Genetics-to-structure multiscale analysis identifies disrupted calcium homeostasis as a mechanism of psychiatric disease

Sherif Gerges[1,2,3], Nikolaj Catois Straarup[8], Mohamed A. El-Brolosy[9,13], F. Kyle Satterstrom[1,2,3], Nolan Kamitaki[1,2,7], Jiayi Yuan[2], Emi Ling[1,2], Raozhou Lin[1], Melissa Goldman[1,2], Tarjinder Singh[4,5,6], Jonathan S. Weissman[9,10,11,12], Sabina Berretta[1,14,], Jen Q. Pan[1], Hilary Finucane[1,2,3], Charlott Stock[8], Poul Nissen[8], Steve McCarroll[1,2,11], Mark Daly[1,2,3]

[1] Stanley Center for Psychiatric Research, Broad Institute of MIT and Harvard, Cambridge, MA 02142, USA
[2] Department of Genetics, Harvard Medical School, Boston, MA 02115, USA
[3] Analytic and Translational Genetics Unit, Massachusetts General Hospital, Boston, 02114, Massachusetts, USA
[4] The Department of Psychiatry at Columbia University Irving Medical Center
[5] The New York Genome Center
[6] Mortimer B. Zuckerman Mind Brain Behavior Institute at Columbia University
[7] Department of Biomedical Informatics, Harvard Medical School, Boston, MA, USA
[8] Danish Research Institute of Translational Neuroscience—DANDRITE, Nordic EMBL Partnership for Molecular Medicine, Department of Molecular Biology and Genetics, Aarhus University, Aarhus, Denmark
[9] Whitehead Institute for Biomedical Research and Department of Biology, Massachusetts Institute of Technology, Cambridge, MA 02142, USA
[10] Department of Biology, Massachusetts Institute of Technology, Cambridge, MA, USA
[11] Howard Hughes Medical Institute, Massachusetts Institute of Technology, Cambridge, MA, USA
[12] David H. Koch Institute for Integrative Cancer Research, Massachusetts Institute of Technology, Cambridge, MA, USA
[13] Society of Fellows, Harvard University, Cambridge, MA, USA
[14] Department of Psychiatry, Harvard Medical School, Boston, MA, USA

Abstract

Polygenic association studies implicate numerous genes in neuropsychiatric disorders, but linkage disequilibrium (LD) and cellular heterogeneity hinder mechanistic interpretation. Here, we integrate single-nucleus RNA-seq from human neurons with network inference and polygenic signal weighting to resolve pathway-level drivers. Neuron-resolved gene co-expression networks constructed across brain regions are reweighted by GWAS-derived polygenic signal (LD-aware heritability enrichment), prioritizing modules that disproportionately contribute to liability. Using this framework, we highlight the dysregulation of Ca^{2+} homeostasis as an etiological driver of neuropsychiatric disorders, and even relative to other neuronal gene sets, Ca^{2+} homeostasis exhibits the greatest concentration of rare variant signal. Furthermore, we find that a critical component of this molecular system, the P-type calcium ATPase *ATP2B2*, exhibits marked expression deficits in both nuclear transcriptomic and synaptic proteomic datasets derived from the dorsolateral prefrontal cortices of individuals with schizophrenia.

To connect sequence variation to structure and mechanism, we developed a residue-centric three-dimensional neighborhood analysis that integrates case–control missense variation with AlphaFold3 structural models to localize mutational hotspots of biological significance for downstream mechanistic interrogation. This approach identified an enrichment of deleterious missense variants - implicated across multiple neuropsychiatric disorders - that changed protein residues in close spatial proximity to both the Ca^{2+} permeation tunnel and the ATP:Mg^{2+} coordination site of ATP2B2. Cellular and biochemical analyses of the canonical Ca^{2+} binding site revealed clear loss-of-function effects, corroborating the earlier functional genomics evidence, and establishing a distinct molecular mechanism that converges on impaired Ca^{2+} extrusion, likely perturbing pre- and post-synaptic Ca^{2+} homeostatic equilibrium in excitatory neurons. Altogether, our study makes a significant contribution by linking genetic risk to neuronal dysfunction through a critical calcium signaling axis, offering mechanistic insight into the pathogenesis of neuropsychiatric disorders. In parallel, we develop a residue-centered 3D neighborhood framework that couples case–control genetics with structural models to discover pathogenic hotspots, generalizable across the proteome to any protein structure.

Main

Neuropsychiatric disorders are among the most debilitating illnesses affecting humankind, imposing a substantial global burden of morbidity that rivals those of cancer and cardiovascular disease, decades of neuropsychiatric drug discovery have been characterized by a failure to develop effective psychopharmacological treatments with novel mechanisms of action[1–4]. Human genetic studies, including genome-wide association and exome sequencing, have identified hundreds of loci associated with these disorders[7,8]. However, most common variant signals fall in noncoding regions, complicated by linkage disequilibrium and the cellular diversity of the brain, making it difficult to resolve their contribution to molecular pathways. Coding variants, in principle more directly interpretable, present a separate challenge: protein-truncating alleles typically abolish function, but most missense variants lie along a continuum from benign to damaging, and the mechanisms by which they alter protein function are often unclear.

Integrative genetic analyses of neuropsychiatric traits suggest that genetic signals are enriched in neuronally expressed genes[5–8]. Beyond cell identity, snRNA-seq also captures dynamic gene-expression programs within cell types. We hypothesized that the gene-expression modules shaping neuronal identities contain latent information linking polygenic risk to biology. To test this, we developed Pairwise Analysis of Neurons (PANs), an interpretable framework that integrates neuronal transcriptomic covariation with per-SNP heritability to identify gene groups associated with neuropsychiatric traits (Figure 1). We focus on neuronal transcriptomes, whose diversity reveals gene-expression modules underlying cell identity and provides a more informative basis for gene and functional prioritization[5–7,9–11]. PANs weights patterns of neuronal co-expression by per-SNP heritability, enabling the prioritization of pathways and gene modules most likely to mediate disease risk. We then validate this framework by incorporating exome sequencing from large case-control cohorts, establishing convergence between common and rare variation on specific biological processes. Finally, we addressed the challenge of interpreting excess missense variation in a neurobiologically relevant gene by developing a structure-guided approach that integrates AlphaFold3-derived protein structures[12] with genetic data to detect localized clustering of deleterious variants, thereby pinpointing structurally constrained regions most likely to underlie disease mechanisms.

.

Results

Development of the pairwise analysis of neurons (PANs) framework

We integrated gene expression from across multiple regions of the human brain from single-nucleus RNA sequencing (snRNA-seq) data of neurons with genome-wide summary statistics to identify co-expressed genes associated with neuropsychiatric disorders (**Figure 1**). Specifically, PANs computes pairwise sets of the top 1,000 expressed genes (Methods) between each pair of neuronal cell types and uses P-values derived from the per-SNP heritability of the surrounding 100kb genomic region of these genes using S-LDSC, while conditioning on a comprehensive set of coding, conserved, and regulatory annotations from the baseline-LD model[13]. PANs then assigns a per-gene score reflecting the frequency with which a gene appears in pairwise comparisons that include genes with significant per-SNP heritability excess. Further details are provided in Methods and Extended Discussion Note.

PANs identifies genes with strong case–control enrichment in exome sequencing

Rare protein truncating variation provides the most interpretable insight into the mutations that perturb genes that give rise to disorders. Since this variation likely modifies protein sequences to elicit large phenotypic effects, it is thus considered the gold-standard for implicating genes and ultimately, the pathways in which they elicit their effect[14,15]. We validated our approach via two distinct and orthogonal publicly available datasets with no relationship to GWAS or transcriptomic data – including the largest available exome-sequencing analysis of schizophrenia, epilepsy, autism, bipolar disorder and developmental disorders[16–19]. First, we tested PANs ability to prioritize genes by comparing the scores of genes with at least nominal significance in exome sequencing studies versus all other genes in the genome, and observed global enrichment of significant genes from sequencing data, relative to non-significant ones (**Figure 2a**, **Supplementary Figure 2**). We then used the individual counts of ultra-rare variants to perform a burden test on the top 500 constrained PANs genes, and find substantial enrichment of damaging variation, above MAGMA and constrained gene subsets of neuronally expressed genes (**Figure 2b**, **Supplementary Table 1**) for all phenotypes except bipolar disorder (likely because it's an underpowered study). In addition to global burden, we were also interested in testing whether we can prioritize a single candidate causal gene within 100kb of an index SNP within complex GWAS-associated regions containing several genes (Methods), using evidence in the exome sequencing or fine-mapped data from the GWAS. We find several examples where we successfully prioritize genes in these gene-dense regions, examples including Bonferroni significant genes from the latest schizophrenia study (*SP4)* and FDR < 5% significant ones (*FAM120A*

and *STAG1*), while *FOXP1* for developmental delay[18], and *SCN1A* for epilepsy[20] (**Supplementary Figure 5**). Taken together, these analyses demonstrate that PANs prioritizes genes from common variation and excess rare variant burden.

Drug targets as orthogonal validation

The existing pharmacopeia of psychiatric drugs share mechanisms that target neurotransmission in the brain[21–23]. Owing to the serendipitous nature of drug development in psychiatry[1,3,24–26], the drug armamentarium have been discovered independently of genetics, and thus can orthogonally validate our approach. We curated drug target gene sets from Open Targets[27], sub-setting to all targets with a drug score > 0 (consisting of either approved or in trial clinical drugs). First, we asked whether these drug targets show excess rare variant burden in exome sequencing and found substantial enrichment for ultra-rare variation, consistent with bona fide biological connections to schizophrenia, bipolar disorder, and epilepsy (**Supplementary Figure 3**).

Second, we then tested whether the top 500 PANs prioritized genes are also implicating these therapeutic targets genes across multiple neuropsychiatric disorders, compared with an equal number of MAGMA genes (ranked by Z-score), and observe that PANs more successfully prioritized of relevant genes compared to the top 500 MAGMA genes (**Figure 2c**). This is likely because PANs is identifying genes that coalesce into biological processes because of their co-expression with GWAS genes, and given that many existing drugs act through non-specific small molecules, making pathway-based prioritization essential for identifying more precise therapeutic targets.

For example, PANs prioritizes a subunit of ionotropic $GABA_A$ receptor, *GABRB2* for depression (drug target of Zuranolone and ranked 8 in PANs and 2,719 in MAGMA). While for bipolar disorder, PANs implicates *SCN2A* and *SCN8A* (targets of lamotrigine, ranked 50 and 57 in PANs and 3,976 and 5,791 in MAGMA, respectively). As negative controls, we also selected drug target gene sets for well-powered phenotypes with low genetic correlation with psychiatric disorders, and observe no enrichment, as expected (**Supplementary Figure 3b**).

Calcium homeostasis in neuropsychiatric and developmental disorders

Neurons co-express hundreds of sets of gene-organized modules which mediate neurotransmitter release across synapses [47,28,29]. Consequently, a primary goal in genetics is to identify genetic targets

linked to disorders and to understand the biological processes through which genetic variations exert their effects. To identify biological processes nominated by PANs, we analyzed public pathway databases to conduct gene ontology analyses[30,31]. We tested the top 500 genes implicated by PANs for 8 neuropsychiatric traits, and observed substantial enrichments which recapitulated robust and related biological processes (**Supplementary Table 5**), consistent with earlier observations that genes with roles in synaptic machinery are likely perturbed in psychiatric disorders[32–34]. For schizophrenia, we identified genes with significant overrepresentation in modules involved with synaptic processes such as regulation of trans-synaptic signaling (GO:0099537, P_{adj} = 9.84 x 10^{-20}), modulation of chemical synaptic transmission (GO:0050804, P_{adj} = 3.18 x 10^{-26}) and regulation of synapse organization (GO:00508807, P_{adj} = 6.08 x 10^{-12}). Amongst neurotransmitter related terms, glutamate receptor signaling pathway (GO:0007215, P_{adj} = 2.4x10^{-8}), and glutamatergic regulation of synaptic transmission (GO:0051966, P_{adj} = 5.3x10^{-7}). Many of the processes involved in schizophrenia were also implicated in other neuropsychiatric traits, consistent with the high degree of genetic correlation between traits.[35–39]

Notably, the highest-ranked disrupted processes were primarily associated with molecular functions and processes related to the maintenance of Ca^{2+}. For schizophrenia, this included the regulation of postsynaptic cytosolic calcium ion concentration (GO:0099566, fold-enrichment = 15.02x, P_{adj} = 1.8x10^{-2}), the regulation of intracellular cytosolic Ca^{2+} (GO:0098703, fold-enrichment = 8.9x, P_{adj} = 2.7 x 10^{-4}) and calcium ion transmembrane import into cytosol (GO:0097553, fold-enrichment = 4.89x, P_{adj} =3.4x10^{-4}) (**Figure 3a**). Notably, we observed this pattern in other brain-related traits as well, including as epilepsy, bipolar disorder, and intelligence (**Supplementary Figure 4**, **Supplementary Table 4**).

Given that some members of voltage-gated ion channels appears in genetic studies of neuropsychiatric disorders[17,40,], we systematically assessed whether these genes, as a class, exhibit an excess burden of ultra-rare variants in schizophrenia, bipolar disorder, autism spectrum disorder, and epilepsy. We stratified burden analyses by voltage-gated ion channel conductance, and found that voltage-gated Ca^{2+} channels harbored the strongest enrichment relative to other channel types, with the most pronounced signal observed in schizophrenia. Within this subclass, 6 of 26 genes surpassed nominal significance ($p < 0.05$) for rare variant burden in schizophrenia. Similar patterns of enrichment were observed in autism spectrum disorder and epilepsy (**Supplementary Figure 6**), although these did not reach statistical significance in bipolar disorder or epilepsy.

We reasoned that the excess rare variant burden in voltage-gated ion channels across neuropsychiatric disorders may reflect ion conductance rather than gating, predicting that other calcium-conducting channels, irrespective of gating class, should also be enriched. To test this, we used HGNC-curated gene sets to identify all protein-coding genes involved in ion transport (see *Methods: Ion Channel Gene Set Selection*)[45]. These genes were then stratified based on their principal ion conductance into five categories: Ca^{2+}, Potassium (K^+), Sodium (Na^+), Chloride (Cl^-) , and non-selective/other (including volume-regulated, acid-sensing, porin, and gap junction channels). Intriguingly, we observed the strongest enrichment in Ca^{2+} channels relative to other ion channel classes (**Figure 3d**).

Notably, PANs also nominates other components of the Ca^{2+} channel 'toolkits' which are distributed at plasma or intracellular organelle membranes are implicated by PANs, and featured prominently in our ontological analysis, and had corresponding signal in exome sequencing studies. These include *RYR2* (ranked 21st in PANs for Schizophrenia exome P-value = 3.4×10^{-3} , Epilepsy P = 2.8×10^{-4}, NDD P-value = 2.86×10^{-6}) and *ITPR1* (ranked 101 for IQ, NDD P-value < 2.2×10^{-16}) (**Table 1**).

Co-expression Networks Reveal Rare Variant Enrichment in Calcium Homeostasis Genes Across Diverse Cellular Compartments

Through many pairwise comparisons, PANs leverages the transcriptomic heterogeneity of neurons. We reasoned that genes consistently co-occurring across neurons are likely to be functionally related and may reflect disease-relevant biological processes. To discover non-randomly co-occurring genes, we implemented a Fisher's exact test that compares the observed overlap of each pair of genes across PANs-enriched sets to the expectation under statistical independence, while controlling for gene-set frequency (see Methods: *Co-occurrence Analysis of Calcium Homeostasis Genes).*

We found that the well-established voltage-gated calcium channel genes co-occur with many Ca^{2+}-related genes, suggesting that PANs prioritizes additional components of the calcium signaling machinery beyond canonical depolarization-activated channels. Several of these genes were also enriched in our gene ontology analysis and showed convergent evidence from exome sequencing studies. For example, *RYR2* (ranked 21st by PANs; schizophrenia P = 3.4×10^{-3}, epilepsy P = 2.8×10^{-4}, NDD P = 2.9×10^{-6}) and *ITPR1* (ranked 40th; NDD P < 2.2×10^{-16}) are intracellular calcium-release channels localized to the endoplasmic reticulum. Beyond these, PANs also highlighted a diverse array of calcium-regulating genes - including ligand-gated channels, ion exchangers and ATP-dependent - many

of which play essential roles in shaping calcium transients and restoring homeostasis after neuronal activation.

Intrigued by the convergence of these diverse calcium-regulating elements, we curated a comprehensive set of genes involved in calcium transport and regulation - spanning both presynaptic and postsynaptic compartments, as well as intracellular organelles. These genes, collectively referred to as Ca^{2+} homeostasis genes[46,47–53], encompass voltage-gated channels, ligand-gated receptors, intracellular release channels, ion exchangers, and pumps. Together, they orchestrate the spatial and temporal control of calcium signaling in neurons, and their prominence in our PANs analysis suggests they may represent a biologically meaningful axis of convergence in neuropsychiatric risk (**Figure 3f**, full list of genes in **Supplementary Table 7)**.

Additionally, we found that this Ca^{2+} homeostasis gene set showed significant co-occurrence with several members of the SCHEMA list of FDR significant genes. **Supplementary Figures 7 and 9b**), as well as autism and NDD genes (**Supplementary Figure 9a**). This pattern suggests that Ca^{2+} homeostasis is a meaningful extension of the core Ca^{2+} signaling channels typically implicated in neuropsychiatric disease. Furthermore, they may function as critical interacting partners that consistently co-occur with risk genes associated with neuropsychiatric disorders, pointing to biologically coordinated relationships that may underlie shared mechanisms.

We then compared Ca^{2+} homeostasis genes to other nervous system gene sets with strong associations to schizophrenia, namely translational targets of FMRP[54], chromatin targets of CHD8[55] and splice targets of RBFOX[56] , and found a substantial enrichment of these genes over what are considered critical and biologically significant benchmarks in neurodevelopmental genetics (**Figure 3d**). Next, to assess the burden of rare de novo coding variants in other neurodevelopmental disorders, we analyzed data from a large developmental disorder exome sequencing study [27]. We observed significant enrichment of de novo missense variants, although less so for PTVs, in constrained Ca^{2+} homeostasis genes, relative to other gene sets involved in nervous system function (**Supplementary Figure 9a)**. The fact that this enrichment was pronounced for missense variants may suggest that Ca^{2+} homeostasis genes may exhibit heightened sensitivity to subtle, function-altering changes, likely due to their roles in fine-tuned intracellular signaling.

In addition to schizophrenia, several of these genes are associated with numerous other neuropsychiatric disorders. For example, *CACNA1G* is Bonferroni significant in schizophrenia exome sequencing ($P = 4.57 \times 10^{-7}$), while *CACNA1E* is significantly associated with NDD[18] and epilepsy with

intellectual disability (missense $P = 1.15 \times 10^{-6}$). Furthermore, a paralog of these genes, *ATP2A2*, was recently a fine-mapped in the PGC schizophrenia phase 3 GWAS, which identified an intronic variant as highly probable of being causal (PP > 0.99) and is the most prioritized gene in the region by PANs (prioritization score of 0.98, **Supplementary Figure 5**). Notably, in the Bipolar Exome (BipEx) sequencing project[16], *ATP2A2* carries a strong missense burden ($P = 5 \times 10^{-3}$ with an odds ratio of 10). *ATP2A2* encodes one of the sarco/endoplasmic reticulum Ca^{2+} ATPase (SERCA) protein family, which are intracellular pumps responsible for catalyzing the hydrolysis of ATP coupled with the translocation of calcium from the cytosol to the lumen[57].

Plasma membrane calcium-transporting ATPase 2

Among the top PANs genes driving our enrichment signals were members of plasma membrane Ca^{2+}-ATPases (PMCAs), *ATP2B1* and *ATP2B2*. Calcium ATPases are key components of calcium extrusion machinery and are pivotal for maintaining the intracellular calcium concentration essential to neuronal function[58]. Two genes within this family have compelling genetic evidence that links their cellular functions to psychiatric and neurodevelopmental disorders. Notably, *ATP2B1* (ranked 8th in PANs) has a SCHEMA PTV and MPC > 3 *P* value of 8.7×10^{-3} and has also been linked to neurodevelopmental delay ($P = 8 \times 10^{-3}$)[19,59]. A second gene *ATP2B2* (ranked 12th in PANs), which contains genetic signals shared amongst schizophrenia, ASD and neurodevelopmental delay. Notably, it's signal in schizophrenia consists of mostly missense, but not protein-truncating, burden (MPC 2-3 odds ratio = 2, $P = 7.0 \times 10^{-4}$, PTV odds ratio = 2.01, $P = 0.2$, ranked 5th in SCHEMA by missense burden)[60]. Additionally, *ATP2B2* is exome-wide significant in ASD and NDD ($q < 5 \times 10^{-6}$ for both disorders).

While human genetics can establish an etiological link between a gene and a disorder, functional data from patients diagnosed with a neuropsychiatric disorder that corroborates the direction of genetic effect can justify the rationale for developing targeted therapies. In particular, disruptions in brain regions strongly implicated in schizophrenia such as the dorsolateral prefrontal cortex (dlPFC)—are believed to play a central role in the cognitive and executive dysfunctions characteristic of the disorder, making them especially critical targets for therapeutic intervention. We examined whether levels of *ATP2B2* RNA and protein levels are reduced in two distinct case-control datasets from the dorsolateral prefrontal cortices of donors with schizophrenia; one transcriptomic[61] and one proteomic[62]. The transcriptomic data consists of single-nucleus RNA sequencing from 191 human donors (94 with schizophrenia and 97 controls)[61]. We observed that *ATP2B2* gene expression was

significantly reduced specifically in glutamatergic neurons in individuals with schizophrenia, with no such reduction seen in other cell types (**Figure 4a**, **Supplementary Figure 16**). We then examined a separate dataset consisting of synaptic proteome fractions from the dlPFC of 69 human donors (not overlapping with the transcriptomic cohort, 34 with schizophrenia and 35 controls)[62]. Again, we found a significant reduction in *ATP2B2* protein expression in synaptic proteome fractions from donors with schizophrenia compared to healthy controls (**Figure 4b**). These observations strongly suggest that deficits, akin to a loss-of-function phenotype in *ATP2B2,* are associated with a specific component of schizophrenia synaptopathobiology in the DLPFC. Considering its broader role in synaptic calcium dynamics, and the tight coupling of calcium and the regulation of gene expression, we hypothesized that the reduced expression of *ATP2B2* may reflect broader disturbances to glutamatergic neuron biology associated with the onset of schizophrenia.

