

Journal Pre-proof

Ultra-rare missense variants implicated in Utah pedigrees multiply affected with schizophrenia

Cathal Ormond, Niamh M. Ryan, Elizabeth A. Heron, Michael Gill, William Byerley, Aiden Corvin



PII: S2667-1743(23)00010-1

DOI: <https://doi.org/10.1016/j.bpsgos.2023.02.002>

Reference: BPSGOS 201

To appear in: *Biological Psychiatry Global Open Science*

Received Date: 24 October 2022

Revised Date: 31 January 2023

Accepted Date: 3 February 2023

Please cite this article as: Ormond C., Ryan N.M, Heron E.A, Gill M., Byerley W. & Corvin A., Ultra-rare missense variants implicated in Utah pedigrees multiply affected with schizophrenia, *Biological Psychiatry Global Open Science* (2023), doi: <https://doi.org/10.1016/j.bpsgos.2023.02.002>.

This is a PDF file of an article that has undergone enhancements after acceptance, such as the addition of a cover page and metadata, and formatting for readability, but it is not yet the definitive version of record. This version will undergo additional copyediting, typesetting and review before it is published in its final form, but we are providing this version to give early visibility of the article. Please note that, during the production process, errors may be discovered which could affect the content, and all legal disclaimers that apply to the journal pertain.

© 2023 Published by Elsevier Inc. on behalf of Society of Biological Psychiatry.

Title

Ultra-rare missense variants implicated in Utah pedigrees multiply affected with schizophrenia

Authors

Cathal Ormond¹, Niamh M Ryan¹, Elizabeth A Heron¹, Michael Gill¹, William Byerley,²
Aiden Corvin¹

Affiliations

1: Neuropsychiatric Genetics Research Group, Department of Psychiatry, Trinity College Dublin, Dublin, Ireland.

2: Department of Psychiatry and Behavioral Sciences, University of California, San Francisco, California, USA.

Corresponding Author

Aiden Corvin, Neuropsychiatric Genetics Research Group, Department of Psychiatry, Trinity Centre for Health Sciences, Trinity College Dublin, James' Street, Dublin 8, Ireland; Tel: +353 1 896 2241; E-mail: acorvin@tcd.ie

Running Title

Ultra-rare variants in schizophrenia pedigrees

Keywords

Schizophrenia; pedigree; WGS; co-segregation; ATP2B2; SCHEMA

Abstract

Introduction

Recent work from the Schizophrenia Exome Sequencing Meta-analysis (SCHEMA) consortium has showed a significant enrichment of ultra-rare variants in schizophrenia cases. Family-based studies offer a unique opportunity to evaluate rare variants, since risk in multiplex pedigrees is more likely to be influenced by the same collection of variants than an unrelated cohort.

Methods

Here we examine whole genome sequencing data from 35 individuals across six pedigrees multiply affected by schizophrenia. We applied a rigorous filtering pipeline to search for classes of protein-coding variants that co-segregated with disease status and these we examined for evidence of enrichment in the SCHEMA dataset. Additionally, we applied a family-based consensus approach to call copy number variants and screen against a list of schizophrenia-associated risk variants.

Results

We identified deleterious missense variants in three genes (*ATP2B2*, *SLC25A28*, and *GSK3A*) that co-segregated with disease in three of the pedigrees. In the SCHEMA the gene *ATP2B2* shows highly suggestive evidence for deleterious missense variants in schizophrenia cases ($p=0.000072$). *ATP2B2* is involved in intracellular calcium homeostasis, is expressed in multiple brain tissue types, and is predicted to be intolerant to loss-of-function and missense variants.

Conclusions

We have identified genes that are likely increasing schizophrenia risk in three of the six pedigrees examined; the strongest evidence being for a gene involved in calcium homeostasis. Further work is required to examine other classes of variants which may be contributing to disease burden.

1 Introduction

Schizophrenia is a clinically heterogeneous brain disorder that impacts on how people perceive and understand the world around them. Existing treatments are partially effective in most people, but outcomes are often poor, and patients die 12-15 years before the average population (1). The underlying etiology is poorly understood, but most of the variation in risk between individuals is genetically mediated, involving a spectrum of risk alleles from many common alleles of small effect to rare copy number and coding variants of larger effect. To date more than 270 common variants have been confirmed by the Psychiatric Genomics Consortium (PGC), but collectively these explain only a small fraction of risk and likely represent <10% of all contributory common variants (2). Twelve known copy number variants (CNV) substantially increase individual risk by 2 to 60-fold (3), but are only present in about 2% of patients (4). The causal mechanism underlying most of these risk loci is still to be determined as the associated loci often overlap multiple genes or regulatory features. Identifying rare, coding variants that contribute to risk may be particularly important in understanding the complex molecular network involved and in developing better treatments for schizophrenia.

