Title	Excess of rare novel loss-of-function variants in synaptic genes in schizophrenia and autism spectrum disorders
Author(s)	Kenny, Elaine M.; Cormican, Paul; Furlong, Sarah; Heron, Eleisa; Kenny, Graham; Fahey, Ciara; Kelleher, Eric; Ennis, Sean; Tropea, Daniela; Anney, Richard; Corvin, Aiden P.; Donohoe, Gary; Gallagher, Louise; Gill, Michael; Morris, Derek W.
Publication Date	2013-10-15
Publication Information	Kenny, E. M.,Cormican, P.,Furlong, S.,Heron, E.,Kenny, G.,Fahey, C.,Kelleher, E.,Ennis, S.,Tropea, D.,Anney, R.,Corvin, A. P.,Donohoe, G.,Gallagher, L.,Gill, M.,Morris, D. W. (2014) 'Excess of rare novel loss-of-function variants in synaptic genes in schizophrenia and autism spectrum disorders'. Molecular Psychiatry, 19 (88):872-9, doi:10.1038/mp.2013.127
Publisher	Springer Nature
Link to publisher's version	https://doi.org/10.1038/mp.2013.127
Item record	http://hdl.handle.net/10379/16269
DOI	http://dx.doi.org/10.1038/mp.2013.127

Excess of rare novel loss-of-function variants in synaptic genes in schizophrenia and autism spectrum disorders

Elaine M. Kenny PhD[1]*, Paul Cormican PhD[1]*, Sarah Furlong PhD[1]*, Eleisa Heron PhD[1], Graham Kenny BSc[1], Ciara Fahey MSc[1], Eric Kelleher MB[1], Sean Ennis PhD[2], Daniela Tropea PhD[1], Richard Anney PhD[1], Aiden P. Corvin PhD[1], Gary Donohoe PhD[1], Louise Gallagher PhD[1], Michael Gill MD[1], Derek W. Morris PhD[1]**.

1 Neuropsychiatric Genetics Research Group, Dept. of Psychiatry and Institute of Molecular Medicine, Trinity College Dublin, Ireland.

2 School of Medicine and Medical Science, University College Dublin, Ireland.

* These authors contributed equally to this work.

** Corresponding author. Mailing address: Room 0.80, Institute of Molecular Medicine, St. James' Hospital, Dublin 8, Ireland. Tel: +353 1 896 2424. Fax: +353 1 896 3405. Email: derek.morris@tcd.ie.

Running title: Rare variant analysis of 215 candidate genes in schizophrenia and autism

This work was supported by grant funding from the Health Research Board (Ireland).

Abstract

Schizophrenia (SZ) and autism spectrum disorders (ASD) are complex neurodevelopmental disorders that may share an underlying pathology suggested by shared genetic risk variants. We sequenced the exonic regions of 215 genes in 147 ASD cases, 273 SZ cases and 287 controls, to identify rare risk mutations. Genes were primarily selected for their function in the synapse and were categorised as: 1) Neurexin and Neuroligin Interacting Proteins, 2) Postsynaptic Glutamate Receptor Complexes, 3) Neural Cell Adhesion Molecules, 4) DISC1 and Interactors, and 5) Functional and Positional Candidates. Thirty-one novel loss-of-function (LoF) variants that are predicted to severely disrupt protein-coding sequence were detected amongst 2,861 rare variants. We found an excess of LoF variants in the combined cases compared to controls (p=0.02). This effect was stronger when analysis was limited to singleton LoF variants (p=0.0007) and the excess was present in both SZ (p=0.002) and ASD (p=0.001). As an individual gene category, Neurexin and Neuroligin Interacting Proteins carried an excess of LoF variants in cases compared to controls (p=0.05). A *de novo* nonsense variant in *GRIN2B* was identified in an ASD case adding to the growing evidence that this is an important risk gene for the disorder. These data support synapse formation and maintenance as key molecular mechanisms for SZ and ASD.

Keywords: schizophrenia, autism, sequencing, mutation, loss-of-function, synapse

Introduction

Schizophrenia and autism spectrum disorders are both neurodevelopmental in origin and are substantially heritable (h^2>0.8)[1, 2]. Schizophrenia (SZ) is characterized by hallucinations, delusions, disordered thinking and cognitive and social deficits. The disorder affects approximately 1% of the population and causes considerable morbidity and mortality[3]. The onset of illness is typically in early adulthood, but the symptoms, severity and course of the disorder are variable. Autism spectrum disorders (ASDs) include autism, Asperger's syndrome and pervasive developmental disorder. They have an onset in childhood and are characterized by impairments in social interaction and communication and a pattern of repetitive behavior and restricted interests [4, 5]. Prototypical ASD is diagnosed in 15–20 per 10,000 children[6], with broader ASD affecting between 60 and 100 in 10,000 [7,8]. Treatments for ASD include behavioural interventions and the use of psychotropic medications to treat comorbid conditions, but core symptoms persist.

SZ and ASD share some clinical features such as cognitive impairment and deficits in social functioning[9] and further support for biological overlap between the disorders comes from epidemiological [10] and neuroimaging studies[11]. The most recent evidence for shared aetiology comes from genetic studies, especially studies of rare copy number variants (CNVs). Many CNVs are common to both disorders, e.g. 1q21.1 [12, 13], 3q29 [14, 15],15q11.2 [16, 17], 15q13.3 [12, 18], 16p11.2 [19, 20], 16p13.11 [21, 22]and 17q12 [23, 24]. There is substantial heterogeneity at these sites in terms of type (deletion or duplication), penetrance and size, and these CNV loci are associated with multiple other neuropsychiatric, developmental and neurological phenotypes [25, 26]. However, in certain instances, mutations in SZ and ASD cases only impact a single gene such as deletions at *NRXN1* suggesting a potential risk mechanism involving synapse function[27-35]. Additional evidence that abnormal synapse formation and maintenance is part of the pathogenesis of both SZ and ASD comes from other CNV studies in SZ [32, 36, 37]and ASD [21, 38], SNP-based group/pathway analysis in SZ [39, 40], transcriptomic analysis of the brain in SZ [41] and ASD [42], and protein interactome analysis in ASD [43]. Where SZ and ASD have been combined for CNV [44] or sequencing [45] analysis, the data supports shared biological pathways for the disorders in synaptogenesis and glutamate neurotransmission.

Based on the emerging evidence that SZ and ASD share common pathogenic mechanisms, we have combined the two disorders in the present study. Here we use next-generation sequencing (NGS) to move beyond CNVs, to the remaining spectrum of potentially rare pathogenic mutations in the form of smaller indels and single nucleotide variants (SNVs). Initial NGS studies in SZ and ASD took the form of whole exome studies of small numbers of trios samples to investigate *de novo* mutation[46-48], family-based exome sequencing in ASD [49] or targeted association studies in SZ of small numbers of candidate genes in pooled DNA samples [50]. These studies indicate a role for rare sequence variation in risk of SZ and ASD. This has been extended by recent and larger exome sequencing studies in ASD [51-53] and SZ [54], which confirmed the importance of *de novo* mutation and the paternal age effect, and for ASD identified new risk genes (e.g. *CHD8, KATNAL2* and *SCN2A*) and provided new support for other strong candidate genes (e.g. *GRIN2B*). Protein-protein interaction network analysis of genes carrying severe *de novo* mutations indicates that a high proportion of these genes have a function in neuronal development [51].

Using our Multiplex Target Enrichment method [55], we adopted a focused approach and sequenced 215 candidate genes, selected primarily for their role in synaptic function and neurodevelopment, in a total sample of 743 individuals to detect rare sequence variations. Genes are grouped into five categories based on the biological basis for their selection, which briefly include (1) Neurexin and Neuroligin Interacting Proteins (n=46), (2) Post-synaptic Glutamate Receptor Complexes (n=58), (3) Neural Cell Adhesion Molecules (n=61), (4) DISC1 and Interacting Proteins (n=23) and (5) other Positional and Functional Candidates (n=27). Within these genes, our primary objective was to detect rare Loss-of-Function (LoF) variants that are predicted to severely disrupt protein-coding sequence. We tested for and found a significant excess of these disruptive mutations in our combined SZ and ASD case sample compared to controls, and for some ASD cases found that mutations were *de novo*. We brought these data forward to further experiments designed to elucidate the biological relevance of these variants in specific gene networks and intermediate cognitive and clinical phenotypes. In addition, we studied all rare missense variants for evidence that this class of mutation increases risk for these neurodevelopmental disorders in our selected networks and genes.

Materials and Methods

Samples

SZ case samples (n=297) were recruited through community mental services and inpatient units in the Republic of Ireland with local ethics approval. All participants were interviewed using a structured clinical interview (Structured Clinical Interview for DSM-IV (SCID-P; ISBN:0880489324)). Diagnosis of a major psychotic disorder was made by the consensus lifetime best estimate method using DSM-IV criteria with all available information (interview, family or staff report and chart review). This sample is described in greater detail elsewhere [56]. The final sample (n=273) used for analysis was 65.2% male and had a mean age at collection of 47.1 years (standard deviation (SD) = 19.4). In selecting the sample, we specifically wanted to include cases with low pre-morbid IQ (n=110 of 188 with available data) and cases that also had another recorded developmental disorders (epilepsy (n=3), speech delay (n=8)). Of the final 273 SZ cases, clinical data on symptom severity, collected using the SAPS and the SANS was available for 245 patients. Neuropsychological data was available for 188 SZ cases, collected using a battery of clinical and neuropsychological measures as previously reported[57]. ASD case samples (n=152) were recruited through schools, parent support groups and clinician referral with local ethics approval. Autism diagnoses were confirmed using Autism Diagnostic Interview – Revised (ADI-R) and the Autism Diagnostic Observation Schedule - Generic (ADOS-G). This childhood sample is described in greater detail elsewhere[58]. The final sample used for analysis was 83.1% male. Control samples (n=294) were ascertained with informed consent from the Trinity Biobank and represent blood donors from the Irish Blood Transfusion Service recruited in the Republic of Ireland[56]. As the lifetime prevalence of schizophrenia or autism is relatively low (<1%), there is no obvious reason for individuals with either disorder to be over represented in the controls. DNA for all samples was extracted from blood. The final sample used for analysis was 65.9%male and had a mean age at collection of 34.0 years (SD = 12.6).

Gene Selection

Definitions of the five gene categories and the method of gene selection are described in full in supplementary information and the full list of gene IDs is detailed in supplementary table A along with the data source that underpinned each selection. In brief, the process involved extensive literature searches, with key references identified in the next sentence, and the use of KEGG (http://www.genome.jp/kegg/pathway.html) and online interaction databases HPRD (http://www.hprd.org/), String (http://string-db.org/), IntAct (http://www.ebi.ac.uk/intact/), BioGRID (http://thebiogrid.org/) and BOND (http://bond.unleashedinformatics.com/). The five gene categories were (1) Neurexin and Neuroligin Interacting Proteins[27-35], (2) Post-synaptic Glutamate Receptor Complexes[59, 60], (3) Neural Cell Adhesion Molecules[60], (4) DISC1 and Interacting Proteins[61] and (5) other Positional and Functional Candidates. The functional categories were used sequentially to select candidate genes. Therefore, "Neurexin and Neuroligin Interacting Proteins" were selected first followed by genes that encoded "Post-synaptic Glutamate Receptor Complexes" that were not already selected for the "Neurexin and Neuroligin Interacting Proteins" category. We next moved to the third category "Neural

Cell Adhesion Molecules" and again selected genes not already picked for categories 1 and 2 and so on. Consequently, there are many instances of genes that could fit in multiple categories. These categories were maintained during association analysis as any re-categorization of genes after data generation could have biased analyses.

Targeted Sequencing, Quality Control and Variant Annotation

The process of sequencing, QC and variant annotation are fully detailed in supplementary information. In brief, samples were indexed and multiplexed in groups of 24. The exons of 215 genes were targeted using Agilent's SureSelect Target Enrichment system (total target = 1,064,238bp) and sequenced on an Illumina Genome Analyzer II. Sequence alignment and calling of both SNVs and indels was performed using GATK (v1.0.5506[62]). The median coverage for all samples included in the final analysis was 41x for SZ, 66x for ASD and 52x for controls (supplementary figure A). Following removal of poorly performing samples and low quality variant calls, variants were classified as rare if they had a minor allele frequency (MAF) <0.01 in the combined case-control sample [63, 64]. The average matching between available GWAS data and sequence data variant calls was >99%. All variants were functionally annotated using SNPeff (v2.0.5; http://snpeff.sourceforge.net/). Analysis of silent SNVs show an average of 167 per SZ sample (SD=12.6), 168 per ASD sample (SD=12.3) and 167 variants per control sample (SD=12.8), indicating an even rate of variant detection across each sample group. Loss-of-Function (LoF) variants are predicted to severely disrupt protein-coding sequence and we used the definition of LoF variants as suggested by MacArthur et al[65]: nonsense SNVs that introduce stop codons, SNVs that disrupt canonical splice sites, and indels that disrupt a transcript's open reading frame or a canonical splice site. We did not consider mutations as putative LoF variants in association analyses if they were located in the last 5% of coding sequence[65]. All rare missense SNVs were assigned a PolyPhen2[66] and SIFT [67] score.

Association Analysis

Our primary analyses was to examine whether there is an excess of rare LoF variants in the combined SZ and ASD case sample versus controls using data from all genes together. This was done using a carrier-based association analysis where case and control samples were categorized as either carriers or non-carriers of at least one rare LoF variant and tested for association using a 2x2 contingency table. Results for chi-square tests are reported except where indicated that a two-tailed Fisher exact test was used because an expected cell count was <5. Where we achieved a nominally significant result ($p<0.05$), we (a) performed the same carrier-based analysis on SZ and ASD cases separately to observe the effect in the individual case groups and (b) tested within each of the gene categories. For the rare missense variants we performed the same carrier-based association analysis for all genes in the combined case group and repeated this for the individual gene categories and the individual genes. We also tested for pairs of interacting genes that were hit by multiple rare missense variants in cases compared to controls.

Results

Figure 1 provides a flowchart of the number of variants detected across all samples and how that number was reduced to a set of variants for inclusion in our association analysis. In total we found 33 rare LoF variants in our sample. All variants were subjected to Sanger sequencing and 31 of 33 were confirmed by this method; 11 nonsense SNVs, 12 frameshift indels, 6 splice site SNVs, 1 splice site indel and 1 stop loss SNV (table 1). All variants were novel of which 27 were singletons and 4 were found in more than one sample. Including data on all genes, we found an excess of individuals carrying LoF variants in our combined SZ and ASD case sample compared to controls (29 in 420 cases v 8 in 287 controls; *p=0.02*) with the effect stronger for ASD (13 in 147 cases; *p=0.005*) than for SZ (16 in 273 cases; *p=0.07;* table 2). In order to focus in on variants that may be most deleterious, we dropped 3 low-frequency variants found in multiple samples that may represent benign variants circulating in the population. All 3 variants were found in both cases and controls. When the analysis is limited to variants that only occur in one individual (singleton variants), the data show a significant excess of LoF variants in the combined case sample versus controls (23 in 420 cases v 2 in 287 controls; *p=0.0007*) and the effect is similar for both ASD (9 in 147 cases; *p=0.001*) and SZ (14 in 273 cases; *p=0.002;* table 2).

