total			83		7		139		222	101

[a]MAF < 0.01% in ExAC v.0.3 and ESP v.0.0.30. [b]denovo-db v.0.9 and smMIPs.

found mutation was confirmed as *de novo*; paternal DNA was not available for the other two but the mutation is not present in either of the patients' mothers. Using array comparative genomic hybridization, we found no evidence for large pathogenic CNVs in any of the three patients for whom we had DNA. While this position is a CpG dinucleotide and therefore prone to recurrent mutation, this variant has not been observed in 60,706 individuals studied by ExAC[18]. Moreover, we identified a second *de novo* missense mutation in close proximity (**Fig. 4a**) in a patient with developmental delay. The dearth of variants in healthy controls and the observation of the same recurrent variant in six unrelated patients (three of which were *de novo*; $P = 5.39 \times 10^{-3}$, one-tailed binomial test, genome-wide correction) suggested that the mutation is pathogenic.

The mutated site maps to the eighth position (p.Ala636) of a highly conserved nine-amino acid motif, SYTANLAAF (**Fig. 4b**), present in the M3 transmembrane domain of all glutamate receptors, which plays a critical role in channel gating[25]. The specific alanine-to-threonine mutation observed in the five patients here has been observed at the functionally equivalent site in other members of the glutamate receptor gene family. It was first identified as a spontaneous mutation in *Grid2* in a mouse line at Jackson Laboratories (Lurcher) that results in a constitutively active channel comprised of homomeric GluRδ2 subunits selectively expressed in cerebellar Purkinje neurons[26]. Mice heterozygous for this mutation in the GluRδ2 receptor

develop severe ataxia as a consequence of neurotoxicity from excess current flux. Notably, humans with the mutation in GluRδ2 also suffer from ataxia[27]. Engineering of the A>T mutation at the homologous site in the rat isoform of the GluA1 receptor produces a similar constitutively active phenotype with altered kinetic and pharmacological properties[28–30].

To confirm constitutive activity or leak current in the human isoform of *GRIA1* identified in affected patients, we synthesized cDNA encoding the human wild-type (WT) and mutant (A636T) at basepair position 1906 (G/A). Leak current was measured using whole-cell voltage-clamp recordings of HEK 293 cells heterologously expressing either WT or A636T in the absence of agonist by applying a voltage ramp from −100 mV to +80 mV. GluA1-mediated current was confirmed by application of the AMPA receptor-selective antagonist 2,3-dihydroxy-6-nitro-7-sulfamoyl-benzo[f]quinoxaline-2,3-dione (NBQX), followed by an additional voltage ramp. Subtracted current in the presence of NBQX revealed a notable constitutive current in A636T-expressing but not WT-expressing cells (**Fig. 4d,f**). Consistent with GluA1-mediated current, inward rectification was abolished following channel blockade with NBQX. No changes in current magnitude or shape were seen in cells expressing the WT channel after NBQX application (**Fig. 4c,f**). Affected patients with the A636T mutation were heterozygous, indicating that a majority of receptors likely contain WT and A636T receptor subunits. To assess the functional
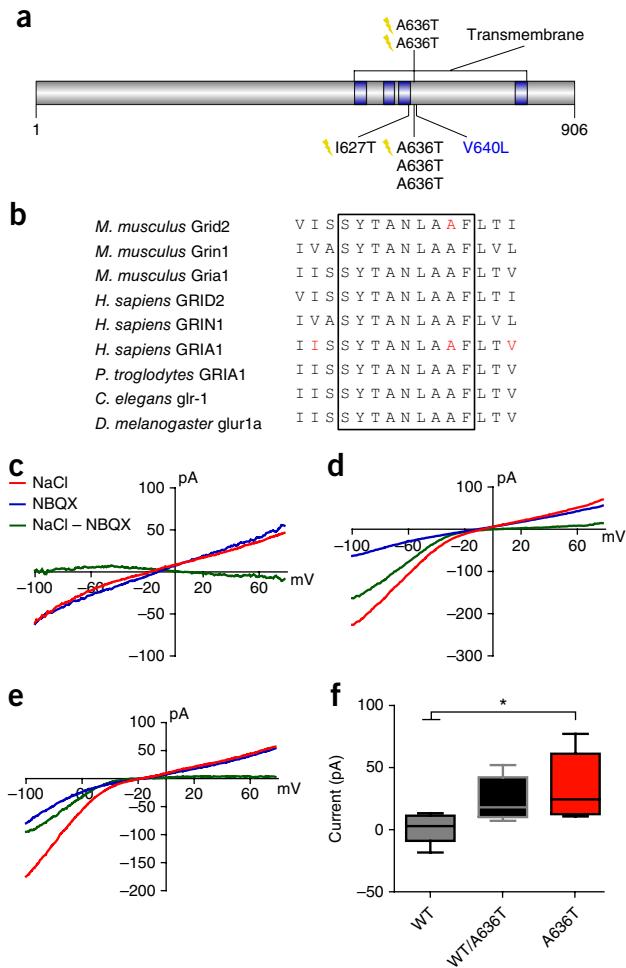
Figure 4 Functional effect of recurrent GRIA1 missense mutations. (**a**) Linear representation of annotated domains in the protein GRIA1, also known as GluA1 (NP_001244950.1), as defined in UniProt. (**b**) A recurrent mutation observed only in NDD cases (*n* = 5 patients) falls within a highly conserved M3 transmembrane domain, important in channel gating. The alanine mutated in these patients (red) is homologous to the one that causes severe ataxia in Lurcher mice in the δ2 subunit of this receptor (Grid2). (**c**–**e**) Example current traces from a 1.8-s voltage-ramp from −100 mV to +80 mV for (**c**) WT, (**d**) A636T and (**e**) heteromeric WT/A636T transfected HEK cells. The three current traces per panel correspond to voltage-ramp currents in the presence of normal extracellular solution (NaCl), extracellular solution supplemented with 50 µM NBQX (NBQX), and the isolated GluA1-dependent current determined by subtracting the NBQX current from the NaCl current (NaCl − NBQX). (**f**) Average leak current at −60 mV. The GluA1-mediated current (NaCl − NBQX) was determined at −60 mV and averaged across cells (*n* = 5 (WT), 7 (A636T), and 5 (heteromeric); *P* = 0.024, *F* = 4.91, 2 d.f., one-way ANOVA). Data are mean ± s.e.m.

