ORCA – Online Research @ Cardiff

This is an Open Access document downloaded from ORCA, Cardiff University's institutional repository:https://orca.cardiff.ac.uk/id/eprint/149119/

This is the author's version of a work that was submitted to / accepted for publication.

Citation for final published version:

Singh, Tarjinder, Schizophrenia Exome Meta-Analysis (SCHEMA) Consortium, Kirov, George , Rees, Elliott , Walters, James , Owen, Michael and O'Donovan, Michael 2022. Rare coding variants in ten genes confer substantial risk for schizophrenia. Nature 604 , pp. 509-516. 10.1038/s41586-022-04556-w

Publishers page: http://doi.org/10.1038/s41586-022-04556-w

Please note:
Changes made as a result of publishing processes such as copy-editing, formatting and page numbers may not be reflected in this version. For the definitive version of this publication, please refer to the published source. You are advised to consult the publisher's version if you wish to cite this paper.

This version is being made available in accordance with publisher policies. See http://orca.cf.ac.uk/policies.html for usage policies. Copyright and moral rights for publications made available in ORCA are retained by the copyright holders.

Rare coding variants in 10 genes confer substantial risk for schizophrenia

Tarjinder Singh on behalf of the Schizophrenia Exome Meta-Analysis (SCHEMA) Consortium*

*The complete list of authors appears at the end of the paper.

Abstract

Rare coding variation has historically provided the most direct connections between gene function and disease pathogenesis. By meta-analyzing the whole-exomes of 24,248 cases and 97,322 controls, we implicate ultra-rare coding variants (URVs) in ten genes as conferring substantial risk for schizophrenia (odds ratios 3 - 50, $P < 2.14 \times 10^{-6}$), and 32 genes at a FDR < 5%. These genes have the greatest expression in central nervous system neurons and have diverse molecular functions that include the formation, structure, and function of the synapse. The associations of NMDA receptor subunit *GRIN2A* and AMPA receptor subunit *GRIA3* provide support for the dysfunction of the glutamatergic system as a mechanistic hypothesis in the pathogenesis of schizophrenia. We observe an overlap of rare variant risk between schizophrenia, autism spectrum disorders (ASD)[1], epilepsy and severe neurodevelopmental disorders (DD/ID)[2], though in some shared genes different mutation types are implicated. Most genes described here however are not implicated in neurodevelopment and we demonstrate that genes prioritized from common variant analyses of schizophrenia are enriched in rare variant risk[3], suggesting that common and rare genetic risk factors at least partially converge on the same underlying pathogenic biological processes. Even after excluding significantly associated genes, schizophrenia cases still carry a substantial excess of URVs, implying that more risk genes await discovery using this approach.

Introduction

Schizophrenia is a severe psychiatric disorder with signs and symptoms that include hallucinations, delusions, disorganized speech and behavior, diminished emotional expression, social withdrawal, and cognitive impairment. The disorder has a lifetime risk of ~0.7%, is often disabling, and reduces life expectancy by nearly 15 years[4,5]. Existing therapies largely address primarily positive symptoms (e.g., hallucinations and delusions) and response to existing antipsychotic medications is highly variable with ~30% of patients classified as treatment resistant[6]. The lack of progress in therapeutic development is in part a consequence of our limited understanding of the molecular etiology of psychiatric disorders[6,7].

It is well-established that schizophrenia has a substantial genetic component with contributions from across the allele frequency spectrum[8–11]. As initially theorized, the high

heritability, consistency of prevalence across populations and increasing risk observed for individuals in more densely affected families suggested that polygenic predisposition should play a dominant role in defining schizophrenia risk in the population[4,12]. This has been borne out by genome-wide association studies (GWAS) which have now, in a companion paper, identified 270 common (minor allele frequency [MAF] > 1%) risk loci of individually small effect (median odds ratio [OR] < 1.05)[13]. As a class of variation, common variants explain ~24% of the variance in disease liability[14]. Several rare (MAF < 0.1%) recurrent copy number variants (CNVs) have also been robustly associated with schizophrenia, as exemplified by the dramatically higher rates of schizophrenia in 22q11.2 deletion carriers[10,15]. This suggests a role for rare gene-disrupting mutations with much larger effects on individual risk (OR 2 - 60). Although the variants we have been able to implicate have large effects on risk in the individual, because they are rare they make only a small contribution to overall heritability in the population. Despite these successes in locus discovery, it remains challenging to move from individual associations to specific genes and disease mechanisms. Because causal variants in schizophrenia GWAS are predominantly non-coding, challenges related to fine-mapping and interpretation of intergenic and intronic elements limit our ability to confidently identify underlying genes, infer the mechanism by which they influence disease risk, and determine the direction of effect. CNVs of large effect, on the other hand, often disrupt hundreds of kilobases of the genome and multiple genes simultaneously, limiting our ability to derive clear functional insights[10].

