Supplementary Materials

Sample collections

The Schizophrenia Exome Meta-Analysis (SCHEMA) consortium was formed in 2017 to aggregate and generate exome-sequencing data for the purposes of schizophrenia gene discovery. To expedite the scale-up of this global effort, our analysis incorporates existing sample collections that had been described and analyzed in earlier and current epidemiological and genetic studies of schizophrenia. Table S1 described contributing collections along with the number of samples sequenced, the number of samples retained in the final analysis, and the technological assays used to generate the sequencing data. For each collection, we referred to earlier publications that described the phenotypic ascertainment and analyzed these data for insights related to schizophrenia. In summary, 7,979 schizophrenia cases from four collections (Swedish schizophrenia study, UK10K schizophrenia study, Bulgarian trios study, Taiwanese trios study) had been exome sequenced and described in earlier publications[1–4]. The remaining 16,269 cases had been previously described in other genetic studies of schizophrenia, primarily through genotyping efforts such as the Psychiatric Genomics Consortium; however, the generation and analysis of their sequencing data are presented here for the first time (Ashkenazi Jewish schizophrenia study, Danish iPSYCH initiative, Genomic Psychiatry Cohort [GPC], Pritzker Neuropsychiatric Disorders Research Consortium (Pritzker NDRC), Mclean Psychosis study, and UK-Ireland collections)[5–9]. Combined, from these collections, 24,248 schizophrenia cases were included in the SCHEMA main analysis.

We included exome samples from the following projects in dbGAP as additional external controls: the Alzheimer's Disease Sequencing Project (dbGAP accession: phs000572.v8.p4), NHLBI Exome Sequencing Project (dbGAP accessions: https://esp.gs.washington.edu/drupal/dbGaP_Releases), and the Myocardial Infarction Genetics Exome Sequencing Consortium (dbGAP accessions: phs001000.v1.p1 and phs000806.v1.p1). In addition, we aggregated and incorporated published *de novo* mutations from 3,402 parent-proband trios of schizophrenia[1,4,10–17]. Two of these collections, the Bulgarian and Taiwanese trios studies, are directly included as part of the SCHEMA consortium call set, and inherited variants from these individuals were incorporated.

Sample recruitment and collection are described extensively in the corresponding studies and collections. To ensure compatibility with Psychiatric Genomics Consortium (PGC) definitions[5], we included samples with a diagnosis of schizophrenia and schizoaffective disorders in our analysis. In the final analysis, 22,781 individuals were diagnosed with schizophrenia, while 1,467 were diagnosed with schizoaffective disorders. Each collection provided the Consortium with individual-level information on case status according to study-specific criteria. Each collection defined sample controls as individuals specifically ascertained to not have psychiatric illness, or individuals that were randomly selected from population registers. Informed consent was obtained from all participants, and the institutional human

subject review and ethics committees relevant to that collection approved the research. Despite the collections originating from a number of partner institutions, 22,548 of the 24,248 schizophrenia cases (and all the new 16,269 samples) were sequenced at the Broad Institute of Harvard and M.I.T. and Massachusetts General Hospital.

Table S1 Sample counts in the SCHEMA collections. The number of cases and controls sequenced are displayed, along with the number of samples retained after quality control steps. The sequencing technology for each collection is highlighted, with the majority sequenced using Illumina or Agilent captures. The Pubmed IDs (PMIDs) of articles that describe ascertainment of each collection is provided.

Study Collection	Cases sequenced	Controls sequenced	Cases in analysis	Controls in analysis	Sequencing assay	Cases (new to this report)	Controls (new to this report)	PMIDs
UK-Ireland schizophrenia study	5593	5559	5232	5362	Illumina Nextera	5232	5362	25056061,29483656
Swedish schizophrenia study	4625	6038	4565	6006	Agilent SureSelect v1, Agilent SureSelect v2	0	0	24463508,27694994
WGSPD case-control study (genomes)	3906	4671	3547	4418	Whole genome sequencing	3547	4418	31591465
Danish iPSYCH initiative	3526	6116	3270	5570	Illumina Nextera	3270	0	31768057
Genomic Psychiatry Cohort (GPC) and Pritzker Neuropsychiatric Disorders Research Consortium (PNDRC) case-control studies (exomes)	3419	5087	3004	4616	Illumina Nextera	3004	4616	31591465
Taiwanese trios study	1781	3382	1716	1575	Agilent SureSelect v2, Illumina Nextera	0	0	31932770
UK10K schizophrenia study	1907	6976	986	4178	Agilent SureSelect v3, Agilent SureSelect v5	0	0	26974950
Ashkenazi Jewish schizophrenia study	770	2156	721	1901	Agilent SureSelect v2, Illumina Nextera	721	1901	26198764
Bulgarian trios study	621	1134	598	599	Agilent SureSelect v2, Agilent SureSelect v3, NimbleGen SeqCap v2	0	0	24463507
McLean case-control study	567	347	493	243	Illumina Nextera	493	243	25740047
FINRISK population study	188	8886	116	7853	Nimblegen SeqCap VCRome	0	0	31367044
dbGAP controls (ADSP, ESP, MIGen)	0	11922	0	8116	Agilent SureSelect v2, Illumina Nextera	0	0	(see Supplementary Text for Accessions)

Sequence data production

Sequencing and alignment

In our meta-analysis, we sought to integrate whole-exome and whole-genome sequence data generated over seven years with a number of different sequencing technologies for psychiatric gene discovery. Most of the cases (22,548 of the 24,248 schizophrenia cases) were sequenced at the Broad Institute of Harvard and MIT using the Illumina HiSeq X platform with the use of 151 base pair paired-end reads. Exome samples new to this study (16,269 of presented cases) were enriched using the Illumina Nextera capture kit. Exome samples were sequenced until 80% of the target capture was covered at 20x, while whole-genome samples were sequenced at 20 or 30x. The majority of the remaining exome samples were sequenced using the Agilent SureSelect v2 capture (Table S1). All sequencing data, including those generated externally and data generated in earlier analyses, were centralized as BAM or FASTQ files, and processed uniformly using Picard sequence processing pipeline[18], before mapped onto the human genome reference build 37 (grch37) using BWA[19]. This exactly followed standard best practice alignment and read processing protocols described in more detail in earlier exome analysis publications[20,21].

Variant joint calling

Joint calling of all SCHEMA samples was performed using the Genome Analysis Toolkit (GATK)[22,23]. First, GATK (version 3.4) was used to perform local realignment around indels and recalibrate base qualities in each sample BAM. We called each sample using HaplotypeCaller, generating gVCF files containing every position of the genome with likelihoods for variants or the genomic reference. We merged samples into batches of 200 using CombineVCFs, and joint-called samples using GenotypeVCFs, all using default settings according to the best-practice pipeline. We annotated all variants using the Variant Quality Score Recalibration (VQSR) tool in GATK (version 3.6). The output consists of a VCF with germline SNVs and indels for all samples used in the core SCHEMA analysis. The variant joint calling of SCHEMA samples was equivalent to the pipeline used in the generation of the gnomAD database[21].

Sequencing coverage

To estimate coverage of the coding region, we randomly sampled 100 individuals from each sequencing wave and extracted depth information from their gVCFs. The gVCFs recorded information for every position of the exome, which we used to calculate the mean read depth (DP) for each coding exon as defined in GENCODE v19 (grch37). Each exonic interval was extended by 8 base pairs to include splice regions variants. Because genotype filters from earlier exome sequencing studies had identified variants with DP < 10 as low quality, we compared the proportion of the coding region with mean DP > 10 in each wave to evaluate differences in sequencing quality[20,21,24]. We observed that exome sequencing waves could be separated into two groups - samples generated using the Agilent v2 capture and Illumina

Nextera capture (Figure S1). Agilent v2 samples were generated before 2017 and covered approximately 85% of the exome at 10x or greater, while Illumina Nextera and whole-genome sequenced samples were generated between 2017 and 2019 and covered over 95% of the exome at 10x or greater. The difference in coverage contributed in part to the differences in the number of variants detected. For each sequencing wave, we defined well-covered regions as segments of exome covered 10x or greater in 80% of all samples. When restricting variant comparisons to these consensus regions, we find that the number of heterozygous variants and transition-to-transversion ratio (TiTv) become comparable between exomes (Figure S2).

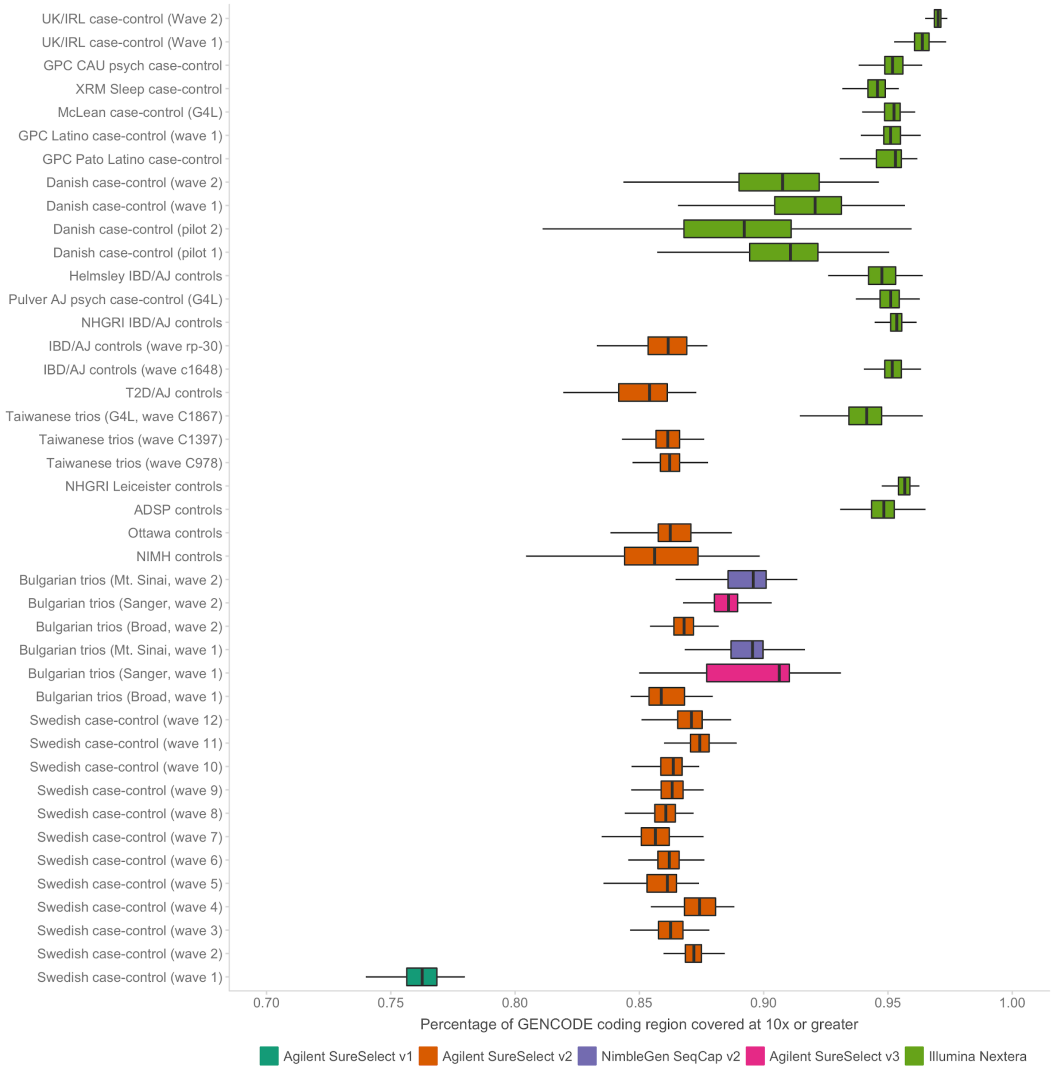


Figure S1 Percentage of the GENCODE coding region covered at 10x or greater in each sequencing wave. The color of each boxplot indicates the exome capture kit used. The sequencing waves are ordered by production date, with the newest waves on top and the oldest at the bottom. The coverages appear consistent between samples of the same capture, except for the Danish case-control samples which were derived from dried blood spots and therefore have slightly lower quality[25].

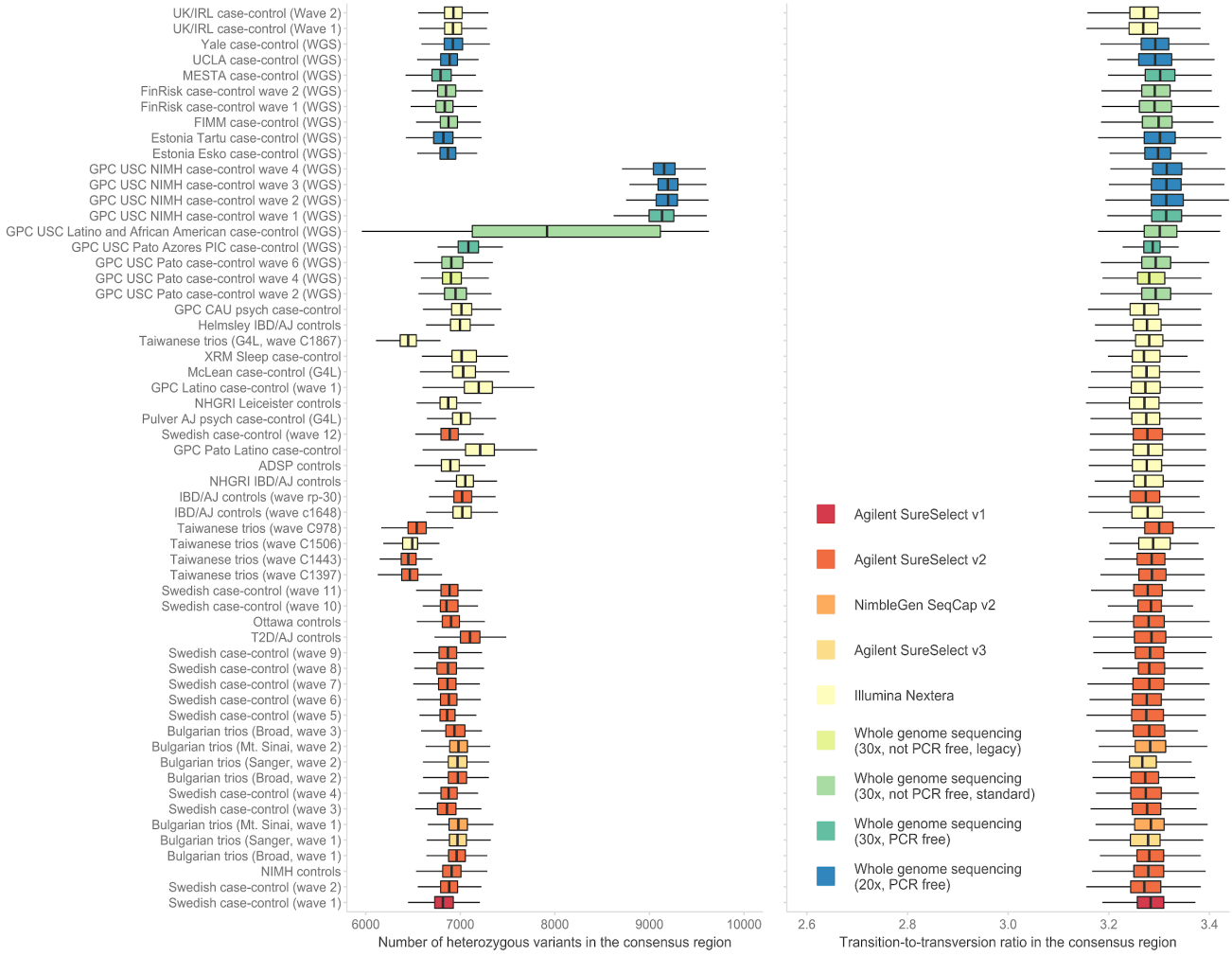


Figure S2 Number of heterozygous variant and transition-to-transversion ratio in each sequencing wave after restricting to coding sequences covered at 10x or greater (within consensus region). The sequencing waves are ordered by production date, with the newest waves on top and the oldest at the bottom. The color of each boxplot indicates the technology used. The number of heterozygous variants restricted to the consensus region appears consistent between most collections and waves, with the exception of the Taiwanese and GPC USC collections, which included individuals of non-European ancestry (see Figure S6).

