

Rare schizophrenia risk variant burden is conserved in diverse human populations

Dongjing Liu^{1*}, Dara Meyer¹, Brian Fennessy¹, Claudia Feng^{1,2}, Esther Cheng¹, Jessica S. Johnson³, You Jeong Park^{1,4}, Marysia-Kolbe Rieder¹, Steven Ascolillo¹, Agathe de Pins¹, Amanda Dobbyn^{1,4}, Dannielle Lebovitch³, Emily Moya¹, Tan-Hoang Nguyen⁵, Lillian Wilkins¹, Arsalan Hassan⁶, Psychiatric Genomics Consortium Phase 3 Targeted Sequencing of Schizophrenia Study Team, Schizophrenia Exome Meta-analysis Consortium, Katherine E. Burdick^{7,8}, Joseph D. Buxbaum⁴, Enrico Domenici^{9,10}, Sophia Frangou^{4,11}, Annette M. Hartmann¹², Dheeraj Malhotra¹³, Carlos N. Pato¹⁴, Michele T. Pato¹⁴, Kerry Ressler^{8,15}, Panos Roussos^{3,16}, Dan Rujescu¹², Celso Arango^{17,18}, Alessandro Bertolino¹⁹, Giuseppe Blasi¹⁹, Luisella Bocchio-Chiavetto^{20,21}, Dominique Campion^{22,23}, Vaughan Cari^{24,25,26}, Janice M. Fullerton^{24,27}, Massimo Gennarelli^{21,28}, Javier González-Peñas^{17,18}, Douglas F. Levinson²⁹, Bryan Mowry^{30,31}, Vishwajit L. Nimgaokar^{32,33}, Giulio Pergola¹⁹, Antonio Rampino¹⁹, Margarita Rivera-Sanchez^{34,35}, Sibylle G. Schwab^{36,37}, Dieter B. Wildenauer³⁸, Mark Daly^{39,40,41,42}, Benjamin Neale^{39,40,41}, Tarjinder Singh^{39,40}, Michael C. O'Donovan⁴³, Michael J. Owen⁴³, James T. Walters⁴³, Muhammad Ayub^{44,45}, Anil K. Malhotra^{46,47,48}, Todd Lencz^{46,47,48}, Patrick F. Sullivan^{49,50}, Pamela Sklar³, Eli A. Stahl^{4,51,52}, Laura M. Huckins^{3*}, Alexander W. Charney^{1,4*}

1 Genetics and Genomics Department, Icahn School of Medicine at Mount Sinai, New York, NY 10029, USA

2 Wellcome Sanger Institute, Hinxton, Cambridgeshire CB10 1SA, UK

3 Pamela Sklar Division of Psychiatric Genomics, Genetics and Genomics Department, Icahn School of Medicine at Mount Sinai, New York, NY 10029, USA

4 Department of Psychiatry, Icahn School of Medicine at Mount Sinai, New York, NY 10029, USA

5 Virginia Institute for Psychiatric and Behavioral Genetics, Department of Psychiatry, Virginia Commonwealth University, Richmond, VA 23298, USA

6 University of Peshawar, Peshawar 25120, Pakistan

7 Department of Psychiatry, Brigham and Women's Hospital, Boston, MA 02115, USA

8 Department of Psychiatry, Harvard Medical School, Boston, MA 02115, USA

9 Fondazione The Microsoft Research University of Trento, Centre for Computational and Systems Biology (COSBI), Rovereto, Italy

10 Department of Cellular, Computational and Integrative Biology (CIBIO), University of Trento, Povo, Italy

11 Djavad Mowafaghian Centre for Brain Health, University of British Columbia, Vancouver, British Columbia, Canada

12 Department of Psychiatry and Psychotherapy, Medical University of Vienna, Vienna, Austria

13 Neuroscience and Rare Diseases, Roche Pharma Research and Early Development, F. Hoffmann-La Roche Ltd, Grenzacherstrasse 124 4070 Basel, Switzerland

14 Department of Psychiatry and Behavioral Sciences, SUNY Downstate College of Medicine, New York, NY 11203, USA

15 Division of Depression and Anxiety Disorders, McLean Hospital, Belmont, MA 02478, USA

16 Mental Illness Research, Education, and Clinical Center (VISN 2 South), James J. Peters VA Medical Center, Bronx, NY 10468, USA

