Research Square

Preprints are preliminary reports that have not undergone peer review.
They should not be considered conclusive, used to inform clinical practice,
or referenced by the media as validated information.

Rare coding variant burden and integrated genetic risk assessment in a Japanese schizophrenia cohort

Itaru Kushima

kushima.itaru.k0@f.mail.nagoya-u.ac.jp

Nagoya University Hospital https://orcid.org/0000-0002-0018-5178

Sawako Furukawa

Yasue Horiuchi

Makoto Arai

Tokyo Metropolitan Institute of Medical Science https://orcid.org/0000-0003-3400-9815

Masanari Itokawa

Daisuke Mori

Yuko Arioka

Nagoya University Graduate School of Medicine https://orcid.org/0000-0002-1792-8812

Hiroki Okumura

Nagoya University Graduate School of Medicine https://orcid.org/0000-0001-5506-9492

Ran Wei

Tetsuaki Arai

Hiroki Ishiguro

Nakao Iwata

Fujita Health University https://orcid.org/0000-0003-3189-6076

Masashi Ikeda

Nagoya University Graduate School of Medicine

Shigenobu Kanba

Toshifumi Kishimoto

Hiroshi Kunugi

Teikyo University School of Medicine https://orcid.org/0000-0002-7209-3790

Masaru Mimura

Toshiya Murai

Kyoto University https://orcid.org/0000-0001-6550-8188

Shusuke Numata

Yuji Ozeki

Ichiro Sora

Tsutomu Takahashi

https://orcid.org/0000-0003-3213-6183
Minoru Takebayashi
Kumamoto University https://orcid.org/0000-0001-9962-8444
Takeo Yoshikawa
RIKEN https://orcid.org/0000-0003-2791-6679
Masao Nagasaki
Yosuke Kawai
Zhenglin Guo
Stanley Center for Psychiatric Research, Broad Institute of MIT and Harvard
Hailiang Huang
Akira Sawa
Johns Hopkins University https://orcid.org/0000-0003-1401-3008
Katsushi Tokunaga
Norio Ozaki
https://orcid.org/0000-0002-7360-4898

Article

Keywords:

Posted Date: October 27th, 2026

DOI: https://doi.org/10.21203/rs.3.rs-10822530/v1

License: This work is licensed under a Creative Commons Attribution 4.0 International License.
Read Full License

Additional Declarations: Yes I.K. has received research/grant support from Novartis Pharma, GlaxoSmithKline, and Takeda Pharmaceutical Co., Ltd., all outside the submitted work. N.O. has received research support or speakers' honoraria from, or has served as a consultant to, Sumitomo Pharma, Otsuka, Viatris, Eisai, Mochida, Kyowa Pharmaceutical Industry, Nihon Medi-Physics, Nippon Chemiphar, Nippon Boehringer Ingelheim, and SUSMED, all outside the submitted work. The remaining authors declare no competing interests.

Abstract

Schizophrenia is a severe psychiatric disorder with a highly complex, polygenic architecture. Although whole-exome sequencing (WES) has identified rare coding variants conferring substantial risk, large-scale studies have predominantly focused on European populations, and the interplay between rare and common genetic risks remains largely elusive. Here, we present a comprehensive WES analysis of a Japanese cohort (3,831 cases and 5,119 controls), confirming a robust enrichment of ultra-rare protein-truncating variants (PTVs) and damaging missense variants in evolutionarily constrained and synaptic genes. To increase discovery power, we conducted a cross-population meta-analysis incorporating 32,729 cases and 108,160 controls, in which 13 previously reported risk genes were reconfirmed at exome-wide significance, while two additional genes, *ZMYM2* and *SLC6A1*, newly reached this threshold. The Japanese cohort strengthened evidence for the established SCHEMA risk gene *RB1CC1* and contributed qualifying variants supporting the associations of both additional genes. Within the same clinical cohort, we jointly evaluated carrier status for ultra-rare PTVs in known risk genes, pathogenic copy number variants (CNVs), and top-decile polygenic risk score (PRS) status. All three risk factors remained associated with schizophrenia after mutual adjustment. Our findings extend rare-variant evidence for schizophrenia to a Japanese population and support integrated assessment across rare coding, structural, and polygenic risk.

INTRODUCTION

Schizophrenia is a severe psychiatric syndrome characterized by a highly complex, polygenic architecture (1). In recent years, whole-exome sequencing (WES) has transformed our understanding of its etiology by uncovering a critical role of ultra-rare coding variants (URVs). Unlike common genetic variants, which exert subtle effects dispersed across thousands of non-coding genomic loci, rare disruptive variants—particularly PTVs—directly implicate specific risk genes. Gene-disruptive, putatively protein-damaging URVs are significantly more abundant in individuals with schizophrenia than in controls, and are strongly concentrated in neuronally expressed and synaptic genes (2). The Schizophrenia Exome Sequencing Meta-Analysis (SCHEMA) consortium has successfully leveraged these rare variants to pinpoint specific risk genes conferring substantial risk, with odds ratios (ORs) ranging from 3 to 50 (3). These discoveries provide unprecedented windows into the underlying neurobiology of the disorder, highlighting disruptions in glutamatergic synaptic signaling and chromatin organization.

Despite the proven utility of rare variant analyses, major challenges remain in fully delineating their impact and positioning within the broader genetic landscape. First, rare-variant classification and association analyses depend critically on ancestry-specific allele frequencies and the representation of diverse populations in reference databases. While large-scale genome-wide association studies (GWAS) and array comparative hybridization (aCGH) studies have successfully elucidated the roles of common variants and CNVs in the Japanese schizophrenia population (4–7), comprehensive WES evaluations of URVs in this specific population remain limited. Most schizophrenia WES studies to date have been

overwhelmingly restricted to populations of European ancestry. Expanding WES to diverse, non-European populations is imperative to capture the full global spectrum of rare risk alleles and maximize power for additional gene discovery. Second, although rare coding variants can confer large effects, their relationships with other established genetic risk factors remain poorly understood. Current estimates suggest that rare coding variants and CNVs collectively explain ~ 5% of the variance in schizophrenia liability (8). CNVs and PRS act in an additive manner (9). However, it remains unclear whether ultra-rare single-nucleotide variants (SNVs) and insertions/deletions (INDELs), including PTVs in risk genes, are associated with schizophrenia after accounting for pathogenic CNVs and common polygenic risk within the same clinical cohort.

Here, we conducted a comprehensive WES analysis of a Japanese schizophrenia cohort comprising 3,831 cases and 5,119 controls to investigate rare coding variant burden and its relationship with established pathogenic CNVs and common polygenic risk factors. We first characterized the exome-wide burden of URVs, demonstrating a significant enrichment of both PTVs and damaging missense variants across evolutionarily constrained, synaptic, and previously implicated risk gene sets. To maximize discovery power, we conducted a cross-population meta-analysis integrating our Japanese cohort with previously published large-scale WES data (3, 10), in which 13 previously reported risk genes were reconfirmed and *ZMYM2* and *SLC6A1* newly reached exome-wide significance. Finally, within the same cohort, we jointly evaluated carrier status for ultra-rare PTVs in 13 autosomal genes among the 14 previously established schizophrenia risk genes, established schizophrenia-associated pathogenic CNVs (11), and top-decile PRS status. This integrated analysis evaluated the mutually adjusted associations of known risk-gene PTVs, pathogenic CNVs, and polygenic risk with schizophrenia.

METHODS

Sample Information

The Japanese case-control cohort initially comprised WES data from 9,521 individuals (4,219 schizophrenia cases and 5,302 controls). After quality control (QC), data from 8,950 individuals remained (3,831 cases and 5,119 controls) (**Supplementary Table 1**). All cases met the Diagnostic and Statistical Manual of Mental Disorders, Fifth Edition (DSM-5) criteria for schizophrenia. Study protocols were approved by the institutional review boards or ethics committees of all participating institutions involved in participant recruitment and sample collection. The names of the institutions and the corresponding approval numbers are provided in **Supplementary Table 1**. Written informed consent for genetic analyses and data sharing was obtained from all participants in accordance with the requirements of the respective ethics committees. The study protocol adhered to the Declaration of Helsinki guidelines.