Sample-level quality control

Hail for data analysis

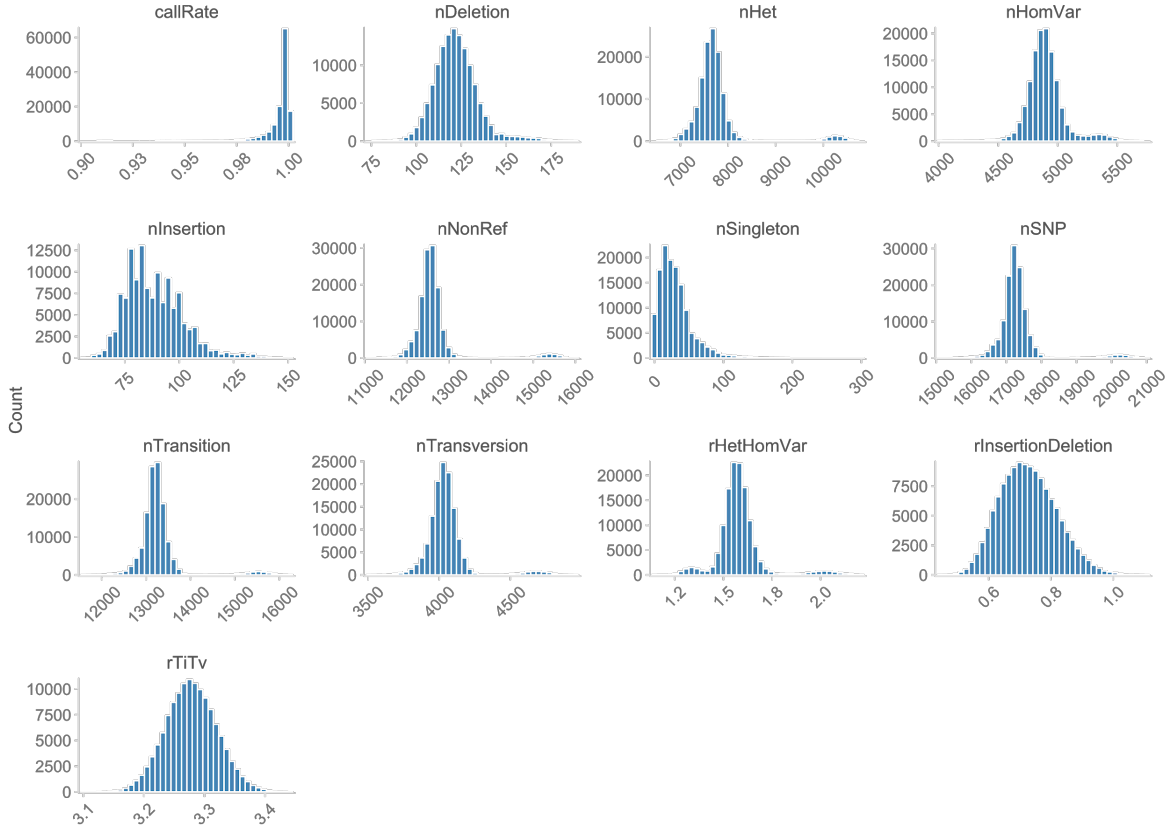
To perform quality control steps on tens of thousands of exomes, we applied and wrote methods based on Hail, an open source Python-based library dedicated to the analysis of large-scale genomic data (website: https://www.hail.is; documentation: https://hail.is/docs/0.1/ and https://hail.is/docs/0.2/; GitHub repository: https://github.com/hail-is/hail). Hail leverages the scalable infrastructure of Apache Spark with a user-friendly interface in Python to enable the processing of terabytes of genomic data. Indeed, this scalable framework has been applied widely in other large sequencing projects, including the latest analyses from the Autism

Sequencing Consortium and the gnomAD Consortium[21,24]. In the following sections, we refer to specific methods in Hail that perform the following sample- and variant-level quality control steps.

Hard Filters

We imported VCFs from joint calling into Hail 0.1 using the import_vcf function, generating Variant Dataset files. Using the consensus regions described in the previous section, we used Hail's sample_qc function to generate the following sample metrics from the raw variant calls: call rate (callRate), number of heterozygous calls (nHet), number of homozygous calls (nHomVar), number of non-reference calls (nNonRef), number of deletions (nDeletion), number of insertions (nInsertion), number of singleton calls (nSingleton), number of SNPs (nSNPs), heterozygous-homozygous call ratio (rHetHomVar), transition-transversion ratio (rTiTv), and insertion-deletion (rInsertionDeletion). Once restricted to consensus regions, the sample metrics across sequencing waves and samples followed comparable distributions except for individual outliers (Figure S3). Based on the observed distributions, we applied the following hard filters to exclude low quality samples: callRate < 0.9, nDeletion < 75 | nDeletion > 190, nHet < 6500 | nHet > 11000, nHomVar < 4000 | nHomVar > 5750, nInsertion < 55 | nInsertion > 150, nNonRef < 11000, nSingleton > 300, nSNP < 15000 | nSNP > 21000, nTransition < 11500 | nTransition > 16500, nTransversion < 3500 | nTransversion > 5000, rHetHomVar < 1.1 | rHetHomVar > 2.2, rInsertionDeletion < 0.4 | rInsertionDeletion > 1.1, and rTiTv < 3.1 | rTiTv > 3.45. The sample metrics were comparable across cohorts after the exclusion of sample outliers (Figure S3).

Figure S3 Distributions of raw variant metrics for all SCHEMA samples, restricted to consensus regions. Descriptions of each metric are provided in the “Hard Filters” subsection (see above). A small subset of samples were observed to have higher total allele counts (e.g. nHet, nNonRef, nTransition, nTransversion). These samples originated from the GPC collections and are of non-European ancestry (see Figure S6).



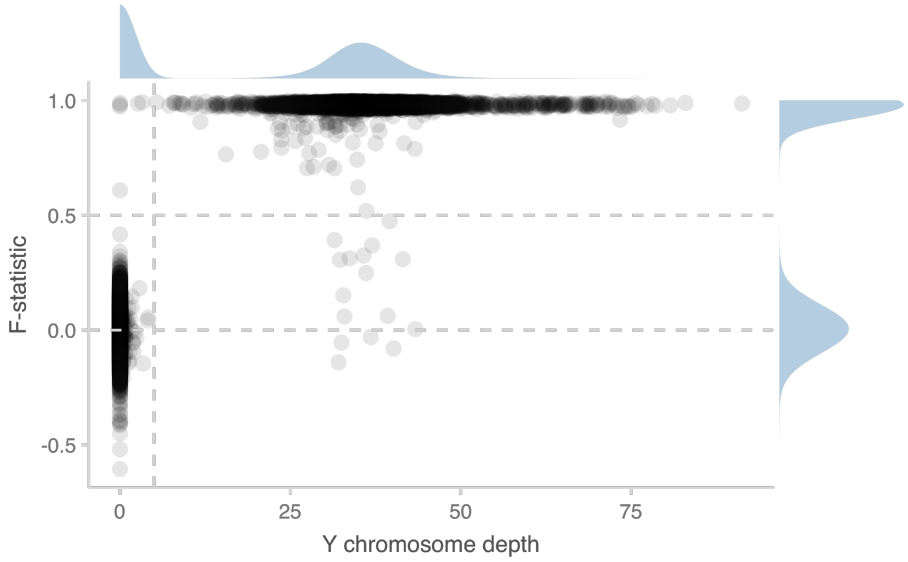
We used Picard to estimate the percentage of chimeric reads per sample, defined as the fraction of reads that map outside a maximum insert size or have two ends mapping to different chromosomes (https://broadinstitute.github.io/picard/picard-metric-definitions.html). We additionally used verifyBamID to estimate the contamination fraction per sample[26], and excluded all samples with a chimeric or contamination fraction > 0.15. The higher-than-usual threshold was chosen due to higher levels of contamination observed in the Taiwanese trio collection, as described in an earlier publication [4]. We applied a higher allelic balance (AB) filter of 0.25 across all samples, as performed in [24] to call *de novo* mutations, to better exclude individual calls that may be contamination artifacts.

Sex imputation

The sex of each sample was inferred using the inbreeding coefficient of chrX variants (F-statistic) and the mean depth of non-pseudoautomal variants on chrY. The *F*-statistic was estimated using Hail’s impute_sex function from chrX variants with call rate ≥ 0.97, allele frequency (AF) ≥ 0.5%, and mean depth (DP) ≥ 7. High-quality variants on chrY were defined as non-PAR variants observed in 1000 Genomes Phase 3, and with a mean DP ≥ 3.5. Samples with F-stat ≥ 0.8 were classified as female, and samples with F-stat ≤ 0.2 were classified as

male. Samples with a mean DP ≥ 10 in chrY variants were inferred to have a Y chromosome; those with a mean DP ≤ 3 were inferred to not have a Y chromosome. Concordance between *F*-statistic and chrY depth inference were required to identify samples as male or female, and we excluded all other individuals from the analysis (Figure S4).

Figure S4 Sample F-statistic plotted against sample chrY variant mean depth. Each dot is a different sample in the SCHEMA call set. Samples with F-statistics ~ 0 and mean chrY coverages ~ 0 are imputed to be male, and samples with F-statistics ~ 1 and a non-zero chrY coverage are imputed to be female. The remaining samples have either low sequencing quality or have chromosomal abnormalities.



Population inference

For population and cryptic relatedness inference, we restricted our analysis to a subset of high-quality SNPs based on genotype and variant-level filters. First, genotypes were required to have depth (DP) ≥ 7, homozygous allelic balance (AB) < 0.1 or 0.2 ≤ heterozygous AB ≤ 0.8, and genotype quality (GQ) ≥ 20. After genotype filters, we retained biallelic SNPs with a mean depth ≥ 7, allele frequency (AF) ≥ 0.01, and call rate ≥ 0.98. We retained 56,818 SNPs for sample QC. These variant filters were only used as a first-pass to identify high-quality samples, and not for the main meta-analysis.

We imputed the ancestry of SCHEMA individuals using principal components analysis on a subset of unrelated samples with known ancestry to form the basis. Most reference samples originated from the 1000 Genomes Project phase 3 release, with additional samples with reported Ashkenazi Jewish, Estonian, and Finnish ancestry[2,27]. We first performed LD-pruning on the remaining SNPs (max $r^2 < 0.1$). We then computed the top 10 principal components (PCs) using these samples, and projected the remaining samples onto these PCs (Figure S5). We trained a random forest model on the reference samples, and used this model to predict global ancestries on the remaining samples, with a threshold probability of 0.7. Samples were assigned the following global populations: European (EUR), Finnish (FIN), Latino American (AMR), African American (AFR), East Asian (EAS), Ashkenazi Jewish (ASJ), Estonian

(EST), and South Asian (SAS). Samples with a probability of < 0.7 were excluded. The proportion of imputed ancestry in our collection appeared reasonable given the institute of origin from which individuals were recruited; for example, a small fraction of FIN individuals were observed in the Estonian and Swedish collections, while the GPC collections contained by design contained individuals of AMR and AFR ancestry (Figure S6).

Figure S5 Principal components analysis of SCHEMA samples using 1000 Genomes individuals with reported ancestries as basis. Cases and controls in our call set are labelled as red and blue respectively. Reference samples from the 1000 Genomes Project are displayed as larger dots in the background, and colored by reported ancestries. We display scatterplots of PC1, PC2, and PC3 to highlight the existing substructure.

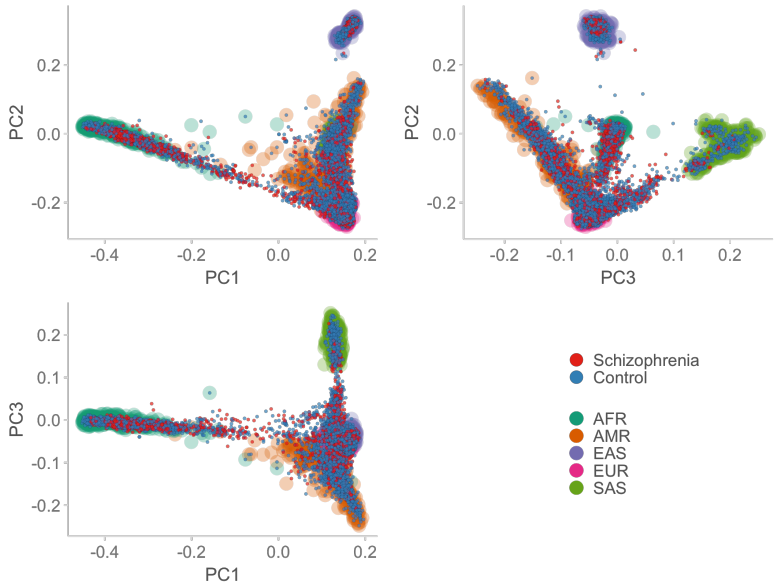
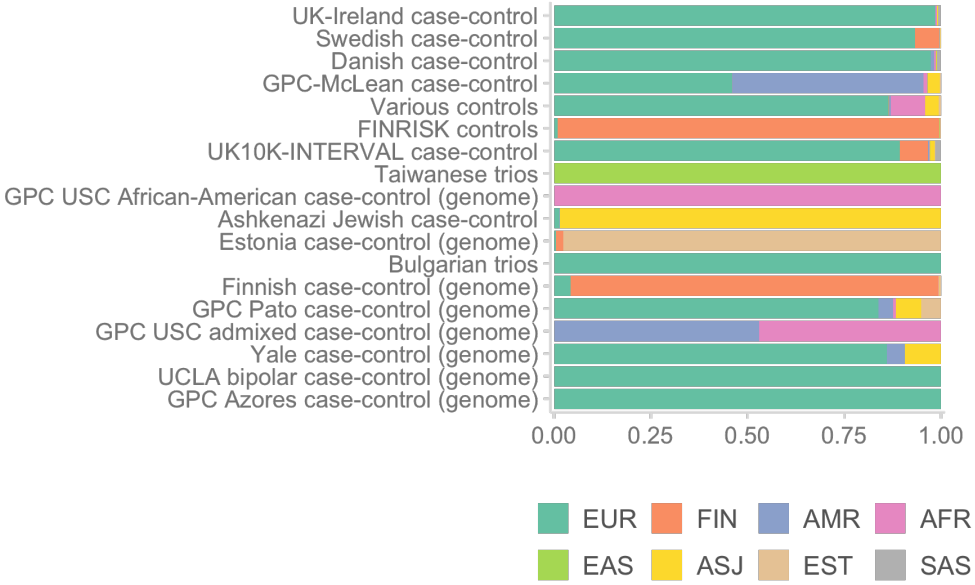


Figure S6 Ancestry assignment by random forest as proportion of each collection. Each sample is assigned a global ancestry based on probabilities from the random forest model. Collections have different ascertainment and design, and we see reasonable concordance between country of origin and imputed ancestries.

