HHS Public Access

Author manuscript

Nat Neurosci. Author manuscript; available in PMC 2017 December 19.

Published in final edited form as:

Nat Neurosci. 2017 August ; 20(8): 1043–1051. doi:10.1038/nn.4589.

Hotspots of missense mutation identify novel 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**[21,22,23], **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], and **Evan E. Eichler**[1,25]

[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

Corresponding author: Evan E. Eichler, Ph.D., Department of Genome Sciences, University of Washington School of Medicine, Foege S-413A, Box 355065, 3720 15th Ave NE, Seattle, WA 98195-5065, eee@gs.washington.edu, Tel: 1-206-543-9526, Fax: 1-206-221-5795.

[26]Present address: Department of Pharmacology, Creighton University School of Medicine, Omaha, Nebraska, USA

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., 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.

COMPETING FINANCIAL INTERESTS

E.E.E. is on the scientific advisory board (SAB) of DNAnexus, Inc.

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). Other data that support the findings of this study are available from the corresponding author upon reasonable request.

ACCESSION CODES

Data from smMIP targeted sequencing is available through NDAR under the project "Sporadic Mutations and Autism Spectrum Disorders" (#2093).

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

Abstract

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 17,689 NDD patients identified 21 new patients with identical missense mutations. One recurrent site (p.Ala636Thr) 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.

INTRODUCTION

Multiple lines of evidence provide strong support for 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]. To date, 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 eightfold more 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% vs. 9%, respectively)[1].

The identification of genes with a significant burden of missense mutations, then, is likely to highlight new classes of neurodevelopmental disorder (NDD) risk genes. 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 novel 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 significant phenotypic and genotypic overlap between ASD, developmental delay (DD) and intellectual disability (ID), epilepsy, congenital heart disease, and schizophrenia[10] to study the pattern and distribution of *de novo* missense mutations more broadly. We focused on clustered and recurrent site 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%) is significantly greater than the fraction of controls (47.8%; p = 0.016, OR = 1.12 [1.02–1.24], two-sided 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 is 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 $p = 9.22\times10^{-4}$, OR = 1.12 [1.06–1.19], 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%) harbor mutations in two or more unrelated cases (Supplementary Table 2). In contrast, among controls, 101 out of 1,362 genes (7.4%) are mutated recurrently (Supplementary Table 3). Matching the number of cases and controls, we observe 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], 1×10^6 permutations) and three or more (one-tailed empirical $p = 3.10\times10^{-5}$, OR = 3.13 [2.22–4.03], 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 are significantly shorter than those with recurrence in controls (mRNA, $p = 5.19\times10^{-3}$; protein, $p = 1.47\times10^{-3}$; two-sided Wilcoxon rank-sum tests). Additionally, the total number of genes with mutations is smaller among cases (1,323 in downsampled cases vs. 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 is significantly positively skewed in cases compared to controls consistent with an increase in deleteriousness ($p = 2.2\times10^{-4}$, two-sided Wilcoxon rank-sum test). Further, at increasing minimum CADD score thresholds, the likelihood that an observed event can be attributed to a case increases (Fig. 2a). At a CADD threshold of 28, the likelihood rises dramatically (>1.2 positive likelihood ratio). Importantly, mutations in genes with higher levels of recurrence in cases also show significantly higher CADD scores ($p = 5.87\times10^{-29}$, $F = 45.12$, 3 degrees of freedom, 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 Methods). We identified 35 genes that had significantly more *de novo* missense mutations in cases than expected (false discovery rate (FDR) < 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 mutation has already been established by copy number variants (CNVs) and LGD mutation (e.g., *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 (e.g., *PACS1* and Schuurs-Hoeijmakers syndrome[17]), the remaining seven warrant additional follow-up.

As a set, the 35 genes with excess *de novo* missense mutations are enriched for aspects of neuronal communication such as postsynaptic membrane potential regulation (6 observed vs. 0.17 expected, 35.3-fold enrichment, p_{adj} (Bonferroni corrected) $= 1.61\times10^{-4}$, two-sided binomial test) and synaptic signaling (8 observed vs. 0.7 expected, 11.4-fold enrichment, $p_{adj} = 3.30\times10^{-3}$, two-sided binomial test), nervous system development (16 observed vs. 3.7 expected, 4.4-fold enrichment, $p_{adj} = 1.05\times10^{-3}$, two-sided binomial test), and gene expression regulation (3 observed vs. 0.03 expected, >100-fold enrichment, $p_{adj} = 2.42\times10^{-2}$, two-sided binomial test). There is also significant enrichment for genes involved in the presynapse (5/336 genes; $p = 3.74\times10^{-4}$, OR = 9.25 [3.40-Inf], one-sided Fisher's exact test) and postsynaptic density (11/1,755 genes; $p = 2.17\times10^{-4}$, OR = 4.50 [2.26-Inf], one-sided Fisher's exact test), and targets of FMRP (14/842 genes; $p = 1.20\times10^{-10}$, OR = 14.37 [7.57-Inf], one-sided 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 (e.g., *PACS1* with six mutations at residue 203) and some genes had more than one recurrently mutated amino acid residue (e.g., *SCN2A*). Sixteen of the amino acid sites involved adjacent mutations in the same codon. Twenty-eight of the 40 sites (36/56 mutations) involve CpG dinucleotides, consistent with their known association with hotspots of single-nucleotide variation[19]. Thirty-four sites had average 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 17,689 patients with a primary diagnosis of ASD or DD (including ID; 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 as an additional control (n = 3,023). We identified and validated 21 recurrent missense variants at 12 sites in 11 genes among cases (Table 1, Fig. 3a–c). 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 determined due to missing parental DNA—six were determined to be *de novo* missense mutations (Table 1; PACS1 p.Arg203 (two mutations), GRIA1 p.Ala636, SCN2A p.Arg937, and SMAD4 p.Ile500 (two mutations)). Interestingly, 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 ID. 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 ALG13 (Fig. 3b), where all six individuals with a mutation at residue 107 have both EPI and DD 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 ID, 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 in *SMAD4* (Fig. 3c). We reexamined our database for regions where multiple recurrent *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 seven variants in five unaffected siblings, representing a significant enrichment ($p = 1.11\times10^{-4}$, OR = 3.93 [1.76–10.89], two-sided Fisher's exact test) (Table 2 and Supplementary Table 7). Twelve of the clustered missense mutations in cases were confirmed *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 cases. While Myhre syndrome has been associated only with residue 500 of SMAD4[22], *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 ID, short stature, and dysmorphic facial features, suggesting that Myhre syndrome is not only limited to one codon[22]. Phenotypic commonalities are also present amongst 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 DD.

De novo missense mutations in *GRIA1*

We identified a recurrently mutated amino acid in *GRIA1* (a.k.a. GluA1; Fig. 4a), a subunit of AMPA glutamate receptors, which was originally reported in one patient with ID[24] and another with ASD[14]. Both patients share an identical *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 found mutation was confirmed as *de novo*; paternal DNA is 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 published by ExAC[18]. Moreover, we identified a second *de novo* missense mutation in close proximity (Fig. 4a) in a patient with DD. 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* ($p = 5.39\times10^{-3}$, one-tailed binomial test, genome-wide correction)) suggested that the mutation was pathogenic.