Whole-Exome Sequencing and Quality Control

Genomic DNA was extracted and subjected to WES. WES was performed at the Broad Institute using the Illumina NovaSeq 6000 platform with the Nextera Exome Kit, and the resulting sequencing data were processed into CRAM files. Sequence reads were aligned to the human reference genome (GRCh38). Variants were joint-called across all samples using the Genome Analysis Toolkit (GATK) HaplotypeCaller V4.1.8 and filtered using GATK Variant Quality Score Recalibration (12). Samples exhibiting outlier values in key sequence quality metrics (e.g. number of singletons, Het/Hom ratio, Ti/Tv ratio, and overall call rate) were excluded (**Supplementary Fig. 1**).

At the variant and genotype levels, subsequent analyses within the Japanese cohort were strictly restricted to biallelic autosomal variants located within the targeted regions of the capture kit to avoid potential confounding from sex-chromosome dosage differences. However, the summary statistics obtained from the SCHEMA (3) and Chick *et al.* (10) cohorts included variants on the X chromosome (excluding pseudoautosomal regions). Therefore, X-linked genes, such as *GRIA3*, were retained in the cross-population meta-analysis based solely on the contributions from these previously published cohorts. We applied stringent genotype-level filters based on read depth (DP), genotype quality (GQ), and alternate allele fraction (AB), with variant-specific thresholds defined separately for SNVs and INDELs (**Supplementary Methods**). Finally, to ensure the absence of systematic sequencing biases between cases and controls in the final dataset (N = 8,950), we evaluated per-sample sequence quality metrics. We confirmed no significant differences in mean Ti/Tv ratio (3.27 for both groups) or average read depth (39.5x vs 39.4x) between cases and controls (**Supplementary Table 2**).

Genotyping Array Data and Population Structure

To investigate cryptic relatedness and population stratification, we employed the Global Screening Array (GSA) dataset. QC of the array data was performed using PLINK (13). We restricted the analysis to high-quality autosomal SNPs and applied standard filtering criteria for missing rates, minor allele frequency, and Hardy-Weinberg equilibrium. Following linkage disequilibrium (LD) pruning, probable duplicates and related individuals were sequentially pruned using pairwise genomic relatedness cutoffs of 0.95 and 0.2, respectively (**Supplementary Fig. 2, Supplementary Table 1**). To evaluate global population stratification, we merged our unrelated cohort with the 1000 Genomes Project reference panels, analyzing both a multi-ancestry set (including JPT, CHB, CEU, and YRI) and an East Asian-specific set (JPT and CHB). The merged datasets were subjected to separate stringent QC and LD-pruning procedures to retain only high-quality common SNPs. Principal component analysis (PCA) confirmed that our samples clustered appropriately with the East Asian reference populations **(Supplementary Fig. 3A, B)**.

Finally, using the initial pruned GSA dataset, we conducted a within-cohort PCA to assess residual population structure and possible case–control differences **(Supplementary Fig. 3C-F)**. The top five PCs were used as covariates in downstream regression analyses.

Rare Variant Annotation and Classification

Variants were annotated using ANNOVAR (14). URVs were defined as cohort singletons (minor allele count = 1) absent from the gnomAD v4.1 genomes (15) (allele frequency = 0). Because potential sample overlap with the present cohort could not be excluded, gnomAD v4.1 exome frequencies were not used for external frequency filtering. URVs were classified into five functional categories: 1) PTVs (nonsense, frameshift, and essential splice-site variants); 2) severe missense variants with an MPC > 3; 3) missense variants with 2 < MPC $\leq$ 3; 4) missense variants with MPC ≤ 2; and 5) synonymous variants.

URV Burden and Geneset Enrichment Analysis

In the cohort-wide URV burden analysis, we aimed to estimate both the excess of URVs per case (β coefficients) and the ORs for the schizophrenia phenotype. Linear and logistic regression models evaluated the association between the number of carried variants in each class and schizophrenia status. Variant-class burden analyses were adjusted for sex, the top five PCs (determined via scree plot; **Supplementary Fig. 3F**), and total URV count per individual; total URV-model was adjusted for sex and the top five PCs. Gene-set enrichment analyses evaluated PTVs, MPC > 2 Missense, and their combination. Genesets included: 1) highly constrained genes (LOEUF < 0.35 in gnomAD; 1386 genes); 2) missense constrained genes (mis_z > 3.09 in gnomAD; 1532 genes); 3) established schizophrenia risk genes from large WES studies (3, 10, 16) (14 genes); 4) schizophrenia risk genes identified in GWAS study (17) (120 genes); 5) SynGO synaptic genes (18) (1233 genes); and 6,7) genes robustly associated with autism spectrum disorder (ASD) (72 genes) and developmental disorders (DD) (309 genes) (19) (**Supplementary Table 3**). To address potential baseline technical confounding, we performed additional sensitivity analyses for both the exome-wide and gene-set URV burdens by further adjusting for the total variant count per sample. Within each variant definition and regression framework, *P*-values across the seven prespecified gene sets were adjusted using the Benjamini–Hochberg procedure, with $q < 0.05$ considered significant.

Meta-analysis

To perform a cohort-stratified gene-level meta-analysis, we combined gene-level variant counts from our Japanese cohort with publicly available summary counts from the SCHEMA consortium (3) and an independent case-control cohort reported by Chick *et al*. (10). The three cohorts were treated as separate strata in Cochran–Mantel–Haenszel (CMH) tests with continuity correction, using the observed 2 × 2 tables without adding pseudo-counts to individual cells. Three nested variant models were evaluated: PTVs alone, PTVs plus missense variants with MPC > 3, and PTVs plus missense variants with MPC > 2. For models with a CMH *P*-value ≤ 0.01 and an available published gene-based de novo *P*-value, Fisher's combined probability test was applied, and the smaller of the CMH and Fisher-combined *P*-values was retained. For each gene, the minimum final *P*-value across the three prespecified variant models was multiplied by three, followed by Benjamini–Hochberg correction across genes. Three-model-adjusted *P*-values < 2.5×10^{-6} and q-values < 0.05 were considered exome-wide significant and FDR-

significant, respectively. Full analytical details and limitations of the summary-count approach are provided in the **Supplementary Methods**.

Visualization of overlap with GWAS loci

To evaluate the physical overlap between the genes newly reaching exome-wide significance identified in our WES meta-analysis and known common variant signals, we generated regional association plots using LocusZoom (20). Specifically, we visualized the genetic landscape surrounding *ZMYM2* and *SLC6A1* by overlaying schizophrenia GWAS summary statistics for East Asian populations derived from a recent large-scale multi-ancestry study by Bigdeli *et al.* (21). To accurately reflect the population-specific haplotype structure, LD was calculated using the 1000 Genomes Project East Asian (EAS) reference panel.

Copy Number Variant Calling and Quality Control

CNVs were identified in the WES CRAM files using the eXome-Hidden Markov Model (XHMM) software tool (22), following its recommended best practices. First, we calculated the mean read depth for each targeted exon. To reduce systematic biases and technical artifacts, we applied strict filtering criteria to the read depth matrix based on GC content, target size, and read depth metrics. After PCA-based normalization to eliminate batch effects, the HMM Viterbi algorithm was applied to discover CNV events (deletions and duplications) from the normalized read depth matrix (**Supplementary Methods**). The resulting CNV calls were subsequently annotated using AnnotSV (23). Pathogenic CNV carrier status was defined by identifying CNVs corresponding to the established schizophrenia-associated loci reported by Marshall *et al*. (11), with concordance of genomic location and deletion or duplication type, based on the AnnotSV annotations. Loci included in this analysis are listed in **Supplementary Table 4**. For a subset of samples, CNV calls from WES were orthogonally validated using aCGH data (4–6).