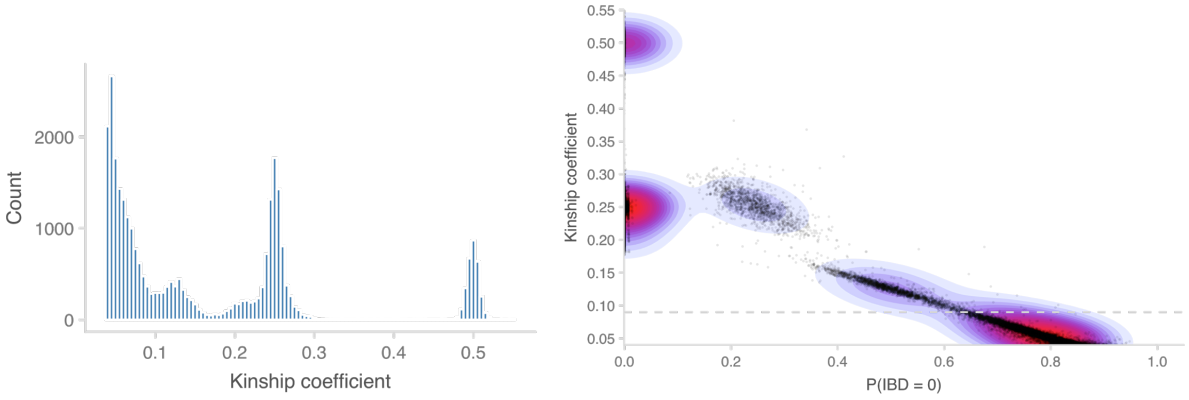


Relatedness and parent-proband trios

We applied the PC-Relate method in Hail (https://hail.is/docs/0.2/methods/genetics.html#hail.methods.pc_relate) to estimate pairwise kinship coefficients (Φ_{ij}) between all pairs of samples with the same assigned global ancestry[28]. We restricted our analysis to LD-pruned SNPs (max $r^2 < 0.1$) with AF > 0.05 and call rate > 0.98. The top 5 PCs were included in the method to correct for population structure in the kinship calculation. We then used Φ_{ij} from PC-Relate to identify clusters of related individuals. We inferred the following relationships based on Φ_{ij}: $\Phi_{ij} > 0.354$ were duplicates or monozygotic twins, $0.177 > \Phi_{ij} > 0.354$ were first-degree relatives, and $0.0884 > \Phi_{ij} > 0.177$ were second-degree relatives (Figure S7)[29]. We pruned clusters of related individuals to ensure that no two samples were second-degree or closer in relations. In edge cases in which multiple options for pruning existed, we prioritized the retention of cases first before prioritizing the total number of unrelated samples.

The Taiwanese and Bulgarian data sets were designed to study *de novo* mutations in schizophrenia through the recruitment of parent-proband trios. Using the calculated kinship coefficients, we were able to confirm the pedigrees of 2,304 trios (Figure S7). *De novo* mutations in these samples had previously been called[1,4]. For inclusion in the case-control meta-analysis, we generated “pseudo-controls” from untransmitted alleles from both parents in each trio. We ensured that the parents used to generate these pseudo-controls were unrelated, and that the resulting pseudo-controls also met the hard filters required for inclusion in the meta-analysis.

Figure S7 Relatedness analysis of samples in the SCHEMA call set using PC-relate. A. Histogram of kinship coefficients (Φ_{ij}) > 0.04. B. Scatter plot and density plot of kinship coefficients v *P*(IBD = 0), or the proportion of loci in which individuals share zero alleles IBD. Dark red indicates a greater density of samples. We infer likely first-, second-, and third-degree relatives in the SCHEMA call set using Φ_{ij}.



Final sample counts for analysis

After all the steps of sample QC, we identify 24,248 cases and 50,437 controls for analysis (Table S2).

Table S2 Number of case-control samples excluded by sample QC criteria. Each QC criterion is described in the Supplementary Text. Is Case: TRUE indicates schizophrenia case, FALSE indicates control. For each of the QC criteria, TRUE indicates the corresponding row of samples have been excluded from the analysis.

Is Case	(Fail) Quality Filters	(Fail) Contamination	(Fail) Ancestry Assignment	(Fail) Sex Imputation	(Fail) Relatedness	Counts
FALSE	FALSE	FALSE	FALSE	FALSE	FALSE	50437
TRUE	FALSE	FALSE	FALSE	FALSE	FALSE	24248
FALSE	FALSE	FALSE	FALSE	FALSE	TRUE	6377
TRUE	FALSE	FALSE	FALSE	FALSE	TRUE	1827
FALSE	TRUE	FALSE	FALSE	FALSE	FALSE	558
FALSE	FALSE	FALSE	TRUE	FALSE	FALSE	542
TRUE	FALSE	FALSE	TRUE	FALSE	FALSE	407
TRUE	TRUE	FALSE	FALSE	FALSE	FALSE	236
FALSE	FALSE	FALSE	FALSE	TRUE	FALSE	154
TRUE	FALSE	FALSE	FALSE	TRUE	FALSE	132
FALSE	TRUE	TRUE	FALSE	FALSE	FALSE	25
TRUE	FALSE	FALSE	TRUE	TRUE	FALSE	20
FALSE	TRUE	FALSE	TRUE	FALSE	FALSE	18
FALSE	TRUE	FALSE	TRUE	TRUE	FALSE	16
FALSE	TRUE	TRUE	FALSE	TRUE	FALSE	11
FALSE	TRUE	FALSE	FALSE	FALSE	TRUE	10

Variant-level quality control

To arrive at a high-quality set of variants for analysis, we applied site and genotype filters in an attempt to increase the number of true positive calls, while reducing the number of false negatives. Analogous genotype and site filters had previously been applied in other sequencing analyses of complex traits[3,24,25,30]. At the genotype level, we retained individual calls if they had a genotype quality (GQ) ≥ 20, allelic balance (AB) < 0.1 in homozygous calls, allelic balance (AB) ≥ 0.25 in heterozygous calls, and depth (DP) ≥ 10. After applying genotype filters, we excluded variants with call rates < 0.9 or if they resided within low-complexity regions (LCR)[31]. In addition to the hard filters, we applied the Genome Analysis Toolkit (GATK) Variant Quality Score Recalibration (VQSR) model to remove low-quality variants based on sequence

annotations from variant calling. We used the recommended annotations and training data sets as suggested by GATK best practices (https://gatkforums.broadinstitute.org/gatk/discussion/23216/how-to-filter-variants-either-with-vqsr-or-by-hard-filtering); the VQSR model applied on our data set was the same as the one applied on the gnomAD database[21]. We excluded variants that failed VQSR for both SNPs and indels, according to default settings.

Variant annotation

The Ensembl Variant Effect Predictor (VEP) v85 was used to annotate all variants according to the GENCODE v19 reference[32,33]. The full pipeline was implemented in Hail 0.1 (https://hail.is/docs/0.1/hail.VariantDataset.html?highlight=vep#hail.VariantDataset.vep), with the LOFTEE annotation provided as default (https://github.com/konradjk/loftee/tree/27b0040f524348baa7f3257f1ce58993529e09ef). Similar to other sequencing analyses, we considered frameshift, stop gained, splice acceptor and donor variants as putative protein-truncating (or loss-of-function) variants[2,20]. All predicted variant consequences were defined according to the GENCODE canonical transcript. To further prioritize missense variants, we annotated all variants using Hail (https://hail.is/docs/0.1/annotationdb.html) with classifiers included in the dbNSFP database (such as CADD and PolyPhen), and the MPC score[34,35]. The MPC score had been successfully used to prioritize missense variants in neurodevelopmental disorders, and we later empirically show that variants with a high score were significantly enriched for schizophrenia risk[24,36]. In total, our case-control data set contained 5,649,811 protein-coding variants after QC steps.

Gene set enrichment analysis

Statistical approach

Gene set enrichment analyses aggregate rare variants across a large number of genes to test specific biological hypotheses related to disease pathophysiology and pathogenesis. Similar to common variant GWAS, these analyses relied on individual-level data for regression models that correct for ancestry-level principal components and technical covariates, and would not include joint analysis with external count data from gnomAD database [3,25,30]. For global analyses of enrichment, we identified subsets of each collection with over 100 samples that had well-matched capture technologies and inferred ancestry. Additionally, we ran a second round of PCA for each global population separately to generate population-specific PCs for use as covariates. In total, we identified 22,444 cases and 39,837 controls forming 16 analytical groups that had matched exome coverage, ancestry, and originated from the same population. This sample definition was more restrictive than the ancestry-assignment used for count-level meta-analysis with gnomAD samples, and would only be used for global analyses aggregating hundreds of genes for genotype-dependent enrichment analyses.

We tested for enrichment by regressing schizophrenia status on the total number of ultra-rare damaging variants in the gene set of interest (X_{in}) while correcting for the total number of ultra-rare coding variants in the genome (X_{all}), sex, and five ancestry-derived principal components. X_{all} controlled for exome-wide differences between schizophrenia cases and controls, ensuring that any result was significant above baseline differences. We performed logistic regression on each of the 16 sample groups using Hail's linreg function (https://hail.is/docs/0.2/guides/genetics.html#single-phenotype), and meta-analyzed the betas and standard errors from each regression using inverse-variance weighting.

We first replicated an earlier result that demonstrated that rare coding variants are concentrated in genes under strong constraint in schizophrenia cases compared to controls[3,20,30]. As in those earlier studies, we tested for enrichment in protein-truncating, damaging missense (MPC > 2), other missense (MPC < 2), and synonymous singleton (minor allele count [MAC] = 1) variants, and found consistent signal in constrained genes in the SCHEMA cohorts (Figure S8). We investigated the strength of enrichment in different variant consequences, and found the strongest signal to be in PTVs (both LOFTEE high-confidence and lower-confidence) and damaging (MPC > 2) missense variants (Figure S9). These results were consistent with the mutability-adjusted proportion of singletons (MAPS) of functional categories from the analysis of large reference exomes, and observations from other sequencing studies of neurodevelopmental disorders[21,24]. As elaborated on later, we will focus on these functional classes in subsequent analyses.

Figure S8 Schizophrenia case-control enrichment in constrained genes (pLI > 0.9) in different SCHEMA cohorts. The odds ratio and standard error of PTVs and Synonymous variants are provided for each cohort. The meta-analyzed odds ratio and standard error is calculated using inverse-variance. PTVs show consistent signals across the different cohorts, and synonymous variants do not deviate from expectation.

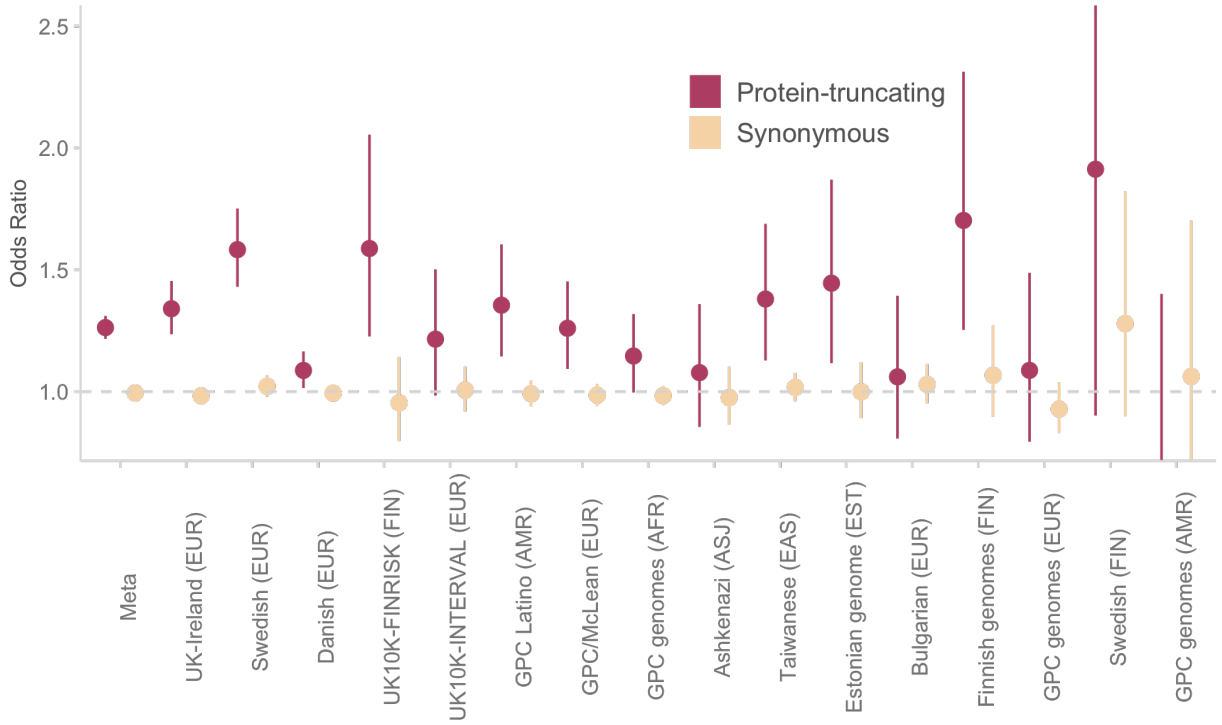
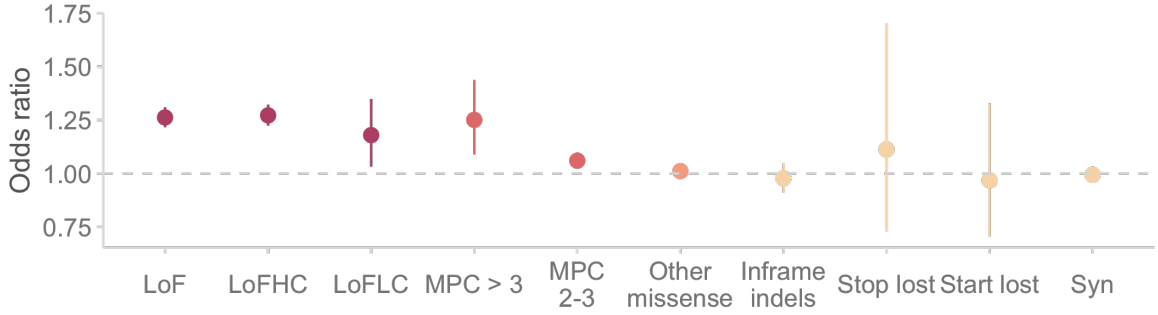


Figure S9 Schizophrenia case-control enrichment in constrained genes (pLI > 0.9) stratified by different variant annotations and inferred consequences. LoF: all loss-of-function or PTVs; LoFHC: high-confidence LOFTEE PTVs; LoFLC: low-confidence based on LOFTEE; MPC > 3: missense variants with MPC > 3; MPC 2 - 3: missense variants with MPC 2 - 3; Other missense: missense variants with MPC < 2; Syn: synonymous variants.