The mutated site maps to the eighth position (p.Ala636) of a highly conserved 9-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 six 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 base-pair 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 −100mV to +80mV. 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, but not WT-expressing cells (Fig. 4d,f). Consistent with GluA1-mediated current, inward rectification is 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 p.Ala636Thr mutation are heterozygous indicating that a majority of receptors are likely comprised of WT and p.Ala636Thr receptor subunits. To assess the functional phenotype of these ‘heteromeric’ receptors, we co-transfected 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 the homomeric p.Ala636Thr 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 demonstrates common features (Supplementary Table 9), including mild to moderate ID (4 of 4 individuals) and ASD (3 of 4 individuals). 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 is noted to have behavioral dysfunction. Two individuals also had delayed motor development. All four have 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 or 1KG (see 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 is significantly associated with aspects of neuronal communication, including regulation of the postsynaptic potential (11 observed vs. 1 expected, 11.0-fold enrichment, $p_{adj} = 6.93\times10^{-5}$, two-sided binomial test) and synaptic signaling (20 observed vs. 4.15 expected, 4.82-fold enrichment, $p_{adj} = 8.38\times10^{-5}$, two-sided binomial test), as well as chromatin-mediated maintenance of transcription (4 observed vs. 0.1 expected, 40.7-fold enrichment, $p_{adj} = 2.96\times10^{-2}$, two-sided binomial test). Many of the genes encode channel proteins and receptors (e.g., *GRIA1, GRIN1, GRIN2A, GRIN2B, KCNH1, KCNQ2*) and exhibit clustering in or near specific functional domains, such as the transmembrane, pore or voltage sensor domains (Fig. 5a–d). Other proteins, such as CTCF, are remarkable in that the clustering pattern of patient missense mutations highlights a subset of the C2H2 ZNF motifs, which are 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, DD, and ID but also patients with epilepsy and schizophrenia due to the extensive comorbidity of these diagnoses. As expected, both recurrence and severity of missense mutations are critical features. The likelihood of a pathogenic mutation rises significantly when three or more missense mutations are observed in a gene ($p = 1.06\times10^{-18}$, two-sided Wilcoxon rank-sum test) and, in particular, when the severity of the missense mutation exceeds a CADD score of 28 (>1.2 positive likelihood ratio). We use these features to identify 35 genes with an excess ($q < 0.05$) of *de novo* missense mutations (Supplementary Table 2). Targeted sequencing of specific protein-coding regions shows that recurrent and clustered amino acid replacements are more common in cases than controls ($p = 1.11\times10^{-4}$, OR = 3.93 [1.76–10.89], two-sided Fisher's exact test). While many of the top-scoring genes are associated with known syndromic and non-syndromic forms of NDD (e.g., *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 identify 200 genes with patterns of *de novo* missense mutations that are more clustered in cases when compared to 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 (e.g., *COL4A3BP, PPP2R5D*) suggesting that LGD events are either not tolerated or 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 DD[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 (e.g., cause embryonic lethality or infertility), although additional 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 (e.g., *CTCF*), transmembrane domains (e.g., *GRIN1*), and voltage sensors and channel pores (e.g., *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 3D protein structure analysis reveals that these three clusters define the cleft of the ligand binding site[34] (Supplementary Fig. 1).

It is interesting that genes associated with hotspots of missense mutation (Supplementary Table 10) are 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 find 35-fold and 11-fold enrichments of genes regulating postsynaptic membrane potential in genes that carry 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 p.Ala636Thr mutation strongly supports a role for *GRIA1* in ASD and related NDDs. This specific *de novo* missense mutation has been observed before at the homologous position in a highly conserved motif in a different glutamate receptor, GluRδ2[26]. 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

plays an important role in learning and memory[31], there is a biologically plausible link between this *de novo* missense mutation in *GRIA1* and ID.

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 prior to 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 ID associated with this mutation. It is interesting to note that 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 p.Ala636Thr mutation in synapse development and function will shed additional light on how alterations in excitatory synaptic function contribute to ASD.

ONLINE METHODS

Exome datasets 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, DD, ID, epilepsy (EPI), schizophrenia (SCZ), and congenital heart disease (CHD) as well as four cohorts of unaffected controls[1,47,65] (n = 2,178) (Supplementary Table 1). Only CHD patients from Homsy et al. (2015) 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 were either 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-sided Fisher's exact tests. For comparisons of mutation rate and recurrence that depended on identical numbers of cases and controls, we performed one 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 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 paradigms based on the analysis type. For significance calculations of whole genes, we utilized the Benjamini-Hochberg FDR 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 (FWER) correction based on the number of amino acids in the genome ($n = 1.1\times10^7$) 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-sided Fisher's exact test for testing the enrichment of specific gene sets, including neuronal compartments such as the post-synaptic density[69] and targets of CHD8[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 6,058 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 an additional four 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 totals to 17,689 cases and 3,023 controls. The study size was not predetermined but based on the maximal 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 DD cohorts. All patients were consented for sequencing and recontacting for inheritance testing at the providing laboratory. Patient samples were acquired from Adelaide (Jozef Gecz, University of Adelaide), Antwerp (Frank Kooy, University of Antwerp), Autism Clinical and Genetic Resources in China (ACGC; Kun Xia), the Autism Genetic Resource Exchange (AGRE), Iowa (Jacob Michaelson, University of Iowa), Leiden (Gijs Santen, Leiden University Medical Center), Leuven (Hilde Peeters, University Hospitals Leuven), Philadelphia (Hakon Hakonarson, Children's Hospital of Philadelphia), Prague (Zdenek Sedlacek, University Hospital Motol), San Diego (Eric Courchesne, UC San Diego), Simons Simplex Collection (SSC), Stockholm (Magnus Nordenskjold, Karolinska University Hospital), the Study of Autism Genetics Exploration (SAGE; Raphe Bernier, University of Washington), The Autism Simplex Collection (TASC), and Troina (Corrado Romano, 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 SF (ATCC ACS-4500) cells, 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. 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 hours later, approximately 0.6 μg of total DNA/coverslip was transfected using the Fugene6 reagent (2 μl/coverslip, Promega). For heteromeric cells, approximately 0.3 μg of WT and A636T were co-transfected. Whole-cell recordings were performed approximately 60 hours 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 −100mV to 80 mV and spanned 1.8 seconds. 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 of 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 ug of DNA. Hybridization was performed using the Agilent 2×400K 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 QC parameters (DLRSD < 0.2), all CNVs over 100 kbp were visually inspected, filtered for known reference sample artifacts, and compared to 29,085 cases of ID/DD 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, DD, ID, or EPI) 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 datasets: 1) missense mutations (MAF < 1%) from Europeans[7] (n = 420; 196,260 mutations) from the 1000 Genomes Project[79] and 2) private missense mutations present in individuals from the ExAC v.0.3 without neuropsychiatric disorders (n = 45,376; 1,466,439 mutations)[18]. All variants were re-annotated using the CRAVAT software to enable exact transcript comparisons[80].