Recently, the Schizophrenia Exome Meta-Analysis (SCHEMA) consortium reported 10 genes in which the burden of ultra-rare variants (URVs) was significantly higher in schizophrenia cases (5). Their study also showed that schizophrenia cases carry an excess of pathogenic URVs in loss-of-function (LoF) intolerant genes compared to controls, even when these ten genes were removed. Highly deleterious missense variants had as strong a signal as protein-truncating variants (PTV). These results suggest that many more genes in which URVs contribute to schizophrenia risk are yet

to be discovered. Family-based studies offer a unique opportunity to identify and evaluate rare variants conferring risk for schizophrenia, since densely affected pedigrees are more likely to be influenced by the same subset of variants compared to an unrelated cohort. Ultra-rare, family-private variants have been shown to be enriched in multiplex compared to simplex pedigrees in autism, indicating that they are good candidates for novel gene discovery (6). Historically, linkage analysis has been used to identify candidate causal genes or regions in pedigrees, but this usually requires data from many individuals across several generations. Instead, a cosegregation analysis can be substituted to analyze variants present in appropriate individuals within each family. Here, we examine whole genome sequencing (WGS) data from six multiplex pedigrees affected by schizophrenia to identify URVs likely contributing to the phenotypic risk.

2 Methods and Materials

2.1 Sample procurement and assessment

Methods used in sample ascertainment and assessment were approved by the University of Utah Institutional Review Board. Multiplex pedigrees were identified by screening hospitalized patients with diagnoses of schizophrenia. Following written informed consent subjects were interviewed by a clinician using the Schedule for Affective Disorders and Schizophrenia-Lifetime Version (“SADS-L”) (7). Medical records were obtained for any individual who received psychiatric care. The interview results and any medical records were then presented to a diagnostic panel comprising two clinicians who played no role in ascertainment or assessment. Consensual diagnoses were made using Research Diagnostic Criteria (“RDC”) (8). Six pedigrees of predicted European ancestry in which schizophrenia was the dominant phenotype were selected from the cohort (Supplementary Figure 1).

2.2 Whole genome sequencing

Batch 1

Twenty-six samples across the six pedigrees had been selected previously for WGS. DNA concentrations were quantified by Qubit and the quality of DNA was determined by agarose gel electrophoresis. WGS was performed by MedGenome, Inc. on an Illumina HiSeqX to an average depth of coverage of at least 30x per sample. Some of these samples had been aligned to the GRCh38 reference genome using the Sentieon Genomics proprietary pipeline but were re-aligned locally to match additional data from these pedigrees.

Batch 2

Twelve additional samples were selected for WGS. DNA concentrations were quantified by Nanodrop and the quality of DNA was determined by agarose gel electrophoresis. WGS was performed by Edinburgh Genomics (Clinical Genomics) on an Illumina HiSeqX to an average depth of coverage of at least 30x per sample.

2.3 Data pre-processing and quality control

All BAM files from Batch 1 were converted to paired-end FASTQ files using the *picard* toolkit so they could be re-aligned using the same pipeline and to the same reference genome as Batch 2. This involved reverting the BAM to remove all mapping information, fixing errors in mate pairs, removing reads without matching pairs and randomizing the reads to remove any potential downstream bias due to the original sort order. All FASTQ files were examined using *FastQC* and *samtools* (9) to identify DNA contamination or degradation. Reads were aligned to the GRCh38 reference genome (GCA_000001405.15, including decoy, HLA and alternative contigs) using *BWA-MEM* (10), following the GATK v3 “Best Practices” (11). This involved marking PCR duplicates with *picard*, base quality score recalibration, local realignment of reads around indels, and variant calling with HaplotypeCaller (GATK version 3.8-0-ge9d806836). All sample were jointly genotyped using the *GenotypeGVCFs* module from GATK using default parameters.

After calling genotypes, variants whose depth of coverage was five standard deviations greater than the average depth of coverage across all sites were removed. Variants were then split by type (SNVs, indels and other), retaining those on the standard 23 pairs of chromosomes. Variant quality score recalibration (VQSR) was

applied to SNVs and indels separately, using truth-sets provided in the GATK resource bundle. A tranche sensitivity threshold of 99.9% for SNVs and 99.0% for indels was used to remove low confidence variant sites. For variants other than SNVs and indels, the following hard filters were applied: $QD < 2.0$; $ReadPosRankSum < -20.0$; $FS > 200.0$; $SOR > 10.0$. Multi-allelic sites were split into multiple bi-allelic sites with *bcftools norm* (12). If any sample had $GQ < 20.0$ or $DP < 10.0$, the genotype for that sample was set to missing (13). Finally, variants marked with a filter and variants with missingness greater than 20% were removed.

The software *peddy* (14) was applied to all samples jointly to perform the following pedigree and data checks: i) relatedness discordance; ii) sex discordance; iii) low median coverage; and iv) ancestry clustering by a principal component analysis (PCA) based on 1000 Genomes Project data (15). To investigate the presence of batch effects and technological stratification, the software *XPAT* (16) was applied to all samples. Given the minimum samples requirements for *XPAT*, additional families recruited from the same cohort and sequenced in the same batches were included.