Insert figure 1, table 1 and table 2 about here

Following analysis of all genes combined, we next tested rare LoF variants in the individual gene categories. The Neurexin and Neuroligin Interacting Proteins grouping contained the highest number of these variants and a significant excess in cases (9 in cases (7xSZ and 2xASD) and 1 in controls, *p=0.05* for SZ+ASD; *p=0.03* for SZ; *p=0.27* for ASD (all Fisher exact tests)). Results for all other gene categories were non-significant but the number of observations are small, e.g. for Post-synaptic Glutamate Receptor Complexes there were 5 LoF variants in SZ+ASD cases and 0 in controls (*p=0.08*; Fisher exact test). At the level of individual genes, only *DST* had enough LoF variants to warrant a test (9 in cases (5xSZ and 4xASD) and 3 in controls, *p=0.38*). The only other gene where we found more than two LoF variants was *INADL* (n= 3) and interestingly all were in cases. The effects of LoF mutations on cognitive and clinical intermediate phenotypes were assessed separately in SZ and ASD by comparing carriers versus non-carriers within each diagnostic group. Across both diagnostic groups little evidence was found to suggest that the LoF variant carriers differed significantly on clinical and cognitive metrics from non-carriers (see supplementary information).

Sanger sequencing of parental DNA that was available for the ASD samples revealed that the LoF variant at *GRIN2B* was *de novo*. The nonsense SNV (Q711*) at *GRIN2B* is located in exon 10 and parent of origin analysis indicated that it was on the maternal chromosome. The previously reported *de novo* LoF variants at *GRIN2B* in autism are a frameshift indel in exon 2, a nonsense SNV in exon 8 and a splice site SNV at exon 11[51]. Initial sequencing of parental samples for the *DISC1* variant indicated that it was *de novo*. This is a frameshift indel that affects transcript variant b (NM_001164538), which lacks two 3' exons of longer transcripts but has an alternate 3' segment. The frameshift occurs in this alternate segment and because of its position towards the end of the coding sequence, it was not included in our association analysis. Molecular analysis will be required to determine the functional impact of this variant. Parent of origin analysis indicated that this variant was on the paternal chromosome but closer study of the paternal DNA

revealed evidence of the LoF allele, suggesting possible mosaicism in the father's blood cells and that the variant is not *de novo* in the proband.

Finally we performed association analysis of the 1,299 rare missense SNVs identified in our sample of which 403 were classified as functional based on PolyPhen2/SIFT scores. Genes were grouped as follows: (a) All Genes, (b) Neurexin and Neuroligin Interacting Proteins, (c) Post-synaptic Glutamate Receptor Complexes, (d) Neural Cell Adhesion Molecules, (e) DISC1 and Interacting Proteins and (f) LoF-containing Genes (n=18 genes that contained a rare LoF variant). For each gene group, we plotted the number of cases (SZ and ASD combined) and controls that carried 0, ≥1, ≥2, ≥3 etc. rare functional missense SNVs (figure 2a-f). We tested the number of samples that carried at least 1 rare functional missense SNV in cases versus controls and did not detect any significant differences for any of the gene categories. Similarly, when we plotted SZ and ASD separately, there were no significant differences between cases and controls. We also tested for a difference between cases and controls for the number of carriers of at least 1 rare functional missense SNV at each individual gene. Q-Q plots indicate a lesser number of nominally associated genes than would have been expected by chance, most likely reflecting the small number of variants included in the analysis of each gene (see supplementary information). None of the 18 genes containing LoF variants had a significant difference in carrier-number of rare functional missense SNVs between cases and controls for either the combined or individual disorders. In addition, within gene categories (b) – (e) above, analysis of interacting gene pairs did not identify any pairs that were hit by mutations at a significantly different rate in cases compared to controls (see supplementary information).

Discussion

By taking a targeted sequencing approach to the detection of rare variants, we add further support to the convergent evidence that synapse formation and maintenance are components of the pathophysiology of SZ and ASD. In our set of 215 candidate genes, we primarily focused on rare LoF variants that are likely to be most disruptive based on their predicted impact on protein-coding sequence. We find a significant excess of novel variants in our combined case sample and in ASD compared to controls. The selection of a MAF<0.1 as a frequency cut-off for rare variants is arbitrary; not all variants above this threshold will be benign and not all variants below this threshold will be pathogenic. But highly pathogenic variants are likely to be rare or even unique. Therefore, to focus on variants that may be most deleterious, we performed an association analysis of singleton variants. There was a significant excess of singleton LoF variants in the combined case sample and for both ASD and SZ when analyzed separately.

When we tested the individual gene categories, we observed a significant excess of variants in Neurexin and Neuroligin Interacting Proteins. Here we found a variant in a male SZ case in the X-linked *NLGN3* gene, which had previously been reported to harbour rare risk variants in ASD[68]. In this category, we found three LoF variants in *INADL*, all in case samples (2xSZ and 1xASD). *INADL* functions to help anchor transmembrane proteins to the cytoskeleton and organize signaling complexes. It interacts with neurexins and neuroligins and is important for cell polarity, migration and may play a role in neurite extension [69, 70]. Also in the Neurexin category is *FYN* where we found a LoF variant in a SZ case that also had epilepsy. *FYN* is a Src family protein tyrosine kinase and is a key regulator of NR2B (encoded by *GRIN2B*) of the NMDA receptor (NMDAR)[71]. Fyn-mutant mice exhibit blunting of long term potentiation and impaired spatial learning plus other neurological defects including uncoordinated hippocampal architecture and reduced neural CAM-dependent neurite outgrowth [72, 73]. Studies using Fyn-deficient mice support a role for *FYN* in the induction of epilepsy [74]. Our data further supports *FYN* as a putative risk gene for SZ and/or epilepsy. Interesting, only two other SZ cases in the study had comorbid epilepsy and both were found to carry LoF variants, in *MACF1* (also in the neurexin category) and in *PLXNA2*. These samples were not included in previous SZ GWAS because of the comorbid epilepsy but highlight the value of taking an inclusive approach when selecting phenotype for rare variant studies.

After the Neurexin and Neuroligin Interacting Proteins gene category, no other categories had a significant excess of LoF variants but the number of observations are small, e.g. for Post-synaptic Glutamate Receptor Complexes there were 5 LoF variants in SZ+ASD cases and 0 in controls. One of these variants was a *de novo* nonsense mutation in an ASD case at *GRIN2B*, which adds to the three recently reported *de novo* LoF mutations in other ASD samples[51] and supports *GRIN2B* as a risk gene for the disorder. Other data indicate that mutation at *GRIN2B* can contribute to various neurodevelopmental disorders. Endele at al[75] identified *de novo* translocations with breakpoints disrupting *GRIN2B* in two individuals, one with mild mental retardation (MR)(46,XY,t(9;12)(p23;p13.1)), one with severe MR (46,XY,t(10;12)(q21.1;p13)). Further screening of*GRIN2B* for mutations in 468 individuals with MR and/or epilepsy identified four individuals with moderate MR and behavioural anomalies who had *de novo GRIN2B* mutations; a missense SNV, splice donor SNV, splice acceptor SNV and a 2bp frameshift deletion. Talkowski et al [76] characterized balanced chromosomal abnormalities in 38 subjects with neurodevelopmental abnormalities and identified a *de novo* translocation in an ASD case (46,XY,inv(12)(p13.1q21.31)dn) that disrupted

GRIN2B.*GRIN2B* encodes the glutamate-binding NR2B subunit of the NMDAR and is important for channel function, organization of postsynaptic macromolecular complexes, dendritic spine formation or maintenance, and regulation of the actin cytoskeleton [77]. Overexpression of the gene in animal models is associated with improved performance in learning and memory [78, 79]. *GRIN2B* mutations in humans may affect brain function and cognition by disturbing the electrophysiological balance of the receptor during neurodevelopment[75].

We detected two LoF variants in *GRIP1* (1xSZ and 1xASD). *GRIP1* is a member of the glutamate receptor interacting protein family and plays a role in receptor trafficking, synaptic organization, and transmission in glutamatergic and GABAergic synapses[80]. A recent study identified 5 rare missense variants in highly conserved regions of the gene in ASD cases only[81]. These variants were shown to be associated with altered *GRIP1* interaction with glutamate receptors, faster recycling and increased surface distribution of GluA2 in neurons *in vitro*, which supports a gain of *GRIP1* function in these variants. Knockout mouse studies demonstrated that *GRIP1* is essential for embryonic development and deficits in the protein lead to increased prepulse inhibition[81].

Finally, the gene with the largest number of rare LoF variants was *DST* (Dystonin), a very large and transcriptionally complex gene that encodes multiple isoforms. It is a member of the plakin family of cytolinker proteins, which link cytoskeletal networks to each other and to junctional complexes. *DST* is expressed throughout mouse development and loss of its function results in neuromuscular dysfunction and early death in the mouse mutant *dystonia musculorum*[82, 83]. Deleterious recessive mutations in *DST* have been identified as the likely cause of a lethal autonomic sensory neuropathy [84]. There is no additional evidence in the literature supporting rare variants at *DST* in SZ or ASD.

Phenotypic analysis of individual LoF carriers in the SZ and ASD samples did not identify any specific phenotypic characteristics. For SZ, it should be noted that when patients were originally chosen for inclusion in this study, we sought to include patients who showed deficits in cognitive performance. By definition, this lowered average cognitive performance scores for this group. Therefore, it is possible that our statistical approach was somewhat biased by comparison with a general SZ population. This reflects a broader issue in the study of symptom severity and cognitive function in rare variant carriers; that is how to classify the performance of individual carriers against an appropriate test group using appropriate statistical approaches. Investigators will want to move away from analysis of individual samples and instead study very large datasets where either multiple samples with rare variants in the same gene or ideally multiple samples with the same rare variant will be available for study.

In conclusion, we have used a focused targeted sequencing study of rare LoF variation to add to the growing volume of data supporting synapse formation and maintenance as key molecular mechanisms in the neurodevelopmental disorders SZ and ASD. We specifically find more evidence that rare variation in genes with Neurexin-related function increases risk of SZ and ASD. The two disorders share some risk genes but there is not yet enough data to suggest that they share the same mutations. A major challenge for genetic analysis of both disorders will be to successfully understand the contribution and possible interaction of both common and rare variants. Synaptic function has been the focus of this rare variant study and an interesting example of how a common risk variant may impact the same molecular

mechanisms has recently been reported in SZ. Knockdown of *ZNF804A*, site of the first genome-wide associated SNP for psychosis [85], alters the expression of genes involved in cell adhesion, suggesting a role for *ZNF804A* in neural migration, neurite outgrowth and synapse formation [86]. In terms of specific genes, our work supports *GRIN2B* as a risk gene in ASD and adds further to data implicating *GRIP1* in ASD. We identify *FYN* as a putative risk gene for SZ and/or epilepsy and highlight multiple genes as potential susceptibility loci for these neurodevelopmental disorders that will require independent support from future sequencing studies.

Supplementary information is available online

Acknowledgements

The authors sincerely thank all patients who contributed to this study and all staff who facilitated their involvement. Funding for this study was provided by the Health Research Board (HRB Ireland; HRA/2009/45) and Science Foundation Ireland (SFI; 08/IN.1/B1916). Next generation sequencing was performed in TrinSeq (Trinity Genome Sequencing Laboratory; http://www.medicine.tcd.ie/sequencing), a core facility funded by SFI under Grant No. [07/RFP/GEN/F327/EC07] to Dr. Morris. Ms Furlong's PhD studentship is funded by the HRB 4-Year PhD Programme in Molecular Medicine at TCD. We acknowledge use of the Trinity Biobank control sample and support from the Trinity Centre for High Performance Computing.

Conflict of interest

All authors declare that they have no competing financial interests in relation to this work.

References

1. Cardno AG, Gottesman, II. Twin studies of schizophrenia: from bow-and-arrow concordances to star wars Mx and functional genomics. *American journal of medical genetics* 2000; **97**(1)**:** 12-17.

2. Freitag CM. The genetics of autistic disorders and its clinical relevance: a review of the literature. *Molecular psychiatry* 2007; **12**(1)**:** 2-22.

3. Tiihonen J, Lonnqvist J, Wahlbeck K, Klaukka T, Niskanen L, Tanskanen A *et al.* 11-year follow-up of mortality in patients with schizophrenia: a population-based cohort study (FIN11 study). *Lancet* 2009; **374**(9690)**:** 620-627.

4. Lord C, Rutter M, Le Couteur A. Autism Diagnostic Interview-Revised: a revised version of a diagnostic interview for caregivers of individuals with possible pervasive developmental disorders. *Journal of autism and developmental disorders* 1994; **24**(5)**:** 659-685.

5. Lord C, Risi S, Lambrecht L, Cook EH, Jr., Leventhal BL, DiLavore PC *et al.* The autism diagnostic observation schedule-generic: a standard measure of social and communication deficits associated with the spectrum of autism. *Journal of autism and developmental disorders* 2000; **30**(3)**:** 205-223.

6. Fombonne E. Epidemiology of pervasive developmental disorders. *Pediatric research* 2009; **65**(6)**:** 591-598.

7. Baron-Cohen S, Scott FJ, Allison C, Williams J, Bolton P, Matthews FE *et al.* Prevalence of autism-spectrum conditions: UK school-based population study. *The British journal of psychiatry : the journal of mental science* 2009; **194**(6)**:** 500-509.

8. Fernell E, Gillberg C. Autism spectrum disorder diagnoses in Stockholm preschoolers. *Research in developmental disabilities* 2010; **31**(3)**:** 680-685.

9. Craddock N, Owen MJ. Data and clinical utility should be the drivers of changes to psychiatric classification. *The British journal of psychiatry : the journal of mental science* 2010; **197**(2)**:** 158; author reply 158-159.

10. Sullivan PF, Magnusson C, Reichenberg A, Boman M, Dalman C, Davidson M *et al.* Family History of Schizophrenia and Bipolar Disorder as Risk Factors for AutismFamily History of Psychosis as Risk Factor for ASD. *Archives of general psychiatry* 2012**:** 1-5.

11. King BH, Lord C. Is schizophrenia on the autism spectrum? *Brain research* 2011; **1380:** 34-41.

12. Rare chromosomal deletions and duplications increase risk of schizophrenia. *Nature* 2008; **455**(7210)**:** 237-241.

