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

EM Kenny1,3, P Cormican1,3, S Furlong1,2, E Heron1, G Kenny1, C Fahey1, E Kelleher1, S Ennis2, D Tropea1, R Anney1, AP Corvin1, G Donohoe1, L Gallagher1, M Gill1 and DW Morris1

Schizophrenia (SZ) and autism spectrum disorders (ASDs) 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 categorized as: (1) Neurexin and Neulroligin Interacting Proteins, (2) Post-synaptic 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 among 2861 rare variants. We found an excess of LoF variants in the combined cases compared with 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 Neulroligin Interacting Proteins carried an excess of LoF variants in cases compared with controls (P = 0.02). 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.

Molecular Psychiatry (2014) 19, 872–879; doi:10.1038/mp.2013.127; published online 15 October 2013

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

INTRODUCTION

Both schizophrenia (SZ) and autism spectrum disorders (ASDs) are neurodevelopmental in origin and are substantially heritable (h2 > 0.8).2 SZ is characterized by hallucinations, delusions, disordered thinking and cognitive and social deficits. The disorder affects ~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. 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 behavioral 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 functioning9 and further support for biological overlap between the disorders comes from epidemiological10 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, for example, 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 and 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.27–29 Additional evidence that abnormal synapse formation and maintenance is a part of the pathogenesis of both SZ and ASD comes from other CNV studies in SZ32,36,37 and ASD,21,38 single nucleotide polymorphism-based group/pathway analysis in SZ39,40 transcriptomic analysis of the brain in SZ41 and ASD,42 and protein interactome analysis in ASD.43 Where SZ and ASD have been combined for CNV44 analysis, the data support shared biological pathways for the disorders in synaptogenesis and glutamate neurotransmission.

On the basis of 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 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 next-generation sequencing studies in SZ and ASD took the form of whole exome sequencing to identify de novo mutation.45,46 More recently, family-based exome sequencing has begun to provide new support for other strong 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 SZ51–53 and SZ,54 which confirmed the importance of de novo mutation and the paternal age effect, and for ASD identified new risk genes (for example, CHD8, KATNAL2 and SCN2A) and provided new support for other strong candidate genes (for example, GRIN2B). Protein–protein interaction network analysis of genes

1Neuropsychiatric Genetics Research Group, Department of Psychiatry, Institute of Molecular Medicine, Trinity College, Dublin, Ireland and 2School of Medicine and Medical Science, University College Dublin, Dublin, Ireland. Correspondence: Dr DW Morris, Neuropsychiatric Genetics Research Group, Department of Psychiatry, Institute of Molecular Medicine, Room 0.80, St James' Hospital, Dublin 8, Ireland.

E-mail: derek.morris@tcd.ie

3These authors contributed equally to this work.

Received 13 November 2012; revised 2 August 2013; accepted 8 August 2013; published online 15 October 2013
Rare variant analysis of 215 candidate genes in schizophrenia and autism
EM Kenny et al

Molecules,42 (4) DISC1 and Interacting Proteins,4 and (5) other Positional and Functional Candidates. The functional categories were used sequentially to select candidate genes. Therefore, ‘Neurexin and Neurolin Interacting Proteins’ were selected first followed by genes that encoded ‘Post-synaptic Glutamate Receptor Complexes’ that were not already selected for the ‘Neurexin and Neurolin 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 analysis.

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 the Agilent’s SureSelect Target Enrichment system (Agilent Technologies, Santa Clara, CA, USA) (total target = 1 064 238 bp) and sequenced on an Illumina Genome Analyzer II (Illumina, San Diego, CA, USA). Sequence alignment and calling of both SNVs and indels was performed using GATK (v1.0.5506; ref. 62). The median coverage for all samples included in the final analysis was 41 x for SZ, 66 x for ASD and 52 x 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) of <0.01 in the combined case–control sample.5,26,28 The average matching between available genome-wide association studies data and sequence data variant calls was >99%. All variants were functionally annotated using SNP eff (v4.0.2; http://www.snpedia.com/). Analysis of silent SNVs shows an average of 167 per SZ sample (s.d. = 12.6), 168 per ASD sample (s.d. = 12.3) and 167 variants per control sample (s.d. = 12.8), indicating an even rate of variant detection across each sample group. LoF variants are predicted to severely disrupt protein-coding sequence and we used the definition of LoF variants as suggested by MacArthur et al.5: 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 analysis if they were located in the last 5% of coding sequence.55 All rare missense SNVs were assigned a PolyPhen266 and SIFT67 score.