Analyzing rare coding variants offers a powerful complementary approach to identify genes in complex traits. Theory predicts that the forces of natural selection will tend to keep large effect risk variants at much lower frequencies in the population, especially in disorders such as schizophrenia that are associated with reduced fecundity[16]. However, most rare variants will have little or no functional consequence or impact on risk, posing a significant challenge in identifying those that are truly causal and complicating required analyses in which rare variants are tested as a group rather than individually. The most natural grouping for rare variants is within a gene, based on predicted functional consequence or evidence for deleteriousness[16,17]. Protein-truncating variants (PTVs) are among the most interpretable associations as they suggest that the effect on disease most commonly tracks with decreasing expression of the gene[18]. Earlier schizophrenia sequencing studies have established that ultra-rare and *de novo* mutations contribute to risk as a category, and have prioritized disease-relevant tissues and processes, specifically observing an enrichment in neuronal genes and synaptic processes[9,11,19–23]. Furthermore, these risk alleles are concentrated in genes with a near-complete depletion of protein-truncating variants in population studies, a result shared with other neurodevelopmental disorders[9,11] and suggesting strong direct selection against such mutations. However, the analysis of URVs has had limited success in delivering individual gene discovery in schizophrenia because of power limitations, with only a single gene, *SETD1A*, identified as robustly associated[16,21].

The Schizophrenia Exome Sequencing Meta-Analysis (SCHEMA) Consortium was formed as a global collaborative effort to analyze sequence data from many studies to advance gene discovery. Here, we generated, aggregated, harmonized variant identification, and meta-analyzed the exome sequences of 24,248 individuals with schizophrenia and 97,322 controls

from seven continental populations. This analysis is, to our knowledge, one of the largest sequencing studies of a complex trait to date. As predicted by apparent rare variant burden in schizophrenia, increasing the sample size has led to the identification of 10 genes with URVs that confer substantial risk at exome-wide significance. Combining these findings with other large-scale sequencing studies, we find shared and distinct genetic signals between schizophrenia and other neurodevelopmental disorders. In tandem with a companion paper from the Psychiatric Genomics Consortium[13], we provide evidence that common and ultra-rare coding variants identify an overlapping set of genes. Finally, we demonstrate that increased scale following this approach will uncover additional risk genes and help complete the genetic architecture of schizophrenia.

Results

Data description, generation, and quality control

We aggregated exome sequence data consisting of 24,248 individuals diagnosed with schizophrenia and 50,437 individuals without a known psychiatric diagnosis, recruited in eleven global collections that had previously contributed to common variant association efforts (Supplementary Methods, Figure 1A, Table S1). The sequence data for 7,979 cases had been previously presented in earlier publications[9,11,19–22], while the remaining 16,269 cases are presented here for the first time. To ensure calibrated analyses, these samples were included in joint re-processing and variant calling using a standardized BWA-Picard-GATK pipeline as part of the larger Genome Aggregation Database (gnomAD) effort (Supplementary Methods); consequently, SCHEMA case-control samples with appropriate permissions are also included in the gnomAD v2 release[24]. After extracting SCHEMA samples from this callset, we performed quality control steps to ensure high quality of sequence data, exclude contaminated samples, identify parent-proband trios and other related individuals, and infer global ancestries (Supplementary Methods, Figure 1B, Figures S1-7, Table S2). We subsequently applied site- and genotype-level filters to generate a robust set of coding SNPs and indels for a well-matched case-control analysis (Supplementary Methods). Previous studies have shown that PTVs are concentrated in 3,063 genes under strong constraint in schizophrenia cases compared to controls[11,25], and we replicated this result with consistent signals across our major cohorts (P_{meta} = 7.6 x 10^{-35}; OR = 1.26, 95% CI = 1.22 - 1.31, Figure 1C, Figure S8).

Analysis approach

To increase power for gene discovery, we incorporated variant counts from additional samples from non-psychiatric and non-neurological collections that were aggregated as part of the gnomAD consortium effort (Supplementary Methods)[24]. We attempted to control for technical and methodological batch effects that may arise from this approach in both variant calling and additionally via permutation testing described below. All samples in gnomAD and SCHEMA consortia were re-processed and joint called using the same pipeline, and the same variant filters were applied to arrive high-quality calls. Importantly, we restricted our analysis to

coding exons with high-quality data across all major exome capture technologies, reducing any artifacts that may arise from coverage differences(Supplementary Methods, Figures S1-2). After incorporating variant counts from additional 46,885 gnomAD controls, our combined discovery data set is composed of 24,248 cases and 97,322 population controls (Figure 1A, 1B, Table S3).