2.4 SNV and Indel Prioritization

Snpsift v5.0 (17) was used to identify private variants in families to reduce the computation burden of annotation. Here, private indicates that the variant is present in one family and absent from all other families. We considered all individuals with a diagnosis of schizophrenia to be cases, and individuals with no diagnosis as controls. Next, variants were retained if there were no Mendelian violations and they followed either a full co-segregation pattern, (carried by all in-family cases, absent from all in-family controls and absent from all marry-in samples) or a reduced co-segregation

pattern (carried by all but one in-family cases, absent from all in-family controls and absent from all marry-in samples). Custom JavaScript code was added to the *FilterVcf* module from *picard* to identify the case/control status and in-family/marry-in status of samples for the co-segregation filter. We removed variants not present in the coding sequence of a protein-coding gene, as defined by the RefSeq ncbiRefSeqCurated table (18), downloaded from the UCSC Table Browser (19).

The Variant Effects Predictor (VEP) v97.0 (20) was used to annotate each variant, taking the canonical transcript of that gene. Where variants overlapped multiple genes, we examined the canonical transcript of each gene separately. As part of the annotation, we included the Genome Aggregation Database (gnomAD) v2.1.1 exome allele frequencies (21) and the database of non-synonymous functional prediction (dbNSFP) v4.1 (22) from which several deleteriousness metrics were extracted, namely MPC (23), SIFT v2.2 (24), PolyPhen2 v2.2.2 (25) and CADD v1.6 (26). Gene-based probability of loss-of-function intolerance (pLI) scores (27), missense Z-scores (28) and LOEUF scores (29) calculated from gnomAD allele frequencies were also annotated from dbNSFP. Custom bash code was used to extract the variant-level scores from dbNSFP that corresponded to the appropriate transcript where scores for different transcripts were supplied.

To prioritize variants likely to be implicated in schizophrenia based on the SCHEMA work, we retained those that satisfied the following: (i) ultra-rare in gnomAD (minor allele count ≤ 5 across all samples); (ii) either PTV (i.e. frameshift, stop gain or splice acceptor/donor); or predicted-deleterious missense variants (MPC > 2); and (iii) present in a highly LoF-intolerant gene (pLI > 0.9).

2.5 Copy Number Variant Calling

We used a family-based consensus of four software tools to call CNVs from the BAM files (Supplementary Figure 5). CNVs were called by tools derived from two classes of calling methods: CNVnator (30) and ERDS (31) (read-depth based callers); LUMPY (32) and Manta (33) (paired-end /split-read based callers). A collapsing strategy was applied to raw CNV calls to eliminate multiple calls which represent the same site, similar to that described in Trost et al. (34). Then, sites that overlapped reciprocally by 50% were merged, first considering within-method callsets (e.g. LUMPY vs Manta) and then between the resulting across-method callsets. Finally, sites for which over 50% of their length comprised of Repeat and Low Complexity Regions (RLCR) were removed. RLCRs are defined in Trost et al. as: (i) assembly gaps, (UCSC “gap” table); (ii) segmental duplications (UCSC “genomicSuperDups” table); (iii) the pseudo-autosomal regions of the sex chromosomes (34).

Within each pedigree, we removed CNV calls that were identified by one calling method and were only found in one individual. The resulting calls have the support of either at least two calling methods or multiple individuals in the same pedigree. For variants identified by one tool and present in multiple related samples, if the proportion of reads supporting an event at the breakpoints was low, such calls were removed.

3 Results

3.1 WGS Data

Thirty-eight samples across six pedigrees were sent for sequencing at a minimum average depth of coverage of 30x (Methods, Supplementary Figure 1). Two samples (K1501_9 and K1527_33) failed sequencing due to low quality DNA. Additionally, sample K1524_3 was found to be unrelated to their four siblings and was removed. The remaining 35 samples passed all pedigree-related checks and were carried forward for analysis (Table 1). A PCA confirmed that all samples clustered with European individuals from the 1000 Genomes Project (Supplementary Figure 2). *XPAT* did not reveal any evidence of technological stratification or batch effects across the samples, with only family-based clustering observed (Supplementary Figures S3-4).

3.2 Prioritized SNVs

In total 11,509,434 variants passed all quality control measures following the joint genotyping of all samples, of which 3,069,960 were private to one of the families. After applying the main prioritization filters, no ultra-rare, functionally relevant variants were identified that had full co-segregation with cases (Table 2). However, three deleterious missense URVs (Class II variants from SCHEMA: $2 \leq \text{MPC} < 3$) that followed a reduced co-segregation pattern were identified, one in each of pedigrees K1546, K1494, and K1524 (Table 3; Figure 1).