13. Mefford HC, Sharp AJ, Baker C, Itsara A, Jiang Z, Buysse K *et al.* Recurrent rearrangements of chromosome 1q21.1 and variable pediatric phenotypes. *The New England journal of medicine* 2008; **359**(16)**:** 1685-1699.

14. Ballif BC, Theisen A, Coppinger J, Gowans GC, Hersh JH, Madan-Khetarpal S *et al.* Expanding the clinical phenotype of the 3q29 microdeletion syndrome and characterization of the reciprocal microduplication. *Molecular cytogenetics* 2008; **1:** 8.

15. Levinson DF, Duan J, Oh S, Wang K, Sanders AR, Shi J *et al.* Copy number variants in schizophrenia: confirmation of five previous findings and new evidence for 3q29 microdeletions and VIPR2 duplications. *The American journal of psychiatry* 2011; **168**(3)**:** 302-316.

16. Stefansson H, Rujescu D, Cichon S, Pietilainen OP, Ingason A, Steinberg S *et al.* Large recurrent microdeletions associated with schizophrenia. *Nature* 2008; **455**(7210)**:** 232-236.

17. Hogart A, Wu D, LaSalle JM, Schanen NC. The comorbidity of autism with the genomic disorders of chromosome 15q11.2-q13. *Neurobiology of disease* 2010; **38**(2)**:** 181-191.

18. Miller DT, Shen Y, Weiss LA, Korn J, Anselm I, Bridgemohan C *et al.* Microdeletion/duplication at 15q13.2q13.3 among individuals with features of autism and other neuropsychiatric disorders. *Journal of medical genetics* 2009; **46**(4)**:** 242-248.

19. Weiss LA, Shen Y, Korn JM, Arking DE, Miller DT, Fossdal R *et al.* Association between microdeletion and microduplication at 16p11.2 and autism. *The New England journal of medicine* 2008; **358**(7)**:** 667-675.

20. McCarthy SE, Makarov V, Kirov G, Addington AM, McClellan J, Yoon S *et al.* Microduplications of 16p11.2 are associated with schizophrenia. *Nature genetics* 2009; **41**(11)**:** 1223-1227.

21. Pinto D, Pagnamenta AT, Klei L, Anney R, Merico D, Regan R *et al.* Functional impact of global rare copy number variation in autism spectrum disorders. *Nature* 2010; **466**(7304)**:** 368-372.

22. Ingason A, Rujescu D, Cichon S, Sigurdsson E, Sigmundsson T, Pietilainen OP *et al.* Copy number variations of chromosome 16p13.1 region associated with schizophrenia. *Molecular psychiatry* 2011; **16**(1)**:** 17-25.

23. Kirov G, Grozeva D, Norton N, Ivanov D, Mantripragada KK, Holmans P *et al.* Support for the involvement of large copy number variants in the pathogenesis of schizophrenia. *Human molecular genetics* 2009; **18**(8)**:** 1497-1503.

24. Moreno-De-Luca D, Mulle JG, Kaminsky EB, Sanders SJ, Myers SM, Adam MP *et al.* Deletion 17q12 is a recurrent copy number variant that confers high risk of autism and schizophrenia. *American journal of human genetics* 2010; **87**(5)**:** 618-630.

25. Cooper GM, Coe BP, Girirajan S, Rosenfeld JA, Vu TH, Baker C *et al.* A copy number variation morbidity map of developmental delay. *Nature genetics* 2011; **43**(9)**:** 838-846.

26. Malhotra D, Sebat J. CNVs: harbingers of a rare variant revolution in psychiatric genetics. *Cell* 2012; **148**(6)**:** 1223-1241.

27. Szatmari P, Paterson AD, Zwaigenbaum L, Roberts W, Brian J, Liu XQ *et al.* Mapping autism risk loci using genetic linkage and chromosomal rearrangements. *Nature genetics* 2007; **39**(3)**:** 319-328.

28. Kirov G, Gumus D, Chen W, Norton N, Georgieva L, Sari M *et al.* Comparative genome hybridization suggests a role for NRXN1 and APBA2 in schizophrenia. *Human molecular genetics* 2008; **17**(3)**:** 458-465.

29. Kim HG, Kishikawa S, Higgins AW, Seong IS, Donovan DJ, Shen Y *et al.* Disruption of neurexin 1 associated with autism spectrum disorder. *American journal of human genetics* 2008; **82**(1)**:** 199-207.

30. Marshall CR, Noor A, Vincent JB, Lionel AC, Feuk L, Skaug J *et al.* Structural variation of chromosomes in autism spectrum disorder. *American journal of human genetics* 2008; **82**(2)**:** 477-488.

31. Vrijenhoek T, Buizer-Voskamp JE, van der Stelt I, Strengman E, Sabatti C, Geurts van Kessel A *et al.* Recurrent CNVs disrupt three candidate genes in schizophrenia patients. *American journal of human genetics* 2008; **83**(4)**:** 504-510.

32. Walsh T, McClellan JM, McCarthy SE, Addington AM, Pierce SB, Cooper GM *et al.* Rare structural variants disrupt multiple genes in neurodevelopmental pathways in schizophrenia. *Science* 2008; **320**(5875)**:** 539-543.

33. Glessner JT, Wang K, Cai G, Korvatska O, Kim CE, Wood S *et al.* Autism genome-wide copy number variation reveals ubiquitin and neuronal genes. *Nature* 2009; **459**(7246)**:** 569-573.

34. Rujescu D, Ingason A, Cichon S, Pietilainen OP, Barnes MR, Toulopoulou T *et al.* Disruption of the neurexin 1 gene is associated with schizophrenia. *Human molecular genetics* 2009; **18**(5)**:** 988-996.

35. Reichelt AC, Rodgers RJ, Clapcote SJ. The role of neurexins in schizophrenia and autistic spectrum disorder. *Neuropharmacology* 2012; **62**(3)**:** 1519-1526.

36. Malhotra D, McCarthy S, Michaelson JJ, Vacic V, Burdick KE, Yoon S *et al.* High frequencies of de novo CNVs in bipolar disorder and schizophrenia. *Neuron* 2011; **72**(6)**:** 951-963.

37. Kirov G, Pocklington AJ, Holmans P, Ivanov D, Ikeda M, Ruderfer D *et al.* De novo CNV analysis implicates specific abnormalities of postsynaptic signalling complexes in the pathogenesis of schizophrenia. *Molecular psychiatry* 2012; **17**(2)**:** 142-153.

38. Gilman SR, Iossifov I, Levy D, Ronemus M, Wigler M, Vitkup D. Rare de novo variants associated with autism implicate a large functional network of genes involved in formation and function of synapses. *Neuron* 2011; **70**(5)**:** 898-907.

39. Lips ES, Cornelisse LN, Toonen RF, Min JL, Hultman CM, Holmans PA *et al.* Functional gene group analysis identifies synaptic gene groups as risk factor for schizophrenia. *Molecular psychiatry* 2012; **17**(10)**:** 996-1006.

40. O'Dushlaine C, Kenny E, Heron E, Donohoe G, Gill M, Morris D *et al.* Molecular pathways involved in neuronal cell adhesion and membrane scaffolding contribute to schizophrenia and bipolar disorder susceptibility. *Molecular psychiatry* 2011; **16**(3)**:** 286-292.

41. Mudge J, Miller NA, Khrebtukova I, Lindquist IE, May GD, Huntley JJ *et al.* Genomic convergence analysis of schizophrenia: mRNA sequencing reveals altered synaptic vesicular transport in post-mortem cerebellum. *PloS one* 2008; **3**(11)**:** e3625.

42. Voineagu I, Wang X, Johnston P, Lowe JK, Tian Y, Horvath S *et al.* Transcriptomic analysis of autistic brain reveals convergent molecular pathology. *Nature* 2011; **474**(7351)**:** 380-384.

43. Sakai Y, Shaw CA, Dawson BC, Dugas DV, Al-Mohtaseb Z, Hill DE *et al.* Protein interactome reveals converging molecular pathways among autism disorders. *Science translational medicine* 2011; **3**(86)**:** 86ra49.

44. Guilmatre A, Dubourg C, Mosca AL, Legallic S, Goldenberg A, Drouin-Garraud V *et al.* Recurrent rearrangements in synaptic and neurodevelopmental genes and shared biologic pathways in schizophrenia, autism, and mental retardation. *Archives of general psychiatry* 2009; **66**(9)**:** 947-956.

45. Myers RA, Casals F, Gauthier J, Hamdan FF, Keebler J, Boyko AR *et al.* A population genetic approach to mapping neurological disorder genes using deep resequencing. *PLoS genetics* 2011; **7**(2)**:** e1001318.

46. Girard SL, Gauthier J, Noreau A, Xiong L, Zhou S, Jouan L *et al.* Increased exonic de novo mutation rate in individuals with schizophrenia. *Nature genetics* 2011; **43**(9)**:** 860-863.

47. O'Roak BJ, Deriziotis P, Lee C, Vives L, Schwartz JJ, Girirajan S *et al.* Exome sequencing in sporadic autism spectrum disorders identifies severe de novo mutations. *Nature genetics* 2011; **43**(6)**:** 585-589.

48. Xu B, Roos JL, Dexheimer P, Boone B, Plummer B, Levy S *et al.* Exome sequencing supports a de novo mutational paradigm for schizophrenia. *Nature genetics* 2011; **43**(9)**:** 864-868.

49. Chahrour MH, Yu TW, Lim ET, Ataman B, Coulter ME, Hill RS *et al.* Whole-exome sequencing and homozygosity analysis implicate depolarization-regulated neuronal genes in autism. *PLoS genetics* 2012; **8**(4)**:** e1002635.

50. Moens LN, De Rijk P, Reumers J, Van den Bossche MJ, Glassee W, De Zutter S *et al.* Sequencing of DISC1 pathway genes reveals increased burden of rare missense variants in schizophrenia patients from a northern Swedish population. *PloS one* 2011; **6**(8)**:** e23450.

51. O'Roak BJ, Vives L, Girirajan S, Karakoc E, Krumm N, Coe BP *et al.* Sporadic autism exomes reveal a highly interconnected protein network of de novo mutations. *Nature* 2012; **485**(7397)**:** 246-250.

52. Neale BM, Kou Y, Liu L, Ma'ayan A, Samocha KE, Sabo A *et al.* Patterns and rates of exonic de novo mutations in autism spectrum disorders. *Nature* 2012; **485**(7397)**:** 242-245.

53. Sanders SJ, Murtha MT, Gupta AR, Murdoch JD, Raubeson MJ, Willsey AJ *et al.* De novo mutations revealed by whole-exome sequencing are strongly associated with autism. *Nature* 2012; **485**(7397)**:** 237-241.

54. Xu B, Ionita-Laza I, Roos JL, Boone B, Woodrick S, Sun Y *et al.* De novo gene mutations highlight patterns of genetic and neural complexity in schizophrenia. *Nature genetics* 2012.

55. Kenny EM, Cormican P, Gilks WP, Gates AS, O'Dushlaine CT, Pinto C *et al.* Multiplex target enrichment using DNA indexing for ultra-high throughput SNP detection. *DNA research : an international journal for rapid publication of reports on genes and genomes* 2011; **18**(1)**:** 31-38.

56. Genome-Wide Association Study Implicates HLA-C*01:02 as a Risk Factor at the Major Histocompatibility Complex Locus in Schizophrenia. *Biological psychiatry* 2012; **72**(8)**:** 620-628.

57. Donohoe G, Walters J, Morris DW, Quinn EM, Judge R, Norton N *et al.* Influence of NOS1 on verbal intelligence and working memory in both patients with schizophrenia and healthy control subjects. *Archives of general psychiatry* 2009; **66**(10)**:** 1045-1054.

58. Cochrane LE, Tansey KE, Gill M, Gallagher L, Anney RJ. Lack of association between markers in the ITGA3, ITGAV, ITGA6 and ITGB3 and autism in an Irish sample. *Autism research : official journal of the International Society for Autism Research* 2010; **3**(6)**:** 342-344.

59. Husi H, Ward MA, Choudhary JS, Blackstock WP, Grant SG. Proteomic analysis of NMDA receptor-adhesion protein signaling complexes. *Nature Neuroscience* 2000; **3**(7): 661-669.

60. Laumonnier F, Cuthbert PC, Grant SG. The role of neuronal complexes in human X-linked brain diseases. *American Journal of Human Genetics* 2007; **80**(2): 205-220.

61. Camargo LM, Collura V, Rain JC, Mizuguchi K, Hermjakob H, Kerrien S, *et al*. Disrupted in Schizophrenia 1 Interactome: evidence for the close connectivity of risk genes and a potential synaptic basis for schizophrenia. *Molecular Psychiatry* 2007; **12**(1): 74-86.

62. DePristo MA, Banks E, Poplin R, Garimella KV, Maguire JR, Hartl C *et al.* A framework for variation discovery and genotyping using next-generation DNA sequencing data. *Nature genetics* 2011; **43**(5)**:** 491-498.

63. Lemire M. Defining rare variants by their frequencies in controls may increase type I error. *Nature genetics* 2011; **43**(5)**:** 391-392; author reply 394-395.

64. Pearson RD. Bias due to selection of rare variants using frequency in controls. *Nature genetics* 2011; **43**(5)**:** 392-393; author reply 394-395.

65. MacArthur DG, Balasubramanian S, Frankish A, Huang N, Morris J, Walter K *et al.* A systematic survey of loss-of-function variants in human protein-coding genes. *Science* 2012; **335**(6070)**:** 823-828.

66. Adzhubei IA, Schmidt S, Peshkin L, Ramensky VE, Gerasimova A, Bork P *et al.* A method and server for predicting damaging missense mutations. *Nature methods* 2010; **7**(4)**:** 248-249.

67. Kumar P, Henikoff S, Ng PC. Predicting the effects of coding non-synonymous variants on protein function using the SIFT algorithm. *Nature protocols* 2009; **4**(7)**:** 1073-1081.

68. Jamain S, Quach H, Betancur C, Rastam M, Colineaux C, Gillberg IC *et al.* Mutations of the X-linked genes encoding neuroligins NLGN3 and NLGN4 are associated with autism. *Nature genetics* 2003; **34**(1)**:** 27-29.

69. Kurschner C, Mermelstein PG, Holden WT, Surmeier DJ. CIPP, a novel multivalent PDZ domain protein, selectively interacts with Kir4.0 family members, NMDA receptor subunits, neurexins, and neuroligins. *Molecular and cellular neurosciences* 1998; **11**(3)**:** 161-172.

70. Barilari M, Dente L. The neuronal proteins CIPP, Cypin and IRSp53 form a tripartite complex mediated by PDZ and SH3 domains. *Biological chemistry* 2010; **391**(10)**:** 1169-1174.