Protein neighborhood analysis

ATP2B2 displays a striking excess of rare missense variants in both schizophrenia and autism, far exceeding the burden of protein-truncating variation (for example in schizophrenia: MPC 2–3 odds ratio = 2.0, $P = 7.0 \times 10^{-4}$; PTV P = 0.2). This suggests a potential mechanism involving regional or domain-specific disruption of pump activity, rather than full gene inactivation. However, without direct experimental assays, predicting the effects of these missense variants on *ATP2B2* function is only speculative. This interpretive challenge**,** known as variant-effect prediction (VEP), is a central bottleneck in human genetics[63–65]. Unlike protein-truncating variants that generally abolish function, missense alleles span a continuum from benign to subtly or profoundly disruptive[66,67]. Furthermore, because they are more abundant than truncating ones and can point to precise biology - developing scalable, structure-aware methods to localize and quantify their effects is essential for building biologically grounded models of disease pathogenesis[68,69].

Importantly, *ATP2B2* and related plasma membrane Ca^{2+}-ATPases (PMCAs) are relatively well-studied, and structural homology to the SERCA family provides a strong functional prior - supported by decades of biochemical data and high-resolution cryo-EM structures[70–75]. Given their essential role in actively exporting Ca^{2+} from the cytosol, PMCAs are central regulators of intracellular calcium dynamics, acting as key lynchpins in shaping ionic fluctuations within neurons. This mechanistic background made *ATP2B2* an ideal case for probing the utility of spatial clustering as a means of interpreting missense variation.

We mapped all case-only and control-only variants from SCHEMA[17] , together with autism-associated and broader neurodevelopmental disorder variants[18,19], onto the AlphaFold3-predicted tertiary structure of *ATP2B2*. Our strategy identifies spatial neighborhoods by constructing 15 Å spheres centered on each missense variant observed in the dataset, without applying any variant-level filtering. The 15 Å radius was selected based on the typical range of interatomic distances that encompass side-chain and backbone interactions, accounting for the van der Waals radii of all relevant heavy atoms[34]. This spatial window is meant to capture the local biochemical environment surrounding each residue, including nearby residues that may participate in functional interactions. For each defined neighborhood, we applied a Fisher's exact test to assess whether it was significantly enriched for case variants relative to the rest of the protein structure (**Figure 5a**; Supplementary Table 6, also see *Methods: 3D protein neighborhood test*).

We find several residues around which there is a Bonferroni significant enrichment of cases variants compared to controls (**Figure 5b**). All of them are around the Ca^{2+} permeation pathway and transmembrane tunnel (V885, p-value = 2.2×10^{-6}, **Figure 5c**). To determine whether any of the enriched spatial neighborhoods overlapped with functionally important residues, we aligned the ATP2B2 sequence to the experimentally resolved cryo-EM structure of human ATP2B1[70] **(Supplementary Figure 15)**. Owing to the high sequence and structural homology between ATP2B2 and ATP2B1, this allowed us to incorporate existing domain annotations from ATP2B1. We found that several of the most enriched neighborhoods contain residues that either constitute or lie within a few angstroms of the three principal Ca^{2+}-binding sites of ATP2B2 (**Figure 5c**). Notably, this includes E457 - which itself formed a neighborhood and demonstrated significant enrichment for case-associated variants ($p = 2.2 \times 10^{-4}$). Permutation of case-control labels indicates that the observed clustering is highly unlikely under the null ($p = 6 \times 10^{-3}$, derived from 1000 permutations, **Supplementary Figure 13**). Importantly, this pattern remained robust when restricting the analysis to schizophrenia-associated variants alone (**Supplementary Figure 11**).

Multiple lines of orthogonal evidence spanning experimental studies, human genetic data, and in silico structural modeling support the pathogenicity and functional impact of E457K in *ATP2B2*. In a mouse model, the E457K substitution was shown to impair Ca^{2+} handling by reducing the rate of intracellular Ca^{2+} increase during K^{+}-induced neuronal depolarization[76]. In the Regeneron Genetics Center Million Exome dataset, which consists of harmonized whole-exome sequencing data from 983,578 individuals, and found that E457 is the second most missense-constrained residue in the entire ATP2B2 protein, indicating extreme intolerance to variation at this site. (**Supplementary Figure 12**)[77]. To examine the structural consequence of this variant, we used Missense3D[78,79], which predicted that the E457K

substitution would expand the local cavity volume (**Figure 5c**). Finally, AlphaMissense assigns it a pathogenicity score of 0.998, among the highest possible in the genome[68].

Integrative Analysis of E457K Reveals Disruption of ATP2B2 Ca^{2+} Extrusion

The support and convergence of these results for a loss-of-function effect lies in E457K's physicochemical properties. Namely, in a negatively charged glutamate is replaced by a positively charged lysine, causing a charge reversal at the point of contact with Ca^{2+}. We hypothesized that this electrostatic disruption repels the positively charged Ca^{2+} and interferes with its coordination, thereby blocking Ca^{2+} transport through the channel, causing a functional loss of ATP2B2 activity. To test this, we used AlphaFold[8] to estimate per-residue Ca^{2+} binding probabilities in wild-type ATP2B2 and ATP2B2-E457K, in the presence of a single Ca^{2+} ion (see Extended Discussion: *Simulating Mutation Effects on Per-Residue Contact Probabilities with AlphaFold3*). The analysis revealed a marked reduction in Ca^{2+} contact probabilities with key residues along the channel pore (**Figure 6a**). This disruption at residues critical for coordination, supports a loss-of-function mechanism. Notably, we performed similar structural analyses for other variants in this region, but none produced a disruption as pronounced as E457K (data not shown).

To experimentally validate these in silico predictions, we evaluated E457K using complementary biochemical assays. Recombinant ATP2B2, expressed in both wild-type and mutant forms, was subjected to ATPase activity and calmodulin-binding assays (see Methods: In Vitro Measurement of ATPase Activity and Calmodulin Activation). These assays directly interrogate two core biochemical properties of the pump: the hydrolytic turnover of ATP that drives Ca^{2+} extrusion and the calmodulin-dependent regulatory interaction that modulates catalytic activity. E457K eliminated both activities in the recombinant protein, producing a functional profile equivalent to a catalytic null mutant (Figure 6c), consistent with a complete loss of function.

Because cells regulate intracellular Ca^{2+} through a complex interplay of influx, buffering, and sequestration, they provide a biologically relevant context for evaluating calcium-handling proteins. We employed a genetically engineered HEK293T model optimized for precise monitoring of submembrane Ca^{2+} dynamics. The system stably expresses a membrane-tethered GCaMP6s-CAAX sensor and Kir2.3 to maintain a hyperpolarized resting membrane potential, enabling controlled depolarization with extracellular KCl and reproducible Ca^{2+} influx[80]. In this system, the E457K variant resulted in an increase

in cytosolic Ca^{2+} , most probably due to significantly impaired Ca^{2+} extrusion, a result once again consistent with a loss-of-function effect **(Figure 6d,e).**

Conditional Analysis Uncovers a Second, Mechanistically Distinct Catalytic Disruption in ATP2B2

Several residues reside far from the Ca^{2+} selectivity pore do not meet Bonferroni significance (Figure 5b). We hypothesized that these may point to a second neighborhood of pathogenic variants far from the Ca^{2+}-binding site , but the signal is likely underpowered and does not reach statistical significance. We ran a conditional analysis whereby we excluded all case and control variants within 15 Å of the primary 3D neighborhood surrounding V885 and repeated the analysis. This revealed a single Bonferroni-significant neighborhood, surrounding residue R508 ($P = 2.01 \times 10^{-4}$; **Figure 7a**), located distally from the transmembrane domain and the Ca^{2+}-binding site, but proximally to the ATP:Mg^{2+} binding site - a region critical to ATP2B2's catalytic function (**Figure 7b**).

Critically, ATP2B2 is a member of the P-type ATPase family and uses ATP hydrolysis to drive a cyclical series of conformational changes, known as the E1–E2 cycle, that actively transport ions across the membrane against their concentration gradient[81] . The enrichment of this region of the protein suggests an alternative mechanism of ATP2B2 perturbation that affects its core function, specifically through interference with the Mg^{2+} dependent coordination of ATP binding and hydrolysis, which is necessary for providing the energy required for pump activity. Within this neighborhood, a pathogenic neurodevelopmental variant, G821R, is the closest in proximity to the ATP:Mg^{2+} cofactor complex (**Figure 7b,c**). We reasoned that the substitution of glycine, a small flexible nonpolar residue, with arginine, a bulky positively charged and conformationally constrained residue, likely introduces significant steric and electrostatic perturbations at the catalytic site. If true, this physicochemical change could disrupt local structural packing and alter the electrostatic environment required for ATP:Mg^{2+} coordination, potentially destabilizing the nucleotide binding pocket.

As with E457K, we simulated ATP2B2 in both the wild-type and G821R contexts in the presence of ATP and a single Mg^{2+} and compared the per-residue contact probabilities. The G821R model showed a pronounced reduction in contacts with the ATP:Mg^{2+} complex (**Figure 7d**), suggesting that this variant may destabilize nucleotide and cofactor interactions that are critical to E1-E2 transition in ATP2B2.

Discussion

A central objective of human genetics is to translate association signals into mechanisms and tractable therapeutic targets, yet for most neuropsychiatric loci, mechanistic attribution remains sparse. To address this gap, we developed PANs, an integrative framework that couples neuronal gene expression with per-SNP heritability across multiple brain regions to prioritize effector genes and pathways. PANs indicates that disease risk converges on Ca^{2+} signaling and ion-homeostatic machinery (**Figure 5d**)[46,82–84] consistent with a model in which dysregulated Ca^{2+} handling perturbs excitation–transcription coupling, neurotransmission, and synaptic plasticity[85] .

Our work introduces a mechanistic advance. PANs nominates the Ca^{2+} efflux pump *ATP2B2* (PMCA2) as a key effector, whose signal is driven by rare, damaging missense variants, which provide a direct handle on mechanism by pinpointing structural elements vulnerable to deleterious variation. We developed a hypothesis-free, structure-anchored 3D neighborhood test which shows that case missense variants cluster within two structurally critical regions, the Ca^{2+}-permeation transmembrane tunnel and the ATP:Mg^{2+} coordinating pocket, predicting defects in ion binding/egress. These observations cohere with transcriptomic (reduced *ATP2B2* mRNA in glutamatergic neurons) and synaptic proteomic evidence (lower protein abundance in cases), supporting a model in which ATP2B2 insufficiency compromises ATP-driven Ca^{2+} extrusion, destabilizes intracellular Ca^{2+} homeostasis, and likely perturbs Ca^{2+}-dependent programs so critical to neuronal biology. Notably, the fact that we see this in so few cases suggests that common regulatory variation is likely operating in the same direction as rare variants, and implies a mechanism across a broad segment of patients. Altogether, this work convergence of genetic, structural, functional, and cell type–specific deficits implicates *ATP2B2* as a key driver of psychiatric neuropathology.

This work also introduces a methodological advance. We introduce a framework that enables structural interpretation and prioritization of missense variation based on three-dimensional protein context. By highlighting structurally indispensable regions and pinpointing individual variants with likely functional consequences, our approach achieves two key goals (1) It narrows the vast search space of missense variants to a focused, high-priority subset for downstream experimental validation; and (2) It generates mechanistic hypotheses of variant pathogenicity and mode-of-action, enabling targeted follow-up in cellular, animal, or organoid models where biological and therapeutic insights are most likely to emerge. The framework is data-agnostic (cryo-EM or AlphaFold), readily augmented with pathogenicity priors

(e.g., AlphaMissense), and applicable across neurodevelopmental/neuropsychiatric gene classes (channels, transporters, synaptic scaffolds, transcriptional regulators, chromatin remodelers).

Our work has several limitations worth noting. First, the brain contains billions of neurons[91], and our survey of the human brain is not exhaustive. Concomitantly, it is likely that genetic studies conducted hitherto have uncovered a fraction of the heritability attributed to complex, polygenic neuropsychiatric traits. As such, uncovering the vastness of human transcriptomes and scaling human genetics will undoubtedly add information to our understanding of neuropsychiatry etiology. Second, we would like to caution against interpreting this analysis to imply that neurons are the only cell class causal for the disorders studied here. While many risk genes arise from genes highly expressed in neurons, disease-colocalizing eQTLs may point to other CNS cell types through which protein activity may be conferred. For example astrocytes, govern key steps in synapse formation and plasticity, and would ostensibly be affected by germline mutations discovered by GWAS. These findings, though derived from neuronal transcriptomes, do not exclude the participation of other cell types in neuropsychiatric disorder biology.

Figures

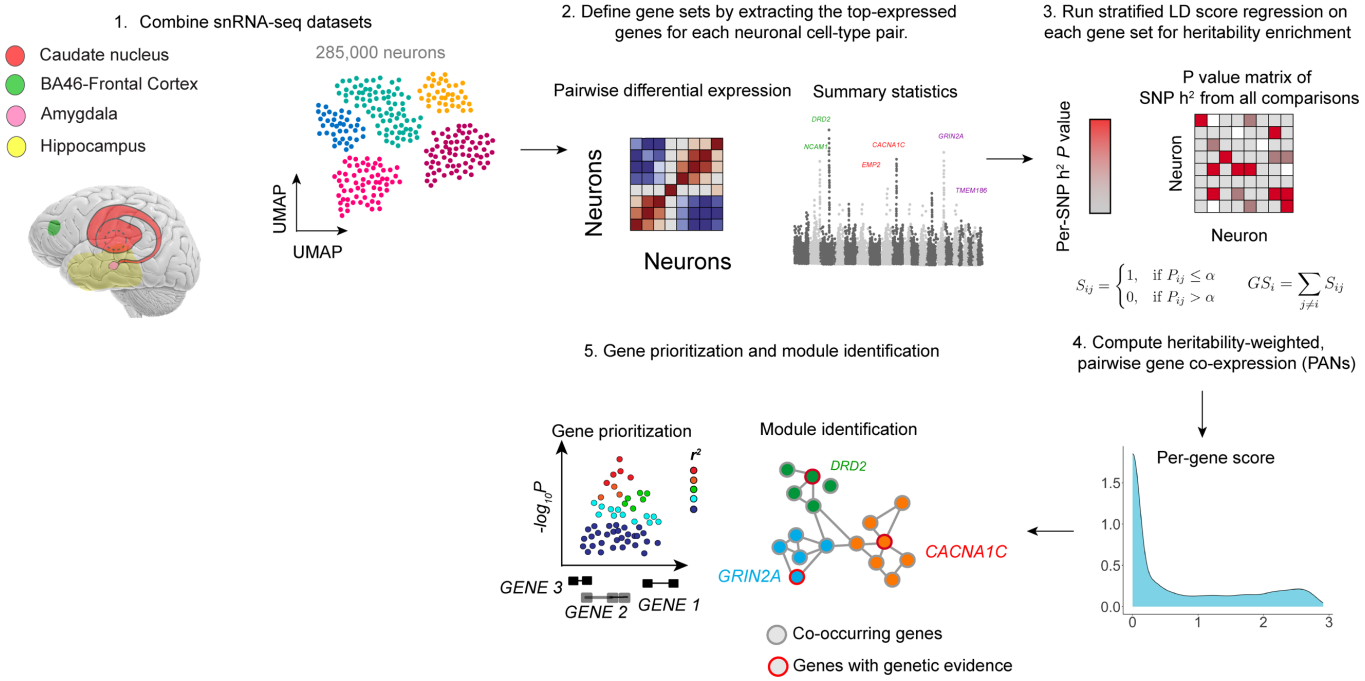


Figure 1. Overview of the PANs Analysis Workflow. 1) Data integration. Single-nucleus RNA-seq datasets from four human brain regions (caudate nucleus, dorsolateral prefrontal cortex [BA46], amygdala, and hippocampus) are merged, quality-controlled and visualized in UMAP space to yield a unified expression matrix of ~285,000 neurons. **2) Gene-set definition.** For every pair of neuronal populations, the top 1,000 highly expressed genes are extracted to form cell-pair–specific gene sets. 3), **Stratified LD score regression.** Each gene set is tested via S-LDSC, conditioning on the baseline-LD v2.2 model to estimate per-SNP heritability enrichment (h^2) within ±100 kb of set genes. The resulting p-value matrix captures enrichment across all pairwise cell comparisons. **(4) PANs scoring.** For each gene pair, we compute a heritability-weighted co-expression score by summing significant S-LDSC outcomes (FWER-controlled) across all neuronal comparisons. **(5) Gene prioritization and module detection.** Genes are ranked by their PANs scores (left), highlighting those with strong genetic evidence (red outlines). High-scoring genes coalesce into modules that pinp

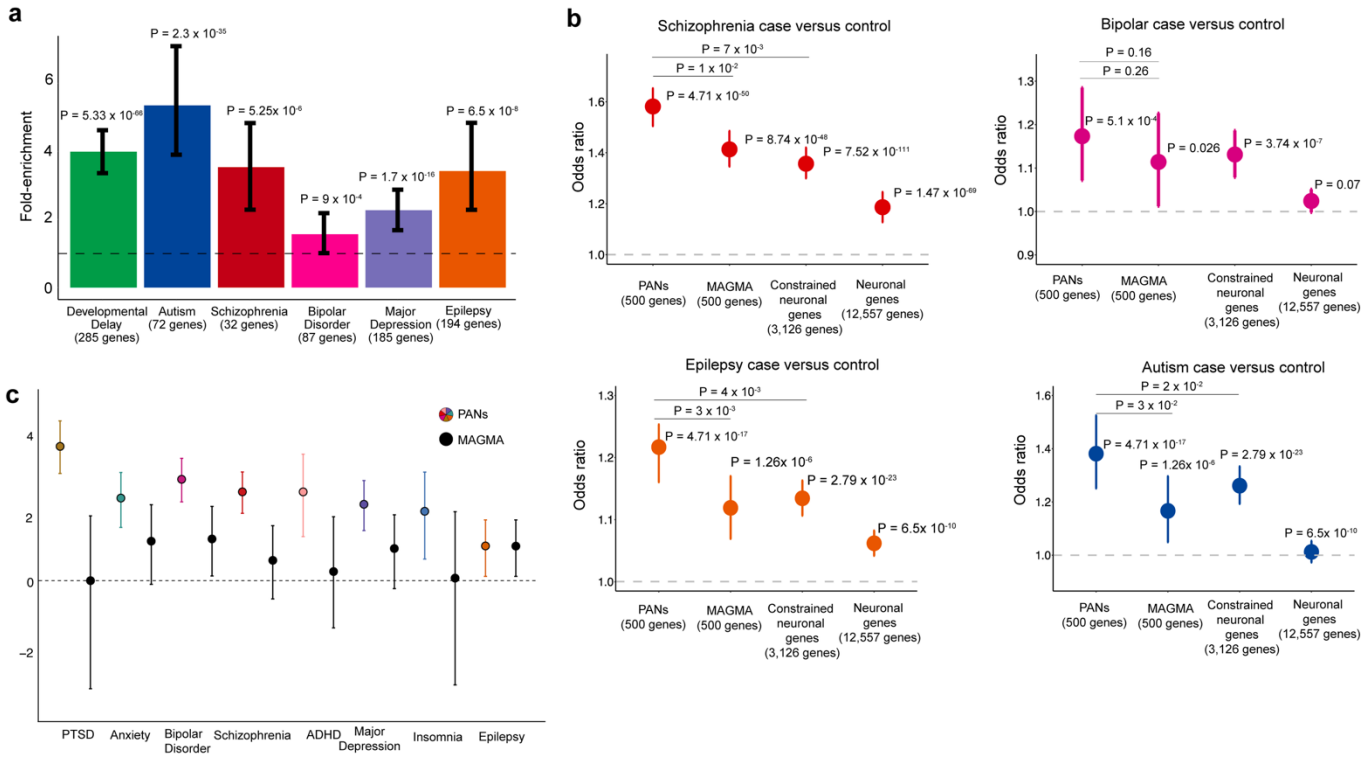


Figure 2. Evaluating PANs. a, Results demonstrating fold-enrichment of PANs prioritized gene scores in exome sequencing data for neuropsychiatric disorders. Error bars represent the 95% CI from a bootstrap of the PAN scores for significant versus non-significant genes in each phenotype (Supplementary Methods). P-values are from a Wilcoxon rank sum test between significant and non-significant genes. For autism, developmental delay and schizophrenia we use genes denoted by their significance threshold in the original publications (FDR ≤ 0.001, Bonferroni significant and FDR ≤ 0.05, respectively). For bipolar disorder, major depression and epilepsy, we use genes with either a protein-damaging or missense p-value < 0.01. **b,** PANs-prioritized genes show stronger ultra-rare variant enrichment across all four disorders. Odds ratios and 95% Cis are shown for the enrichment of rare protein-altering variants in schizophrenia, bipolar disorder, epilepsy, and autism case–control cohorts. Each panel compares the top 500 constrained genes prioritized, as well as global neuronal gene sets. Reported P values reflect two-sided Fisher's exact tests comparing the burden of PTV and MPC > 3 variants in cases versus controls across each gene set. Horizontal P values reflect permutation testing (n = 1,000) comparing the PANs set to random subsets of constrained MAGMA-prioritized or neuronal genes **c,** OpenTargets enrichment in prioritized genes for eight neuropsychiatric disorders. The dots overlapping drug targets (in the top 500 genes in either MAGMA or PANs). The bar represents the 95% CI of the point estimates.represent odds ratio, obtained by applying Fisher's combined probability method on counts of the points estimates.

