Supporting Information Appendix

High impact rare genetic variants in severe schizophrenia

Authors: Anthony W. Zoghbi, MD[1, 2,3,4,5]*, Ryan S. Dhindsa, PhD[2,4,6,7], Terry E. Goldberg, PhD[3,5,8], Aydan Mehralizade, MPH[3,5], Joshua E. Motelow, MD, PhD,[4,7,9] Xinchen Wang, PhD[4,7,10], Anna Alkelai, PhD[4,7], Matthew B. Harms, MD[4,11,12], Jeffrey A. Lieberman, MD[3,5], Sander Markx, MD[3,5†], David B. Goldstein, PhD[4,7*†]

Affiliations:

[1]Menninger Department of Psychiatry and Behavioral Sciences, Baylor College of Medicine, Houston, Texas, USA.

[2]Department of Molecular and Human Genetics, Baylor College of Medicine, Houston, Texas, USA.

[3]Department of Psychiatry, Columbia University Irving Medical Center, New York, New York; New York State Psychiatric Institute, New York, NY, USA.

[4]Institute of Genomic Medicine, Columbia University Irving Medical Center, New York, NY, USA.

[5]New York State Psychiatric Institute, Office of Mental Health, New York, NY, USA.

[6]Jan and Dan Duncan Neurological Research Institute, Houston, TX, USA.

[7]Department of Genetics and Development, Columbia University Irving Medical Center, New York, NY, USA.

[8]Department of Anesthesiology, Columbia University Irving Medical Center, New York, NY, USA.

[9]Department of Pediatrics, Division of Critical Care and Hospital Medicine, Columbia University Irving Medical Center, New York-Presbyterian Morgan Stanley Children's Hospital of New York, New York, NY, USA.

[10]Waypoint Bio, 180 Varick St, 6th Floor, New York, NY, USA.

[11]Department of Neurology, Columbia University Irving Medical Center, New York, NY, USA.

[12]Center for Motor Neuron Biology and Disease, Columbia University Irving Medical Center, New York, NY, USA.

Methods

Fig. S1. UMAP and cluster assignment of SETRS and controls
Fig. S2. UMAP and cluster assignment of typical schizophrenia and controls
Fig. S3. Burden of rare variants in tolerant genes
Fig. S4. SETRS gene-based rare variant collapsing analysis Q-Q plots
Fig. S5. UMAP and cluster assignment of SETRS and controls for genome-only analysis
Fig. S6. UMAP and cluster assignment of typical schizophrenia and controls for genome-only analysis
Fig. S7. Burden of rare variants in intolerant genes for genome-only analysis
Fig. S8. Burden of rare variants in tolerant genes for genome-only analysis
Fig. S9. SETRS gene-based rare variant collapsing analysis Q-Q plots for genome-only analysis
Fig. S10. Synonymous Q-Q plot for typical schizophrenia vs. controls
Fig. S11. Burden of rare variants in SETRS compared to typical schizophrenia and control individuals for genome-only analysis
Fig. S12. Synonymous Q-Q plot for typical schizophrenia vs. controls for genome-only analysis

Table S1. Control and typical schizophrenia phenotypes and sequencing type
Table S2. Case and control cohort ancestries
Table S3. Control phenotypes for genome-only analysis
Table S4. Case and control cohort ancestries for genome-only analysis
Table S5. Qualifying loss-of-function variants in Online Mendelian Inheritance in Man (OMIM) genes
Table S6. Qualifying missense variants in Online Mendelian Inheritance in Man (OMIM) genes
Table S7. Qualifying missense and loss-of-function variants with reported genotype-phenotype associations
Table S8. Top ten genes from gene-based collapsing models
Table S9. Qualifying Missense and Loss-of-Function Variants in Intolerant Genes

References

Methods

Severe, extremely treatment-resistant schizophrenia (SETRS) participant inclusion criteria

1. Provision of informed consent or assent with informed consent from a legally authorized representative. 2. Structured Clinical Interview for DSM-5 (SCID-5) diagnosis of schizophrenia. 3. Continuous residence in a New York State Office of Mental Health inpatient facility for ≥ 5 years. 4. Treatment resistance as defined by continued psychotic symptoms despite adequate trials of two or more antipsychotic medications for ≥ 6 weeks(48).

SETRS participant exclusion criteria

1. Patients hospitalized under Criminal Procedure Law (i.e. Not Guilty by Reason of Insanity). 2. Medical or neurologic disorder (e.g. stroke, traumatic brain injury) that predates schizophrenia onset and better explains symptomatology. 3. Greater than five-year length of stay primarily due to discharge refusal or high risk of violence, suicide, or sexual offense.

Diagnostic methods and phenotypic assessment

For SETRS participants, a psychiatrist (A.Z.) diagnosed schizophrenia using the psychosis module from the Structural Clinical Interview for DSM-5 based on a combination of interviews with the participant, collateral informants (primary clinical team and involved family members), and lifetime medical record review by a psychiatrist (A.Z.) and study coordinator (A.M.). Consensus diagnoses with a second psychiatrist (S.M.) was made for each participant.

We define severe, extremely treatment-resistant schizophrenia (SETRS) individuals as those whose severity of schizophrenia has required continuous hospitalization for at least 5 years. This 5-year cut off is based on prior research on “Kraepelinian” or “Very Poor Outcome” schizophrenia, which differs from typical schizophrenia in that it is characterized by worse

cognitive impairment, treatment resistance, stronger family history of schizophrenia spectrum disorders, and imaging abnormalities(56, 57). We elected not to assess for the presence of schizoaffective disorder (SAD) given the substantial genetic overlap between schizophrenia and schizoaffective disorder(58), low interrater reliability of SAD(59), and difficulty obtaining an accurate retrospective mood and neurovegetative symptom history from severely affected individuals. We calculated total duration of hospitalization in New York State Office of Mental Health state inpatient facilities using the New York State Facilities Enterprise Reporting system. We estimated age of onset of psychosis based on a combination of participant interview, collateral from family and clinical team, first documented hospitalization, and detailed medical record review.

For the typical schizophrenia analysis, we only selected individuals with a diagnosis of schizophrenia from the Genomic Psychiatry Cohort. Diagnostic and phenotyping methods for this cohort have been described in detail elsewhere(53). Thirty samples were ascertained as part of an exome sequencing study of schizophrenia and schizoaffective disorder that has been previously described in detail(54). For individuals ascertained as part of the whole-genome sequencing in psychiatric genetics at Massachusetts General and McLean Hospitals in Boston, the diagnostic process differed slightly for single proband samples versus samples recruited as part of a trio. For single probands, the clinical assessment was based on a review of all hospital records by two independent senior clinicians, which was then used to provide a consensus diagnosis of schizophrenia or schizoaffective disorder based on DSM-5 criteria. For participants enrolled as part of a trio, probands were assessed based on a SCID interview and a review of all hospital records (where applicable). Then, a consensus Axis I DSM-IV diagnoses of schizophrenia or schizoaffective disorder was assigned by a group of four senior clinicians based

on a review of all material. Of note, each of the typical schizophrenia cohorts include some individuals who are treatment resistant. However, given that 30% of all individuals with schizophrenia experience treatment resistance(48), this is not inconsistent with a typically ascertained cohort.

Whole genome/exome sequencing and variant calling

Samples of blood-extracted SETRS individuals and whole genome controls were sequenced according to standard protocols on Illumina's NovaSeq 6000 (Illumina, San Diego, CA, USA) platform with 150 base pair paired-end reads at Columbia's Institute for Genomic Medicine (IGM) and New York Genome Center (NYGC). Typical schizophrenia individuals' whole genome samples were sequenced on a combination of NovaSeq 6000 or HiSeq2500 at the IGM or the HiSeq X at the NYGC. Whole genome typical schizophrenia samples from dbGaP were sequenced at the Broad Institute on Illumina HiSeq2500. Exome samples were sequenced with IDT Exome Research Panel Version 1, Roche NimbleGen EZCap Version 3, Illumina TruSeq 65MB, or Agilent SureSelect Human All Exon V5 capture kits using Illumina GAIIx, HiSeq 2000, or HiSeq 2500 sequencers following standard protocols.

We aligned reads to human reference GRCh37 using DRAGEN (Edico Genome, San Diego, CA, USA)(61) and duplicates were marked and removed using Picard tools (Broad Institute, Boston, MA, USA). Variants were called as per the Genome Analysis Toolkit (GATK - Broad Institute, Boston, MA, USA) Best Practices recommendations v3.6(62). Variant type/function were annotated with ClinEff(63) and the IGM's in-house Analysis Tool for Annotated Variants(64) was used to add custom annotations including gnomAD v2.1 frequencies, regional-intolerance metrics, and clinical annotations provided by the Human Gene

Mutation Database (HGMD)(65), ClinVar(66), and Online Mendelian Inheritance in Man (OMIM)(29).

Sample and variant quality control and filtering

The initial samples included 116 SETRS individuals, 230 typical schizophrenia individuals, and 5,950 controls for the combined whole exome and genome analysis. For the genome-only analysis, we included 116 SETRS individuals, 200 typical schizophrenia, and 4,817 controls. We removed samples with a discordance between sequence-derived X:Y coverage ratio and their self-reported gender and >5% contamination according to VerifyBamID(67). Case and control cohorts were screened for cryptic relatedness with KING to remove related individuals (second-degree relatives or closer in relatedness) electing to retain cases over controls in each pair. Samples were also removed if they had less than 90% 10-fold coverage of the consensus coding sequence (CCDS) regions (release 20). SETRS (97.6%), typical schizophrenia (97.2%), and whole exome/genome controls (95.6%) genome-only controls (97.5%) were similar in the percentage of CCDS bases with 10-fold coverage. Control samples were also removed if they were included in the gnomAD database or were sequenced using a capture kit with insufficient coverage of the exome. The final SETRS cohort included 112 SETRS individuals, 218 typical schizophrenia individuals, and 4,929 controls. The final genome-only cohort included 114 SETRS individuals, 198 typical schizophrenia individuals, and 4,146 controls.

Variant calls were required to have at least 10-fold coverage, quality-by-depth score (QD) ≥ 5, quality score (QUAL) ≥ 50, mapping quality score (MQ) ≥ 40, genotype quality score (GQ) ≥ 20, read position rank sum (RPRS) ≥ -3, mapping quality rank sum (MQRS) ≥ -10. Single nucleotide variants (SNVs) were required to have Fisher's strand bias (FS) of < 60 and

Strand Odds Ratio (SOR) < 3. Indels were required to have FS < 200 and SOR < 10. Variants were removed if they did not pass GATK's standard variant quality score recalibration (VQSR) threshold. We also removed known sequencing artifacts as described in Petrovski et al. 2017(51) and variants previously identified as problematic by ExAC(69), gnomAD(21), or EVS (http://evs.gs.washington.edu/EVS/HelpDescriptions.jsp). For heterozygous genotypes, the alternative allele ratio (allelic balance) was required to be ≥ 30%. We restricted the analyses to variants within the CCDS or the 2 base pair canonical splice sites.

To remove confounding due to differential coverage between cases and controls, we used site-coverage harmonization as previously described(51). Briefly, we removed any bases that had a greater than 7% absolute difference in 10-fold coverage between cases and controls as coverage differences between cases and controls could potentially introduce bias. This coverage harmonization technique reduces potential bias that can stem from differences in sequencing depth, coverage between sequencers, and differential coverage between whole genome sequencing and whole exome sequencing capture kits.

We externally filtered for variants with a minor allele frequency (MAF) $\leq 1 \times 10^{-4}$ in each represented ancestral population in the non-psychiatric ("non-neuro") subset of gnomAD. We then filtered using internal minor allele frequency of 0.0005 to control for site and sequencer-specific artifacts as previously described(52). We only included SNVs and indels receiving "PASS" for the gnomAD random forest filter. We then removed variants falling under the recommended cutoffs for the gnomAD random forest true probability for SNVs: 0.1 (exome)/0.4 (genome) and indels: 0.2 (exome)/0.4 (genome). We removed loss-of-function variants flagged as low confidence by LOFTEE and those falling in low complexity, segmental duplication, and decoy regions as defined by gnomAD(21). We filtered out loss-of-function variants expressed in

less than 10% of transcripts using the proportion expressed across transcripts (PEXT) score(70) with the recommended cutoff of 0.1. To normalize PEXT values across a gene, we divided each base pair's PEXT score by the maximum PEXT value for that gene to create a PEXT ratio. This procedure accounts for the spuriously low PEXT values across certain genes in gnomAD (e.g. *CNTNAP2*). Qualifying loss-of-function variants included rare variants that passed the above quality control metrics and were annotated by ClinEff as stop gain, canonical splice donor/acceptor, or frameshift variants.