Polygenic Risk Score Analysis

Following standard genotype QC, GSA variants were aligned to the GRCh37 1000 Genomes Project Phase 3 v5a reference panel using conform-gt and were subsequently phased and imputed using Beagle 5.5 (24, 25) (**Supplementary Methods**). The primary East Asian (EAS)-derived PRS was generated using PRS-CS (26) with summary statistics from the PGC3 EAS schizophrenia GWAS (17) and the 1000 Genomes EAS LD reference panel. The PRS-CS n_gwas parameter was set to 30,515, and the global shrinkage parameter (phi) was set to 1×10^{-2}. Posterior SNP effect sizes were applied to imputed allele dosages to calculate individual PRS. To assess the robustness of the results to potential sample overlap with the EAS discovery GWAS, we separately generated a European (EUR)-derived PRS using PRS-CS with EUR schizophrenia GWAS summary statistics (21), the 1000 Genomes EUR linkage disequilibrium reference panel, an n_gwas value of 250,326, and the same phi value. Each PRS was standardized across the 8,948 participants included in the PRS analyses to have a mean of 0 and a standard deviation

of 1. Model discrimination was evaluated using the area under the receiver operating characteristic curve from logistic regression models including standardized PRS, sex, and the top five PCs. PRS data were available for 8,948 of the 8,950 participants who passed WES quality control (3,829 cases and 5,119 controls).

Joint Analysis of Rare, Structural, and Polygenic Risk Factors

The joint analysis was restricted to the 8,948 participants with available PRS data (3,829 cases and 5,119 controls). To evaluate PTVs in previously identified risk genes alongside other genetic risk, we conducted an integrated analysis of known risk-gene PTVs, pathogenic CNVs, and common polygenic risk within the Japanese cohort. Among the 14 genes that had previously reached exome-wide significance in schizophrenia sequencing studies, known risk-gene PTV carrier status was defined based on the presence of at least one ultra-rare PTV in any of the 13 autosomal genes evaluated in the present study: *SETD1A*, *CUL1*, *XPO7*, *SP4*, *GRIN2A*, *HERC1*, *RB1CC1*, *TRIO*, *CACNA1G*, *SRRM2*, *AKAP11*, *STAG1*, and *ZNF136* (3, 10, 16). *GRIA3*, located on chromosome X, was not evaluated because the Japanese exome analysis was restricted to autosomes. For each individual, we defined three binary variables: 1) carrier status for ultra-rare PTVs in these 13 known autosomal risk genes, 2) carrier status of established schizophrenia-associated pathogenic CNVs reported by Marshall *et al.* (11), and 3) presence in the top decile of schizophrenia PRS. The top-decile threshold was defined within the 8,948 participants included in the joint analysis.

We constructed a joint logistic regression model incorporating all three variables and adjusted for sex and the top five PCs to estimate mutually adjusted ORs. As a sensitivity analysis, the model was repeated by replacing EAS-derived PRS top-decile status with EUR-derived PRS top-decile status, which was defined separately using the EUR-derived PRS distribution. Case proportions were descriptively summarized by the number of risk factors carried.

Use of AI tools

OpenAI ChatGPT and Google Gemini were used for language editing and assistance with code review; all analyses and outputs were independently verified by the authors.

RESULTS

URV burden in the Japanese schizophrenia cohort

We analyzed Japanese WES data comprising 3,831 schizophrenia cases and 5,119 controls. We focused on autosomal URVs, that is, singletons in our dataset that are not observed in the gnomAD v4.1 genomes. After filtering and QC, we extracted 355,769 URVs (346,147 SNVs and 9,622 INDELs). We performed burden analyses across five functional categories: PTVs, missense variants with MPC > 3, missense variants with 2 < MPC $\leq$ 3, missense variants with MPC ≤ 2, and synonymous variants.

Among these, ultra-rare PTVs and damaging missense variants with MPC > 2 were enriched in cases (Fig. 1, **Supplementary Table 5**). PTVs exhibited a significant excess burden per case and increased disease risk (β = 0.17, 95% CI = 0.11–0.24, $P = 4.9 \times 10^{-7}$; OR = 1.07, 95% CI = 1.04–1.10, $P = 5.1 \times 10^{-7}$). Missense variants with 2 < MPC $\leq$ 3 were also enriched (β = 0.08, 95% CI = 0.05–0.11, $P = 4.1 \times 10^{-6}$; OR = 1.13, 95% CI = 1.08–1.20, $P = 4.4 \times 10^{-6}$). Missense variants with an MPC > 3 showed the largest OR but did not reach nominal significance (β = 0.01, 95% CI = 0.00 − 0.02, *P* = 0.053; OR = 1.19, 95% CI = 1.00–1.41, *P* = 0.054). Combining damaging missense variants with MPC > 2 yielded β = 0.09 (95% CI = 0.06–0.13, $P = 6.3 \times 10^{-7}$) and OR = 1.14 (95% CI = 1.08–1.20, $P = 6.9 \times 10^{-7}$) (**Supplementary Fig. 4**).

Missense variants with MPC ≤ 2 or synonymous URVs showed no risk-increasing enrichment (Fig. 1, **Supplementary Fig. 4, Supplementary Table 5**). The total number of all variants (including common variants) in the capture region was slightly lower in cases than in controls (β = − 61.15, 95% CI = − 104.63 to − 17.67, *P* = 0.0058; **Supplementary Table 2**). By contrast, total URV count did not differ significantly (β = 0.12, 95% CI = − 0.21 to 0.45, *P* = 0.47; OR = 1.00, 95% CI = 1.00–1.01, *P* = 0.47; **Supplementary Table 5**), confirming that the observed disease association is specific to deleterious URVs. Because the overall number of called variants was lower in cases, we also fitted sensitivity models with total variant count per sample as an additional covariate. Enrichment of ultra-rare PTVs and damaging missense variants remained significant in these models (**Supplementary Table 5)**.

Gene-level and gene-set enrichment analyses highlight neurodevelopmental pathways

We then conducted gene-level association testing in the Japanese cohort across three variant inclusion criteria: PTVs alone, PTVs + highly damaging missense variants (MPC > 3), and PTVs + damaging missense variants (MPC > 2). Given the limited statistical power for individual rare-variant associations in a single cohort, no gene achieved exome-wide significance ($P < 2.5 \times 10^{-6}$), and no gene had a *q*-value < 0.05 in any model (**Supplementary Table 6**).

To assess the biological relevance of the observed URV burden, we performed gene-set enrichment analyses for PTVs, damaging missense variants (MPC > 2), and their combination (PTV + MPC > 2). Case-enriched PTVs were significantly overrepresented in highly constrained genes (LOEUF < 0.35; β = 0.045, $P = 1.2 \times 10^{-12}$; OR = 1.69, $P = 2.5 \times 10^{-12}$), missense constrained genes (β = 0.039, $P = 2.2 \times 10^{-5}$; OR = 1.24, $P = 2.3 \times 10^{-5}$), established schizophrenia WES genes (β = 0.007, $P = 1.2 \times 10^{-6}$; OR = 7.49, $P = 3.6 \times 10^{-5}$), SynGO genes (β = 0.038, $P = 2.3 \times 10^{-6}$; OR = 1.31, $P = 2.5 \times 10^{-6}$), ASD genes (β = 0.005, $P = 3.4 \times 10^{-3}$; OR = 2.09, $P = 4.0 \times 10^{-3}$) and DD genes (β = 0.016, $P = 2.8 \times 10^{-5}$; OR = 1.66, $P = 3.6 \times 10^{-5}$; Fig. 2A, **Supplementary Table 7**).

Damaging missense variants (MPC > 2) were also enriched in similar gene sets, including highly constrained genes (β = 0.042, $P = 1.9 \times 10^{-5}$; OR = 1.23, $P = 2.0 \times 10^{-5}$), missense constrained genes (β = 0.053, $P = 2.7 \times 10^{-6}$; OR = 1.22, $P = 2.9 \times 10^{-6}$), SynGO synaptic genes (β = 0.021, $P = 2.4 \times 10^{-3}$; OR =

1.23, $P = 2.4 \times 10^{-3}$), and DD-associated genes (β = 0.023, $P = 2.1 \times 10^{-4}$; OR = 1.32, $P = 2.2 \times 10^{-4}$; **Supplementary Fig. 5, Supplementary Table 7**).

Combining all deleterious coding variants (PTV + MPC > 2) showed enrichment in highly constrained genes (β = 0.087, $P = 7.3 \times 10^{-14}$; OR = 1.35, $P = 1.1 \times 10^{-13}$), missense constrained genes (β = 0.092, $P = 2.0 \times 10^{-10}$; OR = 1.23, $P = 2.4 \times 10^{-10}$), schizophrenia WES genes (β = 0.009, $P = 7.9 \times 10^{-7}$; OR = 3.85, $P = 4.4 \times 10^{-6}$), SynGO synaptic genes (β = 0.060, $P = 2.2\times 10^{-8}$; OR = 1.28, $P = 2.4 \times 10^{-8}$), ASD genes (β = 0.011, $P = 2.8 \times 10^{-3}$; OR = 1.45, $P = 2.9 \times 10^{-3}$), and DD genes (β = 0.039, $P = 7.1\times 10^{-8}$; OR = 1.41, $P = 8.6\times 10^{-8}$; Fig. 2B, **Supplementary Table 7**).