71. Trepanier CH, Jackson MF, MacDonald JF. Regulation of NMDA receptors by the tyrosine kinase Fyn. *The FEBS journal* 2012; **279**(1)**:** 12-19.

72. Grant SG, O'Dell TJ, Karl KA, Stein PL, Soriano P, Kandel ER. Impaired long-term potentiation, spatial learning, and hippocampal development in fyn mutant mice. *Science* 1992; **258**(5090)**:** 1903-1910.

73. Beggs HE, Soriano P, Maness PF. NCAM-dependent neurite outgrowth is inhibited in neurons from Fyn-minus mice. *The Journal of cell biology* 1994; **127**(3)**:** 825-833.

74. Cain DP, Grant SG, Saucier D, Hargreaves EL, Kandel ER. Fyn tyrosine kinase is required for normal amygdala kindling. *Epilepsy research* 1995; **22**(2)**:** 107-114.

75. Endele S, Rosenberger G, Geider K, Popp B, Tamer C, Stefanova I *et al.* Mutations in GRIN2A and GRIN2B encoding regulatory subunits of NMDA receptors cause variable neurodevelopmental phenotypes. *Nature genetics* 2010; **42**(11)**:** 1021-1026.

76. Talkowski ME, Rosenfeld JA, Blumenthal I, Pillalamarri V, Chiang C, Heilbut A *et al.* Sequencing chromosomal abnormalities reveals neurodevelopmental loci that confer risk across diagnostic boundaries. *Cell* 2012; **149**(3)**:** 525-537.

77. Akashi K, Kakizaki T, Kamiya H, Fukaya M, Yamasaki M, Abe M *et al.* NMDA receptor GluN2B (GluR epsilon 2/NR2B) subunit is crucial for channel function, postsynaptic macromolecular organization, and actin cytoskeleton at hippocampal CA3 synapses. *The Journal of neuroscience : the official journal of the Society for Neuroscience* 2009; **29**(35)**:** 10869-10882.

78. Tang YP, Shimizu E, Dube GR, Rampon C, Kerchner GA, Zhuo M *et al.* Genetic enhancement of learning and memory in mice. *Nature* 1999; **401**(6748)**:** 63-69.

79. Wang D, Cui Z, Zeng Q, Kuang H, Wang LP, Tsien JZ *et al.* Genetic enhancement of memory and long-term potentiation but not CA1 long-term depression in NR2B transgenic rats. *PloS one* 2009; **4**(10)**:** e7486.

80. Mohrluder J, Schwarten M, Willbold D. Structure and potential function of gamma-aminobutyrate type A receptor-associated protein. *The FEBS journal* 2009; **276**(18)**:** 4989-5005.

81. Mejias R, Adamczyk A, Anggono V, Niranjan T, Thomas GM, Sharma K *et al.* Gain-of-function glutamate receptor interacting protein 1 variants alter GluA2 recycling and surface distribution in patients with autism. *Proceedings of the National Academy of Sciences of the United States of America* 2011; **108**(12)**:** 4920-4925.

82. Guo L, Degenstein L, Dowling J, Yu QC, Wollmann R, Perman B *et al.* Gene targeting of BPAG1: abnormalities in mechanical strength and cell migration in stratified epithelia and neurologic degeneration. *Cell* 1995; **81**(2)**:** 233-243.

83. Sonnenberg A, Liem RK. Plakins in development and disease. *Experimental cell research* 2007; **313**(10)**:** 2189-2203.

84. Edvardson S, Cinnamon Y, Jalas C, Shaag A, Maayan C, Axelrod FB *et al.* Hereditary sensory autonomic neuropathy caused by a mutation in dystonin. *Annals of neurology* 2012; **71**(4)**:** 569-572.

85. O'Donovan MC, Craddock N, Norton N, Williams H, Peirce T, Moskvina V *et al.* Identification of loci associated with schizophrenia by genome-wide association and follow-up. *Nature genetics* 2008; **40**(9)**:** 1053-1055.

86. Hill MJ, Jeffries AR, Dobson RJ, Price J, Bray NJ. Knockdown of the psychosis susceptibility gene ZNF804A alters expression of genes involved in cell adhesion. *Human molecular genetics* 2012; **21**(5)**:** 1018-1024.

Figures:

Figure 1

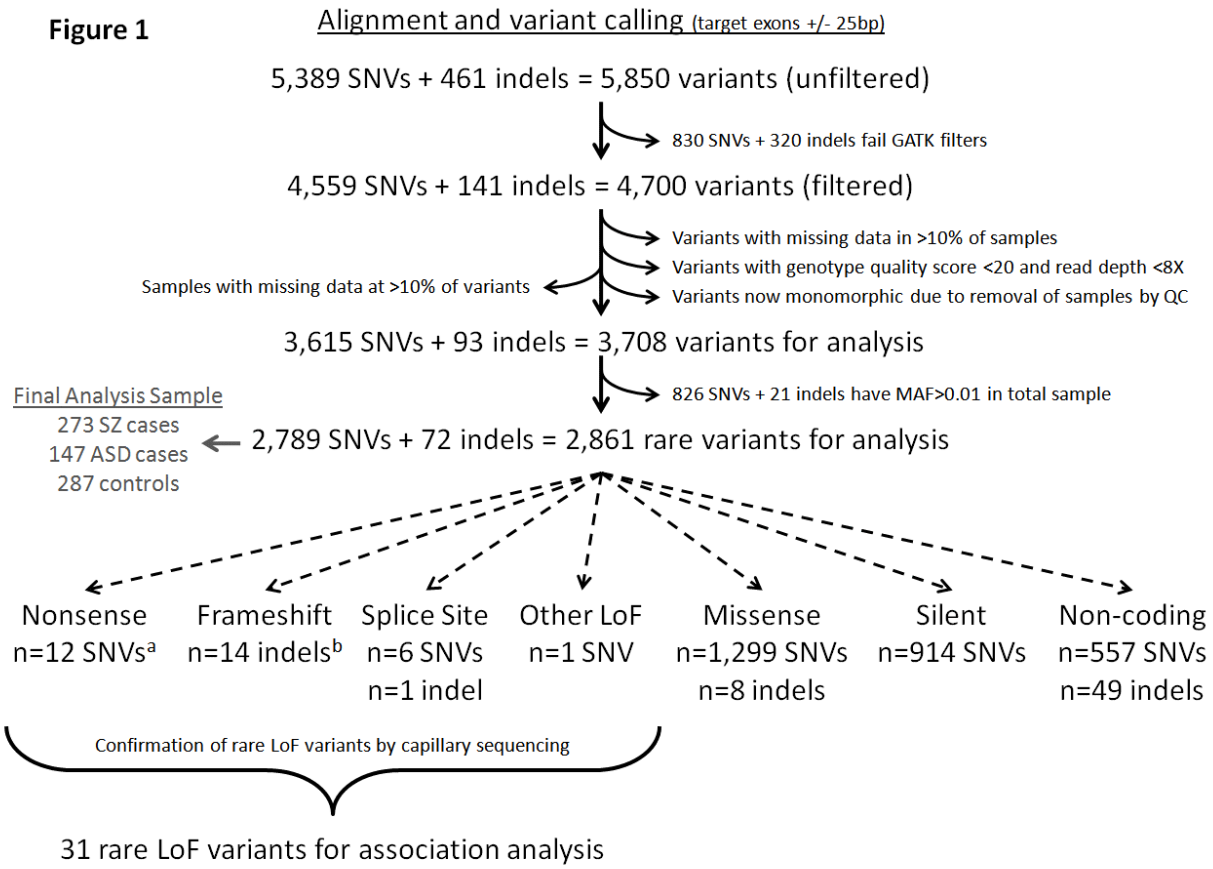


Figure 1: Flowchart displaying the number of variants and processes involved in reducing the total of 5,850 unfiltered variants to a set of 2,861 rare variants for analysis. [a]One nonsense SNV and 1 frameshift indel were called separately but were found to be in the same SZ case and located adjacent to each other in the MACF1 gene. Following confirmation by Sanger sequencing, these two were combined, analyzed and reported as a single frameshift indel in MACF1 (see table 1). Therefore, the total number of rare LoF variants detected was 33. [b]Two LoF indels were not confirmed by Sanger sequencing. The final number of LoF variants for association analysis was 31 (11 nonsense SNVs, 12 frameshift indels, 6 splice site SNVs, 1 splice site indel and 1 stop loss SNV (table 1)).

Figure 2

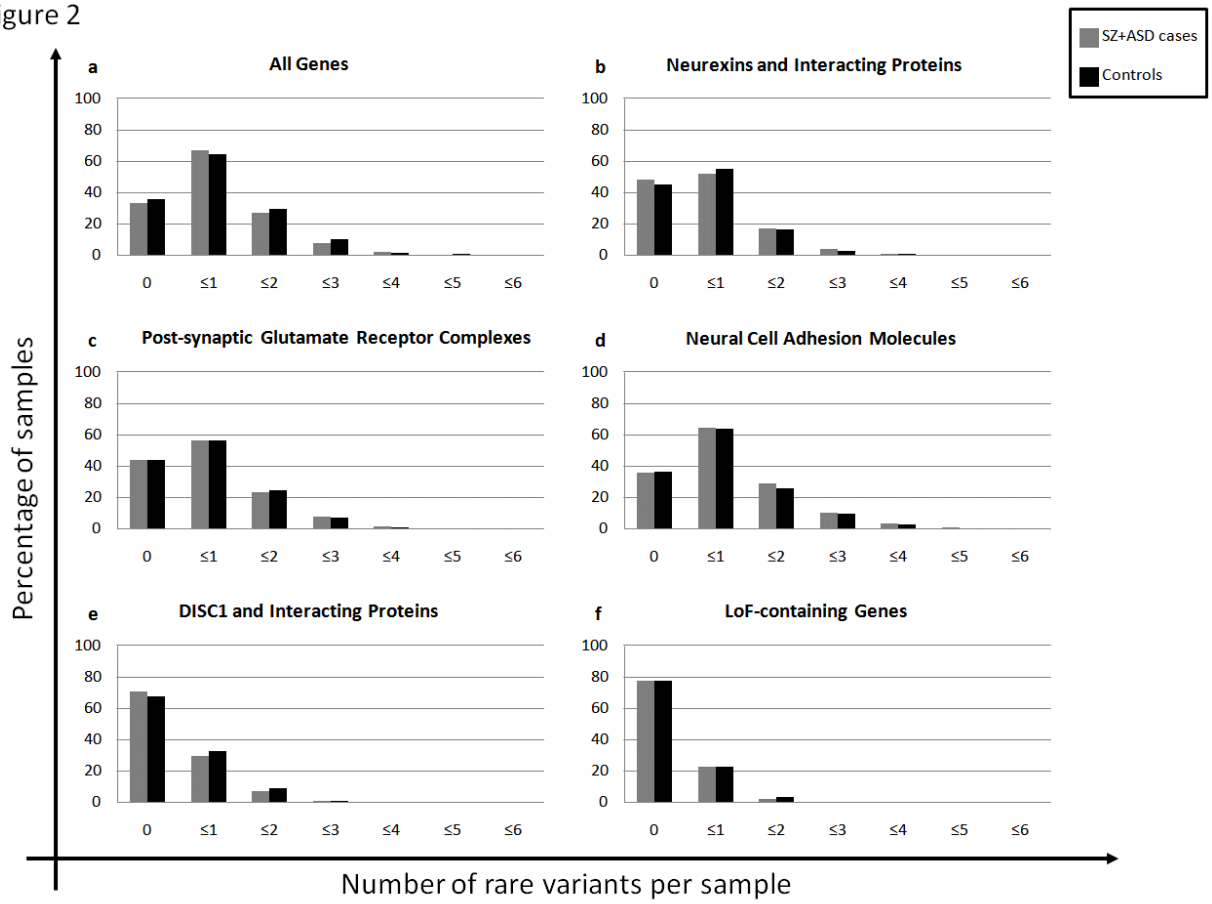


Figure 2: Data on rare functional missense variants (based on PolyPhen2 and SIFT scores) are plotted for the following groups of genes: (a) All Genes, (b) Neurexins and Interacting Proteins, (c) Post-synaptic Glutamate Receptor Complexes, (d) Neural Cell Adhesion Molecules, (e) DISC1 and Interacting Proteins and (f) LoF-containing Genes (n=18 genes that contained a rare LoF variant). For each gene group, we plotted the percentage of cases (SZ and ASD combined) and controls that carried 0, ≥1, ≥2, ≥3 etc. variants. For example, for figure 2a, across all genes 33.1% of cases had 0 rare functional missense SNVs whereas 35.5% of controls had 0 SNVs. Consequently, 66.9% of cases had ≥1 SNV and 64.5% of controls had ≥1 SNV. After that 26.7% of cases had ≥2 SNVs and 29.3% of controls had ≥2 SNVs and so on.