phenotype of these heteromeric receptors, we cotransfected equal ratios of WT and A636T DNA and performed the same voltage-ramp recordings (**Fig. 4e**). While a noticeable constitutive current was still present, it was smaller than with the homomeric A636T channel, demonstrating that the overall effects of the mutation are mitigated by the presence of the WT subunits (**Fig. 4f**).

Consistent with the prevalent role of GluA1 homomeric channels in synapse development and synaptic plasticity[31], phenotypic analysis of four of the individuals with the A636T mutation demonstrated common features (**Supplementary Table 9**), including mild
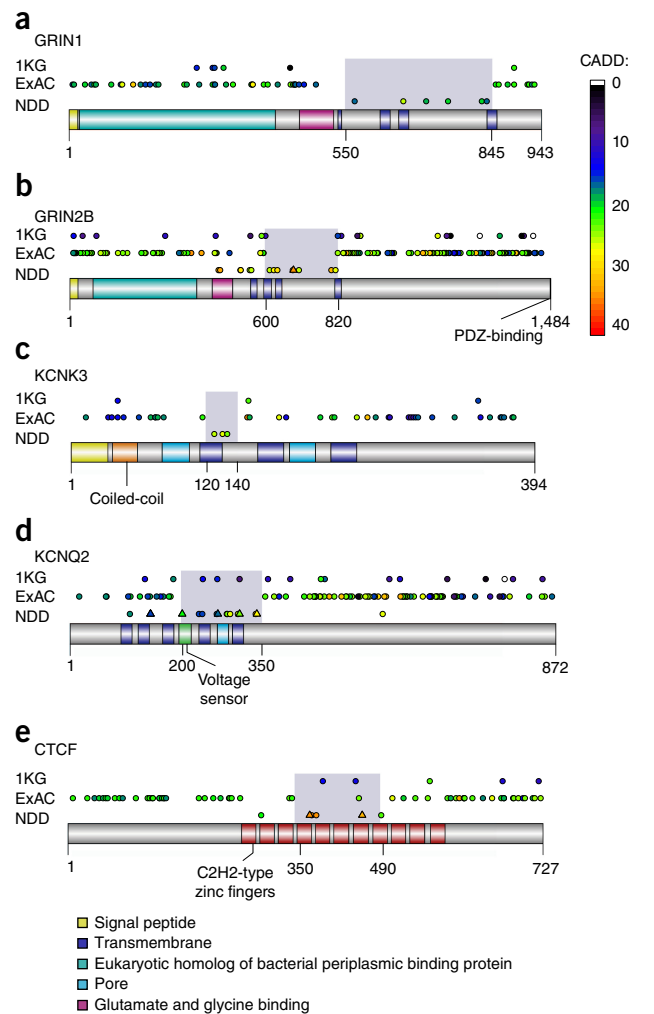
Figure 5 Proteins with excessive clustering of missense mutations in NDD cases. The pattern of *de novo* missense mutations in cases with NDDs is contrasted with that of rare missense variants from the 1000 Genomes Project (1KG) and private missense mutations from ExAC excluding those associated with neuropsychiatric diagnoses. Missense mutations are colored by severity (CADD heat map) and recurrent *de novo* mutations at a specific amino acid position are indicated (triangle). Significance of clustering was calculated based on comparison to ExAC using CLUMP. (**a**) GRIN1 (NP_001172019.1) shows greater missense mutation clustering in NDD patients (CLUMP = 1.68, *P* = 0.013), with region-specific significance corresponding to the transmembrane domains (amino acids 550–845; Fisher's exact test *P* = 5.6×10^{-8}). (**b**) Similarly, missense mutations cluster for GRIN2B (NP_000825.2; CLUMP = 1.34, *P* = 0.003), in particular between the second and fourth transmembrane domains (amino acids 600–820; *P* = 2.0×10^{-9}). (**c**) KCNK3 (NP_002237.1) patient missense mutations cluster (CLUMP = 0.54, *P* = 0.036) near the first transmembrane domain (amino acids 120–140, *P* = 9.4×10^{-5}, OR = ∞). The average per-base rate of ExAC samples with ≥10-fold coverage across the exon harboring mutations in cases was 79.1%. (**d**) KCNQ2 (NP_742105.1) shows several missense mutation hotspots (CLUMP = 0.36, *P* < 1×10^{-3}) corresponding to the pore and voltage sensor of the channel (amino acids 200–350, *P* = 2.0×10^{-14}). (**e**) Finally, patients show more severe CTCF (NP_006556.1) missense mutations that cluster (CLUMP = 1.0, *P* = 0.007) at two locations between the fourth and seventh C2H2 zinc finger motifs (amino acids 350–490, *P* = 9.1×10^{-8}).

to moderate intellectual disability (all four individuals) and ASD (three of the four). Three of the four for whom information is available had delayed language development, with two (both with ASD)

demonstrating persistent difficulties with pronunciation and vocabulary. These two individuals were also noted to have highly similar facial features and were diagnosed with ADHD. Similarly, the individual without ASD was noted to have behavioral dysfunction. Two individuals also had delayed motor development. All four had normal MRIs. Collectively our evidence suggests that this specific missense mutation dictates a common pathological brain development trajectory and supports the idea that specific missense mutations contribute to NDD pathogenesis.