Schizophrenia GWAS-prioritized genes were not significantly enriched for URVs under any of the three variant models in the primary analyses (all $P \geq 0.051$; Fig. 2; **Supplementary Fig. 5; Supplementary Table 7**).

Cross-population meta-analysis identified exome-wide significant schizophrenia risk genes

To increase the statistical power for gene discovery, we integrated the Japanese cohort with published large-scale schizophrenia WES datasets, including the SCHEMA consortium (3) and the independent case-control cohort reported by Chick et al. (10), yielding a combined meta-analysis of 32,729 cases, 108,160 controls, and 3,444 proband-parent trios. We used CMH tests across the three cohort strata, followed by integration with published *de novo* variant data via Fisher's combined probability method. After accounting for the three nested variant categories, 15 genes met the exome-wide significance threshold ($P < 2.5 \times 10^{-6}$), comprising 13 previously reported risk genes and two additional genes, *ZMYM2* and *SLC6A1*, that newly reached this threshold in the combined analysis (Fig. 3A, B, **Table 1**). In total, 40 genes met the FDR threshold (q < 0.05), of which 25 did not reach exome-wide significance (Fig. 3B). Illustrative clinical observations for individuals carrying PTVs in *SRRM2* and *PCLO* are provided in the **Supplementary Note**.

The Japanese cohort also strengthened evidence for *RB1CC1*, an established SCHEMA risk gene: seven Japanese cases and no controls carried a qualifying PTV (Fig. 3C), and the three-model-adjusted *P* value was 3.02×10^{-8} versus 3.17×10^{-6} in a sensitivity analysis excluding the Japanese stratum (**Table 1**). Evidence for *AKAP11* was also strengthened (four Japanese case PTV carriers and no control carriers; **Table 1**). These findings represent cross-population support rather than novel gene discovery.

Among the two additional genes, the exome-wide significance of *ZMYM2* was supported by PTVs identified in the Japanese cohort (Fig. 3C). *ZMYM2* encodes a nuclear zinc finger protein, part of a transcriptional complex acting as a corepressor by interacting with different nuclear receptors and the LSD1-CoREST-HDAC1 complex on chromatin (27). Clinically, pathogenic *ZMYM2* variants cause a neurodevelopmental-craniofacial syndrome with variable renal and cardiac abnormalities (OMIM 619522) (28).

For *SLC6A1*, exome-wide significance was reached only after integration with published de novo evidence: the three-model-adjusted *P*-value was 1.1×10^{-3} in the case-control-only analysis and 1.6×10^{-6} after integration of de novo evidence (**Supplementary Table 8**). No case PTV carriers were observed in any of the three case-control cohorts, and the Japanese cohort contributed one case carrying a qualifying damaging missense variant and no control carriers (Fig. 3C). In the fixed-model sensitivity analysis integrating de novo evidence but excluding the Japanese stratum, the three-model-adjusted *P*-value was 2.8×10^{-6}, slightly above the exome-wide significance threshold. *SLC6A1* encodes the GABA transporter GAT-1 and is an established causative gene for myoclonic-atonic epilepsy (OMIM 616421) (29).

Neither *ZMYM2* nor *SLC6A1* overlapped with genome-wide significant loci identified in recent common variant GWAS (21) (**Supplementary Fig. 6**), highlighting how URV studies can pinpoint loci not implicated by current common-variant GWAS.

CACNA1G, which reached exome-wide significance in the original SCHEMA study (3), fell below this threshold in our meta-analysis, consistent with the expanded analysis reported by Chick *et al.* (10) (Fig. 3B).

Joint analysis of rare coding variants, pathogenic CNVs, and polygenic risk

Among the 8,948 participants with available PRS data (3,829 cases and 5,119 controls), we jointly evaluated ultra-rare PTVs in 13 autosomal genes among the 14 previously identified schizophrenia risk genes (hereafter referred to as known risk-gene PTVs) (3, 10, 16), established schizophrenia-associated pathogenic CNVs (11), and top decile of schizophrenia PRS. Before the main joint analysis, we evaluated the association and model discrimination of the continuous PRS. In a logistic regression model adjusted for sex and the top five PCs, the PRS was highly associated with schizophrenia status (OR per 1 SD = 1.63, $P = 8.6 \times 10^{-95}$), achieving a full-model AUC of 0.67 (**Supplementary Fig. 7**).

In the mutually adjusted joint logistic regression model, all three binary risk factors remained associated with schizophrenia: known risk-gene PTV carrier status (OR = 8.05, 95% CI = 3.09–20.94, $P = 1.9 \times 10^{-5}$), pathogenic CNV carrier status (OR = 3.63, 95% CI = 2.27–5.79, $P = 6.3 \times 10^{-8}$), and top-decile PRS status (OR = 2.50, 95% CI = 2.16–2.89, $P = 5.0 \times 10^{-35}$) (Fig. 4A, 4B**)**.

The proportion of schizophrenia cases was descriptively higher among individuals carrying ≥ 1 genetic risk factor than among those carrying none. Case proportions were also summarized according to the number of risk factors carried; these comparisons were descriptive because some multi-risk groups contained few individuals (Fig. 4C, **Supplementary Fig. 8**). To assess the robustness of the findings to potential sample overlap, we repeated the analysis using an independent European-derived PRS. This sensitivity analysis yielded broadly similar mutually adjusted associations and descriptive case-proportion patterns across risk-factor categories (**Supplementary Fig. 9**). Collectively, these findings

indicate that known risk-gene PTVs, pathogenic CNVs, and polygenic risk each remained associated with schizophrenia after mutual adjustment.

DISCUSSION

Here, we performed a comprehensive investigation of rare coding variation in schizophrenia in a non-European population. This study has three major findings. First, we confirmed a robust enrichment of ultra-rare PTVs and damaging missense variants in schizophrenia cases. Second, our cross-population meta-analysis reconfirmed 13 previously reported exome-wide significant risk genes, while two additional genes, *ZMYM2* and *SLC6A1*, newly reached the exome-wide significance threshold in the combined analysis. Finally, in the joint analysis, known risk-gene PTVs, pathogenic CNVs, and polygenic risk each remained associated with schizophrenia after mutual adjustment, while descriptive case-proportion patterns were consistent with multiple components of schizophrenia liability.

The enrichment of ultra-rare PTVs and damaging missense variants in the Japanese cohort replicates findings from European populations (2, 3, 10), supporting a shared role of strongly deleterious variants across diverse genetic ancestries. By contrast, no significant enrichment was detected in the schizophrenia GWAS-prioritized gene set in this Japanese cohort (**Supplementary Table 7**), although this null result does not establish an absence of overlap between rare- and common-variant targets.

Beyond increasing the total sample size, the Japanese cohort made a measurable contribution to gene-level evidence. For *RB1CC1*, an established SCHEMA risk gene, the inclusion of seven Japanese case PTV carriers and no control carriers reduced the three-model-adjusted *P* value from 3.17×10^{-6} without the Japanese stratum to 3.02×10^{-8} in the full analysis (**Table 1**). Because the Japanese cohort alone was underpowered for individual-gene discovery, this should be viewed as cross-population support rather than a population-specific association. Together with *ZMYM2* and *SLC6A1* crossing the threshold after inclusion of the Japanese stratum, this illustrates how adding ancestrally diverse cohorts can reinforce established signals and enable additional gene discovery.

Neither *ZMYM2* nor *SLC6A1* lies within an established schizophrenia GWAS locus (**Supplementary Fig. 6**). *ZMYM2* supports a role for disrupted chromatin organization and transcriptional regulation, consistent with *SETD1A* (30), *STAG1*, and *SP4* (3). *SLC6A1* is consistent with perturbed GABAergic signaling in schizophrenia (31, 32); together with *GRIN2A*, *GRIA3*, and *TRIO*, it supports involvement of both inhibitory and excitatory synaptic pathways (33). Rare de novo *SLC6A1* missense variants have been reported in schizophrenia (34), but exome-wide significance here depended on integration of published de novo evidence, and the Japanese cohort contributed only one qualifying case carrier; the result therefore warrants cautious interpretation.