- 17 Department of Child and Adolescent Psychiatry, Institute of Psychiatry and Mental Health, Hospital General Universitario Gregorio Marañón, Instituto de Investigación Sanitaria Gregorio Marañón (IiSGM), Madrid, Spain.
- 18 Centro de Investigación Biomédica en Red de Salud Mental (CIBERSAM), Madrid, Spain.
- 19 Department of Basic Medical Science, Neuroscience and Sense Organs, University of Bari 'Aldo Moro', Bari, Italy.
- A Department of Neurological and Psychiatric Sciences, University of Bari 'Aldo Moro', Bari, Italy
- 20 Department of Theoretical and Applied Sciences, eCampus University, 22060 Novedrate, Como, Italy
- 21 Genetics Unit, IRCCS Istituto Centro S. Giovanni di Dio Fatebenefratelli, 25125 Brescia, Italy
- 22 Normandie Univ, UNIROUEN, Inserm U1245 and Rouen University Hospital, Department of Genetics and CNR-MAJ, Normandy Center for Genomic and Personalized Medicine, Rouen, Normandie, France
- 23 Department of research, Centre hospitalier du Rouvray, Sotteville-lès-Rouen, Normandie, France
- 24 Neuroscience Research Australia, Sydney, New South Wales, Australia
- 25 School of Psychiatry, University of New South Wales (UNSW), Sydney, NSW, 2052, Australia
- 26 Department of Psychiatry, School of Clinical Sciences, Monash University, Clayton, VIC, Australia
- 27 School of Medical Sciences, University of New South Wales, Sydney, New South Wales, Australia
- 28 Department of Molecular and Translational Medicine, University of Brescia, 25123 Brescia, Italy
- 29 Department of Psychiatry, Stanford University, Stanford, CA 94305, USA
- 30 Queensland Brain Institute, The University of Queensland, Brisbane, QLD, Australia
- 31 Queensland Centre for Mental Health Research, The University of Queensland, Brisbane, QLD, Australia
- 32 Department of Psychiatry, University of Pittsburgh School of Medicine, Western Psychiatric Hospital, Pittsburgh, PA 15213, USA
- 33 Department of Human Genetics, Graduate School of Public Health, University of Pittsburgh, Pittsburgh, PA 15261, USA
- 34 Department of Biochemistry and Molecular Biology II, Faculty of Pharmacy, University of Granada, 18071 Granada, Spain
- 35 Institute of Neurosciences, Biomedical Research Centre (CIBM), University of Granada, 18071 Granada, Spain
- 36 Molecular Horizons, Faculty of Science, Medicine and Health, University of Wollongong, Wollongong, New South Wales, Australia
- 37 Illawarra Health and Medical Research Institute, Wollongong, New South Wales, Australia
- 38 University of Western Australia, Perth, Western Australia, Australia
- 39 Analytic and Translational Genetics Unit, Department of Medicine, Massachusetts General Hospital, Boston, Massachusetts 02114, USA
- 40 Stanley Center for Psychiatric Research, Broad Institute of MIT and Harvard, Cambridge, MA 02142, USA

41 Program in Medical and Population Genetics, Broad Institute of Harvard and MIT, Cambridge, MA 02142, USA

42 Institute for Molecular Medicine Finland, University of Helsinki, Helsinki, Finland

43 MRC Centre for Neuropsychiatric Genetics and Genomics, Division of Psychological Medicine and Clinical Neurosciences, Cardiff University, Cardiff, UK

44 University College London, London, UK.

45 Department of Psychiatry, Queen's University, Kingston, ON, Canada

46 Department of Psychiatry, Zucker School of Medicine at Hofstra/Northwell, Hempstead, NY 11549, USA

47 Center for Psychiatric Neuroscience, Feinstein Institutes for Medical Research, Manhasset, NY 11030, USA

48 Division of Psychiatry Research, The Zucker Hillside Hospital, Northwell Health, Queens, NY 11004, USA.

49 Departments of Genetics and Psychiatry, University of North Carolina, Chapel Hill, NC 27599, USA

50 Department of Medical Epidemiology and Biostatistics, Karolinska Institutet, Stockholm, Sweden

51 Regeneron Pharmaceuticals, Inc., Tarrytown, NY 10591, USA

52 Department of Genetics and Genomic Science and Institute for Multiscale Biology, Icahn School of Medicine at Mount Sinai, New York, NY 10029, USA

*Correspondence:

Dongjing Liu (dongjing.liu@mssm.edu)

Laura M. Huckins (laura.huckins@mssm.edu)

Alexander W. Charney (alexander.charney@icahn.mssm.edu)

Abstract

Schizophrenia is a chronic mental illness that is amongst the most debilitating conditions encountered in medical practice. A recent landmark schizophrenia study of the protein-coding regions of the genome identified a causal role for ten genes and a concentration of rare variant signals in evolutionarily constrained genes¹. This study -- and most other large-scale human genetic studies -- was mainly composed of individuals of European ancestry, and the generalizability of the findings in non-European populations is unclear. To address this gap in knowledge, we designed a custom sequencing panel based on current knowledge of the genetic architecture of schizophrenia and applied it to a new cohort of 22,135 individuals of diverse ancestries. Replicating earlier work, cases carried a significantly higher burden of rare protein-truncating variants among constrained genes (OR=1.48, p-value = 5.4×10^{-6}). In meta-analyses with existing schizophrenia datasets totaling up to 35,828 cases and 107,877 controls, this excess burden was largely consistent across five continental populations. Two genes (*SRRM2* and *AKAP11*) were newly implicated as schizophrenia risk genes, and one gene (*PCLO*) was identified as a shared risk gene for schizophrenia and autism. Overall, our results lend robust support to the rare allelic spectrum of the genetic architecture of schizophrenia being conserved across diverse human populations.

Main

Schizophrenia (SCZ) is a severe, chronic psychiatric illness associated with lifelong progression and early mortality²⁻⁴. The genetic architecture of SCZ has been deeply characterized over the past fifteen years, with clear genetic contribution from common single-nucleotide polymorphisms (SNPs)⁵, large copy number variants (CNVs)⁶, and rare protein-truncating variants (PTVs)^{1,7-14}. Amongst the classes of genetic variation linked to SCZ, rare PTVs provide unique value by linking disease risk to individual genes unambiguously. Most recently, the Schizophrenia Exome Sequencing Meta-Analysis (SCHEMA) consortium increased the sequenced sample size to 24,248 SCZ cases and 97,322 controls, consolidated the enrichment of rare PTVs in SCZ cases across genes under strong evolutionary constraint, and identified ten genes with excess burden of rare disruptive variants in cases compared to controls¹. When considered alongside earlier studies, these results suggest that with greater sample sizes additional genes will be found to harbor excess rare PTVs in SCZ. Whole-exome sequencing (WES) and whole-genome sequencing (WGS) allow a hypothesis-free approach to risk gene discovery. However, applying these methods at the scale required to achieve the power necessary to confidently link genes to disease remains cost prohibitive. Targeted sequencing of genes chosen through data-driven algorithms that take into account prior knowledge of the genic PTV burden provided by studies such as SCHEMA is an alternative approach to rapidly achieve the required sample size for novel risk gene discovery.