3.3 Schizophrenia-associated CNVs

We called CNVs in all samples using a consensus of four calling tools to rule out any variants with a known association with schizophrenia (Supplementary Table 1).

Several such variants were initially identified in the cohort, most were excluded when examining the level of support at the breakpoints (Supplementary Table 2). Only one rare variant was retained, a duplication on chromosome 16p11.2 in sample K1524_5. This CNV was called by both read-depth based callers and was not observed in any other samples in this pedigree.

Journal Pre-proof

4 Discussion

We examined whole genome sequencing data from 35 individuals across six pedigrees recruited from Utah that were multiply affected by schizophrenia. In the absence of sufficient number of individuals to perform linkage analysis we performed a co-segregation analysis to identify variants which are likely to increase disease burden. Following recent work from the SCHEMA consortium, we investigated the presence of ultra-rare, deleterious variants in LoF-intolerant genes. While no fully co-segregating pathogenic URVs were found, we did observe three missense variants with a reduced co-segregation pattern in three families. All three variants were predicted to be deleterious by additional pathogenicity metrics (Table 2). None of the three genes survived FDR correction in the reported SCHEMA analysis, but there was a suggestive excess of the same class of missense variants at *ATP2B2* in the schizophrenia cases compared to controls in the SCHEMA dataset (Table 4). Additionally, the variant in *ATP2B2* identified here was observed only in one schizophrenia sample in the SCHEMA dataset. *ATP2B2* has the highest missense Z-score of the three genes, indicating that it is the most intolerant to missense variants.

ATP2B2 (ATPase plasma membrane Ca²⁺ transporting 2) is a member of the plasma membrane Ca²⁺ ATPase (PMCA) protein family which is involved in intracellular calcium homeostasis (18). It is found to be expressed in multiple brain tissue types in the GTEx project (35). In a GWAS meta-analysis of autism spectrum disorder (ASD) and schizophrenia, an intronic variant in this gene (rs9879311) was found to be genome-wide significant (36). Additionally, damaging *de novo* variants in *ATP2B2* have been shown to be significantly enriched in ASD cases compared to unaffected siblings in a Japanese cohort (37). A protein-protein interaction analysis of genes

implicated in schizophrenia from both rare variants (CNVs and *de novo* SNVs) and common SNPs pointed to N-methyl-D-aspartate receptor (NMDAR) genes as having significant combined effects between rare and common variants (38). *ATP2B2* was found to be connected to the core members of this NMDAR interactome. A paralog of this gene is *ATP2A2*, which is a member of the sarco/endoplasmic reticulum CA²⁺ ATPase (SERCA) protein family. Variants in this gene cause Darier's disease, which is known to increase risk for schizophrenia and bipolar disorder (39). Fine-mapping of the significant loci from the PGC schizophrenia phase 3 GWAS identified an intronic variant of *ATP2A2* as highly probable of being causal (2).

SLC25A28 ("Solute carrier Family 25 Member 28") is part of the mitochondrial carrier sub-group of the SLC gene family. It is a mitochondrial iron transporter that mediates iron uptake, and is expressed in most tissue types, including several brain tissues (35). There is no previous evidence of association between *SLC25A28* and schizophrenia or related disorders. *GSK3A* ("glycogen synthase kinase-3 α ") is one of the two isoforms of the GSK-3 protein kinase and is expressed in multiple brain tissues (35). Lithium, used to treat bipolar disorder, inhibits the activity of the paralog of this gene (*GSK3B*) (40). The variant in this gene was also present in the SCHEMA analysis, where one allele was observed in an individual with schizophrenia. Only one individual across the six families was found to carry a rare, schizophrenia associated CNV: a duplication on chromosome 16p11.2 was called in sample K1524_5. This individual was the only sample in the pedigree not to carry the URV in *SLC25A28*.

This study represents an exploratory analysis of these pedigrees to screen for classes of gene-affecting variants that are known to be associated with schizophrenia. Full co-

segregation would provide the strongest evidence for candidate causal variants, but within each pedigree the inheritance patterns of the three SNVs are consistent with a variant inherited from a common ancestor of most of the affected individuals. The majority of the unaffected individuals from these three pedigrees were last contacted after the typical age of onset of schizophrenia, so are likely to have remained as such (Supplementary Table 3). It is worth noting that the number of samples sequenced in each pedigree is modest, so additional evidence could be gained by the availability of other affected and unaffected family members. Further work is needed to examine other fully co-segregating variants, as well as variants outside the protein-coding regions. We employed strict cut-off thresholds for our filtering and acknowledge that other analytical strategies may identify plausible candidate causal variants missed by the filtering approach.

The methodology implemented here is robust and follows practices established by large-scale genomics consortia, but the SNVs and CNV identified need to be validated with additional sequencing methods. Given how rare these variants are in the population, large samples sizes are required in the discovery analysis to achieve statistical significance, and future work by the SCHEMA consortium may implicate *ATP2B2* as a schizophrenia risk gene.