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgments

We thank the individuals and their families for participation in this study. This research was supported, in part, by the following: Simons Foundation Autism Research Initiative (SFARI 303241) to E.E.E., National Institutes of Health (R01MH101221 to E.E.E., R01MH100047 to R.A.B., R01MH104450 to L.S.Z., RO1MH105527 and R01DC014489 to J.J.M.), the NHGRI Interdisciplinary Training in Genome Science Grant (T32HG00035) to H.A.F.S. and T.N.T., postdoctoral fellowship grant from the Autism Science Foundation (16-008) to 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 31671114) to K.X. and H.G., the China Scholarship Council (201406370028) and the Fundamental Research Funds for the Central Universities (2012zzts110) to T.W., 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, the Stockholm County Council, grants (KL2TR00099 and 1KL2TR001444) from the University of California, San Diego Clinical and Translational Research Institute to T.P., the Research Fund - Flanders (FWO) to R.F.K. and G.V., and 00064203 and NF-CZ11-PDP-3-003-2014 to Z.S. We are grateful to all of the families at the participating Simons Simplex Collection (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, E. Wijsman). We appreciate obtaining access to phenotypic data on SFARI Base. Approved researchers can obtain the SSC population dataset described in this study (http://sfari.org/resources/simons-simplex-collection) by applying at https://base.sfari.org. 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 Clara M. Lajonchere (PI). We thank J. Gerdts, S. Trinh, and B. McKenna for

their contributions and T. Brown for assistance in editing this manuscript. H.P. is a Senior Clinical Investigator of The Research Foundation-Flanders (FWO). E.E.E. is an investigator of the Howard Hughes Medical Institute.

References

1. Iossifov I, et al. The contribution of de novo coding mutations to autism spectrum disorder. Nature. 2014; 515:216–221. [PubMed: 25363768]
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. 2014; 15:133–41. [PubMed: 24430941]
3. Bernier R, et al. Disruptive CHD8 mutations define a subtype of autism early in development. Cell. 2014; 158:263–276. [PubMed: 24998929]
4. Stessman HA, Bernier R, Eichler EE. A genotype-first approach to defining the subtypes of a complex disease. Cell. 2014; 156:872–877. [PubMed: 24581488]
5. Sanders SJ, et al. Insights into Autism Spectrum Disorder Genomic Architecture and Biology from 71 Risk Loci. Neuron. 2015; 87:1215–1233. [PubMed: 26402605]
6. Packer A. Neocortical neurogenesis and the etiology of autism spectrum disorder. Neurosci. Biobehav. Rev. 2016; 64:185–195. [PubMed: 26949225]
7. Turner TN, et al. Proteins linked to autosomal dominant and autosomal recessive disorders harbor characteristic rare missense mutation distribution patterns. Hum. Mol. Genet. 2015; 24:5995–6002. [PubMed: 26246501]
8. van Bon BWM, et al. Disruptive de novo mutations of DYRK1A lead to a syndromic form of autism and ID. Mol. Psychiatry. 2015:1–7. [PubMed: 25648202]
9. Helsmoortel C, et al. A SWI/SNF-related autism syndrome caused by de novo mutations in ADNP. Nat. Genet. 2014; 46:380–4. [PubMed: 24531329]
10. Buxbaum JD. DSM-5 and psychiatric genetics - Round hole, meet square peg. Biol. Psychiatry. 2015; 77:766–768. [PubMed: 25843333]
11. Kircher M, et al. A general framework for estimating the relative pathogenicity of human genetic variants. Nat. Genet. 2014; 46:310–315. [PubMed: 24487276]
12. Turner TN, et al. Genome Sequencing of Autism-Affected Families Reveals Disruption of Putative Noncoding Regulatory DNA. Am. J. Hum. Genet. 2015; 98:58–74. [PubMed: 26749308]
13. O'Roak BJ, et al. Multiplex targeted sequencing identifies recurrently mutated genes in autism spectrum disorders. Science. 2012; 338:1619–22. [PubMed: 23160955]
14. De Rubeis S, et al. Synaptic, transcriptional and chromatin genes disrupted in autism. Nature. 2014; 515:209–215. [PubMed: 25363760]
15. McRae JF, et al. Prevalence and architecture of de novo mutations in developmental disorders. Nature. 2017
16. Stessman HAF, Turner TN, Eichler EE. Molecular subtyping and improved treatment of neurodevelopmental disease. Genome Med. 2016; 8:22. [PubMed: 26917491]
17. Schuurs-Hoeijmakers JHM, 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. 2012; 91:1122–1127. [PubMed: 23159249]
18. Lek M, et al. Analysis of protein-coding genetic variation in 60,706 humans. Nature. 2016; 536:285–291. [PubMed: 27535533]
19. Lynch M. Rate, molecular spectrum, and consequences of human mutation. Proc. Natl. Acad. Sci. U. S. A. 2010; 107:961–8. [PubMed: 20080596]
20. Maheshwari M, et al. PTPN11 mutations in noonan syndrome type I: Detection of recurrent mutations in exons 3 and 13. Hum. Mutat. 2002; 20:298–304. [PubMed: 12325025]
21. Myhre SA, Ruvalcaba RHA, Graham CB. A new growth deficiency syndrome. Clin. Genet. 2008; 20:1–5.
22. Le Goff C, et al. Mutations at a single codon in Mad homology 2 domain of SMAD4 cause Myhre syndrome. Nat. Genet. 2012; 44:85–8.
23. Landrum MJ, et al. ClinVar: Public archive of interpretations of clinically relevant variants. Nucleic Acids Res. 2016; 44:D862–D868. [PubMed: 26582918]

24. de Ligt J, et al. Diagnostic Exome Sequencing in Persons with Severe Intellectual Disability. N. Engl. J. Med. 2012; 367:1921–1929. [PubMed: 23033978]
25. Yuan H, Erreger K, Dravid SM, Traynelis SF. Conserved structural and functional control of N-methyl-D-aspartate receptor gating by transmembrane domain M3. J. Biol. Chem. 2005; 280:29708–29716. [PubMed: 15970596]
26. Zuo J, et al. Neurodegeneration in Lurcher mice caused by mutation in delta2 glutamate receptor gene. Nature. 1997; 388:769–773. [PubMed: 9285588]
27. Coutelier M, et al. GRID2 mutations span from congenital to mild adult-onset cerebellar ataxia. Neurology. 2015; 84:1751–1759. [PubMed: 25841024]
28. Kohda K, Wang Y, Yuzaki M. Mutation of a glutamate receptor motif reveals its role in gating and delta2 receptor channel properties. Nat. Neurosci. 2000; 3:315–322. [PubMed: 10725919]
29. Taverna F, et al. The Lurcher mutation of an alpha-amino-3-hydroxy-5-methyl-4-isoxazolepropionic acid receptor subunit enhances potency of glutamate and converts an antagonist to an agonist. J. Biol. Chem. 2000; 275:8475–8479. [PubMed: 10722683]
30. Klein RM, Howe JR. Effects of the lurcher mutation on GluR1 desensitization and activation kinetics. J. Neurosci. 2004; 24:4941–4951. [PubMed: 15163686]
31. Kessels HW, Malinow R. Synaptic AMPA Receptor Plasticity and Behavior. Neuron. 2009; 61:340–350. [PubMed: 19217372]
32. Tavassoli T, et al. De novo SCN2A splice site mutation in a boy with Autism spectrum disorder. BMC Med. Genet. 2014; 15:35. [PubMed: 24650168]
33. Allen AS, et al. De novo mutations in epileptic encephalopathies. Nature. 2013; 501:217–21. [PubMed: 23934111]
34. Niknafs N, et al. MuPIT interactive: Webserver for mapping variant positions to annotated, interactive 3D structures. Hum. Genet. 2013; 132:1235–1243. [PubMed: 23793516]
35. Banerjee S, Riordan M, Bhat MA. Genetic aspects of autism spectrum disorders: insights from animal models. Front. Cell. Neurosci. 2014; 8:58. [PubMed: 24605088]
36. Durand CM, et al. Mutations in the gene encoding the synaptic scaffolding protein SHANK3 are associated with autism spectrum disorders. Nat Genet. 2007; 39:25–27. [PubMed: 17173049]
37. Hamdan FF, et al. De novo syngap1 mutations in nonsyndromic intellectual disability and autism. Biol. Psychiatry. 2011; 69:898–901. [PubMed: 21237447]
38. Peñagarikano O, et al. Absence of CNTNAP2 leads to epilepsy, neuronal migration abnormalities, and core autism-related deficits. Cell. 2011; 147:235–246. [PubMed: 21962519]
39. Varoqueaux F, et al. Neuroligins Determine Synapse Maturation and Function. Neuron. 2006; 51:741–754. [PubMed: 16982420]
40. Zanjani HS, et al. Death and survival of heterozygous lurcher purkinje cells in vitro. Dev. Neurobiol. 2009; 69:505–517. [PubMed: 19294643]
41. Andrásfalvy BK, Smith Ma, Borchardt T, Sprengel R, Magee JC. Impaired regulation of synaptic strength in hippocampal neurons from GluR1-deficient mice. J. Physiol. 2003; 552:35–45. [PubMed: 12878757]
42. Wiedholz LM, et al. Mice lacking the AMPA GluR1 receptor exhibit striatal hyperdopaminergia and ‘schizophrenia-related’ behaviors. Mol. Psychiatry. 2008; 13:631–640. [PubMed: 17684498]
43. Barcia G, et al. De novo gain-of-function KCNT1 channel mutations cause malignant migrating partial seizures of infancy. Nat. Genet. 2012; 44:1255–9. [PubMed: 23086397]
44. Deciphering Developmental Disorders Study. Large-scale discovery of novel genetic causes of developmental disorders. Nature. 2015; 519:223–8. [PubMed: 25533962]
45. Dimassi S, et al. Whole-exome sequencing improves the diagnosis yield in sporadic infantile spasm syndrome. Clin. Genet. 2015:198–204.
46. Fromer M, et al. De novo mutations in schizophrenia implicate synaptic networks. Nature. 2014; 506:179–184. [PubMed: 24463507]
47. Gulsuner S, et al. Spatial and temporal mapping of de novo mutations in schizophrenia to a fetal prefrontal cortical network. Cell. 2013; 154:518–529. [PubMed: 23911319]
48. Hashimoto R, et al. Whole-exome sequencing and neurite outgrowth analysis in autism spectrum disorder. J. Hum. Genet. 2015; 61:1–8.