Because only summary-level variant counts were available for the 46,885 external controls, we tested for an excess of disruptive variants per gene using a Fisher's exact test in which statistical significance was determined by case-control permutations within each strata (Supplementary Methods, Table S3). As in other sequencing studies, we enriched for pathogenic variants by restricting our analysis to ultra-rare variants (defined as minor allele count [MAC] ≤ 5) that are also either PTVs (defined as stop-gained, frameshift, and essential splice donor or acceptor variants) or damaging missense variants as defined by the MPC pathogenicity score[1,26](Supplementary Methods). We found that missense variants with MPC > 3 have a global signal on par with PTVs in schizophrenia, autism spectrum disorders, and severe neurodevelopmental disorders, while variants with MPC 2 - 3 has a significant but weaker signal than PTVs and were therefore analyzed separately (Figure 1C, Figures S9, S10, S11, Table S4, Supplementary Methods). Motivated by these observations, we performed a burden test of PTVs and MPC > 3 variants (Class I) to generate a *P* value for 18,321 protein-coding genes (Supplementary Methods). In the 4,512 genes with MPC 2 - 3 (Class II) variants, we perform an additional test aggregating these variants, and meta-analyze these gene statistics with Class I *P* values using a weighted *Z*-score method (Supplementary Methods). To ensure the robustness of the results generated by this approach, we observed the expected null distribution of *P* values in gene-based tests of synonymous variants in each strata and in the meta-analysis (Figure S12, S13). Additionally, we observed no inflation of synonymous *P* values using the Mantel-Haenszel test even after limiting our analysis to genes with larger total numbers of alleles (gene-wide MAC > 10, 50, or 100), where we had greater power to detect potential artifacts (Figure S14, S15).

Previous studies had integrated case-control and trio-based *de novo* mutations for gene discovery[1,21], and to this end, we aggregated and re-annotated *de novo* mutations from 3,402 published parent-proband trios (Supplementary Methods). Despite the sizable number of trios, there were few *de novo* mutations for analysis with only 325 genes with one or more *de novo* PTV and only 449 with at least one Class I or Class II mutations. Using Poisson rate tests based on expected mutation rate[27], we found these *de novo* mutations are enriched for the 244 genes with $P < 0.01$ in our case-control analysis (Figure S16, Table S5), with limited or no signal in the remaining genes in the genome (Figure 1D). The most striking enrichment was observed for the 52 genes with case-control $P < 0.001$ (Class I mutations: $P = 2.1 \times 10^{-11}$; Rate ratio = 8.3, 95% CI = 4.9 - 13), which provides additional reassurance of the robustness of our case-control gene results. Motivated by these observations, we calculated *de novo* Class I and II *P* values in the 244 genes with $P_{\text{case-control}} < 0.01$ using the Poisson rate test and meta-analyzed them with our case-control test statistic using a weighted Z-score method to increase power (Supplementary Methods).

Risk genes implicated by rare protein-coding variants

Combined, our meta-analysis of 24,248 cases, 97,322 controls, and *de novo* mutations from 3,402 trios implicates 10 genes in which ultra-rare coding variants are significantly associated with schizophrenia ($P < 2.14 \times 10^{-6}$ corresponding to 0.05/23,321 tests; Figure 2A, 2B). These top associations as a group are supported by complementary types of variation that include case-control PTVs, damaging missense variants, and *de novo* mutations (Table 1, Table S5). Although confidence intervals are wide, URVs in these genes appear to confer substantial risk, with odds ratio of PTVs and Class I variants ranging from 3 to 50. As expected, all ten genes are among the most constrained genes in the genome, with a substantial depletion of PTVs compared to chance expectation[24]. The annotated functions of these genes are diverse and include ion transport (*CACNA1G*, *GRIN2A*, and *GRIA3*), neuronal migration and growth (*TRIO*), transcriptional regulation (*SP4, RB1CC1*, and *SETD1A*), nuclear transport (*XPO7*), and ubiquitin ligation (*CUL1*, *HERC1*). We include a brief discussion of the known biological functions of these genes in Box 1. Beyond these ten genes, we identify 22 additional genes at a False Discovery Rate (FDR) < 5% (Figure 2A, Table S5). We observe notable deviation at the tail of the distribution beyond the associated genes, suggesting that more genes remain to be discovered (Figure 2B). We report all high-quality variants, relevant annotations, and gene-level results on a public browser at https://schema.broadinstitute.org.