Association analysis

Our primary analysis 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 2 x 2 contingency table. Results for \( \chi^2 \) tests are reported except where indicated that a two-tailed Fisher’s exact test was used because an expected cell count was <5. Where we achieved a nominally significant result (\( P < 0.05 \)), we (1) performed the same carrier-based analysis on SZ and ASD case samples separately to observe the effect in the individual case groups and (2) 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 with 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; 17 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 with controls (29 in 420 cases versus 8 in 287 controls; \( P = 0.02 \)) with the effect stronger for ASD (13 in 147
Alignment and variant calling (target exons +/- 20bp)

<table>
<thead>
<tr>
<th>SNVs</th>
<th>Indels</th>
<th>Total Variants</th>
</tr>
</thead>
<tbody>
<tr>
<td>5,389</td>
<td>461</td>
<td>5,850 (unfiltered)</td>
</tr>
<tr>
<td>4,559</td>
<td>141</td>
<td>4,700 (filtered)</td>
</tr>
<tr>
<td>3,615</td>
<td>93</td>
<td>3,708 for analysis</td>
</tr>
</tbody>
</table>

Final Analysis Sample

<table>
<thead>
<tr>
<th>Variants</th>
<th>Number</th>
</tr>
</thead>
<tbody>
<tr>
<td>Nonsense</td>
<td>12 SNVs</td>
</tr>
<tr>
<td>Frameshift</td>
<td>14 indels</td>
</tr>
<tr>
<td>Splice Site</td>
<td>6 SNVs</td>
</tr>
<tr>
<td>Other LoF</td>
<td>1 SNV</td>
</tr>
<tr>
<td>Missense</td>
<td>1,299 SNVs</td>
</tr>
<tr>
<td>Silent</td>
<td>914 SNVs</td>
</tr>
<tr>
<td>Non-coding</td>
<td>557 SNVs</td>
</tr>
</tbody>
</table>

31 rare LoF variants for association analysis

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. One nonsense single nucleotide variant (SNV) and one frameshift indel were called separately but were found to be in the same schizophrenia (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 loss-of-function (LoF) variants detected was 33. 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).

cases; \( P = 0.005 \) than for SZ (16 in 273 cases; \( P = 0.07 \); Table 2). To focus on variants that may be most deleterious, we dropped three low-frequency variants found in multiple samples that may represent benign variants circulating in the population. All three 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 versus 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).

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 147 cases; \( P = 0.001 \)) and SZ (14 in 273 cases; \( P = 0.002 \); Table 2).

To test whether the evidence for a LoF effect for the DISC1 variant indicated that it was de novo, we determined 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: (1) All Genes, (2) Neurexin and Neuroligin Interacting Proteins, (3) Post-synaptic Glutamate Receptor Complexes, (4) Neural Cell Adhesion Molecules, (5) DISC1 and Interacting Proteins and (6) 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, and so on 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 one rare functional missense SNP in individual disorders. In addition, within gene categories (2)–(5)
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 with controls. The selection of an MAF of <0.1 as a frequency cutoff 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

Above, analysis of interacting gene pairs did not identify any pairs that were hit by mutations at a significantly different rate in cases compared with controls (see Supplementary Information).
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. 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 to organize signaling complexes. It interacts with neurexins and neuroligins and is important for cell polarity, migration and may have a role in neurite extension. Also in the Neurexin category is FYN where we found an LoF variant in an 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. Fyn-mutant mice exhibit blunting of long-term potentiation and impaired spatial learning plus other neurological defects including uncoordinated hippocampal architecture and reduced neural cell adhesion molecule-dependent neurite outgrowth. Studies using Fyn-deficient mice support a role for FYN in the induction of epilepsy. Our data further support FYN as a putative risk gene for SZ and/or epilepsy. Interestingly, 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 genome-wide association studies because of the comorbid epilepsy but highlight the value of taking an inclusive approach when selecting phenotype for rare variant studies.