Missense variants were filtered for pathogenicity/deleteriousness using the rare exome variant ensemble learner (REVEL) tool(23). We used the recommended cutoff of > 0.5 to classify variants as "damaging." We selected a cutoff of < 0.15 to classify variants as "benign," corresponding to a false-positive rate of 5%. We further filtered missense variants using the missense tolerance ratio (MTR) tool(25). MTR identifies regional intolerance to missense variation, thereby highlighting functionally important genic sub-regions. We used the most conservative estimate of regional intolerance, MTR FDR, with the authors' recommended cutoff of MTR FDR < 0.1(25). Qualifying missense variants for gene set burden analyses were rare variants that passed the above quality control metrics with REVEL > 0.5 and MTR FDR < 0.1.

Gene set curation

To create the missense and loss-of-function intolerant gene sets, we used the recommended cutoffs of pLI > 0.9 and missense Z score > 3.09 based on constraint metrics from gnomAD v2.1 (https://storage.googleapis.com/gnomad-public/release/2.1.1/constraint/gnomad.v2.1.1.lof_metrics.by_gene.txt.bgz). Our missense and

loss-of-function tolerant gene sets included genes with missense Z < 1 and pLI < 0.001, respectively.

We created our OMIM gene set based on OMIM data downloaded on November 21st, 2019. We began with a list of all protein-coding OMIM genes (4,231 genes) and removed any gene with an “equivocal,” “non-disease,” or “susceptibility” tag, which resulted in an established disease gene set of 3,509 genes. Of these 3,509 genes, 3,114 have a detailed clinical synopsis with phenotypic annotations across organ systems in OMIM. To create our OMIM “behavioral” gene set, we took the subset of established disease genes that had the requisite annotation in OMIM: “neurologicBehavioralPsychiatricManifestations.” We reviewed all phenotypic annotations in this list and removed any non-behavioral annotations including "ALDH2*2 carriers suffer more severe hangovers | Increased intoxicating symptoms after alcohol consumption", and "No behavior problems." This filtering produced a list of 498 behaviorally-annotated OMIM genes. The remaining 2616 phenotypically annotated genes in OMIM were used to create our “Non-Behavioral OMIM” gene set.

As most of the rare variant burden in schizophrenia and our sample lies in intolerant genes, we restricted our OMIM and behavioral OMIM analyses to our previously described intolerant gene sets of pLI > 0.9 and missense Z score > 3.09. This resulted in final gene sets of OMIM loss-of-function intolerant: 805 genes, OMIM Missense Intolerant: 350 genes, OMIM Behavioral loss-of-function Intolerant: 226 genes, Non-Behavioral OMIM loss-of-function intolerant: 511 genes, OMIM Behavioral Missense Intolerant: 127 genes, Non-Behavioral OMIM missense intolerant: 201 genes.

To create the autism spectrum disorder and developmental delay gene set, we combined the most recently published genome-wide significant autism spectrum disorder (n=102) and

developmental delay (n=299) genes for a total of 340 unique genes in the gene set(6, 7). We tested for a case-control burden of qualifying missense and loss-of-function variants in this combined gene set. To create the gene set of genes previously implicated in typical schizophrenia, we used data from (accessed October 7th, 2019) the Schizophrenia Exome Meta-Analysis Consortium (SCHEMA) browser (https://schema.broadinstitute.org/) (accessed October 7th, 2019), which published specific *p* values for missense and loss-of-function enrichment in nearly 25,000 typical schizophrenia cases and 100,000 controls. We used these variant specific *p* values from SCHEMA rather than the meta-analysis *p* values (which combine counts of missense and loss-of-function variants) because of importance matching variant type with the genetic mechanism (e.g. analyzing missense variants in missense-driven gene). As there are only 10 genome-wide significant genes in schizophrenia, we elected to use *p* values of nominal significance for missense-driven ($p < 0.05$, n=144) and loss-of-function-driven ($p < 0.01$, n=198) SCHEMA genes to create roughly equivalently sized gene sets. We also filtered this list by removing any genes that reached nominal significance due to enrichment in controls as opposed to cases. As previously described, we further restricted these two gene sets to their respective intolerant subset with pLI > 0.9 for loss-of-function-driven SCHEMA genes (n=94) and missense Z > 3.09 for missense-driven SCHEMA genes (n=45).

Quantile-quantile plots and Genomic Inflation Factor (λ)

Quantile-quantile (QQ) plots were generated using a previously described method (19, 51). Observed *p* values were generated for each gene for each model using the Cochran-Mantel-Haenszel test as described above (see Gene-based rare variant collapsing analysis). Expected *p* values and 95% confidence intervals were generated in the following manner. Briefly, for each

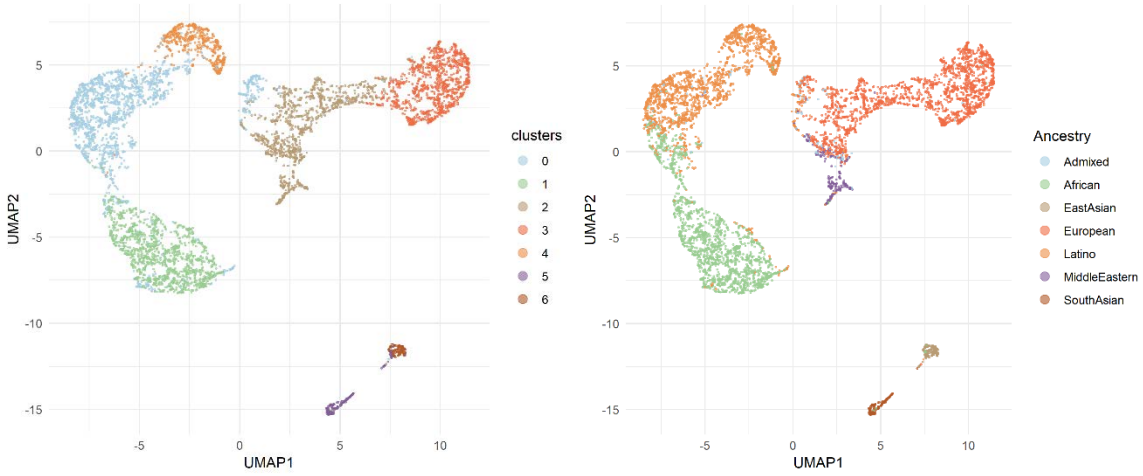
collapsing model within each cluster, case and control labels were randomly permuted while keeping the qualifying variant gene-by-sample matrix fixed. We then used the CMH test to compare the new case and controls with and without qualifying variants in each gene to test for an association for case vs. control status. We repeated this process 1,000 times and used the mean rank-ordered *p* values as empirical estimates of the expected *p* values. We then generated QQ plots using the negative logarithm of the permutation-based expected vs. observed distribution of the *p* values. The genomic inflation factor (λ) was estimated from the permutation-based expected *p* values using a regression method previously described (19, 51).

Clinical diagnostic analysis

For our diagnostic analysis, we analyzed single nucleotide and insertion/deletion variants within the CCDS and 2 base pair splice sites in SETRS individuals only. Variant quality control was as described above with several exceptions. We allowed a lower allelic balance (25%) and did not perform site coverage harmonization as it is irrelevant for diagnostic analysis. Variants were filtered for frequency with allele counts of less than five in gnomAD. We also analyzed all variants annotated as potentially damaging as defined by ClinEff as opposed to only missense and loss-of-function variants.

We then prioritized variants that have previously been reported as pathogenic/likely pathogenic or affect the same amino acid as a pathogenic variant. Our pipeline incorporates curated data from ClinVar(66), Human Gene Mutation Database (HGMD), and internal IGM cases to annotate all variants previously reported as pathogenic. We also prioritize loss-of-function variants (stop gain/lost, start lost, splice site acceptor/donors, frameshift indels) in genes with known pathogenic loss-of-function variants or reported as haploinsufficient in ClinGen(71).

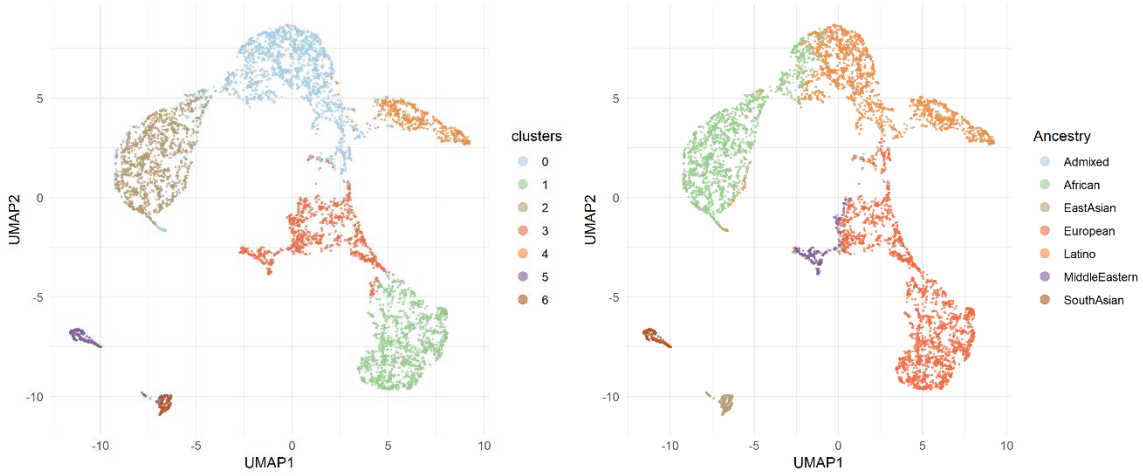
We also individually reviewed all qualifying missense and loss-of-function variants in OMIM disease genes (Table S5 and S6). We then evaluated each candidate variant using ACMG guidelines(30) with respect to their schizophrenia diagnosis only. As such, we did not report pathogenic variants for unrelated conditions (e.g. carrier for Cystic Fibrosis) for this manuscript.



Cluster	Cases	Controls
0	13	1709
1	37	1205
2	32	1009
3	30	1006
4	3	510
5	NA	170
6	1	136

Fig. S1. UMAP and cluster assignment of SETRS and controls

UMAP and cluster assignments with genetically determined ancestry of SETRS cases and controls. Clusters highlighted in red were excluded from the analysis due to insufficient sample size.



Cluster	Cases	Controls
0	8	1705
1	153	1102
2	8	1204
3	49	913
4	6	515
5	7	170
6	NA	136

Fig. S2. UMAP and cluster assignment of typical schizophrenia and controls

UMAP and cluster assignments with genetically determined ancestry of typical schizophrenia cases and controls. Clusters highlighted in red were excluded from the analysis due to insufficient sample size and lack of corresponding SETRS cluster.