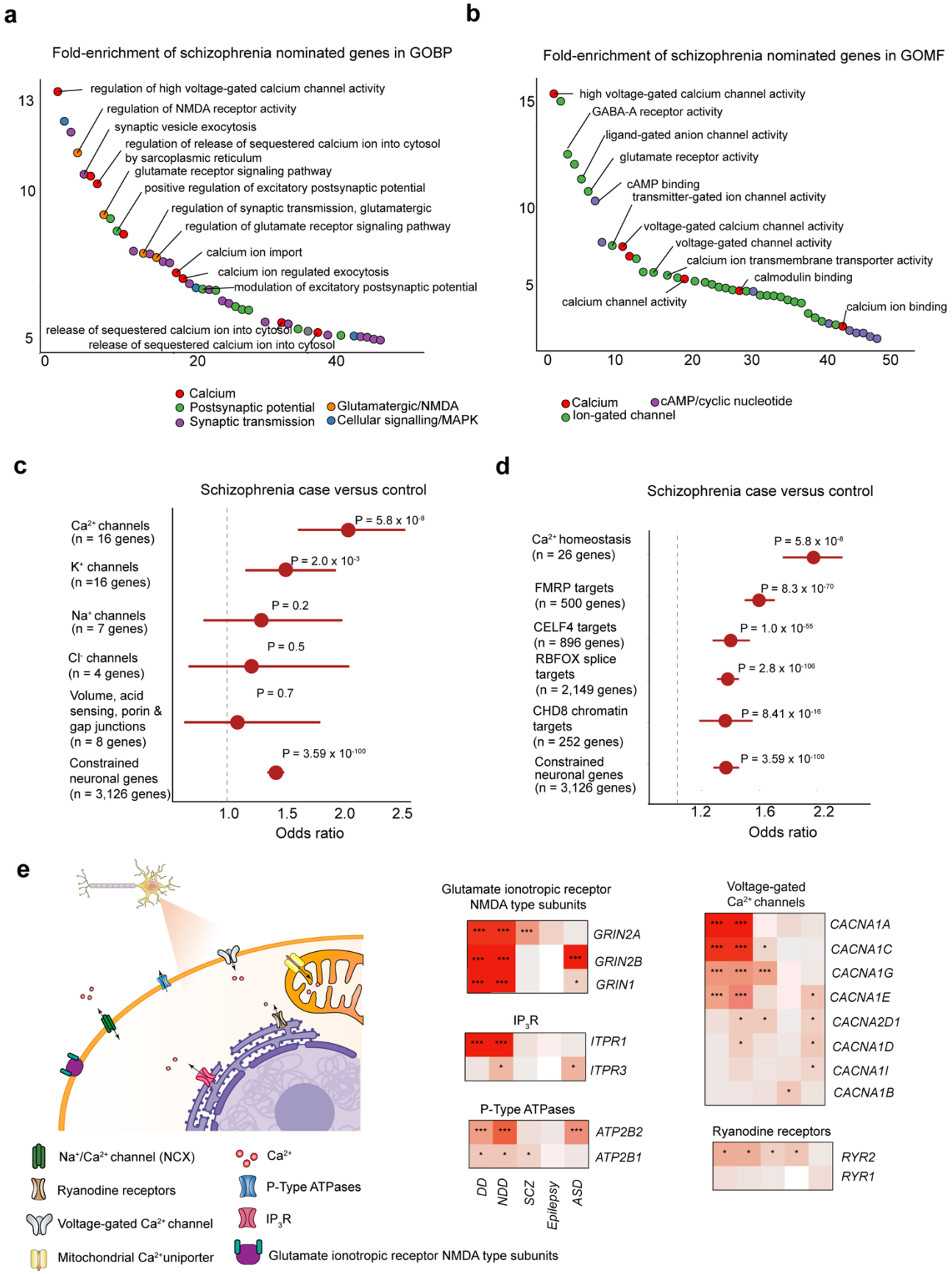

a
Fold-enrichment of schizophrenia nominated genes in GOBP
regulation of high voltage-gated calcium channel activity
regulation of NMDA receptor activity
synaptic vesicle exocytosis
regulation of release of sequestered calcium ion into cytosol by sarcoplasmic reticulum
glutamate receptor signaling pathway
positive regulation of excitatory postsynaptic potential
regulation of synaptic transmission, glutamatergic
regulation of glutamate receptor signaling pathway
calcium ion import
calcium ion regulated exocytosis
modulation of excitatory postsynaptic potential
release of sequestered calcium ion into cytosol
release of sequestered calcium ion into cytosol
Calcium
Postsynaptic potential
Synaptic transmission
Glutamatergic/NMDA
Cellular signalling/MAPK
b
Fold-enrichment of schizophrenia nominated genes in GOMF
high voltage-gated calcium channel activity
GABA-A receptor activity
ligand-gated anion channel activity
glutamate receptor activity
cAMP binding
transmitter-gated ion channel activity
voltage-gated calcium channel activity
voltage-gated channel activity
calcium ion transmembrane transporter activity
calmodulin binding
calcium channel activity
calcium ion binding
Calcium
Ion-gated channel
cAMP/cyclic nucleotide
c
Schizophrenia case versus control
Ca2+ channels (n = 16 genes)
K+ channels (n =16 genes)
Na+ channels (n = 7 genes)
Cl- channels (n = 4 genes)
Volume, acid sensing, porin & gap junctions (n = 8 genes)
Constrained neuronal genes (n = 3,126 genes)
P = 5.8 x 10-8
P = 2.0 x 10-3
P = 0.2
P = 0.5
P = 0.7
P = 3.59 x 10-100
Odds ratio
d
Schizophrenia case versus control
Ca2+ homeostasis (n = 26 genes)
FMRP targets (n = 500 genes)
CELF4 targets (n = 896 genes)
RBFOX splice targets (n = 2,149 genes)
CHD8 chromatin targets (n = 252 genes)
Constrained neuronal genes (n = 3,126 genes)
P = 5.8 x 10-8
P = 8.3 x 10-70
P = 1.0 x 10-55
P = 2.8 x 10-106
P = 8.41 x 10-16
P = 3.59 x 10-100
Odds ratio
e
Glutamate ionotropic receptor NMDA type subunits
GRIN2A
GRIN2B
GRIN1
IP3R
ITPR1
ITPR3
P-Type ATPases
ATP2B2
ATP2B1
DD
NDD
SCZ
Epilepsy
ASD
Voltage-gated Ca2+ channels
CACNA1A
CACNA1C
CACNA1G
CACNA1E
CACNA2D1
CACNA1D
CACNA1I
CACNA1B
Ryanodine receptors
RYR2
RYR1
Na+/Ca2+ channel (NCX)
Ryanodine receptors
Voltage-gated Ca2+ channel
Mitochondrial Ca2+uniporter
Ca2+
P-Type ATPases
IP3R
Glutamate ionotropic receptor NMDA type subunits

Figure 3. Integrated single cell and genetic analysis reveal Ca^{2+} homeostasis as an etiological driver of schizophrenia. a, b. Fold-enrichment of the top 500 schizophrenia-nominated genes in Gene Ontology Biological Process (GOBP) and Gene Ontology Molecular Function (GOMF) terms. Each point corresponds to a GOBP/GOMF term. The highest fold-enrichment is observed in terms related to Ca^{2+} ion channel processes. **c**, Ion channels were stratified by conductance type: calcium, potassium, sodium, chloride, and non-selective/other (e.g., volume-sensing, porins, gap junctions). Calcium channels showed the strongest enrichment for rare damaging variants in schizophrenia cases versus controls (see Methods: Ion Channel Gene Set Selection for details as to how gene sets were defined). **d,** Case–control enrichment of schizophrenia coding variants in constrained members of the Ca^{2+} homeostasis genes set compared to neuronal and regulatory gene sets. Two-sided P-values from Fisher's exact test compare variant burden between cases and controls. Dots represent odds ratios; bars show 95% confidence intervals. **e,** Select members of the calcium homeostasis gene list (left) and their corresponding p values in neuropsychiatric exome studies (right). full list of genes in **Supplementary Table 7.**

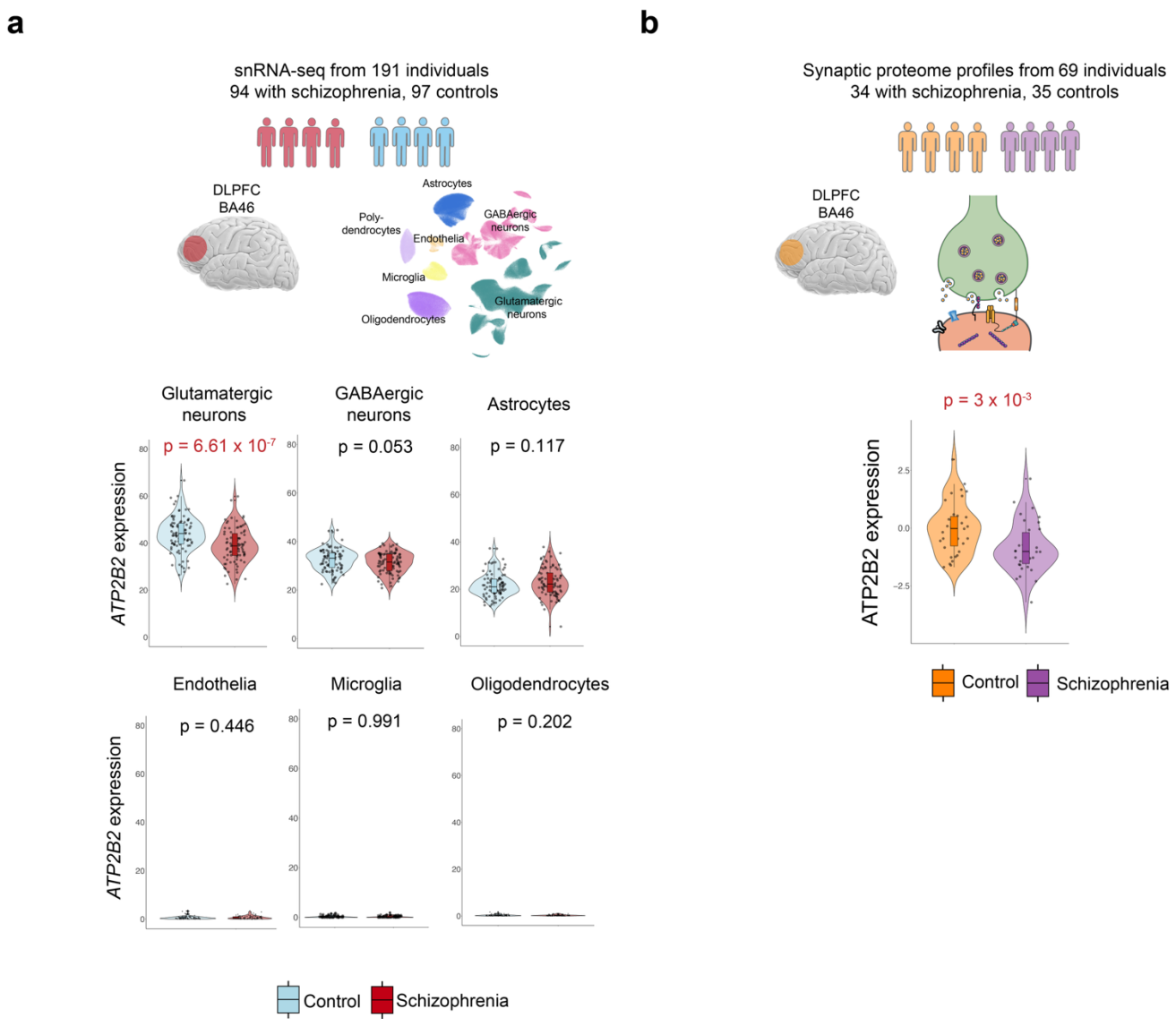


Figure 4. ***ATP2B2* is significantly under-expressed in donors with schizophrenia versus donors without**. **a**, Shown is *ATP2B2* expression (per 10^5 detected nuclear transcripts) in each cell type in individual donors. P-values are computed by a linear regression of normalized expression while controlling for PMI, age and sex. Box plots display interquartile ranges, whiskers extend to 1.5 times the interquartile range, central lines represent medians, and notches indicate confidence intervals around the medians. **b,** ATP2B2 protein abundances are significantly lower in synapses in the DLPFC of donors with schizophrenia. Shown are covariate adjusted ATP2B2 synaptic protein abundance measurements. P-values are computed from a linear regression model. Box plots display interquartile ranges, whiskers extend to 1.5 times the interquartile range, central lines represent medians, and notches indicate confidence intervals around the medians.

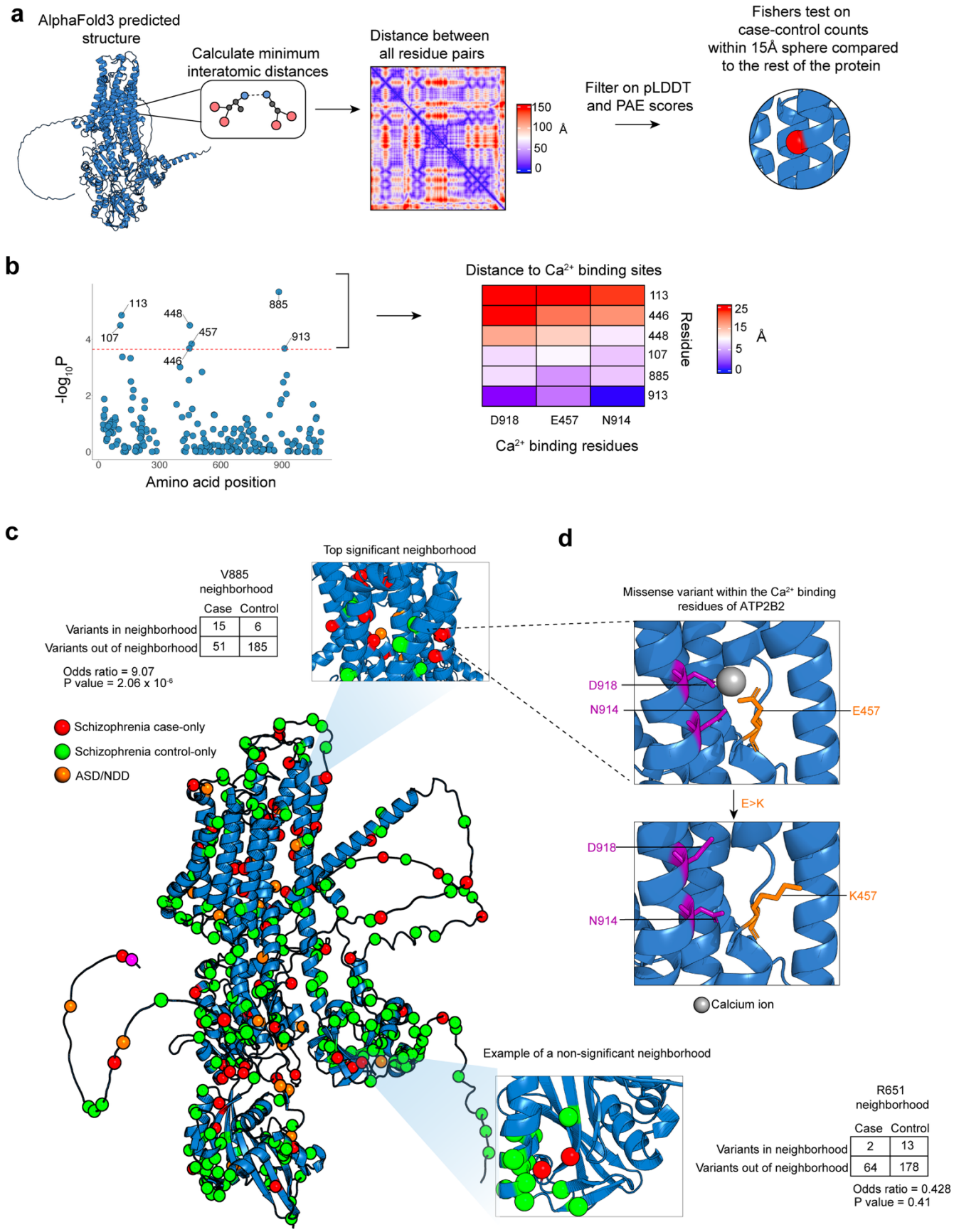

a
AlphaFold3 predicted structure
Calculate minimum interatomic distances
Distance between all residue pairs
150
100
50
0
Å
Filter on pLDDT and PAE scores
Fishers test on case-control counts within 15Å sphere compared to the rest of the protein
b
113
448
885
107
457
913
446
-log10P
Amino acid position
Distance to Ca2+ binding sites
113
446
448
107
885
913
Residue
D918
E457
N914
Ca2+ binding residues
25
15
5
0
Å
c
Top significant neighborhood
V885 neighborhood
Case Control
Variants in neighborhood 15 6
Variants out of neighborhood 51 185
Odds ratio = 9.07
P value = 2.06 x 10-6
Schizophrenia case-only
Schizophrenia control-only
ASD/NDD
d
Missense variant within the Ca2+ binding residues of ATP2B2
D918
N914
E457
E>K
K457
Calcium ion
Example of a non-significant neighborhood
R651 neighborhood
Case Control
Variants in neighborhood 2 13
Variants out of neighborhood 64 178
Odds ratio = 0.428
P value = 0.41

Figure 5. 3D-neighborhood analysis of neuropsychiatric associated variants reveals impact on ATP2B2 Ca^{2+} binding and activity. a, Using the AF3 solved structure of ATP2B2, we calculate the shortest interatomic distances between each pair of amino acid residues, before removing low confidence pLDDT and PAE scores. We then construct a 15Å sphere around each residue and perform Fisher's exact tests to test for enrichment of schizophrenia, autism and NDD associated variants within the sphere compared to the rest of the protein (See Methods: 3D protein neighborhood test for details). **b**, Left: P-values from the per-residue neighborhood burden enrichment analysis. The red line represents the Bonferroni significance threshold (P value = 2.2×10^{-4}), residues reaching statistical significance are labelled. Numerical results are in **Supplementary table 10**. Right: Heatmap of the distances (in Å) of each of the Bonferroni significant residues to the three canonical Ca^{2+}-binding residues (D918, E457, and N914). in ATP2B2. Several enriched residues lie within <10 Å of the binding pocket. **C,** Case-only schizophrenia, ASD and NDD-associated variants are mapped onto the AlphaFold-predicted structure of ATP2B2. Two regions are highlighted: the most significant site (V885, top), enriched for case variants, and a representative site (R651, bottom) that is depleted for case variants and likely less relevant to disease-related protein function. 2x2 tables adjacent to the plot consist of the case-control counts from the neighborhood analysis corresponding to that highlighted residue. **d,** Ca^{2+}-binding site in ATP2B2 illustrating the Ca^{2+} binding residues in transmembrane domains IV and VI. The E457K variant disrupts one of the three Ca^{2+} binding residues in the channel. In purple are the two other residues, D918 and N914.

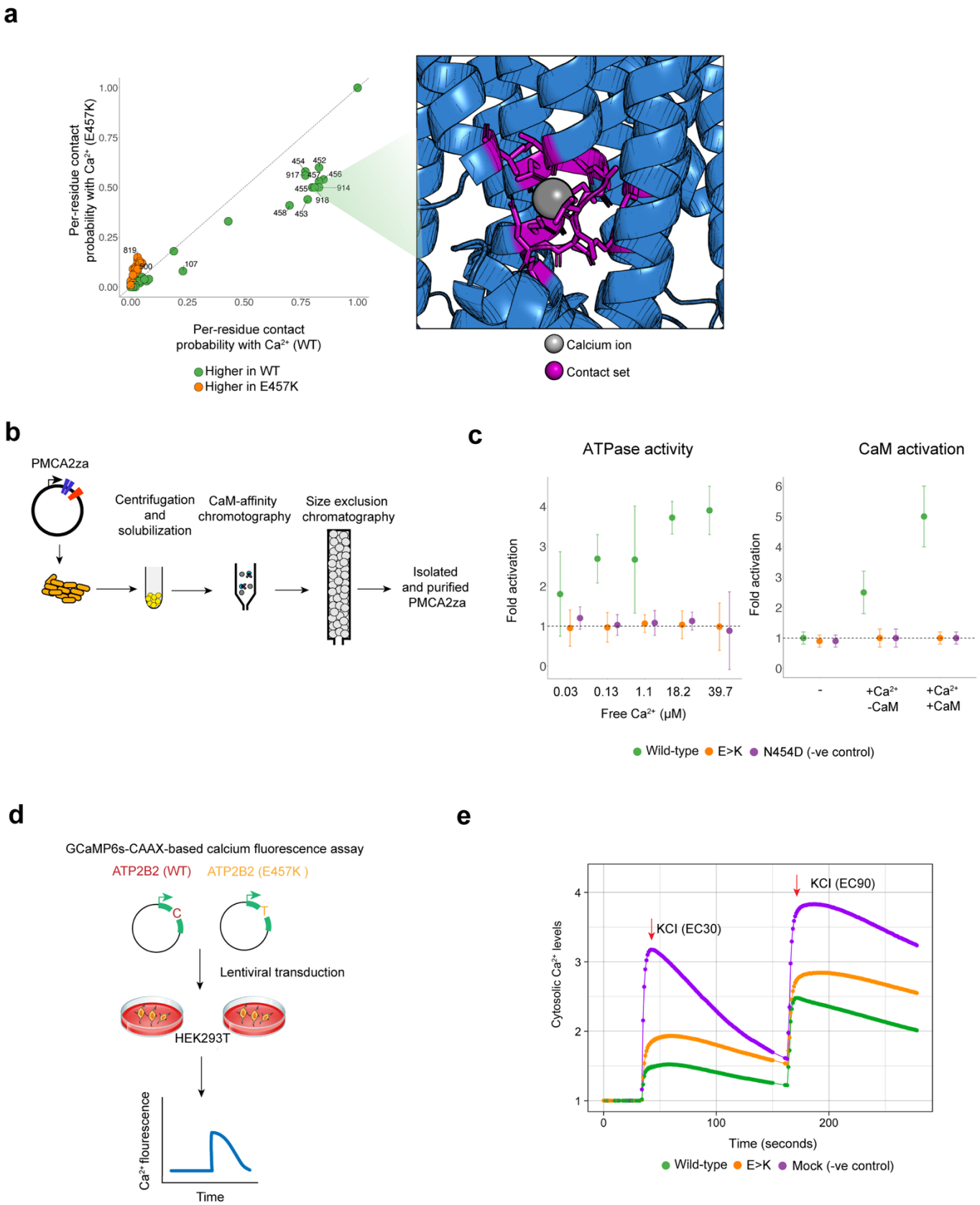

a
Per-residue contact probability with Ca2+ (E457K)
Per-residue contact probability with Ca2+ (WT)
Higher in WT
Higher in E457K
Calcium ion
Contact set
b
PMCA2za
Centrifugation and solubilization
CaM-affinity chromotography
Size exclusion chromatography
Isolated and purified PMCA2za
c
ATPase activity
CaM activation
Fold activation
Free Ca2+ (μM)
Wild-type
E>K
N454D (-ve control)
d
GCaMP6s-CAAX-based calcium fluorescence assay
ATP2B2 (WT)
ATP2B2 (E457K)
Lentiviral transduction
HEK293T
Ca2+ flourescence
Time
e
KCl (EC30)
KCl (EC90)
Cytosolic Ca2+ levels
Time (seconds)
Wild-type
E>K
Mock (-ve control)

Figure 6: E457K Mutation in ATP2B2 Impairs Calcium Binding, Extrusion and ATPase Activity. a, AF3 simulations of contact probabilities for each residue indicate that the E457K mutation reduces Ca^{2+} binding. The simulation was performed using the AF3 structure of ATP2B2 with a single Ca^{2+} ion, comparing the wild-type structure to the E457K mutant. The residues most affected by the mutation, particularly those within the calcium channel, are highlighted and labeled in purple in the AF3 model. ***b,c*** The equivalent substitution of E457K in an isoform of ATP2B2 eliminates its ability to extrude calcium, as demonstrated by a lack of ATPase activity and activation in the presence of calmodulin (CaM). Left panel: A colorimetric assay measuring phosphate release shows no detectable ATPase activity for the E457K mutant. Right panel: Fold activation of ATP2B2 under three conditions: in the absence of Ca^{2+}, with Ca^{2+} but no CaM, and with both Ca^{2+} and CaM. The N454D mutant is a non-functional variant and serves as a negative control. Data are normalized to the ATPase activity of wild-type PMCA2za in the absence of calcium.**d**,Synthesized DNA constructs containing both ATP2B2 alleles were inserted into a lentiviral vector and transduced into HEK293T cells engineered to express Kir2.3 channels and the membrane-tethered genetically encoded calcium sensor GCaMP6s-CAAX. Following transduction, cells were hyperpolarized by extracellular KCl, leading to calcium influx. Intracellular calcium levels were then quantified using a high-throughput fluorometric imaging plate reader. **e**, The ATP2B2-E457K impairs Ca^{2+} extrusion mechanisms and disrupts cytosolic Ca^{2+} levels. HEK293T cells were transduced with wild-type ATP2B2 (green), E>K mutant ATP2B2 (orange), or mock control (purple) were subjected to sequential KCl stimulation at EC30 and EC90 (indicated by red arrows). Cytosolic calcium levels were measured over time using the genetically encoded calcium sensor GCaMP6s-CAAX. Calcium measurements are normalized to baseline calcium levels. Numerical results are in **Supplementary table 9**.