Table 1: Rare LoF variants in SZ, ASD and controls											
Chr	**Position (hg19)**	**Gene[a]**	**Gene Category[b]**	**Ref allele**	**Alt allele**	**Type**	**SZ n=273**	**ASD n=147**	**CON n=287**	**Single-ton?**	**Effect[c]**
1	62,321,741	INADL (1)	1	TC	T	Coding Indel		1		Yes	Frameshift in exon 18 of 43, premature stop 2 codons downstream
1	62,349,979	INADL (2)	1	GC	G	Coding Indel	1			Yes	Frameshift in exon 22 of 43, premature stop 44 codons downstream
1	62,456,007	INADL (3)	1	C	T	Nonsense SNV	1			Yes	R1280* in exon 28 of 43
1	208,216,512	PLXNA2	5	GT	G	Coding Indel	1			Yes	Frameshift in exon 21 of 32, premature stop 37 codons downstream
1	39,788,292	MACF1	1	CAAC	TA	Coding Indel	1			Yes	Frameshift in exon 32 of 102, premature stop 7 codons downstream
2	187,519,413	ITGAV	3	A	AG	Coding Indel	1			Yes	Frameshift in exon 16 of 30, premature stop 7 codons downstream
2	239,257,490	TRAF3IP1	4	G	T	Splice Site SNV		1		Yes	Donor site of exon 11 of 17, premature stop 30 codons downstream
3	57,282,220	APPL1	2	G	T	Splice Site SNV		1		Yes	Acceptor site of exon 10 of 22, exon 10 skipped, transcript continues in frame
4	187,628,509	FAT1	2	C	A	Nonsense SNV	1			Yes	E825* in exon 2 of 27
6	56,358,939	DST (1)	3	TA	T	Coding Indel	1			Yes	Frameshift in exon 83 of 102, immediate premature stop
6	56,472,474	DST (2)	3	G	A	Nonsense SNV	1			Yes	Q2285* in exon 39 of 102
6	56,479,284	DST (3)	3	T	C	Splice Site SNV		1		Yes	Acceptor site of exon 36 of 102, exon 36 skipped, transcript continues in frame
6	56,482,783	DST (4)	3	C	CCT	Splice Site Indel		1		Yes	Donor site of exon 23 of 102, premature stop 23 codons downstream
6	56,483,170	DST (5)	3	C	A	Nonsense SNV	1			Yes	E1888* in exon 23 of 24
6	56,483,389	DST (6)	3	G	A	Nonsense SNV			1	Yes	Q1815* in exon 23 of 24
6	56,507,564	DST (7)	3	TA	T	Coding Indel	2	2	2	No	Frameshift in exon 1 of 84, premature stop 33 codons downstream
6	112,025,283	FYN	1	G	A	Nonsense SNV	1			Yes	R156* in exon 7 of 14
8	27,463,990	CLU	3	CTG	C	Coding Indel		1	1	No	Frameshift in exon 4 of 9, premature stop 4 codons downstream
10	79,584,235	DLG5	1	C	G	Splice Site SNV			1	Yes	Acceptor site of exon 14 of 32, exon 14 skipped, transcript continues in frame

10	79,614,016	DLG5	1	C	A	Nonsense SNV	1			Yes	E217* in exon 4 of 32
12	13,724,778	GRIN2B	2	G	A	Nonsense SNV		1		Yes	Q711* in exon 10 of 13
12	66,765,472	GRIP1 (1)	2	A	T	Splice Site SNV	1			Yes	Donor site of exon 23 of 25, premature stop 25 codons downstream
12	66,923,668	GRIP1 (2)	2	G	A	Nonsense SNV		1		Yes	R149* in exon 5 of 25
13	20,797,556	GJB6	3	TC	T	Coding Indel		1	3[d]	No	Frameshift in exon 5 of 5, premature stop 11 codons downstream
13	109,610,055	MYO16	1	C	T	Nonsense SNV	1			Yes	Q627* in exon 16 of 34
17	40,844,654	CNTNAP1	1	C	T	Nonsense SNV		1		Yes	R890* in exon 17 of 24
X	32,429,867	DMD	4	G	A	Splice site SNV		1		Yes	Donor site of exon 30 of 79, premature stop 24 codon downstream
X	70,367,905	NLGN3	1	TC	T	Coding Indel	1			Yes	Frameshift in exon 2 of 8, premature stop 42 codons downstream
						Total LoF variants	**16**	**13**	**8**		
						Total singleton LoF variants	**14**	**9**	**2**		
Other protein-truncating rare variants located in last 5% of coding sequence and not included in LoF association analysis											
1	232,144,803	DISC1	4	CT	C	Coding Indel		1		Yes	Frameshift in exon 11 of 11, premature stop 24 codons downstream
4	72,433,527	SLC4A4	2	G	GT	Coding Indel			1	Yes	Frameshift in exon 25 of 25, premature stop 2 codons downstream
18	74,728,772	MBP	2	A	G	Stop Loss SNV	1	1[d]		No	Stop codon lost, new stop 16 codons downstream

[a] Numbers in parenthesis after gene names are to identify variants in phenotypic analyses (supplementary figures A-F)

[b] 1=Neurexin and Neuroligin Interacting Proteins, 2=Post-synaptic Glutamate Receptor Complexes, 3=Neural Cell Adhesion Molecules, 4=DISC1 and Interacting Proteins, 5=Positional and Functional Candidates

[c] Position of variant is reported for largest protein-coding transcript containing that variant based on Ensembl

[d] One sample is homozygous for this variant

Table 2: Carrier-based association analysis of rare LoF variants in all genes														
	SZ + ASD (n=420)	**CON (n=287)**	***P* value**	***OR* 95% CI**		**ASD (n=147)**	**CON (n=287)**	***P* value**	**OR 95% CI**		**SZ (n=273)**	**CON (n=287)**	***P* value**	**OR 95% CI**
# of rare LoF variant carriers	29	8	*0.02*	2.59 1.11,6.24		13	8	*0.005*	3.38 1.27,9.17		16	8	*0.07*	2.17 0.86,5.64
# of singleton LoF variant carriers	23	2	*0.0007*	8.26 1.87,51.06		9	2	*0.001*[a]	9.29 1.85,63.14		14	2	*0.002*	7.70 1.65,49.53
[a] Fisher exact test														

Supplementary Information

Excess of rare novel loss-of-function variants in synaptic genes in schizophrenia and autism spectrum disorders

Elaine M. Kenny PhD[1]*, Paul Cormican PhD [1]*, Sarah Furlong BSc[1]*, Eleisa Heron PhD [1], Graham Kenny BSc[1], Ciara Fahey MSc[1], Eric Kelleher MB[1], Sean Ennis PhD [2], Daniela Tropea PhD [1], Richard Anney PhD [1], Aiden P. Corvin PhD [1], Gary Donohoe PhD [1], Louise Gallagher PhD [1], Michael Gill MD [1], Derek W. Morris PhD [1]**.

1 Neuropsychiatric Genetics Research Group, Dept. of Psychiatry and Institute of Molecular Medicine, Trinity College Dublin, Ireland.

2 School of Medicine and Medical Science, University College Dublin, Ireland.

* These authors contributed equally to this work.

** Corresponding author. Mailing address: Room 0.80, Institute of Molecular Medicine, St. James' Hospital, Dublin 8, Ireland. Tel: +353 1 896 2424. Fax: +353 1 896 3405. Email: derek.morris@tcd.ie.

Supplementary Information Contents **Page**

1. Gene Selection

A total of 215 genes were grouped into five categories that define how they were selected for analysis. Some genes could be included in more than one category but are listed based on method of first selection. The full list of gene IDs is detailed in supplementary table A along with the data source that underpinned each selection.

1) Neurexin and Neuroligin Interacting Proteins

Neurexins are a family of neuronal cell adhesion molecules (CAMs) that play a major role in synaptic function. Neurexins are predominantly presynaptic CAMS and bind to the postsynaptic neuroligins (NLGNs; Sudhof 2008; Betancur et al. 2009). Several studies have reported that rare deletions at NRXN1 are risk factors for SZ and ASD (AGP Consortium 2007; Kirov et al, 2008; Kim et al 2008; Marshall et al, 2008; Vrijenhoek et al, 2008; Walsh et al, 2008; Glessner at al, 2009; Rujescu et al, 2009), and rare risk mutations at functionally-related genes have also been reported for ASD (e.g. NRXN2 (Gauthier et al, 2011),NRXN3 (Vaags et al, 2012), NLGN3 and NLGN4X (Jamain et al, 2003), SHANK2 (Berkel et al, 2010), SHANK3 (Durand et al, 2007; Moessner et al, 2007)). Using an extensive literature search and the online interaction databases HPRD (http://www.hprd.org/), String (http://string-db.org/), IntAct (http://www.ebi.ac.uk/intact/), BioGRID (http://thebiogrid.org/) and BOND (http://bond.unleashedinformatics.com/), we compiled a list of Neurexins, Neuroligins and interacting genes (n=46). In supplementary table A we indicate if the gene is a direct interactor of the core gene(s) (e.g. NRXN or NLGN), or if it is a secondary interaction in which case we identify the intermediary gene, which would also have been selected. In each case, we identify the interaction database that is the source of this information and the experimental method used. By using the database and pulling up data on any of the pairs of interacting genes, the database provides a link to the original study responsible for identifying the interaction.

2) Post-synaptic Glutamate Receptor Complexes

Our next functional targets in the synapse were glutamate receptor complexes. Glutamate is the most abundant excitatory neurotransmitter and it plays a critical role in synaptic plasticity and cognitive processes, disruption of which may contribute to neurodevelopmental disorders (Grant et al, 2005). Laumonnier et al. (2007) detail 1,180 proteins in the mouse post-synaptic proteome and categorize them by structure and function. There are three multi-protein glutamate receptor complexes: N-methyl-D-aspartic acid (NMDA), metabotropic glutamate receptor 5 (mGluR5) and α-amino-3-hydroxyl-5-methyl-4-isoxazole-propionate (AMPA). A number of x-linked genes within the NMDA receptor complex are known to be involved in psychiatric disorders (Laumonnier et al. 2007) and recent data on *de novo* CNVs in SZ showed enrichment of de novo events in cases at genes within the NMDA receptor complex (Kirov et al. 2012). We selected genes within these receptor complexes that encode structural proteins: channels, receptors, membrane-associated guanylate

kinases (MAGUKs)/adaptors/scaffolders, cytoskeletal molecules and CAMs (n=58). In supplementary table A we identify which Glutamate Receptor Complex each gene is associated with and its functional classification within that complex.

3) Neural Cell Adhesion Molecules

Given the important role of CAMs in synapse structure and function, and the genetic data implicating neural CAMs in neurodevelopmental disorders from studies of rare variation (e.g. NRXN1) and studies of common variation using pathway methods (O'Dushlaine et al, 2011), we used data from the post-synaptic proteome (Laumonnier et al. 2007) and KEGG pathway annotation (http://www.genome.jp/kegg/pathway.html) to identify additional genes that encode neural CAMS (n=61; see supplementary table A).

4) DISC1 and Interacting Proteins

The identification of a balanced t(1:11) chromosomal translocation in a large Scottish family that segregated with a wide spectrum of psychiatric phenotypes identified DISC1(Disrupted in Schizophrenia 1) (St. Clair et al, 1990; Hennah et al, 2009). Subsequent work has established multiple roles for DISC1 and its 'Interactome' in neurodevelopment and synapse regulation (Camargo et al, 2007; Brandon and Sawa 2011). Early sequencing studies of DISC1 and its interacting proteins have identified rare variation as a risk factor for schizophrenia (Song et al, 2008; Moens et al, 2011). Therefore, using online interaction databases and the extensive DISC1 literature (e.g. Camargo et al. 2007; Chubb et al, 2008; Porteous & Millar 2009), we selected genes that encode known interacting proteins of DISC1 (n=23). In supplementary table A we indicate if the gene is a direct interactor of DISC1, or if it is a secondary interaction in which case we identify the intermediary gene, which would also have been selected. In each case, we identify the interaction database that is the source of this information and the experimental method used. By using the database and pulling up data on any of the pairs of interacting genes, the database provides a link to the original study responsible for identifying the interaction.

5) Positional and Functional Candidates

We finally selected additional candidate genes based on recent GWAS data (e.g. ZNF804A for SZ (O'Donovan et al, 2008), MACROD2 for ASD (Anney et al, 2010)), data from severe neurodevelopmental disorders (e.g. MECP2, Amir et al, 1999) and data from neurodevelopmental biology (e.g. SEMA6A (Runker et al, 2011), n=27). All source papers are identified via their PubMed ID in supplementary table A.

2. Targeted Sequencing, Quality Control and Variant Annotation

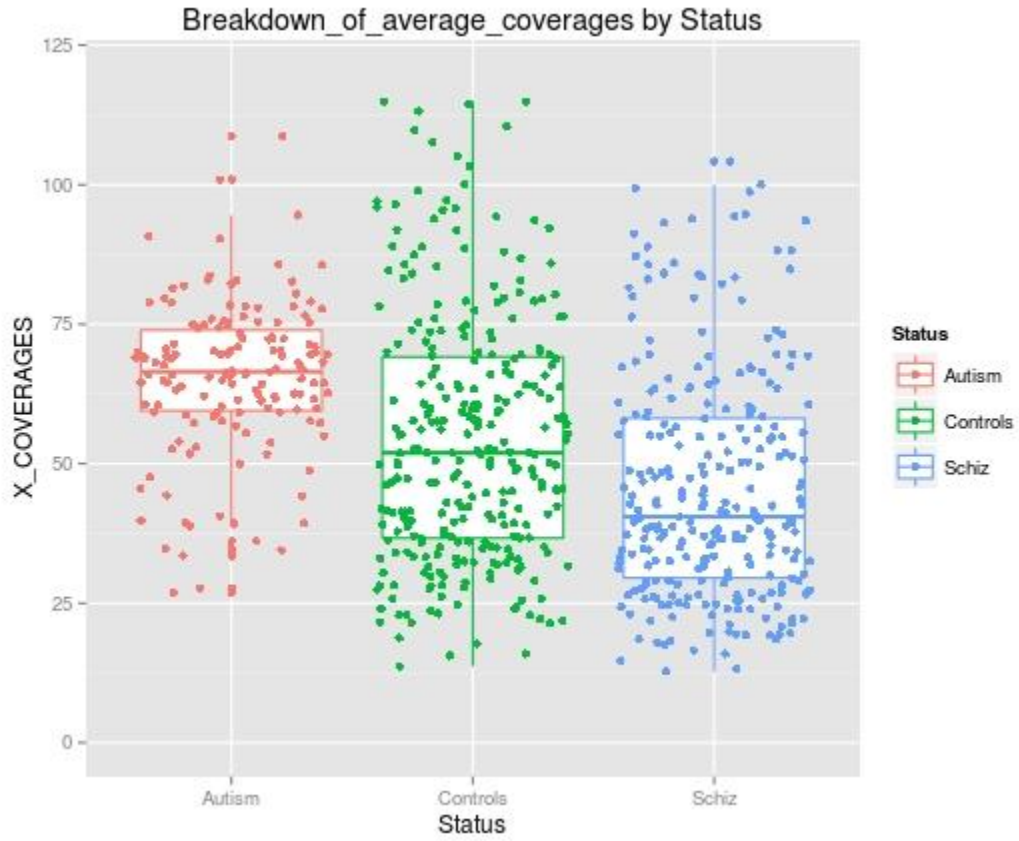
The hg18 exon co-ordinates for all genes were uploaded to the Galaxy browser (http://main.g2.bx.psu.edu/) from tracks available from the UCSC browser for the 215 genes targeted for this study. This final co-ordinate list was uploaded to Agilent Technology's eArray design website to design the SureSelect Target Enrichment array. In order to increase the likelihood that we could target each region with at least two SureSelect 120mer probes, we artificially inflated any target region under 120 bases to 121 bases. Using default settings we successfully generated baits for 3,672 of 3,709 target exons (99%). The total quantity of baited sequence was 1,064,238bp targeting 693,593bp of exonic sequence. Samples were indexed and multiplexed in groups of 24 and sequenced using the multiplex target enrichment method described previously (Kenny et al, 2011), followed by 80bp paired-end sequencing on an Illumina Genome Analyzer II. The median coverage for all samples included in the final analysis was 41x for SZ, 66x for ASD and 52x for controls (supplementary figure A). 95% of samples had sequence coverage ≥8x at 80% of successfully targeted bases (supplementary figure B).