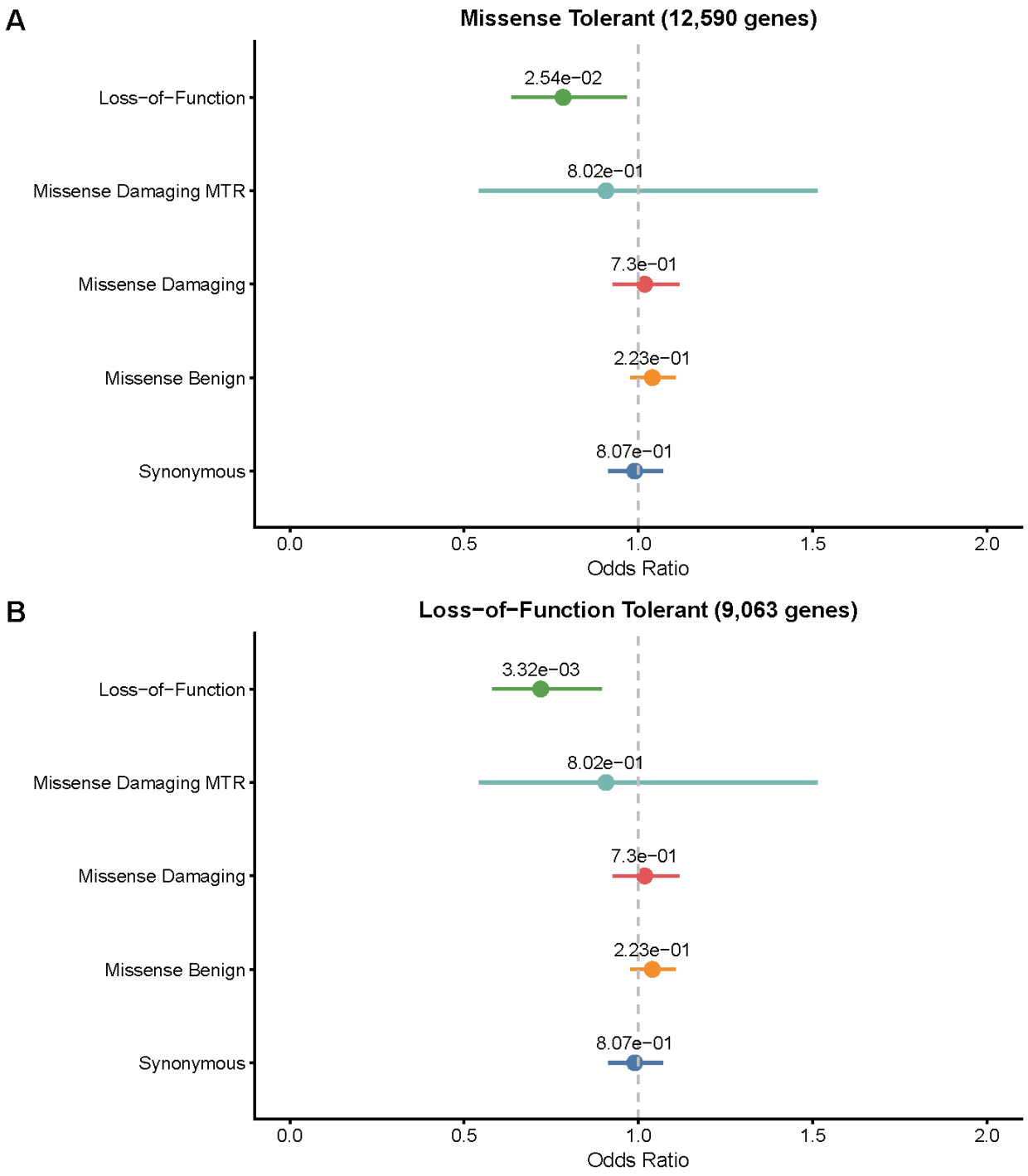
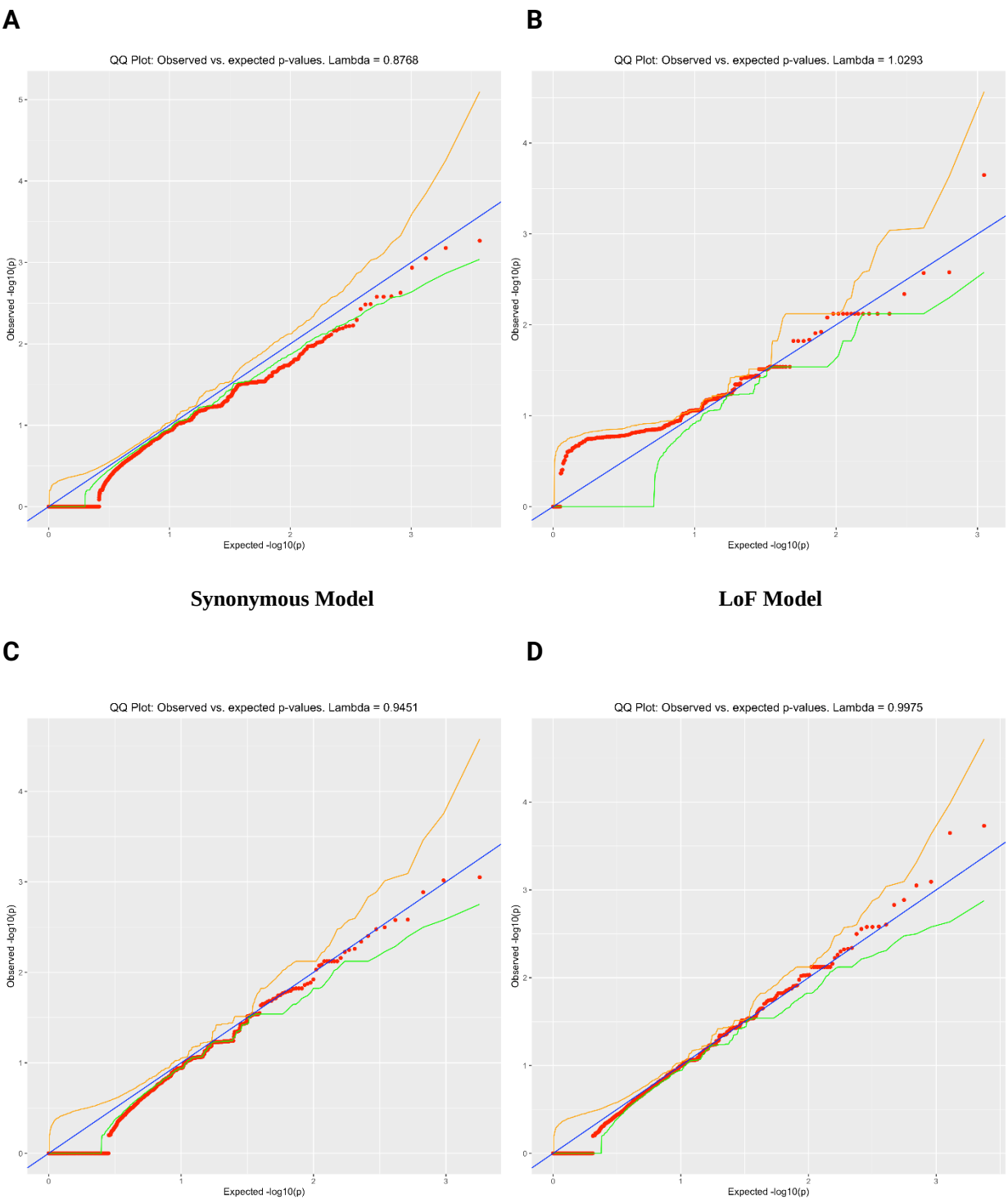


Fig. S3. Burden of rare variants in tolerant genes

The burden of rare variants in missense and loss-of-function (LoF) tolerant genes in 112 SETRS cases and 4,929 controls. **A.** Missense tolerant genes were those with missense Z < 1 and **B.** LoF tolerant genes were those with probability of being loss-of-function intolerant (pLI) score <

0.001 in gnomAD. Missense benign: REVEL score < 0.15, Missense damaging: REVEL score > 0.5, Missense damaging MTR: REVEL score > 0.5 and missense tolerance ratio false discovery rate (MTR FDR) < 0.1. Unadjusted two-sided Cochran–Mantel–Haenszel exact *p* values and horizontal bars indicating 95% confidence intervals are shown.



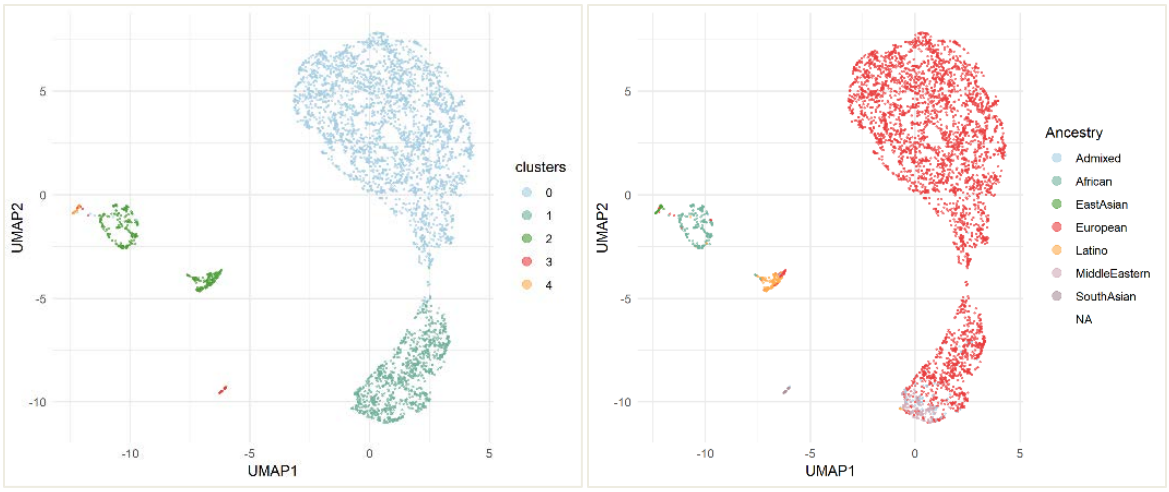
Synonymous Model

LoF Model

Damaging Missense (REVEL>0.5) Model

Damaging Missense + LoF Model

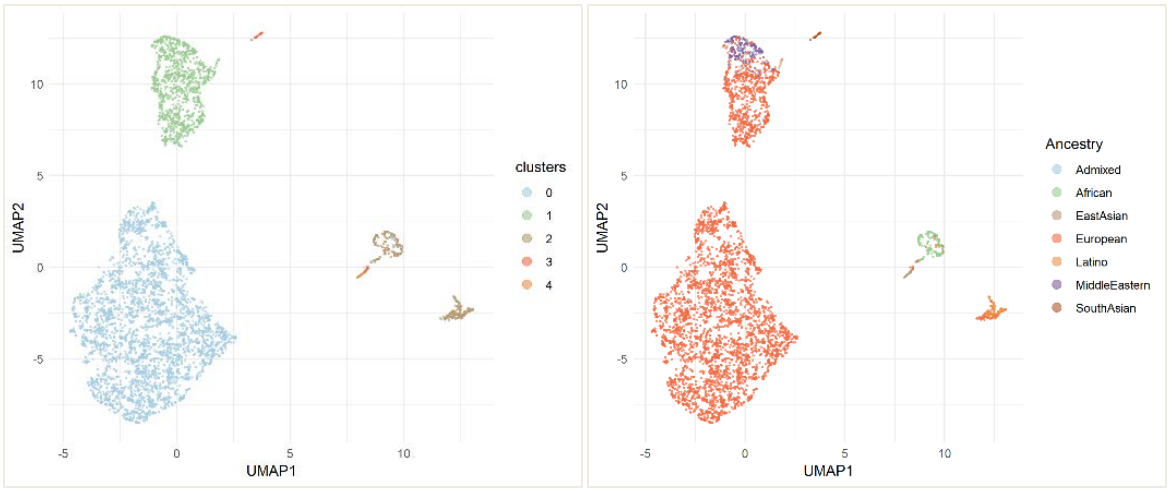
Fig. S4. SETRS gene-based rare variant collapsing analysis Q-Q plots



Cluster	Cases	Controls
0	28	2937
1	33	925
2	53	284
3	1	45
4	1	31

Fig. S5. UMAP and Cluster Assignment of SETRS and controls for genome-only analysis

UMAP and cluster assignments with genetically determined ancestry of SETRS cases and whole genome sequenced controls. Clusters highlighted in red were excluded from the analysis due to insufficient sample size.



Cluster	Cases	Controls
0	136	2939
1	48	922
2	14	283
3	2	47
4	NA	31

Fig. S6. UMAP and Cluster Assignment of typical schizophrenia and controls for genome-only analysis

UMAP and cluster assignments with genetically determined ancestry of typical schizophrenia cases and whole genome sequenced controls. Clusters highlighted in red were excluded from the analysis due to insufficient sample size.