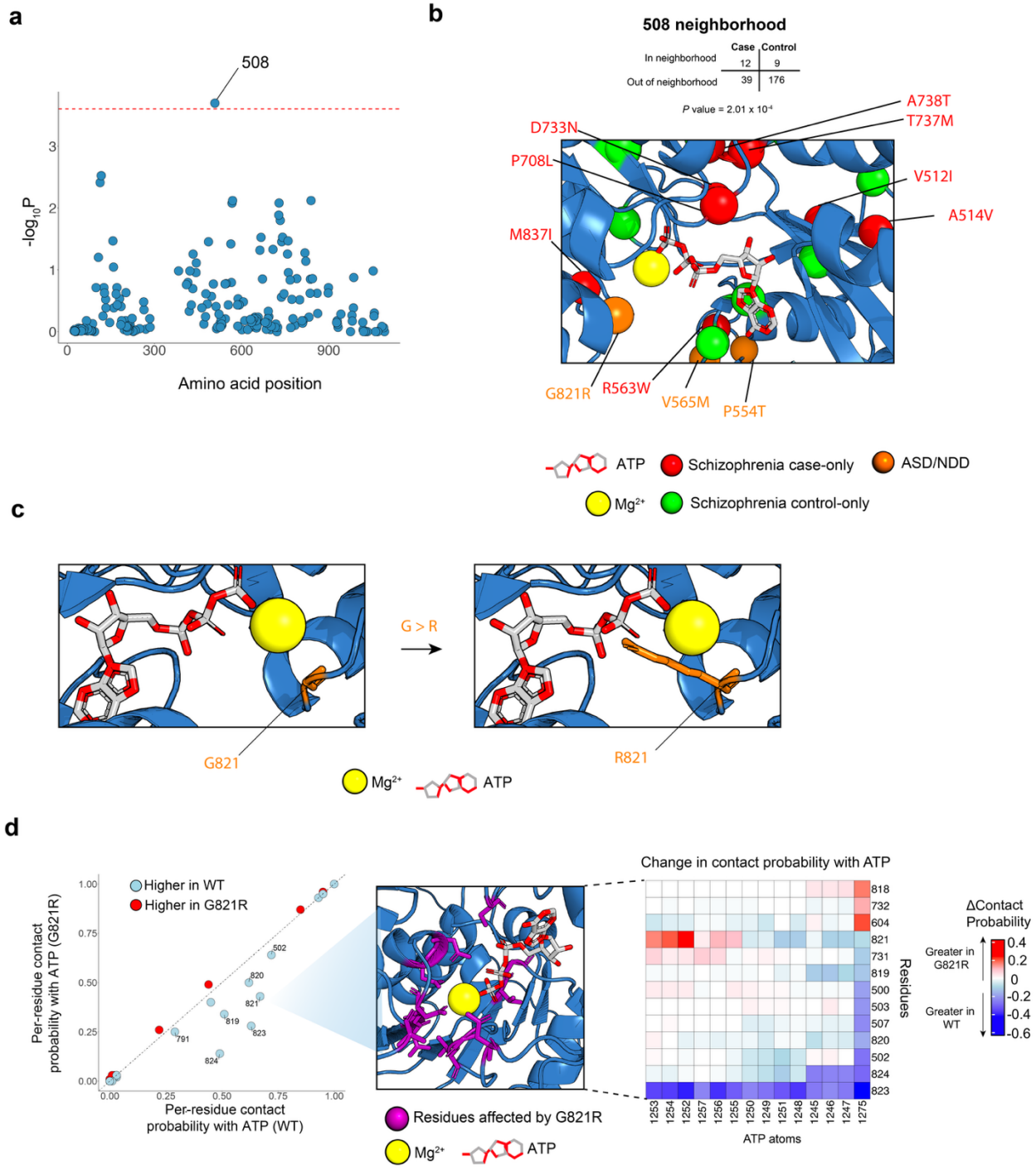


Figure 7: Conditional analysis uncovers a signal near the ATP:Mg^{2+} binding site. Conditional analysis on the Ca^{2+} binding site revealed an additional signal at ATP2B2's ATP:Mg^{2+} coordination site, suggesting potential impairment of ATPase activity through disrupted ATP hydrolysis. The panel shows P-values (as in Fig. 5B) from this conditional analysis, with the red line marking the Bonferroni significance

threshold and significant residues labeled. The red line represents the Bonferroni significance threshold, with labeled residue indicating those reaching statistical significance. **b**, Case-only schizophrenia, ASD and NDD-associated variants are mapped onto the AlphaFold-predicted structure of *ATP2B2 around the* ATP:Mg^{2+} binding site. 2x2 tables adjacent to the plot indicate the case-control counts from the neighborhood analysis corresponding to that highlighted residue. **c**, Visualization of the G821R variant near the ATP:Mg^{2+} binding site reveals that the introduction of a large arginine side chain likely alters the local environment. **d**, The substitution G821R modifies the per-residue contact probabilities of residues near the Mg^{2+} binding site (left) and the ATP binding site (heatmap: right), as computed using AlphaFold3. Middle: Visualization of the residues most impacted by G821R.

Methods

LDSC

S-LDSC assesses the contribution of a genomic annotation to the heritability of a trait. Namely, it assumes that the per-SNP heritability or variance of effect size (of standardized genotype on trait) of each SNP is equal to the linear contribution of each annotation:

$$Var(\beta_i) = \sum_k \tau_k 1\{i \in C_k\}$$

Where the marginal Chi-square association statistic for SNP *i* reflects the causal contributions of all SNPs in LD with SNP *i*.

$$E[\chi_i^2] = 1 + Na + N\sum_k \tau_k \ell(i,k)$$

Where a is a constant that reflects sources of confounding and N is the GWAS sample size [89], and $\ell(i,k)$ is the LD score of SNP *i* to category C_k. The regression coefficient τ_k quantifies the importance of annotation C_k, correcting for all other annotations in the model. Specifically, τ_k which is the proportionate change in per-SNP heritability associated with a standard deviation change in the value of the annotation, conditional on other annotations included in the model [7,13].

PANs

Given a dataset of neuronal cell types, we first computed a conservative expression estimate for each gene within a cluster. For gene *g*, let the mean expression across all cells be μ_g, the standard deviation be σ_g, and the number of cells in the cluster be n_c. The standard error of the mean was defined as:

$$\mathrm{se}(g) = \frac{\sigma_g}{\sqrt{n_c}}$$

We then defined the adjusted expression value as:

$$T(g) = \mu_g - \mathrm{se}(g)$$

For each pair of cell types, we defined two one-sided differential expression gene sets: the top 1,000 genes enriched in cell type A relative to cell type B, and conversely the top 1,000 genes enriched in B relative to A. Genes were ranked by the $\log_2$ fold-change of , and we applied a floor on expression to exclude genes supported by very few reads.

All of these pairwise-defined gene sets were then used as annotations in stratified LD score regression to compute , (conditioning on the full baseline model). We applied family-wise error rate correction across all tested pairs. We then defined an indicator function:

$$S_{ij} = \begin{cases} 1, & \text{if } P_{ij} \leq \alpha \\ 0, & \text{if } P_{ij} > \alpha \end{cases}$$

where P_{ij} is the raw p-value and α is the FWER-adjusted threshold.

For each gene, we consider all gene sets G in which g appears. From stratified LD score regression, each gene set G has an estimated per-SNP heritability enrichment coefficient $\tau_G$and an associated standard error $se(\tau_G)$. The PANs score is calculated by summing standardized enrichment coefficients across all gene sets in which a gene appears, restricted to those gene sets that survive FWER[94] correction as such:

$$\mathrm{PANs}(g) = \sum_{G \ni g} S_{ij} \frac{\tau_G}{\mathrm{se}(\tau_G)}.$$

.

MAGMA

MAGMA implements a ‘snp-wise-mean model’, which computes a gene-level test statistic from the full set of SNPs from GWAS summary statistics. It does so by computing the mean of the χ^2 statistic for all SNPs in a gene, and then uses a variant of Brown's method to approximate an empirical p-values for each gene while accounting for the LD structure of the genome. For studies for which SNPs to genes were mapped[93] , we used the per-gene p-values provided in the study. Otherwise, summary statistics

from all phenotypes were uploaded to FUMA ([94]webserver at https://fuma.ctglab.nl). Gene-level Z-scores were calculated using the following parameter settings. The reference panel population was set to "1000G Phase3 EUR". The MHC region was excluded due to its unusual genetic architecture and L and the P value threshold of 5×10^{-8} and r^2 threshold were set to 0.1, while MAF was set to > 0.01. MAGMA Z-scores were used for downstream analyses.

Prioritization Score

Both PANs and MAGMA compute per-gene scores for all protein coding genes. Let G be the top nominated gene's score and S_i be the sum of the scores of all genes i within 200kb of a genome-wide associated locus. The prioritization score P is defined as:

$$P = \frac{G}{\sum_i S_i}$$

By definition, since MAGMA outputs Z-scores, some genes may have a Z-score < 0. We always omit genes with Z-scores < 0 since those are unlikely to be interpreted as genes associated with any single trait.

ScRNA-seq data pre-processing

All scRNA-seq datasets utilized in this study are publicly accessible cell-by-gene expression matrices, aligned to the GRCh38 human transcriptome. For the caudate dataset, we used the caudate dataset from[95], while the DLPFC snRNA-seq data is from[614]. We generated the amygdala dataset as per previously published protocols. Namely, the amygdala data is from frozen tissue blocks from the amygdala of four healthy brain donors. Comprehensive health records, detailed family questionnaires, toxicology screening and neuropathology reports were reviewed to exclude donors with neurodegenerative or psychiatric disorders, as well as recent history of substance use. For each donor, the full amygdala in its rostro-caudal, dorso-ventral, and medio-lateral extent was contained in two to three coronal blocks. These blocks were sectioned using a cryostat. Approximately every 100 μm, a section was mounted on a glass slide and stained with Luxol Blue to delineate the boundaries of the amygdala. These boundaries were then lightly etched (~ 100 μm depth) on the tissue block surface using a surgical blade to allow precise dissection of the amygdala in subsequent sections. Ten to twelve serial sections (12 μm each) were collected into Eppendorf tubes, ensuring that each tube contained a representative sampling from the rostral, middle, and posterior amygdala. One vial per case was used for RNA sequencing.

To address variability in sequencing depth across cells from all brain regions, we applied uniform normalization techniques to each count matrix. This involved normalizing the total number of unique molecular identifiers (UMIs) per cell, converting these counts to transcripts per 10,000 (TP10K), and then taking the logarithm to derive the final expression unit: log2(TP10K + 1). Seurat was employed for various analytical processes including data scaling, transformation, clustering and dimensionality reduction[96]. The SCTransform() function was used to scale and transform the data, as well as to identify variable genes[97]. Linear regression was performed to eliminate unwanted variations related to cellular complexity (such as the number of genes and UMIs per cell) or cell quality (percentage of mitochondrial and rRNA reads). Principal component analysis (PCA) was conducted using the identified variable genes, with the first 20 principal components (PCs) utilized to perform UMAP, embedding the dataset into two dimensions.

Subsequently, these first 20 PCs were used to construct a shared nearest-neighbor (SNN) graph via the FindNeighbors() function. This SNN was then leveraged to cluster the dataset using the FindClusters() function, implementing a graph-based modularity optimization algorithm based on the Louvain method for community detection.

GWAS summary statistics

We analyzed publicly available GWAS summary statistics for unique traits genetic correlation <0.9. We also used the summary statistics for SNPs with minor allele count >5 in a 1000 Genomes Project European reference panel[98].

Gene set and drug target analyses

To broadly prioritize schizophrenia-relevant processes, we used GOrilla[99], which identifies enriched GO terms in a given list of genes, to identify genes highly nominated for each phenotype by PANs. We excluded terms that contained less than 10 genes. We used the Ensembl identifiers of the underlying genes from either analysis as foreground set and all genes used in the PANs analysis as background. Analysis of gene sets specific to nervous system biology which have been previously implicated in neuropsychiatric disorders were also included, such as translational targets of FMRP[54], chromatin targets of CHD8[55], splice targets of RBFOX[56] and *CELF4* target genes, (defined as genes with an iCLIP occupancy > 0.2)[100].

For drug target analysis, we manually extracted drug target gene sets from Open Targets[101] for all eight neuropsychiatric disorders. For each disorder, we selected all genes with drug score > 0 (clinical trial phase 0 and above); each trait was well powered containing at least 43 genes (anorexia) and up to 309 genes (epilepsy). For MAGMA, the per-gene p-values are calculated using the SNP2GENE function

from FUMA for each trait, except for insomnia, which was extracted from Watanabe et al.[93], as this was substantially more well powered than the publicly available summary statistics for that trait.

Ion Channel Gene Set Selection

To conduct ion channel analyses, we retrieved 215 ion channel genes from HGNC, grouped by major permeant ion (e.g., Ca^{2+}, potassium, sodium, chloride). In addition to canonical channel types, we defined a composite category comprising volume-sensing channels, acid-sensing ion channels (ASICs), porins, and gap junction proteins (connexins and pannexins), reflecting their distinct functional roles and smaller representation in the genome; this group is labeled as "Volume & acid sensing, porin & gap junctions." We restricted all analyses to genes that were expressed in the brain, defined by aggregating all frontal cortical neurons (excitatory and inhibitory) and selecting genes with logFC > 0, padj < 0.01, and avgExpr > 0.01 using the presto Wilcoxon AUC method[102]. Furthermore, within each ion channel class, we only tested genes that were also constrained (defined as having LOEUF observed/expected bin < 2). This filtered, neuronally expressed, and constrained gene set served as the background for all comparative analyses.

Selective Constraint

We used LOEUF (loss-of-function observed/expected upper bound fraction, O/E) measures of intolerance to loss-of-function mutations extracted from gnomAD.v2.1's pLoF metrics by gene data[69]. We used genes in the O/E bin of < 2 to denote constraint.

Transcriptomic and proteomic schizophrenia case-control analysis

To assess the relationship between *ATP2B2* expression and schizophrenia status, we used snRNA-seq samples from[61] , as provided by the authors, consisting of counts scaled to 100k UMIs per donor. We performed a multiple linear regression with ATP2B2 expression as the dependent variable, and schizophrenia status, age, sex, and post-mortem interval (PMI) as independent variables. Diagnostic tests were performed to ensure the assumptions of linearity, independence, homoscedasticity, normality of residuals, and absence of multicollinearity were met.

For proteomic analyses, we used synaptic protein abundance measurements generated by mass spectrometry as in[62], which was log-transformed and processed through median centering as well as median absolute deviation scaling applied on a sample-by-sample basis. We constructed a linear model

to examine the relationships between ATP2B2 protein abundance and covariates including diagnosis (schizophrenia versus control), age, sex, and post-mortem interval (PMI) across 69 individuals. To plot the effect of diagnosis controlling for confounding variables, we multiplied the matrix of covariates' values (age, sex, and PMI) with their corresponding regression coefficients and subsequently adjusted the ATP2B2 levels by subtracting these offsets. This adjustment allowed us to isolate the effects attributable to the diagnosis.

Co-occurrence Analysis of Calcium Homeostasis Genes

To quantify gene–gene co-occurrence, we constructed a binary matrix indicating whether each gene was present in each PANs-derived gene set. For every SCHEMA–calcium gene pair, we computed a 2×2 contingency table representing their pattern of co-occurrence across gene sets: the number of sets containing both genes, each gene alone, or neither. We then applied a one-sided Fisher's exact test to assess whether the observed co-occurrence exceeded what would be expected under independence, correcting for each gene's individual frequency of appearance. P-values from all pairwise tests were corrected for multiple comparisons using the Benjamini–Hochberg procedure, and gene pairs with FDR < 0.01 were considered significantly co-occurring.

To assess whether the number of significant SCHEMA–calcium gene connections exceeded chance expectations, we performed a permutation test, whereby in each of 1,000 iterations, we sampled a random gene set of the same size as the calcium gene set from the PANs universe and repeated the co-occurrence analysis with SCHEMA genes. An empirical p-value was computed as the fraction of permutations in which the number of significant co-occurrences was greater than or equal to the observed value (**Figure 2a***).*

Burden Testing

For case-control analyses, we analyzed gene-level variant counts in autism from[19] , bipolar disorder from[16] , schizophrenia from[17] major depression from[103] and epilepsy from[20] which included protein-truncating (PTV) and high-impact missense variants. Genes were filtered for strong LoF constraint (LOEUF bin < 2) and merged with PANs-derived or MAGMA scores, The top 500 genes were selected. We summed the case and control counts of variants in these genes and tested for enrichment using a 2×2 Fisher's exact test based on total variant counts and sample sizes. To assess whether the PANs gene set exhibits stronger enrichment for rare damaging variants in autism cases compared to MAGMA, we performed a permutation-based burden test, where we compute the observed odds ratio for rare

protein-truncating variants (PTVs) and missense variants with MPC > 3 in the top 500 PANs genes using a 2×2 Fisher's exact test. We then randomly sampled 500 genes from the same universe (all high pLI autism-relevant genes) 1,000 times and computed the OR for each random set. The empirical p-value was defined as the proportion of permuted gene sets with an OR greater than or equal to the observed PANs OR.

For de novo burden analysis of rare variants in neurodevelopmental disorders. We used the latest trio data from Kaplanis et al[18]. We assessed enrichment of de novo coding mutations in by comparing observed mutation counts to expected rates based on gene-specific mutation models. For each gene, the observed number of de novo variants was derived by summing reported missense class 1 and class 2 mutations (defined by a combination of MPC and AlphaMissense scores), or PTVs, across all probands in the dataset. Expected mutation rates for each consequence class were obtained from published per-gene mutation rate estimates from gnomad.

Expected de novo counts were computed by summing the relevant mutation rates across the test gene set and scaling by twice the number of probands (to account for diploidy). Enrichment was calculated as the ratio of observed to expected mutations. Statistical significance was assessed using a two-sided Poisson test under the null hypothesis of no enrichment (rate ratio = 1). All analyses were performed separately for missense and PTV classes. Only constrained genes (LOEUF o/e < 2) were included in the analysis.

In Vitro Measurement of ATPase Activity and Calmodulin Activation

The DNA sequence encoding the human ATP2B2 protein (PMCA2za) was cloned into an expression vector and transformed into yeast cells for overexpression. Following induction, yeast cells were harvested and lysed, and membrane fractions were isolated by differential centrifugation. Membrane proteins were then solubilized using 0.1% n-Dodecyl-β-D-maltoside (DDM) for extraction. The solubilized proteins were subjected to a calmodulin affinity column. To further purify the sample and ensure exclusive presence of ATP2B2, the calmodulin-binding proteins were then separated by size exclusion chromatography (SEC). After SEC, ATP2B2 was confirmed as the only protein in the sample by western blot. This preparation was then used for activity measurements shown in Figure 4D.

Testing ATP2B2 variants in HEK293T cells

Cell culture

HEK293T/17 cells were used for generation of lentivirus, and grown in DMEM supplemented with 10% FBS, 100 U/ml penicillin and 100 µg/ml streptomycin. HEK293 CaV3/Kir2.3/GCaMP6 cells were grown in DMEM/F12 media supplemented with 10% FBS, 100 U/ml penicillin and 100 µg/ml streptomycin, and were cultured as previously described[80]. All cells were grown at 37 °C, 95% humidity with 5% CO2.

Lentivirus production

Lentivirus was produced by co-transfecting HEK293T/17 cells with the desired transfer plasmids (or pooled library) and two packaging vectors (psPAX2 and pMD2.G, Addgene #12260 and #12259) using Fugene HD transfection reagent. 6-12 hours post transfection, fresh media was added to the HEK cells, and virus was collected through collecting the media supernatant at 48 hours post transfection then flash frozen. Virus was rapidly thawed prior to transduction.

Plasmids and cell line generation

gblocks containing the WT or E457K ATP2B2 sequences were purchased from Integrated DNA Technologies (IDT) and cloned into a lentiviral expression vector under the PGK promoter, generating constructs with an mCherry-P2A sequence upstream of either the WT or E457K ATP2B2 coding sequence. Lentivirus was produced from each construct and used to transduce HEK293 CaV3/Kir2.3/GCaMP6 cells at a multiplicity of infection (MOI) of 0.1 by adding the virus to the media following cell splitting. Given the potential for lentiviral transgene silencing in HEK cells, transduced cells were sorted based on mCherry expression (as a proxy for ATP2B2 expression) one day before seeding for the FLIPR assay.

FLIPR assay for testing calcium levels

A day after sorting cells based on mCherry expression, 15,000 cells per well were seeded into poly-D-lysine-coated 384-well clear-bottom plates with 1 µg/mL doxycycline. Each plate included 128 wells per condition: (1) mock WT HEK293 CaV3/Kir2.3/GCaMP6 cells, (2) HEK293 CaV3/Kir2.3/GCaMP6 cells expressing WT ATP2B2, and (3) HEK293 CaV3/Kir2.3/GCaMP6 cells expressing E457K ATP2B2. Two test plates were used, along with a KCl titration plate to determine the EC30 and EC90 values for the assay. Two days after plating, the FLIPR assay was performed as previously described[80].