5 Acknowledgements

The authors acknowledge the support of the Trinity Centre for High Performance Computing (ResearchIT). This work as part of the Psychiatric Genomics Consortium was supported by the National Institutes of Health [5U01MH 109499-04 to A.C], and Science Foundation Ireland [16/SPP/3324 to A.C].

Journal Pre-proof

6 Disclosures

The authors report no biomedical financial interests or potential conflicts of interest.

Journal Pre-proof

7 References

1. Tiihonen J, Lonnqvist J, Wahlbeck K, Klaukka T, Niskanen L, Tanskanen A, et al. (2009): 11-year follow-up of mortality in patients with schizophrenia: a population-based cohort study (FIN11 study). *Lancet*. 374:620-627.
2. Ripke S, Walters JT, O'Donovan MC (2020): Mapping genomic loci prioritises genes and implicates synaptic biology in schizophrenia. *medRxiv.2020.2009.2012.20192922*.
3. Rees E, Kirov G (2021): Copy number variation and neuropsychiatric illness. *Current opinion in genetics & development*. 68:57-63.
4. Kirov G (2015): CNVs in neuropsychiatric disorders. *Human molecular genetics*. 24:R45-49.
5. Singh T, Poterba T, Curtis D, Akil H, Al Eissa M, Barchas JD, et al. (2020): Exome sequencing identifies rare coding variants in 10 genes which confer substantial risk for schizophrenia. *medRxiv.2020.2009.2018.20192815*.
6. Wilfert AB, Turner TN, Murali SC, Hsieh P, Sulovari A, Wang T, et al. (2021): Recent ultra-rare inherited variants implicate new autism candidate risk genes. *Nature genetics*. 53:1125-1134.
7. Endicott J, Spitzer RL (1978): A diagnostic interview: the schedule for affective disorders and schizophrenia. *Archives of general psychiatry*. 35:837-844.
8. Spitzer RL, Endicott J, Robins E (1978): Research diagnostic criteria: rationale and reliability. *Archives of general psychiatry*. 35:773-782.
9. Li H, Handsaker B, Wysoker A, Fennell T, Ruan J, Homer N, et al. (2009): The Sequence Alignment/Map format and SAMtools. *Bioinformatics*. 25:2078-2079.
10. Li H (2013): Aligning sequence reads, clone sequences and assembly contigs with BWA-MEM.
11. Van der Auwera GA, Carneiro MO, Hartl C, Poplin R, Del Angel G, Levy-Moonshine A, et al. (2013): From FastQ data to high confidence variant calls: the Genome Analysis Toolkit best practices pipeline. *Current protocols in bioinformatics*. 43:11.10.11-33.
12. Danecek P, Bonfield JK, Liddle J, Marshall J, Ohan V, Pollard MO, et al. (2021): Twelve years of SAMtools and BCFtools. *Gigascience*. 10.
13. Vadgama N, Pittman A, Simpson M, Nirmalanathan N, Murray R, Yoshikawa T, et al. (2019): De novo single-nucleotide and copy number variation in discordant monozygotic twins reveals disease-related genes. *European journal of human genetics : EJHG*.
14. Pedersen BS, Quinlan AR (2017): Who's Who? Detecting and Resolving Sample Anomalies in Human DNA Sequencing Studies with Peddy. *American journal of human genetics*. 100:406-413.
15. Auton A, Brooks LD, Durbin RM, Garrison EP, Kang HM, Korbel JO, et al. (2015): A global reference for human genetic variation. *Nature*. 526:68-74.
16. Yu Y, Hu H, Bohlender RJ, Hu F, Chen JS, Holt C, et al. (2018): XPAT: a toolkit to conduct cross-platform association studies with heterogeneous sequencing datasets. *Nucleic acids research*. 46:e32.
17. Cingolani P, Patel VM, Coon M, Nguyen T, Land SJ, Ruden DM, et al. (2012): Using *Drosophila melanogaster* as a Model for Genotoxic Chemical Mutational Studies with a New Program, SnpSift. *Front Genet*. 3:35.