The majority of large-scale human genetics research initiatives to date have failed to include diverse populations. Over 80% of genome-wide association studies (GWAS) participants are of European ancestry, despite this group comprising less than one-fourth of the total human

population^{15,16}. Studies of mental illness have contributed to this disparity with almost exclusive European GWAS cohorts despite roughly equal prevalence of psychiatric disorders worldwide¹⁷. The limited evidence from SCZ GWAS and CNV studies of non-European populations suggests broadly shared genetic architecture with that of European populations, but ancestry-specific genetic risk factors are also present¹⁸⁻²³. So far, studies of rare PTVs of complex human traits have been largely consistent across ancestries²⁴⁻³⁰, although no studies have yet shown this for SCZ.

Here, to diversify the population profiled in rare PTV studies of SCZ and achieve the power needed to discover novel risk genes, we designed a custom sequencing panel of 161 putative SCZ risk genes and applied it to case-control cohorts totaling 22,135 individuals from diverse ancestries (40% non-European; Figure 1, Table S1). This study, outlined in Figure 1A and hereafter referred to as the Psychiatric Genomics Consortium Phase 3 Targeted Sequencing of Schizophrenia Study (PGC3SEQ), was limited to cohorts that were not part of earlier SCZ sequencing initiatives such as SCHEMA. A data-driven algorithm was used to construct the sequencing panel that considered all current knowledge of the genetic architecture of SCZ, including a preliminary version of the SCHEMA gene-level burden statistics^{31,32}, with the goal of enriching for genes likely to harbor excess rare PTVs in SCZ but did not reach exome-wide significance in earlier studies due to a lack of power. Specifically, this algorithm was based on gTADA^{33,34}, a Bayesian framework which prioritizes genes by integrating genic rare variant burden statistics (from SCHEMA) with gene membership in gene sets that have implicated in SCZ through a variety of approaches (e.g., GWAS, CNV studies) (Figure 1B, Table S2-S3). The exonic regions of the 161 prioritized genes were sequenced on the Ion Torrent platform and rigorous quality control procedures were carried out subsequently (Supplementary Figure S1-S6). Analyses comparing SCZ to controls were limited to rare PTVs (stop-gain, frameshift indels, or essential splicing donor/acceptor) and deleterious missense (placed into tiers based on the MPC score³⁵ (tier 1: MPC>3; tier 2: MPC 2~3; non-damaging: MPC<2) variants, and synonymous variants were analyzed as a negative control. To maximize power, PGC3SEQ was meta-analyzed with SCHEMA data (available SCHEMA datasets summarized in Table S4 and Figure S7) and sequencing datasets for bipolar disorder and autism. Two broad types of analysis were performed: (1) global enrichment of all constrained genes on the targeted panel (n=80 genes) to demonstrate the overall role of rare disruptive variants in diverse ancestries and (2) gene-level burden test to identify novel SCZ risk genes.