The two genes also show contrasting spatiotemporal expression profiles (35); *ZMYM2* is predominantly expressed during early fetal neurodevelopment, whereas *SLC6A1* expression sharply increases postnatally (**Supplementary Fig. 10**). Pathogenic variants in *ZMYM2* and *SLC6A1* cause severe

monogenic neurodevelopmental disorders, such as syndromic craniofacial malformations (28) and myoclonic-atonic epilepsy (29), respectively. This highlights a pervasive genetic pleiotropy, suggesting that the early neurodevelopmental perturbations caused by these variants can manifest as severe developmental delays in childhood or, perhaps modulated by genetic background or environmental factors, as schizophrenia in adulthood.

Rare coding variants, pathogenic CNVs, and common polygenic risk have often been evaluated separately. By integrating these three classes of genetic variation within a single clinical cohort, we found that known risk-gene PTV carrier status, pathogenic CNV carrier status, and top-decile PRS status each remained associated with schizophrenia after mutual adjustment. These findings suggest that the associations of rare coding variants and pathogenic CNVs were not explained by the measured polygenic risk or by each other in this cohort. However, given the small numbers of PTV and CNV carriers and the resulting uncertainty in their effect estimates, these analyses do not establish additivity, interaction, differences in effect magnitude, or the relative contributions of these classes of genetic variation.

Several limitations should be noted. First, while our cross-population meta-analysis identified risk genes, the sample size of this non-European cohort remains considerably smaller than that of current European efforts. Larger WES studies of ancestrally diverse cohorts will therefore be needed to increase power for risk-gene discovery and to characterize population-specific rare coding variation. Second, while our joint model successfully integrated major variant classes, it did not capture complex structural variants or non-coding regulatory variation, which may also play critical roles. Third, there was a potential sample overlap in our primary PRS analysis. Base summary statistics derived from the East Asian GWAS (17) may include a maximum of ~ 1,000 individuals who overlap with our current study cohort. To assess the robustness of the findings to potential sample overlap, we repeated the analysis using a EUR-derived PRS based on an independent discovery GWAS. The sensitivity analysis yielded broadly similar mutually-adjusted associations and descriptive case-proportion patterns (**Supplementary Fig. 9**), suggesting that the findings were not only due to sample overlap. Fourth, some groups carrying multiple genetic risk factors contained only a small number of individuals. Accordingly, analyses based on cumulative numbers and specific combinations of risk factors should be interpreted as descriptive; larger cohorts are required to evaluate dose-response trends, additivity, and interactions with adequate statistical power. Finally, the mechanisms through which ultra-rare coding variants in *ZMYM2* or *SLC6A1* contribute to schizophrenia pathophysiology remain to be elucidated. Future functional studies using *in vitro* human cellular models or *in vivo* animal models are required to translate these genomic discoveries into the underlying pathophysiological mechanisms.

In conclusion, this study extends rare-coding-variant evidence for schizophrenia to a Japanese cohort and demonstrates the value of cross-population meta-analysis for risk-gene discovery. The findings implicate chromatin regulation and GABAergic signaling and show that the association of known risk-gene PTV carrier status was robust to adjustment for measured CNV and polygenic risk factors.

Declarations

ACKNOWLEDGEMENTS

This work was supported by the Japan Agency for Medical Research and Development (AMED) (Grant Nos. JP23ak0101215, JP24zf0127011, JP22gm1410011, JP23gm1910005, JP23ek0109678, JP25wm0625518 [to N.O.], JP21km0405216 [to I.K.], JP22tm0424222 [to S.F., Y.H., R.W., M.N., M.I., Y.K., K.T., and I.K.], and JP24wm0625302 [to T.T.]), the Japan Society for the Promotion of Science (JSPS KAKENHI Grant Nos. 23KJ1112 [to S.F.], 21K07543 [to I.K.], 21H00194 [to I.K.], 25K22598 [to N.O.], and 26K02355 [to N.O.]), and the SENSHIN Medical Research Foundation [to I.K.]. Computations were performed on the SHIROKANE supercomputer at the Human Genome Center, The Institute of Medical Science, The University of Tokyo.

CONFLICT OF INTEREST

I.K. has received research/grant support from Novartis Pharma, GlaxoSmithKline, and Takeda Pharmaceutical Co., Ltd., all outside the submitted work. N.O. has received research support or speakers' honoraria from, or has served as a consultant to, Sumitomo Pharma, Otsuka, Viatris, Eisai, Mochida, Kyowa Pharmaceutical Industry, Nihon Medi-Physics, Nippon Chemiphar, Nippon Boehringer Ingelheim, and SUSMED, all outside the submitted work. The remaining authors declare no competing interests.

AUTHOR CONTRIBUTIONS

S.F., Y.H., Masanari Itokawa, I.K., N.O. conceptualized and designed the study. S.F. developed the analytical framework, performed the analysis, interpreted the results, and wrote the original manuscript. D.M., Y.A., H.O., and R.W. contributed to the interpretation of the findings through scientific discussions within the AMED GRIFIN project and critically reviewed the manuscript. Y.H., Masanari Itokawa, T.A., H.I., N.I., Masashi Ikeda, S.K., T.K., H.K., M.M., T.M., S.N., Y.O., I.S., T.T., M.T., T.Y., A.S. and N.O. contributed to participant recruitment and sample collection. M.A. supervised the research activities and provided institutional oversight for the study. Z.G. contributed to the generation and curation of the WES and GSA data. H.H. supervised the data generation and provided relevant resources. M.N., Y.K., K.T., I.K. and N.O. supervised the study.

DATA AVAILABILITY

Schizophrenia genotype data are available upon application to the Stanley Global Asia Initiatives: SGAI@broadinstitute.org. These data must be under controlled access due to the data use limitation imposed by the compliance, participant consent and national laws. Application to access such data requires a short research proposal that will be reviewed by principal investigator of the constituent study and, if necessary, by the respective ethic committee. The principal investigator review process takes 2 weeks.

CODE AVAILABILITY

The software and code used in this study are described in the Methods. The code used to generate the results is available at https://github.com/sawakof/SCZ_WES.

References

1. Jauhar S, Johnstone M, McKenna PJ. Schizophrenia. Lancet. 2022;399(10323):473–86.
2. Genovese G, Fromer M, Stahl EA, Ruderfer DM, Chambert K, Landen M, et al. Increased burden of ultra-rare protein-altering variants among 4,877 individuals with schizophrenia. Nat Neurosci. 2016;19(11):1433–41.
3. Singh T, Poterba T, Curtis D, Akil H, Al Eissa M, Barchas JD, et al. Rare coding variants in ten genes confer substantial risk for schizophrenia. Nature. 2022;604(7906):509–16.
4. Kushima I, Aleksic B, Nakatochi M, Shimamura T, Shiino T, Yoshimi A, et al. High-resolution copy number variation analysis of schizophrenia in Japan. Mol Psychiatry. 2017;22(3):430–40.
5. Kushima I, Aleksic B, Nakatochi M, Shimamura T, Okada T, Uno Y, et al. Comparative Analyses of Copy-Number Variation in Autism Spectrum Disorder and Schizophrenia Reveal Etiological Overlap and Biological Insights. Cell Rep. 2018;24(11):2838–56.
6. Kushima I, Nakatochi M, Aleksic B, Okada T, Kimura H, Kato H, et al. Cross-Disorder Analysis of Genic and Regulatory Copy Number Variations in Bipolar Disorder, Schizophrenia, and Autism Spectrum Disorder. Biol Psychiatry. 2022; 92: 362–374.
7. Ikeda M, Takahashi A, Kamatani Y, Momozawa Y, Saito T, Kondo K, et al. Genome-Wide Association Study Detected Novel Susceptibility Genes for Schizophrenia and Shared Trans-Populations/Diseases Genetic Effect. Schizophr Bull. 2019;45(4):824–34.
8. Owen MJ, Legge SE, Rees E, Walters JTR, O'Donovan MC. Genomic findings in schizophrenia and their implications. Mol Psychiatry. 2023;28(9):3638–47.
9. Bergen SE, Ploner A, Howrigan D, Group CNVA, the Schizophrenia Working Group of the Psychiatric Genomics C, O'Donovan MC, et al. Joint Contributions of Rare Copy Number Variants and Common SNPs to Risk for Schizophrenia. Am J Psychiatry. 2019;176(1):29–35.
10. Chick SL, Holmans P, Cameron D, Grozeva D, Sims R, Williams J, et al. Whole-exome sequencing analysis identifies risk genes for schizophrenia. Nat Commun. 2025;16(1):7102.
11. Marshall CR, Howrigan DP, Merico D, Thiruvahindrapuram B, Wu W, Greer DS, et al. Contribution of copy number variants to schizophrenia from a genome-wide study of 41,321 subjects. Nat Genet. 2017;49(1):27–35.
12. Van der Auwera GA, Carneiro MO, Hartl C, Poplin R, Del Angel G, Levy-Moonshine A, et al. From FastQ data to high confidence variant calls: the Genome Analysis Toolkit best practices pipeline. Curr Protoc Bioinformatics. 2013; 43: 11.10.1–11.10.33.
13. Purcell S, Neale B, Todd-Brown K, Thomas L, Ferreira MA, Bender D, et al. PLINK: a tool set for whole-genome association and population-based linkage analyses. Am J Hum Genet. 2007;81(3):559–75.