Clustered missense mutations and functional domains

Our sequencing results as well as the *GRIA1* analysis strongly suggest that clustered and recurrent missense mutations have the potential to highlight functional protein domains important in NDD pathology. We previously developed a tool, CLUMP[7], to assess the significance of clustered mutations, and we applied it to an updated version of denovo-db (v.1.2) to identify genes and functional domains for future investigation. Overall, we examined 8,917 *de novo* missense mutations in cases and calculated raw CLUMP scores for 1,699 proteins containing at least two mutations in cases. We performed case–control analyses comparing the pattern of private alleles in ExAC and separately among European individuals from the 1000 Genomes Project (see Online Methods). Twenty-eight out of 34 genes we initially identified were testable by this approach and 18 of them showed nominally significant clustering of *de novo* missense mutations ($P < 0.05$, CLUMP, one-tailed permutation test). Altogether, we identified 200 genes with significant clustering of missense mutations at the protein level (**Supplementary Table 10**). Once again, this set was significantly associated with aspects of neuronal communication, including regulation of the postsynaptic potential (11 observed versus 1 expected, 11.0-fold enrichment, $P_{adj} = 6.93 \times 10^{-5}$, two-tailed binomial test) and synaptic signaling (20 observed versus 4.15 expected, 4.82-fold enrichment, $P_{adj} = 8.38 \times 10^{-5}$, two-tailed binomial test), as well as chromatin-mediated maintenance of transcription (4 observed versus 0.1 expected, 40.7-fold enrichment, $P_{adj} = 2.96 \times 10^{-2}$, two-tailed binomial test). Many of the genes encoded channel proteins and receptors (for example, *GRIA1*, *GRIN1*, *GRIN2A*, *GRIN2B*, *KCNH1*, *KCNQ2*) and exhibited clustering in or near specific functional domains, such as the transmembrane, pore or voltage sensor domains (**Fig. 5a–d**). Other proteins, such as CTCF, were remarkable in that the clustering pattern of missense mutations in cases highlighted a subset of the C2H2 zinc finger motifs, which were never mutated in controls (**Fig. 5e**). These pockets of patient-only missense mutations will be increasingly important in characterizing pathogenic genes and functional domains.

DISCUSSION

The objective of this research study was twofold: define the features of likely disease-causing *de novo* missense mutations and identify new genes and functional domains relevant to the pathology of NDDs. To increase sample size, we broadly defined NDDs to include not only data from patients with ASD, developmental delay and intellectual disability but also patients with epilepsy and schizophrenia because of the extensive comorbidity of these diagnoses. As expected, both recurrence and severity of missense mutations were critical features. The likelihood of a pathogenic mutation rose significantly when three or more missense mutations were observed in a gene ($P = 1.06 \times 10^{-18}$, two-tailed Wilcoxon rank-sum test) and, in particular, when the severity of the missense mutation exceeded a CADD score of 28 (>1.2 positive likelihood ratio). We use these features to identify 35 genes with an excess (false discovery rate < 5%) of *de novo* missense mutations (**Supplementary Table 2**). Targeted sequencing of specific

protein-coding regions showed that recurrent and clustered amino acid replacements were more common in cases than controls ($P = 1.11 \times 10^{-4}$, OR = 3.93 [1.76–10.89, 95% CI], two-tailed Fisher's exact test). While many of the top-scoring genes were associated with known syndromic and nonsyndromic forms of NDD (for example, *SCN2A* with ASD[32], *PACS1* with Schuurs-Hoeijmakers syndrome[17] and *ALG13* with epilepsy[33]), seven of these candidates have not been previously reported in ClinVar or OMIM. We also identified 200 genes with patterns of *de novo* missense mutations that were more clustered in cases than in population controls (**Supplementary Table 10**), 79% (*n* = 157) of which have not yet been associated with an NDD in OMIM or ClinVar databases.

Among the 35 genes with a significant excess of recurrent missense mutations, 37% (*n* = 13) have not yet been associated with a *de novo* LGD mutation (for example, *COL4A3BP*, *PPP2R5D*), suggesting that LGD events either are not tolerated or are associated with a different diagnostic outcome. In support of this observation, 71% (*n* = 25) of genes were also recently highlighted as likely pathogenic in an exome sequencing study of 3,287 individuals with developmental delay[15]. Of the 200 genes with significant clustering of missense mutations, 67% (*n* = 134) did not show any evidence of LGD mutation in NDDs in denovo-db v.1.2, OMIM or ClinVar; 45% (*n* = 89) have been shown to be loss-of-function intolerant in the ExAC database[18], suggesting that LGD mutations in them may be genetically lethal (for example, cause embryonic lethality or infertility), although more experiments will be required to make this determination. In many cases, the clustering of *de novo* mutations highlights protein functional domains (**Fig. 5**), such as specific zinc-finger motifs (for example, in *CTCF*), transmembrane domains (for example, in *GRIN1*) and voltage sensors and channel pores (for example, in *KCNQ2*). As the number of exomes increases, these hotspots of pathogenic missense mutation will become more transparent and may be better understood in the context of protein structure. *PTPN11*, associated with Noonan syndrome[20], is predicted, for example, to have three clusters by CLUMP, and three-dimensional protein structure analysis revealed that these three clusters define the cleft of the ligand binding site[34] (**Supplementary Fig. 1**).