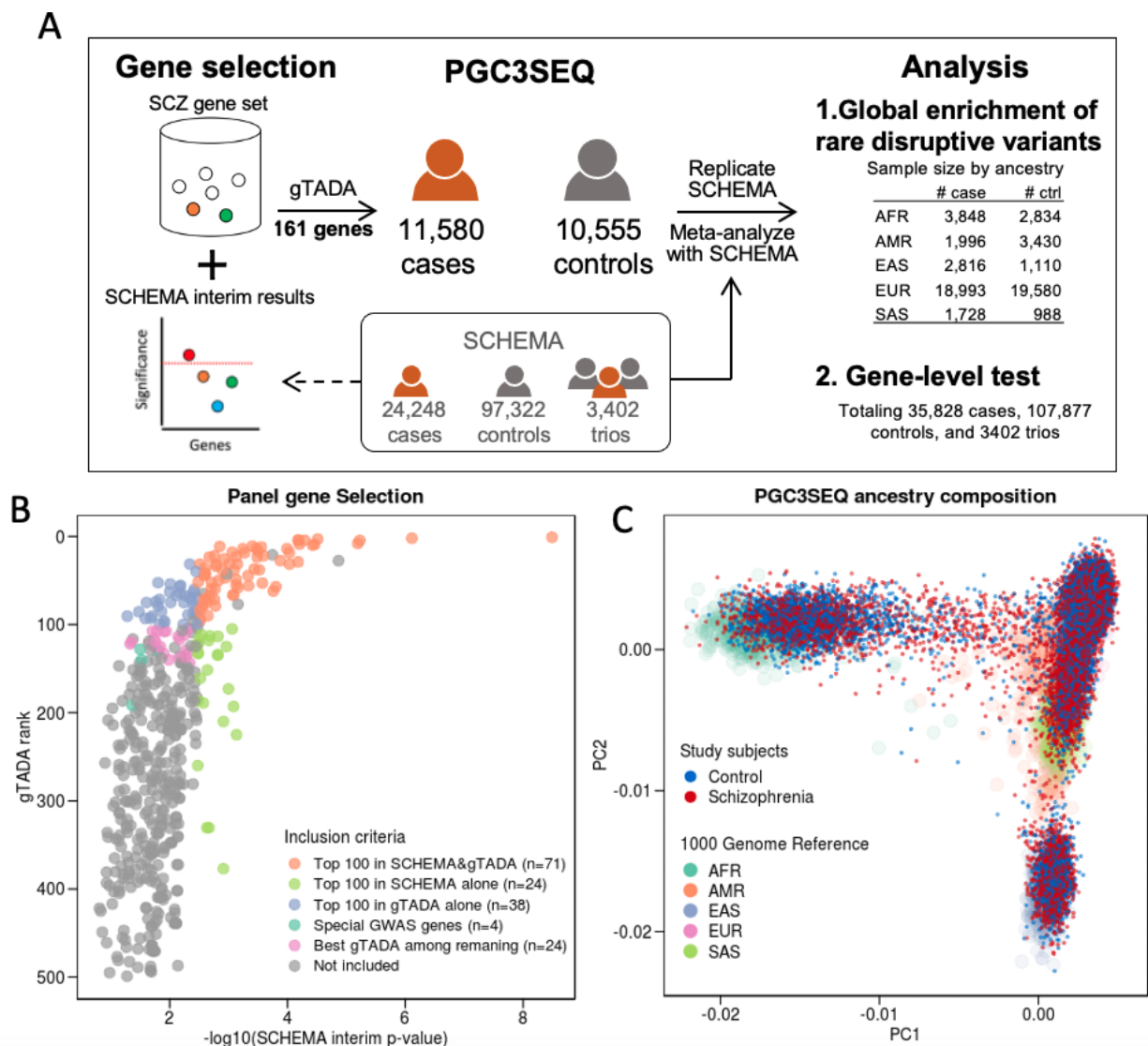


Figure 1. Study design and cohort ancestry composition. **(A)** An overview of the study design. **(B)** Gene selection for the targeted sequencing panel. Genes were selected based on a combination of prior association statistics (SCHEMA), gTADA rankings, and GWAS associations. Specially, we included (1) genes in the top 100 based on the gTADA rank and/or the top 100 based on SCHEMA p-value (Top 100 in SCHEMA&gTADA, Top 100 in SCHEMA alone, Top 100 in gTADA alone, total = 133 genes); (2) genes with evidence for association with SCZ in both GWAS and SCHEMA (special GWAS genes, n=4 genes); (3) an additional 24 genes which had the best 24 gTADA rankings of the remaining genes with a burden p-value < 0.05, to fill up the target panel. X-axis, gene-level p-value using SCHEMA interim data based on which the panel was constructed (different from the final published version); y-axis, gTADA rank of genes. Only the top 500 genes are plotted for a clear display. Some highly ranked genes were excluded (gray dots) due to logistic issues during panel construction. **(C)** PGC3SEQ samples include substantial non-European ancestry. The first two principal components are plotted along the axes, colored by SCZ case control status. 1000 Genome samples are colored by superpopulation. AFR: African, AMR: Admixed American, EAS: East Asian, EUR: European, SAS: South Asian.

PGC3SEQ SCZ cases carried a significantly higher burden of rare PTVs among the 80 constrained genes on the targeted sequencing panel after adjusting for counts of rare synonymous variants and five genotype-derived ancestry PCs (OR=1.48, p-value=5.4 × 10⁻⁶,

Figure 2A, Table S5), replicating in a large independent cohort the excess burden of rare PTVs observed in 3,063 constrained genes in SCHEMA. The higher effect size seen in PGC3SEQ compared to SCHEMA ($OR_{PGC3SEQ} = 1.48$ in 80 genes; $OR_{SCHEMA} = 1.22$ in 3063 genes) demonstrates the effectiveness of our gene prioritization strategy. Limiting SCHEMA data to the 80 genes tested in PGC3SEQ showed that the enrichment in PGC3SEQ is much attenuated compared to SCHEMA ($OR_{PGC3SEQ} = 1.48$ vs $OR_{SCHEMA} = 3.0$, Figure 2A), indicating an effect overestimation in SCHEMA. Tier 1 and tier 2 missense variants were not significantly enriched in SCZ relative to controls in PGC3SEQ. This failure to replicate one of the primary SCHEMA findings may be due to a lack of power, as the effects in two studies were directionally consistent. The burden of rare synonymous variants, which were analyzed as a negative control, was significantly higher in SCZ relative to controls. Sensitivity analysis showed that this signal was due to an overall higher burden of any rare coding variants in SCZ relative to controls, rather than due to technical bias or variability between contributing cohorts (Methods, Figure S8). The global PTV enrichment remained significant after accounting for this overall higher baseline burden ($OR = 1.4$, $p\text{-value} = 1.2 \times 10^{-4}$, Figure S8C, Table S5).