After replicating the known constrained gene result and confirming consistent signals across SCHEMA cohorts, we applied the method to test for enrichment in gene sets based on public databases, other more targeted hypotheses about schizophrenia risk, and other genome-wide screens investigating rare variants in other neurodevelopmental disorders. The strongest signals came from PTVs (LOFTEE high-confidence and low-confidence) and damaging missense (MPC > 3) variants. Motivated by this, we generated enrichment *P*-values for Class I (PTV + MPC > 3) and Class II (MPC 2 - 3) variants separately for each gene set. To calculate a

single test statistic for each gene set, we meta-analyzed the *P*-values using Fisher's combined probability method with *df* = 4.

Description of gene sets

When defining gene sets for analysis, we mapped all gene identifiers to the GENCODE v.19 release, and all non-coding genes were excluded. Here, we describe the main gene sets used in our analysis. **Genic constraint** We used the pLI metric calculated in the gnomAD database as a measure of per-gene selective constraint[20,21]. In total, 3,063 genes had a probability of pLI > 0.9, which we defined as loss-of-function intolerant. We additionally tested for enrichment in the remaining set of 16,134 pLI < 0.9 genes. These enrichment results are displayed in Table S4. **Public pathway databases** To broadly prioritize schizophrenia-relevant processes, we tested for enrichment in 1,732 gene sets defined in an earlier study and aggregated from five public databases, including Gene Ontology, KEGG, PANTHER, REACTOME, and the Molecular Signatures Databases (MSigDB) hallmark processes[30]. In addition, a number of gene sets specific to nervous system biology were also included, including translational targets of FMRP[37], chromatin targets of CHD8[38], splice targets of RBFOX[39], components of the postsynaptic density[40], and neuronal gene lists from the Gene2Cognition database[41]. These results are displayed in Table S8. **SynGO** Because of the specific enrichment of synaptic processes in common and rare risk alleles in schizophrenia, we tested for enrichment in ontology terms defined by the SynGO consortium[42]. The resource systematically annotated synaptic genes and processes after accumulating available research about synaptic biology. We acquired the bulk release from the SynGo website, and tested for enrichment in the 110 biological process (BP) and cellular components (CC) terms that contain at least 10 genes. These results are displayed in Table S7. **Gene expression from GTEx** To test for tissue-specific enrichments, we used gene sets defined in an earlier report that prioritized disease-relevant tissues using common variant data[43]. Briefly, that study computed a per-gene *t*-statistic for specific expression in a focal tissue compared against all other tissues, and identified the 10% of genes with the highest *t*-statistic as enriched in that tissue type. They defined enriched genes in this manner for 43 GTEx tissues[44]. They additionally identified enriched genes between the 13 brain tissues in GTEx for prioritizing specific regions for brain disorders. Results on these gene sets are displayed in Table S8 and S9.

Association studies We specifically explored schizophrenia rare variant enrichment in genes implicated by other genome-wide scans of common variants and broader neurodevelopmental disorders. To investigate the shared genetic signal with common risk variants, we tested for enrichment in genes prioritized from the largest GWAS of schizophrenia to date, which identified associations at 270 distinct loci[45]. Statistical fine-mapping using FINEMAP prioritized 76 genes, of which 69 are protein-coding[46]. A recent analysis of *de novo* mutations from 31,058 DD/ID trios implicated 299 genes[47], and a parallel analysis of 11,986 ASD cases identified 102 genes at FDR < 10%[24]. These genes are listed in Table S11, and the corresponding enrichment results are displayed in Table S12.

Incorporation of external controls

To increase power for gene discovery, we incorporated external samples from non-psychiatric and non-neurological collections aggregated and called as part of the gnomAD consortium effort. The inclusion of these samples introduced technological and methodological challenges that need to be addressed. First, only aggregated variant counts could be acquired due to the design and mission of the gnomAD consortium; as such, analytical methods applied here would necessarily be limited to count-based data. Second, the external count data must be stratified in a way that allows for the control of potential batch effects from ancestry and capture. Fortunately, gnomAD samples were well-matched with the SCHEMA samples for the following reasons. First, the majority of the gnomAD data were also generated at the Broad Institute of Harvard and M.I.T., and Agilent v2 and Illumina Nextera were similarly the most commonly used capture kits (35.47% and 47.42% of all samples respectively). Second, all samples in gnomAD and SCHEMA efforts were re-processed and joint called using the same pipeline. We additionally applied the same genotype and site filters to the gnomAD subset that excluded SCHEMA samples. Therefore, we need only aggregate variant counts by exome capture and global ancestry before jointly analyzing with SCHEMA.

Concordance with imputed platform and ancestry

The gnomAD consortium performed platform imputation to infer the capture and sequencing platform metadata from all samples[21]. Briefly, they ran PCA on variant call rates in exome capture intervals and identified sample clusters using HBDSCAN. Because a large number of our samples were included as part of the gnomAD v2 release, we checked the concordance between imputed and known capture in 47,345 samples. Here, we observed a concordance of 99.9%, suggesting that we could accurately impute platform from gnomAD samples for case-control matching.

Similarly, appropriate ancestry matching was required to jointly analyze our samples with counts from the gnomAD database. In samples shared between SCHEMA and gnomAD, we checked the concordance of imputed ancestry between gnomAD and our random forest model. Concordance of 97.7% was observed between SCHEMA and gnomAD imputed ancestries, suggesting that the population assignment was reasonably well-matched between the two data sets.

Final sample counts for meta-analysis

After excluding all SCHEMA samples, we generated variant-level counts from gnomAD stratified by imputed capture and ancestry. The gnomAD platform assignment identified 16 different captures, of which 82.89% of samples were generated using only two captures (Agilent v2 and Illumina Nextera)[21]. Because some of the remaining captures may not be well-represented in our exome coverage analysis, we excluded all gnomAD samples that were not generated using the Agilent v2 and Illumina Nextera captures. In total, we identified counts from 46,885 gnomAD samples (46,465 exomes and 420 genomes) for inclusion in our meta-analysis.

When combined with counts from SCHEMA, we identified 11 strata based on sequencing platforms and inferred population (Table S3). In total, the meta-analysis included variants counts from 24,248 cases, 50,437 internal controls, and 46,885 external controls.

Table S3 Number of case-control samples included in the count-based case-control meta-analysis. We provided sample counts per stratum, and control numbers are split between samples included in the SCHEMA call set and external samples from gnomAD.

Stratum	Cases	Controls (SCHEMA)	Controls (gnomAD)
All samples	24248	50437	46885
EUR (Exomes)	8874	19074	23561
EUR (Exomes, non-Nextera)	7277	11187	0
AMR (Exomes)	1388	3146	12008
FIN (Exomes, non-Nextera)	944	7984	3542
EAS (Exomes)	1730	1607	6806
AFR (Whole Genomes)	2245	1170	420
ASJ (Exomes)	869	2415	548
EST (Whole Genomes)	261	2281	0
FIN (Whole Genomes)	423	655	0
AFR (Exomes)	127	765	0
SAS (Exomes)	110	153	0

Strategy for meta-analysis

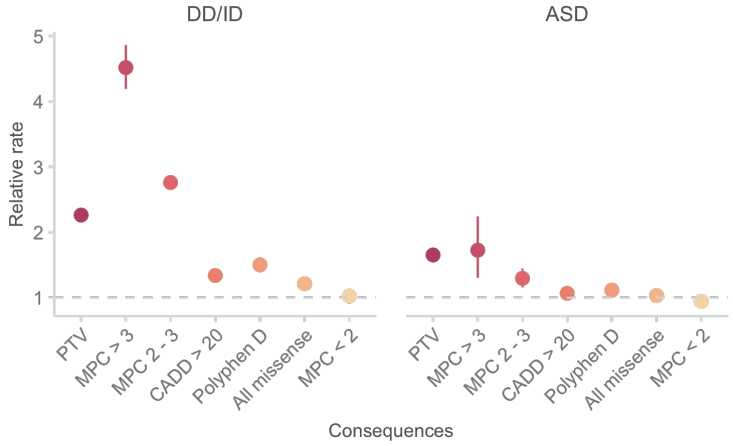
Incorporation of two classes of pathogenic variants for gene discovery

To increase power for discovery, we enriched for pathogenic variants by restricting our analysis to ultra-rare variants (minor allele count [MAC] ≤ 5 studywide) that are either PTVs (defined as stop-gained, frameshift, and essential splice donor and acceptor variants) or missense variants prioritized by a pathogenicity classifier. We sought to empirically evaluate which classifier most significantly enriched for damaging missense variants using external sequencing data sets of other neurodevelopmental disorders. We annotated *de novo* mutations from published ASD, DD/ID, and unaffected (sibling) trios with CADD, PolyPhen, and MPC missense annotations[24,48]. Using the Poisson exact test, we determined which classifier resulted in the most significant difference in *de novo* rates between affected and unaffected (sibling)

probands. The MPC classifier most powerfully prioritized damaging missense variants in both ASD and DD/ID trios (Figure S10). Furthermore, missense variants at the top bin of pathogenicity (MPC ≥ 3) have a global signal on par with PTVs in schizophrenia and neurodevelopmental disorders, while variants with MPC ≥ 2 have a significant but weaker signal than PTVs. We confirmed this result by analyzing the case-control burden of an independent case-control data set of autism and ADHD[49], and observed that MPC ≥ 3 and MPC ≥ 2 were also significantly enriched for rare variant risk (Figure S11). Motivated by consistent observations across data sets and types of genetic variation, we chose to meta-analyze three classes of variants (PTVs, missense variants with MPC > 3 and MPC 2 - 3) for the purposes of gene discovery.

Figure S10 Results from Poisson exact tests quantifying the strength of enrichment after stratifying *de novo* mutations by variant consequences and missense classifiers. For demonstrative purposes, we show the difference in enrichment using MPC, CADD and PolyPhen-2. **A**: The enrichment analysis is across all protein-coding genes. **B:** The comparison is restricted to constrained genes (pLI > 0.9). All tests other than PTV restrict to missense variants. PolyPhen D: predicted damaging by Polyphen; All missense: no classifier applied.

A.



B.

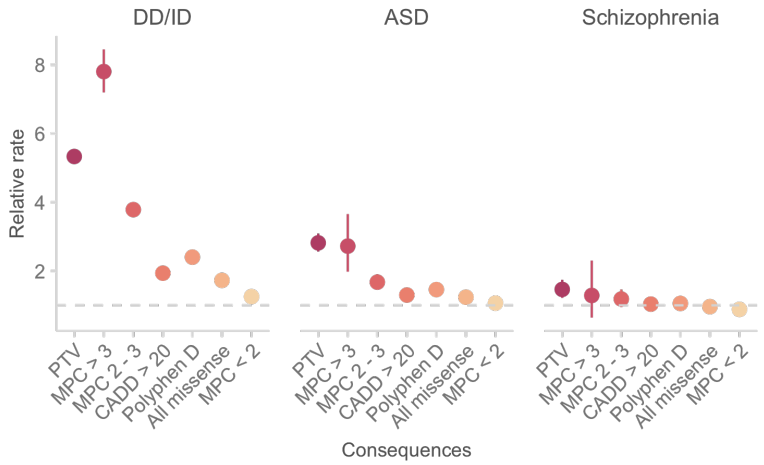
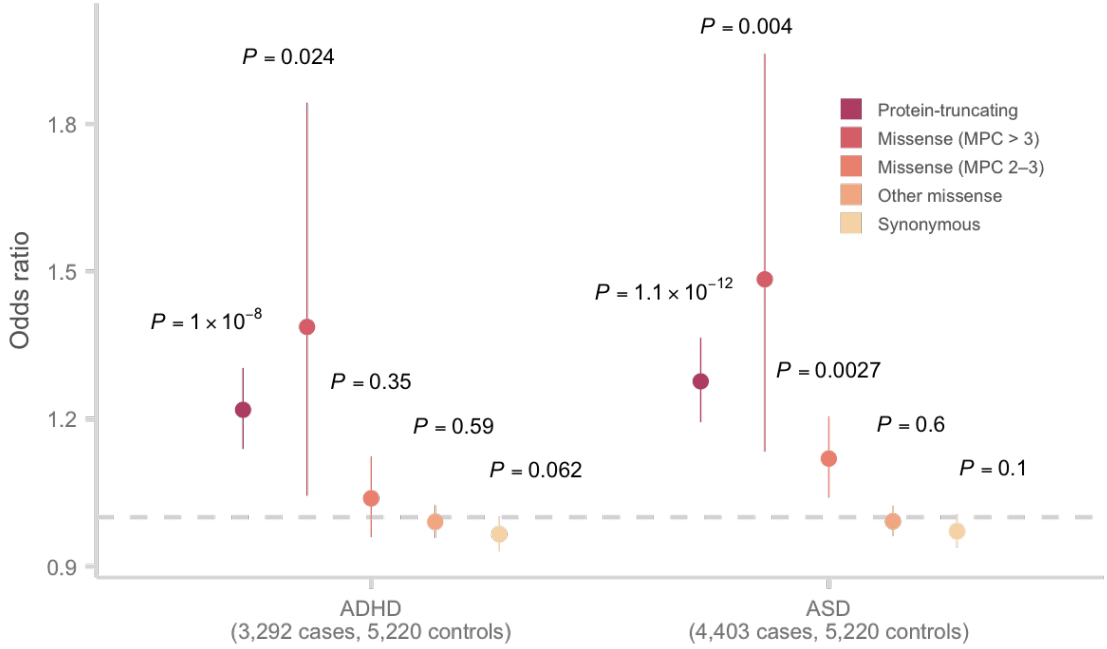


Figure S11 Enrichment of URVs in ASD and ADHD cases compared to controls stratified by variant annotation and consequences in constrained genes (pLI > 0.9). Results were calculated using logistic regression as described for schizophrenia case-control samples in the Supplementary Text.



In earlier exome sequencing studies, various allele frequency thresholds had been applied to enrich for pathogenic protein-coding variation, including restricting analyses to singletons (MAC = 1)[3,40], MAC <= 5[25], MAF < 0.1% or 0.5%[40], and exclusion based on presence in a public reference[50]. To better understand which frequency cut-off to use in our analysis, we tested for enrichment of PTV and missense > 2 variants in constrained (pLI > 0.9) genes in two additional frequency bins: MAC 2 - 5 and MAC 6 - 10 as defined in our data set. We applied the same approach used for gene set enrichment analyses, meta-analyzing the signal from Class I and Class II variants. We observed a signal for both Class I and II variants in the MAC 2 - 5 bin (P_{meta} = 9.6 x 10^{-4}; $P_{Class\ I}$ = 0.023; $P_{Class\ II}$ = 0.0075), and limited to no enrichment in the MAC 6 - 10 bin (P_{meta} = 0.046; $P_{Class\ I}$ = 0.061; $P_{Class\ II}$ = 0.23). This result is consistent with the allele frequencies of PTVs disrupting the 3,063 constrained genes in our call set: of the 28,096 total PTVs, 20,632 (73.4%) were singletons, and 27,067 (96.3%) were MAC < 5. Therefore, similar to an earlier analysis of ASD exome sequence data[24,25], we restricted our gene discovery analysis to variants with MAC < 5. At this frequency cut-off, the genotype matrix for PTVs and MPC > 2 variants was very sparse, with most samples carrying no PTVs or MPC > 2 variants in a single gene. In the 24,248 cases and 50,437 controls for which we have full genotype data, the gene-by-sample matrix (74,685 samples x 18,321 genes = 1,368,303,885 entries) created by aggregating up the total number of PTV and MPC > 2 alleles in each gene and each individual had only 318,562 non-zero entries, of which nearly all (315,624 or 99.1%) had a total of one allele.