199.5Gb of sequence data was generated for a total sample of 743 individuals. Sequence alignment and calling of both SNVs and indels was performed using GATK (v1.0.5506; DePristo et al, 2011). Post-GATK filtering, the dataset included 4,700 variants. The dataset was further reduced by applying quality control (QC) filters to both samples and variants. Samples that had missing data at >10% of variants were removed. Variants with genotype quality score <20, read depth <8x or had missing data in >10% of samples were removed. The removal of samples also rendered some variants monomorphic. Variants were classified as rare if they had a minor allele frequency (MAF) <0.01 in the combined case-control sample. Frequency was determined across the full sample and not either cases or controls on their own to avoid biases and potential increased type I error (Lemire 2011; Pearson 2011). All variants were called with hg18-aligned data but were lifted over to hg19 co-ordinates and are reported as such below. Affymetrix 6.0 GWAS data including imputation data using 1000 Genomes data (1000 Genomes Project Consortium 2010) was available for 277 of 297 SZ cases and all controls (Strange et al, 2012). Illumina Human 1M-single Infinium BeadChip GWAS data was available for 135 of 152 ASD cases (Anney et al. 2010). GWAS data for variants within our target regions were used for comparison against sequence data. For samples where GWAS data was available, the average matching between GWAS and sequence data variant calls was >99%.Analysis of silent SNVs show an average of 167 per SZ sample (SD=12.6), 168 per ASD sample (SD=12.3) and 167 variants per control sample (SD=12.8). This indicates an even rate of variant detection across all three sample groups.

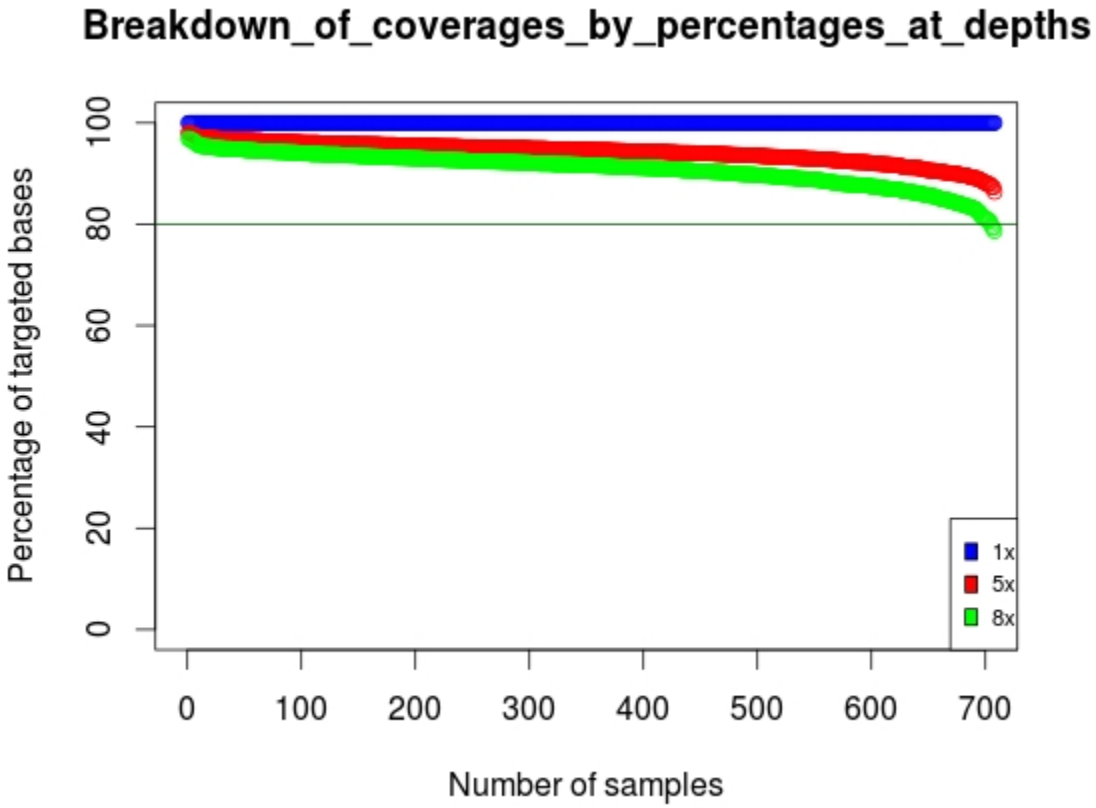
All variants were functionally annotated using SNPeff (v2.0.5; http://snpeff.sourceforge.net/). Loss-of-Function (LoF) variants are predicted to severely disrupt protein-coding sequence and we used the definition of LoF variants as suggested by a recent comprehensive study of LoF variation (MacArthur et al, 2012): nonsense SNVs that introduce stop codons, SNVs that disrupt splice sites and indels that disrupt a transcript's open reading frame or a splice site. We detected one other variant type that we considered as potential LoF; a stop loss SNV that change the predicted stop codon into an amino acid codon and extended the reading frame into the 3'UTR. As per MacArthur et al (2012), we categorized variants as LoF as long as they were located in a proportion of protein-coding transcripts and we did not consider mutations as putative LoF variants in association analyses if they were located in the last 5% of coding sequence. All rare

missense SNVs were uploaded to the Variant Effect Predictor Ensembl database (http://www.ensembl.org/tools.html; McLaren et al, 2010) which assigns a PolyPhen2 (Adzhubei et al, 2010) and SIFT (Kumar et al, 2009) score to each transcript in which the variant resides. Where multiple transcripts contained the same variant, the transcript containing the most deleterious PolyPhen2 score (or SIFT score if PolyPhen2 score was not available) was kept for analysis. We classified variants as "functional missense" if they had a SIFT score that was categorized as deleterious and a PolyPhen2 score that was categorized as deleterious or potentially damaging.

We also implemented an additional analysis of deleterious genetic mutation load across interacting genes. No individual sample carried more than 1 LoF variant, so we focused this analysis on missense variants. Firstly, we stuck with our method of classifying missense variants as functional/deleterious using the PolyPhen2/SIFT scores. We next identified all samples that carried at least 2 functional missense mutations and used this data to construct a list of all pairs of genes that were hit by 2 mutations in the same individual (n=306 pairs of genes in our data). We considered the gene categories to represent networks of interacting genes even though the interactions can be direct or indirect. Each category has a set number of interacting genes, e.g. the "Neurexin and Neuroligin Interacting Proteins" category (n=46 genes) has 1,035 possible pairs of interacting genes. For this category, using the data above, we could count for each pair of interacting genes how many times they were hit by mutations in a case or a control. What we were looking for was a pair of interacting genes that were hit by mutations to a significant excess in cases v. controls (suggesting that a double hit in these two genes increases risk of illness), or a pair of interacting genes that were hit by mutations to a significant excess in controls v. cases (suggesting that a double hit in these two genes has a protective effect). What we found in the Neurexin category and in each of the other categories was that no pair of interacting genes differed significantly between cases and controls in terms of the number of mutation hits. When we considered all 215 genes together, 306 of 23,005 possible pairs of genes were hit. For all gene pairs that were hit, ~98% were only hit in 1 or 2 samples. The gene pair most often hit was ANK3-RELN, which was hit by mutations in 1 case and in 5 controls (p=0.043). This gene pair falls within the other Positional and Functional Candidates category and after checking interaction databases, we note that they are not known to interact. The only other gene pair hit in more than 3 samples was DST-FAT1, which was hit by mutations in 2 cases and in 2 controls (p=1.0).



Supplementary Figure A: Box plots of the sequence coverage for each of the three phenotype groups in the study.The median coverage for all samples included in the final analysis was 41x for SZ, 66x for ASD and 52x for controls.

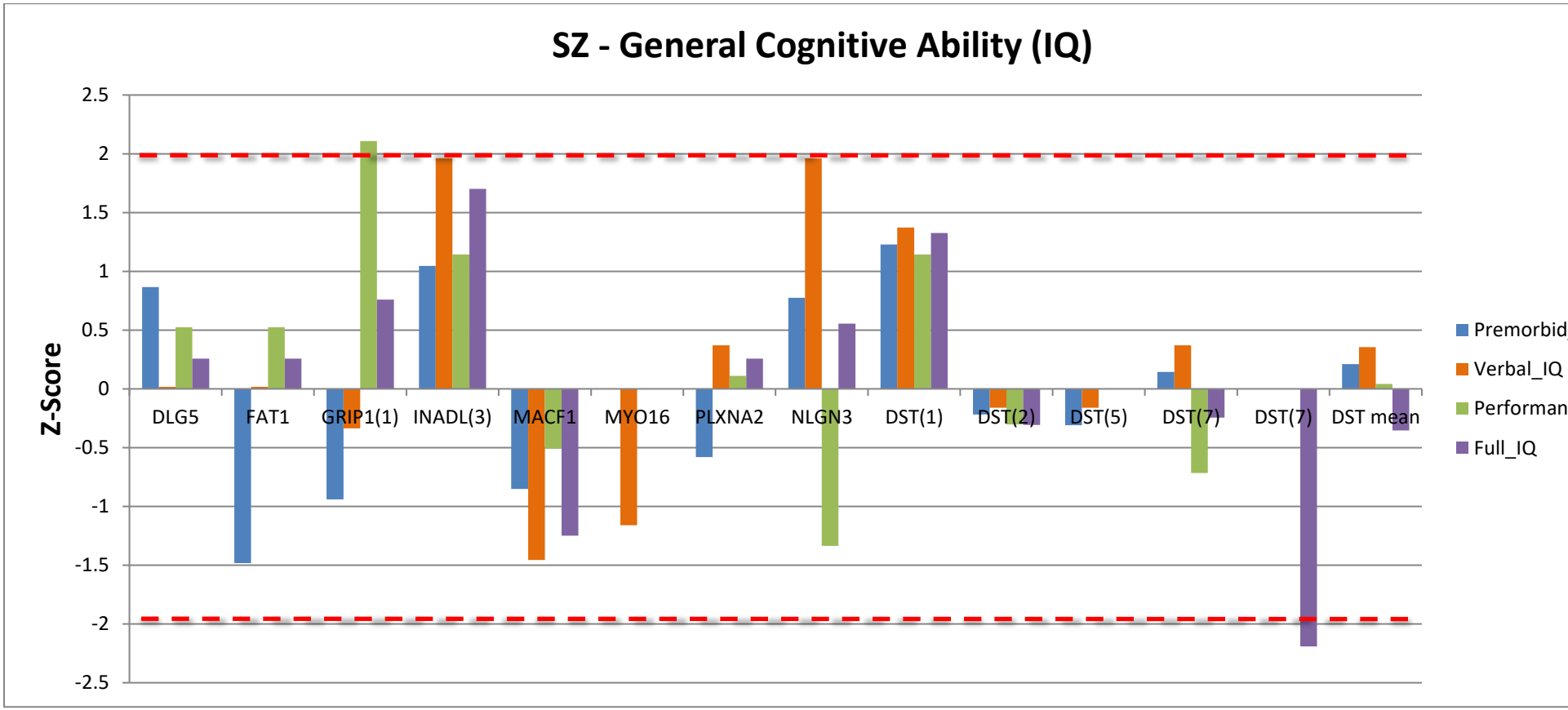


Supplementary Figure B: Plot of sequence coverage for percentage of target bases across all samples in the study. For the final sample of 707 individuals, >95% of samples had sequence coverage ≥8x (green line) at 80% of successfully targeted bases.

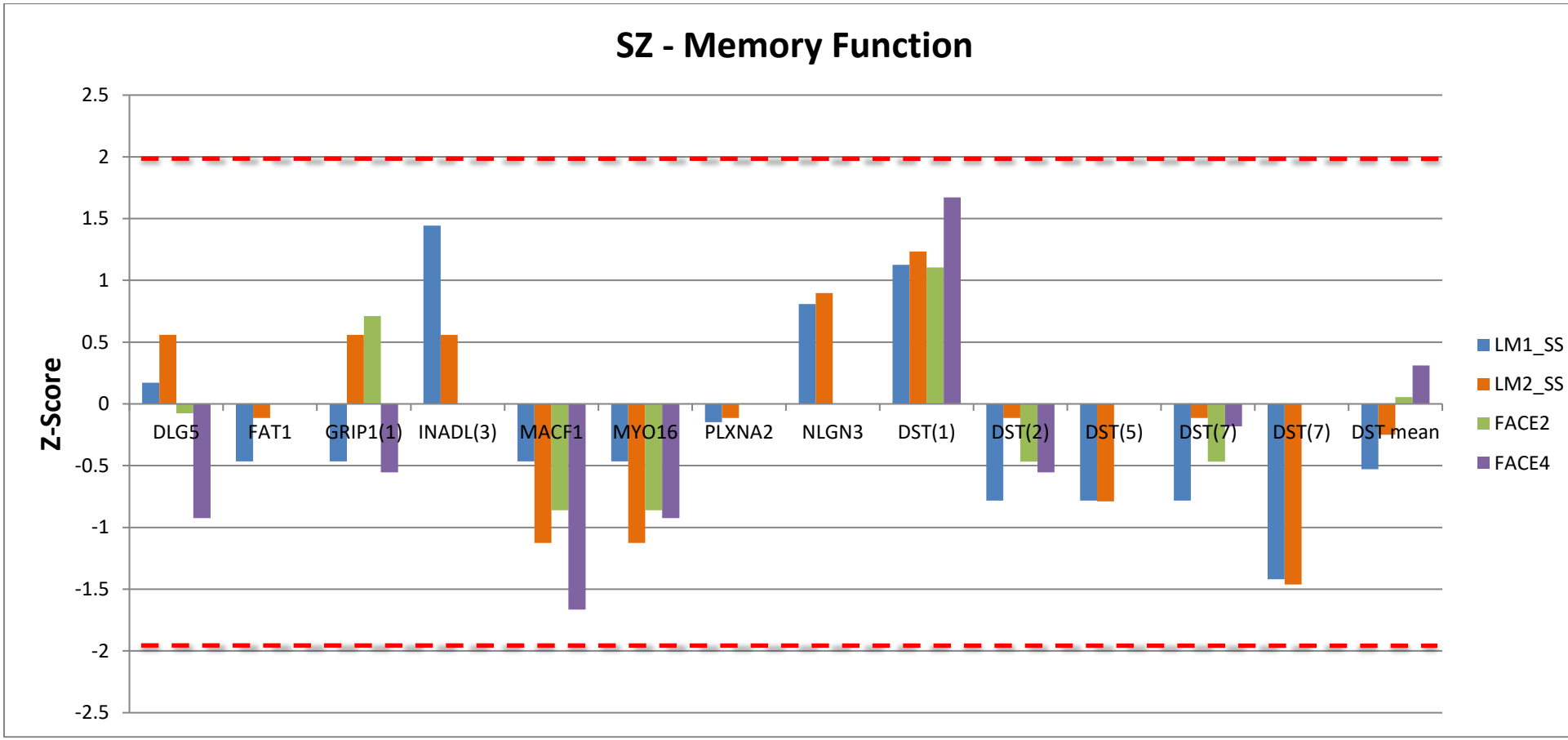
3. Phenotype analysis

Of the final 273 SZ cases, clinical data on symptom severity, collected using the SAPS and the SANS was available for 245 patients. However, due to missing data on individual questions, only data for the hallucinations and delusions questions were judged to have sufficiently complete data (>95%) to allow follow-up analyses for individual variant carriers. Neuropsychological data was available for 188 SZ cases. Briefly, neuropsychological function was assessed in terms of general cognitive ability (using the WTAR and the WAIS-III), episodic memory (using the Logical memory and faces subtest from the Wechsler Memory test-III; WMS-III), working memory (using the WMS-III letter number sequence scale and the CANTAB spatial working memory scale), and attentional control (using the CPT-IP). The phenotype data available for follow-up analyses for ASD variant carriers included the autism symptom severity using the ADOS severity metric, adaptive functioning using the Vineland Adaptive Behavior Composite - Standard Score (ABC-SS) and verbal status as determined by the ADI-R.