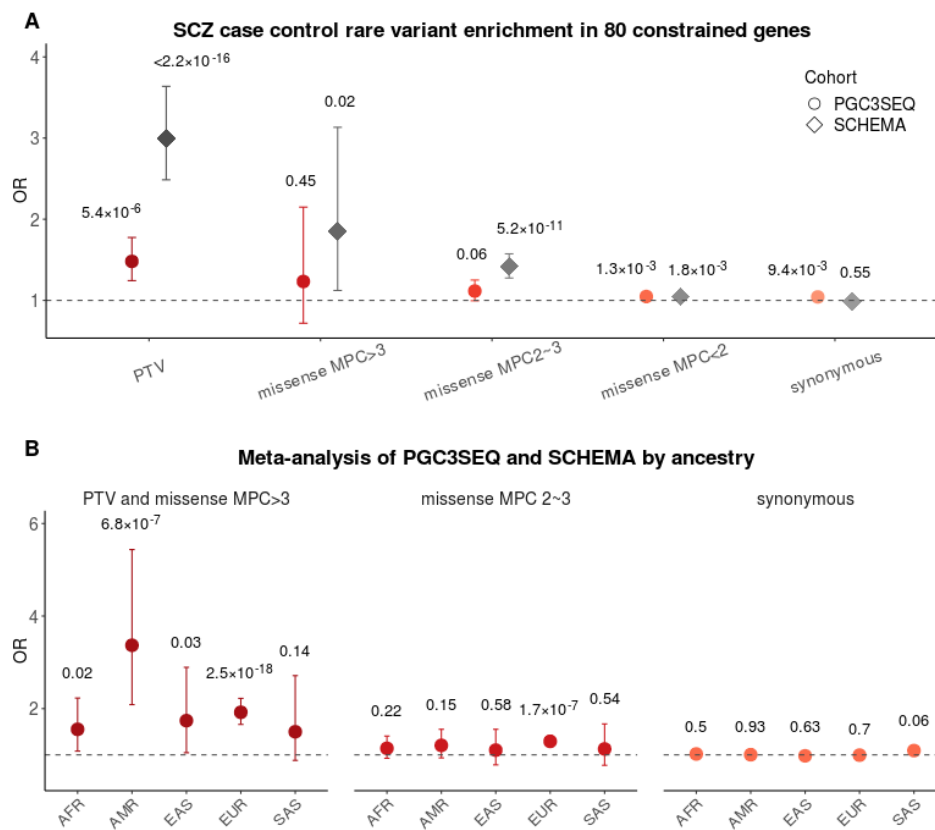


Figure 2. Global enrichment in 80 panel genes under strong constraint ($pLI > 0.9$). (A) Case-control enrichment of rare (minor allele count ≤ 5) PTV, missense, and synonymous variants in all ancestries combined, PGC3SEQ results shown in red-orange. We conducted the same analysis in the SCHEMA samples (gray) which we had access to for comparison. The enrichment folds (OR) are on plotted on the y-axis, and bars represent the 95% CIs. P-values were calculated using Firth logistic regression controlling for five ancestry PCs and either rare synonymous variant count (for PTV and missense) or rare non-synonymous variant count (for synonymous variants) to control for potential unknown technical biases. (B) Ancestry-stratified rare variant ($MAF < 0.1\%$) enrichment in the meta-analysis

of PGC3SEQ and SCHEMA. Three groups of variants were analyzed: PTV + MPC>3 missense (combined to increase power); MPC 2~3 missense; and synonymous variants.

Meta-analysis of PGC3SEQ and SCHEMA were performed to test whether the global enrichment signal was consistent across diverse ancestries (total N=57,323, N by ancestry in Figure 1A). Samples were assigned into five superpopulations used in the 1000 Genome Project (Methods). At the aggregate level, four of the five populations tested displayed a higher burden of rare disruptive variants (PTV + MPC>3 missense) in SCZ cases compared to controls at p -value<0.05 (Figure 2B left, Table S6). Although we did not find a nominally significant enrichment in the fifth population (SAS), the magnitude of enrichment was similar to that in the AFR population (OR=1.5), indicating that non-significance is likely due to a lack of statistical power (see power analysis in Methods, Figure S9). When considered separately, PGC3SEQ and SCHEMA provided independent support for the ancestry-stratified enrichments (all ancestries had OR>1 in both datasets, Table S6). Indeed, the PGC3SEQ alone showed nominal significance for AMR, EAS, and EUR, exempt from any potential effect overestimation in SCHEMA. Differences, if any, in the strength of enrichment between pairs of populations were not sizable enough to achieve statistical significance at the current sample sizes. Across five populations, burden of tier 2 missense variants was evaluated although not significant in most (ORs from 1.1 to 1.2, Figure 2B middle), whereas synonymous variants were not enriched in any (Figure 2B right).

Having replicated the global rare PTV enrichment in PGC3SEQ and established its conservation across diverse populations, we then tested for individual genes for harboring excess burden of rare PTVs in SCZ relative to controls. In the PGC3SEQ data alone, none of the 161 genes sequenced were significant after Bonferroni correction (Table S7). The directions of effects of these 161 genes were consistent with the directions observed in SCHEMA (binomial test p -value=0.016) and this observation became more pronounced when considering only those 44 genes with a SCHEMA p -value < 0.01 (binomial test p -value=0.002). Of the ten significant genes identified in SCHEMA, nine were included in the PGC3SEQ panel (*GRIA3* was not). There was an enrichment of rare PTVs on these nine genes collectively (OR=1.66, p =0.03, 49 PTVs in cases vs 24 in controls), and two genes had a p -value<0.05 when considered individually (*RB1CC1* and *CUL1*, Table 1). Notably, *SETD1A*, the gene with the strongest enrichment in SCHEMA, was not replicated in PGC3SEQ, suggesting that its effect size may have been overestimated in SCHEMA (OR_{PGC3SEQ}=1.6 vs. OR_{SCHEMA}=20.1). Another SCHEMA gene that PGC3SEQ did not support is *CACNA1G*, which among the nine SCHEMA genes on the PGC3SEQ panel had the largest number of PTV events in PGC3SEQ (n=19) yet had an OR of 0.42, directionally inconsistent with its effect in SCHEMA (OR_{SCHEMA}=3.1). Despite some evidence of winner's curse, altogether the gene-level replication tests in PGC3SEQ suggest many of the SCHEMA genes likely confer genuine disease risk, including those not yet reaching the exome-wide level, and their significance might be reached by increasing sample sizes and/or the incorporation of very recent mutations from family data.