Genes associated with hotspots of missense mutation (**Supplementary Table 10**) were particularly enriched for presynaptic active zone proteins, FMRP-binding targets and covalent chromatin modification, although not CHD8 target genes. Accumulating evidence supports a link between the development and function of excitatory synapses in NDD and ASD[35]. Consistent with this, we found 35-fold and 11-fold enrichments of genes regulating postsynaptic membrane potential in genes that carried a significant burden and genes with significant clustering of *de novo* missense events, respectively. While several scaffolding and intracellular signaling proteins have been associated with ASD and disruption of synaptic function, including SH3 and multiple ankyrin repeat domain (SHANK) proteins[36], synaptic Ras GTPase-activating (SYNGAP) proteins[37], neurexins[38], neuroligins[39] and others[35], a functional mutation in an essential pore-forming subunit of an excitatory ionotropic glutamate receptor has not been described to our knowledge.

The fact that five patients with phenotypic similarity were identified with a gain-of-function A636T mutation strongly supports a role for *GRIA1* in ASD and related NDDs. This specific *de novo* missense mutation has been observed before[26] at the homologous position in a highly conserved motif in a different glutamate receptor, GluRδ2. The mutation has a gain-of-function effect, causing constitutive channel opening, neurotoxicity, and degeneration of the cerebellar Purkinje cells in which GluRδ2 is selectively expressed[40]. Both mice and humans with this mutation in GluRδ2 develop ataxia as a direct

consequence[27]. This mutation in rodent GluA1 (the product of *Gria1*) has the same effect on channel gating[29,30], and here we have replicated this finding in human GluA1. As GluA1 is important in learning and memory[31], there is a biologically plausible link between this *de novo* missense mutation in *GRIA1* and intellectual disability.

GRIA1 has been demonstrated to play a key role in early synapse development, with GluA1 homomeric channels being inserted into nascent synapses to provide a calcium-permeable, high-conductance channel before being replaced by GluA2-containing channels that mediate long-term synaptic connectivity. Continuing into adulthood, long-term potentiation of excitatory synapses, associated with learning and memory, requires initial insertion of GluA2-absent, calcium-permeable AMPA receptors followed by replacement with GluA2-containing receptors[31]. The developmental and adult function of GluA1 in these contexts likely contributes to the intellectual disability associated with this mutation. Loss-of-function of GluA1 in *Gria1* knockout mice leads to impaired synaptic function[41] and behavioral phenotypes, including social behavior deficits and impulsivity[42], which suggests that bidirectional aberration in excitatory signaling can result in similar ASD and NDD phenotypes. Future studies investigating the impact of the gain-of-function, Lurcher-like A636T mutation in synapse development and function will shed further light on how alterations in excitatory synaptic function contribute to ASD.

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.

AUTHOR CONTRIBUTIONS
E.E.E., L.S.Z., M.R.G., G.H., T.N.T., B.P.C., H.A.F.S. and K.X. designed the study; M.R.G., G.H., T.N.T., B.P.C., T.W. and K.H. performed the experiments; B.P.C. and T.N.T. helped with MIP design and data analysis; M.K., M.N., M.S., J.G., C.B., E.M.T., G.V., R.F.K., T.P., S.N., H.P., C.R., R.A.B., K.X. and H.H. tested inheritance and provided clinical follow-up on select patients; other authors participated in the sample collection and DNA extraction and/or preparation. M.R.G., E.E.E., L.S.Z., G.H., B.P.C. and T.N.T. wrote the manuscript with input from all authors.