49. Helbig KL, et al. Diagnostic exome sequencing provides a molecular diagnosis for a significant proportion of patients with epilepsy. Genet. Med. 2016:1–8.
50. Homsy J, et al. De novo mutations in congenital heart disease with neurodevelopmental and other congenital anomalies. Science (80-.). 2015; 350:1262–1266.
51. Jiang Y, et al. Detection of clinically relevant genetic variants in autism spectrum disorder by whole-genome sequencing. Am. J. Hum. Genet. 2013; 93:249–63. [PubMed: 23849776]
52. Kranz TM, et al. De novo mutations from sporadic schizophrenia cases highlight important signaling genes in an independent sample. Schizophr. Res. 2015; 166:119–124. [PubMed: 26091878]
53. Krumm N, et al. Excess of rare, inherited truncating mutations in autism. Nat. Genet. 2015; 47:582–588. [PubMed: 25961944]
54. Lee H, Lin M, chin A, Kornblum HI, Papazian DM, Nelson SF. 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. 2014; 23:3481–3489. [PubMed: 24501278]
55. Lelieveld SH, et al. Meta-analysis of 2,104 trios provides support for 10 novel candidate genes for intellectual disability. Nat. Neurosci. 2016; 19:1194–1196. [PubMed: 27479843]
56. McCarthy SE, et al. De novo mutations in schizophrenia implicate chromatin remodeling and support a genetic overlap with autism and intellectual disability. Mol. Psychiatry. 2014; 19:652–8. [PubMed: 24776741]
57. Michaelson JJ, et al. Whole-genome sequencing in autism identifies hot spots for de novo germline mutation. Cell. 2012; 151:1431–1442. [PubMed: 23260136]
58. O'Roak BJ, et al. Recurrent de novo mutations implicate novel genes underlying simplex autism risk. Nat. Commun. 2014; 5:5595. [PubMed: 25418537]
59. Rauch A, et al. Range of genetic mutations associated with severe non-syndromic sporadic intellectual disability: an exome sequencing study. Lancet. 2012; 380:1674–1682. [PubMed: 23020937]
60. Veeramah KR, 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. 2012; 90:502–510. [PubMed: 22365152]
61. Veeramah KR, et al. Exome sequencing reveals new causal mutations in children with epileptic encephalopathies. Epilepsia. 2013; 54:1270–1281. [PubMed: 23647072]
62. Yuen RKC, et al. Whole-genome sequencing of quartet families with autism spectrum disorder. Nat. Med. 2015; 21:185–91. [PubMed: 25621899]
63. Zaidi S, et al. De novo mutations in histone-modifying genes in congenital heart disease. Nature. 2013; 498:220–3. [PubMed: 23665959]
64. Turner TN, et al. Denovo-Db: a Compendium of Human *De Novo* Variants. Nucleic Acids Res. 2016:gkw865.
65. The Genome of the Netherlands Consortium. Whole-genome sequence variation, population structure and demographic history of the Dutch population. Nat. Genet. 2014; 46:818–825. [PubMed: 24974849]
66. Constantino J. Social Responsiveness Scale (SRS-2). West. Psychol. Serv. 2012
67. Ng SB, et al. Targeted Capture and Massicely Parallel Sequencing of twelve human exomes. Nature. 2010; 461:272–276.
68. Ezkurdia I, et al. Multiple evidence strands suggest that theremay be as few as 19 000 human protein-coding genes. Hum. Mol. Genet. 2014; 23:5866–5878. [PubMed: 24939910]
69. Pirooznia M, et al. SynaptomeDB: An ontology-based knowledgebase for synaptic genes. Bioinformatics. 2012; 28:897–899. [PubMed: 22285564]
70. Subtil-Rodríguez A, et al. The chromatin remodeller CHD8 is required for E2F-dependent transcription activation of S-phase genes. Nucleic Acids Res. 2014; 42:2185–2196. [PubMed: 24265227]
71. Darnell JC, et al. FMRP stalls ribosomal translocation on mRNAs linked to synaptic function and autism. Cell. 2011; 146:247–261. [PubMed: 21784246]

72. Hiatt JB, Pritchard CC, Salipante SJ, O'Roak BJ, Shendure J. Single molecule molecular inversion probes for targeted, high-accuracy detection of low-frequency variation. Genome Res. 2013; 23:843–854. [PubMed: 23382536]

73. Boyle, Ea, O'Roak, BJ., Martin, BK., Kumar, A., Shendure, J. MIPgen: optimized modeling and design of molecular inversion probes for targeted resequencing. Bioinformatics. 2014; 30:2670–2. [PubMed: 24867941]

74. Hoischen A, et al. De novo mutations of SETBP1 cause Schinzel-Giedion syndrome. Nat. Genet. 2010; 42:483–485. [PubMed: 20436468]

75. Li H, Durbin R. Fast and accurate short read alignment with Burrows-Wheeler transform. Bioinformatics. 2009; 25:1754–1760. [PubMed: 19451168]

76. Girirajan S, et al. Refinement and discovery of new hotspots of copy-number variation associated with autism spectrum disorder. Am. J. Hum. Genet. 2013; 92:221–237. [PubMed: 23375656]

77. Coe BP, et al. Refining analyses of copy number variation identifies specific genes associated with developmental delay. Nat. Genet. 2014; 46

78. Moreno-Ramos OA, Olivares AM, Haider NB, De Autismo LC, Lattig MC. Whole-exome sequencing in a south American cohort links ALDH1A3, FOXN1 and retinoic acid regulation pathways to autism spectrum disorders. PLoS One. 2015; 10:1–13.