Table 1: Attempted replication of the nine significant SCHEMA genes in PGC3SEQ

Gene	PGC3SEQ						SCHEMA	
	# of PTV allele in case	# of PTV allele in ctrl	# of allele in case	# of allele in ctrl	OR (PTV)	Fisher's exact test p-value	OR (PTV)	P-value
<i>SETD1A</i>	9	5	23160	21110	1.64	0.431	20.1	2.00x10 ⁻¹²
<i>CUL1</i>	6	0	23160	21110	Inf	0.032	36.1	2.01x10 ⁻⁹
<i>XPO7</i>	5	0	23158	21110	Inf	0.064	52.2	7.18x10 ⁻⁹
<i>TRIO</i>	3	3	23160	21110	0.91	1.000	5.0	6.35x10 ⁻⁸
<i>CACNA1G</i>	6	13	23160	21110	0.42	0.105	3.1	4.57x10 ⁻⁷
<i>SP4</i>	1	0	23150	21104	Inf	1.000	9.4	5.08x10 ⁻⁷
<i>GRIN2A</i>	0	1	23152	21104	0.00	0.477	18.1	7.37x10 ⁻⁷
<i>HERC1</i>	9	2	23160	21110	4.10	0.069	3.5	1.26x10 ⁻⁶
<i>RB1CC1</i>	10	0	23148	21108	Inf	0.002	10.0	2.00x10 ⁻⁶

Combining SCHEMA and PGC3SEQ (totaling 35,828 cases and 107,877 controls) via a p-value based meta-analysis of gene-level statistics identified two new disease genes at the exome-wide significance threshold (Table 2, Table S7): *SRRM2* (p-value=7.2 × 10⁻⁷) and *AKAP11* (p-value=4.2 × 10⁻⁷). In previous work, *SRRM2* has been shown to play a role in the tauopathy of Alzheimer's disease³⁶⁻³⁸, and *de novo* mutations in this gene have been linked to developmental disorders³⁹. *AKAP11* was suggested as a trans-gene linking to a SCZ GWAS signal in a recent study⁴⁰, which considered together with our results, adds to examples of convergence of common and rare variant associations on the same gene. A recent meta-analysis of SCHEMA and a bipolar disorder (BD) dataset also found exome-wide significance for *AKAP11*⁴¹, suggesting a role of this gene in the shared etiology of SCZ and BD. The current study consolidates the role of *AKAP11* in SCZ independent of other psychiatric disorders.

Table 2: Novel exome-wide significant SCZ genes

Gene	pLI ^a	PGC3SEQ			SCHEMA ^b		Meta P-value ^c	
		# of PTV	OR (PTV)	P-value	OR (PTV)	P-value	SCZ	SCZ & Autism
<i>AKAP11</i>	0.98	17	4.26	0.014	5.25	8.28x10 ⁻⁶	4.15x10 ⁻⁷	-
<i>SRRM2</i>	1	10	9.12	0.013	7.14	1.53x10 ⁻⁵	7.19x10 ⁻⁷	-
<i>PCLO</i>	1	8	5.01	0.024	4.02	9.36x10 ⁻⁴	1.06x10 ⁻⁵	5.84x10 ⁻⁸

^a Probability of loss-of-function intolerance

^b The SCHEMA p-values were retrieved from SCHEMA summary statistics and represent strength of evidence from both case-control and patient-proband trio (*de novo* mutation) data

^c Meta P-values were determined by Stouffer's method and weighted by sample size. SCZ, meta of PGC3SEQ and SCHEMA; SCZ&Autism, further meta with Autism Sequencing Consortium WES

Lastly, gene-level rare disruptive variant statistics from SCZ, autism spectrum disorder (ASD)⁴² and BD⁴¹ were meta-analyzed to identify pleiotropic risk genes that are not detectable at the sample sizes attained by studies of any single disorder. This identified *PCLO* as a shared risk gene for SCZ and ASD for the first time (p-value=5.8 × 10⁻⁸, Table 2). *PCLO* was not significant

in the meta-analysis of SCZ and BD samples; however, this may be due to a lack of power in the BD study ($OR_{BD}=5.2$, $p\text{-value}=0.12$). The association of *PCLO* reported here suggests this gene may be driving the common variant association at nearby loci reported in GWAS of SCZ⁴³ and other psychiatric disorders⁴⁴⁻⁴⁷.

To date, the general lack of case-control exome sequencing studies of non-Europeans has made it difficult to assess the degree to which rare PTV associations are susceptible to the well-known confounding effects of ancestry in GWAS and polygenic prediction studies⁴⁸⁻⁵². Without this knowledge, a complete view of the genetic architecture of complex diseases in human populations cannot be established. Here, we have addressed this gap in knowledge with respect to mental illnesses, showing that in SCZ rare PTV burden is conserved across diverse human populations, and therefore the biological processes disrupted by those genes are likely important in the pathogenesis of SCZ across populations. Larger sample sizes in diverse populations will be vital in further elucidating the rare and common genetic architecture of schizophrenia.

There are limitations to the current study. An interim version of the SCHEMA results was used to construct the targeted sequencing panel, and subsequent changes in the SCHEMA analytical strategy led to differences in gene-level statistics used to build the panel and ultimately in the SCHEMA publication. Specifically, the interim SCHEMA statistics^{31,32} at the time of panel design did not include *de novo* mutations from trios, used a different strategy to combine PTV and missense variants than that ultimately used in the SCHEMA publication, and were compiled before the incorporation of external Genome Aggregation Database (gnomAD) subjects that doubled the size of controls in SCHEMA (the case cohort in SCHEMA did not change between the interim and final version). Comparing the interim and the published SCHEMA results, gene ranks underwent nontrivial changes, with only 27 overlapping genes between the top 100 lists in the two versions of SCHEMA results. Consequently, our panel likely included more random noise than it would have if panel construction had waited until SCHEMA was complete (64% of panel genes dropped out of top 500 in the published version) and missed some important genes (e.g., *GRIA3*, originally ranked far below 161 yet later changed dramatically to be one of the significant ten). As whole-exome sequencing studies of other diseases approach the sample size achieved for SCHEMA and strategies are considered for how to increase power, the current report offers valuable lessons, and we note that results on datasets as large as 24,000 cases and 50,000 controls can still change substantially as more samples are added. The possibility of such changes makes the targeted panel approach vulnerable, and perhaps WES and WGS as the safest strategies despite their cost.