### Table 2. Carrier-based association analysis of rare LoF variants in all genes

<table>
<thead>
<tr>
<th></th>
<th>SZ + ASD (n = 420)</th>
<th>CON (n = 287)</th>
<th>P-value</th>
<th>OR (95% CI)</th>
<th>ASD (n = 147)</th>
<th>CON (n = 287)</th>
<th>P-value</th>
<th>OR (95% CI)</th>
<th>SZ (n = 273)</th>
<th>CON (n = 287)</th>
<th>P-value</th>
<th>OR (95% CI)</th>
</tr>
</thead>
<tbody>
<tr>
<td># Of rare LoF variant carriers</td>
<td>29</td>
<td>8</td>
<td>0.02</td>
<td>2.59</td>
<td>(1.11, 6.24)</td>
<td>13</td>
<td>8</td>
<td>0.005</td>
<td>3.38</td>
<td>(1.27, 9.17)</td>
<td>16</td>
<td>8</td>
</tr>
<tr>
<td># Of singleton LoF variant carriers</td>
<td>23</td>
<td>2</td>
<td>0.0007</td>
<td>8.26</td>
<td>(1.87, 51.06)</td>
<td>9</td>
<td>2</td>
<td>0.001*</td>
<td>9.29</td>
<td>(1.85, 63.14)</td>
<td>14</td>
<td>2</td>
</tr>
</tbody>
</table>

Abbreviations: ASD, autism spectrum disorder; CI, confidence interval; LoF, loss-of-function; OR, odds ratio; SNV, single nucleotide variant; SZ, schizophrenia.

Fisher’s exact test.
mutations in 468 individuals with MR and/or epilepsy identified four individuals with moderate MR and behavioral anomalies who had de novo GRIN2B mutations; a missense SNV, splice donor SNV, splice acceptor SNV and a 2-bp frameshift deletion. Taftowski et al.76 characterized balanced chromosomal abnormalities in 38 subjects with neurodevelopmental abnormalities and identified a de novo translocation in an ASD case (46,X,Y,inv (12)(p13.1q21.31)dn) that disrupted GRIN2B. GRIN2B encodes the glutamate-binding NR2B subunit of the NMDA receptor and is important for channel function, organization of post-synaptic macromolecular complexes, dendritic spine formation or maintenance and regulation of the actin cytoskeleton.77 Overexpression of the gene in animal models is associated with impaired performance in learning and memory.76,79 GRIN2B mutations in humans may affect brain function and cognition by disturbing the electrophysiological balance of the receptor during neurodevelopment.78

We detected two LoF variants in GRIP1 (1xSZ and 1xASD). GRIP1 is a member of the glutamate receptor interacting protein family and has a role in receptor trafficking, synaptic organization and transmission in glutamatergic and GABAergic synapses.80 A recent study identified five 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 function of 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 autosomal 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 data sets where either multiple samples with rare variants from 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 the 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 single nucleotide polymorphism 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.

CONFLICT OF INTEREST

The authors declare no conflict of interest.

ACKNOWLEDGMENTS

We 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/ IN1.1B1916). 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/FEN327/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. This work was supported by grant funding from the Health Research Board (Ireland).

REFERENCES

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
Rare variant analysis of 215 candidate genes in schizophrenia and autism

EM Kenny et al

Molecular Psychiatry (2014), 872 – 879

2014 Macmillan Publishers Limited
Rare variant analysis of 215 candidate genes in schizophrenia and autism
EM Kenny et al

71 Trepanier CH, Jackson MF, MacDonald JF. Regulation of NMDA receptors by the tyrosine kinase Fyn. FEBS J 2012; 279: 12–19.

77 Akashi K, Kakizaki T, Kamiya H, Fukaya M, Yamasaki M, Abe M et al. NMDA receptor GluN2B (GluRII epsilon 2/NR2B) subunit is crucial for channel function, postsynaptic macromolecular organization, and actin cytoskeleton at hippocampal CA3 synapses. J Neurosci 2009; 29: 10869–10882.

Supplementary Information accompanies the paper on the Molecular Psychiatry website (http://www.nature.com/mp)