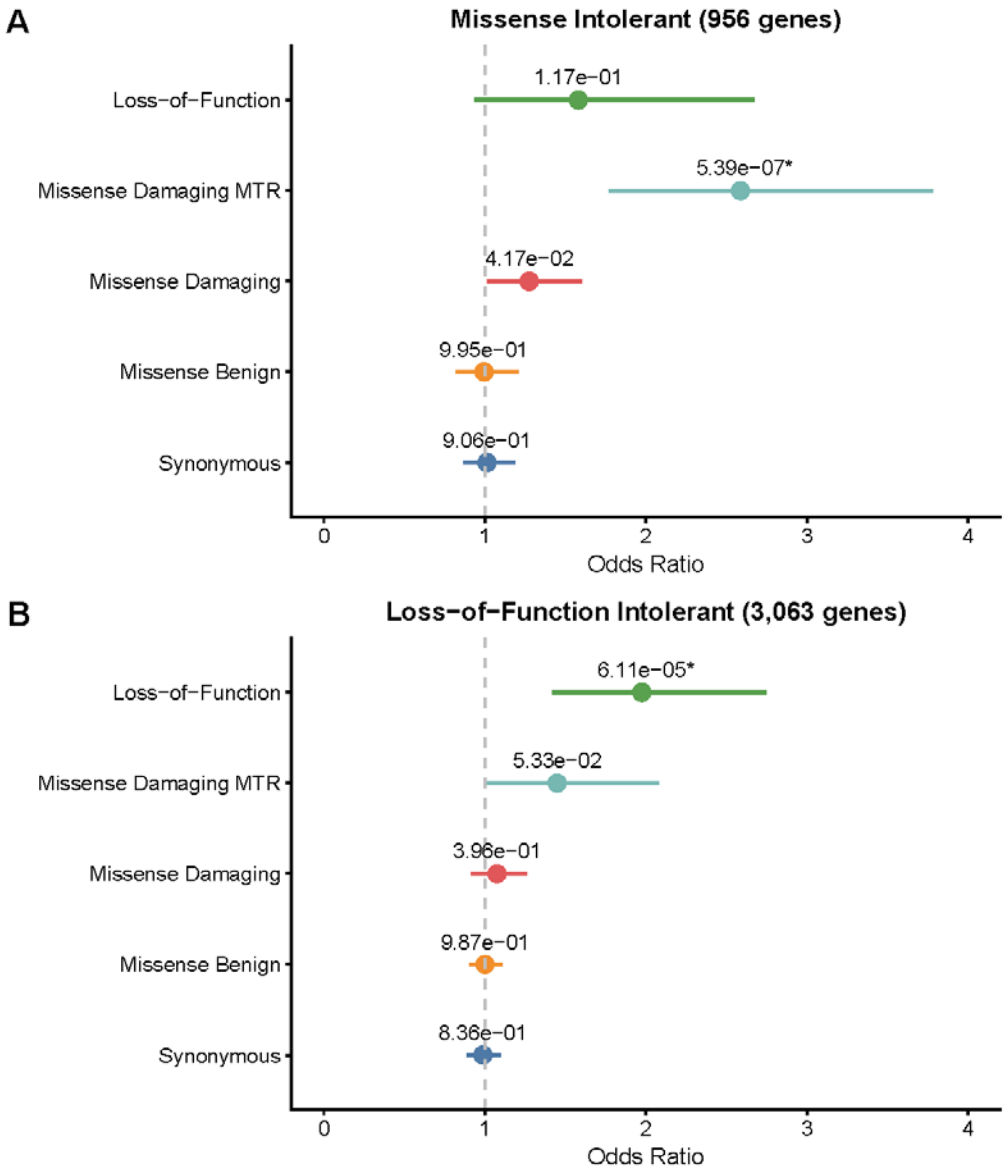


Fig. S7 Burden of rare variants in intolerant genes for genome-only analysis

The burden of rare variants in missense and loss-of-function intolerant genes in 114 SETRS cases and 4,146 whole genome sequenced controls. (**A**) Missense intolerant genes were those with missense Z > 3.09 and (**B**) loss-of-function intolerant genes were those with probability of being loss-of-function intolerant (pLI) score > 0.9 as per gnomAD. Missense benign: REVEL score < 0.15, Missense damaging: REVEL score > 0.5, Missense damaging MTR: REVEL score > 0.5 and missense tolerance ratio false discovery rate (MTR FDR) < 0.1. Unadjusted two-sided Cochran–Mantel–Haenszel exact *p* values and horizontal bars indicating 95% confidence intervals are shown. *: FDR < 0.1

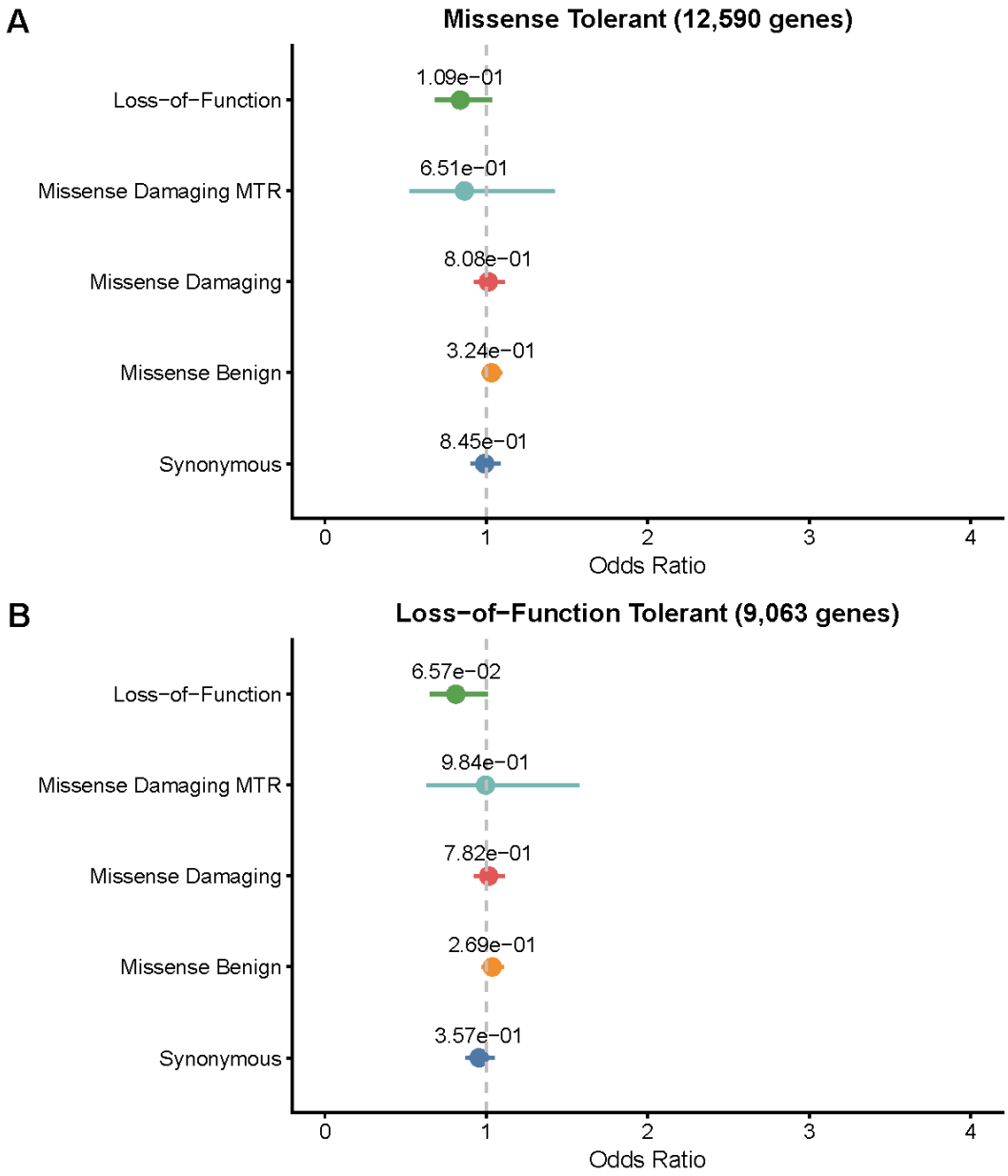
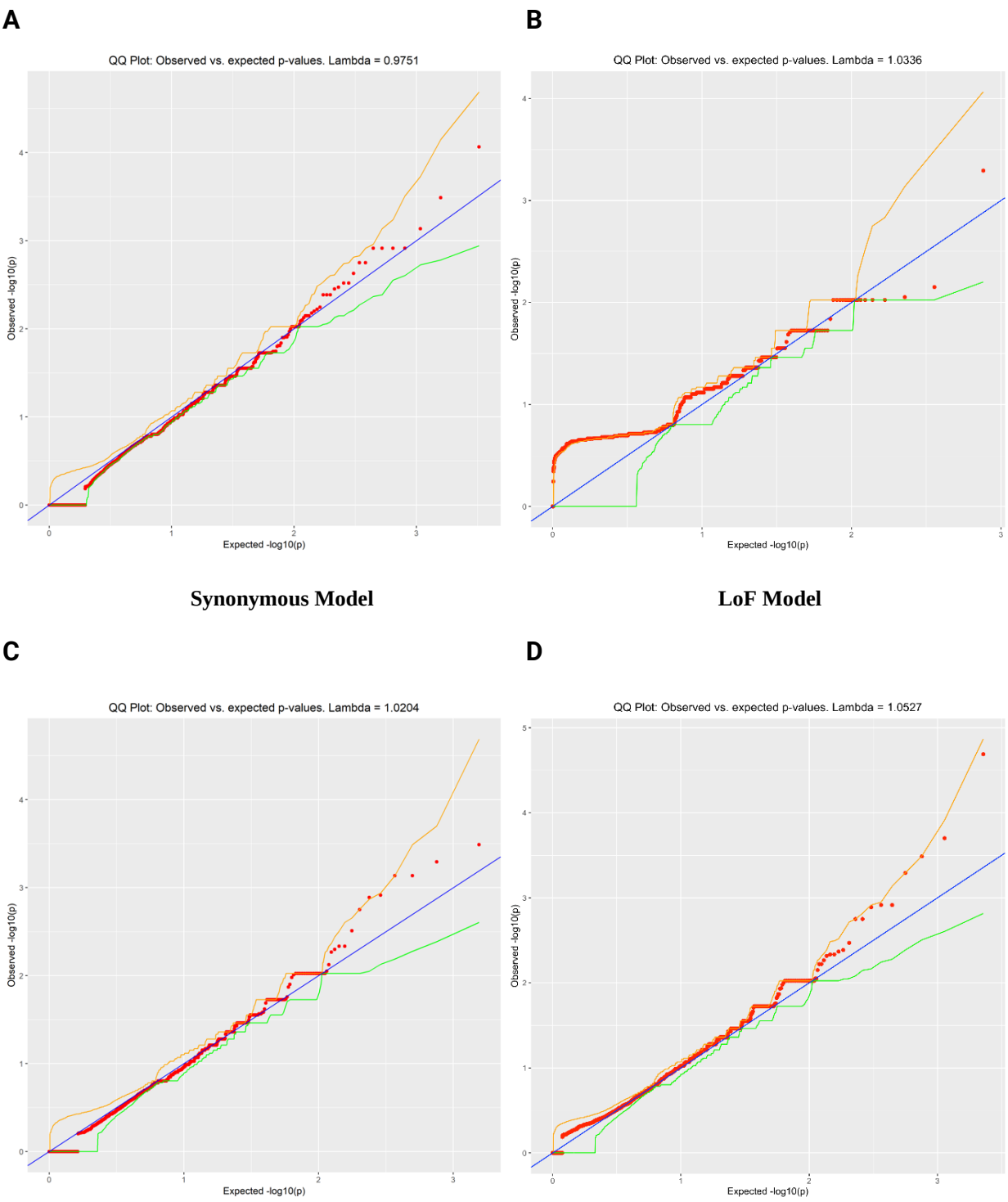


Fig. S8. Burden of rare variants in tolerant genes for genome-only analysis

The burden of rare variants in missense and loss-of-function (LoF) tolerant genes in 114 SETRS cases and 4,146 whole genome sequenced controls. **A.** Missense tolerant genes were those with missense Z < 1 and **B.** LoF tolerant genes were those with probability of being loss-of-function intolerant (pLI) score < 0.001 in gnomAD. Missense benign: REVEL score < 0.15, Missense damaging: REVEL score > 0.5, Missense damaging MTR: REVEL score > 0.5 and missense tolerance ratio false discovery rate (MTR FDR) < 0.1. Unadjusted two-sided Cochran–Mantel–Haenszel exact *p* values and horizontal bars indicating 95% confidence intervals are shown.