14. Wang K, Li M, Hakonarson H. ANNOVAR: functional annotation of genetic variants from high-throughput sequencing data. Nucleic Acids Res. 2010;38(16):e164.
15. Chen S, Francioli LC, Goodrich JK, Collins RL, Kanai M, Wang Q, et al. A genomic mutational constraint map using variation in 76,156 human genomes. Nature. 2024;625(7993):92–100.
16. Liu D, Meyer D, Fennessy B, Feng C, Cheng E, Johnson JS, et al. Schizophrenia risk conferred by rare protein-truncating variants is conserved across diverse human populations. Nat Genet. 2023;55(3):369–76.
17. Trubetskoy V, Pardinas AF, Qi T, Panagiotaropoulou G, Awasthi S, Bigdeli TB, et al. Mapping genomic loci implicates genes and synaptic biology in schizophrenia. Nature. 2022;604(7906):502–8.
18. Koopmans F, van Nierop P, Andres-Alonso M, Byrnes A, Cijsouw T, Coba MP, et al. SynGO: An Evidence-Based, Expert-Curated Knowledge Base for the Synapse. Neuron. 2019;103(2):217–234. e4.
19. Fu JM, Satterstrom FK, Peng M, Brand H, Collins RL, Dong S, et al. Rare coding variation provides insight into the genetic architecture and phenotypic context of autism. Nat Genet. 2022;54(9):1320–31.
20. Boughton AP, Welch RP, Flickinger M, VandeHaar P, Taliun D, Abecasis GR, et al. LocusZoom.js: interactive and embeddable visualization of genetic association study results. Bioinformatics. 2021;37(18):3017–8.
21. Bigdeli TB, Chatzinakos C, Bendl J, Barr PB, Venkatesh S, Gorman BR, et al. Biological insights into schizophrenia from ancestrally diverse populations. Nature. 2026;651(8105):404–13.
22. Fromer M, Purcell SM. Using XHMM Software to Detect Copy Number Variation in Whole-Exome Sequencing Data. Curr Protoc Hum Genet. 2014; 81: 7.23.1–7.23.21.
23. Geoffroy V, Herenger Y, Kress A, Stoetzel C, Piton A, Dollfus H, et al. AnnotSV: an integrated tool for structural variations annotation. Bioinformatics. 2018;34(20):3572–4.
24. Browning BL, Tian X, Zhou Y, Browning SR. Fast two-stage phasing of large-scale sequence data. Am J Hum Genet. 2021;108(10):1880–90.
25. Browning BL, Zhou Y, Browning SR. A One-Penny Imputed Genome from Next-Generation Reference Panels. Am J Hum Genet. 2018;103(3):338–48.
26. Ge T, Chen CY, Ni Y, Feng YA, Smoller JW. Polygenic prediction via Bayesian regression and continuous shrinkage priors. Nat Commun. 2019;10(1):1776.
27. Gocke CB, Yu H. ZNF198 stabilizes the LSD1-CoREST-HDAC1 complex on chromatin through its MYM-type zinc fingers. PLoS One. 2008;3(9):e3255.
28. Connaughton DM, Dai R, Owen DJ, Marquez J, Mann N, Graham-Paquin AL, et al. Mutations of the Transcriptional Corepressor ZMYM2 Cause Syndromic Urinary Tract Malformations. Am J Hum Genet. 2020;107(4):727–42.
29. Johannesen KM, Gardella E, Linnankivi T, Courage C, de Saint Martin A, Lehesjoki AE, et al. Defining the phenotypic spectrum of SLC6A1 mutations. Epilepsia. 2018;59(2):389–402.

30. Singh T, Kurki MI, Curtis D, Purcell SM, Crooks L, McRae J, et al. Rare loss-of-function variants in SETD1A are associated with schizophrenia and developmental disorders. Nat Neurosci. 2016;19(4):571–7.
31. Lewis DA, Curley AA, Glausier JR, Volk DW. Cortical parvalbumin interneurons and cognitive dysfunction in schizophrenia. Trends Neurosci. 2012;35(1):57–67.
32. Nakazawa K, Zsiros V, Jiang Z, Nakao K, Kolata S, Zhang S, et al. GABAergic interneuron origin of schizophrenia pathophysiology. Neuropharmacology. 2012;62(3):1574–83.
33. Lisman J. Excitation, inhibition, local oscillations, or large-scale loops: what causes the symptoms of schizophrenia? Curr Opin Neurobiol. 2012;22(3):537–44.
34. Rees E, Han J, Morgan J, Carrera N, Escott-Price V, Pocklington AJ, et al. De novo mutations identified by exome sequencing implicate rare missense variants in SLC6A1 in schizophrenia. Nat Neurosci. 2020;23(2):179–84.
35. Kang HJ, Kawasawa YI, Cheng F, Zhu Y, Xu X, Li M, et al. Spatio-temporal transcriptome of the human brain. Nature. 2011;478(7370):483–9.

Table 1

Table 1 is available in the Supplementary Files section.