In summary, rare PTVs have a robust role in SCZ risk, and across ancestries their effect is consistently concentrated in genes under strong evolutionary constraint. The deconvolution of this overall contribution into individual genes, especially those that may display ancestry-specific effects, will require sequencing more individuals of diverse backgrounds. Achieving diversity in

human genetic research must be the top priority of the field in order to prevent health disparities from worsening as the findings from genetic research begin to be translated into clinical practice.

1. Singh, T. Exome sequencing identifies rare coding variants in 10 genes which confer substantial risk for schizophrenia. *medRxiv* 1–28 (2020).
2. Laursen, T. M., Nordentoft, M. & Mortensen, P. B. Excess early mortality in schizophrenia. *Annu Rev Clin Psychol* **10**, 425–448 (2014).
3. Owen, M. J., Sawa, A. & Mortensen, P. B. Schizophrenia. *The Lancet* **388**, 86–97 (2016).
4. Hjorthøj, C., Stürup, A. E., McGrath, J. J. & Nordentoft, M. Years of potential life lost and life expectancy in schizophrenia: a systematic review and meta-analysis. *Lancet Psychiatry* **4**, 295–301 (2017).
5. The Schizophrenia Working Group of the Psychiatric Genomics Consortium. Mapping genomic loci prioritises genes and implicates synaptic biology in schizophrenia. *medRxiv* 2020.09.12.20192922–50 (2020). doi:10.1101/2020.09.12.20192922
6. Marshall, C. R. *et al.* Contribution of copy number variants to schizophrenia from a genome-wide study of 41,321 subjects. *Nat Genet* **49**, 27–35 (2017).
7. Lescai, F. *et al.* Meta-analysis of Scandinavian Schizophrenia Exomes. *bioRxiv* **49**, 1166–15 (2019).
8. Singh, T. *et al.* The contribution of rare variants to risk of schizophrenia in individuals with and without intellectual disability. *Nat Genet* 1–10 (2017). doi:10.1038/ng.3903
9. Genovese, G. *et al.* Increased burden of ultra-rare protein-altering variants among 4,877 individuals with schizophrenia. *Nat. Neurosci.* **19**, 1433–1441 (2016).
10. Swedish Schizophrenia Study *et al.* Rare loss-of-function variants in SETD1A are associated with schizophrenia and developmental disorders. *Nat. Neurosci.* **19**, 571–577 (2016).
11. Fromer, M. *et al.* De novo mutations in schizophrenia implicate synaptic networks. *Nature* **506**, 179–184 (2014).
12. Purcell, S. M. *et al.* A polygenic burden of rare disruptive mutations in schizophrenia. *Nature* 1–17 (2014). doi:10.1038/nature12975
13. Steinberg, S. *et al.* Truncating mutations in RBM12 are associated with psychosis. *Nat Genet* **49**, 1251–1254 (2017).
14. Takata, A. *et al.* Loss-of-function variants in schizophrenia risk and SETD1A as a candidate susceptibility gene. *Neuron* **82**, 773–780 (2014).
15. Publications, U. N. *World Population Prospects 2019: Highlights*. (2019).
16. Popejoy, A. B. & Fullerton, S. M. Genomics is failing on diversity. *Nature* **538**, 161–164 (2016).
17. Peterson, R. E. *et al.* Genome-wide Association Studies in Ancestrally Diverse Populations: Opportunities, Methods, Pitfalls, and Recommendations. *Cell* **179**, 589–603 (2019).
18. Li, Z. *et al.* Genome-wide Analysis of the Role of Copy Number Variation in Schizophrenia Risk in Chinese. *Biological Psychiatry* **80**, 331–337 (2016).
19. Li, Z. *et al.* Genome-wide association analysis identifies 30 new susceptibility loci for schizophrenia. *Nat Genet* **49**, 1576–1583 (2017).
20. Lam, M. *et al.* Comparative genetic architectures of schizophrenia in East Asian and European populations. *Nat Genet* **51**, 1670–1678 (2019).
21. de Candia, T. R. *et al.* Additive genetic variation in schizophrenia risk is shared by populations of African and European descent. *Am. J. Hum. Genet.* **93**, 463–470 (2013).
22. Gulsuner, S. *et al.* Genetics of schizophrenia in the South African Xhosa. *Science* **367**, 569–573 (2020).
23. Bigdeli, T. B. *et al.* Contributions of common genetic variants to risk of schizophrenia among individuals of African and Latino ancestry. *Mol Psychiatry* **25**, 2455–2467 (2020).