Robustness of combined analysis

We conducted the following analyses to ensure the robustness of the results generated by this approach. First, we find that Fisher's exact tests of synonymous variants performed on each stratum controlled well for Type I error rates (Figure S12). In these tests, we observed a deflation of test statistics in strata with small sample sizes due to low observed alleles in each gene. Second, we observed no inflation of synonymous *P* values in genes with total alleles counts > 10, > 50, and > 100 where we have greater power to detect potential artifacts, even when using the Mantel–Haenszel test (Figure S14). Third, we observed the expected null distribution of *P* values in our gene-based tests when analyzing synonymous variants (Figure S13).

Figure S12 QQ plot of gene burden *P*-values calculated in each stratum using Fisher's exact test. We display gene burden *P*-values for Class I (PTV and MPC > 3; dark red), Class II (MPC > 2; orange), and synonymous variants (yellow). Strata were sorted by total sample size from left to right, top to bottom. Some deflation is observed due to low per-gene counts based on the asymptotic properties of the Fisher's exact test.

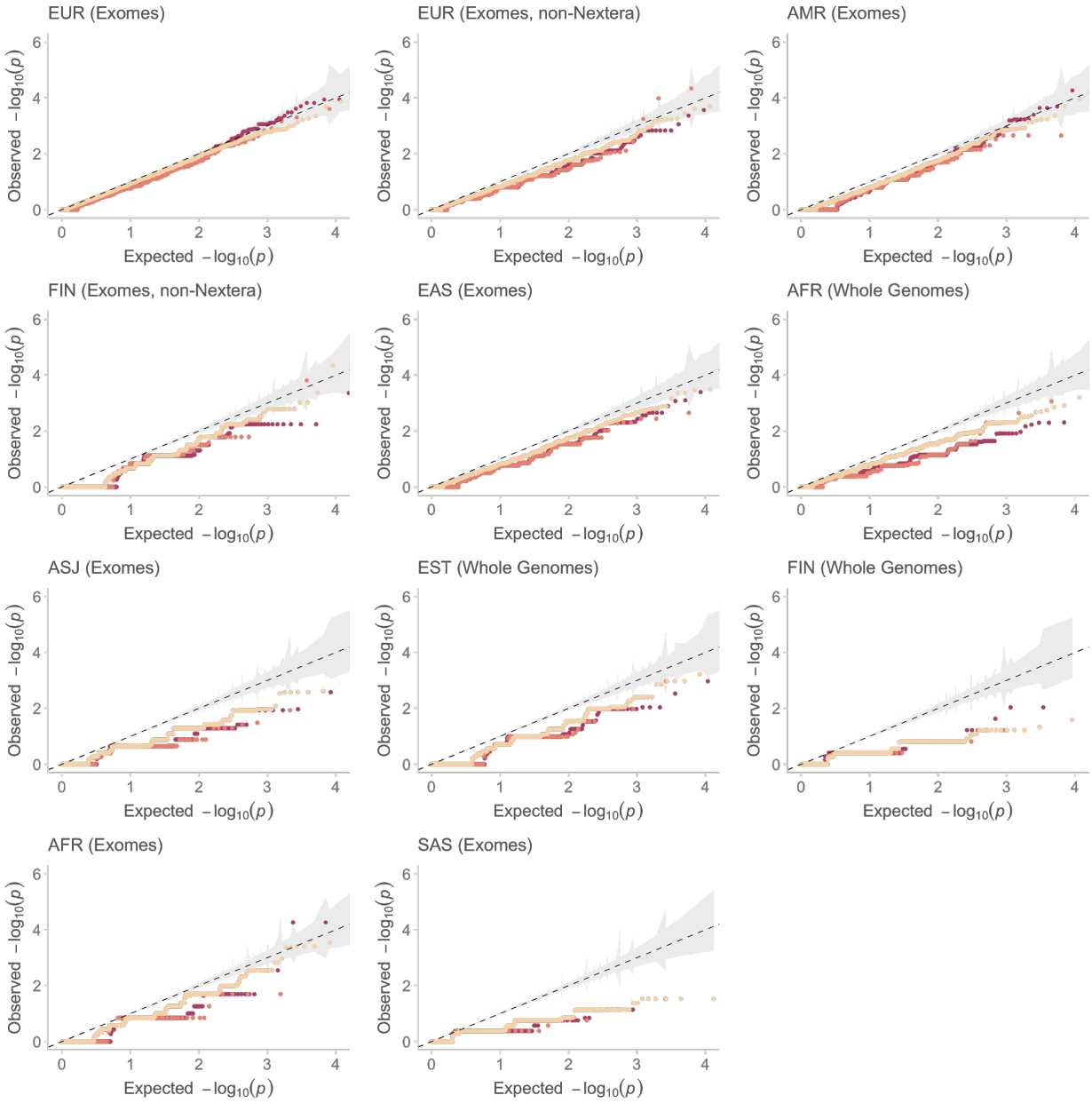


Figure S13 QQ plot of gene burden *P*-values of synonymous variants with and without strata adjustment. The uncorrected P-values were calculated from the naive Fisher's exact test not adjusting for strata, and the permuted Fisher's exact test adjusted for strata.

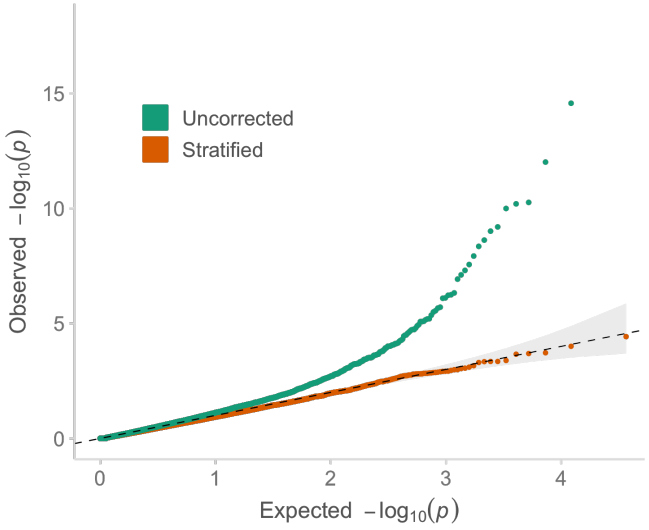


Figure S14 QQ plot of gene burden *P*-values of synonymous variants calculated using the Mantel-Haenszel test at four different gene-wide allele count thresholds. When restricting to Alleles > 0, all genes were included. When restricting to Alleles > 10, only genes with a total allele count of > 10 across cases and controls are included. Gene-wide count thresholds of 50 and 100 are also included.

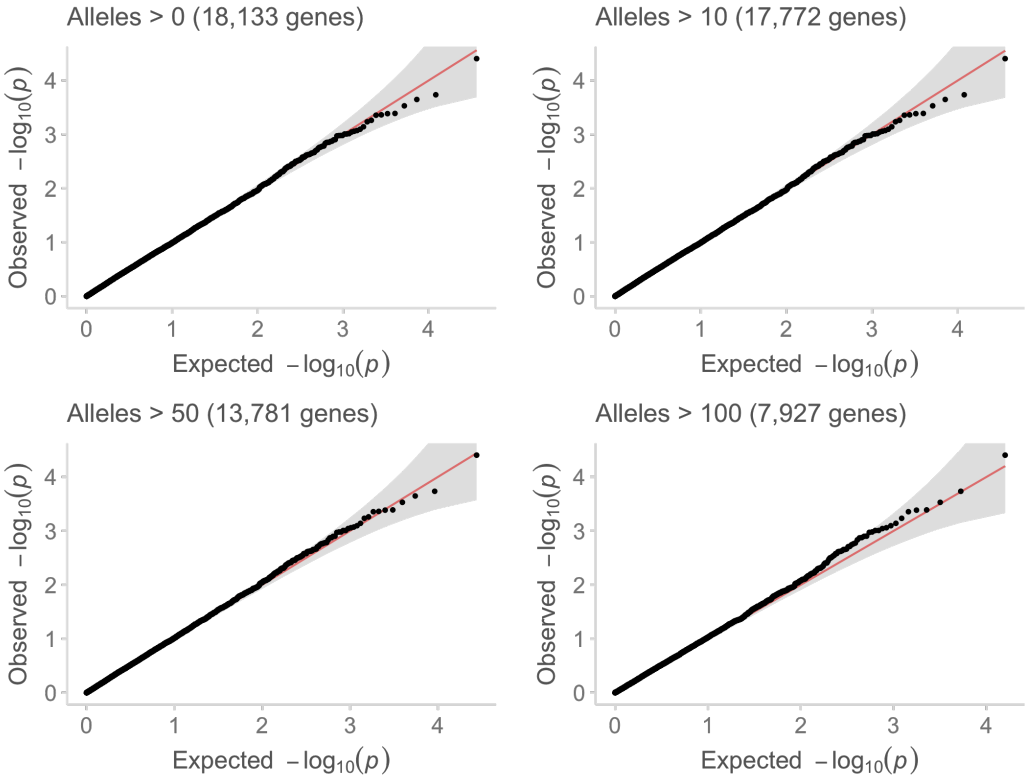
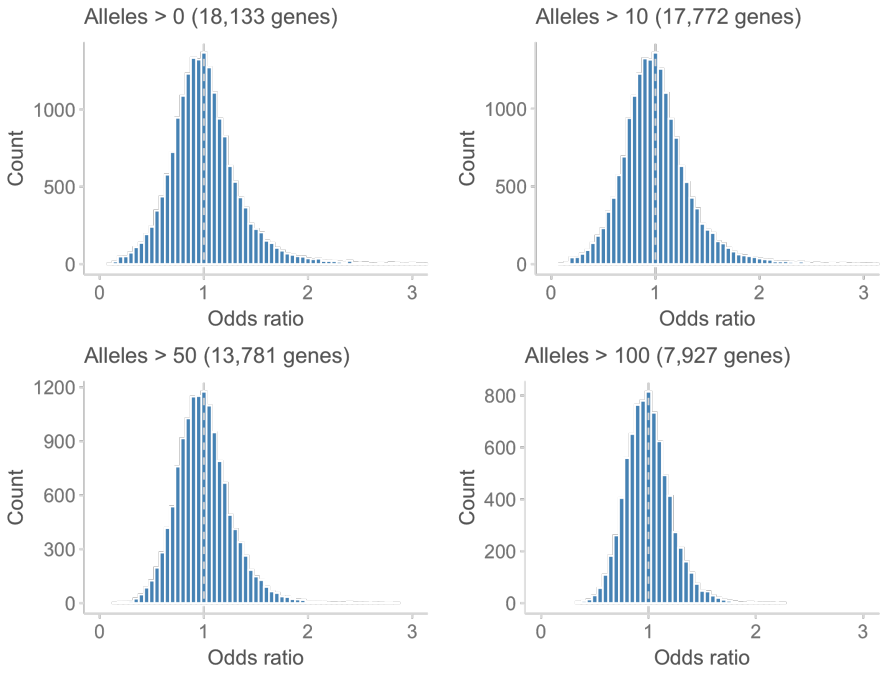


Figure S15 Histogram of odds ratios from the synonymous Mantel-Haenszel test at four different gene-wide allele count thresholds.



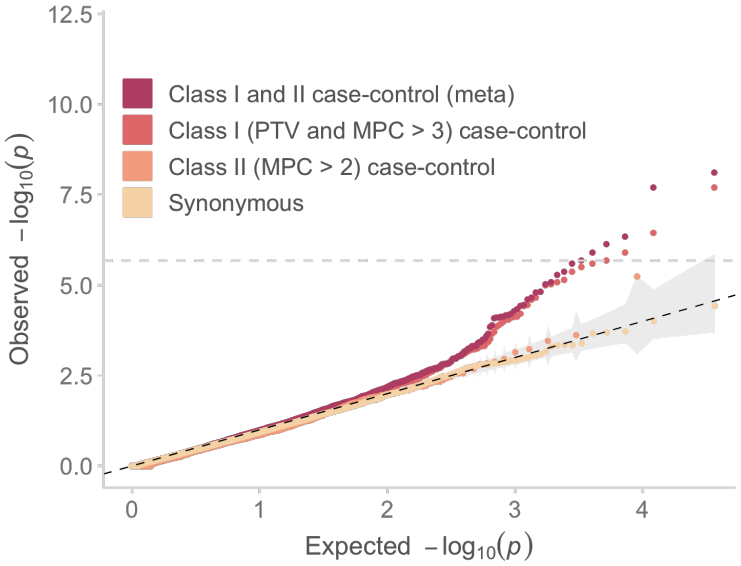
Permutation-based Fisher exact test for gene discovery

Because only summary-level variant counts were available for the external controls, we tested for an excess of disruptive variants per gene using a Fisher's exact test in which statistical significance was determined by case-control permutations within each population stratum (Table S3). The sparsity of the observed count data, as discussed in the previous section, suggested that the exact test would be an appropriate choice to test for burden. To evaluate significance, we first calculated the observed *P*-value (P_0) from the exact test of the observed counts. We then randomly shuffle the observed counts between case and control samples within each stratum to calculate null *P*-values that controlled for ancestry and batch-specific differences. We performed 1 million permutations to generate the null distribution, and calculated per-gene *P*-value as the proportion of permuted *P* that was less than or equal to P_0. For genes with $P \leq 1 \times 10^{-3}$, we performed 50 million permutations in a similar manner to generate a more accurate estimate of the true *P*-value. When performed on synonymous variants, stratified permutation sufficiently controlled for ancestry and batch-specific differences when compared to an unstratified exact test, providing confidence that our approach adequately controlled for false positives (Figure S13).

We first tested the burden of Class I variants (PTVs and mis3) to generate a *P*-value for 18,321 genes. Shown earlier, we found the strongest enrichment in PTVs (both high- [HC] and low-confidence [LC] LOFTEE) and MPC > 3 variants (Figure S9, Figure S10, Figure S11). Notably, LC PTVs as a class were marginally less enriched than high-confidence PTVs, and only a small fraction of genes have any MPC > 3 missense variants. While 18,051 genes have

at least one PTV in cases or controls within our sample, only 1,131 genes have any MPC > 3 missense variants, and 7,594 genes have one or more LC PTV variants. Thus, it would be useful to consider a test for high-confidence PTV burden alone for all genes in addition to tests that include LC PTVs and MPC > 3 variants. Crucially, these tests would be highly correlated since a test of all PTVs and MPC > 3 variants contain all variants included in a test of HC PTVs. To leverage this property, we adopted the min-*P* procedure to empirically correct for multiple testing, as described in [40,51]. For each gene, we performed burden tests of HC PTVs, all PTVs, HC PTVs and MPC > 3 variants, and all PTVs and MPC > 3 variants. The same order of permutations was applied for across all tests of the same gene, and the minimum *P* value for each permutation across all tests were retained. A joint null distribution of these minimal *P* values was used to determine the significance of each gene, providing control of the family-wise Type I error rate across tests. For each gene, we report the smallest *P*-value from the tests of Class I variants.