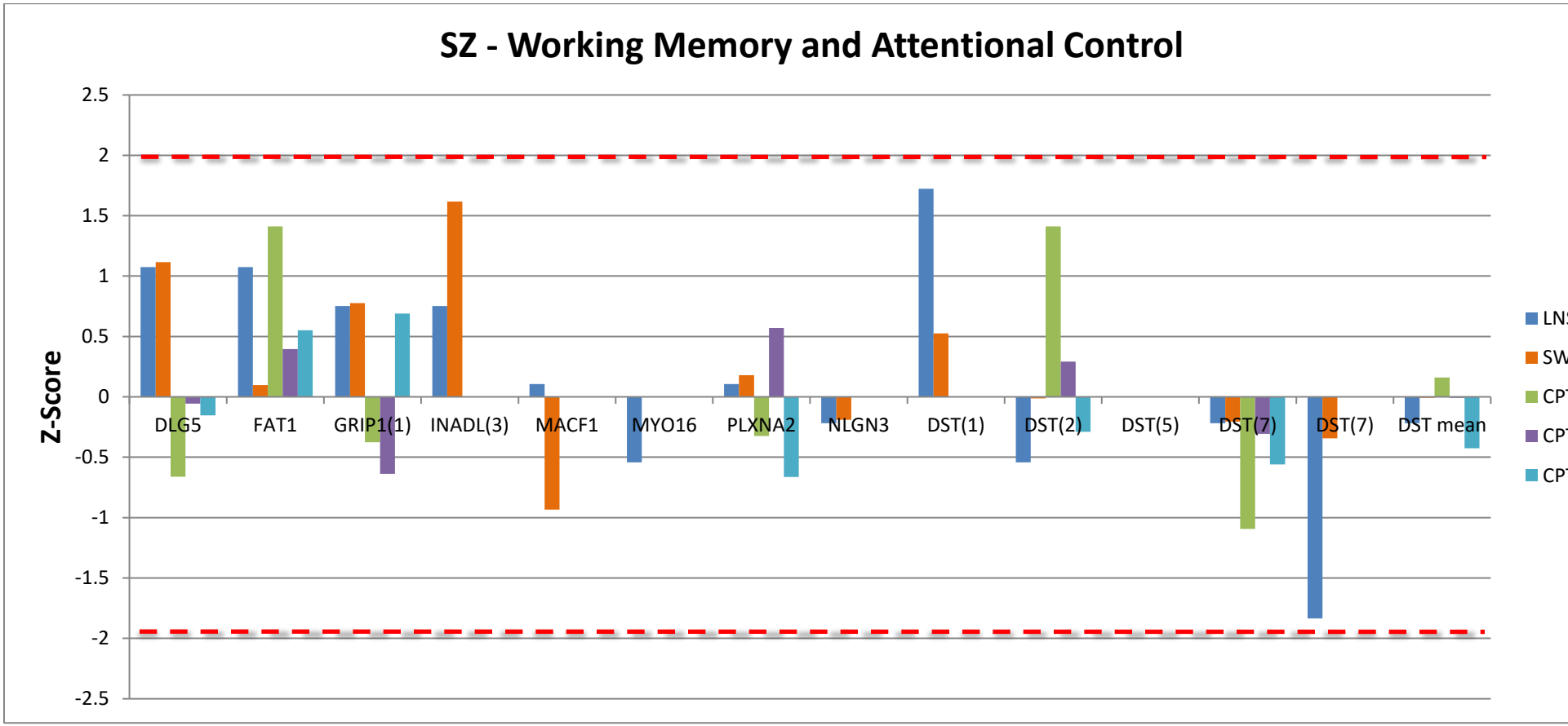
Effects of LOF mutations on cognitive and clinical intermediate phenotypes were assessed separately in SZ and ASD. This was based on a comparison of carriers versus non-carriers within each diagnostic group. In SZ, this analysis was performed for symptom severity, general cognitive ability (IQ), memory function, working memory and attentional control by converting individual scores for all participants to Z-scores and then plotting the Z-scores scores for each individual LoF carrier to show their distance (in standard deviations) from the overall group mean of the non-carrier group (supplementary figures C-F). In ASD, Z-scores were similarly calculated for the ADOS severity metric and the Vineland ABC-SS measure of adaptive functioning (supplementary figures G-H). Because measures of verbal status in ASD cases do not have a normal distribution, these data are presented in a scatter plot (supplementary figure I). Across both diagnostic groups little evidence was found to suggest that the LoF variant carriers differed significantly on clinical and cognitive metrics from non-carriers.



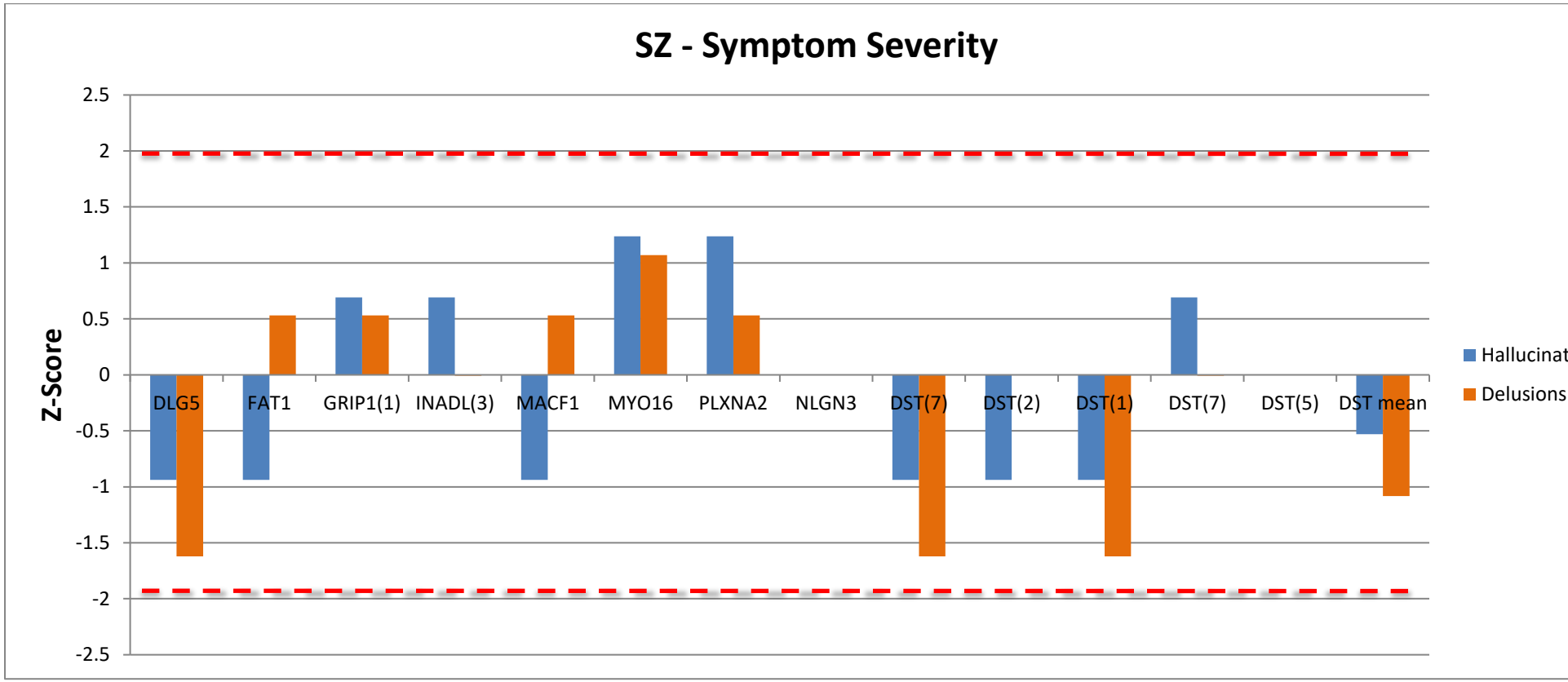
Supplementary Figure C: Z-scores are plotted for four IQ variables for SZ LoF variant carriers. Red dotted lines indicate levels at which Z-scores (+/-1.96) are nominally significant ($p<0.05$). For LoF non-carriers: mean Premorbid IQ = 90.41 (SD = 11.07), mean Verbal IQ = 84.70 (SD = 16.98),mean Performance IQ = 82.39 (SD = 14.52),mean Full IQ = 82.67 (SD = 14.98). See table 1 for information on LoF variants.



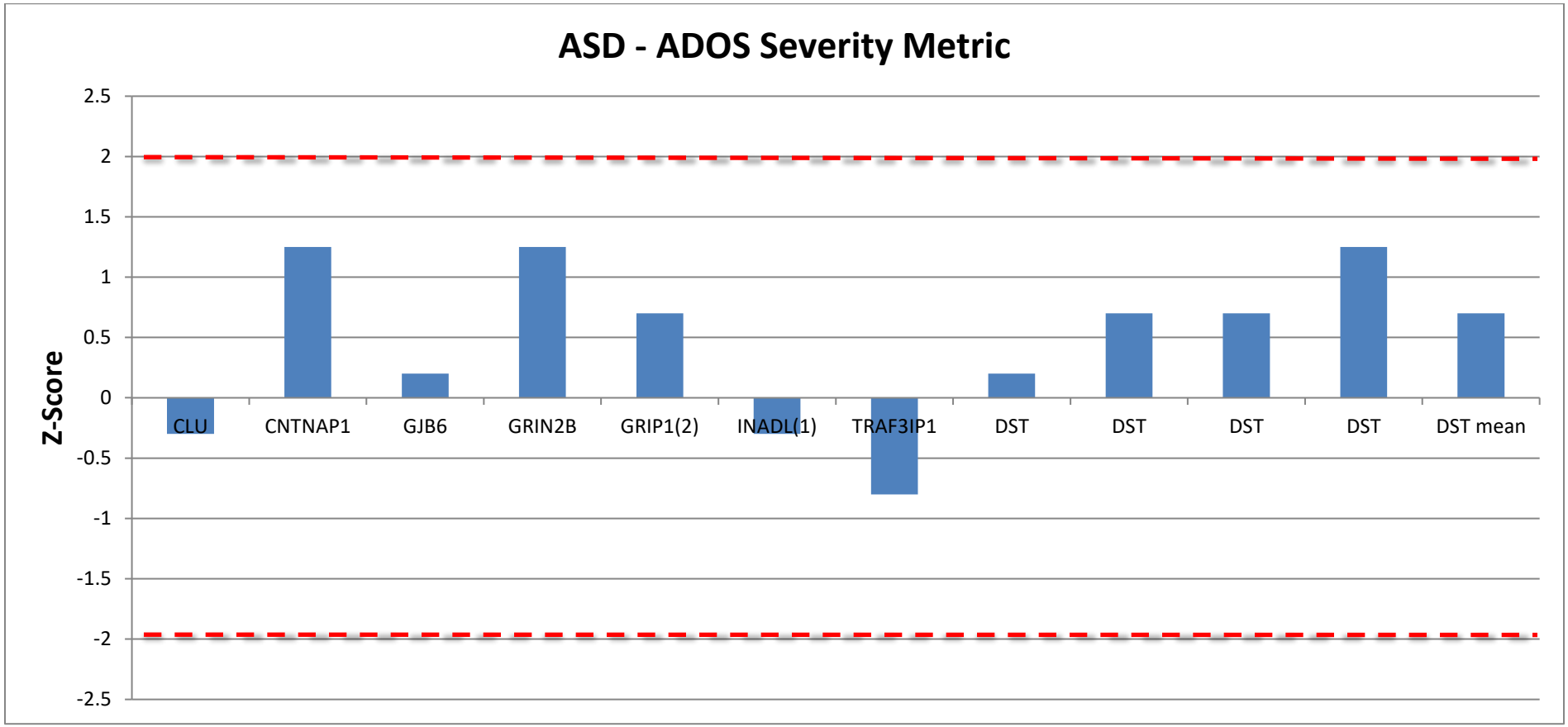
Supplementary Figure D: Z-scores are plotted for four variables for memory function for SZ LoF variant carriers. Red dotted lines indicate levels at which Z-scores (+/- 1.96) are nominally significant ($p<0.05$). For LoF non-carriers: mean logical memory 1 total recall scaled score (LM1_SS) = 5.46 (SD = 3.14), mean LM1_SS = 6.34 (SD = 2.97),mean Faces 2 scaled score = 8.19 (SD = 2.54),mean Faces 4 scaled score = 8.49 (SD = 2.67). See table 1 for information on LoF variants.