79. Auton A, et al. A global reference for human genetic variation. Nature. 2015; 526:68–74. [PubMed: 26432245]

80. Douville C, et al. CRAVAT: Cancer-related analysis of variants toolkit. Bioinformatics. 2013; 29:647–648. [PubMed: 23325621]

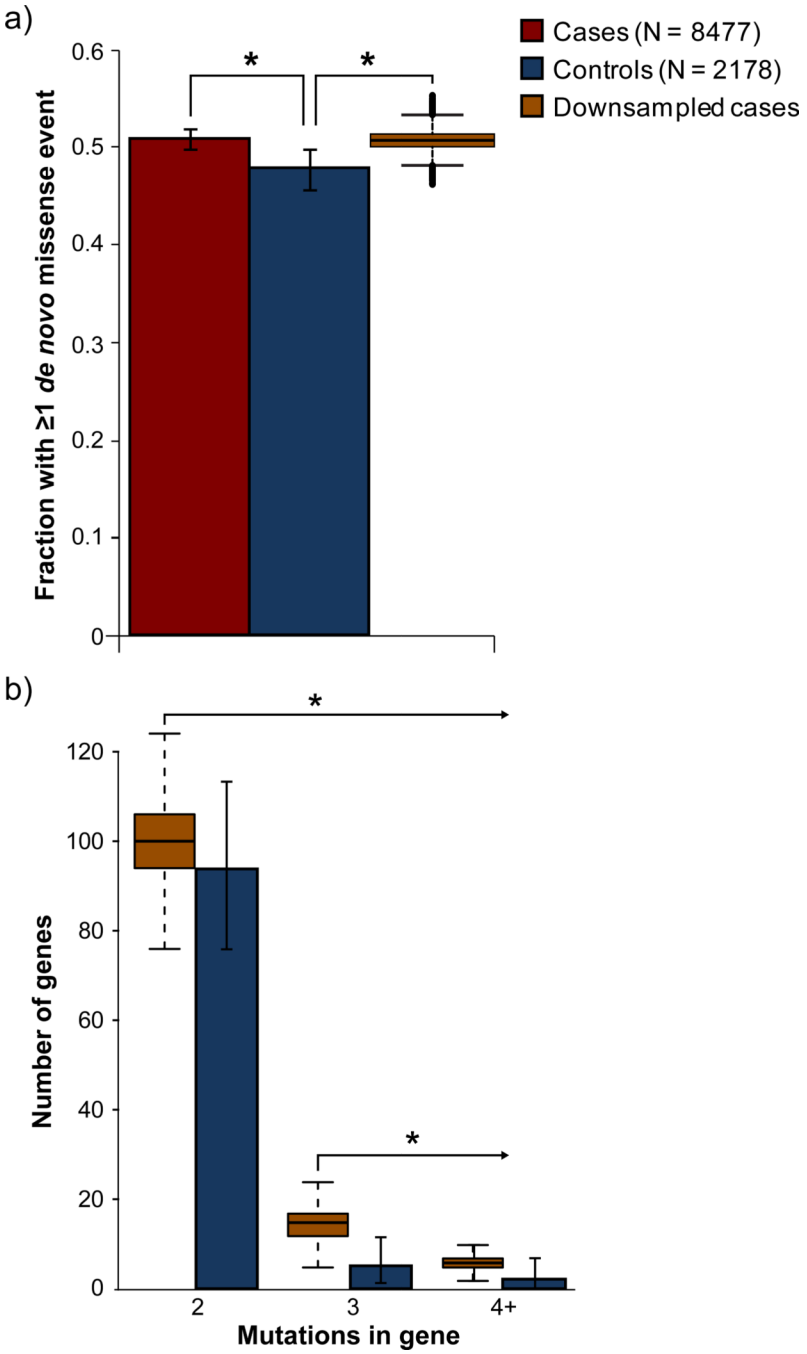

Figure 1. Burden and recurrence of *de novo* missense mutations

Bars for cases and controls represent observed data and error bars indicate the 95% confidence interval (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. Asterisks indicate $p < 0.05$. **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, two-sided Fisher's exact test) even after downsampling (empirical $p = 9.22\times10^{-4}$, OR = 1.12 [1.06–1.19]). **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]), as is the number of genes with three or more mutations (empirical $p = 3.1\times10^{-5}$, OR = 3.13 [2.22–4.03]).