Figure S16 QQ plot of case-control *P*-values of gene burden tests of Class I, Class II, and synonymous variants. Gene *P*-values for Class I, Class II, and synonymous variants are calculated using the permuted Fisher's exact test. The meta-analysis gene *P*-value is calculated from the weighted Z-score method.



Combining Class I and Class II variants with empirical weights

For each gene, we additionally calculated a combined *P*-value of Class I and Class II variants. Using the permuted Fisher's exact test, we first calculated a burden *P*-value of Class II missense variants in 4,512 genes with MPC > 2 variants. We then meta-analyzed Class I and Class II *P* values using the weighted *Z*-score method, or Stouffer's method[52]. To appropriately consider the information content between Class I and II variants, we defined the relative weight as the ratio of the standardized effect sizes of these classes of variants as observed from enrichment analyses of constrained genes (Figure 1C, Figure S11). For Class I variants, we find $w_1 = \beta / \sigma^2 = 0.233/0.0183$; for Class II variants, $w_2 = \beta / \sigma^2 = 0.058/0.018$. The relative weight between the two classes is 3.97. In total, we performed 18,321 independent tests for Class I

variants, and 4,512 tests for a combined test of Class I and II variants, for a total of 22,833 tests (Figure S16). The reported case-control odds ratios and confidence intervals were calculated from raw counts without taking into account ancestry and batch-specific differences.

Incorporation of *de novo* mutations from schizophrenia trios

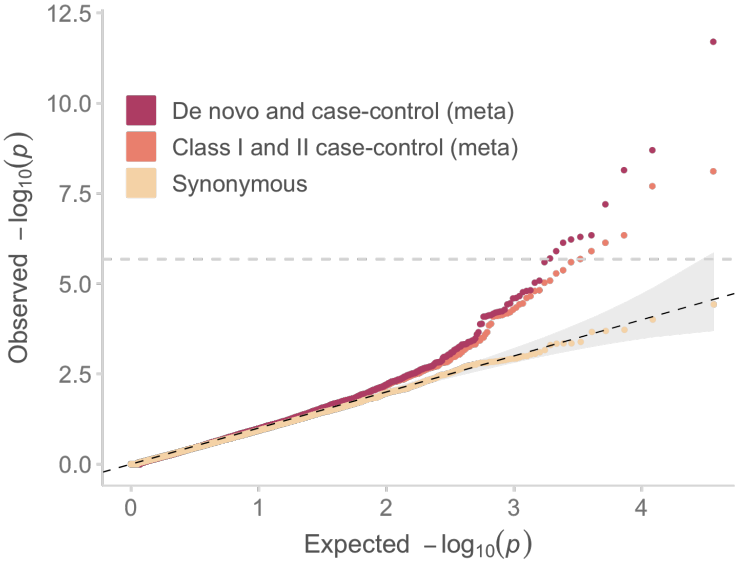
We aggregated and re-annotated validated *de novo* mutations from 10 published studies of schizophrenia trios for analysis with our case-control cohort[1,4,10–17]. *De novo* mutations from 3,402 parent-proband trios were available for joint analysis. We modeled the recurrence of *de novo* mutations as the Poisson probability of observing ≥ N *de novo* mutations in a gene given a baseline gene-specific mutation rate[53]. The gene-specific mutation rates had been applied in analyses of large-scale neurodevelopmental disorders and autism spectrum disorders[24,47]. We additionally adapted the model to produce mutations rates for MPC > 2 and MPC > 3 missense variants, which allow us to calculate the Poisson probabilities of observing Class I and Class II *de novo* mutations. For each gene of interest, we calculated a Poisson *P*-value for Class I mutations and a *P*-value for Class I and II mutations.

Previous studies have successfully integrated case-control and *de novo* mutations for gene discovery in neurodevelopmental disorders. In our study design, the case-control analysis has greater power for gene discovery due to much larger sample size. Furthermore, unlike other neurodevelopmental disorders which are significantly enriched for *de novo* mutations, *de novo* mutations in schizophrenia are apparently much sparser[2]. Indeed, only 325 genes had one or more *de novo* protein-truncating variants, while only 449 genes had one or more Class I (PTV + MPC > 3) and II (3 > MPC > 2) mutations. The remaining genes in the genome had no signal in the class of variants most enriched for schizophrenia risk. In addition, only 18 genes had the more unlikely event of two or more *de novo* PTVs, while 35 genes had two or more Class I and II *de novo* mutations. Despite this, we found that *de novo* Class I and Class II variants are enriched for 244 genes with $P < 0.01$ in our case-control analysis (Figure 1D).

Motivated by these observations, we used *de novo P* values to increase power for the 244 genes with $P_{meta} < 0.01$ in our case-control analysis (Figure S17). For each of these genes, we calculated a Poisson *P*-value for Class I mutations and a *P*-value for Class I and II mutations, and meta-analyzed the *de novo* and the case-control burden *P*-values using the weighted Z-score method. We report the minimum *P* value across these tests (case-control only [no *de novo*, 22,833 tests], *de novo* Class I and case-control [244 tests], *de novo* Class I and II and case-control [244 tests]) as the study-wide *P* value. To identify the ideal weight, we simulated uniform chi-square values for the *de novo* test, and calculated the equivalent chi-square value for the case-control test as a multiple of the *de novo* value. We set this multiple as 6 since there was approximately six times the number of cases in our case-control data set than in the trio data set. We then transformed the chi-square values to *P*-values and meta-analyzed them using the following weights in the weighted Z-score method: w_i: {0.5, 1.0, 2.0, 5.0, 10.0, 15.0}. The minimal *P*-value was not achieved using neither equal weighting (1.0) nor by dramatically over-weighting (> 5.0) the case-control *P*-value. From these results, we applied a $w_i = 2$ to combine the case-control and *de novo P*-values, although we note that the top 10 genes

remain exome-wide significant even if equal weighting ($w_i = 1$) was applied. When including the *de novo* stage, we performed 23,321 (18,321 + 4,512 + 244 x 2) tests for the purposes of gene discovery, which results in a Bonferroni threshold of 2.14×10^{-6}.

Figure S17 QQ plot of case-control and *de novo* combined gene *P*-values, case-control only gene *P*-values, and synonymous gene *P*-values.



Gene set enrichment results

Full enrichment results were displayed in Table S4, Table S6 to S9, and Table S11. In Figure S18, we show the distribution of test statistics from the 1,732 gene sets derived from databases of biological pathways and experimental data. As expected, synonymous variants had limited enrichment while a strong aggregate signal was observed for PTVs and MPC > 2 missense variants in synaptic and neuronal pathways. We additionally displayed rare variant enrichment in GTEx tissues in Figure S19, which highlighted the notable enrichment of risk URVs in brain tissues.

Figure S18 QQ plot of the enrichment analyses of the 1,732 gene sets derived from databases of biological pathways and experimental data. We display the gene set burden statistics of different variant classes (Class I, Class II, other missense [MPC < 2], and synonymous variants) alongside the meta-analyzed test statistics integrating Class I and Class II variants (Combined). We label the gene sets with $P < 2.9 \times 10^{-5}$, the Bonferroni threshold correcting for the described tests (Table S6).

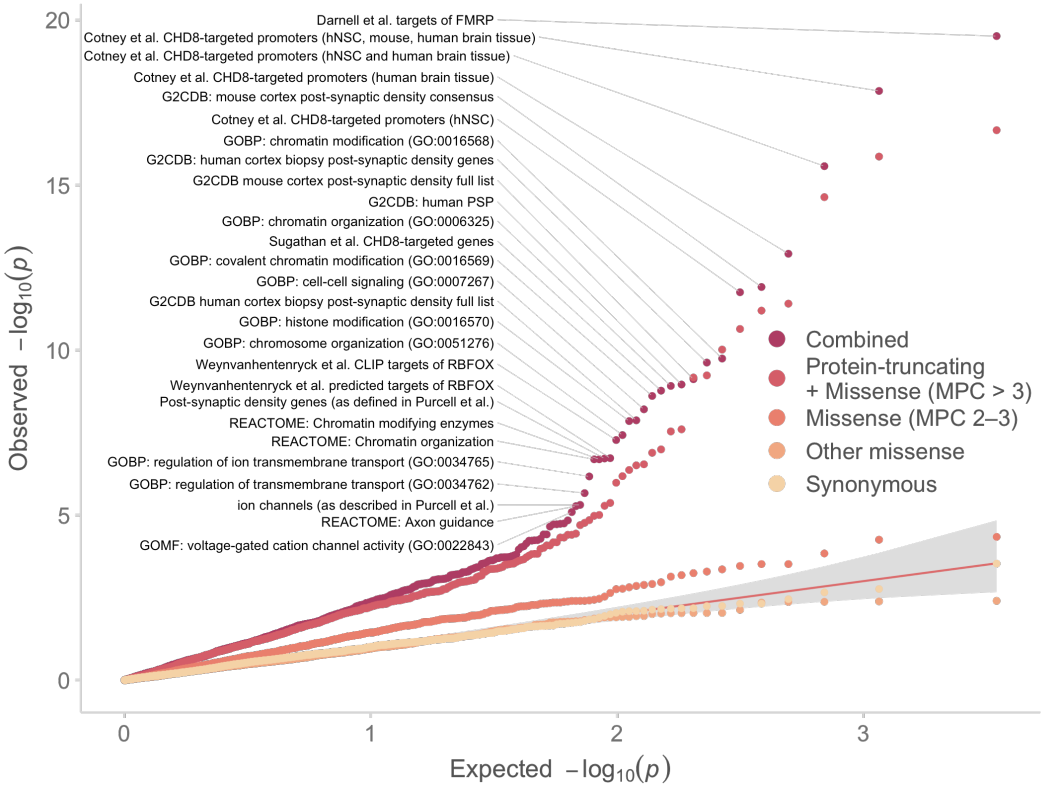
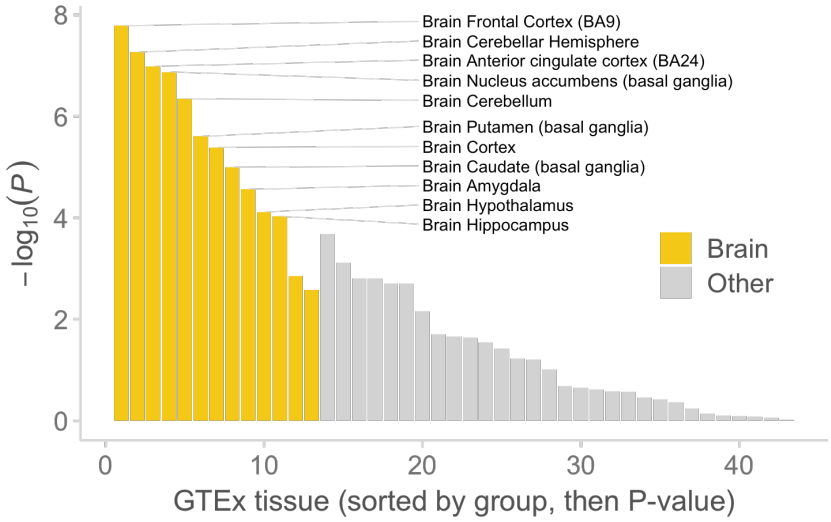


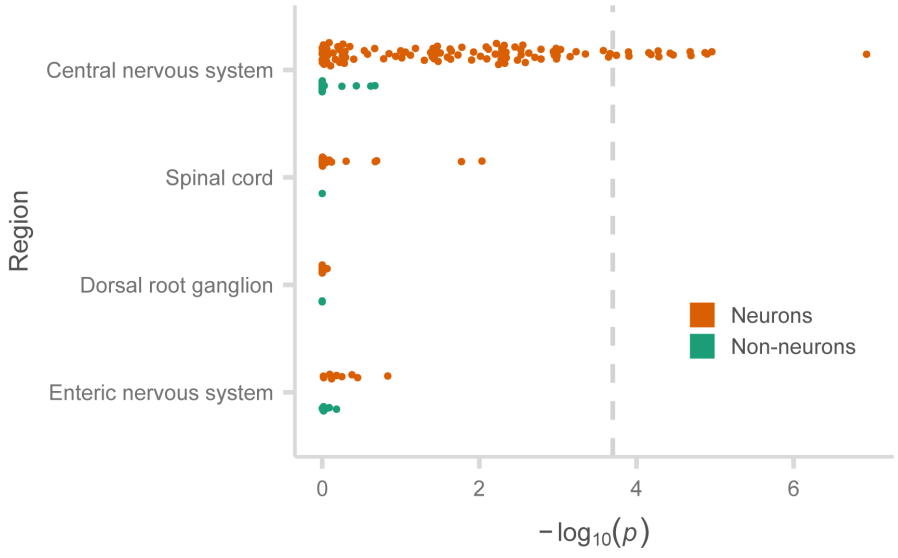
Figure S19 Schizophrenia case-control gene set enrichment in brain and non-brain GTEx tissues. We test for the burden of rare PTVs in genes with the strongest specific expression in that tissue type relative to other tissues as defined in Finucane *et al.* Each bar is a different tissue in GTEx, grouped by whether it is part of the central nervous system and sorted by *P*-value (Table S8).



Spatiotemporal expression of risk genes

Single cell atlas To prioritize cell types as possibly involved in schizophrenia, we investigated which cells show the highest specific expression for the 32 FDR < 5% schizophrenia genes. We re-processed data from a single cell atlas of the mouse nervous system based on the RNA sequencing of half a million cells[54]. First, we downloaded the raw cell count data with cluster annotations from that study. Within each defined cell type, we summed the UMIs for each gene, and normalized these counts by dividing by the total UMIs across all genes in that cell type. This value, which we called *p*, represented the proportion of total UMIs in a cell type that were of the gene of interest. To enable comparisons of genes between clusters, we re-scaled the *p* of each gene to have mean 0 and standard error of 1, essentially turning *p*s into *Z*-scores. A cell type with a high *Z*-score for a particular gene suggested that the expression of that gene was higher in that cluster than in the remaining clusters. We restricted subsequent analyses to mouse genes that mapped to equivalent human protein-coding genes, leading to a *Z*-score matrix of 16,198 genes by 265 cell types. For each cell type, we performed a one-sided Wilcoxon rank sum test comparing the expression *Z*-scores of the 32 FDR < 5% genes against the values of all remaining genes. We reported the *P*-values of these 265 tests in Table S10. Because of the correlated expression patterns of different neuronal types and incompleteness of existing cell atlases, it would be difficult to prioritize individual cell types. However, when subdividing these cells into neurons and non-neurons (glial, astrocytes, vascular, and immune cells) and by region (CNS and PNS), it was clear that the bulk of the signal is for neurons in the CNS and not for non-neurons or PNS cells (Figure S20).

Figure S20 Strongest relative expression of schizophrenia-associated genes in single cell data from the mouse central nervous system. For each cell type, the distribution of normalized expression values is compared between schizophrenia risk genes (FDR < 5%) and all other genes using a Wilcoxon rank sum test. Each dot is a single cell type as defined in Zeisel et al. (2018) stratified by four anatomical locations.