18. O'Leary NA, Wright MW, Brister JR, Ciufu S, Haddad D, McVeigh R, et al. (2016): Reference sequence (RefSeq) database at NCBI: current status, taxonomic expansion, and functional annotation. *Nucleic acids research*. 44:D733-745.
19. Haeussler M, Zweig AS, Tyner C, Speir ML, Rosenbloom KR, Raney BJ, et al. (2019): The UCSC Genome Browser database: 2019 update. *Nucleic acids research*. 47:D853-d858.
20. McLaren W, Gil L, Hunt SE, Riat HS, Ritchie GR, Thormann A, et al. (2016): The Ensembl Variant Effect Predictor. *Genome biology*. 17:122.
21. Karczewski KJ, Francioli LC, Tiao G, Cummings BB, Alföldi J, Wang Q, et al. (2019): Variation across 141,456 human exomes and genomes reveals the spectrum of loss-of-function intolerance across human protein-coding genes. *bioRxiv*.531210.
22. Liu X, Li C, Mou C, Dong Y, Tu Y (2020): dbNSFP v4: a comprehensive database of transcript-specific functional predictions and annotations for human nonsynonymous and splice-site SNVs. *Genome medicine*. 12:103.
23. Samocha KE, Kosmicki JA, Karczewski KJ, O'Donnell-Luria AH, Pierce-Hoffman E, MacArthur DG, et al. (2017): Regional missense constraint improves variant deleteriousness prediction. *bioRxiv*.148353.
24. Vaser R, Adusumalli S, Leng SN, Sikic M, Ng PC (2016): SIFT missense predictions for genomes. *Nat Protoc*. 11:1-9.
25. Adzhubei I, Jordan DM, Sunyaev SR (2013): Predicting functional effect of human missense mutations using PolyPhen-2. *Curr Protoc Hum Genet*. Chapter 7:Unit7.20.
26. Rentzsch P, Schubach M, Shendure J, Kircher M (2021): CADD-Splice-improving genome-wide variant effect prediction using deep learning-derived splice scores. *Genome medicine*. 13:31.
27. Lek M, Karczewski KJ, Minikel EV, Samocha KE, Banks E, Fennell T, et al. (2016): Analysis of protein-coding genetic variation in 60,706 humans. *Nature*. 536:285-291.
28. Samocha KE, Robinson EB, Sanders SJ, Stevens C, Sabo A, McGrath LM, et al. (2014): A framework for the interpretation of de novo mutation in human disease. *Nature genetics*. 46:944-950.
29. Karczewski KJ, Francioli LC, Tiao G, Cummings BB, Alföldi J, Wang Q, et al. (2020): The mutational constraint spectrum quantified from variation in 141,456 humans. *Nature*. 581:434-443.
30. Abyzov A, Urban AE, Snyder M, Gerstein M (2011): CNVnator: an approach to discover, genotype, and characterize typical and atypical CNVs from family and population genome sequencing. *Genome research*. 21:974-984.
31. Zhu M, Need AC, Han Y, Ge D, Maia JM, Zhu Q, et al. (2012): Using ERDS to infer copy-number variants in high-coverage genomes. *American journal of human genetics*. 91:408-421.
32. Layer RM, Chiang C, Quinlan AR, Hall IM (2014): LUMPY: a probabilistic framework for structural variant discovery. *Genome biology*. 15:R84.
33. Chen X, Schulz-Trieglaff O, Shaw R, Barnes B, Schlesinger F, Kallberg M, et al. (2016): Manta: rapid detection of structural variants and indels for germline and cancer sequencing applications. *Bioinformatics*. 32:1220-1222.
34. Trost B, Walker S, Wang Z, Thiruvahindrapuram B, MacDonald JR, Sung WWL, et al. (2018): A Comprehensive Workflow for Read Depth-Based Identification of Copy-Number Variation from Whole-Genome Sequence Data. *American journal of human genetics*. 102:142-155.

35. (2013): The Genotype-Tissue Expression (GTEx) project. *Nature genetics*. 45:580-585.
36. (2017): Meta-analysis of GWAS of over 16,000 individuals with autism spectrum disorder highlights a novel locus at 10q24.32 and a significant overlap with schizophrenia. *Mol Autism*. 8:21.
37. Takata A, Miyake N, Tsurusaki Y, Fukai R, Miyatake S, Koshimizu E, et al. (2018): Integrative Analyses of De Novo Mutations Provide Deeper Biological Insights into Autism Spectrum Disorder. *Cell Rep*. 22:734-747.
38. Chang X, Lima LA, Liu Y, Li J, Li Q, Sleiman PMA, et al. (2018): Common and Rare Genetic Risk Factors Converge in Protein Interaction Networks Underlying Schizophrenia. *Front Genet*. 9:434.
39. Cederlöf M, Bergen SE, Långström N, Larsson H, Boman M, Craddock N, et al. (2015): The association between Darier disease, bipolar disorder, and schizophrenia revisited: a population-based family study. *Bipolar disorders*. 17:340-344.
40. Young W (2009): Review of lithium effects on brain and blood. *Cell Transplant*. 18:951-975.

8 Tables

Table 1: The total number of individuals sequenced who were retained for analysis, broken down by diagnosis and whether they were members of the family (In-Family) or marry-in (Marry-In) samples. SCZ: schizophrenia; UN: unaffected.