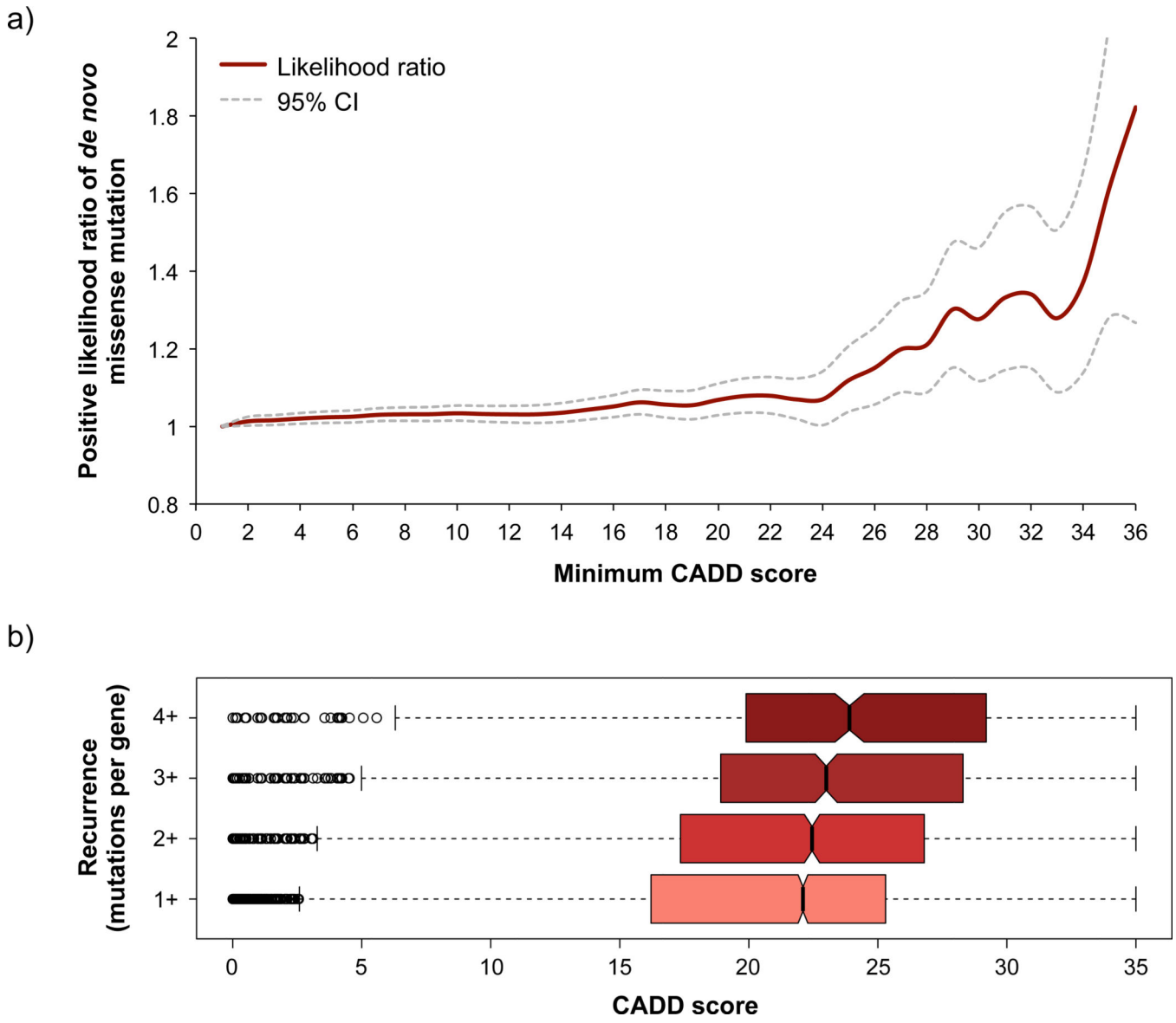

Figure 2. Severity of *de novo* missense mutations

a) *De novo* missense mutations are more likely to be deleterious in cases (n = 5,807 mutations) versus controls (n = 1,475 mutations) and the positive likelihood ratio increases as severity increases (as measured by Combined Annotation Dependent Depletion (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.

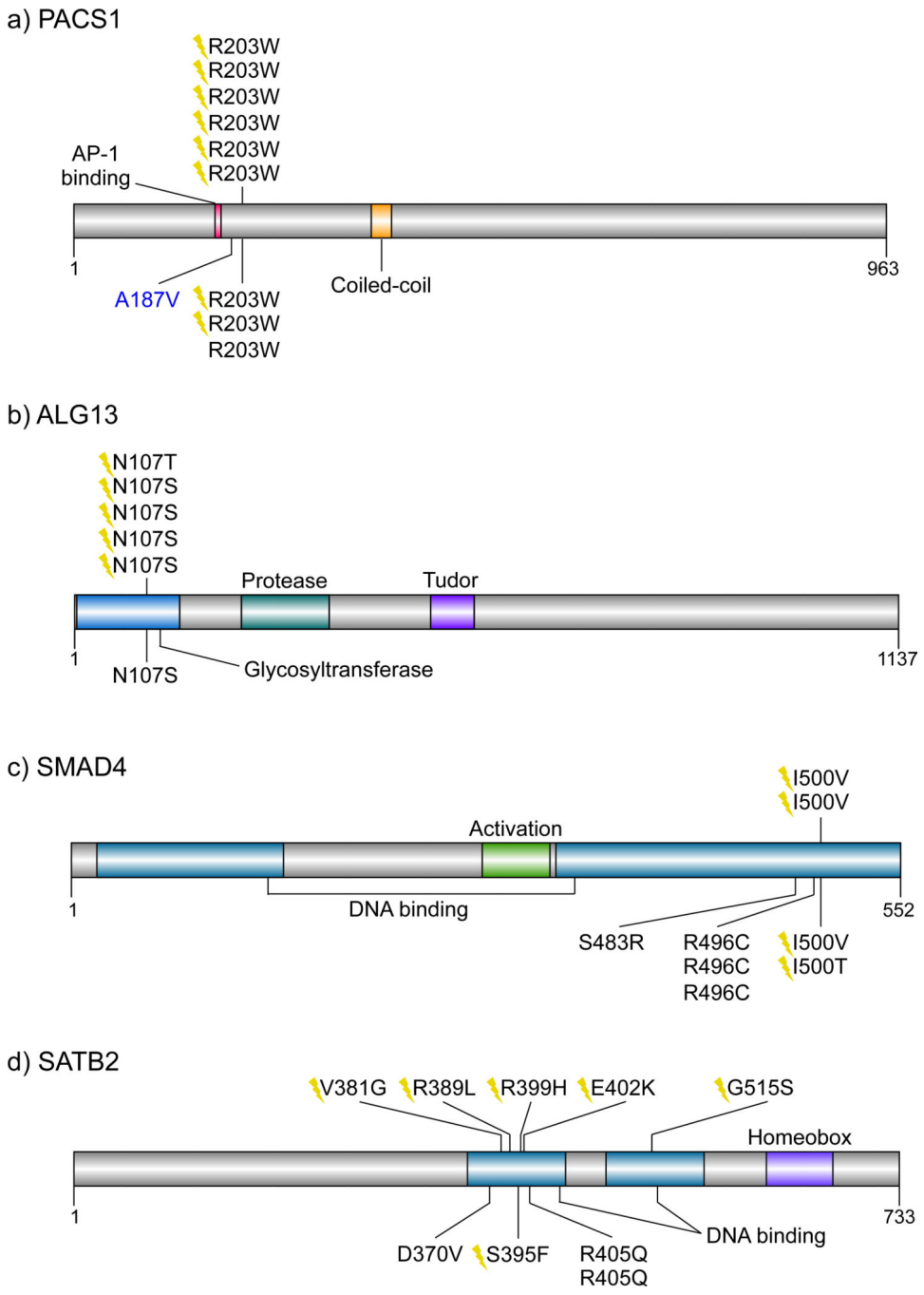

Figure 3. Recurrent mutations fall in or near functional domains

Published mutations in NDD patients are above the protein and new mutations identified by targeted sequencing are below the protein. *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.