Synonymous Model **LoF Model**

Damaging Missense (REVEL>0.5) Model **Damaging Missense + LoF Model**

Fig. S9. SETRS gene-based rare variant collapsing analysis Q-Q plots for genome-only analysis

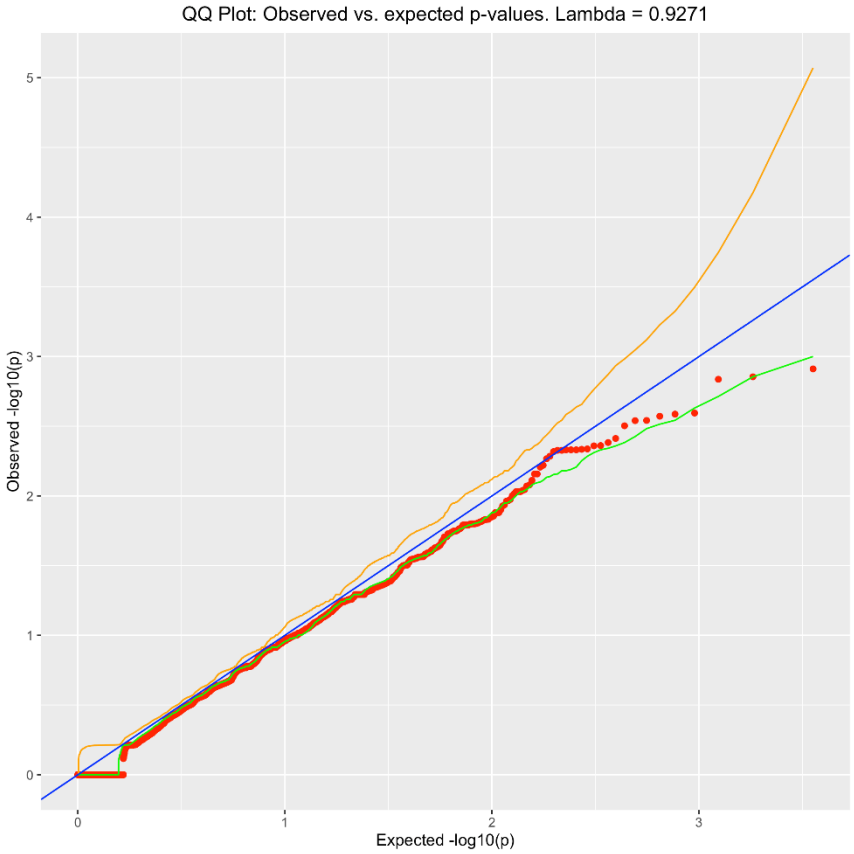


Fig. S10. Synonymous Q-Q plot for typical schizophrenia vs. controls

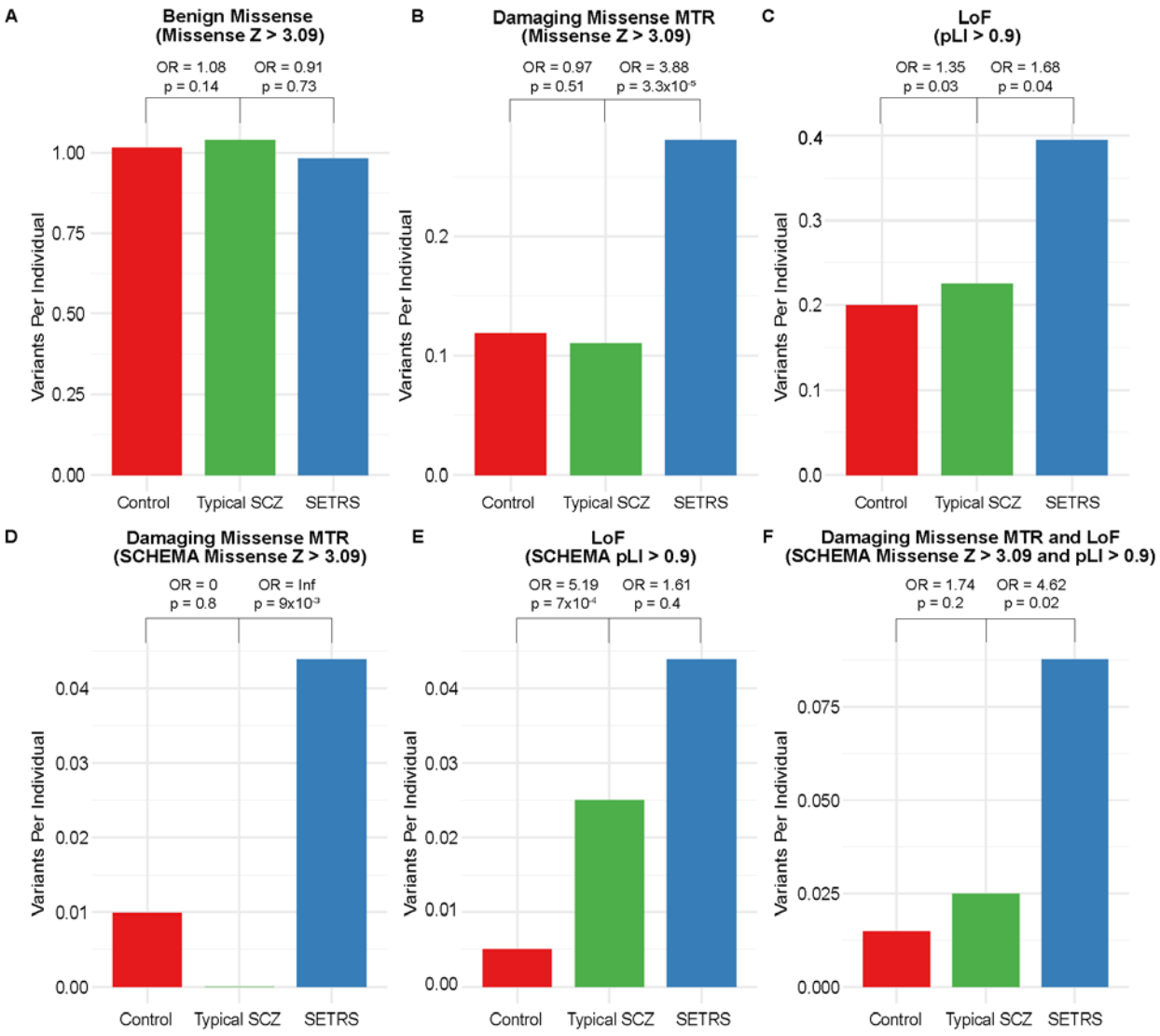


Fig. S11. Burden of rare variants in SETRS compared to typical schizophrenia and control individuals for genome-only analysis

(**A**) No difference observed between SETRS, typical SCZ, and controls for benign missense variants. (**B)** Increased burden of damaging missense MTR variants in SETRS compared to typical SCZ in missense intolerant genes. (**C**) Increased burden of loss-of-function variants in SETRS compared to typical SCZ in loss-of-function intolerant genes. (**D**) Increased burden of damaging missense MTR variants in SETRS compared to typical SCZ in missense intolerant SCHEMA genes. (**E**) Nonsignificant increase in burden of loss-of-function variants in SETRS compared to typical SCZ in loss-of-function intolerant SCHEMA genes. (**F**) Increased burden of damaging missense MTR and loss-of-function in SETRS compared to typical SCZ in intolerant

SCHEMA genes. Odds ratios (OR) and *p* values shown are based on a comparison of typical SCZ individuals to controls and SETRS individuals to typical SCZ using one-sided Cochran–Mantel–Haenszel test. Benign missense: missense variants with REVEL score < 0.15, Missense damaging MTR: missense variants with REVEL score > 0.5 and missense tolerance ratio false discovery rate (MTR FDR) < 0.1.

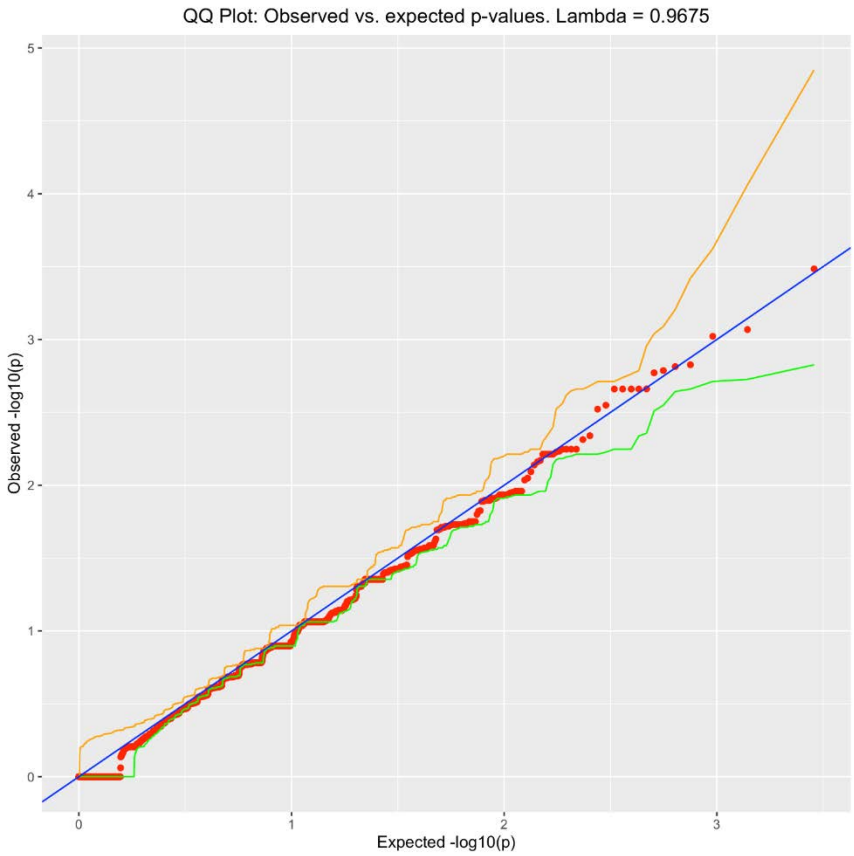


Fig. S12. Synonymous Q-Q plot for typical schizophrenia vs. controls for genome only analysis

Table S1. Control and typical schizophrenia phenotypes and sequencing type

Phenotype	Whole Exome Sequencing	Whole Genome Sequencing
Healthy Control	2637	33
Healthy Family Member of Proband with Non-Psychiatric Illness	2207	52
Typical Schizophrenia	25	193

Table S2. Case and control cohort ancestries

Cohort	Admixed	African	European	Latino	Middle Eastern
SETRS	15	39	50	8	0
Typical Schizophrenia	13	10	189	5	1
Controls	473	1590	1700	1121	45

Table S3. C**ontrol phenotypes for genome-only analysis**

Phenotype	**N**
IgA nephropathy	247
Membranous Nephropathy	239
Nephritis	93
Parents of children with posterior urethral valves	161
Posterior Urethral valves	7
Adults with HIV	118
Adults with amyotrophic lateral sclerosis	3,281
Total	**4,146**

Table S4. Case and control cohort ancestries for genome-only analysis

Cohort	Admixed	African	European	Latino	Middle Eastern
SETRS	14	39	50	11	0
Typical Schizophrenia	10	2	176	9	1
Controls	272	145	3627	101	1

Table S5. Qualifying Loss-of-Function Variants in Online Mendelian Inheritance in Man (OMIM) Genes

Gene	Variant	OMIM Phenotype (Inheritance)	ACMG determination and rationale
C3	19-6678035-C-T	C3 Deficiency (AR)	VOUS - AR disease
CSNK1D	17-80207447-C-CG	Advanced Sleep Phase, Familial (AD)	VOUS- Missense driven disease
DIS3L2	2-233001277-CCT-C	Perlman Syndrome (AR)	VOUS - AR disease
DST	6-56498989-C-CT	Neuropathy, sensory/autonomic (AR) and Epidermolysis bullosa simplex (AR)	VOUS - AR disease
FBXO11	2-48132709-GCT-G	Intellectual Developmental Disorder with Dysmorphic Facies and Behavioral Abnormalities (AD)	VOUS – canonical transcript is less expressed than others, no pathogenic LoF variants in this exon, no phenotype match
FOXP2	7-114284810-G-T	Childhood speech dyspraxia (AD)	Pathogenic
KLC2	11-66031422-A-AG	Spastic Paraparesis and Optic Atrophy (AR)	VOUS - AR disease
NR4A3	9-102590382-C-T	Chondrosarcoma susceptibility (AD)	VOUS - Susceptibility locus, mutational mechanism is gene fusion
PRKCG	19-54403863-A-T	Spinocerebellar Ataxia 14 (AD)	VOUS -Missense driven disease
RLIM	X-73811754-T-TGGAA	X Linked ID/Behavior Disorder (X-linked)	VOUS - X-Linked carrier
SETX	9-135187206-TTCTC-T	Amyotrophic lateral sclerosis (ALS), juvenile (AD), Spinocerebellar ataxia with axonal neuropathy (AR)	VOUS - ALS is missense driven, ataxia is AR disease
SCARF2	22-20783543-C-CG	Van den Ende-Gupta/Marden Walker Syndrome (AR)	VOUS - AR disease
TCF7L2	10-114925401-TC-T	Diabetes Susceptibility (AD)	VOUS - Questionable LoF, no phenotype match