24. Mensah-Ablorh, A. *et al.* Meta-Analysis of Rare Variant Association Tests in Multiethnic Populations. *Genet. Epidemiol.* **40**, 57–65 (2016).
25. Huffman, J. E. *et al.* Rare and low-frequency variants and their association with plasma levels of fibrinogen, FVII, FVIII, and vWF. *Blood* **126**, e19–29 (2015).
26. Zhao, J. *et al.* A Burden of Rare Variants Associated with Extremes of Gene Expression in Human Peripheral Blood. *Am. J. Hum. Genet.* **98**, 299–309 (2016).
27. Ulirsch, J. C. *et al.* The Genetic Landscape of Diamond-Blackfan Anemia. *Am. J. Hum. Genet.* **104**, 356 (2019).
28. Guo, M. H., Plummer, L., Chan, Y.-M., Hirschhorn, J. N. & Lippincott, M. F. Burden Testing of Rare Variants Identified through Exome Sequencing via Publicly Available Control Data. *Am. J. Hum. Genet.* **103**, 522–534 (2018).
29. Shaw, N. D. *et al.* SMCHD1 mutations associated with a rare muscular dystrophy can also cause isolated arhinia and Bosma arhinia microphthalmia syndrome. *Nat Genet* **49**, 238–248 (2017).
30. Hindy, G. *et al.* Rare coding variants in 35 genes associate with circulating lipid levels: a multi-ancestry analysis of 170,000 exomes. *bioRxiv* 2020.12.22.423783 (2021). doi:10.1101/2020.12.22.423783
31. Singh, T., Neale, B., Daly, M. J. & Consortium, S. Exome sequencing of 23,851 cases implicates novel risk genes and provides insights into the genetic architecture of schizophrenia. **29**, S1098 (Elsevier, 2018).
32. Singh, T., Neale, B., Daly, M. J. & Consortium, S. Initial results from the meta-analysis of the whole-exomes of over 20,000 schizophrenia cases and 45,000 controls. **29**, S813–S814 (Elsevier, 2017).
33. Nguyen, H. T. *et al.* Integrative analysis of rare variants and pathway information shows convergent results between immune pathways, drug targets and epilepsy genes. **101**, 516–28 (2018).
34. Nguyen, H. T. *et al.* Integrated Bayesian analysis of rare exonic variants to identify risk genes for schizophrenia and neurodevelopmental disorders. *Genome Med* **9**, 114–22 (2017).
35. Samocha, K. E. *et al.* Regional missense constraint improves variant deleteriousness prediction. *bioRxiv* 148353 (2017). doi:10.1101/148353
36. McMillan, P. J. *et al.* Pathological tau drives ectopic nuclear speckle scaffold protein SRRM2 accumulation in neuron cytoplasm in Alzheimer's disease. *Acta Neuropathol Commun* **9**, 117–14 (2021).
37. Lester, E. *et al.* Tau aggregates are RNA-protein assemblies that mislocalize multiple nuclear speckle components. *Neuron* **109**, 1675–1691.e9 (2021).
38. Fazeli, S. *et al.* A compound downregulation of SRRM2 and miR-27a-3p with upregulation of miR-27b-3p in PBMCs of Parkinson's patients is associated with the early stage onset of disease. *PLoS ONE* **15**, e0240855 (2020).
39. Kaplanis, J. *et al.* Evidence for 28 genetic disorders discovered by combining healthcare and research data. *Nature* **586**, 757–762 (2020).
40. Liu, S. *et al.* Illuminating links between cis-regulators and trans-acting variants in the human prefrontal cortex. *bioRxiv* 2021.09.07.459322 (2021). doi:10.1101/2021.09.07.459322
41. Palmer, D. S. *et al.* Exome sequencing in bipolar disorder reveals shared risk gene &AKAP11 with schizophrenia. *medRxiv* 2021.03.09.21252930 (2021).
42. Satterstrom, F. K. *et al.* Large-Scale Exome Sequencing Study Implicates Both Developmental and Functional Changes in the Neurobiology of Autism. *Cell* **180**, 568–584.e23 (2020).
43. Bigdeli, T. B. *et al.* Genome-Wide Association Studies of Schizophrenia and Bipolar Disorder in a Diverse Cohort of US Veterans. *Schizophr Bull* **47**, 517–529 (2021).

44. Sullivan, P. F. *et al.* Genome-wide association for major depressive disorder: a possible role for the presynaptic protein piccolo. *Mol Psychiatry* **14**, 359–375 (2009).
45. Minelli, A. *et al.* PCLO gene: its role in vulnerability to major depressive disorder. *J Affect Disord* **139**, 250–255 (2012).
46. Choi, K. H. *et al.* Gene expression and genetic variation data implicate PCLO in bipolar disorder. *Biological Psychiatry* **69**, 353–359 (2011).
47. Cross-Disorder Group of the Psychiatric Genomics Consortium. Genomic Relationships, Novel Loci, and Pleiotropic Mechanisms across Eight Psychiatric Disorders. *Cell* **179**, 1469–1482.e11 (2019).
48. Curtis, D. Polygenic risk score for schizophrenia is more strongly associated with ancestry than with schizophrenia. *Psychiatric Genetics* **28**, 85–89 (2018).
49. Martin, A. R. *et al.* Clinical use of current polygenic risk scores may exacerbate health disparities. *Nat Genet* **51**, 584–591 (2019).
50. Guo, J. *et al.* Quantifying genetic heterogeneity between continental populations for human height and body mass index. *Sci rep* **11**, 5240–9 (2021).
51. Lu, H. *et al.* Evaluating marginal genetic correlation of associated loci for complex diseases and traits between European and East Asian populations. *Hum Genet* **140**, 1285–1297 (2021).
52. Brown, B. C., Asian Genetic Epidemiology Network Type 2 Diabetes Consortium, Ye, C. J., Price, A. L. & Zaitlen, N. Transethnic Genetic-Correlation Estimates from Summary Statistics. *Am. J. Hum. Genet.* **99**, 76–88 (2016).