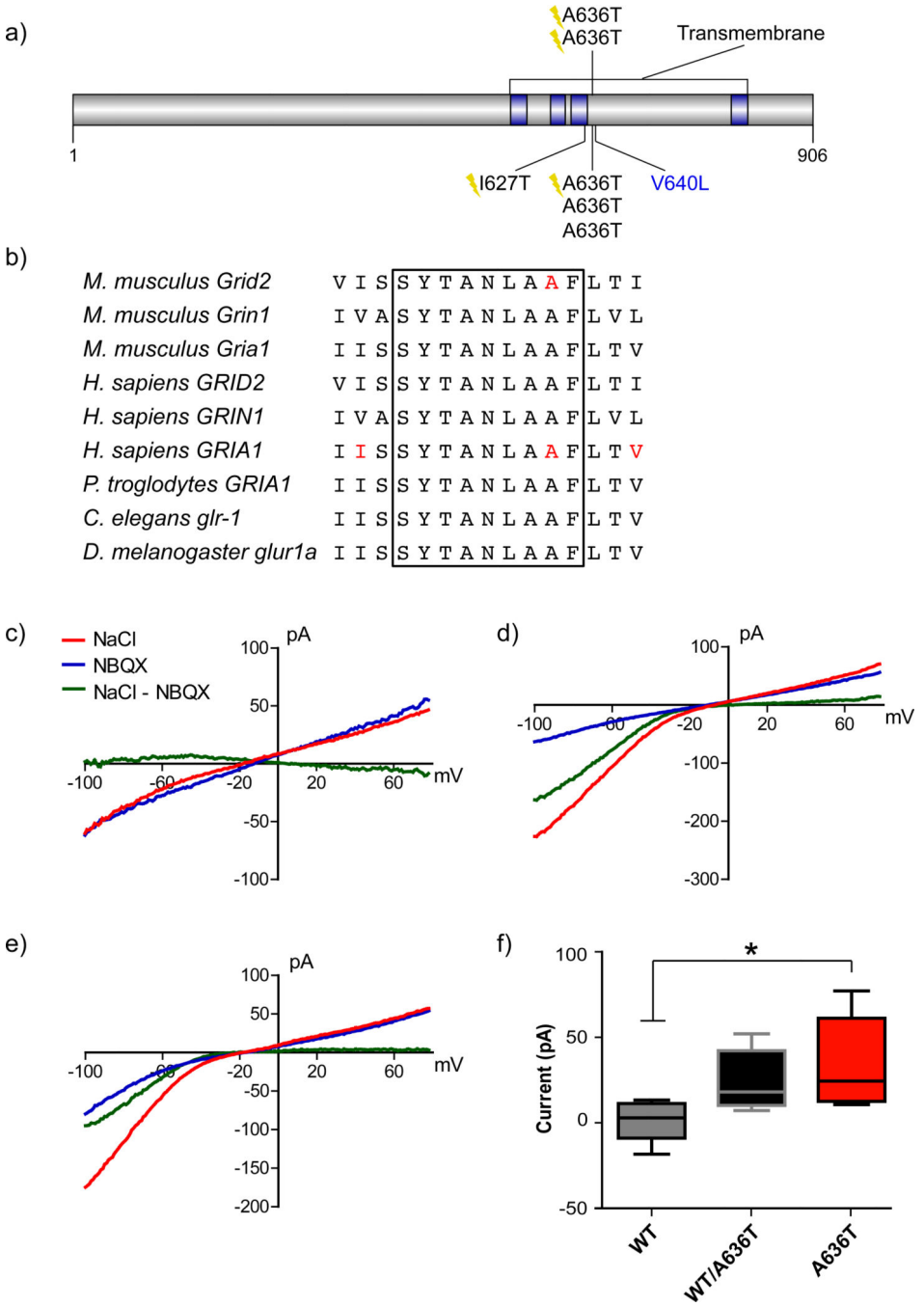

Figure 4. Functional effect of recurrent GRIA1 missense mutations

a) Linear representation of annotated domains in the protein GRIA1, a.k.a. GluA1 (NP_001244950.1) as defined in UniProt. **b)** A recurrent mutation observed only in NDD cases (n = 6 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 delta-2 subunit of this receptor (GRID2). **c–e)** Example current traces from a 1.8s voltage-ramp from −100mV to +80mV 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 uM 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 −60mV. The GluA1-mediated current (NaCl – NBQX) was determined at −60mV and averaged across cells (n = 5 (WT), 7 (A636T), and 5 (heteromeric); p = 0.024, F = 4.91, 2 degrees of freedom, one-way ANOVA). Data are mean +/− S.E.M.

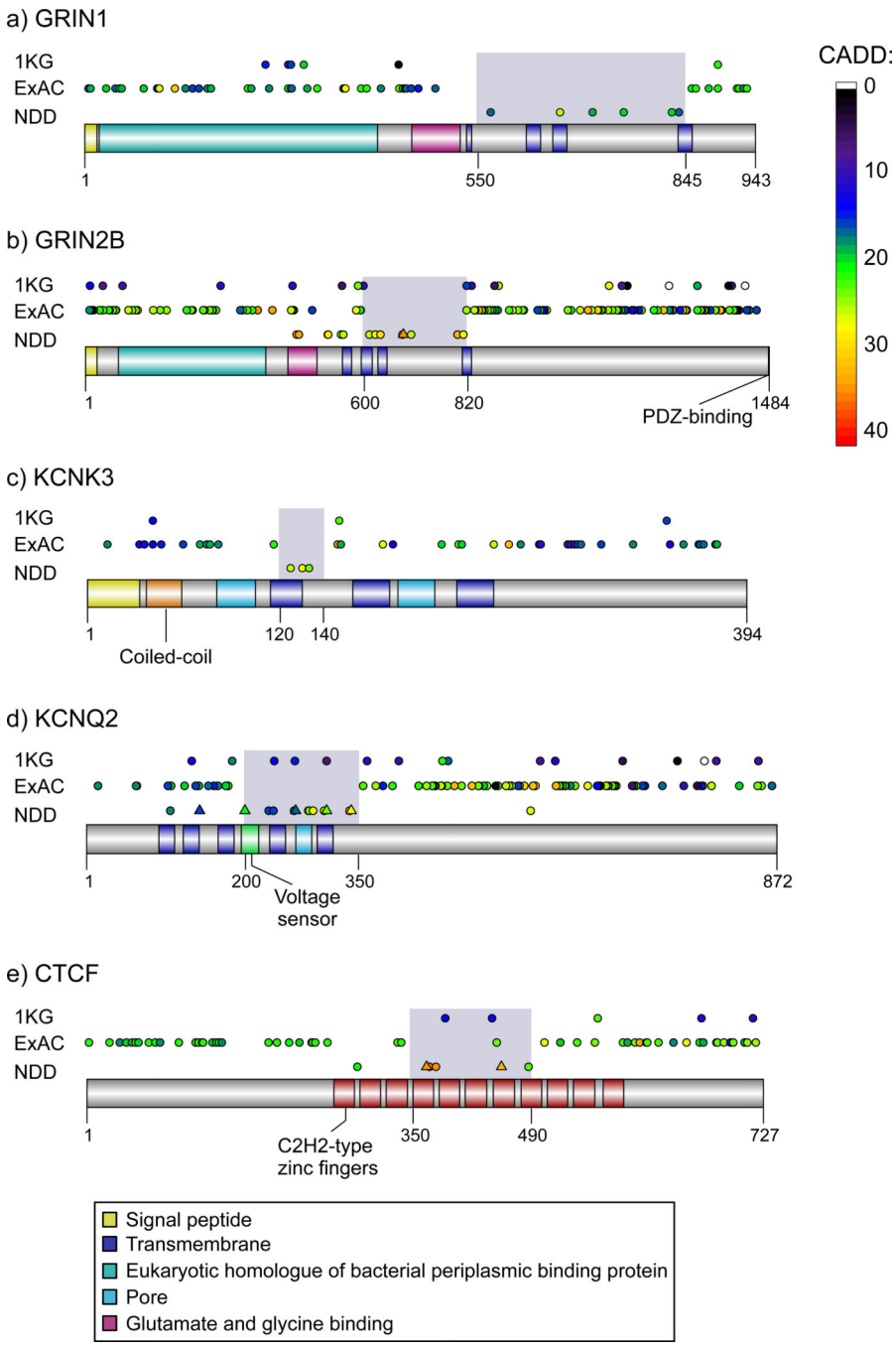

Figure 5. Proteins with excessive clustering of missense mutations in NDD cases

The pattern of *de novo* missense mutations in cases with NDDs is contrasted with rare missense variants from the 1000 Genomes Project (1KG) and private missense mutations from the Exome Aggregation Consortium (ExAC) excluding neuropsychiatric cases. Missense mutations are colored by severity (CADD heatmap) 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 $p = 5.6\times10^{-8}$). **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; $p = 2.0\times10^{-9}$). **c)** KCNK3 (NP_002237.1) patient missense mutations cluster (CLUMP = 0.54, p = 0.036) near the first transmembrane domain (amino acids 120–140, $p = 9.4\times10^{-5}$, OR = Inf). The average per-base rate of ExAC samples with ≥10× coverage across the exon harboring mutations in cases was 79.1%. **d)** KCNQ2 (NP_742105.1) shows several missense mutation hotspots (CLUMP = 0.36, $p < 1\times10^{-3}$) corresponding to the pore and voltage sensor of the channel (amino acids 200–350, $p = 2.0\times10^{-14}$). **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, $p = 9.1\times10^{-8}$).