TRIM2	4-154191485-AG-A	Charcot Marie Tooth, Type 2R (AR)	VOUS - AR disease and missense driven disease
WBP11	12-14943528-G-A	Vertebral, cardiac, tracheoesophageal, renal, and limb defects (AD)	Likely Pathogenic
ZFHX2	14-23994453-TTCAGGGGCACGGG-T	Marsili Syndrome (AD pain insensitivity syndrome)	VOUS - Missense driven phenotype

Table S6. Qualifying Missense Variants in Online Mendelian Inheritance in Man (OMIM) Genes

Gene	Variant	OMIM Phenotype (Inheritance)	ACMG Determination and Rationale
ACACA	17-35627664-T-G	Acetyl-CoA carboxylase deficiency (AR)	VOUS - No pathogenic missense variants, AR inheritance
ACTN4	19-39212264-G-A	FSGS (AD)	VOUS - No nearby pathogenic variants
CACNA1C	12-2717777-A-G	Timothy Syndrome (AD), Brugada Syndrome (AD)	VOUS - No nearby pathogenic variants
CHD8	14-21866091-T-C	Autism (AD)	VOUS - No nearby pathogenic variants
GABRA2	4-46263990-T-C	Alcohol Dependence, Epileptic Encephalopathy (AD)	VOUS - No nearby pathogenic variants
JAG1	20-10625568-C-T	Alagille Syndrome (AD)	VOUS – No nearby pathogenic variants
KCNA1	12-5021197-C-G	Episodic ataxia/myokymia syndrome (AD)	VOUS – No nearby pathogenic variants
KCNB1	20-48098611-C-T	Epileptic encephalopathy (AD)	VOUS – No nearby pathogenic variants
MED23	6-131917146-G-A	Mental Retardation (AR)	VOUS – No nearby pathogenic variants, AR inheritance
NTRK2	9-87570260-C-T	Obesity, Hyperphagia, and Developmental Delay (AD)	VOUS – No nearby pathogenic variants
PIK3CD	1-9787026-T-A	Immunodeficiency (AD)	VOUS - Pathogenic missense variant for immunodeficiency is 3 base pairs away but on a different amino acid
PLCB1	20-8709793-G-C	Epileptic Encephalopathy (EE) (AR)	VOUS – Epileptic encephalopathy with AR inheritance due to LoF
POLR3B	12-106895135-G-A	Leukodystrophy, hypomyelinating, 8 (AR)	VOUS - No nearby pathogenic missense variants and AR inheritance

PPP3CA	4-102030139-T-C	EE (AD), Arthrogryposis/ ID (AD)	VOUS – No nearby pathogenic variants
PTEN	10-89653847-A-G	Cowden Syndrome, Macrocephaly/Autism (AD)	VOUS - in a mutational hotspot with 20 pathogenic and 2 benign, adjacent AA has a pathogenic variant.
SMARCA2	9-2039708-G-A	Nicolaides-Baraitser syndrome (AD)	VOUS - No nearby pathogenic missense variants
SMARCA2	9-2073297-C-T	Nicolaides-Baraitser syndrome (AD)	VOUS - No nearby pathogenic missense variants
VCP	9-35061044-T-G	Inclusion Body Myopathy (AD), Charcot-Marie-Tooth Disease, Type 2Y (AD)	VOUS - No nearby pathogenic missense variants

Table S7. Qualifying missense and loss-of-function variants with reported genotype-phenotype associations

Gene	Variant Type	SCZ, DD, ASD	OMIM Disease Phenotype [Inheritance Pattern]
ABCA2	Missense	SCZ (Miss)	-
ACACA	Missense	-	Acetyl-COA Carboxylase Deficiency [AR]
ACTN4	Missense	-	Focal Segmental Glomerulosclerosis [AD]
CACNA1C	Missense	DD	Timothy Syndrome [AD]
CHD5	Missense	SCZ (Miss)	-
CHD8	Missense	DD, ASD	-
DPYSL2	Missense	ASD	-
GABRA2	Missense		Alcohol Dependence, Epileptic Encephalopathy [AD]
GRM5	Missense	SCZ (Miss)	-
IPO5	Missense	SCZ (Miss)	-
JAG1	Missense	-	Alagille Syndrome [AD] | Tetralogy of Fallot [AD]
KCNA1	Missense	-	Episodic Ataxia, Type 1 [AD]
KCNB1	Missense	DD	Epileptic Encephalopathy, Early infantile [AD]
MED23	Missense		Mental Retardation [AR]
NTRK2	Missense		Obesity, Hyperphagia, and Developmental Delay [AD]
PIK3CD	Missense	SCZ (Miss)	Immunodeficiency [AD]
PLCB1	Missense	-	Epileptic Encephalopathy, Early Infantile [AR]
POLR3B	Missense	-	Hypomyelinating Leukodystrophy with/without oligodontia and/or Hypogonadotropic Hypogonadism [AR]
PPP3CA	Missense	SCZ (LoF), DD	Epileptic Encephalopathy [AR] | Arthrogryposis, Cleft Palate, Craniosynostosis and Impaired Development [AR]

PTEN	Missense	DD, ASD	Cowden Syndrome [AD] | Macrocephaly/Autism Syndrome [AD]
SMARCA2	Missense	DD	Nicolaides-Baraitser Syndrome [AD]
SMARCA2	Missense	DD	Nicolaides-Baraitser Syndrome [AD]
AKAP11	LoF	SCZ (LoF)	-
C3	LoF	-	C3 Deficiency [AR]
CSNK1D	LoF	-	Familial Advanced Sleep Phase Syndrome [AD]
DIS3L2	LoF	-	Perlman Syndrome [AR]
DST	LoF	-	Hereditary Neuropathy Sensory and Autonomic, Type VI [AR] | Epidermolysis Bullosa Simplex [AR]
FBXO11	LoF	DD	Intellectual Developmental Disorder with Dysmorphic Facies and Behavioral Abnormalities [AD]
FOXP2	LoF	DD, ASD	Speech-Language Disorder [AD]
HMGCR	LoF	SCZ (LoF)	-
KLC2	LoF	-	Spastic Paraplegia, Optic Atrophy, and Neuropathy [AR]
PRKCG	LoF	-	Spinocerebellar Ataxia [AD]
PTK2	LoF	SCZ (LoF)	-
PUM2	LoF	SCZ (LoF)	-
RLIM	LoF	-	Tonne-Kalscheuer Syndrome [X-linked]
SCARF2	LoF	-	Van Den Ende-Gupta Syndrome [AR]
SETX	LoF	-	Amyotrophic Lateral Sclerosis, Juvenile [AD] | Spinocerebellar Ataxia with Axonal Neuropathy [AR]
TCF7L2	LoF	DD, ASD	Diabetes Mellitus, Non-Insulin-Dependent [AD]
TERF1	LoF	SCZ (LoF)	-
TRIM2	LoF	-	Axonal Charcot-Marie-Tooth Disease, Type 2R [AR]

WBP11	LoF	-	Vertebral, Cardiac, Tracheoesophageal, Renal, and Limb defects [AD]
ZFHX2	LoF	-	Marsili Syndrome [AD]

Qualifying missense and loss-of-function variants in intolerant genes (missense Z > 3.09 and pLI > 0.9, respectively) with known genotype-phenotype relationships. SCZ (Miss): Gene with evidence of enrichment for missense variation in the SCHEMA study of schizophrenia (n = 45 genes). SCZ (LoF): Gene with prior evidence of enrichment for loss-of-function variation in SCHEMA (n = 75 genes). DD: One of 299 genome-wide significant genes associated with developmental delay. ASD: One of 102 genome-wide significant genes associated with autism spectrum disorder. OMIM: Genes known to cause Mendelian disorders based on the Online Mendelian Inheritance in Man database. Of these qualifying variants in OMIM genes, only the loss-of-function variants in FOXP2 and WBP11 meet American College of Medical Genetics (ACMG) diagnostic criteria for a “pathogenic” or “likely pathogenic” variant. AR: Autosomal Recessive; AD: Autosomal Dominant

Table S8. Top ten genes from gene-based collapsing models

Rank	Gene	Case QV	Ctrl QV	Case QV Freq %	Ctrl QV Freq %	CMH Exact P Value
1	'MTA1'	2	0	1.79	0.00	8.90E-04
2	'PDE1C'	3	5	2.68	0.10	9.64E-04
3	'NDUFAF4'	2	1	1.79	0.02	1.30E-03
4	'HDAC6'	2	1	1.79	0.02	2.62E-03
5	'KCNJ2'	2	1	1.79	0.02	2.64E-03
6	'MEOX2'	2	2	1.79	0.04	3.18E-03
7	'CYP39A1'	2	2	1.79	0.04	3.34E-03
8	'B4GALT2'	2	3	1.79	0.06	3.97E-03
9	'HBP1'	2	4	1.79	0.08	4.58E-03
10	'ZDHHC5'	2	3	1.79	0.06	5.50E-03

Top ten genes from qualifying missense (REVEL > 0.5) collapsing analysis

Rank	Gene	Case QV	Ctrl QV	Case QV Freq %	Ctrl QV Freq %	CMH Exact P Value
1	'B4GALT2'	3	3	2.68	0.06	1.86E-04
2	'CCDC125'	2	0	1.79	0.00	2.25E-04
3	'PLTP'	3	4	2.68	0.08	8.09E-04
4	'MTA1'	2	0	1.79	0.00	8.90E-04
5	'NDUFAF4'	2	1	1.79	0.02	1.30E-03
6	'PDE1C'	3	6	2.68	0.12	1.48E-03
7	'GPR39'	2	1	1.79	0.02	2.49E-03
8	'HDAC6'	2	1	1.79	0.02	2.62E-03
9	'NUP37'	2	1	1.79	0.02	2.64E-03
10	'KCNJ2'	2	1	1.79	0.02	2.64E-03

Top ten genes from qualifying missense (REVEL > 0.5) and LoF variant collapsing analysis

Rank	Gene	Case QV	Ctrl QV	Case QV Freq %	Ctrl QV Freq %	CMH Exact P Value
1	'CCDC125'	2	0	1.79	0	2.25E-04

2	'NUP37'	2	1	1.79	0.02	2.64E-03
3	'SLCO4A1'	2	1	1.79	0.02	2.69E-03
4	'HLTF'	2	4	1.79	0.08	4.58E-03
5	'RAD51D'	1	0	0.89	0	7.55E-03
6	'UQCRH'	1	0	0.89	0	7.55E-03
7	'MCMBP'	1	0	0.89	0	7.55E-03
8	'SCARF2'	1	0	0.89	0	7.55E-03
9	'VANGL2'	1	0	0.89	0	7.55E-03
10	'UFSP1'	1	0	0.89	0	7.55E-03

Top ten genes from qualifying LoF variant collapsing analysis

Rank	Gene	Case QV	Ctrl QV	Case QV Freq %	Ctrl QV Freq %	CMH Exact P Value
1	'LMOD3'	3	6	2.68	0.12	5.41E-04
2	'WDR63'	3	5	2.68	0.10	6.65E-04
3	'UGT2B11'	2	0	1.79	0.00	8.90E-04
4	'AVIL'	4	16	3.57	0.32	1.16E-03
5	'KIAA1211L'	3	8	2.68	0.16	2.36E-03
6	'CTSG'	2	1	1.79	0.02	2.62E-03
7	'PRKACG'	2	1	1.79	0.02	2.64E-03
8	'UTP15'	2	1	1.79	0.02	2.64E-03
9	'NUDC'	2	2	1.79	0.04	3.25E-03
10	'FGFBP3'	2	2	1.79	0.04	3.29E-03