The identification of individual genes provides support for more specific mechanistic hypotheses underlying schizophrenia pathogenesis. Developed from neuropharmacological and neuropathological observations, the glutamatergic hypothesis postulates that the hypofunction of glutamatergic signaling through NMDA receptors is a possible mechanism of disease[28] (Box 1). Here, we find that PTV and damaging missense variants in NMDA receptor subunit *GRIN2A* confer substantial risk for schizophrenia ($P = 7.37 \times 10^{-7}$; Class I [PTV and MPC > 3] OR 24.1, 95% CI 5.36 - 221; Class II [MPC > 2] OR 2.37, 95% CI 1.1 - 4.92). Schizophrenia GWAS also identified a common variant at *GRIN2A* (OR = 1.057, $P = 1.57 \times 10^{-10}$), providing an allelic series in which different perturbation of gene function results in severity of disease risk (Figure 3A)[8].The NMDA receptor changes in composition during prenatal to postnatal neurodevelopment with *GRIN2A* predominantly expressed during late childhood and adolescence, recapitulating expected epidemiological observations on schizophrenia age-of-onset (Supplementary Methods, Figure 3B)[29]. We additionally find that risk URVs in AMPA receptor subunit *GRIA3* confer substantial risk ($P = 5.98 \times 10^{-7}$; Class I [PTV and MPC > 3] OR 20.1 95% CI 4.28 - 188; Table 1). Combined, our results from exome sequencing support the dysregulation of the glutamatergic system as a mechanistic hypothesis for the development of schizophrenia, and that the specific identification of genes by coding variation may provide new avenues of understanding disease pathogenesis.

Shared genes and processes identified by common variants and ultra-rare protein-coding variants

Pathway analyses of common variants have prioritized disease-relevant tissues and cell types, and in some cases, independently recapitulating known biology[8,30,31]. To derive insights from global patterns of rare coding variants, we tested for an excess burden of URVs in schizophrenia cases compared to controls in 1,732 broadly-defined gene sets from databases of biological pathways (e.g. Gene Ontology, REACTOME, KEGG) and experimental data (Supplementary Methods)[11]. We observed significant enrichment of URVs in 33 gene sets ($P < 2.9 \times 10^{-5}$) that recapitulated consistent and overlapping cellular compartments and biological processes, including definitions of the postsynaptic density (human cortex biopsy post-synaptic density; $P = 1.2 \times 10^{-12}$), chromatin modification (GO:0016568; $P = 1.8 \times 10^{-12}$), regulation of ion transmembrane transport (GO:0034765; $P = 6.7 \times 10^{-7}$), axon guidance ($P = 5.4 \times 10^{-6}$), voltage-gated cation channel activity (GO:0022843; $P = 8.1 \times 10^{-6}$), and synaptic transmission (GO:0007268; $P = 1.79 \times 10^{-5}$) (Table S6, Figure S19). Because of the clear synaptic signal, we investigated in the refined synaptic ontology defined by the SynGO consortium[32], and found consistent enrichment for postsynaptic components and processes (GO:0098794; $P = 3.9 \times 10^{-6}$; Table S7). These global observations are consistent with the known functions of the individual risk genes now implicated by rare variation (Box 1). Following earlier reports studying heritability enrichment in GTEx tissues[8,31], we found that genes with the highest specific expression in brain regions showed the strongest enrichment of risk URVs, most significantly in the human frontal cortex ($P = 1.63 \times 10^{-8}$) and with limited signal in the other tissue types (Figure S20, Table S8, S9). To further deconvolute this signal, we investigated which single cell types in the mouse nervous system show the highest specific expression for the 32 (FDR < 5%) schizophrenia risk genes (Supplementary Methods)[33,34]. Here, we found widespread enrichments across central nervous system neurons with limited to no signal in glial cells and peripheral nervous system neurons (Table S10, Figure S21). Thus, at a high level, global analysis of ultra-rare protein-coding variation independently recapitulated known biology related to schizophrenia pathogenesis, including processes, cellular components, and tissues previously implicated by common variant analyses.

To evaluate the overlap of schizophrenia associations from common variants and ultra-rare coding variant analyses, we jointly analyzed our results with the largest GWAS of schizophrenia to date, which identified common variant associations at 270 distinct loci from the analysis of 69,369 cases and 236,642 controls[13]. Statistical fine-mapping prioritized the likely underlying protein-coding gene at 64 of these associations (Table S11, Figure S22), and we found a case-control enrichment of URVs in these genes ($P_{meta} = 3.9 \times 10^{-4}$; $OR_{Class\ I} = 1.46$, 1.2 - 1.77 95% CI; Figure 4A, Table S12). Beyond the statistical enrichment, *GRIN2A* and *SP4,* two of the ten significant rare variant genes, had clear associations in schizophrenia GWAS (Figure 3A, Figure 4B). Furthermore, *FAM120A* and *STAG1* resided in more complex GWAS-associated regions containing multiple genes but were prioritized among their neighbors as FDR < 5% in our sequencing study (Figure 4C, 4D). Combined, these results suggest there is at least partial convergence in the genes and biological processes implicated by common and

ultra-rare genetic variation, and that ultra-rare coding variants can be leveraged to prioritize genes within GWAS loci.