Nat Neurosci. Author manuscript; available in PMC 2017 December 19.

Table 1

New recurrent mutations at targeted missense sites.

Gene	Site	Alternate amino acid(s)	Protein ID	Mutations in denovo-db v.0.9 (N = 8477)	Codon *de novo* p	Codon *de novo* p, genome-wide correction*	Mutations identified with smMIPs (N = 17850)			Total *de novo*	Codon *de novo* p	Codon *de novo* p, genome-wide correction*	ExAC v. 0.3 allele count (N = 45376)
							Inherited	*De novo*	Unknown				
PACS1	p.Arg203	Trp	NP_060496.2	6	1.03E-24	1.13E-17		2 (St, Tr)	1 (Ad)	8	1.37E-29	1.51E-22	1
PPP2R5D	p.Glu198	Lys	NP_006236.1	4	3.84E-18	4.22E-11			1 (Tr)	4	3.48E-16	3.83E-09	0
ALG13	p.Asn107	Ser, Thr	NP_060936.1	5	5.26E-17	5.79E-10			1 (Tr)	5	1.47E-14	1.62E-07	0
SCN2A	p.Arg937	Cys, His	NP_001035232.1	3	9.47E-12	1.04E-04		1 (ACGC)	1 (ACGC)	4	8.26E-14	9.09E-07	0
SMAD4	p.Ile500	Val, Thr	NP_005350.1	2	1.12E-07	1		2 (An, Le)		4	1.88E-13	2.07E-06	0
PTPN11	p.Gly503	Arg, Glu	NP_002825.3	3	1.66E-11	1.83E-04	1† (AGRE)			3	5.10E-10	5.62E-03	0
GRIA1	p.Ala636	Thr	NP_000818.2	2	1.11E-07	1		1 (St)	2 (St)	3	4.89E-10	5.39E-03	0
SCN2A	p.Arg379	His	NP_001035232.1	2	7.39E-08	8.14E-01			1 (Ad)	2	7.04E-07	1	0
CLCN4	p.Arg718	Trp, Gln	NP_001821.2	2	9.57E-08	1	1 (St)			2	9.11E-07	1	0/2††
KCNQ3	p.Arg230	Cys	NP_004510.1	2	3.36E-07	1			1 (Tr)	2	3.20E-06	1	0
ZNF215	p.Arg473	Gln	NP_037382.2	2	6.54E-07	1			3 (Ad)	2	3.49E-06	1	5
CUX2	p.Glu590	Lys	NP_056082.2	2	5.38E-07	1			1 (Tr)	2	5.12E-06	1	0

*Bonferroni family-wise error rate (FWER) correction based on 1.1E7 codons in genome.

†In two affected siblings.

††Allele 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

Nat Neurosci. Author manuscript; available in PMC 2017 December 19.

Table 2

Rare* clustered missense mutations identified by targeted sequencing (CADD >20).

				Controls		Cases		All cases**		
Gene	**Site or cluster**	**Protein ID**	**denovo-db v.0.9**	**N**	**Missense variants**	**N**	**Missense variants**	**N**	**Missense variants**	**Known *de novo***
PACS1	p.Arg203	NP_060496.2	6	3023	0	17689	4	26166	10	8
PPP2R5D	p.Glu198	NP_006236.1	4	3023	1	17689	3	26166	7	4
SCN2A	p.Arg937	NP_001035232.1	3	3023	0	17689	5	26166	8	5
DEAF1	p.Gln264	NP_066288.2	2	3023	0	17689	7	26166	9	5
ALG13	p.Asn107	NP_060936.1	5	3023	0	17689	1	26166	6	5
GRIA1	p.Ala636	NP_000818.2	2	3023	0	17689	5	26166	7	4
COL4A3BP	p.Ser260	NP_001123577.1	3	3023	0	17689	2	26166	5	3
SCN8A	p.Gly214-p.Asn215	NP_055006.1	3	169	0	11631	1	20108	4	3
DEAF1	p.Leu219-p.Gly220	NP_066288.2	2	169	0	11631	2	20108	4	3
SATB2	p.Arg399-p.Glu402	NP_001165988.1	2	169	0	11631	3	20108	5	3
PPP2R1A	p.Arg182	NP_055040.2	2	3023	0	17689	2	26166	4	3
SCN8A	p.Arg1617-p.Gly1625	NP_055006.1	2	169	0	11631	2	20108	4	3
PTPN11	p.Gly503	NP_002825.3	3	3023	0	17689	6	26166	9	3
STXBP1	p.Gly544	NP_001027392.1	2	3023	0	17689	5	26166	7	3
BCL11A	p.Thr47-p.Cys48	NP_075044.2	2	169	0	11631	4	20108	6	2
KCNQ3	p.Arg230	NP_004510.1	2	3023	0	17689	3	26166	5	3
TRRAP	p.Trp1848	NP_003487.1	2	3023	2	17689	9	26166	11	3
CLTCL1	p.Val641-p.Asn647	NP_009029.3	2	169	0	11631	2	20108	4	2
TRPM7	p.Thr379-p.Glu387	NP_060142.3	2	169	0	11631	2	20108	4	2
SCN2A	p.Arg853	NP_001035232.1	2	3023	1	17689	2	26166	4	2
DUSP15	p.Thr4	NP_001012662.1	2	3023	0	11113	3	19590	5	2
SCN2A	p.Arg379	NP_001035232.1	2	3023	0	17689	6	26166	8	2
SATB2	p.Val381-p.Arg389	NP_001165988.1	2	169	0	11631	1	20108	3	2
SMAD4	p.Ile500	NP_005350.1	2	3023	0	17689	6	26166	8	4
SLC35C2	p.Ser173-p.Gly176	NP_057029.8	2	169	0	11631	3	20108	5	2
CLCN4	p.Arg718	NP_001821.2	2	3023	1	17689	4	26166	6	2
PCGF2	p.Pro65	NP_009075.1	2	3023	0	17689	5	26166	7	2

				Controls		Cases		All cases**		
Gene	**Site or cluster**	**Protein ID**	**denovo-db v.0.9**	**N**	**Missense variants**	**N**	**Missense variants**	**N**	**Missense variants**	**Known *de novo***
KAT6B	p.Ser1380-p.Glu1389	NP_001243398.1	2	169	0	11631	4	20108	6	2
TRIO	p.Pro1461	NP_009049.2	2	3023	0	17689	1	26166	3	2
NCAN	p.Pro1219-p.Val1221	NP_004377.2	2	169	0	11631	8	20108	10	2
ITPR1	p.Thr267-p.Arg269	NP_001161744.1	2	169	0	11631	1	20108	3	2
CUX2	p.Glu590	NP_056082.2	2	3023	1	17689	3	26166	5	2
ZNF215	p.Arg473	NP_037382.2	2	3023	1	11113	10	19590	12	2
SMARCA2	p.Arg525	NP_003061.3	2	3023	0	17689	3	26166	5	2
TBR1	p.Trp271	NP_006584.1	1	3023	0	17689	3	26166	4	1
PDCD11	p.Arg964	NP_055791.1	1	3023	0	17689	8	26166	9	1
		TOTAL	**83**		**7**		**139**		**222**	**101**

*Minor allele frequency (MAF) < 0.01% in ExAC v.0.3 and ESP v.0.0.30.

**denovo-db v.0.9 and smMIPs.

Nat Neurosci. Author manuscript; available in PMC 2017 December 19.