3D Neighborhood analysis

Our pairwise residue distance matrix was computed by extracting all atom coordinates from the *ATP2B2* PDB file generated by AF3, with the distance being the shortest euclidean distance between all atoms of all residue pairs. Residues were filtered based on AlphaFold-predicted pLDDT scores, computed as the average B-factor of their atoms, and only residues with pLDDT scores >50 were included in the analysis. Distances with PAE scores greater than 15 in both directions were excluded. For each residue, case and control counts within its neighborhood were compared using a chi-squared test, with neighborhoods were defined as residues within a 15 Å radius. A Bonferroni correction adjusted the significance to account for the number of tested residues. Residues with p-values below the corrected threshold were considered significant and are reported in Figure 5.

Graphics

We used R to generate all plots (R version 4.1, 4.2 and 4.3). We generated enrichment heatmaps, gene term enrichment, error plots, box plots, distribution plots and scatterplots using a combination of ggplot2 (v.3.3.6) and ggpubr (v.0.4.15).

Statistics and reproducibility

All data used in the present study were generated and designed by the original studies and no statistical method was used to predetermine sample size. No data were excluded from the analyses. The experiments were not randomized. The Investigators were not blinded to allocation during experiments and outcome assessment.

Code availability

The 3D neighborhood test is available as an open-source Python package at *https://github.com/sherifgerges/neighborhood_test*

Table 1. List of Ca^{2+} homeostasis genes.

Full Ca^{2+} homeostasis gene list (related to figure 3A) with corresponding P-values from exome sequencing studies for schizophrenia, ASD, NDD, DD and Epilepsy.

Acknowledgements

We thank Jacob Ulirsch and Tushar Kamath for helpful discussions regarding ATP2B2 analyses. We are also grateful to members of the McCarroll laboratory and the Stanley Center, including Marta Florio, Steven Burger, Yong Hoon Kim, and Avin Veerakumar, for their advice and discussions. We further thank Beryl Cummings, Steve Hyman, Ed Scolnick, and Benjamin Neale for valuable input, and we are deeply grateful to the brain tissue donors and their families, without whom this study would not have been possible.

Contributions

S.G, S.A.M and M.J.D designed the study. S.G and M.J.D performed the single-cell analysis with input from E.L., M.G., T.S., N.K. and S.A.M. J.Y and N.K. performed the synaptic proteomic analysis of ATP2B2. N.C.S, P.N and C.S designed and tested ATP2B2 variants in the colorimetric assay. M.A.B designed the cellular assay to test ATP2B2 variants with input from S.G, M.J.D, J.W. and J.Q.P. M.A.B and R.L performed the assay. F.K.S collected ATP2B2 variants for the ASD/NDD analysis. H.F. and S.G performed the 3D neighborhood enrichment analysis with input from M.J.D.

Corresponding authors

Correspondence to Steven A. McCarroll or Mark Daly.

Supplementary Materials

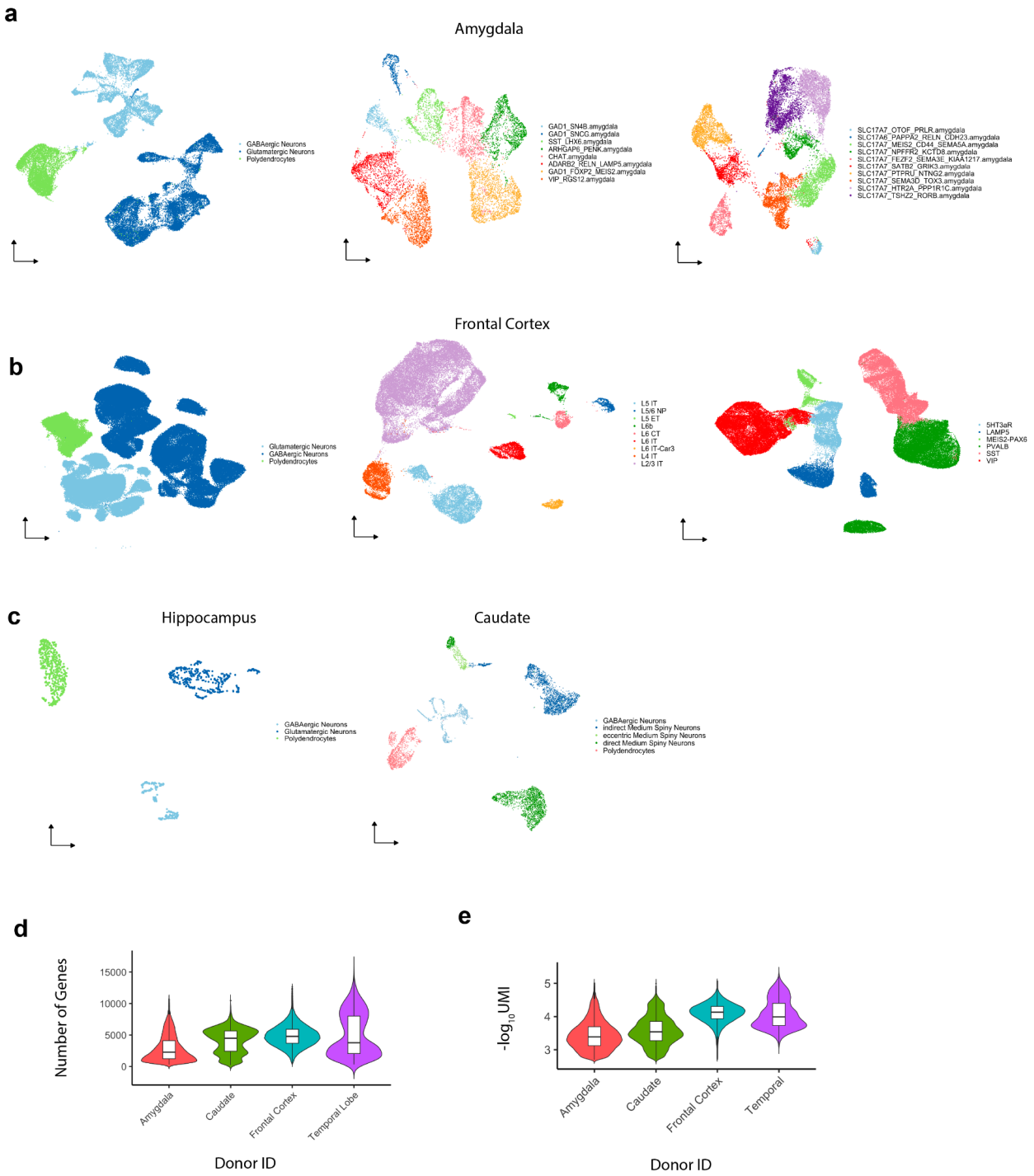
1. Supplementary Figure 1. Single-nucleus RNA-seq analysis of brain tissue from four brain regions
2. Supplementary Figure 2. Benchmarking the PANs method.
3. Supplementary Figure 3. Rare variant drug target analysis.
4. Supplementary Figure 4. Gene ontology analysis of epilepsy, bipolar disorder and IQ
5. Supplementary Figure 5. LocusZoom plots of hand-selected GWAS genes alongside their corresponding rare-variant signals
6. Supplementary Figure 6. Rare variant burden in ion channel gene sets across schizophrenia, epilepsy, bipolar disorder, and autism
7. Supplementary Figure 7. Significant Co-occurrence Between Calcium Homeostasis and Disease Genes.
8. Supplementary Figure 8. Permutation-based co-occurrence analysis of calcium homeostasis genes with NDD, SCZ, and ASD gene sets
9. Supplementary Figure 9. Additional calcium analysis in neurodevelopment disorders and statistical analysis of PANs network.
10. Supplementary Figure 10. Comparisons of AlphaMissense pathogenicity scores in cases versus controls of schizophrenia.
11. Supplementary Figure 11. Result of the schizophrenia only 3D neighborhood analysis
12. Supplementary Figure 12. Missense tolerance ratio (MTR) track of ATP2B2 from the Regeneron Genetics Center Million Exome dataset
13. Supplementary Figure 13. Permutation testing the 3D neighborhood method.
14. Supplementary Figure 14. AlphaFold3-based modeling of ion–ATP2B2 contact probabilities
15. Supplementary Figure 15. Identifying the calcium binding site via structural alignment of ATP2B2 and ATP2B1.
16. Supplementary Figure 16. ATP2B2 expression in excitatory neurons subtypes.

17. Extended discussion of Ca^{2+} homeostasis gene selection

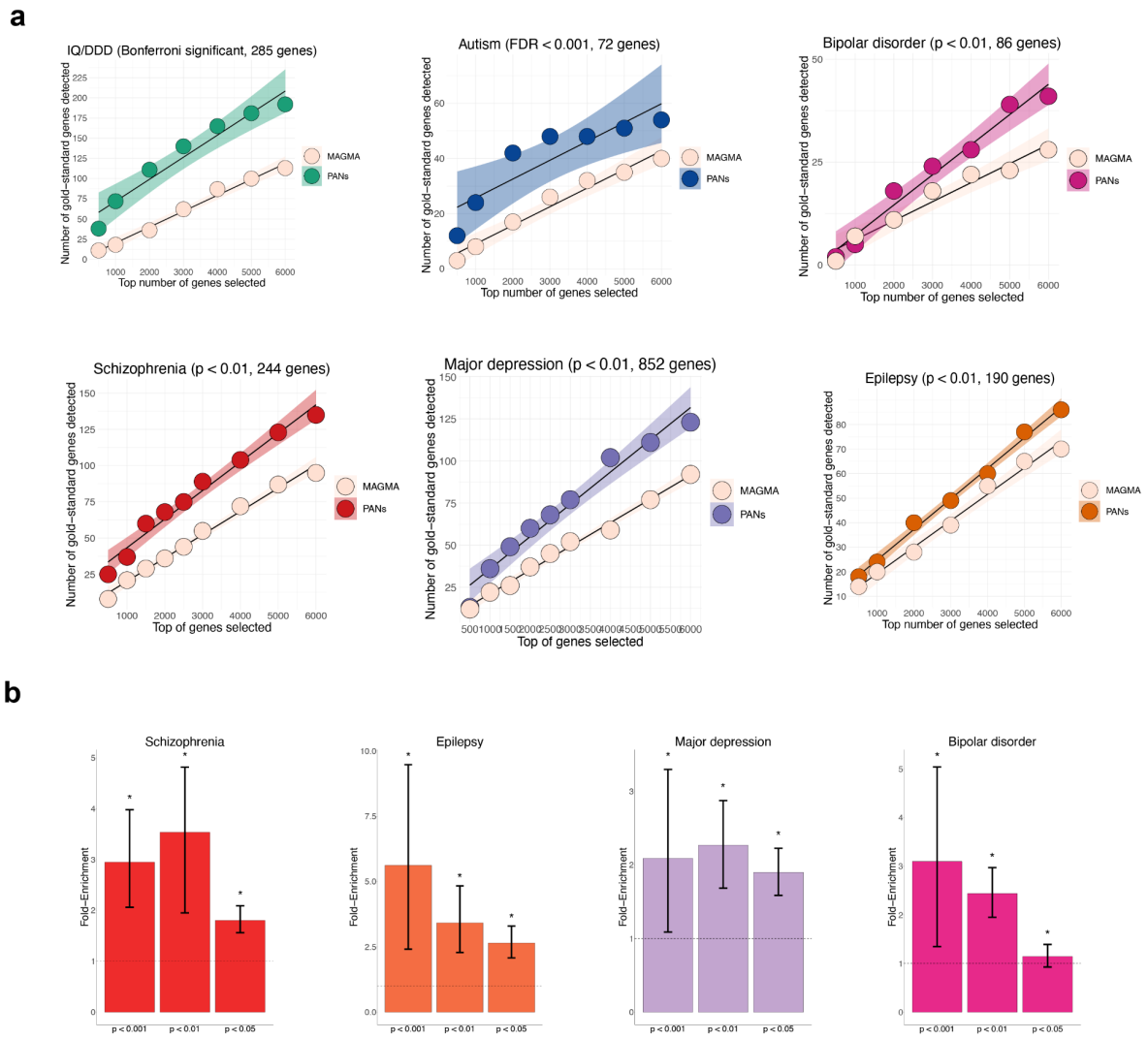
18. Extended Discussion: Simulating Mutation Effects on Per-Residue Contact Probabilities with AlphaFold3

19. Extended discussion of limitations

20. Supplementary methods

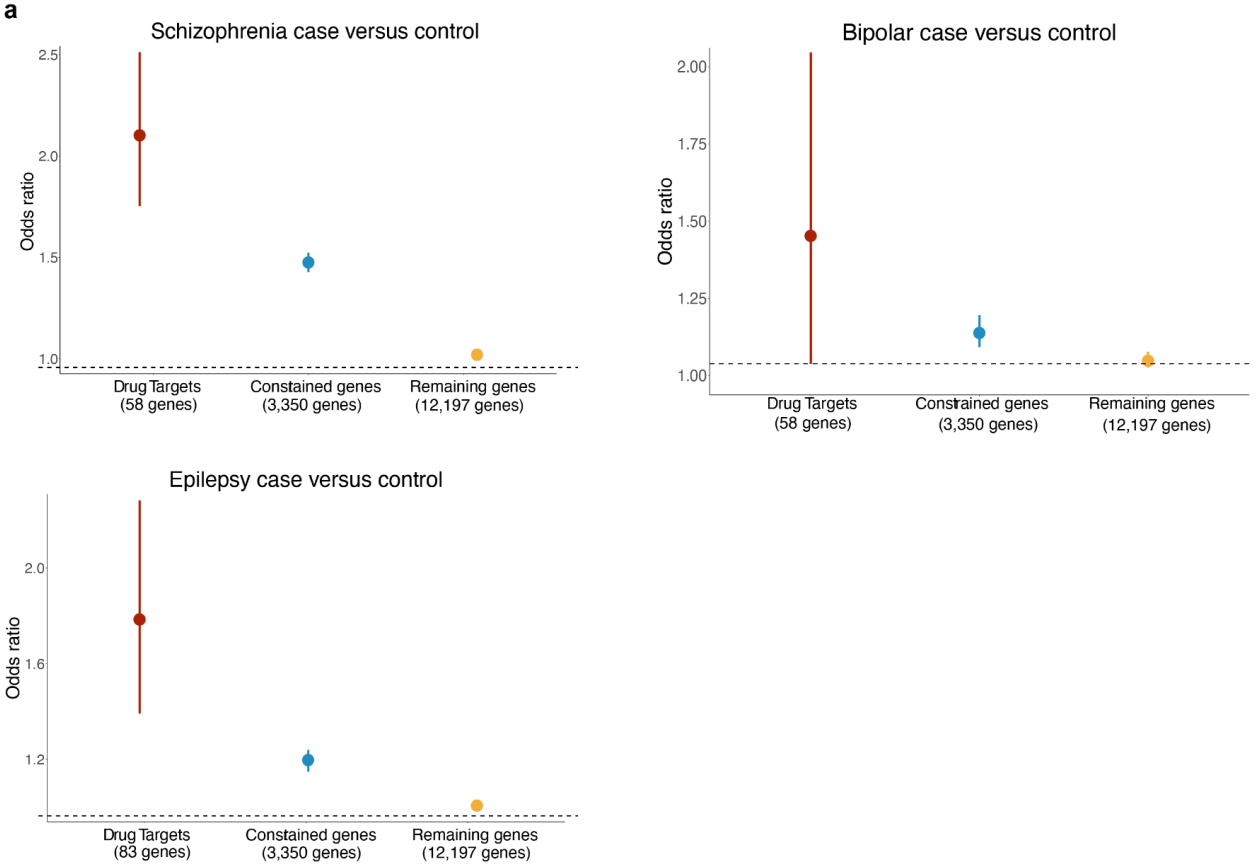


Supplementary Figure 1a-1c. UMAP embeddings of snRNA-seq profiles colored by cell type annotations. Amygdala (A) and frontal cortex (B), neuronal subtypes are included. **d,e**, Violin plots showing the distribution of the number of genes detected per nucleus across donor brainsfor each region and Violin plots showing the distribution of log-transformed UMI counts per nucleus, indicating consistent sequencing depth across donors and regions

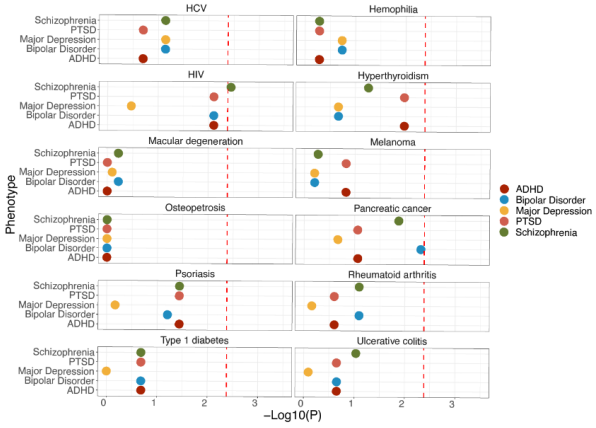


Supplementary Figure 2. For neuropsychiatric traits with available exome sequencing data, we demonstrate that PANs nominates a large number of genes across multiple exome sequencing studies. The x-axis represents the gene set size when ranked by top N scores (top 500, 1000, 1500 genes etc). For phenotypes with a large number of bonferroni significant genes (DD and ASD), we use the original thresholds in those studies. For other phenotypes (schizophrenia, bipolar disorder, major depression and epilepsy), we use a threshold P value of 0.01. For MAGMA analysis, we took all MAGMA genes with Z-score > 0 and corresponding P value < 0.05 for all protein coding genes. Shaded regions around the trend line represent 95% confidence intervals. We demonstrate fold-enrichment of PANs prioritized gene scores in exome sequencing data for neuropsychiatric traits at different p-value cutoffs (similar to **Figure 4-2a**). Error bars represent the 95% CI from a bootstrap of the PAN scores for significant versus non-significant genes in each phenotype (Supplementary Methods). P-value asterisks are from a wilcoxon rank sum test between significant and non-significant genes at each significant threshold from the sequencing data.

a

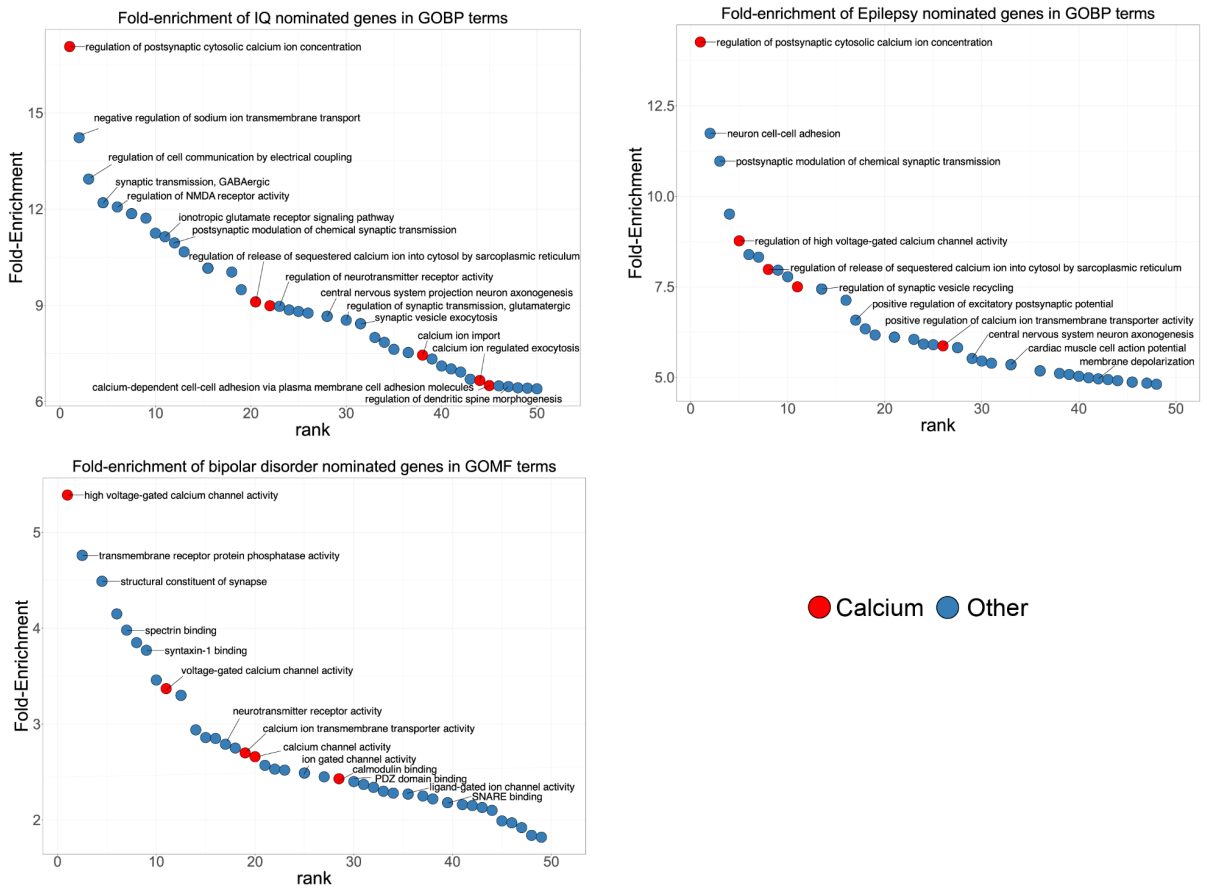

Schizophrenia case versus control
Odds ratio
2.5
2.0
1.5
1.0
Drug Targets (58 genes)
Constained genes (3,350 genes)
Remaining genes (12,197 genes)
Bipolar case versus control
Odds ratio
2.00
1.75
1.50
1.25
1.00
Drug Targets (58 genes)
Constrained genes (3,350 genes)
Remaining genes (12,197 genes)
Epilepsy case versus control
Odds ratio
2.0
1.6
1.2
Drug Targets (83 genes)
Constained genes (3,350 genes)
Remaining genes (12,197 genes)