Shared and distinct genetic risk with other neurodevelopmental disorders

Exome sequencing studies of autism spectrum disorders (ASD) and severe neurodevelopmental disorders (DD/ID) have leveraged ultra-rare coding variants to identify risk genes. These studies have established that the genetic signals were concentrated in constrained genes and shared between the two disorders[35,36]. Most recently, the analysis of *de novo* mutations from 31,058 DD/ID trios implicated 299 genes, while the analysis of 11,986 ASD cases identified 102 genes at FDR < 10% (Table S11)[1,37]. We found a significant excess of URVs in schizophrenia cases compared to controls in the 299 DD/ID-associated genes (P_{meta} = 1.5 x 10^{-14}; $OR_{Class\ I}$ = 1.44, 1.3 - 1.6 95% CI), and in the 102 ASD-associated genes (P_{meta} = 3.7 x 10^{-7}; $OR_{Class\ I}$ = 1.45, 1.23 - 1.72 95% CI; Figure 5A; Table S12). Thus, some schizophrenia rare variant risk appears to be shared with other neurodevelopmental disorders.

With 31,058 trios, the scale of gene discovery in severe DD/ID provided sufficient power to evaluate the individual schizophrenia risk genes associated in our study for a role in broader neurodevelopmental disorders. Nine of the ten schizophrenia genes showed limited *de novo* PTV signal in DD/ID, with a combined 8 *de novo* PTVs observed in these genes (X_{exp} = 4.98; P_{Pois} = 0.13; Figure 5B; Table S13). *SETD1A* had a significant *de novo* PTV signal in DD/ID (X_{obs} = 8, X_{exp} = 0.41; P = 1.3 x 10^{-8}), supporting an earlier report that described *SETD1A* as a gene associated with both schizophrenia and broader neurodevelopmental disorders[21]. We also observed a missense signal in *SETD1A* in our study (Table 1; Figure S23). Extending this analysis to the additional 22 FDR < 5% genes, we found that six genes (*STAG1*, *ASH1L*, *ZMYM2*, *KDM6B*, *SRRM2*, and *HIST1H1E*) were significantly associated with DD/ID in addition to schizophrenia (Figure 5B; Table S13). Among these FDR < 5% genes, *ASH1L*, *KDM6B* and *NR3C2* were associated with ASD [1] (Table S13). Broadly speaking, while PTV mutations in certain genes are joint risk factors for schizophrenia and DD/ID, the majority of schizophrenia associations reported here appear to have little or no role in DD/ID despite the enormous power of published DD/ID studies to date.

Notably, three of the ten risk genes for schizophrenia (*TRIO*, *GRIN2A*, and *CACNA1G*) were associated with risk of severe DD/IDs exclusively through *de novo* missense mutations that cluster within each gene (Figure 5B; Table S13), while the schizophrenia signal was largely driven by PTVs. *De novo* missense mutations in *TRIO* significantly disrupted the exons preceding or containing the RhoGEF domain (Figure 5C)[37,38], and *de novo* missense mutations in *GRIN2A* cluster at the base of the ion channel with the most mutations in the exon encoding for the pore of the complex (Figure 5D). *STAG1*, which had a common and rare variant signal in schizophrenia (Figure 4D), was associated with DD/ID primarily through *de novo* missense mutations (Figure 5B; Table S13). These observations suggest schizophrenia and childhood onset neurodevelopmental disorders share some genes and biological processes, but that at

least in some cases, the severity or the nature of the functional impairment differs between disorders.

We explored what properties may differ between schizophrenia- and DD/ID-associated risk genes, and hypothesized that DD/ID genes were under stronger evolutionary constraint with a bias towards prenatal expression when compared to schizophrenia genes. While schizophrenia genes (FDR < 5%) were under substantial genic constraint compared to expectation (Figure S24; M-W U test; $P = 2.9 \times 10^{-7}$; Supplementary Methods), they are significantly less constrained than DD/ID-associated genes (M-W U test; $P = 3.5 \times 10^{-5}$). Furthermore, schizophrenia genes as a group did not show pre- or postnatal bias in brain expression ($P = 0.21$; Figure S25), while DD/ID-associated genes were overwhelmingly prenatal in expression ($P = 7.5 \times 10^{-20}$). Indeed, individual genes like *SETD1A*, *TRIO,* and *SP4* exhibited prenatal expression while *GRIN2A* and *GRIA3* showed postnatal expression (Figure S26). These observations offer the possibility that certain properties may differentiate genes for adult psychiatric disorders and more severe DD/IDs.