Temporal expression To investigate temporal expression of risk genes, we acquired bulk RNA-seq data from Brainspan, an atlas of the developing human brain[55]. The available data included brain tissue collected from 42 individuals that profiled up to 16 distinct regions spanning pre- and postnatal-development. We acquired the expression matrix from the BrainSpan website, and converted the RPKM to TPM values by normalizing the total transcripts per sample to one million. Because available brain regions differed between early prenatal and postnatal samples, we averaged the TPM across all brain regions per donor to serve as an estimate of temporal expression across the whole human brain. Each donor was assigned to a specific developmental period as defined by the BrainSpan data set (Figure 3B and Figure S24). Because of low numbers for postnatal donors, we created larger temporal groupings for those samples, defining infancy as birth to 1 years old (yo), childhood as 1 yo to 12 yo, adolescence as 12 yo to 20 yo, and adulthood as greater than 20 yo. Figure 3B and Figure S24 for schizophrenia-associated genes are boxplots of these TPM values and categories, with the smoothed line calculated using loess (y ~ x) across all samples. We additionally sought to create an estimator for the degree of prenatal versus postnatal expression for each gene. For each gene, we performed a *T*-test between the TPMs from early to mid-pre-natal samples against adolescence and adulthood samples. Genes with negative *T*-values were expressed more prenatally while those with positive *T*-values were expressed more postnatally. All protein-coding genes were ranked according to these *T*-values, which were used in Figure S23 to test for pre- and postnatal biases in expression in associated genes.

Comparisons between schizophrenia and other neurodevelopmental disorders

We aggregated and annotated *de novo* mutations from 31,058 DD/ID probands, 6,430 ASD probands, and 2,179 unaffected sibling (control) probands from two published studies[24,47]. Variant annotation was performed in the same manner as our schizophrenia case-control analysis. We identified *de novo* PTV and MPC > 2 missense mutations in schizophrenia-associated genes, and displayed them in gene plots in Figure 5C, 5D, and S22. To calculate *de novo P*-values in DD/ID and ASD probands, we applied the same test as for schizophrenia gene discovery and calculated *P*-value as the Poisson probability of observing ≥ N *de novo* mutations in a gene given a baseline gene-specific mutation rate. We calculated *de novo* gene *P*-values for Class I and Class II variants for comparison between disorders, with a focus on genes associated with schizophrenia (Figure 5B, Table S14). To compare the *de novo* enrichment between disorders in pLI genes (Figure S25), we used the Poisson exact test to calculate differences in *de novo* mutation rates between schizophrenia, ASD, and DD/ID probands and sibling (control) probands. Counts in PTVs, MPC > 2 missense, other missense, and synonymous variants were tested separately. We displayed the one-sided *P*-values, rate ratio, and 95% confidence interval in Figure S25, and reported the precise figures in the main text under section “Contribution of ultra-rare PTVs to schizophrenia risk”.

We explored what properties may differ between schizophrenia and DD/ID-associated risk genes, and hypothesized that DD/ID genes were under stronger evolutionary constraint with

a bias towards prenatal expression when compared to schizophrenia genes. To perform this analysis, we annotated each protein-coding gene with observed/expected (o/e) values from gnomAD and rank of prenatal expression bias as estimated from Brainspan (described earlier)[21,55]. o/e was calculated as the degree of depletion of PTVs in the gnomAD database, and could be seen as a continuous measure of genic constraint. Because it was clear that schizophrenia and DD/ID risk genes were constrained, we tested if DD/ID genes were under stronger constraint by performing a Wilcoxon Rank Sum test between the o/e values of DD/ID risk genes and schizophrenia FDR < 5% genes[47]. We additionally compared the o/e values of ASD and schizophrenia risk genes, and displayed the results in Figure S22[24]. We then attempted to test if schizophrenia, DD/ID, and ASD risk genes displayed pre- or postnatal bias. We performed a Wilcoxon rank sum test comparing the prenatal percentiles of each set of risk genes against the genome-wide background, and displayed the results in Figure S23. We found that schizophrenia risk genes did not display a prenatal bias in expression, while DD/ID genes were overwhelmingly prenatal in expression. We additionally plotted the expression trajectories of individual schizophrenia risk genes in Figure S24, in which individual genes like *SETD1A*, *TRIO,* and *SP4* exhibit prenatal expression while *GRIN2A* and *GRIA3* show postnatal expression.

Figure S21 The gene plot displays the protein-coding variants that contribute to the exome signal in *SETD1A*. Variants discovered in cases are plotted above the gene, and those from controls are plotted below. Each variant is colored based on inferred consequence, and the protein domains of the gene are also labelled.

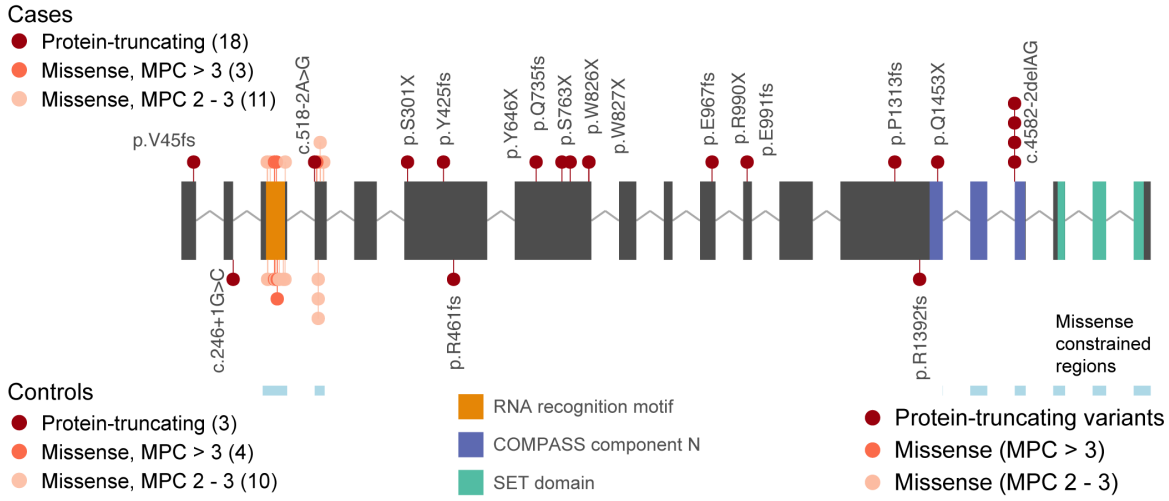


Figure S22 Constraint score distributions of schizophrenia, ASD, and DD/ID genes. The observed-to-expected (o/e) constraint values from the gnomAD database are plotted for associated genes in each trait. Genes with o/e closer to 0 are more constrained, while genes with o/e closer to 1 are less constrained. The *P*-value is calculated from comparing the o/e distribution of schizophrenia genes to each of the other disorders using a Wilcoxon rank sum test.

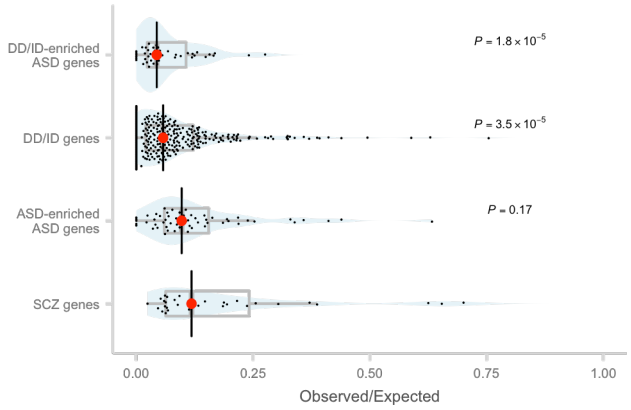


Figure S23 Temporal expression patterns of schizophrenia, ASD, and DD/ID genes. All genes are ranked according to degree of pre- to postnatal expression as calculated from Brainspan data[55]. Genes ranked closer to the 0th percentile are maximally prenatally, while genes ranked closer to the 100th percentile are maximally expressed postnatally. We plot the temporal expression rank of associated genes in each trait. The *P* value is calculated from comparing the rank of each set of associated genes to the remaining genes in the genome using a Wilcoxon rank sum test.

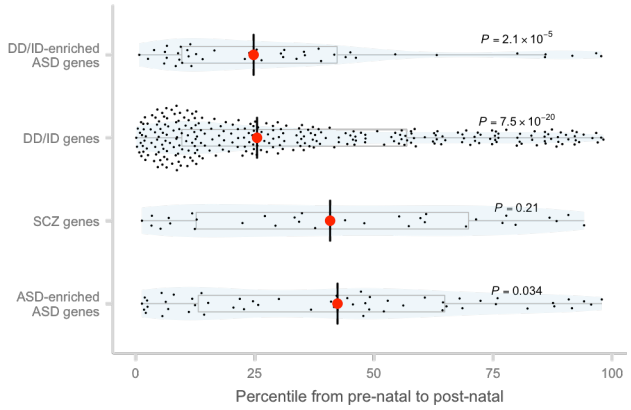


Figure S24 Temporal expression of significant genes in the human brain. We show expression in four prenatal and four postnatal periods derived from whole-brain tissue in BrainSpan[55]. The expression values plotted are in transcript-per-million (TPM).

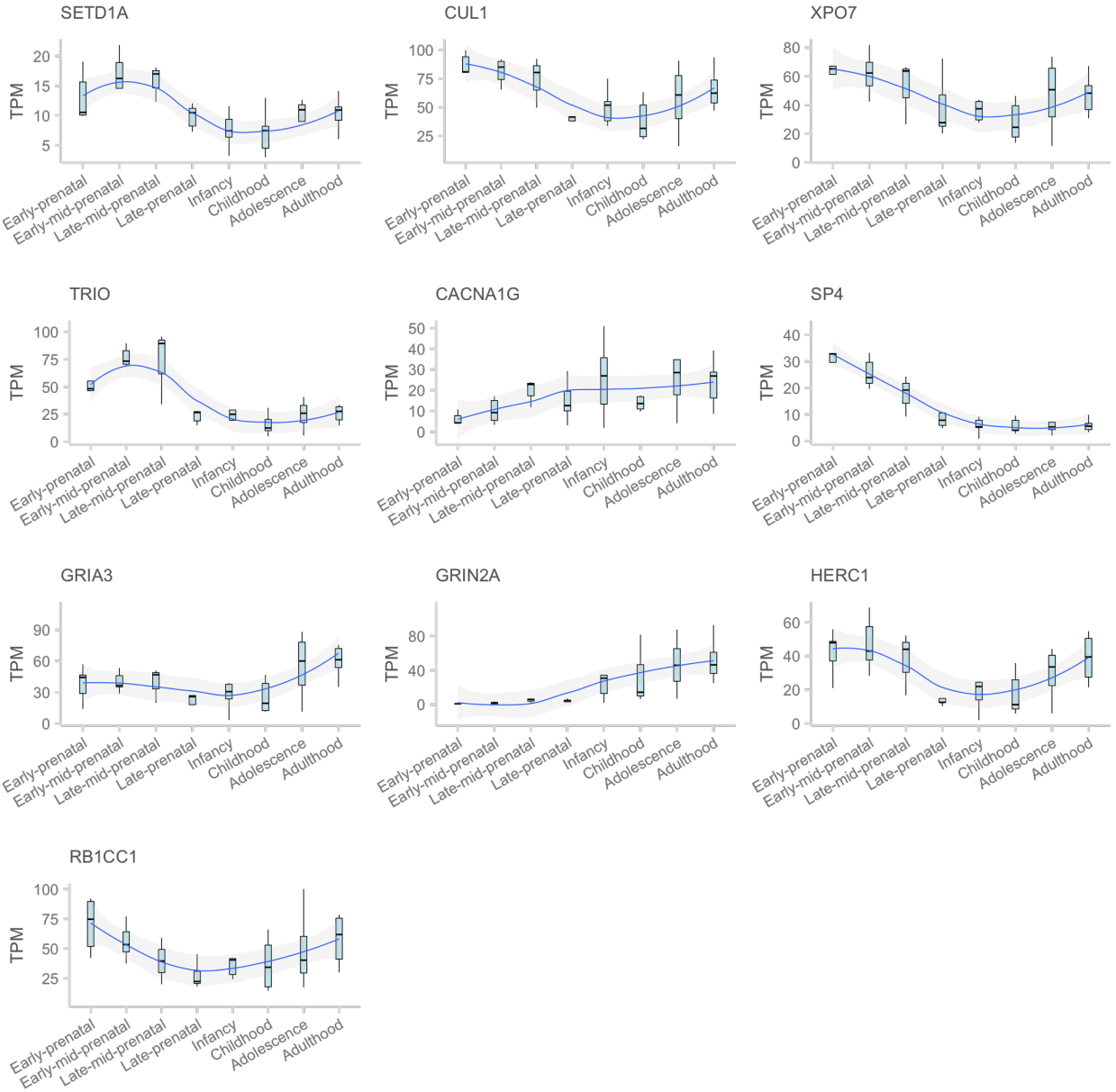
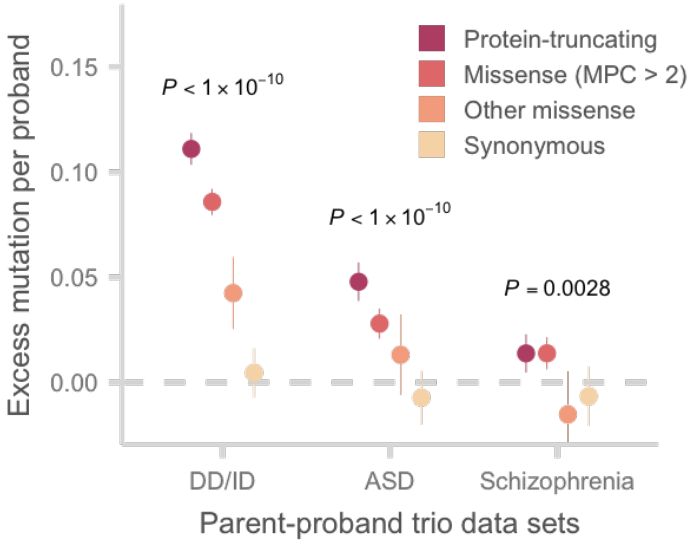


Figure S25 *De novo* mutations in constrained genes in each neurodevelopmental disorder. The rate of *de novo* mutations in DD/ID, ASD, and SCZ are compared to control probands to estimate the number of extra *de novo* mutations in affected individuals compared to controls. The analysis was restricted to constrained genes with pLI > 0.9. A Poisson rate test was used to calculate the rate difference and *P*-value. We display the *P* value for PTVs. Bars represent the 95% CIs of the point estimates.



Descriptions of Supplementary Tables

Table S1 to S3 are embedded in the Supplementary Materials.

Table S4, S6, S7, S8, S9, S11 Enrichment results for different gene sets. S4: constrained genes (pLI > 0.9); S6: Public pathway databases; S7: SynGO; Table S8: all GTEx tissues; Table S9: GTEx brain tissues; Table S11: Disease-associated genes from GWAS, DD/ID, and ASD.
Name: gene set of interest. Consequence: variant annotation used in analysis; N genes: size of gene set; Beta, standard error: inverse-weighted meta-analyzed beta and standard error from logistic regressions; odds ratio: calculated from the meta-analyzed beta and standard error; *Z*: meta-analyzed *Z*-score; *P*-value: gene set *P*-value

Table S5 Gene burden results from the main analysis, including complete case-control and *de novo* counts of protein-coding genes. PTV: protein-truncating variant, mis3: missense variants with MPC > 3, mis2: missense variants with MPC 2 - 3; P ca/co (Class 1): case-control *P* value of Class I variants; P ca/co (Class 2): case-control *P* value of Class II variants; P ca/co (comb): Meta-analyzed Class I and II case-control *P* values; *P* de novo: minimum *de novo P* value from the Poisson rate test of Class I or Class I and II variants; *P* meta: studywide *P* value; Q value: adjusted *P* value after FDR adjustment; OR: odds ratio; Class I: PTV and missense variants (MPC > 3); Class II: missense variants (MPC 2 - 3).