b

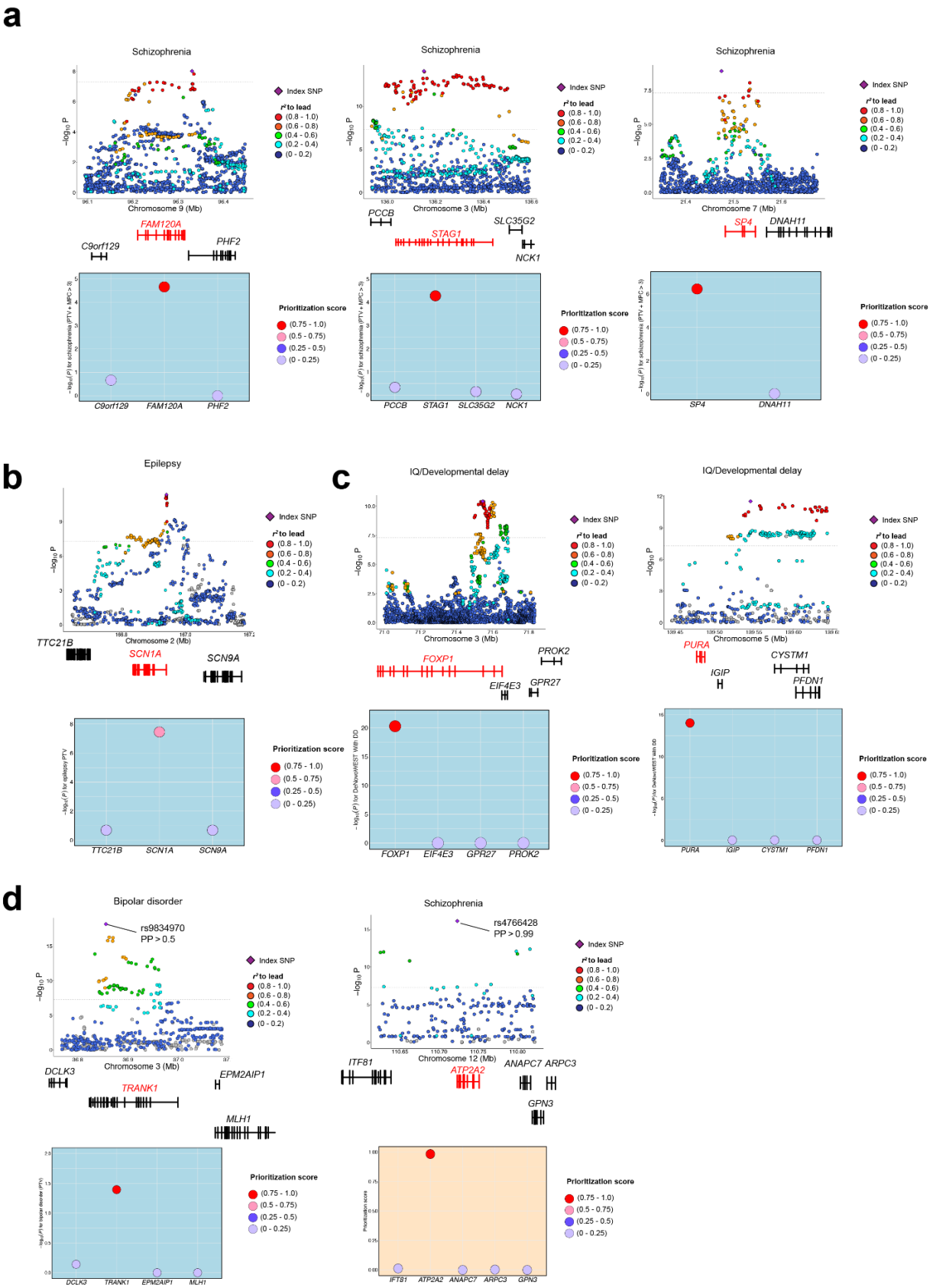

HCV
Hemophilia
HIV
Hyperthyroidism
Macular degeneration
Melanoma
Osteopetrosis
Pancreatic cancer
Psoriasis
Rheumatoid arthritis
Type 1 diabetes
Ulcerative colitis
Schizophrenia
PTSD
Major Depression
Bipolar Disorder
ADHD
Phenotype
0
1
2
3
−Log10(P)
ADHD
Bipolar Disorder
Major Depression
PTSD
Schizophrenia

Supplementary Figure 3. Reported $-\log_{10}(P)$ values reflect two-sided Fisher's exact tests comparing the burden of ultra-rare protein-truncating and damaging missense variants in cases versus controls for constrained drug target genes implicated in schizophrenia, and damaging missense variants only for bipolar disorder. Each dot represents the odds ratio (OR), and bars indicate 95% confidence intervals. Numerical results in **Supplementary Table 2**. **b,** We tested whether genes implicated in five psychiatric disorders (ADHD, bipolar disorder, major depression, PTSD, and schizophrenia) show enrichment for rare variant associations in a selection of the most powered non-psychiatric traits from the Open Targets Genetics portal. Each point represents the $-\log_{10}(P)$ value from gene-level burden testing for a psychiatric gene set against a given non-psychiatric trait. The red dashed line indicates nominal significance (P = 0.05). No systematic enrichment was observed.

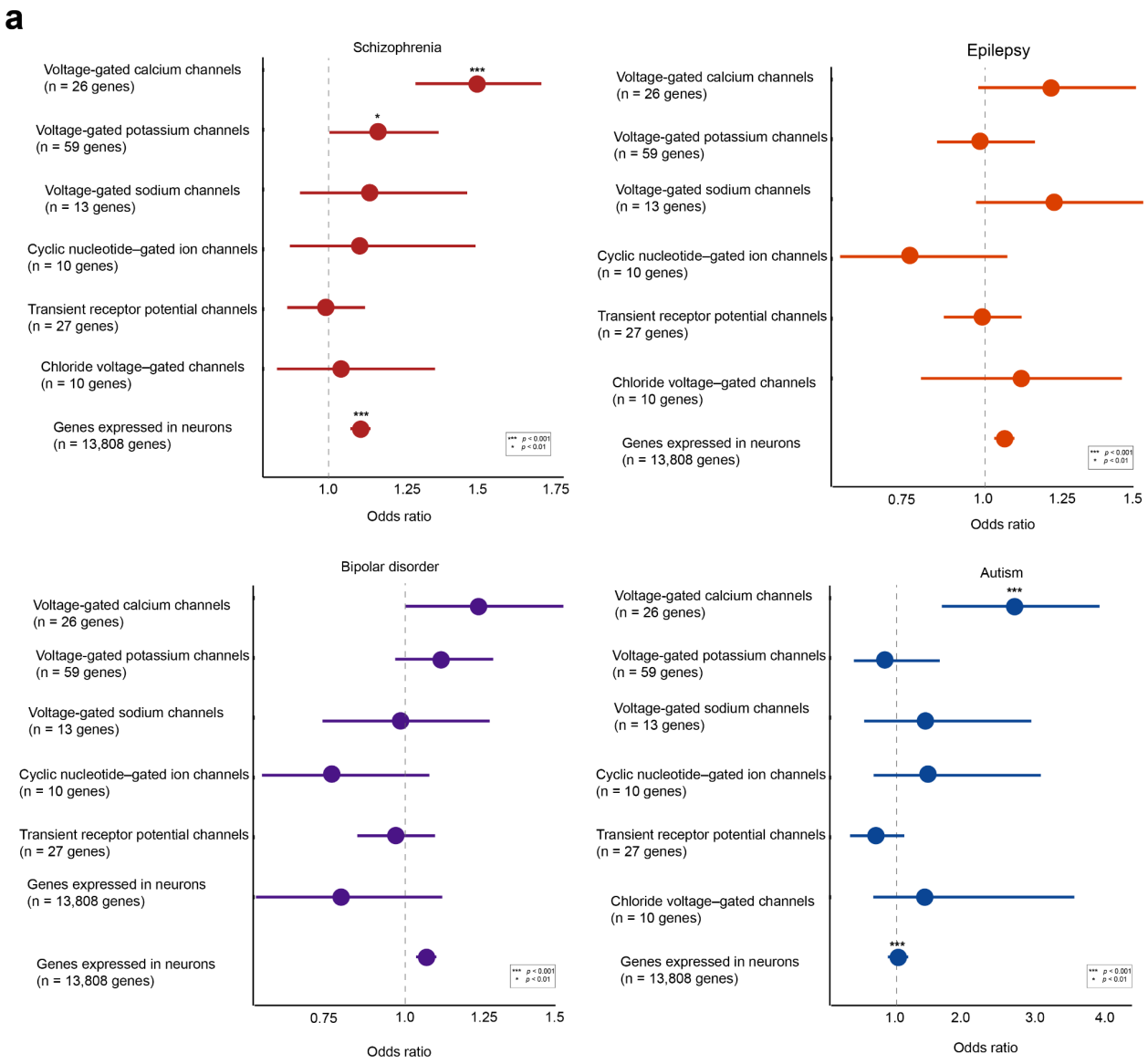
a



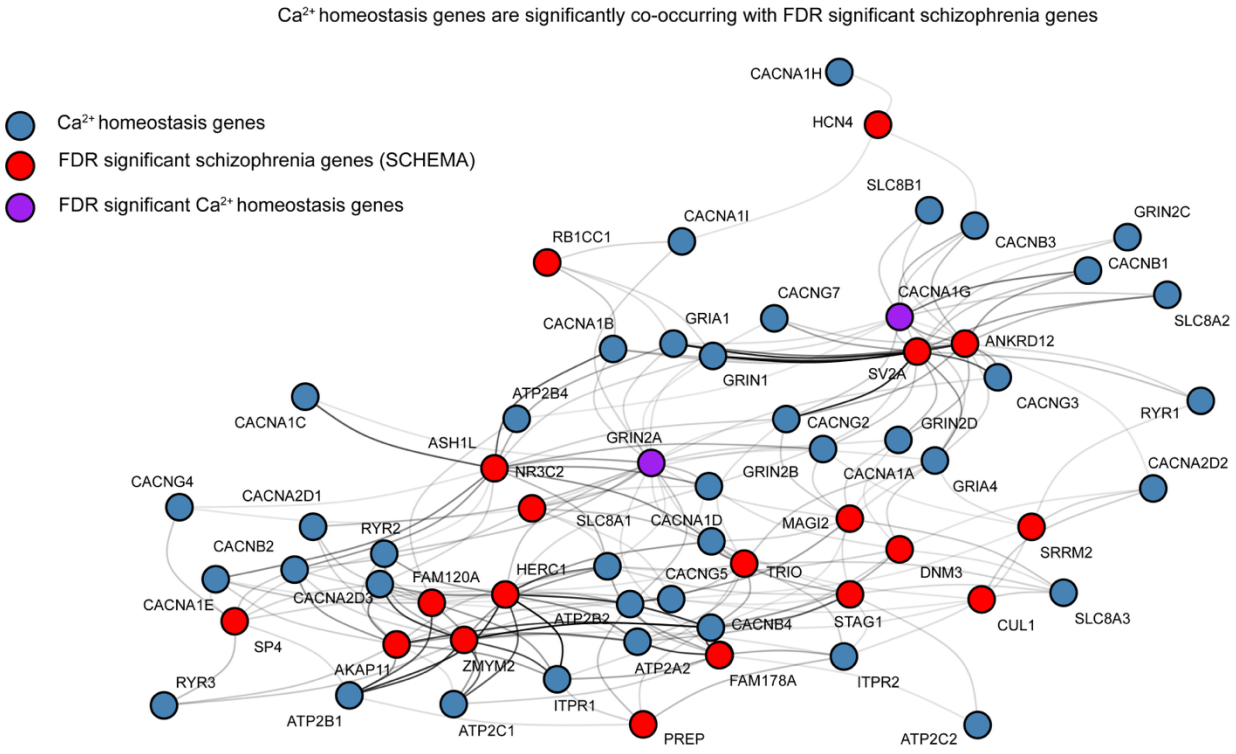
Supplementary Figure 4. Fold-enrichment of nominated genes in Gene Ontology terms in IQ, Bipolar Disorder, and Epilepsy (corresponding to **Figure 3a**). Terms related to Ca^{2+} signaling and homeostasis (red) are notably enriched. In all panels, terms not related to Ca^{2+} are shown in blue.


a
Schizophrenia
Chromosome 9 (Mb)
FAM120A
C9orf129
PHF2
Schizophrenia
Chromosome 3 (Mb)
PCCB
STAG1
SLC35G2
NCK1
Schizophrenia
Chromosome 7 (Mb)
SP4
DNAH11
Index SNP
r² to lead
(0.8 - 1.0)
(0.6 - 0.8)
(0.4 - 0.6)
(0.2 - 0.4)
(0 - 0.2)
Prioritization score
(0.75 - 1.0)
(0.5 - 0.75)
(0.25 - 0.5)
(0 - 0.25)
b
Epilepsy
Chromosome 2 (Mb)
TTC21B
SCN1A
SCN9A
c
IQ/Developmental delay
Chromosome 3 (Mb)
FOXP1
PROK2
EIF4E3
GPR27
IQ/Developmental delay
Chromosome 5 (Mb)
PURA
IGIP
CYSTM1
PFDN1
d
Bipolar disorder
rs9834970
PP > 0.5
Chromosome 3 (Mb)
DCLK3
TRANK1
EPM2AIP1
MLH1
Schizophrenia
rs4766428
PP > 0.99
Chromosome 12 (Mb)
IFT81
ATP2A2
ANAPC7
ARPC3
GPN3

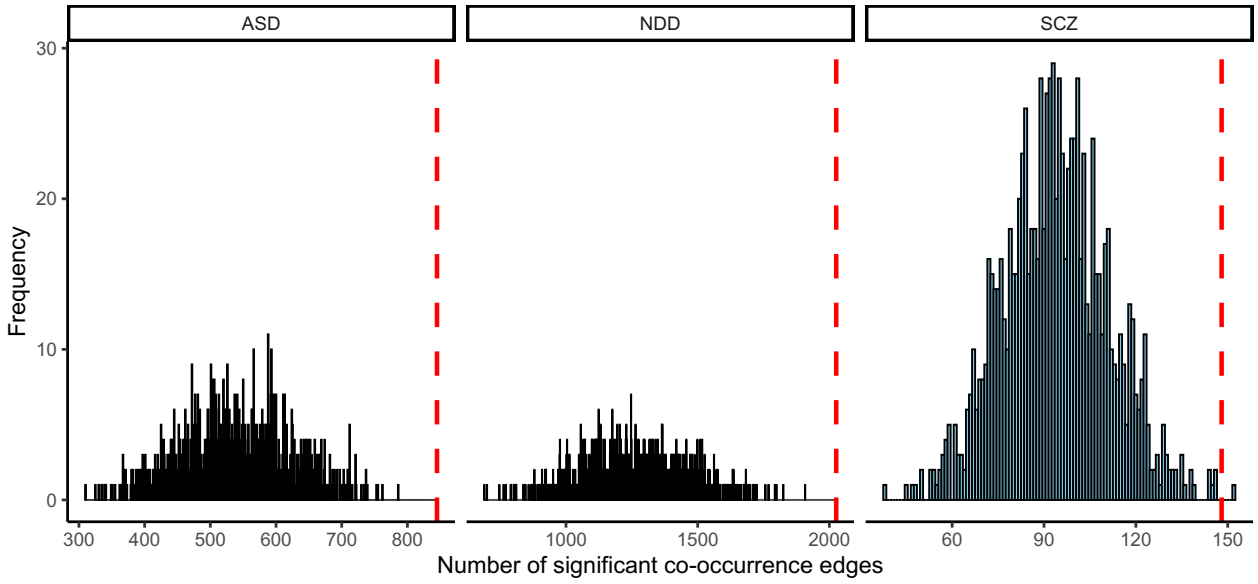
Supplementary Figure 5a-d. Top, LocusZoom plots for loci from across three GWAS with evidence in exome sequencing (A,C and D) or fine mapping (C) are displayed (GWAS from [1–3]). The two-sided P values of each SNP from the GWAS meta-analysis are shown along the y axis. Each SNP is colored by LD to the lead variant. Below, we show the along with the two-sided burden test meta-analysis P-values of ultra-rare variants from exome or de novo sequencing analysis (from [4–6]). For schizophrenia, FAM120A and STAG1 are FDR 5% significant, while SP4 is genome-wide significant. For ATP2A2, which the PANs prioritization score is shown. Color denotes prioritization strength by PANs.



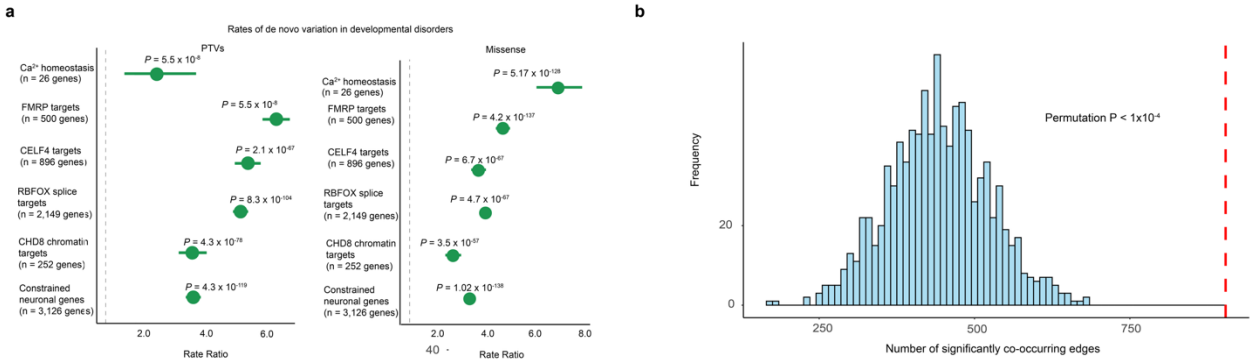
Supplementary Figure 6. Case–control enrichment and excess case rare-variant burden in Ca^{2+} , K^{+} and Na^{2+} voltage gated channel genes in schizophrenia, epilepsy, autism and bipolar disorder. Significant enrichment was observed in autism. The dot represents the OR and the bar represents the 95% CI of the point estimates. We used neuronal genes as a background, defined as genes that are expressed in at least 5% of any one neuronal cluster from single cell RNA-seq data. Gene lists obtained from the HUGO gene nomenclature committee.[7]



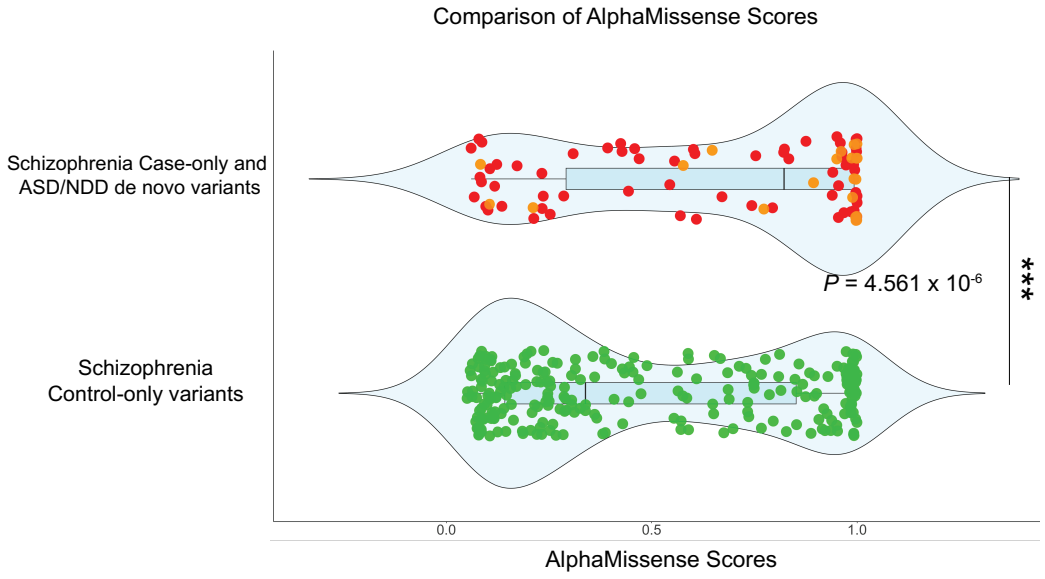
Supplementary Figure 7. Co-occurrence network between SCHEMA genes and Ca^{2+} homeostasis genes. Nodes in red represent FDR-significant schizophrenia risk genes. Edges represent gene pairs with significant co-occurrence (FDR < 0.001) across PANs-derived neuronal gene sets.



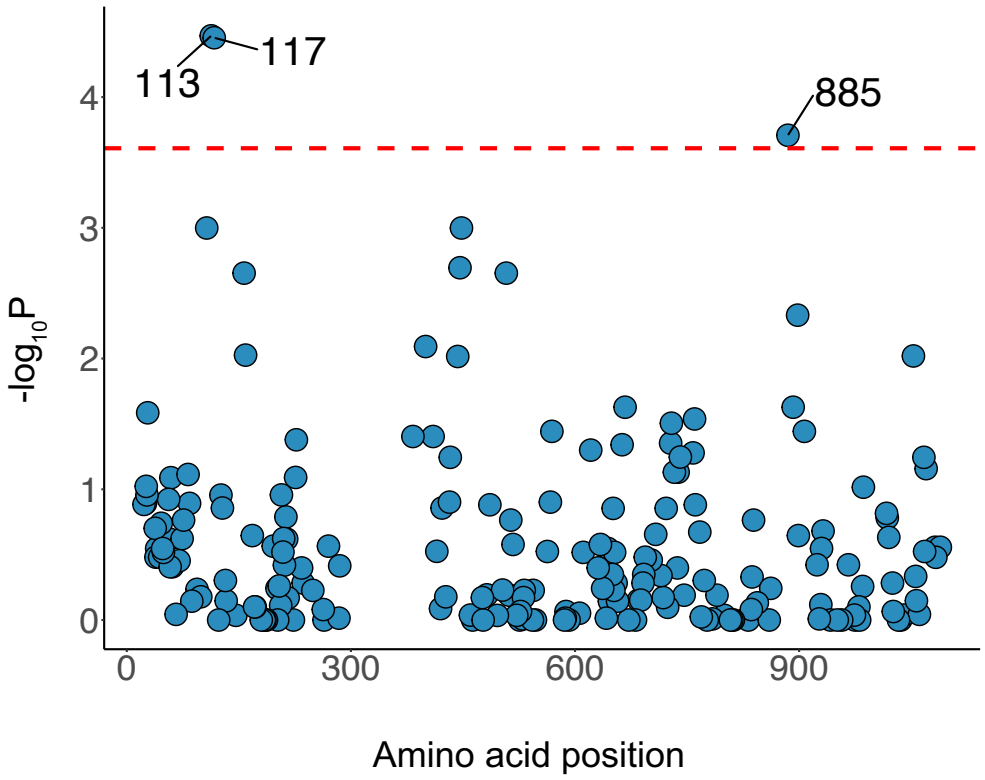
Supplementary Figure 8. **Significant Co-occurrence Between Calcium Homeostasis and Disease Genes.** Permutation-based co-occurrence testing between calcium homeostasis genes and disease gene sets reveals significant enrichment across neuropsychiatric phenotypes. Using heritability-informed PANs gene sets, we observed significantly more co-occurring gene pairs than expected by chance for all tested phenotypes: neurodevelopmental disorder (NDD; 2,026 significant pairs, $P < 0.001$), schizophrenia (SCZ; 148 pairs, $P = 0.001$), and autism spectrum disorder (ASD; 845 pairs, $P < 0.001$).