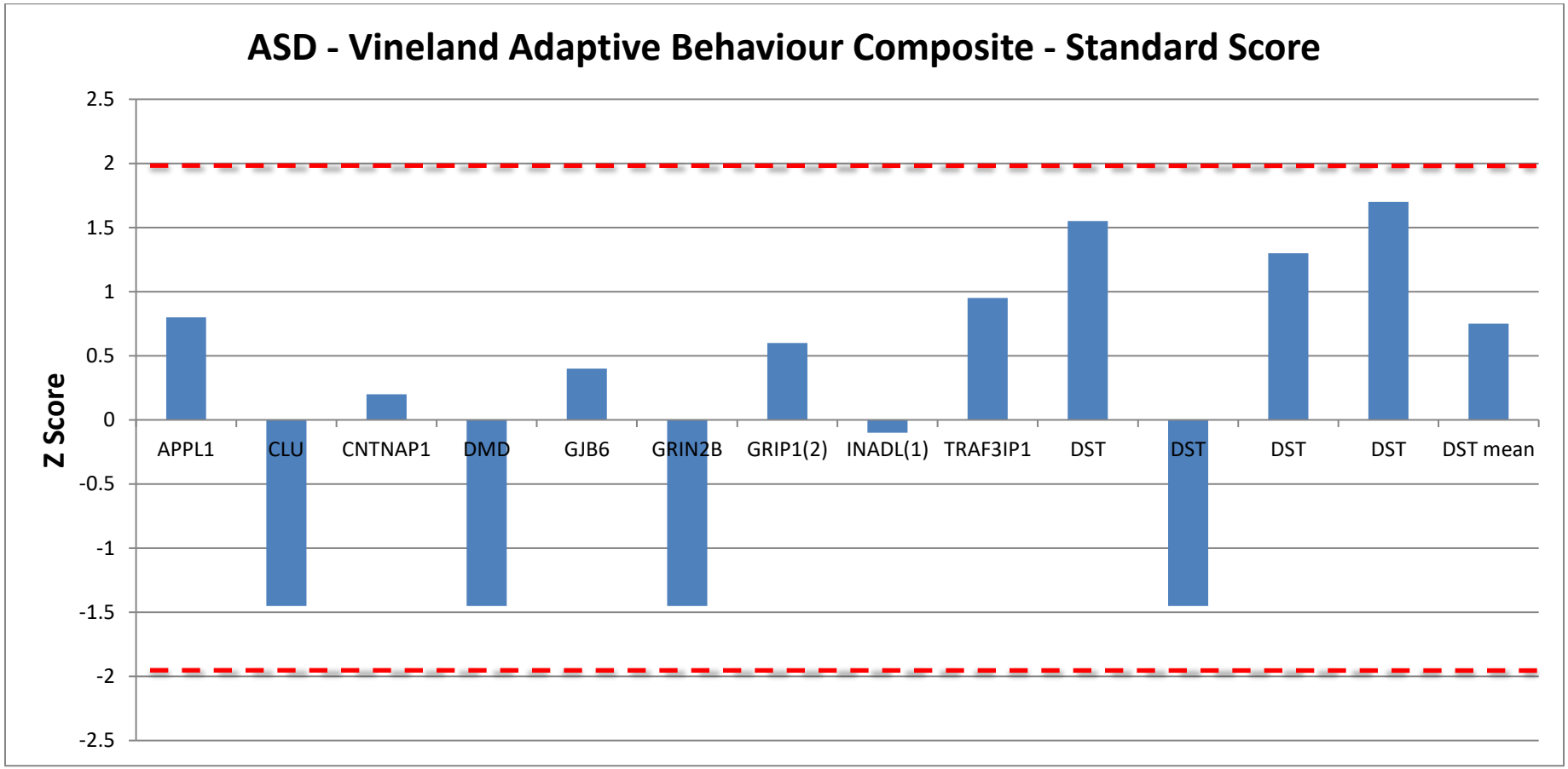
Supplementary Figure E: Z-scores are plotted for five variables for working memory and attentional control for SZ LoF variant carriers. Red dotted lines indicate levels at which Z-scores (+/- 1.96) are nominally significant ($p<0.05$). For LoF non-carriers: mean letter number sequencing scaled score (LNS2) = 6.67 (SD = 3.09), mean between errors standard score (SWM) = -1.21 (SD = 1.36),mean continuous performance test (CPT) 14 score = 2.42 (SD = 1.08),mean CPT28 score = 1.59 (SD = 0.85), mean CPT42 score = 0.79 (SD = 0.67). See table 1 for information on LoF variants.



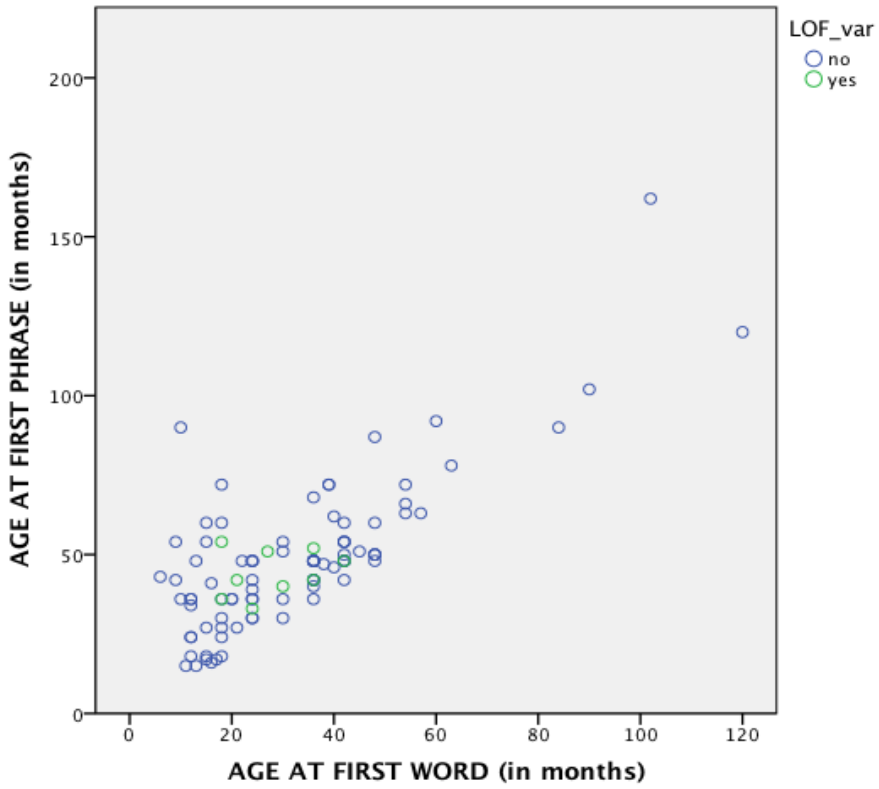
Supplementary Figure F: Z-scores are plotted for symptom severity measures of hallucinations and delusions for SZ LoF variant carriers. Red dotted lines indicate levels at which Z-scores (+/- 1.96) are nominally significant ($p<0.05$). For LoF non-carriers: mean hallucinations score = 2.73 (SD = 1.84), mean delusions score = 4.01 (SD = 1.86). See table 1 for information on LoF variants.



Supplementary Figure G: Z scores are plotted for the ADOS severity metric for ASD LoF variant carriers. Red dotted lines indicate levels at which Z-scores (+/- 1.96) are nominally significant ($p<0.05$). For LoF non-carriers the mean score =7.64 (SD=1.90). See table 1 for information on LoF variants.



Supplementary Figure H: Z scores are plotted for the Vineland Adaptive Behaviour Composite - Standard Score for ASD LoF variant carriers. Red dotted lines indicate levels at which Z-scores (+/- 1.96) are nominally significant ($p<0.05$). For LoF non-carriers the mean score =42.47, (SD=16.33). See table 1 for information on LoF variants.

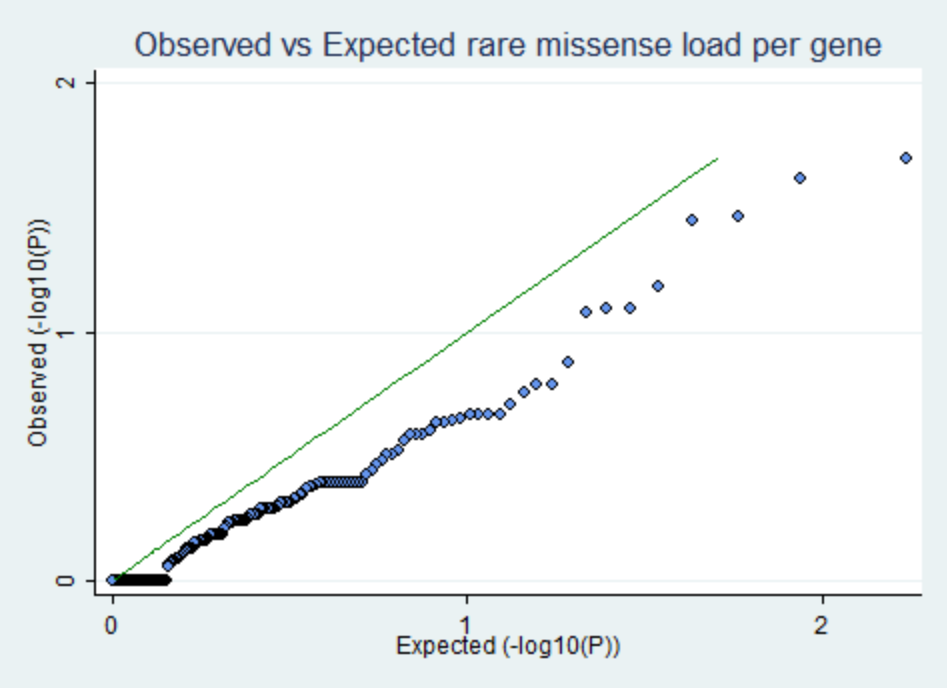


Supplementary Figure I: ADI items 'verbal status', 'age at first words' and 'age at first phrases were used to evaluate verbal status in the ASD cases. Contrary to expectations that LOF carriers might have more severe impairments in language functioning, a larger proportion of non-carriers (32.8%) were non verbal compared with LOF carriers (23%). For cases that had acquired speech there appeared to be no difference between LoF carriers and non-carriers with respect to age of language acquisition.

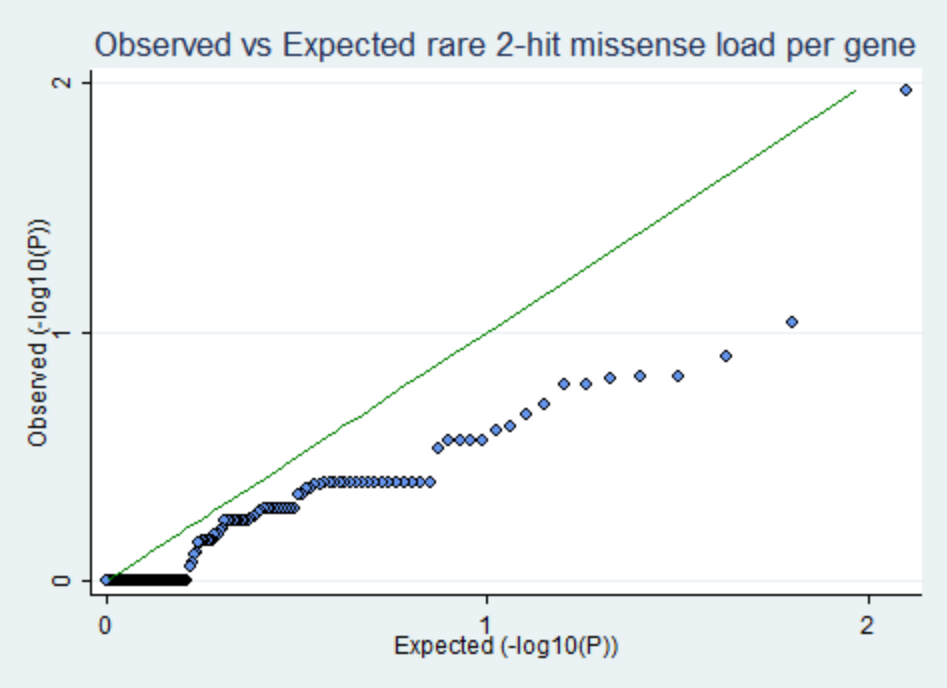
4. Analysis of Missense Variants at Individual Genes

All rare missense SNVs were uploaded to the Variant Effect Predictor Ensembl database (http://www.ensembl.org/tools.html; McLaren et al, 2010) which assigns a PolyPhen2 (Adzhubei et al, 2010) and SIFT (Kumar et al, 2009) score to each transcript in which the variant resides. Where multiple transcripts contained the same variant, the transcript containing the most deleterious PolyPhen2 score (or SIFT score if PolyPhen2 score was not available) was kept for analysis. We classified variants as "functional missense" if they had a SIFT score that was categorized as deleterious and a PolyPhen2 score that was categorized as deleterious or potentially damaging. Of 1,299 rare missense SNVs identified in our sample, 403 were classified as functional based on PolyPhen2/SIFT scores.

We tested for a difference between the combined cases and controls for the number of carriers of (a) and at least 1 rare missense SNV and (b) at least 1 rare functional missense SNV at each individual gene. The Q-Q plots below (supplementary figures J and K) indicate a lesser number of nominally associated genes than would have been expected by chance, most likely reflecting the small number of variants included in the analysis of each gene.



Supplementary Figure J: Q-Q plot of observed and expected p values for carrier-based association analysis of rare missense SNVs (n=1,299) at each individual gene.



Supplementary Figure K: Q-Q plot of observed and expected p values for carrier-based association analysis of rare missense functional SNVs (n=403) at each individual gene. "2-hit" indicates that a missense variant is classified as functional if it had a SIFT score that was categorized as deleterious and a PolyPhen2 score that was categorized as deleterious or potentially damaging.

5. References (with PubMed ID (PMID))

Sudhof 2008; PMID: 18923512

Betancur et al, 2009; PMID: 19541375

Autism Genome Project Consortium 2007; PMID: 17322880

Kirov et al, 2008; PMID: 17989066

Kim et al, 2008; PMID: 18179900

Marshall et al, 2008; PMID: 18252227

Vrijenhoek et al, 2008; PMID: 18940311

Walsh et al, 2008; PMID: 18369103

Glessner at al, 2009; PMID: 19404257

Rujescu et al, 2009; PMID: 18945720

Gauthier et al, 2011; PMID: 21424692

Vaags et al, 2012; PMID: 22209245

Jamain et al, 2003; PMID: 12669065

Berkel et al, 2010; PMID: 20473310

Durand et al, 2007; PMID: 17173049

Moessner et al, 2007; PMID: 17999366

Grant et al, 2005; PMID: 16150739

Laumonnier et al, 2007; PMID: 17236127

Kirov et al, 2012; PMID: 22083728

O'Dushlaine et al, 2011; PMID: 20157312

St. Clair et al, 1990; PMID: 1973210

Hennah et al, 2009; PMID: 18317464

Camargo et al. 2007; PMID: 17043677

Brandon &Sawa 2011; PMID: 22095064

Song et al, 2008; PMID: 18164685

Chubb et al, 2008; PMID: 17912248

Porteous & Miller 2009; PMID: 19778497

O'Donovan et al, 2008; PMID: 18677311

Anney et al, 2010; PMID: 20663923

Amir et al, 1999; PMID: 10508514

Runker et al, 2011; PMID: 22132072

Kenny et al, 2011; PMID: 21163834

DePristo et al, 2011; PMID: 21478889

Lemire 2011; PMID: 21522172

Pearson 2011; PMID: 21522173

1000 Genomes Project Consortium 2010; PMID: 20981092

Strange et al, 2012; PMID: 22883433

MacArthur et al, 2012; PMID: 22344438

McLaren et al, 2010; PMID: 20562413

Adzhubei et al, 2010; PMID: 20354512

Kumar et al, 2009; PMID: 19561590