Figures

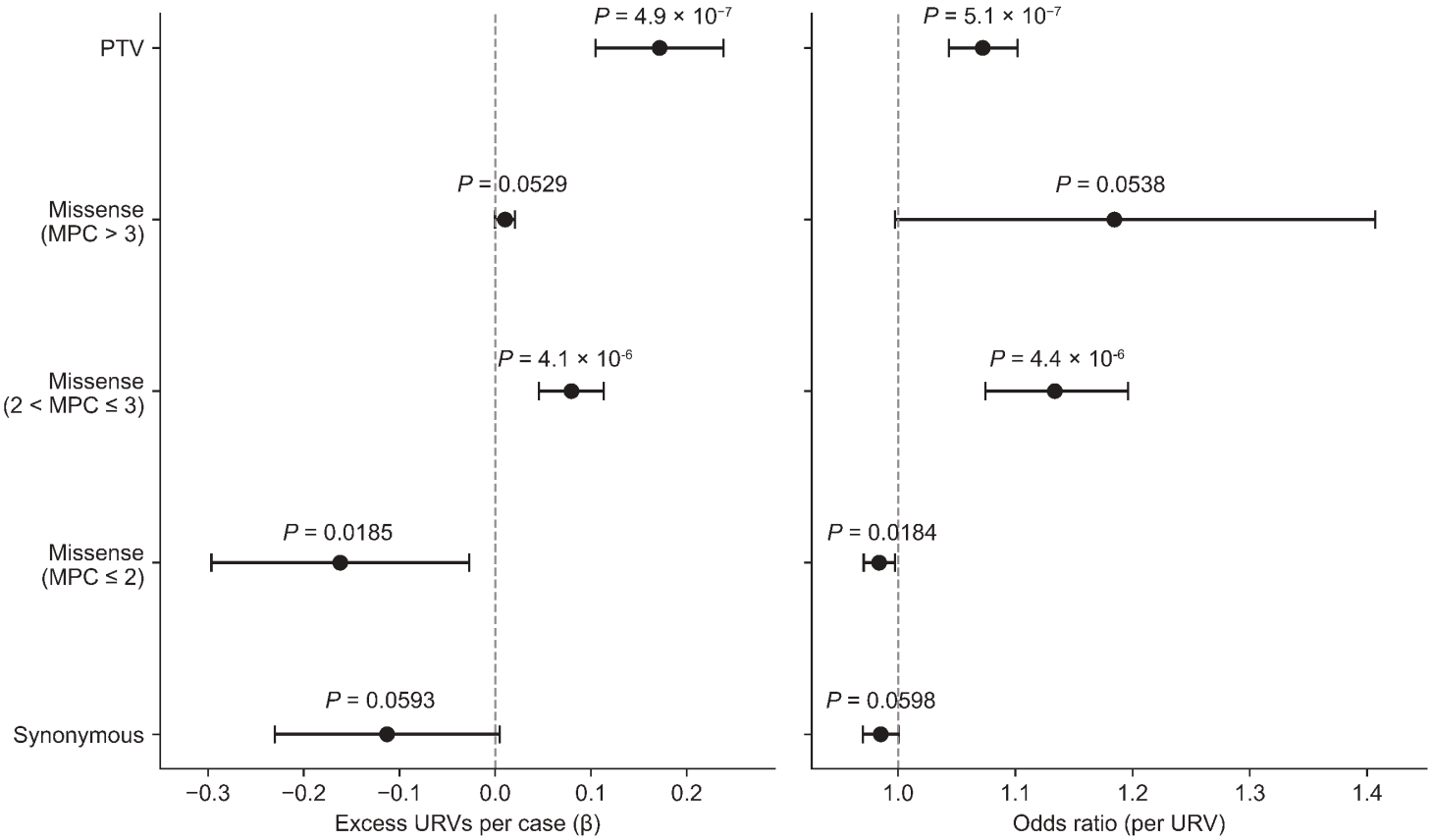

Figure 1

Exome-wide burden of ultra-rare variants in the Japanese schizophrenia cohort.

Forest plot showing the effect size of URVs across different functional annotation classes. The left x-axis represents the excess number of URVs per case (linear regression coefficient, β), and the right x-axis represents the corresponding OR. Points indicate the effect estimates derived from regression models adjusted for sex, the top five PCs and total URV counts, with error bars representing 95% confidence intervals. The vertical dashed line denotes an effect size of zero (β= 0, OR = 1). *P*-values are derived from the respective regression models.

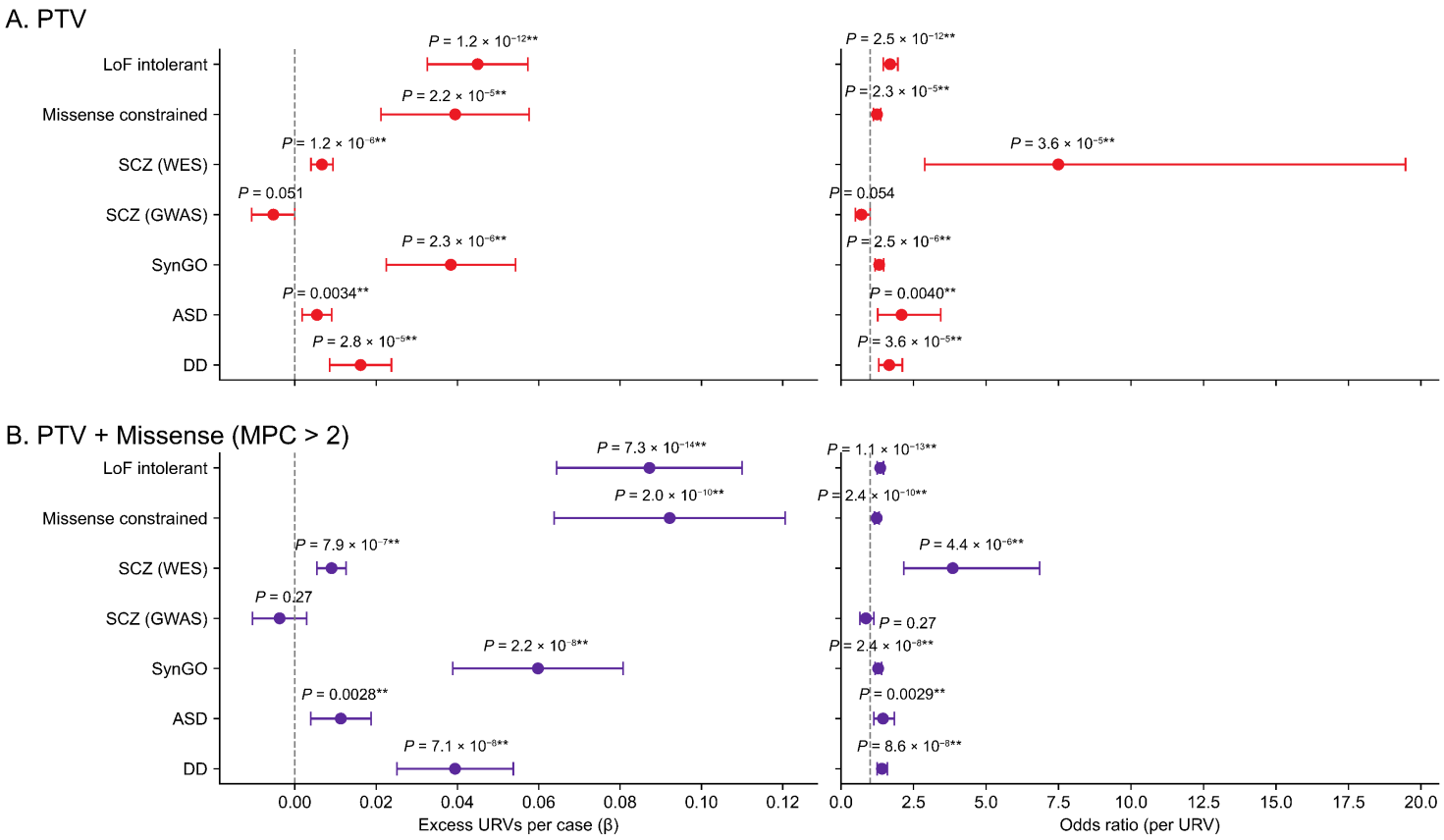

Figure 2

Enrichment of ultra-rare variants across functionally and genetically defined gene sets.

Burden of URVs in specific gene sets for (A) PTVs and (B) PTVs plus damaging missense variants (MPC > 2). The x-axes represent the excess number of URVs per case (linear β, left) and the OR (right). Evaluated gene sets include LoF intolerant genes (LOEUF < 0.35), missense constrained genes (Z-score > 3.09), previously implicated schizophrenia risk genes from WES (SCZ WES) and GWAS (SCZ GWAS), synaptic genes (SynGO), and genes implicated in ASD and DD. Points represent effect estimates with 95% confidence intervals (error bars). Statistical significance was assessed using regression models adjusted for sex, the top five PCs, and total URV counts, with asterisks denoting the FDR-adjusted *q*-values (* $q < 0.05$, ** $q < 0.01$).