1. Iossifov, I. *et al.* The contribution of *de novo* coding mutations to autism spectrum disorder. *Nature* **515**, 216–221 (2014).
2. Ronemus, M., Iossifov, I., Levy, D. & Wigler, M. The role of *de novo* mutations in the genetics of autism spectrum disorders. *Nat. Rev. Genet.* **15**, 133–141 (2014).
3. Bernier, R. *et al.* Disruptive CHD8 mutations define a subtype of autism early in development. *Cell* **158**, 263–276 (2014).
4. 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).
5. Sanders, S.J. *et al.* Insights into autism spectrum disorder genomic architecture and biology from 71 risk loci. *Neuron* **87**, 1215–1233 (2015).
6. Packer, A. Neocortical neurogenesis and the etiology of autism spectrum disorder. *Neurosci. Biobehav. Rev.* **64**, 185–195 (2016).
7. Turner, T.N. *et al.* Proteins linked to autosomal dominant and autosomal recessive disorders harbor characteristic rare missense mutation distribution patterns. *Hum. Mol. Genet.* **24**, 5995–6002 (2015).
8. van Bon, B.W.M. *et al.* Disruptive *de novo* mutations of *DYRK1A* lead to a syndromic form of autism and ID. *Mol. Psychiatry* **21**, 126–132 (2016).
9. Helsmoortel, C. *et al.* A SWI/SNF-related autism syndrome caused by *de novo* mutations in *ADNP*. *Nat. Genet.* **46**, 380–384 (2014).
10. Buxbaum, J.D. DSM-5 and psychiatric genetics — round hole, meet square peg. *Biol. Psychiatry* **77**, 766–768 (2015).
11. Kircher, M. *et al.* A general framework for estimating the relative pathogenicity of human genetic variants. *Nat. Genet.* **46**, 310–315 (2014).
12. 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).
13. O'Roak, B.J. *et al.* Multiplex targeted sequencing identifies recurrently mutated genes in autism spectrum disorders. *Science* **338**, 1619–1622 (2012).
14. De Rubeis, S. *et al.* Synaptic, transcriptional and chromatin genes disrupted in autism. *Nature* **515**, 209–215 (2014).
15. McRae, J.F. *et al.* Prevalence and architecture of *de novo* mutations in developmental disorders. *Nature* http://dx.doi.org/10.1038/nature21062 (2017).
16. Stessman, H.A.F., Turner, T.N. & Eichler, E.E. Molecular subtyping and improved treatment of neurodevelopmental disease. *Genome Med.* **8**, 22 (2016).
17. Schuurs-Hoeijmakers, J.H.M. *et al.* Recurrent *de novo* mutations in *PACS1* cause defective cranial-neural-crest migration and define a recognizable intellectual-disability syndrome. *Am. J. Hum. Genet.* **91**, 1122–1127 (2012).
18. Lek, M. *et al.* Analysis of protein-coding genetic variation in 60,706 humans. *Nature* **536**, 285–291 (2016).
19. Lynch, M. Rate, molecular spectrum, and consequences of human mutation. *Proc. Natl. Acad. Sci. USA* **107**, 961–968 (2010).
20. Maheshwari, M. *et al.* PTPN11 mutations in Noonan syndrome type I: detection of recurrent mutations in exons 3 and 13. *Hum. Mutat.* **20**, 298–304 (2002).
21. Myhre, S.A., Ruvalcaba, R.H.A. & Graham, C.B. A new growth deficiency syndrome. *Clin. Genet.* **20**, 1–5 (1981).
22. Le Goff, C. *et al.* Mutations at a single codon in Mad homology 2 domain of *SMAD4* cause Myhre syndrome. *Nat. Genet.* **44**, 85–88 (2011).
23. Landrum, M.J. *et al.* ClinVar: public archive of interpretations of clinically relevant variants. *Nucleic Acids Res.* **44**, D1, D862–D868 (2016).
24. de Ligt, J. *et al.* Diagnostic exome sequencing in persons with severe intellectual disability. *N. Engl. J. Med.* **367**, 1921–1929 (2012).
25. Yuan, H., Erreger, K., Dravid, S.M. & Traynelis, S.F. Conserved structural and functional control of *N*-methyl-d-aspartate receptor gating by transmembrane domain M3. *J. Biol. Chem.* **280**, 29708–29716 (2005).
26. Zuo, J. *et al.* Neurodegeneration in Lurcher mice caused by mutation in δ2 glutamate receptor gene. *Nature* **388**, 769–773 (1997).
27. Coutelier, M. *et al.* GRID2 mutations span from congenital to mild adult-onset cerebellar ataxia. *Neurology* **84**, 1751–1759 (2015).
28. Kohda, K., Wang, Y. & Yuzaki, M. Mutation of a glutamate receptor motif reveals its role in gating and δ2 receptor channel properties. *Nat. Neurosci.* **3**, 315–322 (2000).

29. Taverna, F. *et al*. The Lurcher mutation of an α-amino-3-hydroxy-5-methyl-4-isoxazolepropionic acid receptor subunit enhances potency of glutamate and converts an antagonist to an agonist. *J. Biol. Chem.* **275**, 8475–8479 (2000).

30. Klein, R.M. & Howe, J.R. Effects of the lurcher mutation on GluR1 desensitization and activation kinetics. *J. Neurosci.* **24**, 4941–4951 (2004).

31. Kessels, H.W. & Malinow, R. Synaptic AMPA receptor plasticity and behavior. *Neuron* **61**, 340–350 (2009).

32. Tavassoli, T. *et al*. *De novo SCN2A* splice site mutation in a boy with autism spectrum disorder. *BMC Med. Genet.* **15**, 35 (2014).

33. Allen, A.S. *et al*. *De novo* mutations in epileptic encephalopathies. *Nature* **501**, 217–221 (2013).

34. Niknafs, N. *et al*. MuPIT interactive: webserver for mapping variant positions to annotated, interactive 3D structures. *Hum. Genet.* **132**, 1235–1243 (2013).

35. Banerjee, S., Riordan, M. & Bhat, M.A. Genetic aspects of autism spectrum disorders: insights from animal models. *Front. Cell. Neurosci.* **8**, 58 (2014).

36. Durand, C.M. *et al*. Mutations in the gene encoding the synaptic scaffolding protein SHANK3 are associated with autism spectrum disorders. *Nat. Genet.* **39**, 25–27 (2007).

37. Hamdan, F.F. *et al*. *De novo SYNGAP1* mutations in nonsyndromic intellectual disability and autism. *Biol. Psychiatry* **69**, 898–901 (2011).

38. Peñagarikano, O. *et al*. Absence of CNTNAP2 leads to epilepsy, neuronal migration abnormalities, and core autism-related deficits. *Cell* **147**, 235–246 (2011).

39. Varoqueaux, F. *et al*. Neuroligins determine synapse maturation and function. *Neuron* **51**, 741–754 (2006).

40. Zanjani, H.S. *et al*. Death and survival of heterozygous Lurcher Purkinje cells *in vitro*. *Dev. Neurobiol.* **69**, 505–517 (2009).

41. Andrásfalvy, B.K., Smith, M.A., Borchardt, T., Sprengel, R. & Magee, J.C. Impaired regulation of synaptic strength in hippocampal neurons from GluR1-deficient mice. *J. Physiol. (Lond.)* **552**, 35–45 (2003).

42. Wiedholz, L.M. *et al*. Mice lacking the AMPA GluR1 receptor exhibit striatal hyperdopaminergia and 'schizophrenia-related' behaviors. *Mol. Psychiatry* **13**, 631–640 (2008).