Contribution of ultra-rare PTVs to schizophrenia risk

Efforts in the past decade are beginning to generate a more comprehensive view of the genetic architecture for schizophrenia, composed of common variants of small effects, large CNVs with elevated frequencies driven by genomic instability, and now, URVs of large effect implicating individual genes (Figure 6A)[8,10]. Because schizophrenia as a trait is under strong selection[39–41], we expect that URVs of large effect to be frequently *de novo* or of very recent origin and contribute to risk in only a fraction of diagnosed patients. We quantified the contribution of PTVs to risk first in our full schizophrenia data set, and then partitioned the *de novo* and inherited contributions in 2,304 parent-proband trios. We restrict these analyses to the 3,063 PTV-intolerant (pLI > 0.9) genes in which schizophrenia risk URVs are concentrated. We observed 0.057 (0.049 - 0.065 95% CI) extra singleton PTV variants per individual in cases compared to controls, suggesting ~5.7% of cases carried a PTV relevant to disease risk. In the 2,304 trios, 0.0394 (0.014 - 0.065 95% CI; 74%) extra singleton PTV variants were inherited per proband, and 0.0121 (0.0022 - 0.02 95% CI; or 26%) extra *de novo* PTV mutations in constrained genes were identified in cases compared to controls. In contrast, DD/ID probands have 0.111 (0.103 - 0.119 95% CI) extra *de novo* PTV mutations in constrained genes, while ASD individuals have 0.0478 (0.0387 - 0.0568 95% CI) extra *de novo PTV* mutations (Figure S27; Supplementary Methods). In the ten schizophrenia-associated genes, 7 *de novo* mutations and 13 transmitted variants are observed in 2,304 trios, suggesting that 0.86% of patients are carriers and ~35% of variants are *de novo*. Finally, the genome-wide signal in constrained genes (pLI > 0.9: OR = 1.26, $P = 7.6 \times 10^{-35}$) remains significant even after excluding the 32 FDR < 5% genes (OR = 1.23, $P = 4.3 \times 10^{-27}$; Figure 6B, Table S4), reaffirming the genetic heterogeneity underlying schizophrenia risk and suggesting that the majority of schizophrenia risk genes in which rare variants confer risk remain to be discovered.

Discussion

In one of the largest exome sequencing studies to date, we identify genes in which disruptive coding variants confer substantial risk for schizophrenia at exome--wide significance. This effort required re-processing a decade of sequence data, harmonization of variant calling and quality control, inclusion of external controls, and integration of PTV, damaging missense, and *de novo* variants. Global, collaborative efforts such as this provide a template for tackling the genetic contributions in other complex diseases.

Genome-wide analyses recapitulated known biological processes and reaffirm that schizophrenia risk genes are involved in the postsynaptic density and broader synaptic function, and enriched in expression in neuronal tissues. Furthermore, the identification of specific genes supports more specific mechanistic hypotheses. The association of PTVs in the NMDA receptor subunit *GRIN2A* to schizophrenia risk provides genetic support for the dysregulation of glutamatergic signaling as a possible mechanism of disease. A natural dose-response curve occurs at this gene in which common regulatory variants modestly influence disease risk and PTV and predicted damaging missense variants increase risk more substantially. Interestingly, the NMDA receptor is composed of two *GRIN2* units (*GRIN2A* and/or *GRIN2B*) along with two constitutive *GRIN1* units, and *GRIN2A* increases dramatically in expression later in childhood and adolescence, mimicking the age of onset of disease for schizophrenia. *De novo* mutations in *GRIN2B* conversely are associated with more severe disorders of neurodevelopment that manifest in childhood, including intellectual disability and autism[42]. Such findings provide a unique opportunity to identify experiments of nature which help to build and support mechanistic hypotheses that may lead to a better understanding of disease biology.

Joint analysis with genetic data from DD/ID and ASD consortia have provided evidence for shared genes between neuropsychiatric and broader neurodevelopmental disorders. Indeed, seven of the 32 FDR < 5% genes are also associated with DD/ID, providing additional confidence in those associations. The shared genes suggest that there is at least some contribution from early brain developmental processes that predisposes to schizophrenia. Despite this sharing, PTVs in 9 of the 10 most confidently associated genes are associated with schizophrenia and not for DD/ID, which may provide avenues for identifying disease-specific processes. Of further interest, we observe allelic series in *GRIN2A, TRIO,* and *CACNA1G* in which PTVs increase schizophrenia risk and *de novo* missense mutations confer strong DD/ID risk. *De novo* missense mutations in these genes clustered in specific domains and are associated with more severe neurodevelopmental, syndromic disorders with cognitive impairment, suggesting an alternate or gain-of-function effect. Analyses estimating relative penetrance for different phenotypes will increase in power as consortium efforts studying specific diseases and biobank efforts continue to grow, all of which would be fruitful in informing what is shared and distinct across disorders.