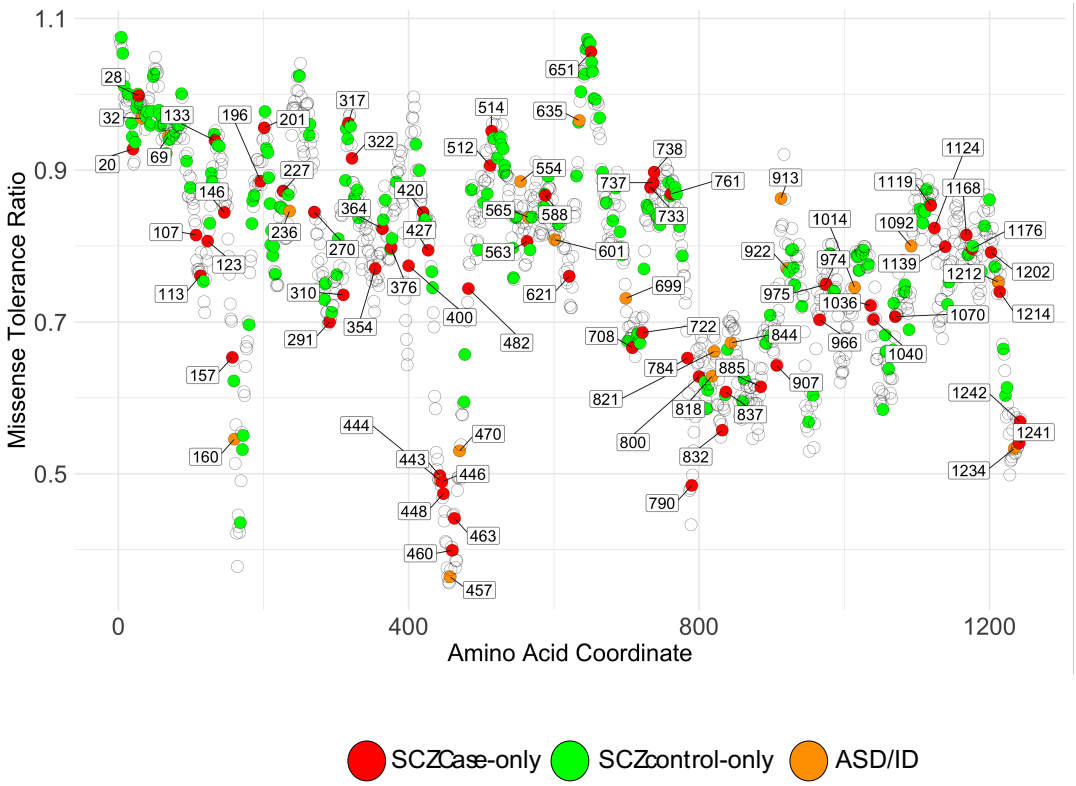
Supplementary Figure 9: a, Rates of de novo protein-truncating (left) and missense (right) variants in NDD. Ca^{2+} homeostasis genes show strong missense enrichment and significant PTV burden relative to other nervous system gene sets. **b,** Related to Supplementary Figure 7. , Histogram showing the null distribution of the number of significantly co-occurring edges (FDR < 0.001) between SCHEMA genes and randomly sampled background gene sets of equal size from PANs background (n = 1,000 permutations). The red dashed line indicates the observed number of significant co-occurrences between SCHEMA and Ca^{2+} homeostasis genes curated from the literature. No permutation exceeded the observed value ($P < 1\times10^{-4}$).



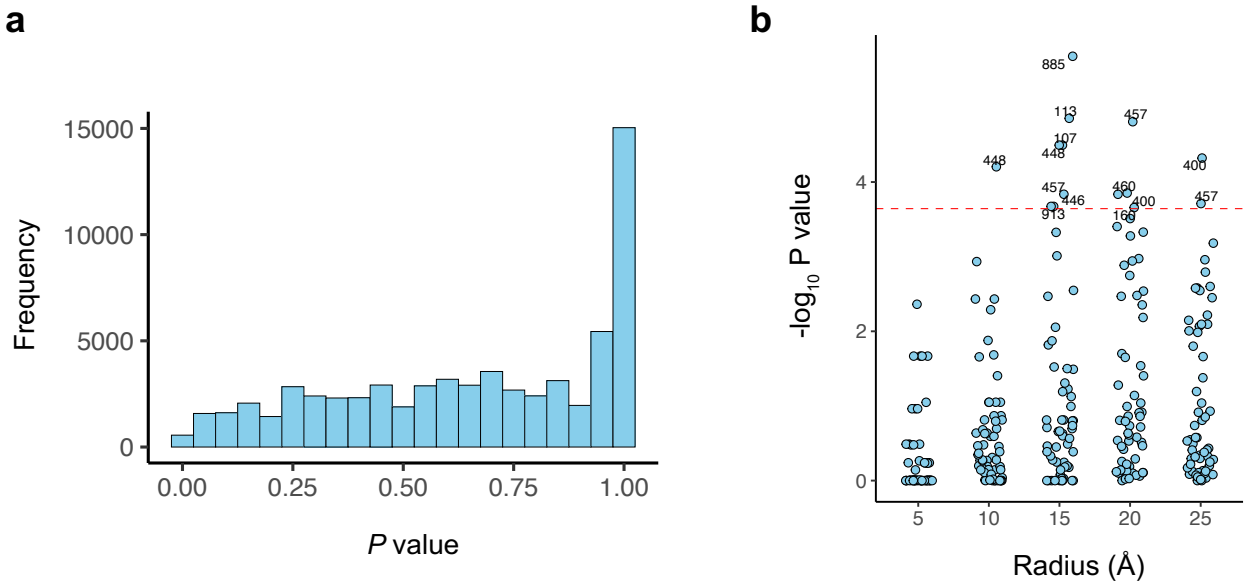
Supplementary Figure 10. Comparisons of AlphaMissense pathogenicity scores of the non-filtered case-only schizophrenia, autism and neurodevelopmental disorder associated variants to control-only variants. The P value is computed from a one-sided Wilcoxon test.



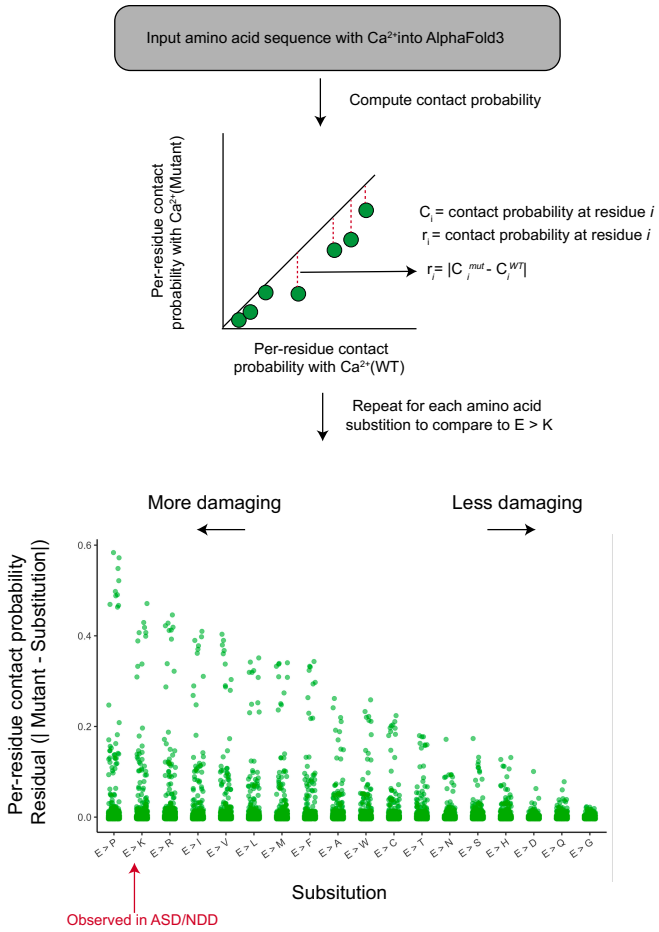
Supplementary Figure 11. Result of the schizophrenia only 3D neighborhood analysis (related to **Figure 4b**). All missense case-only and control-only variants from the schizophrenia exome (SCHEMA) were used. All three residues here are also significant in the schizophrenia, autism and neurodevelopmental disorder joint analysis in **Figure 4-2b.**



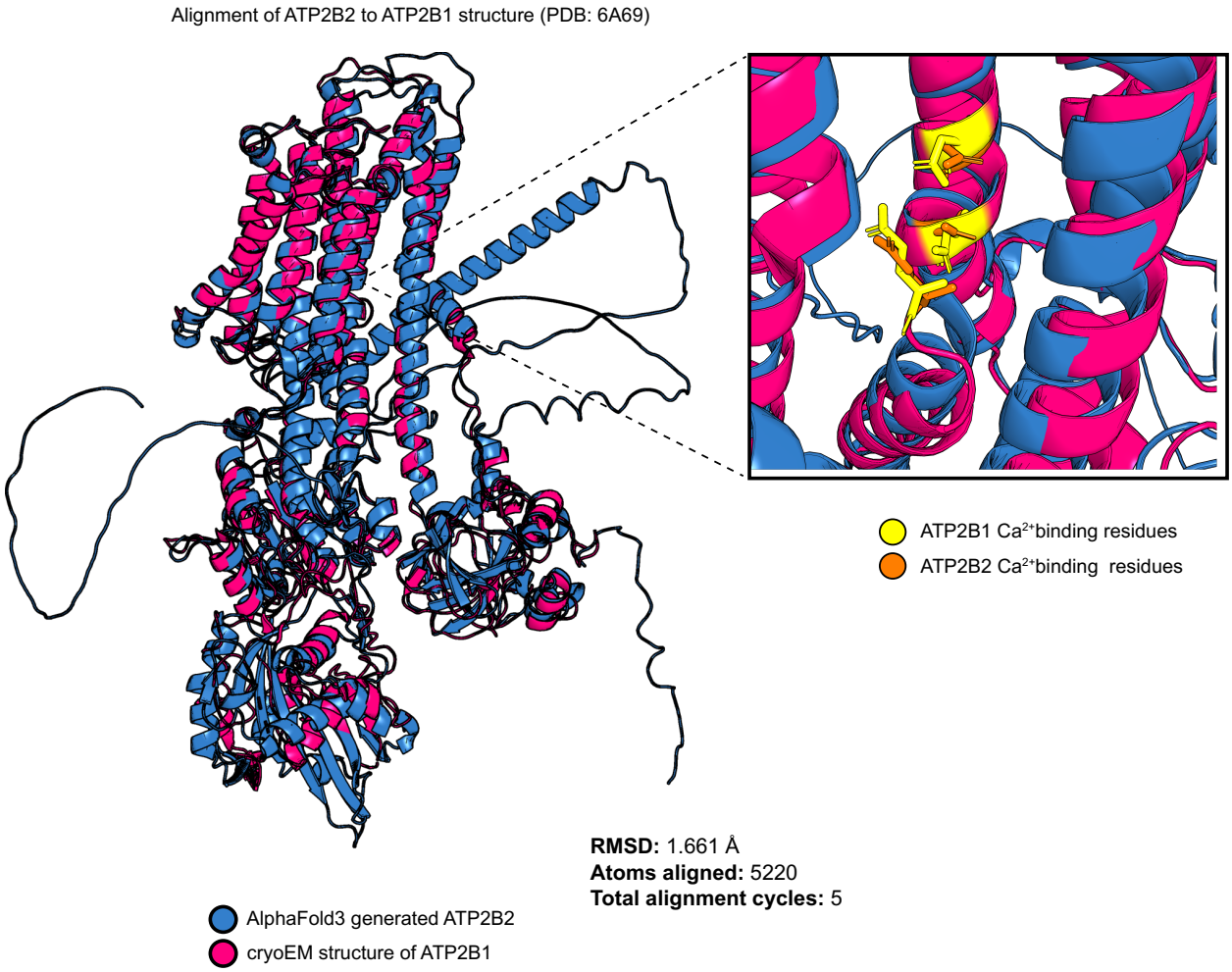
Supplementary Figure 12. Missense tolerance ratio (MTR)[1,2] track of ATP2B2 from the Regeneron Genetics Center Million Exome dataset [1]. Residues with highly damaging variants associated with schizophrenia case-only, ASD or NDD studies are colored. Glu457 is a residue located within one of the most regionally constrained regions of the gene. The data is taken from the Regeneron Genetics Center (RGC) browser.



Supplementary Figure 13, a, Histogram of permutation test results from shuffling the case-control labels 1000 times. The permutation P value is < 1 x 10^{-4}, indicating no single permutation produced a result as significant as P = 3.59 x 10^{-6}. A total of 6 out of 1000 permutations yielded residues reaching Bonferroni significance, though none matched the most significant residues shown in **Figure 4-2b**. **b,** Burden testing results across varying radii. A radius of 15 Å yielded the most significant p-values.

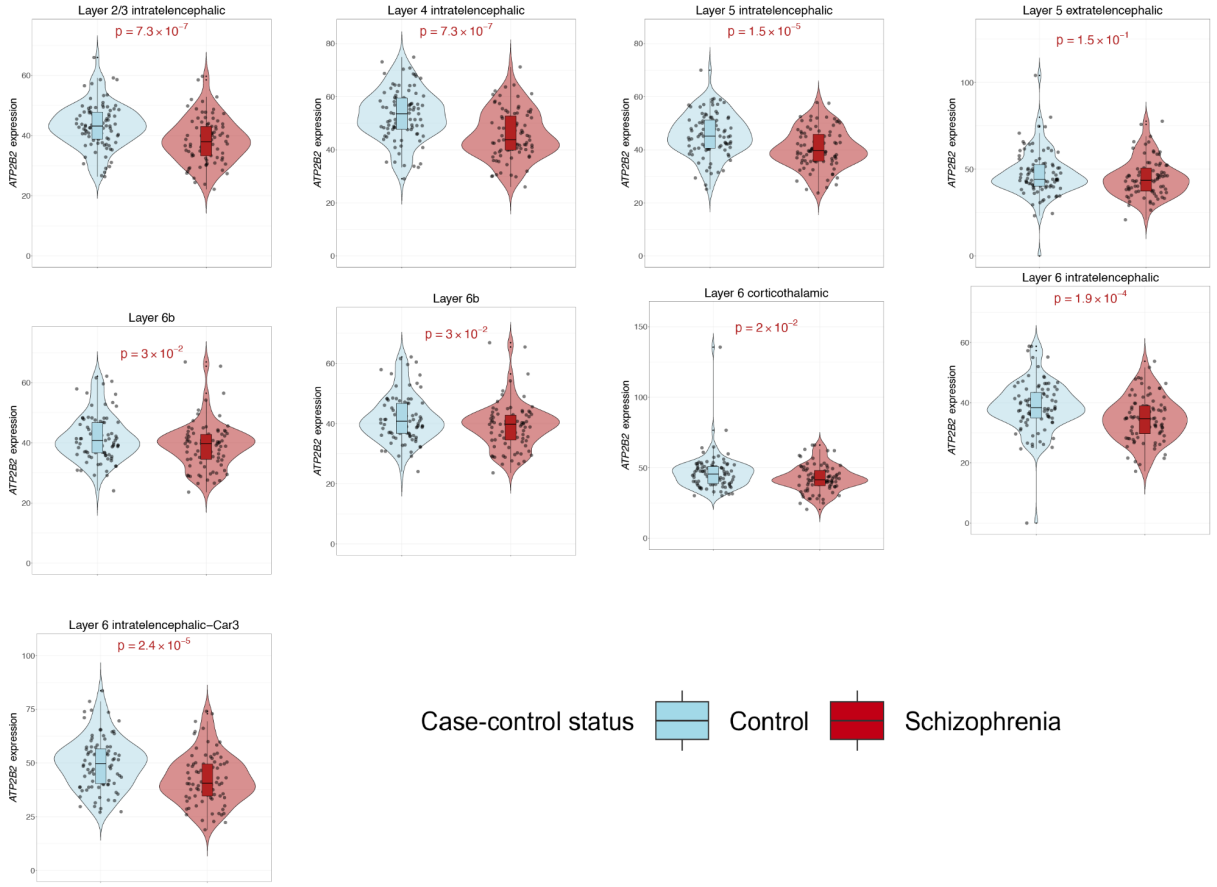


Supplementary Figure 14. Top. Workflow illustrating how AlphaFold3 was used to model mutant structures of ATP2B2 with bound Ca^{2+}, compute per-residue Ca^{2+} contact probabilities, and quantify disruption relative to wild type. Residual contact disruption was calculated for each residue. Bottom: Distribution of per-residue residuals for all substitutions at glutamic acid 457 (E457), sorted by the maximum disruption observed per mutation. Mutations are ordered from most to least structurally disruptive to Ca^{2+} contacts. Substitutions that differ substantially from glutamic acid in charge, size, or hydropathy - such as proline and arginine - were among the most disruptive. In contrast, substitutions to chemically similar amino acids, such as glycine, glutamine, and asparagine, caused minimal disruption. The E>K mutation, which was seen in probands with autism and neurodevelopmental disorders is among the most disruptive.



Supplementary Figure 15: **Structural alignment of ATP2B2 (orange) and ATP2B1 (grey, PDB: 6A69).** Left: ATP2B2 was aligned to ATP2B1 using PyMOL, yielding an RMSD of 1.661Å across the full structure. Right: Alignment of both structures at the calcium binding sites., we pinpointed the putative calcium-binding site in the predicted model of ATP2B2.

a.



Supplementary Figure 16. ATP2B2 expression (per 10^5 detected nuclear transcripts) comparison in persons with and without schizophrenia (n = 93 controls, 87 cases) across glutamatergic neuron subtypes of distinct cortical layers. Sequencing data obtained from[84] . Shown is ATP2B2 expression (per 10^5 detected nuclear transcripts) in each cell type in individual donors. P-values are computed by a logistic regression of normalized expression while controlling for PMI, age and sex. Box plots display interquartile ranges, whiskers extend to 1.5 times the interquartile range, central lines represent medians, a

Supplementary Discussion

Extended discussion of Ca^{2+} homeostasis gene selection

In defining our core set of "Ca^{2+} homeostasis genes" we intentionally focused on genes encoding calcium-selective channels excluding other elements of the neuronal Ca^{2+}-signaling toolkit. Notably, we omitted N-methyl-D-aspartate-type receptors (NMDARs) and AMPA- and kainate-type glutamate receptors. While these receptors play a crucial role in transmitting calcium signals to the nucleus, triggering action potentials, and leading to plasma membrane depolarization and have roles in excitation-transcription coupling (reviewed in [82]). However, these receptors are non-selective channels, and have already been implicated in previous neuropsychiatric studies.

Extended Discussion: Simulating Mutation Effects on Per-Residue Contact Probabilities with AlphaFold3

AlphaFold3 can model protein–ligand interactions, including metal ions like calcium, with improved accuracy by integrating physical and structural constraints into its deep learning framework. When modeling calcium-binding proteins, AlphaFold3 provides residue-level contact probabilities, reflecting the likelihood that specific amino acids are spatially coordinated with calcium ions. The Critical Assessment of Protein Structure Prediction (CASP) defines two nonadjacent residues to be in contact if the distance between their Cβ atoms (or Cα for glycine) is less than 8 Å in the folded structure[106]. These predictions provide a quantitative basis for identifying calcium-coordinating residues, and allows us to simulate the effects of substituting individual residues to gauge the change in per-residue contact probabilities. To further explore the structural and functional consequences of the E457K mutation, we revisited the underlying biophysical principles that might explain its disruptive effects. As previously noted, this mutation (Figure 6a) induces a marked reduction in per-residue contact probabilities, which could be attributed to the substitution of a negatively charged glutamic acid with a positively charged lysine. To systematically investigate this hypothesis, we systematically mutated a glutamic acid (E) residue in the ATP2B2 protein to every other amino acid and measured the per-residue change in calcium contact probabilities (compared to wild type). This was quantified as the difference (residual) in contact

probability (not, all. A larger residual reflected a larger perturbation between calcium ions and contact sites in ATP2B2.

As expected, mutations to Proline, Arginine, and bulky hydrophobic residues (e.g., Isoleucine, Valine, Methionine) caused the largest disruptions in calcium contact probabilities (Figure S12). These likely disturb local protein structure or calcium-binding sites due to charge reversal or steric clash. Reassuringly, mutations to chemically similar or small residues (e.g., Aspartic acid, Glutamine, Glycine) showed minimal disruption, suggesting partial retention of the wild-type interaction. These conservative substitutions produced minimal changes in predicted structure and calcium-binding behavior, suggesting that AF3 reasonably captures the structural impact of substitutions based on core biochemical properties. Taken together, these results suggest that substitutions that alter charge, polarity, or hydrophobicity tend to cause the most pronounced structural disruptions, whereas substitutions that retain the fundamental properties of glutamic acid result in comparatively minor perturbations.

Extended discussion of limitations

PANs uses S-LDSC to leverage underlying cell types and the power of GWAS to detect the genetic variation that modulates cellular processes in the brain. Thus, its first limitation is in the tissues and vast cell types and states represented. Second, PANs method relies on identifying highly expressed gene programs, which may not reflect biological causality if a causal cell type is not assayed, PANs may identify co-expressed cell types and genes as causal. However, these non-causal genes might be informative for drug development purposes by nominating a causal pathway. Third, PANs does not utilize cell type specific eQTL data [107–109], which may reflect the underlying biology more appropriately. Fourth, PANs does not distinguish whether two gene processes implicated in neuropsychiatric disorders are conditionally independent signals.