ONLINE METHODS

Exome data sets and missense mutation annotation. We initially analyzed all *de novo* missense mutations available from 24 published cohorts[1,13,14,24,32,33,43–63] of *de novo* mutations in individuals with NDDs (denovo-db v.0.9; **Supplementary Table 11**)[64]. The NDD set included 8,477 individuals diagnosed with ASD, developmental delay, intellectual disability, epilepsy, schizophrenia or congenital heart disease (CHD), as well as four cohorts of unaffected controls[1,47,65] ($n = 2,178$) (**Supplementary Table 1**). Only patients with CHD from Homsy *et al.*[50] with a secondary diagnosis of NDD were included in this study; we also excluded unaffected siblings of ASD patients as controls if they had a Social Responsiveness Scale (SRS) score ≥60 to remove controls on the autism spectrum[66]. Variants were annotated with SeattleSeq[67] version 138, which provides annotation for all available RefSeq transcripts in GRCh37/hg19. In the case of multiple transcripts, we selected the transcript for which the majority of missense mutations were annotated in both cases and controls. All *de novo* missense mutations either were previously validated or investigators relied on a high (>95%) validation rate in a subset of mutations to ensure specificity. As some individuals with ASD were assayed as part of multiple cohorts, we took care to remove any duplicate entries. When possible, we compared the global identifier given to the samples that were housed at Rutgers (RUID). Three duplicate entries were found in this manner. For other shared mutations in ASD cohorts, we performed PCR amplification and Sanger sequencing on in-house DNA samples to confirm secondary variants. Five out of six pairs tested—two ASC (Autism Sequencing Consortium)–SSC (Simons Simplex Collection) pairs and three ASC–TASC (The Autism Simplex Collection) pairs—shared a second variant and we therefore assumed them to be duplicates. The presence of uniquely identifying secondary-site mutations was also used to eliminate potential duplicates for globally dispersed samples. We excluded high-frequency mutations (MAF > 0.1%) observed in NHLBI GO ESP Exome Variant Server (Exome Variant Server, NHLBI GO Exome Sequencing Project (ESP), Seattle, WA; http://evs.gs.washington.edu/EVS/, August 2016).

Statistical analyses. Wherever possible, non-parametric tests were used. Data collection and analysis were not performed blind to the conditions of the experiments. Burden was compared between cases and controls for rare (MAF < 0.1% in ESP) *de novo* missense mutations. Comparisons in rate of mutation and gene recurrence were made using two-tailed Fisher's exact tests. For comparisons of mutation rate and recurrence that depended on identical numbers of cases and controls, we performed 1 million downsamplings and used permutation tests, reporting the empirical *P*-values. Data distribution was assumed to be normal but was not formally tested. To identify significant enrichments for missense mutations within genes and genic regions, we applied a probabilistic model that incorporates sequence context and human–chimpanzee fixed differences to generate a null model for the distribution of missense variation across the genome and applied a one-tailed binomial test to test for enrichment[13]. For examination of individual codons and specific target regions, we applied the same method but restricted it to the sequence context of the target region and normalized by the gene-specific human–chimpanzee divergence. For all tests we assumed a mutation rate of 1.8 *de novo* coding variants per generation[12]. Multiple-testing corrections were applied using two procedures based on the analysis type. For significance calculations of whole genes, we used the Benjamini-Hochberg false discovery rate correction based on an estimated 19,000 genes in the human genome[68] and report the *q*-values for each test. For codon analysis we applied the conservative Bonferroni family-wise error rate correction based on the number of amino acids in the genome ($n = 1.1 \times 10^7$) to generate genome-wide significance estimates and report the adjusted *P*-value (P_{adj}). Gene ontology enrichment was assessed using PANTHER (database 2017-04-13) for GO biological process annotation and corrected for multiple testing (Bonferroni, reported as P_{adj}). We also applied a one-tailed Fisher's exact test for testing the enrichment of specific gene sets, including neuronal compartments such as the postsynaptic density[69] and targets of CHD8 (ref. 70) and FMRP in brain tissue[71].

Targeted sequencing. smMIPs[72] were designed with the MIPgen program[73] to capture sequences of interest. To maximize coverage, we designed one smMIP for each strand for each target. We first used smMIPs targeting 24 sites to sequence eight cohorts containing a total of 6,057 cases and 2,854 controls. We also used smMIPs targeting two amino acids thought to be sites (amino acids with *de novo* missense mutations in two or more unrelated individuals with NDD) but later

discovered to be duplicate database entries (*TBR1*) or present in both the case and her unaffected sister (*PDCD11*). As clusters of missense mutations have been associated with NDDs[74], we then designed a set of smMIPs targeting 17 clusters in denovo-db v.0.9 (**Supplementary Table 11**). These cluster smMIPs, along with the 24 site smMIPs, were used on four more cohorts, containing 5,055 cases and 169 controls. A final set of smMIPs was created that excluded those targeting four sites that had no brain expression (*AGER, ZNF215*), low CADD scores (*ALDH5A1*) or high frequency in control populations (*DUSP15*). This final set, targeting 20 sites and 17 clusters, was used on five new cohorts, containing a total of 6,576 cases (**Supplementary Tables 5** and **6**). Across all three designs, this totaled to 17,688 cases and 3,023 controls. The study size was not predetermined but based on the maximum number of samples that could be screened. Reads were aligned using BWA-MEM[75] to GRCh37/hg19. All 146 rare (MAF < 0.01%) variants with CADD score >20 were validated with Sanger sequencing (**Supplementary Table 7**). Patients were initially identified through targeted sequencing in anonymized ASD and developmental delay cohorts. All patients consented to sequencing and recontacting for inheritance testing at the providing laboratory. Patient samples were acquired from Adelaide (J.G., University of Adelaide), Antwerp (R.F.K., University of Antwerp), Autism Clinical and Genetic Resources in China (ACGC; K.X.), the Autism Genetic Resource Exchange (AGRE), Iowa (J.J.M., University of Iowa), Leiden (G.W.E.S., Leiden University Medical Center), Leuven (H.P., University Hospitals Leuven), Philadelphia (H.H., Children's Hospital of Philadelphia), Prague (Z.S., University Hospital Motol), San Diego (E.C., UC San Diego), Simons Simplex Collection (SSC), Stockholm (M.N., Karolinska University Hospital), the Study of Autism Genetics Exploration (SAGE; R.A.B., University of Washington), The Autism Simplex Collection (TASC) and Troina (C.R., Associazione Oasi Maria Santissima).