We show for the first time that common regulatory variants from GWAS and ultra-rare coding variants disrupt an overlapping set of genes, including an allelic series in four genes in

which common variants and rare coding variants increase risk to varying degrees. Combined, these results suggest that exome sequencing identifies some common, shared underlying biology that is dysregulated across the allele frequency spectrum, rather than syndromic forms of disease with unrelated biology regulated by common variation. Furthermore, because of this sharing, coding variants can help refine and fine-map common variant associations like at the *STAG1* and *FAM120A* loci. As common and rare variant association studies continue to grow, we can better determine the actual degree of overlap of genes that are regulated by both types of variation. Ultimately, the emerging evidence of an overlap between common and ultra-rare variation gives confidence that the integration of results from sequencing consortia with the GWAS efforts will have significant value for identifying specific genes beyond what any single strategy can achieve on its own.

A decade of genotyping and sequencing studies now establish specific genetic contributions from common variants, copy number variants, and ultra-rare coding variants as conferring risk for schizophrenia. Despite this progress, it is clear that we are still in the early stages of gene discovery[13]. The vast majority of risk alleles, their direction and magnitude of effect, mode of action, and responsible genes are yet to be discovered. These emerging genetic findings will serve in part to direct and motivate mechanistic studies that begin to unravel disease biology. The success of common variant association studies, and now exome sequencing, suggest concrete progress towards understanding the causes of human complex traits and diseases, and provide a clear roadmap towards understanding the genetic architecture of schizophrenia.

Box 1

Here, we provide a brief review of literature on exome-wide significant genes.

***SETD1A* (SET Domain Containing 1A, Histone Lysine Methyltransferase)**
SETD1A encodes a catalytic subunit of the COMPASS histone methyltransferase complex that performs mono-, di-, and tri-methylation at Histone H3 Lysine 4. Through epigenetic modification, SETD1A influences transcriptional regulation that downstream affects axonal branching and cortical synaptic dynamics[43]. Paralogs of this gene, including *KMT2A* (Wiedemann-Steiner syndrome), *KMT2B*, *KMT2C*, and *KMT2D* (Kabuki syndrome), have been associated with severe neurodevelopmental disorders[21,44].

***CUL1* (Cullin 1)**
CUL1 is a core component of a E3 ubiquitin-protein ligase involved in the ubiquitination of proteins broadly involved in signal transduction, gene regulation, and cell cycle progression[45,46]. CUL1, RBX1, and SKP1 form the invariable components of the SCF complex that recruits proteins for degradation[47]. Mouse studies have demonstrated that CUL1 is required for early development, and mutant mice fail to regulate G1 cyclin (cyclin E) during embryogenesis[47].

***XPO7* (Exportin-7)**
XPO7 is involved in the trafficking of specific proteins through the nuclear pores to the cytoplasm as part of the Ran-GTP pathway[48]. In a GTP-bound form, exportins bind to nuclear export signal sequences of select proteins, interact with nucleoporins to transit into the cytoplasm, release the cargo protein, and relocate to the nuclear compartment. While exportins are generally involved in nuclear export, XPO7 may be involved in the nuclear import of select substrates as well[48].

***TRIO* (Trio Rho Guanine Nucleotide Exchange Factor)**
TRIO encodes for a Rho GDP to GTP exchange factor that promote actin cytoskeleton reorganization through the direct activation of the signaling G protein RAC1. TRIO has been demonstrated to regulate neuronal migration, axonogenesis, axon guidance, and synaptogenesis through actin cytoskeleton remodeling. Mutations in *TRIO* are associated with intellectual disability and neurodevelopmental disorders, with mutational hotspots at the seventh spectrin repeat and RAC1-activating GEF1 domain[38,49]. Specific mutations in these domains cause hyper- or hypo-activation of RAC1, resulting in varying severities of developmental delay[38]. TRIO is been shown to directly affect glutamatergic neurotransmission in rodent neurons *in vitro*: specific hypomorphs in the GEF1/Rac1 activating domain reduce synaptic AMPA receptor expression, while hyperfunctional mutations enhance glutamatergic signaling and synaptogenesis[50].

***CACNA1G* (Calcium Voltage-Gated Channel Subunit Alpha-1G subunit)**
CACNA1G, also known as *Cav3.1*, encodes for the alpha 1 G subunit of a T-type voltage-sensitive calcium channel that mediates calcium influx into excitable cells[51]. T-type channels produce transient and small currents that modulate the firing patterns of neurons and cardiac

nodal cells, which has been shown to influence muscle contraction, neurotransmitter release, cell division and growth, cell motility, and sleep stabilization[52,53]. A specific functional mutation in *CACNA1G* has been linked to hereditary cerebellar ataxia in a single family[54].