Top ten genes from qualifying synonymous variant collapsing analysis

QV: Qualifying variants that pass quality control and filtering

CMH: Cochran–Mantel–Haenszel Test

Table S9. Qualifying Missense and Loss-of-Function Variants in Intolerant Genes

Variant ID	Effect	HGVS_p	Gene Name	Sample Name	gnomAD Exome non_neuro_AF	gnomAD Gene pLI	gnomAD Gene mis_z	MTR	MTR FDR	REVEL
9-139906176-C-G	missense_variant	p.Cys1854Ser	ABCA2	chapscz099	NA	1	4.91	0.563	0.007	0.682
17-35627664-T-G	missense_variant	p.Asp251Ala	ACACA	chapscz046	NA	1	7.24	0.479	0.093	0.861
19-39212264-G-A	missense_variant	p.Glu460Lys	ACTN4	chapscz092	4.81E-06	1	4.16	0.635	0.027	0.567
13-42874887-C-CA	frameshift_variant	p.Thr670fs	AKAP11	chapscz123	NA	0.979	0.313	NA	NA	NA
11-120343825-G-T	stop_gained	p.Glu989*	ARHGEF12	chapscz167	NA	1	3.25	NA	NA	NA
20-3565405-A-AT	frameshift_variant	p.Lys1022fs	ATRN	chapscz042	NA	1	2.69	NA	NA	NA
19-6678035-C-T	splice_acceptor_variant	NA	C3	chapscz022	5.29E-05	0.904	2.75	NA	NA	NA
12-2717777-A-G	missense_variant	p.Ile1153Val	CACNA1C	chapscz015	4.81E-06	1	6.47	0.601	0.042	0.729
1-6172218-C-T	missense_variant	p.Ala1708Thr	CHD5	chapscz068	4.81E-06	1	5.32	0.732	0.031	0.595
20-40079664-T-C	missense_variant	p.Asp1202Gly	CHD6	chapscz030	2.40E-05	1	4	0.671	0.093	0.711
14-21866091-T-C	missense_variant	p.Met1648Val	CHD8	chapscz116	NA	1	5.95	0.406	0.031	0.678
17-80207447-C-CG	frameshift_variant	p.Arg306fs	CSNK1D	chapscz133	NA	0.996	2.7	NA	NA	NA
4-151142828-G-A	splice_acceptor_variant	NA	DCLK2	chapscz042	NA	0.957	1.76	NA	NA	NA
15-65954266-TCTTGA-T	frameshift_variant	p.Val1837fs	DENND4A	chapscz068	1.47E-05	0.996	3.67	NA	NA	NA
12-31586177-C-T	stop_gained	p.Trp673*	DENND5B	chapscz132	NA	1	3.39	NA	NA	NA
2-233001277-CCT-C	frameshift_variant	p.Leu267fs	DIS3L2	chapscz081	NA	0.944	0.944	NA	NA	NA
8-26501008-C-A	missense_variant	p.His289Asn	DPYSL2	chapscz126	1.45E-05	0.994	3.88	0.59	0.059	0.642
6-56498989-C-CT	frameshift_variant	p.Val651fs	DST	chapscz050	NA	1	2.22	NA	NA	NA
7-143053910-G-T	stop_gained	p.Tyr244*	FAM131B	chapscz144	NA	0.991	1.42	NA	NA	NA
2-48132707-CT-C	frameshift_variant	p.Gln51fs	FBXO11	chapscz117	NA	1	4.38	NA	NA	NA
2-48132709-GCT-G	frameshift_variant	p.Gln50fs	FBXO11	chapscz117	NA	1	4.38	NA	NA	NA
19-17883356-A-ATTGGGCAGGTGAG	frameshift_variant	p.Tyr218fs	FCHO1	chapscz098	NA	0.998	1.4	NA	NA	NA
7-114284810-G-T	stop_gained	p.Glu354*	FOXP2	chapscz107	NA	1	1.9	NA	NA	NA

Variant ID	Effect	HGVS_p	Gene Name	Sample Name	gnomAD Exome non_neuro_AF	gnomAD Gene pLI	gnomAD Gene mis_z	MTR	MTR FDR	REVEL
4-48529660-GAAAAT-G	frameshift_variant	p.Ser2384fs	FRYL	chapscz016	NA	0.997	2.96	NA	NA	NA
10-35928610-T-C	missense_variant	p.Tyr583Cys	FZD8	chapscz081	NA	0.79	3.17	0.707	0.048	0.782
4-46263990-T-C	missense_variant	p.Thr338Ala	GABRA2	chapscz103	4.81E-06	0.996	3.13	0.559	0.061	0.8
9-74865664-A-C	splice_acceptor_variant	NA	GDA	chapscz121	NA	0.999	1.46	NA	NA	NA
11-88386431-C-T	missense_variant	p.Arg351Gln	GRM5	chapscz072	4.81E-06	0.999	3.21	0.688	0.058	0.816
5-74646765-ATCTC-A	frameshift_variant	p.Ser313fs	HMGCR	chapscz013	NA	0.999	3.76	NA	NA	NA
13-98664545-G-A	missense_variant	p.Glu719Lys	IPO5	chapscz012	1.44E-05	1	3.36	0.478	0.024	0.615
21-35257811-C-T	missense_variant	p.Pro1610Ser	ITSN1	chapscz082	NA	1	3.61	0.483	0.004	0.507
20-10625568-C-T	missense_variant	p.Gly763Ser	JAG1	chapscz022	3.36E-05	1	3.25	0.723	0.082	0.672
12-5021197-C-G	missense_variant	p.Thr218Arg	KCNA1	chapscz084	4.80E-06	0.076	3.33	0.718	0.081	0.526
20-48098611-C-T	missense_variant	p.Arg136His	KCNB1	chapscz099	NA	1	4.27	0.568	0.053	0.823
2-155555474-G-C	missense_variant	p.Glu63Gln	KCNJ3	chapscz073	NA	0.991	4.06	0.248	0.001	0.788
12-21926333-A-G	missense_variant	p.Leu73Pro	KCNJ8	chapscz028	NA	0.38	3.45	0.556	0.091	0.982
16-15698201-G-A	stop_gained	p.Arg1413*	KIAA0430	chapscz011	NA	1	2.73	NA	NA	NA
11-66031422-A-AG	frameshift_variant	p.Ile311fs	KLC2	chapscz107	NA	0.975	2.02	NA	NA	NA
4-18023353-TTCTC-T	frameshift_variant	p.Arg7fs	LCORL	chapscz074	NA	0.924	1.66	NA	NA	NA
19-49005737-C-G	missense_variant	p.Glu278Asp	LMTK3	chapscz044	NA	1	4.32	0.606	0.037	0.556
20-3147741-TG-T	frameshift_variant	p.Pro23fs	LZTS3	chapscz129	NA	0.996	1.75	NA	NA	NA
10-121598071-C-CT	frameshift_variant	p.Glu464fs	MCMBP	chapscz095	NA	0.996	1.98	NA	NA	NA
1-204515996-T-TAC	frameshift_variant	p.Ser300fs	MDM4	chapscz020	NA	1	1.71	NA	NA	NA
6-131917146-G-A	missense_variant	p.Pro985Leu	MED23	chapscz162	NA	3.10E-08	4.73	0.631	0.088	0.505
14-105905053-C-T	missense_variant	p.Arg25Trp	MTA1	chapscz129	NA	1	3.5	0.502	0.024	0.743
20-62836451-G-C	splice_donor_variant	NA	MYT1	chapscz036	9.62E-06	0.978	2.56	NA	NA	NA
12-124819821-CG-C	frameshift_variant	p.Val2081fs	NCOR2	chapscz011	NA	1	2.01	NA	NA	NA
9-102590382-C-T	stop_gained	p.Gln31*	NR4A3	chapscz047	NA	0.913	2.81	NA	NA	NA

Variant ID	Effect	HGVS_p	Gene Name	Sample Name	gnomAD Exome non_neuro_AF	gnomAD Gene pLI	gnomAD Gene mis_z	MTR	MTR FDR	REVEL
9-87570260-C-T	missense_variant	p.Ser667Leu	NTRK2	chapscz074	4.87E-06	1	3.73	0.61	0.065	0.634
1-9787026-T-A	missense_variant	p.Phe1019Leu	PIK3CD	chapscz044	NA	1	4.27	0.552	0.026	0.544
20-8709793-G-C	missense_variant	p.Gln620His	PLCB1	chapscz012	NA	0.983	3.83	0.414	0.06	0.592
11-64025983-C-T	missense_variant	p.Arg351Cys	PLCB3	chapscz128	4.82E-06	0.94	3.27	0.499	0.021	0.645
12-106895135-G-A	missense_variant	p.Val1007Met	POLR3B	chapscz080	3.85E-05	1.68E-23	3.2	0.525	0.014	0.592
4-102030139-T-C	missense_variant	p.Tyr119Cys	PPP3CA	chapscz038	NA	1	3.63	0.506	0.079	0.656
19-54403863-A-T	splice_acceptor_variant	NA	PRKCG	chapscz130	NA	1	3.06	NA	NA	NA
10-89653847-A-G	missense_variant	p.Asn49Asp	PTEN	chapscz133	NA	0.257	3.49	0.472	0.019	0.716
8-141762340-G-A	stop_gained	p.Gln470*	PTK2	chapscz046	NA	1	3.44	NA	NA	0.457
2-20494234-AC-A	frameshift_variant	p.Val352fs	PUM2	chapscz113	NA	0.995	2.59	NA	NA	NA
X-73811754-T-TGGAA	frameshift_variant	p.Ser466fs	RLIM	chapscz042	NA	0.993	2.42	NA	NA	NA
X-73811755-C-CTG	frameshift_variant	p.Ser466fs	RLIM	chapscz042	NA	0.993	2.42	NA	NA	NA
22-20783543-C-CG	frameshift_variant	p.Leu509fs	SCARF2	chapscz098	4.81E-06	1	3.02	NA	NA	NA
9-135187206-TTCTC-T	frameshift_variant	p.Glu1770fs	SETX	chapscz120	2.89E-05	0.955	-0.112	NA	NA	NA
9-2073297-C-T	missense_variant	p.Ala611Val	SMARCA2	chapscz122	NA	1	5.05	0.288	0	0.713
9-2039708-G-A	missense_variant	p.Glu200Lys	SMARCA2	chapscz046	9.68E-06	1	5.05	0.583	0.017	0.509
17-45992756-C-T	splice_donor_variant	NA	SP2	chapscz059	2.42E-05	0.997	2.37	NA	NA	NA
19-46341718-C-T	splice_donor_variant	NA	SYMPK	chapscz142	NA	1	2.86	NA	NA	NA
10-114925401-TC-T	frameshift_variant	p.Ser490fs	TCF7L2	chapscz122	3.88E-05	0.995	2.4	NA	NA	NA
6-35446210-C-G	splice_donor_variant	NA	TEAD3	chapscz065	NA	0.994	2.76	NA	NA	NA
8-73958260-G-A	stop_gained	p.Trp383*	TERF1	chapscz052	NA	0.907	0.372	NA	NA	NA
4-154191485-AG-A	splice_acceptor_variant	NA	TRIM2	chapscz116	NA	0.996	3.57	NA	NA	NA
7-138946487-AGTAAGT-A	splice_donor_variant	NA	UBN2	chapscz013	NA	1	0.187	NA	NA	NA
19-35762011-C-T	stop_gained	p.Arg232*	USF2	chapscz036	4.81E-06	0.993	1.89	NA	NA	NA
9-35061044-T-G	missense_variant	p.Asn443His	VCP	chapscz134	4.81E-06	1	5.41	0.636	0.062	0.58
12-14943528-G-A	stop_gained	p.Gln391*	WBP11	chapscz075	NA	1	2.98	NA	NA	NA

Variant ID	Effect	HGVS_p	Gene Name	Sample Name	gnomAD Exome non_neuro_AF	gnomAD Gene pLI	gnomAD Gene mis_z	MTR	MTR FDR	REVEL
9-137019605-G-A	missense_variant	p.Val217Met	WDR5	chapscz151	NA	1	3.39	0.242	0	0.715
14-23994453-TTCAGGGGCACGGG-T	frameshift_variant	p.Thr1562fs	ZFHX2	chapscz044	0	0.997	2.95	NA	NA	NA