Table S10 Cell type enrichment of schizophrenia FDR < 5% in mouse single cell atlas. Cell cluster definitions acquired from Zeisel et al[54]. Class: major cell types in nervous system, mainly neurons and non-neurons; TaxonomyRank2: describes neuron and non-neuron cell types and location in CNS and PNS; ClusterName: Encoding of the specific cell type (unique identifier); CellSubclass: more resolved subdivision of cell types that include excitatory and inhibitory labels; Description: full description of cell type; p_value: Enrichment from Wilcoxon rank sum test of enrichment *Z*-scores between schizophrenia FDR < 5% genes and the remaining genes in the genome.

Table S11 List of associated genes from DD/ID and ASD exome sequencing studies, and schizophrenia GWAS[24,45,47]. Category: source of gene sets; Name: gene set label; Gene ID: Ensembl gene ID; Gene name: HGNC ID; pLI: probability of loss-of-function intolerant from gnomAD; o/e: observed-to-expected ratio from gnomAD[21].

Table S13 Shared and distinct genetic signals between schizophrenia, DD/ID, and ASD. We showed the burden results from ASD and DD/ID for the schizophrenia FDR < 5% genes as presented by the corresponding sequencing studies[24,47]. Yellow: indicates signal in DD/ID or ASD. Orange: indicates the possibility of an allelic series. gene_id, gene_name, gene_description, and p_studywide (SCZ): columns as defined in Table S5. Significant (DD/ID), p_meta (DD/ID): study-wide significance and *P*-value as presented in the DD/ID study. p_cluster: test of significant allelic clustering as presented in the DD/ID study; dn_{ptv, mis3, mis2}, case_ptv, ctrl_ptv: *de novo* and case-control counts as described in the DD/ID and ASD studies; p_dn_ptv (DD/ID), p_dn_mis (DD/ID): *P*-values re-calculated from Poisson rate test as described in the Supplementary Text.

Collections

Ashkenazi Jewish schizophrenia study

David C. Glahn, Ann E. Pulver

Danish iPSYCH initiative

Anders D. Børglum, Jonas Bybjerg-Grauholm, Jakob Grove, David M. Hougaard, Francesco Lescai, Preben Bo Mortensen, Merete Nordentoft, Thomas M. Werge

McLean psychosis study

Bruce M. Cohen, Mei-Hua Hall, Steven A. McCarroll, Dost Ongur

Pritzker Neuropsychiatric Disorders Research Consortium (Pritzker NDRC)

Huda Akil, Jack D. Barchas, William E. Bunney, William F. Byerley, Lynn DeLisi, Francis S. Lee, Richard M. Myers, Brandi Rollins, Alan Schatzberg, Marquis P. Vawter, Stanley J. Watson

Swedish schizophrenia study

Christina M. Hultman, Jordan W. Smoller, Patrick F. Sullivan

Taiwanese trios study

Wei J. Chen, Stephen V. Faraone, Stephen J. Glatt, Hai-Gwo Hwu, Ming T. Tsuang

UK and Ireland schizophrenia collections

Mariam Al Eissa, Nicholas Bass, Douglas H Blackwood, Gerome Breen, Aiden P. Corvin, Nicholas Craddock, Charles Curtis, David Curtis, Alessia Fiorentino, George Kirov, Andrew McQuillin, Derek W. Morris, Niamh L. O'Brien, Michael C. O'Donovan, Willem H. Ouwehand, Michael J. Owen, Aarno Palotie, Digby Quested, Sally I. Sharp, David St. Clair, James T. Walters

Whole Genome Sequencing in Psychiatric Disorders (WGSPD) consortium and the Genomic Psychiatry Cohort (GPC)

Tim B. Bigdeli, Michael Boehnke, Evelyn J. Bromet, Peter F. Buckley, Michael A. Escamilla, Saana Eskelinen, Tõnu Esko, Ayman H. Fanous, Nelson B. Freimer, Sarah A. Gagliano Taliun, Andrea Ganna, Eija Hämäläinen, Matti Holi, René S. Kahn, Hyun Min Kang, James A. Knowles, Douglas S. Lehrer, Dolores Malaspina, Stephen R. Marder, Helena Medeiros, Lili Milani, Christopher P. Morley, Roel A. Ophoff, Carlos N. Pato, Michele T. Pato, Tiina Paunio, Mark H. Rapaport, Laura J. Scott, Janet l. Sobell, Jaana Suvisaari

Acknowledgements

We would like to thank the patients and families who participated in our studies in the past two decades, without whom our research and findings would not be possible. Research reported in this publication was supported by the National Institute of Mental Health, and the National Human Genome Research Institute of the National Institutes of Health under award numbers: U01MH10564, U01MH105578, R01MH085548 and U54HG003067.

We would also like to acknowledge the generous support from the Stanley Family Foundation, Kent and Elizabeth Dauten, and The Dalio Foundation who have enabled us to rapidly expand our data generation collections with the goal of moving towards better treatments for schizophrenia and other psychiatric disorders.

Competing Interests

M.J.D. is a founder of Maze Therapeutics. B.M.N. is a member of the scientific advisory board at Deep Genomics and consultant for Camp4 Therapeutics, Takeda Pharmaceutical, and Biogen. A.P. is a member of Astra Zeneca's Genomics Advisory Board. M.C.O, M.J.O, and J.T.W. are supported by a collaborative research grant from Takeda Pharmaceuticals.

References

1. Fromer, M. *et al.* De novo mutations in schizophrenia implicate synaptic networks. *Nature* **506**, 179–184 (2014).
2. Singh, T. *et al.* Rare loss-of-function variants in SETD1A are associated with schizophrenia and developmental disorders. *Nat. Neurosci.* **19**, 571–577 (2016).
3. Genovese, G. *et al.* Increased burden of ultra-rare protein-altering variants among 4,877 individuals with schizophrenia. *Nat. Neurosci.* **19**, 1433–1441 (2016).
4. Howrigan, D. P. *et al.* Exome sequencing in schizophrenia-affected parent-offspring trios reveals risk conferred by protein-coding de novo mutations. *Nat. Neurosci.* **23**, 185–193 (2020).
5. Schizophrenia Working Group of the Psychiatric Genomics Consortium. Biological insights from 108 schizophrenia-associated genetic loci. *Nature* **511**, 421–427 (2014).
6. Goes, F. S. *et al.* Genome-wide association study of schizophrenia in Ashkenazi Jews. *Am. J. Med. Genet. B Neuropsychiatr. Genet.* **168**, 649–659 (2015).
7. Hall, M.-H. *et al.* Genomewide association analyses of electrophysiological endophenotypes for schizophrenia and psychotic bipolar disorders: a preliminary report. *Am. J. Med. Genet. B Neuropsychiatr. Genet.* **168B**, 151–161 (2015).
8. Pardiñas, A. F. *et al.* Common schizophrenia alleles are enriched in mutation-intolerant genes and in regions under strong background selection. *Nat. Genet.* **50**, 381–389 (2018).
9. Bigdeli, T. B. *et al.* Contributions of common genetic variants to risk of schizophrenia among individuals of African and Latino ancestry. *Mol. Psychiatry* (2019) doi:10.1038/s41380-019-0517-y.
10. Girard, S. L. *et al.* Increased exonic de novo mutation rate in individuals with schizophrenia. *Nat. Genet.* **43**, 860–863 (2011).
11. Xu, B. *et al.* De novo gene mutations highlight patterns of genetic and neural complexity in

schizophrenia. *Nat. Genet.* **44**, 1365–1369 (2012).

12. Gulsuner, S. *et al.* Spatial and temporal mapping of de novo mutations in schizophrenia to a fetal prefrontal cortical network. *Cell* **154**, 518–529 (2013).
13. Guipponi, M. *et al.* Exome Sequencing in 53 Sporadic Cases of Schizophrenia Identifies 18 Putative Candidate Genes. *PLoS One* **9**, e112745 (2014).
14. McCarthy, S. E. *et al.* De novo mutations in schizophrenia implicate chromatin remodeling and support a genetic overlap with autism and intellectual disability. *Mol. Psychiatry* **19**, 652–658 (2014).
15. Takata, A. *et al.* Loss-of-function variants in schizophrenia risk and SETD1A as a candidate susceptibility gene. *Neuron* **82**, 773–780 (2014).
16. Ambalavanan, A. *et al.* De novo variants in sporadic cases of childhood onset schizophrenia. *Eur. J. Hum. Genet.* **24**, 1–5 (2015).
17. Rees, E. *et al.* De novo mutations identified by exome sequencing implicate rare missense variants in SLC6A1 in schizophrenia. *Nat. Neurosci.* **23**, 179–184 (2020).
18. Picard. Picard. http://sourceforge.net/projects/picard/.
19. Li, H. & Durbin, R. Fast and accurate short read alignment with Burrows-Wheeler transform. *Bioinformatics* **25**, 1754–1760 (2009).
20. Lek, M. *et al.* Analysis of protein-coding genetic variation in 60,706 humans. *Nature* **536**, 285–291 (2016).
21. Karczewski, K. J. *et al.* The mutational constraint spectrum quantified from variation in 141,456 humans. *Nature* **581**, 434–443 (2020).
22. Mckenna, A. *et al.* The Genome Analysis Toolkit: a MapReduce framework for analyzing next-generation DNA sequencing data. *Genome Res.* **20**, 1297–1303 (2010).
23. Van der Auwera, G. A. *et al.* From FastQ data to high confidence variant calls: the Genome Analysis Toolkit best practices pipeline. *Curr. Protoc. Bioinformatics* **11**, 11.10.1–11.10.33 (2013).

24. Satterstrom, F. K. *et al.* Large-Scale Exome Sequencing Study Implicates Both Developmental and Functional Changes in the Neurobiology of Autism. *Cell* (2020) doi:10.1016/j.cell.2019.12.036.
25. Satterstrom, F. K. *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).
26. Jun, G. *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).
27. Rivas, M. A. *et al.* Insights into the genetic epidemiology of Crohn's and rare diseases in the Ashkenazi Jewish population. *PLoS Genet.* **14**, e1007329 (2018).
28. Conomos, M. P., Reiner, A. P., Weir, B. S. & Thornton, T. A. Model-free Estimation of Recent Genetic Relatedness. *Am. J. Hum. Genet.* **98**, 127–148 (2016).
29. Thornton, T. *et al.* Estimating kinship in admixed populations. *Am. J. Hum. Genet.* **91**, 122–138 (2012).
30. Singh, T. *et al.* The contribution of rare variants to risk of schizophrenia in individuals with and without intellectual disability. *Nat. Genet.* 1–10 (2017).
31. Li, H. Toward better understanding of artifacts in variant calling from high-coverage samples. *Bioinformatics* **30**, 2843–2851 (2014).
32. McLaren, W. *et al.* Deriving the consequences of genomic variants with the Ensembl API and SNP Effect Predictor. *Bioinformatics* **26**, 2069–2070 (2010).
33. Derrien, T. *et al.* The GENCODE v7 catalog of human long noncoding RNAs: analysis of their gene structure, evolution, and expression. *Genome Res.* **22**, 1775–1789 (2012).
34. Liu, X., Jian, X. & Boerwinkle, E. dbNSFP v2.0: A database of human non-synonymous SNVs and their functional predictions and annotations. *Hum. Mutat.* **34**, 2393–2402 (2013).
35. Samocha, K. E. *et al.* Regional missense constraint improves variant deleteriousness prediction. *bioRxiv* 148353 (2017) doi:10.1101/148353.

36. Feng, Y.-C. A. *et al.* Ultra-Rare Genetic Variation in the Epilepsies: A Whole-Exome Sequencing Study of 17,606 Individuals. *Am. J. Hum. Genet.* **105**, 267–282 (2019).
37. Darnell, J. C. *et al.* FMRP stalls ribosomal translocation on mRNAs linked to synaptic function and autism. *Cell* **146**, 247–261 (2011).
38. Sugathan, A. *et al. CHD8* regulates neurodevelopmental pathways associated with autism spectrum disorder in neural progenitors. *Proceedings of the National Academy of Sciences* **111**, E4468–E4477 (2014).
39. Weyn-Vanhentenryck, S. M. *et al.* HITS-CLIP and Integrative Modeling Define the Rbfox Splicing-Regulatory Network Linked to Brain Development and Autism. *Cell Rep.* **6**, 1139–1152 (2014).
40. Purcell, S. M. *et al.* A polygenic burden of rare disruptive mutations in schizophrenia. *Nature* **506**, 185–190 (2014).
41. Croning, M. D. R., Marshall, M. C., McLaren, P., Armstrong, J. D. & Grant, S. G. N. G2Cdb: the Genes to Cognition database. *Nucleic Acids Res.* **37**, D846–51 (2009).
42. Koopmans, F. *et al.* SynGO: An Evidence-Based, Expert-Curated Knowledge Base for the Synapse. *Neuron* **103**, 217–234.e4 (2019).
43. Finucane, H. K. *et al.* Heritability enrichment of specifically expressed genes identifies disease-relevant tissues and cell types. *Nat. Genet.* **50**, 621–629 (2018).
44. The GTEx Consortium. The Genotype-Tissue Expression (GTEx) pilot analysis: Multitissue gene regulation in humans. *Science* **348**, 648–660 (2015).
45. Schizophrenia Working Group of the Psychiatric Genomics Consortium. Mapping genomic loci prioritises genes and implicates synaptic biology in schizophrenia. *Submitted* (2020).
46. Benner, C. *et al.* FINEMAP: efficient variable selection using summary data from genome-wide association studies. *Bioinformatics* **32**, 1493–1501 (2016).
47. Kaplanis, J. *et al.* Integrating healthcare and research genetic data empowers the discovery of 49 novel developmental disorders. *bioRxiv* 797787 (2019) doi:10.1101/797787.

48. McRae, J. F. *et al. Prevalence, phenotype and architecture of developmental disorders caused by de novo mutation*. 1–39 http://biorxiv.org/lookup/doi/10.1101/049056 (2016) doi:10.1101/049056.
49. Kyle Satterstrom, F. *et al.* ASD and ADHD have a similar burden of rare protein-truncating variants. *bioRxiv* 277707 (2018) doi:10.1101/277707.
50. Kosmicki, J. *et al.* Refining the role of de novo protein truncating variants in neurodevelopmental disorders using population reference samples. *bioRxiv* 052886 (2016).
51. Westfall, P. H. & Stanley Young, S. *Resampling-Based Multiple Testing: Examples and Methods for p-Value Adjustment*. (John Wiley & Sons, 1993).
52. Willer, C. J., Li, Y. & Abecasis, G. R. METAL: fast and efficient meta-analysis of genomewide association scans. *Bioinformatics* **26**, 2190–2191 (2010).
53. Samocha, K. E. *et al.* A framework for the interpretation of de novo mutation in human disease. *Nat. Genet.* **46**, 944–950 (2014).
54. Zeisel, A. *et al.* Molecular Architecture of the Mouse Nervous System. *Cell* **174**, 999–1014.e22 (2018).
55. Kang, H. J. *et al.* Spatio-temporal transcriptome of the human brain. *Nature* **478**, 483–489 (2011).