***GRIA1* transfection and patch-clamp recording assays.** *GRIA1* wild-type and A636T mutant DNA sequences were synthesized (GenScript) and cloned into mammalian expression vectors. Human embryonic kidney (HEK) 293T/17 cells (ATCC CRL-11268), routinely used for transient transfection and electrophysiological recordings as they allow robust heterologous expression, were cultured in DMEM (Invitrogen) supplemented with 10% fetal bovine serum and 1% streptomycin up to a maximum passage number of 15. Cells were validated and certified as mycoplasma free by ATCC (Manassas, VA). For transient transfection, cells were split and plated onto 12-mm glass coverslips (Carolina Scientific) coated with poly-L-lysine (50 ng/μl). Then, 4–6 h later, approximately 0.6 μg of total DNA per coverslip was transfected using the Fugene6 reagent (2 μl per coverslip, Promega). For heteromeric cells, approximately 0.3 μg of WT and A636T were cotransfected. Whole-cell recordings were performed approximately 60 h after transfection using a Multiclamp 700B amplifier (Molecular Devices) with glass micropipettes of resistance 2–5 MΩ. Extracellular solution contained (in mM) 150 NaCl, 2.5 CaCl$_2$, 2.5 KCl, 1 MgCl$_2$, 10 D-glucose, 10 HEPES, pH to 7.4 with NaOH. Intracellular pipette solution contained (in mM) 140 CsCl, 2 MgCl$_2$, 10 HEPES, 10 EGTA, pH to 7.3 with CsOH. Voltage-ramp recordings ranged from −100 mV to 80 mV and spanned 1.8 s. Data were collected with sampling at 10 kHz, and only cells with whole-cell access resistance that remained less than 15 MΩ across recordings were included in analysis. To verify channel expression, a saturating concentration of glutamate (1 mM) was applied with 100 μM CX614, and only cells with detectable current were included. NBQX and CX614 were acquired from Tocris Biosciences. Sample size was chosen based on previous literature and variance observed in ion channel studies of similar nature.

Array comparative genomic hybridization. Array labeling and hybridization was performed as previously described[76]. Briefly, 250 ng of sample DNA was labeled with Cy3 using a NimbleGen labeling kit (Roche). Reference DNA (NA12878) was labeled in a pooled reaction for four arrays with Cy5 using 1 μg of DNA. Hybridization was performed using the Agilent 2x400K array platform using standard reagents, imaged using an Agilent Scanner, and processed using Agilent Feature Extraction. CNV calls were generated using Agilent CytoGenomics 4.03.12 and the ADM2 calling algorithm with default parameters. For samples passing standard Agilent quality control parameters (DLRSD < 0.2), all CNVs over 100 kbp were visually inspected, filtered for known reference sample artifacts, and compared to those seen among 29,085 cases of intellectual disability or developmental delay and 19,584 controls[77] to identify rare CNVs that may contribute to pathogenicity in these cases.

Missense clustering. Genes with significant clustering of missense mutations were identified by CLUMP[7] (clustering by mutation position; https://github.com/karchinlab/clump/), which applies an unsupervised clustering algorithm based on partitioning around medoid distances between mutations. We implemented the permutation (-z 1000) and minimum mutation options (-m 2) and calculated a P-value based on the null distribution of case and control CLUMP score differences. The case set included individuals with an NDD primary phenotype (ASD, developmental delay, intellectual disability, or epilepsy) from denovo-db[64] v.1.2 (**Supplementary Table 11**) and consisted of 22 studies[1,12–14,24,32,33,43,44,48,49,51,53–55,57–62,78] with 9,997 affected individuals (8,917 de novo missense variants). We compared against two control missense data sets: (i) missense mutations (MAF < 1%) from Europeans[7] ($n = 420$; 196,260 mutations) from the 1000 Genomes Project[79] and (ii) private missense mutations present in individuals from ExAC v.0.3 without neuropsychiatric disorders ($n = 45,376$; 1,466,439 mutations)[18]. All variants were reannotated using the CRAVAT software to enable exact transcript comparisons[80].

Data availability. De novo mutations used for discovery may be obtained from http://denovo-db.gs.washington.edu/denovo-db/. Data from smMIP targeted sequencing is available through the National Database of Autism Research (NDAR) under the project "Sporadic Mutations and Autism Spectrum Disorders" (#2093; https://ndar.nih.gov/edit_collection.html?id=2093). 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/. Other data that support the findings of this study are available from the corresponding author upon reasonable request.

A **Supplementary Methods Checklist** is available.