References

1. G. B. D. Disease, I. Injury, C. Prevalence, Global, regional, and national incidence, prevalence, and years lived with disability for 310 diseases and injuries, 1990-2015: a systematic analysis for the Global Burden of Disease Study 2015. *Lancet* **388**, 1545-1602 (2016).
2. P. F. Sullivan, K. S. Kendler, M. C. Neale, Schizophrenia as a complex trait: evidence from a meta-analysis of twin studies. *Arch. Gen. Psychiatry* **60**, 1187-1192 (2003).
3. A. Sekar *et al.*, Schizophrenia risk from complex variation of complement component 4. *Nature* **530**, 177-183 (2016).
4. C. Schizophrenia Working Group of the Psychiatric Genomics, Biological insights from 108 schizophrenia-associated genetic loci. *Nature* **511**, 421-427 (2014).
5. T. Singh *et al.*, Exome sequencing identifies rare coding variants in 10 genes which confer substantial risk for schizophrenia. *medRxiv* 10.1101/2020.09.18.20192815, 2020.2009.2018.20192815 (2020).
6. J. Kaplanis *et al.*, Evidence for 28 genetic disorders discovered by combining healthcare and research data. *Nature* 10.1038/s41586-020-2832-5 (2020).
7. F. K. Satterstrom *et al.*, Large-Scale Exome Sequencing Study Implicates Both Developmental and Functional Changes in the Neurobiology of Autism. *Cell* **180**, 568-584 e523 (2020).
8. K. Ahn *et al.*, High rate of disease-related copy number variations in childhood onset schizophrenia. *Mol. Psychiatry* **19**, 568-572 (2014).
9. K. C. Epi *et al.*, De novo mutations in epileptic encephalopathies. *Nature* **501**, 217-221 (2013).
10. I. K. Kotowski *et al.*, A spectrum of PCSK9 alleles contributes to plasma levels of low-density lipoprotein cholesterol. *Am. J. Hum. Genet.* **78**, 410-422 (2006).
11. T. N. Turner *et al.*, Loss of delta-catenin function in severe autism. *Nature* **520**, 51-56 (2015).
12. I. J. Barnett, S. Lee, X. Lin, Detecting rare variant effects using extreme phenotype sampling in sequencing association studies. *Genet. Epidemiol.* **37**, 142-151 (2013).
13. A. Ambalavanan *et al.*, De novo variants in sporadic cases of childhood onset schizophrenia. *Eur. J. Hum. Genet.* **24**, 944-948 (2016).
14. E. J. Gardner *et al.*, Sex-biased reduction in reproductive success drives selective constraint on human genes. *bioRxiv* (2020).
15. S. Gulsuner *et al.*, Genetics of schizophrenia in the South African Xhosa. *Science* **367**, 569-573 (2020).
16. E. F. Torrey (2016) Going, Going Gone: Trends and Consequences of Eliminating State Psychiatric Beds. (Treatment Advocacy Center, Arlington, Virginia), pp 1-38.
17. T. Singh *et al.*, The contribution of rare variants to risk of schizophrenia in individuals with and without intellectual disability. *Nat. Genet.* **49**, 1167-1173 (2017).
18. American Psychiatric Association., *Diagnostic and statistical manual of mental disorders : DSM-5* (American Psychiatric Association, Washington, D.C., ed. 5th, 2013), pp. xliv, 947 p.
19. E. Collaborative, Sub-genic intolerance, ClinVar, and the epilepsies: A whole-exome sequencing study of 29,165 individuals. *Am. J. Hum. Genet.* **108**, 965-982 (2021).
20. F. K. Satterstrom *et al.*, Autism spectrum disorder and attention deficit hyperactivity disorder have a similar burden of rare protein-truncating variants. *Nat. Neurosci.* **22**, 1961-1965 (2019).
21. K. J. Karczewski *et al.*, The mutational constraint spectrum quantified from variation in 141,456 humans. *Nature* **581**, 434-443 (2020).
22. M. Halvorsen *et al.*, Exome sequencing in obsessive-compulsive disorder reveals a burden of rare damaging coding variants. *Nat. Neurosci.* **24**, 1071-1076 (2021).
23. N. M. Ioannidis *et al.*, REVEL: An Ensemble Method for Predicting the Pathogenicity of Rare Missense Variants. *Am. J. Hum. Genet.* **99**, 877-885 (2016).
24. R. Ghosh, N. Oak, S. E. Plon, Evaluation of in silico algorithms for use with ACMG/AMP clinical variant interpretation guidelines. *Genome Biol.* **18**, 225 (2017).

25. M. Silk, S. Petrovski, D. B. Ascher, MTR-Viewer: identifying regions within genes under purifying selection. *Nucleic Acids Res.* **47**, W121-W126 (2019).
26. B. B. Cummings *et al.*, Transcript expression-aware annotation improves rare variant interpretation. *Nature* **581**, 452-458 (2020).
27. S. Benjamin, M. D. Lauterbach, A. L. Stanislawski, Congenital and acquired disorders presenting as psychosis in children and young adults. *Child Adolesc. Psychiatr. Clin. N. Am.* **22**, 581-608 (2013).
28. M. D. Lauterbach, A. L. Stanislawski-Zygaj, S. Benjamin, The differential diagnosis of childhood- and young adult-onset disorders that include psychosis. *J. Neuropsychiatry Clin. Neurosci.* **20**, 409-418 (2008).
29. Online Mendelian Inheritance in Man, OMIM® (2020) (McKusick-Nathans Institute of Genetic Medicine, Johns Hopkins University (Baltimore, MD)).
30. S. Richards *et al.*, Standards and guidelines for the interpretation of sequence variants: a joint consensus recommendation of the American College of Medical Genetics and Genomics and the Association for Molecular Pathology. *Genet. Med.* **17**, 405-424 (2015).
31. K. D. MacDermot *et al.*, Identification of FOXP2 truncation as a novel cause of developmental speech and language deficits. *Am. J. Hum. Genet.* **76**, 1074-1080 (2005).
32. E. Martin *et al.*, Heterozygous loss of WBP11 function causes multiple congenital defects in humans and mice. *Hum. Mol. Genet.* **29**, 3662-3678 (2020).
33. T. Balakrishna, D. Curtis, Assessment of Potential Clinical Role for Exome Sequencing in Schizophrenia. *Schizophr. Bull.* 10.1093/schbul/sbz057 (2019).
34. S. Srivastava *et al.*, Meta-analysis and multidisciplinary consensus statement: exome sequencing is a first-tier clinical diagnostic test for individuals with neurodevelopmental disorders. *Genet. Med.* **21**, 2413-2421 (2019).
35. D. Astuti *et al.*, Germline mutations in DIS3L2 cause the Perlman syndrome of overgrowth and Wilms tumor susceptibility. *Nat. Genet.* **44**, 277-284 (2012).
36. V. Sriretnakumar, R. Harripaul, J. B. Vincent, J. L. Kennedy, J. So, Enrichment of pathogenic variants in genes associated with inborn errors of metabolism in psychiatric populations. *Am. J. Med. Genet. B Neuropsychiatr. Genet.* **180**, 46-54 (2019).
37. L. Bastarache *et al.*, Phenotype risk scores identify patients with unrecognized Mendelian disease patterns. *Science* **359**, 1233-1239 (2018).
38. G. Genovese *et al.*, Increased burden of ultra-rare protein-altering variants among 4,877 individuals with schizophrenia. *Nat. Neurosci.* **19**, 1433-1441 (2016).
39. F. Lescai *et al.*, Meta-analysis of Scandinavian Schizophrenia Exomes. *bioRxiv* (2019).
40. S. Perry, B. Kiragasi, D. Dickman, A. Ray, The Role of Histone Deacetylase 6 in Synaptic Plasticity and Memory. *Cell Rep* **18**, 1337-1345 (2017).
41. H. F. Chan *et al.*, A novel neuropsychiatric phenotype of KCNJ2 mutation in one Taiwanese family with Andersen-Tawil syndrome. *J. Hum. Genet.* **55**, 186-188 (2010).
42. W. S. Stone *et al.*, Association Between the Duration of Untreated Psychosis and Selective Cognitive Performance in Community-Dwelling Individuals With Chronic Untreated Schizophrenia in Rural China. *JAMA Psychiatry* 10.1001/jamapsychiatry.2020.1619 (2020).
43. E. B. Robinson *et al.*, Autism spectrum disorder severity reflects the average contribution of de novo and familial influences. *Proc. Natl. Acad. Sci. U. S. A.* **111**, 15161-15165 (2014).
44. A. Reichenberg *et al.*, Discontinuity in the genetic and environmental causes of the intellectual disability spectrum. *Proc. Natl. Acad. Sci. U. S. A.* **113**, 1098-1103 (2016).
45. H. Guo *et al.*, Genome sequencing identifies multiple deleterious variants in autism patients with more severe phenotypes. *Genet. Med.* **21**, 1611-1620 (2019).
46. S. M. Myers *et al.*, Insufficient Evidence for "Autism-Specific" Genes. *Am. J. Hum. Genet.* **106**, 587-595 (2020).
47. S. L. Bishop *et al.*, Identification of Developmental and Behavioral Markers Associated With Genetic Abnormalities in Autism Spectrum Disorder. *Am. J. Psychiatry* **174**, 576-585 (2017).

48. J. Kane, G. Honigfeld, J. Singer, H. Meltzer, Clozapine for the treatment-resistant schizophrenic. A double-blind comparison with chlorpromazine. *Arch. Gen. Psychiatry* **45**, 789-796 (1988).
49. G. Abraham, Y. Qiu, M. Inouye, FlashPCA2: principal component analysis of Biobank-scale genotype datasets. *Bioinformatics* **33**, 2776-2778 (2017).
50. S. Cameron-Christie *et al.*, Exome-Based Rare-Variant Analyses in CKD. *J. Am. Soc. Nephrol.* **30**, 1109-1122 (2019).
51. S. Petrovski *et al.*, An Exome Sequencing Study to Assess the Role of Rare Genetic Variation in Pulmonary Fibrosis. *Am. J. Respir. Crit. Care Med.* **196**, 82-93 (2017).
52. G. Povysil *et al.*, Rare-variant collapsing analyses for complex traits: guidelines and applications. *Nat Rev Genet* **20**, 747-759 (2019).
53. M. T. Pato *et al.*, The genomic psychiatry cohort: partners in discovery. *Am. J. Med. Genet. B Neuropsychiatr. Genet.* **162B**, 306-312 (2013).
54. A. C. Need *et al.*, Exome sequencing followed by large-scale genotyping suggests a limited role for moderately rare risk factors of strong effect in schizophrenia. *Am. J. Hum. Genet.* **91**, 303-312 (2012).
55. W. J. B. W. First M.B., Karg R.S., Spitzer R.L., *Structured Clinical Interview for DSM-5—Research Version (SCID-5 for DSM-5, Research Version; SCID-5-RV)* (American Psychiatric Association, Arlington, VA, 2015).
56. R. S. Keefe *et al.*, Clinical characteristics of Kraepelinian schizophrenia: replication and extension of previous findings. *Am. J. Psychiatry* **153**, 806-811 (1996).
57. S. A. Mitelman, M. S. Buchsbaum, Very poor outcome schizophrenia: clinical and neuroimaging aspects. *Int. Rev. Psychiatry* **19**, 345-357 (2007).
58. A. G. Cardno, M. J. Owen, Genetic relationships between schizophrenia, bipolar disorder, and schizoaffective disorder. *Schizophr. Bull.* **40**, 504-515 (2014).
59. H. Santelmann, J. Franklin, J. Busshoff, C. Baethge, Interrater reliability of schizoaffective disorder compared with schizophrenia, bipolar disorder, and unipolar depression - A systematic review and meta-analysis. *Schizophr. Res.* **176**, 357-363 (2016).
60. M. First, Spitzer, RL, Gibbon, M, Williams, JBW, Structured Clinical Interview for DSM-IV-TR Axis I Disorders, Research Version, Patient Edition. (SCID-I/P). New York: Biometrics Research, New York State Psychiatric Institute. (2002).
61. N. A. Miller *et al.*, A 26-hour system of highly sensitive whole genome sequencing for emergency management of genetic diseases. *Genome Med.* **7**, 100 (2015).
62. G. A. Van der Auwera *et al.*, From FastQ data to high confidence variant calls: the Genome Analysis Toolkit best practices pipeline. *Curr Protoc Bioinformatics* **43**, 11 10 11-33 (2013).
63. P. Cingolani *et al.*, A program for annotating and predicting the effects of single nucleotide polymorphisms, SnpEff: SNPs in the genome of Drosophila melanogaster strain w1118; iso-2; iso-3. *Fly (Austin)* **6**, 80-92 (2012).
64. Z. Ren *et al.*, ATAV: a comprehensive platform for population-scale genomic analyses. *BMC Bioinformatics* **22**, 149 (2021).
65. P. D. Stenson *et al.*, Human Gene Mutation Database (HGMD): 2003 update. *Hum. Mutat.* **21**, 577-581 (2003).
66. M. J. Landrum *et al.*, ClinVar: public archive of relationships among sequence variation and human phenotype. *Nucleic Acids Res.* **42**, D980-985 (2014).
67. G. Jun *et al.*, Detecting and estimating contamination of human DNA samples in sequencing and array-based genotype data. *Am. J. Hum. Genet.* **91**, 839-848 (2012).
68. A. Manichaikul *et al.*, Robust relationship inference in genome-wide association studies. *Bioinformatics* **26**, 2867-2873 (2010).
69. M. Lek *et al.*, Analysis of protein-coding genetic variation in 60,706 humans. *Nature* **536**, 285-291 (2016).
70. B. B. Cummings *et al.*, Transcript expression-aware annotation improves rare variant discovery and interpretation. *bioRxiv* (2019).

71. H. L. Rehm *et al.*, ClinGen--the Clinical Genome Resource. *N. Engl. J. Med.* **372**, 2235-2242 (2015).