Fifth, the LD score regression framework is primarily applicable to common and low-frequency variants, and less applicable to rare variant enrichments (we instead use rare variants to validate our findings). Sixth, we have focused on human scRNA- seq data; however, incorporating data from other modalities such as ATAC-seq (Assay for Transposase-Accessible Chromatin using sequencing, [110]) and DNase-seq (DNase I hypersensitive sites sequencing [111]) can more specifically map candidate cis-regulatory sequences. In turn, relevant animal models could allow experimental validation of neuropsychiatric disorders mechanisms in model organisms [112]. Seventh, we identify programs by gene co- variation using only healthy tissues. Future work could utilize disorder relevant tissue to nominate more precise sets of genes and the cell types in which they are active [6,113,114]. Eighth, PANs

mainly detects genes with high levels of expression, and thus genes with low expression in neurons would not be detected by our framework [115,116]. Finally, our analysis relies on sequencing data for validation, which remains underpowered to implicate most genes.

Supplementary methods

Gene Enrichment

For each trait with a GWAS and a corresponding exome-sequencing data, we define enrichment as the mean PANs scores for the nominated genes passing a significance threshold, divided by the scores of genes below that threshold. Given the heterogeneity in power between different neuropsychiatric traits (e.g developmental disorders versus bipolar disorders), we maximized power by taking different thresholds. In the case of well powered sequencing data, we used FDR significant genes for Autism Spectrum Disorder and neurodevelopmental disorders, as per the original publications (FDR < 0.001 for ASD[19] leading to 72 genes, 1.3e-06 for NDD[18]). For schizophrenia, bipolar disorder and epilepsy, we used a threshold of $p < 0.01$.

Bootstrapping

Resampling methods were employed to evaluate the robustness of PANs scoring metrics for genes implicated in neuropsychiatric phenotypes (Figure 2a). Specifically, 10,000 bootstrap resamples were generated to estimate the distribution of PANs scores within two distinct gene sets: those designated as significant and those deemed not significant from the corresponding exome-sequencing study. For each resample iteration, genes from the significant group and the not-significant group were sampled with replacement to compute the average PANs Score. This process allowed for the construction of two empirical distributions of mean scores, from which 95% confidence intervals were derived. The lower and upper bounds of these intervals were then calculated for both the significant and not-significant gene sets. The final comparative metric was the ratio of the lower bound of the not-significant group to the upper bound of the significant group and vice versa, providing a measure of overlap and distinction between the two gene sets' PANs scores.

SNAP analysis

To identify all genes with relationships to SNAP in glutamatergic neurons ($SNAP_{glut}$), we utilized the matrix of gene- and donor-level loadings onto the SNAP-neuron component (SNAP-n) provided in [102] . We performed a genome-wide regression analysis in which the scaled expression of each gene was regressed against the SNAP-n donor loading, adjusting for relevant biological and technical covariates including age, postmortem interval (PMI), and sex.

$$\text{Scaled Expression} \sim \beta_0 + \beta_1 * SNAP_n + \beta_2 * Age + \beta_3 * PMI + \beta_4 * Sex + \epsilon$$

This approach enabled the identification of genes whose expression patterns covary with the SNAP-n axis across donors, thereby implicating them in the broader regulatory architecture of the SNAP program within glutamatergic neuronal populations.

To mitigate bias introduced by genes with high absolute expression levels but limited inter-individual variability, we employed scaled expression values in our regressions. This normalization ensures that genes with extreme expression magnitudes do not disproportionately influence regression coefficients solely due to scale. This was applied independently to both glutamatergic and GABAergic neuronal populations. We subsequently restricted our analysis to genes exhibiting statistically significant associations with SNAP-n donor loading in glutamatergic neurons, while remaining non-significant in GABAergic neurons, thereby isolating genes with cell type-specific relationships in glutamatergic neurons to the SNAP transcriptional program.

References

1. Hyman, S. E. Revitalizing psychiatric therapeutics. *Neuropsychopharmacology* **39**, 220–229 (2014).

2. Kesselheim, A. S., Hwang, T. J. & Franklin, J. M. Two decades of new drug development for central nervous system disorders. *Nat. Rev. Drug Discov.* **14**, 815–816 (2015).

3. Miller, A. H. & Raison, C. L. Burning down the house: reinventing drug discovery in psychiatry for the development of targeted therapies. *Mol. Psychiatry* **28**, 68–75 (2022).

4. Willsey, H. R., Willsey, A. J., Wang, B. & State, Matthew W. Genomics, convergent neuroscience and progress in understanding autism spectrum disorder. *Nat. Rev. Neurosci.* **23**, 323–341 (2022).

5. Skene, N. G. *et al.* Genetic identification of brain cell types underlying schizophrenia. *Nat. Genet.*

50, 825–833 (2018).

6. Jagadeesh, K. A. *et al.* Identifying disease-critical cell types and cellular processes by integrating single-cell RNA-sequencing and human genetics. *Nat. Genet.* **54**, 1479–1492 (2022).

7. Finucane, H. K. *et al.* Heritability enrichment of specifically expressed genes identifies disease-relevant tissues and cell types. *Nat. Genet.* **50**, 621–629 (2018).

8. Zhang, M. J. *et al.* Polygenic enrichment distinguishes disease associations of individual cells in single-cell RNA-seq data. *Nat. Genet.* **54**, 1572–1580 (2022).

9. Macosko, E. Z. *et al.* Highly Parallel Genome-wide Expression Profiling of Individual Cells Using Nanoliter Droplets. *Cell* **161**, 1202–1214 (2015).

10. Bryois, J. *et al.* Genetic identification of cell types underlying brain complex traits yields insights into the etiology of Parkinson's disease. *Nat. Genet.* **52**, 482–493 (2020).

11. Emont, M. P. *et al.* A single-cell atlas of human and mouse white adipose tissue. *Nature* **603**, 926–933 (2022).

12. Abramson, J. *et al.* Accurate structure prediction of biomolecular interactions with AlphaFold 3. *Nature* **630**, 493–500 (2024).

13. Finucane, H. K. *et al.* Partitioning heritability by functional annotation using genome-wide association summary statistics. *Nat. Genet.* **47**, 1228–1235 (2015).

14. Abul-Husn, N. S. *et al.* A Protein-Truncating HSD17B13 Variant and Protection from Chronic Liver Disease. *N. Engl. J. Med.* **378**, 1096–1106 (2018).

15. Cohen, J. C., Boerwinkle, E., Mosley, T. H., Jr & Hobbs, H. H. Sequence variations in PCSK9, low LDL, and protection against coronary heart disease. *N. Engl. J. Med.* **354**, 1264–1272 (2006).

16. Palmer, D. S. *et al.* Exome sequencing in bipolar disorder identifies AKAP11 as a risk gene shared with schizophrenia. *Nat. Genet.* **54**, 541–547 (2022).

17. Singh, T. *et al.* Rare coding variants in ten genes confer substantial risk for schizophrenia. *Nature* **604**, 509–516 (2022).

18. Kaplanis, J. *et al.* Evidence for 28 genetic disorders discovered by combining healthcare and

research data. *Nature* **586**, 757–762 (2020).

19. Fu, J. M. *et al.* Rare coding variation provides insight into the genetic architecture and phenotypic context of autism. *Nat. Genet.* **54**, 1320–1331 (2022).

20. Epi25 Collaborative, Chen, S., Neale, B. M. & Berkovic, S. F. Shared and distinct ultra-rare genetic risk for diverse epilepsies: A whole-exome sequencing study of 54,423 individuals across multiple genetic ancestries. *medRxiv* (2023) doi:10.1101/2023.02.22.23286310.

21. Yun, S. *et al.* Antipsychotic drug efficacy correlates with the modulation of D1 rather than D2 receptor-expressing striatal projection neurons. *Nat. Neurosci.* **26**, 1417–1428 (2023).

22. Carlsson, A. & Lindqvist, M. Effect of chlorpromazine or haloperidol on formation of 3methoxytyramine and normetanephrine in mouse brain. *Acta Pharmacol. Toxicol.* **20**, 140–144 (1963).

23. Creese, I., Burt, D. R. & Snyder, S. H. Dopamine Receptor Binding Predicts Clinical and Pharmacological Potencies of Antischizophrenic Drugs. *Science* **192**, 481–483 (1976).

24. Hyman, S. E. Psychiatric drug development: diagnosing a crisis. *Cerebrum* **2013**, 5 (2013).

25. Cade, J. F. J. Lithium salts in the treatment of psychotic excitement. *Med. J. Aust.* **2**, 349–352 (1949).

26. Ramachandraiah, C. T., Subramaniam, N. & Tancer, M. The story of antipsychotics: Past and present. *Indian J. Psychiatry* **51**, 324–326 (2009).

27. Ochoa, D. *et al.* Open Targets Platform: supporting systematic drug–target identification and prioritisation. *Nucleic Acids Res.* **49**, D1302–D1310 (2020).

28. Greer, P. L. & Greenberg, M. E. From Synapse to Nucleus: Calcium-Dependent Gene Transcription in the Control of Synapse Development and Function. *Neuron* **59**, 846–860 (2008).

29. West, A. E. & Greenberg, M. E. Neuronal activity-regulated gene transcription in synapse development and cognitive function. *Cold Spring Harb. Perspect. Biol.* **3**, (2011).

30. Aleksander, S. A. *et al.* The Gene Ontology knowledgebase in 2023. *Genetics* **224**, iyad031 (2023).

31. Thomas, P. D. *et al.* PANTHER: Making genome-scale phylogenetics accessible to all. *Protein*

Sci. **31**, 8–22 (2022).

32. Trubetskoy, V. *et al.* Mapping genomic loci implicates genes and synaptic biology in schizophrenia. *Nature* **604**, 502–508 (2022).

33. Mullins, N. *et al.* Genome-wide association study of more than 40,000 bipolar disorder cases provides new insights into the underlying biology. *Nat. Genet.* **53**, 817–829 (2021).

34. Howard, D. M. *et al.* Genome-wide meta-analysis of depression identifies 102 independent variants and highlights the importance of the prefrontal brain regions. *Nat. Neurosci.* **22**, 343–352 (2019).

35. Watanabe, K. *et al.* A global overview of pleiotropy and genetic architecture in complex traits. *Nat. Genet.* **51**, 1339–1348 (2019).

36. Bulik-Sullivan, B. *et al.* An atlas of genetic correlations across human diseases and traits. *Nat. Genet.* **47**, 1236–1241 (2015).

37. Grotzinger, A. D. *et al.* Genetic architecture of 11 major psychiatric disorders at biobehavioral, functional genomic and molecular genetic levels of analysis. *Nat. Genet.* **54**, 548–559 (2022).

38. Werme, J., van der Sluis, S., Posthuma, D. & de Leeuw, C. A. An integrated framework for local genetic correlation analysis. *Nat. Genet.* **54**, 274–282 (2022).

39. Consortium, T. B. *et al.* Analysis of shared heritability in common disorders of the brain. *Science* **360**, eaap8757 (2018).

40. Cross-Disorder Group of the Psychiatric Genomics Consortium. Electronic address: plee0@mgh.harvard.edu & Cross-Disorder Group of the Psychiatric Genomics Consortium. Genomic Relationships, Novel Loci, and Pleiotropic Mechanisms across Eight Psychiatric Disorders. *Cell* **179**, 1469–1482.e11 (2019).

41. Wisedchaisri, G. & Gamal El-Din, T. M. Druggability of Voltage-Gated Sodium Channels-Exploring Old and New Drug Receptor Sites. *Front. Pharmacol.* **13**, 858348 (2022).

42. Kaczorowski, G. J., McManus, O. B., Priest, B. T. & Garcia, M. L. Ion channels as drug targets: the next GPCRs. *J. Gen. Physiol.* **131**, 399–405 (2008).

43. Purcell, S. M. *et al.* A polygenic burden of rare disruptive mutations in schizophrenia. *Nature* **506**, 185–190 (2014).

44. Dedic, N. *et al.* Cross-disorder risk gene CACNA1C differentially modulates susceptibility to psychiatric disorders during development and adulthood. *Mol. Psychiatry* **23**, 533–543 (2018).

45. Seal, R. L. *et al.* Genenames.org: the HGNC resources in 2023. *Nucleic Acids Res.* **51**, D1003–D1009 (2023).

46. Brini, M. & Carafoli, E. The Plasma Membrane Ca^{2+} ATPase and the Plasma Membrane Sodium Calcium Exchanger Cooperate in the Regulation of Cell Calcium. *Cold Spring Harb. Perspect. Biol.* **3**, (2011).

47. Walker, D. *et al.* A New β Subtype-specific Interaction in α1ASubunit Controls P/Q-type Ca^{2+} Channel Activation *. *J. Biol. Chem.* **274**, 12383–12390 (1999).

48. Clapham, D. E. Calcium signaling. *Cell* **131**, 1047–1058 (2007).

49. Bagur, R. & Hajnóczky, G. Intracellular Ca^{2+} sensing: Its role in calcium homeostasis and signaling. *Mol. Cell* **66**, 780–788 (2017).

50. Berridge, M. J., Lipp, P. & Bootman, M. D. The versatility and universality of calcium signalling. *Nat. Rev. Mol. Cell Biol.* **1**, 11–21 (2000).

51. De Stefani, D., Raffaello, A., Teardo, E., Szabò, I. & Rizzuto, R. A forty-kilodalton protein of the inner membrane is the mitochondrial calcium uniporter. *Nature* **476**, 336–340 (2011).

52. Cooper, D. & Dimri, M. Biochemistry, Calcium Channels. in *StatPearls* (StatPearls Publishing, Treasure Island (FL), 2023).

53. Chen, J., Sitsel, A., Benoy, V., Sepúlveda, M. R. & Vangheluwe, P. Primary Active Ca^{2+} Transport Systems in Health and Disease. *Cold Spring Harb. Perspect. Biol.* **12**, (2020).

54. Darnell, J. C. *et al.* FMRP stalls ribosomal translocation on mRNAs linked to synaptic function and autism. *Cell* **146**, 247–261 (2011).

55. Sugathan, A. *et al.* CHD8 regulates neurodevelopmental pathways associated with autism spectrum disorder in neural progenitors. *Proc. Natl. Acad. Sci. U. S. A.* **111**, E4468–77 (2014).

56. Fogel, B. L. *et al.* RBFOX1 regulates both splicing and transcriptional networks in human neuronal development. *Hum. Mol. Genet.* **21**, 4171–4186 (2012).

57. Vandecaetsbeek, I., Vangheluwe, P., Raeymaekers, L., Wuytack, F. & Vanoevelen, J. The Ca^{2+} pumps of the endoplasmic reticulum and Golgi apparatus. *Cold Spring Harb. Perspect. Biol.* **3**, (2011).

58. Strehler, E. E. & Thayer, S. A. Evidence for a role of plasma membrane calcium pumps in neurodegenerative disease: Recent developments. *Neurosci. Lett.* **663**, 39–47 (2018).

59. Rahimi, M. J. *et al.* De novo variants in ATP2B1 lead to neurodevelopmental delay. *Am. J. Hum. Genet.* **109**, 944–952 (2022).

60. Smits, J. J. *et al.* De novo and inherited loss-of-function variants of ATP2B2 are associated with rapidly progressive hearing impairment. *Hum. Genet.* **138**, 61–72 (2019).

61. Ling, E. *et al.* A concerted neuron–astrocyte program declines in ageing and schizophrenia. *Nature* **627**, 604–611 (2024).

62. Aryal, S. *et al.* Deep proteomics identifies shared molecular pathway alterations in synapses of patients with schizophrenia and bipolar disorder and mouse model. *Cell Rep.* **42**, 112497 (2023).

63. Frazer, J. *et al.* Disease variant prediction with deep generative models of evolutionary data. *Nature* **599**, 91–95 (2021).

64. Rehm, H. L. & Fowler, D. M. Keeping up with the genomes: scaling genomic variant interpretation. *Genome Med.* **12**, 5 (2019).

65. Richards, S. *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).

66. Miosge, L. A. *et al.* Comparison of predicted and actual consequences of missense mutations. *Proc. Natl. Acad. Sci. U. S. A.* **112**, E5189–98 (2015).

67. Brandes, N., Goldman, G., Wang, C. H., Ye, C. J. & Ntranos, V. Genome-wide prediction of disease variant effects with a deep protein language model. *Nat. Genet.* **55**, 1512–1522 (2023).

68. Cheng, J. *et al.* Accurate proteome-wide missense variant effect prediction with AlphaMissense.

Science **381**, eadg7492 (2023).

69. Karczewski, K. J. *et al.* The mutational constraint spectrum quantified from variation in 141,456 humans. *Nature* **581**, 434–443 (2020).

70. Gong, D. *et al.* Structure of the human plasma membrane Ca^{2+}-ATPase 1 in complex with its obligatory subunit neuroplastin. *Nat. Commun.* **9**, 1–9 (2018).

71. James, P., Inui, M., Tada, M., Chiesi, M. & Carafoli, E. Nature and site of phospholamban regulation of the Ca^{2+} pump of sarcoplasmic reticulum. *Nature* **342**, 90–92 (1989).

72. Toyoshima, C., Nakasako, M., Nomura, H. & Ogawa, H. Crystal structure of the calcium pump of sarcoplasmic reticulum at 2.6 Å resolution. *Nature* **405**, 647–655 (2000).

73. Kühlbrandt, W. Biology, structure and mechanism of P-type ATPases. *Nat. Rev. Mol. Cell Biol.* **5**, 282–295 (2004).

74. Sørensen, D. M., Buch-Pedersen, M. J. & Palmgren, M. G. Structural divergence between the two subgroups of P5 ATPases. *Biochim. Biophys. Acta* **1797**, 846–855 (2010).

75. Li, P. *et al.* The structure and function of P5A-ATPases. *Nat. Commun.* **15**, 9605 (2024).

76. Ueno, T. *et al.* A mouse with a point mutation in plasma membrane Ca^{2+}-ATPase isoform 2 gene showed the reduced Ca^{2+} influx in cerebellar neurons. *Neurosci. Res.* **42**, 287–297 (2002).

77. Sun, K. Y. *et al.* A deep catalogue of protein-coding variation in 983,578 individuals. *Nature* **631**, 583–592 (2024).

78. Ittisoponpisan, S. *et al.* Can predicted protein 3D structures provide reliable insights into whether missense variants are disease associated? *J. Mol. Biol.* **431**, 2197–2212 (2019).

79. Khanna, T., Hanna, G., Sternberg, M. J. E. & David, A. Missense3D-DB web catalogue: an atom-based analysis and repository of 4M human protein-coding genetic variants. *Hum. Genet.* **140**, 805–812 (2021).

80. Zhang, Y.-L. *et al.* Novel fluorescence-based high-throughput FLIPR assay utilizing membrane-tethered genetic calcium sensors to identify T-type calcium channel modulators. *ACS Pharmacol. Transl. Sci.* **5**, 156–168 (2022).

81. Palmgren, M. P-type ATPases: Many more enigmas left to solve. *J. Biol. Chem.* **299**, 105352 (2023).

82. Zheng, S., Wang, X., Zhao, D., Liu, H. & Hu, Y. Calcium homeostasis and cancer: insights from endoplasmic reticulum-centered organelle communications. *Trends Cell Biol.* **33**, 312–323 (2023).

83. Brini, M., Calì, T., Ottolini, D. & Carafoli, E. Intracellular calcium homeostasis and signaling. *Met. Ions Life Sci.* **12**, 119–168 (2013).

84. Ebert, D. H. & Greenberg, M. E. Activity-dependent neuronal signalling and autism spectrum disorder. *Nature* **493**, 327–337 (2013).

85. Ma, H. *et al.* Excitation–transcription coupling, neuronal gene expression and synaptic plasticity. *Nat. Rev. Neurosci.* **24**, 672–692 (2023).

86. Gleichmann, M. & Mattson, M. P. Neuronal calcium homeostasis and dysregulation. *Antioxid. Redox Signal.* **14**, 1261–1273 (2011).

87. Brini, M., Calì, T., Ottolini, D. & Carafoli, E. Neuronal calcium signaling: function and dysfunction. *Cell. Mol. Life Sci.* **71**, 2787–2814 (2014).

88. West, A. E. *et al.* Calcium regulation of neuronal gene expression. *Proc. Natl. Acad. Sci. U. S. A.* **98**, 11024–11031 (2001).

89. Pankevich, D. E., Altevogt, B. M., Dunlop, J., Gage, F. H. & Hyman, S. E. Improving and accelerating drug development for nervous system disorders. *Neuron* **84**, 546–553 (2014).

90. Trajanoska, K. *et al.* From target discovery to clinical drug development with human genetics. *Nature* **620**, 737–745 (2023).

91. Shapson-Coe, A. *et al.* A petavoxel fragment of human cerebral cortex reconstructed at nanoscale resolution. *Science* **384**, eadk4858 (2024).

92. Bulik-Sullivan, B. K. *et al.* LD Score regression distinguishes confounding from polygenicity in genome-wide association studies. *Nat. Genet.* **47**, 291–295 (2015).

93. Watanabe, K. *et al.* Genome-wide meta-analysis of insomnia prioritizes genes associated with metabolic and psychiatric pathways. *Nat. Genet.* **54**, 1125–1132 (2022).

94. Watanabe, K., Taskesen, E., van Bochoven, A. & Posthuma, D. Functional mapping and annotation of genetic associations with FUMA. *Nat. Commun.* **8**, 1826 (2017).

95. Krienen, F. M. *et al.* Innovations present in the primate interneuron repertoire. *Nature* **586**, 262–269 (2020).

96. Butler, A., Hoffman, P., Smibert, P., Papalexi, E. & Satija, R. Integrating single-cell transcriptomic data across different conditions, technologies, and species. *Nat. Biotechnol.* **36**, 411–420 (2018).

97. Hafemeister, C. & Satija, R. Normalization and variance stabilization of single-cell RNA-seq data using regularized negative binomial regression. *Genome Biol.* **20**, 296 (2019).

98. Roadmap Epigenomics Consortium *et al.* Integrative analysis of 111 reference human epigenomes. *Nature* **518**, 317–330 (2015).

99. Eden, E., Navon, R., Steinfeld, I., Lipson, D. & Yakhini, Z. GOrilla: a tool for discovery and visualization of enriched GO terms in ranked gene lists. *BMC Bioinformatics* **10**, 48 (2009).

100. Wagnon, J. L. *et al.* CELF4 regulates translation and local abundance of a vast set of mRNAs, including genes associated with regulation of synaptic function. *PLoS Genet.* **8**, e1003067 (2012).

101. Koscielny, G. *et al.* Open Targets: a platform for therapeutic target identification and validation. *Nucleic Acids Res.* **45**, D985–D994 (2017).

102. Korsunsky, I., Nathan, A., Millard, N. & Raychaudhuri, S. Presto scales Wilcoxon and auROC analyses to millions of observations. *bioRxiv* 653253 (2019) doi:10.1101/653253.

103. Tian, R. *et al.* Whole-exome sequencing in UK Biobank reveals rare genetic architecture for depression. *Nat. Commun.* **15**, 1755 (2024).