43. Barcia, G. et al. De novo gain-of-function KCNT1 channel mutations cause malignant migrating partial seizures of infancy. Nat. Genet. **44**, 1255–1259 (2012).
44. Deciphering Developmental Disorders Study. Large-scale discovery of novel genetic causes of developmental disorders. Nature **519**, 223–228 (2015).
45. Dimassi, S. et al. Whole-exome sequencing improves the diagnosis yield in sporadic infantile spasm syndrome. Clin. Genet. **89**, 198–204 (2016).
46. Fromer, M. et al. De novo mutations in schizophrenia implicate synaptic networks. Nature **506**, 179–184 (2014).
47. Gulsuner, S. et al. Spatial and temporal mapping of de novo mutations in schizophrenia to a fetal prefrontal cortical network. Cell **154**, 518–529 (2013).
48. Hashimoto, R. et al. Whole-exome sequencing and neurite outgrowth analysis in autism spectrum disorder. J. Hum. Genet. **61**, 199–206 (2016).
49. Helbig, K.L. et al. Diagnostic exome sequencing provides a molecular diagnosis for a significant proportion of patients with epilepsy. Genet. Med. **18**, 898–905 (2016).
50. Homsy, J. et al. De novo mutations in congenital heart disease with neurodevelopmental and other congenital anomalies. Science **350**, 1262–1266 (2015).
51. Jiang, Y.H. et al. Detection of clinically relevant genetic variants in autism spectrum disorder by whole-genome sequencing. Am. J. Hum. Genet. **93**, 249–263 (2013).
52. Kranz, T.M. et al. De novo mutations from sporadic schizophrenia cases highlight important signaling genes in an independent sample. Schizophr. Res. **166**, 119–124 (2015).
53. Krumm, N. et al. Excess of rare, inherited truncating mutations in autism. Nat. Genet. **47**, 582–588 (2015).
54. Lee, H., Lin, M.C., Kornblum, H.I., Papazian, D.M. & Nelson, S.F. Exome sequencing identifies de novo gain of function missense mutation in KCND2 in identical twins with autism and seizures that slows potassium channel inactivation. Hum. Mol. Genet. **23**, 3481–3489 (2014).
55. Lelieveld, S.H. et al. Meta-analysis of 2,104 trios provides support for 10 new genes for intellectual disability. Nat. Neurosci. **19**, 1194–1196 (2016).
56. McCarthy, S.E. et al. De novo mutations in schizophrenia implicate chromatin remodeling and support a genetic overlap with autism and intellectual disability. Mol. Psychiatry **19**, 652–658 (2014).
57. Michaelson, J.J. et al. Whole-genome sequencing in autism identifies hot spots for de novo germline mutation. Cell **151**, 1431–1442 (2012).
58. O'Roak, B.J. et al. Recurrent de novo mutations implicate novel genes underlying simplex autism risk. Nat. Commun. **5**, 5595 (2014).
59. 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).
60. Veeramah, K.R. et al. De novo pathogenic SCN8A mutation identified by whole-genome sequencing of a family quartet affected by infantile epileptic encephalopathy and SUDEP. Am. J. Hum. Genet. **90**, 502–510 (2012).
61. Veeramah, K.R. et al. Exome sequencing reveals new causal mutations in children with epileptic encephalopathies. Epilepsia **54**, 1270–1281 (2013).
62. Yuen, R.K.C. et al. Whole-genome sequencing of quartet families with autism spectrum disorder. Nat. Med. **21**, 185–191 (2015).
63. Zaidi, S. et al. De novo mutations in histone-modifying genes in congenital heart disease. Nature **498**, 220–223 (2013).
64. Turner, T.N. et al. Denovo-db: a compendium of human de novo variants. Nucleic Acids Res. **45**, D1, D804–D811 (2017).
65. Genome of the Netherlands Consortium. Whole-genome sequence variation, population structure and demographic history of the Dutch population. Nat. Genet. **46**, 818–825 (2014).
66. Constantino, J. Social Responsiveness Scale (SRS-2). West. Psychol. Serv. (2012).
67. Ng, S.B. et al. Targeted capture and massively parallel sequencing of 12 human exomes. Nature **461**, 272–276 (2009).
68. Ezkurdia, I. et al. Multiple evidence strands suggest that there may be as few as 19,000 human protein-coding genes. Hum. Mol. Genet. **23**, 5866–5878 (2014).
69. Pirooznia, M. et al. SynaptomeDB: an ontology-based knowledgebase for synaptic genes. Bioinformatics **28**, 897–899 (2012).
70. Subtil-Rodríguez, A. et al. The chromatin remodeller CHD8 is required for E2F-dependent transcription activation of S-phase genes. Nucleic Acids Res. **42**, 2185–2196 (2014).
71. Darnell, J.C. et al. FMRP stalls ribosomal translocation on mRNAs linked to synaptic function and autism. Cell **146**, 247–261 (2011).
72. 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).
73. Boyle, E.A., O'Roak, B.J., Martin, B.K., Kumar, A. & Shendure, J. MIPgen: optimized modeling and design of molecular inversion probes for targeted resequencing. Bioinformatics **30**, 2670–2672 (2014).
74. Hoischen, A. et al. De novo mutations of SETBP1 cause Schinzel-Giedion syndrome. Nat. Genet. **42**, 483–485 (2010).
75. Li, H. & Durbin, R. Fast and accurate short read alignment with Burrows-Wheeler transform. Bioinformatics **25**, 1754–1760 (2009).
76. Girirajan, S. et al. Refinement and discovery of new hotspots of copy-number variation associated with autism spectrum disorder. Am. J. Hum. Genet. **92**, 221–237 (2013).
77. Coe, B.P. et al. Refining analyses of copy number variation identifies specific genes associated with developmental delay. Nat. Genet. **46**, 1063–1071 (2014).
78. Moreno-Ramos, O.A., Olivares, A.M., Haider, N.B., de Autismo, L.C. & Lattig, M.C. Whole-exome sequencing in a south American cohort links ALDH1A3, FOXN1 and retinoic acid regulation pathways to autism spectrum disorders. PLoS One **10**, e0135927 (2015).
79. Auton, A. et al. A global reference for human genetic variation. Nature **526**, 68–74 (2015).
80. Douville, C. et al. CRAVAT: cancer-related analysis of variants toolkit. Bioinformatics **29**, 647–648 (2013).