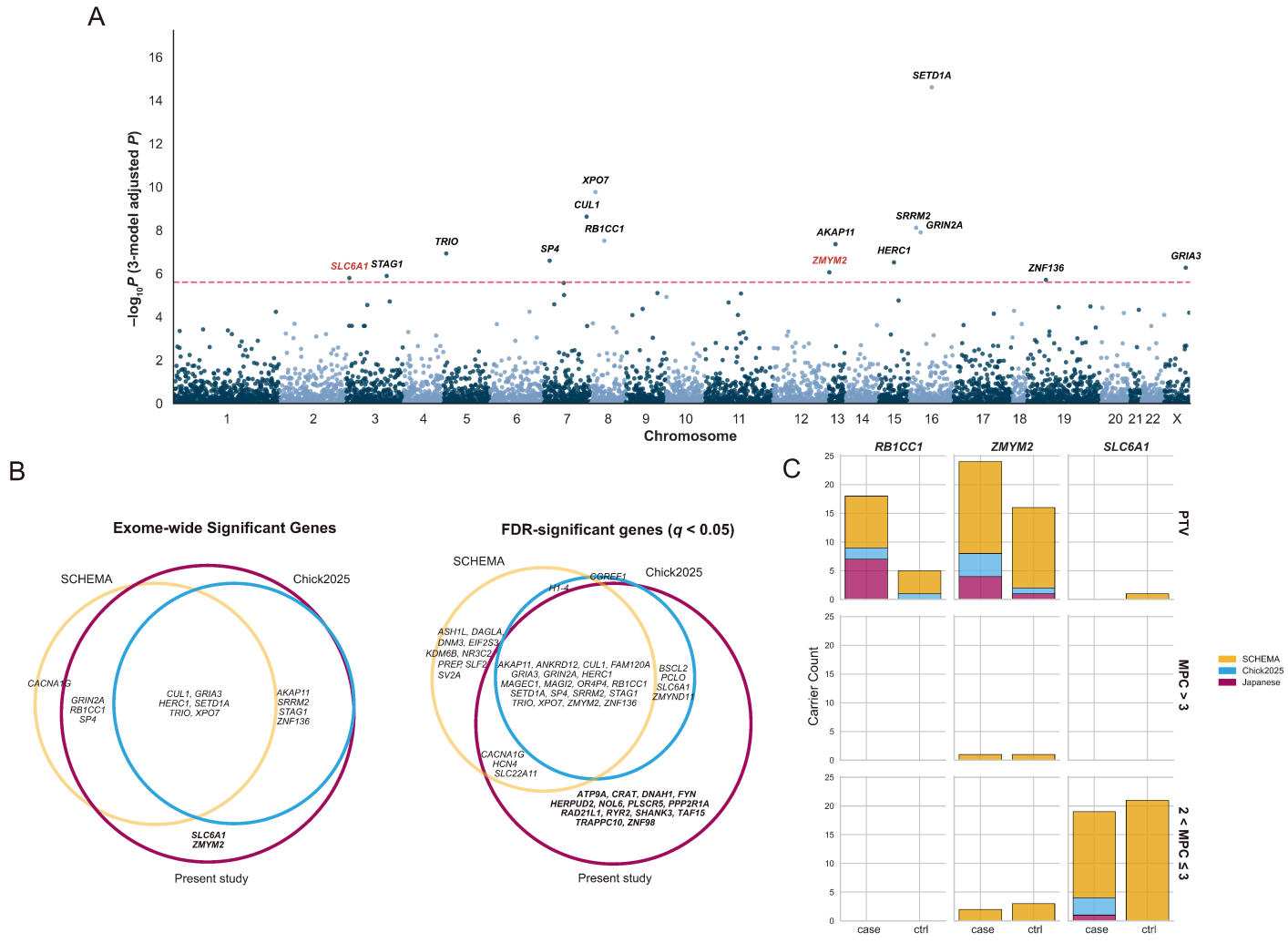

Figure 3

Cross-population meta-analysis identifies exome-wide significant schizophrenia risk genes.

(A) Manhattan plot of the cross-population meta-analysis incorporating the Japanese, SCHEMA, and the independent case-control cohort reported by Chick *et al*. (32,729 cases, 108,160 controls, and 3,444 proband-parent trios). The y-axis displays the $-\log_{10}$ model-adjusted meta-analysis *P*-values derived from cohort-stratified CMH tests with conditional integration of published de novo *P*-values. The horizontal dashed line indicates the exome-wide significance threshold ($P < 2.5 \times 10^{-6}$). Gene labels shown in red indicate genes that newly reached exome-wide significance in the present meta-analysis. **(B)** Venn diagrams illustrating the overlap of exome-wide significant genes (left) and FDR-significant genes (q < 0.05) (right) identified in the SCHEMA, Chick et al. (2025), and the present cross-population meta-analysis. **(C)** Qualifying carrier counts for *RB1CC1*, *ZMYM2*, and *SLC6A1*, stratified by variant class (PTV, missense MPC > 3, and missense 2 < MPC ≤ 3) and contributing cohort (SCHEMA, the independent Chick et al. cohort, and the Japanese cohort). *RB1CC1* is an established SCHEMA risk gene for which evidence was strengthened by the Japanese cohort; *ZMYM2* and *SLC6A1* newly reached exome-wide significance in the present meta-analysis.

(Note: The genetic data for the Japanese cohort were restricted to autosomes; thus, the association signals for X-linked genes, such as *GRIA3*, in the meta-analysis are derived exclusively from the SCHEMA and Chick et al. cohorts).

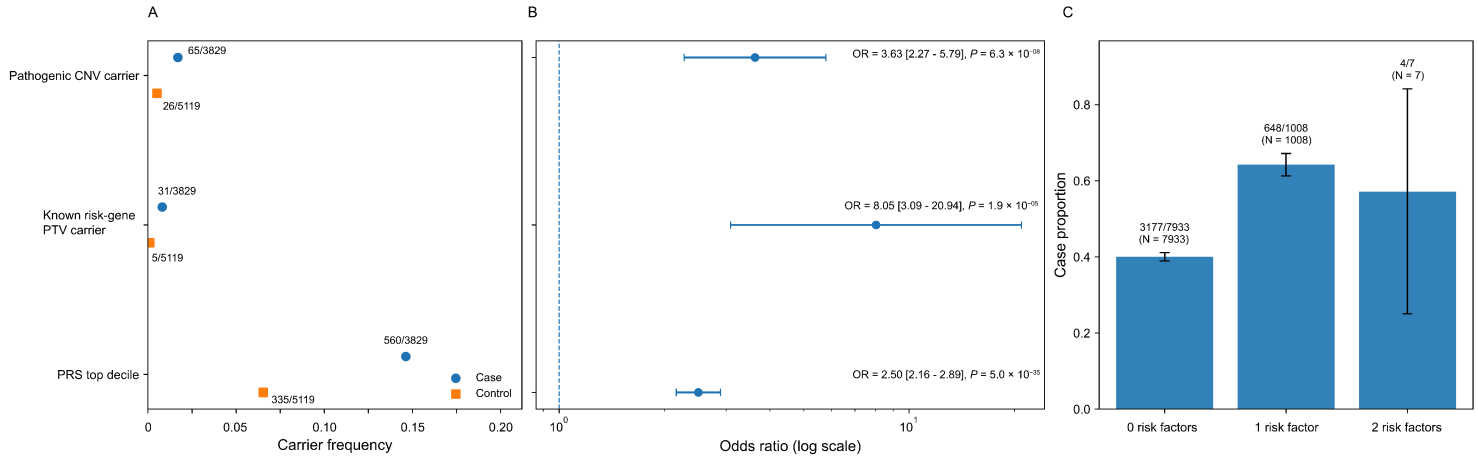

Figure 4

Joint associations of ultra-rare coding variants, structural variants, and polygenic risk.

(A) Carrier frequencies of pathogenic CNVs, known risk-gene PTVs, and the top decile of the primary PRS in Japanese schizophrenia cases and controls. Known risk-gene PTV carrier status was based on 13 autosomal genes among the 14 previously identified schizophrenia risk genes; the X-linked gene *GRIA3* was not evaluated because the Japanese exome analysis was restricted to autosomes. **(B)** Forest plot displaying the OR and 95% confidence intervals from a joint logistic regression model incorporating all three binary risk factors, adjusted for sex and the top five PCs. **(C)** Proportion of schizophrenia cases stratified by the cumulative number of genetic risk factors carried by an individual (0–2 risk factors). The exact number of cases and the total number of individuals (N) for each group are indicated above the respective bars. Error bars represent 95% confidence intervals calculated using the Wilson score interval. All panels were based on 8,948 participants with available PRS data (3,829 cases and 5,119 controls).

Supplementary Files

This is a list of supplementary files associated with this preprint. Click to download.

- SupplementaryTable1.SampleInformationv3.xlsx
- Table1.ExomewidesignificantGenes2608262.xlsx
- SupplementaryTable6.GenelevelBurdenfixed.xlsx
- SupplementaryTable4.SchizophreniaassociatedpathogenicCNVlociusedinthejointanalysis.xlsx
- SupplementaryTable8.MetaAnalysis260802.xlsx
- SupplementaryTable7.GenesetBurdenfixed.xlsx

- SupplementaryInformationSF20.pdf
- SupplementaryTable2.SamplelevelQualityControlandBaselineVariantStatistics.xlsx
- SupplementaryTable5.URVBurdenfixed.xlsx
- SupplementaryTable3.Geneset.xlsx