Pedigree	Total	In-Family		Marry-In
		SCZ	UN	
K1480	5	4	-	1
K1494	5	4	-	1
K1501	7	4	2	1
K1524	4	4	-	-
K1527	6	5	-	1
K1546	8	5	2	1
Total	35	26	4	5

Table 2: The number of variants remaining after each stage of the prioritization process. Counts on the left are for full co-segregation and on the right are for reduced co-segregation. LoF: loss-of-function

Description	Variants	
	<i>Full</i>	<i>Reduced</i>
Quality control filters	11,509,434	
Family-private variants	3,069,960	
Co-segregation pattern	<i>Full</i>	<i>Reduced</i>
	57,253	172,748
In coding sequence	515	1,904
Ultra-rare in gnomAD	50	210
Functional relevance	0	10
LoF intolerant gene	0	3

Table 3: Details of the three prioritized variants with reduced co-segregation. Positions are given on GRCh38. Included are the protein sequence ID (HGVS_p), the minor allele count (MAC) from gnomAD (exome), and several deleteriousness prediction metrics. For SIFT and PolyPhen2, D represents “damaging” and “deleterious” respectively. CADD: combined annotation dependent depletion.

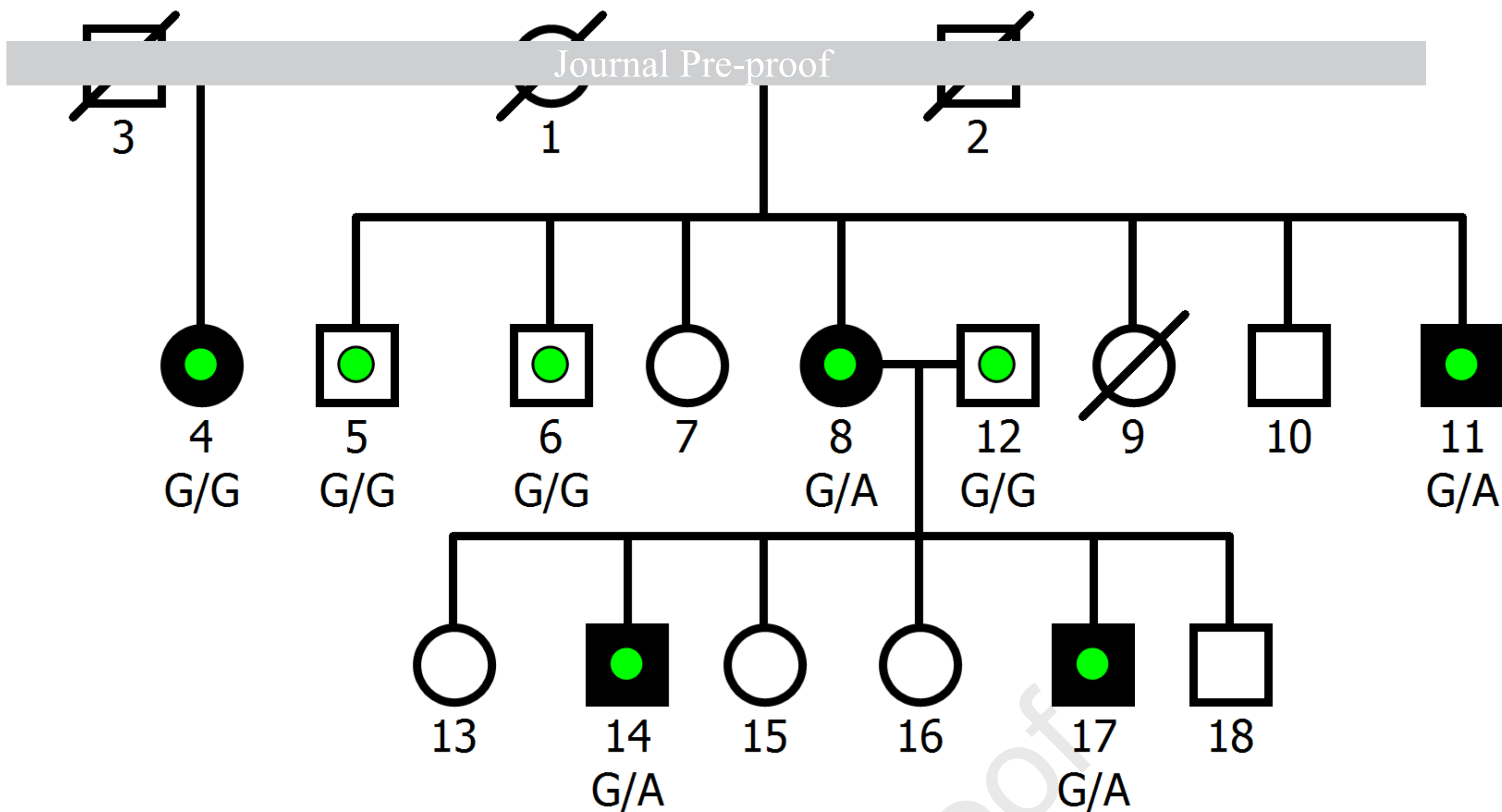
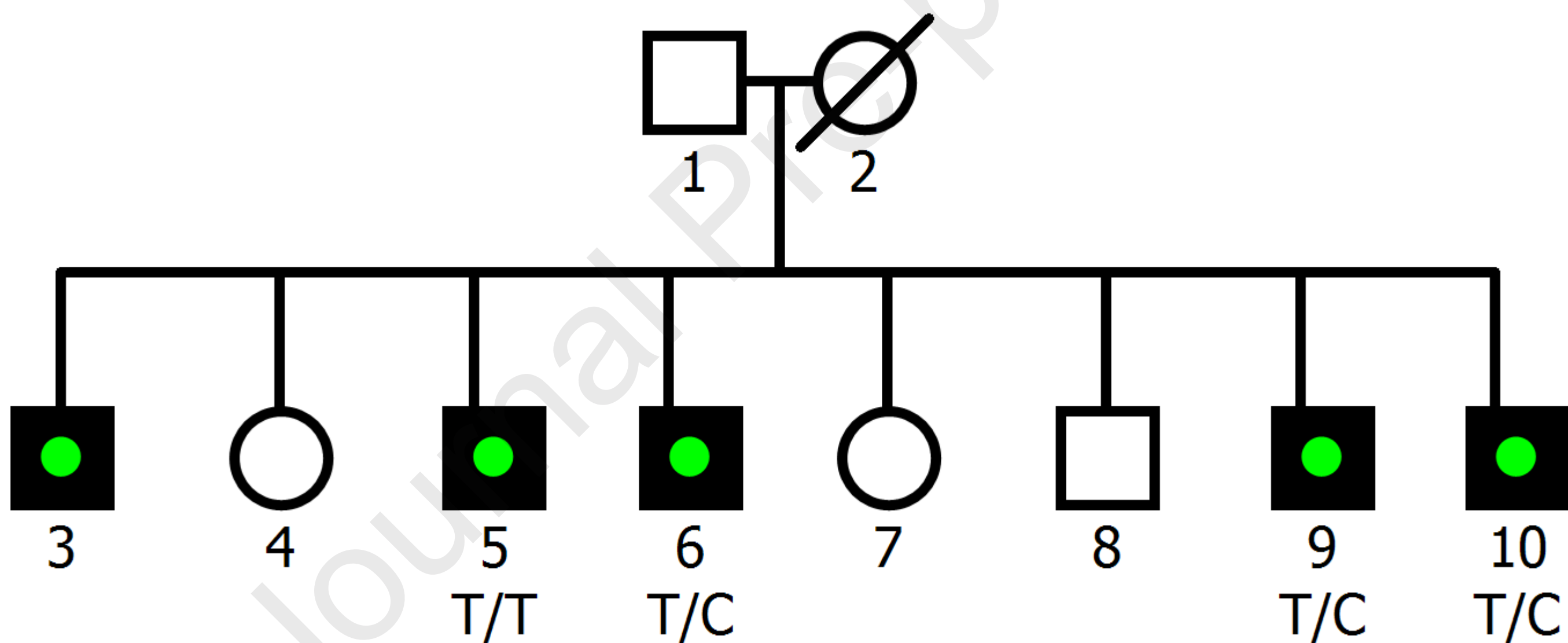
Pedigree	Chr	Position	Variant	Gene	Exon	HGVS_p	MAC	MPC	CADD	SIFT	PolyPhen2
K1546	3	10360021	G>A	<i>ATP2B2</i>	13/23	R588C	1	2.23	31.0	D	D
K1524	10	99610923	T>C	<i>SLC25A28</i>	4/4	I341V	0	2.11	25.6	D	D
K1494	19	42232651	A>G	<i>GSK3A</i>	9/11	I377T	0	2.39	26.9	D	D

Table 4: Gene-level constraint information from gnomAD, including the probability of LoF-intolerant (pLI) score (27), the missense Z-score (28), the LOEUF metric (29), and results (odds ratio (OR), p-value) for Class II variants from the SCHEMA analysis (5).

Gene	Constraint			SCHEMA (Class II)	
	pLI	mis_Z	LOEUF	OR	p-value
<i>ATP2B2</i>	1.00	4.55	0.15	1.920	0.000719
<i>SLC25A28</i>	0.93	2.92	0.37	0.617	0.744000
<i>GSK3A</i>	1.00	3.22	0.13	0.830	0.835000

9 Figure Legends

Figure 1: Pedigree images of the families that harbor an ultra-rare missense variant with reduced co-segregation. Fully shaded boxes denote individuals with a diagnosis of schizophrenia, and sequenced individuals are marked with a colored dot. The genotype of the identified SNV is shown beneath all sequenced individuals that were carried forward for analysis. Additional details for the variants are given in Table 3. (A) K1546 (ATP2B2:p.Arg588Cys); (B) K1524 (SLC25A28:p.Ile341Val); and (C) K1494 (GSK3A:p.Ile377Thr)

A**B****C**