***SP4* (Sp4 Transcription Factor)**

SP4 encodes a transcription factor that binds to GT and GC promoter elements to activate transcription. SP1 and SP3 transcription factors recognize similar DNA elements, but SP4 has stronger specific abundance in the central nervous system[55]. SP4 knockout mice develop until birth without abnormalities but two-thirds die within four weeks of birth and the remaining third suffer severe developmental delay[55,56]. These mice also experience robust deficits in sensorimotor gating and contextual memory associated with hippocampal vacuolization[57]. Furthermore, SP4 knockout causes dramatically decreased expression of *GRIN1*, the constitutive subunit of the NMDA receptor, which may impact NMDA neurotransmission[58].

***GRIA3* (Glutamate Ionotropic Receptor AMPA Type Subunit 3)**

GRIA3, or *GluA3*, encodes for a subunit of the tetrametric AMPA-sensitive glutamate receptor, which, as part of the glutamate receptor family, serve the predominant excitatory neurotransmitter receptors in the mammalian brain. AMPA receptors are well-characterized in their involvement mammalian central nervous system function, and moderates fast excitatory synaptic transmission and are involved in long-term potential and synaptic plasticity[59–61]. Individual mutations in two AMPARs, *GRIA2* and *GRIA3*, have been linked with intellectual disability in small patient cohorts[62,63].

***GRIN2A* (Glutamate Ionotropic Receptor NMDA Type Subunit 2A)**

GRIN2A, also known as *NR2A* or *GluN2A*, encodes for the 2A subunit of the glutamatergic N-methyl-D-aspartate receptors (NMDARs) which, along with AMPARs, are implicated in learning, memory and synaptic plasticity involving long-term potentiation. NMDA receptor dysregulation has been long postulated as a hypothesis for schizophrenia etiology[64]. NMDAR channel blockers (ketamine and phencyclidine) induce schizophrenia-like symptoms and cognitive deficits in healthy individuals[65,66], and anti-NMDAR encephalitis, an autoimmune disorder originating from a antibody-mediated attack on NMDARs, result in psychosis-like symptoms and memory deficits[67]. Protein-truncating and specific missense mutations in *GRIN2A and GRIN2B* cause neurodevelopmental disorders with prominent speech features and epilepsies[42,68]. A patient cohort of 248 individuals show that missense variants in the transmembrane and linker domains are associated with severe developmental phenotypes, while PTVs and missense variants outside the pore result in mild cognitive defects[69].

***HERC1* (HECT And RLD Domain Containing E3 Ubiquitin Protein Ligase Family Member 1)**

HERC1 is a large E3 ubiquitin protein ligase with HECT and RCC1-like as its characteristic protein domains. *HERC1* is involved in membrane trafficking and cell proliferation through its interactions with clathrin, M2-pyruvate, and TSC2[70–72]. A specific recessive missense mutation in the RCC1-like domain causes severe ataxia and early death through progressive Purkinje cell degeneration[72].

***RB1CC1* (RB1 Inducible Coiled-Coil 1)**

RB1CC1 encodes for a DNA-binding transcriptional factor that coordinates cell proliferation, autophagy, and cell migration[73,74]. It binds to a GC-rich region upstream of the RB1 promoter and forms a complex with p53 to activate RB1 expression. Deletions of *RB1CC1* have increased frequencies in human breast cancers, with its involvement in cancer development primarily attributed through its regulation of RB1[73].

References

1. Satterstrom, F. K. *et al.* Large-Scale Exome Sequencing Study Implicates Both Developmental and Functional Changes in the Neurobiology of Autism. *Cell* (2020) doi:10.1016/j.cell.2019.12.036.
2. Kaplanis, J. *et al.* Evidence for 28 genetic disorders discovered by combining healthcare and research data. *Nature* (2020) doi:10.1038/s41586-020-2832-5.
3. Schizophrenia Working Group of the Psychiatric Genomics Consortium, Ripke, S., Walters, J. T. R. & O'Donovan, M. C. Mapping genomic loci prioritises genes and implicates synaptic biology in schizophrenia. *medRxiv* (2020) doi:10.1101/2020.09.12.20192922.
4. McGrath, J., Saha, S., Chant, D. & Welham, J. Schizophrenia: A Concise Overview of Incidence, Prevalence, and Mortality. *Epidemiol. Rev.* **30**, 67–76 (2008).
5. Hjorthøj, C., Stürup, A. E., McGrath, J. J. & Nordentoft, M. Years of potential life lost and life expectancy in schizophrenia: a systematic review and meta-analysis. *The Lancet Psychiatry* **4**, 295–301 (2017).
6. Lehman, A. F. *et al.* Practice guideline for the treatment of patients with schizophrenia, second edition. *Am. J. Psychiatry* **161**, 1–56 (2004).
7. Hyman, S. E. Revolution stalled. *Sci. Transl. Med.* **4**, 155cm11 (2012).
8. Schizophrenia Working Group of the Psychiatric Genomics Consortium. Biological insights from 108 schizophrenia-associated genetic loci. *Nature* **511**, 421–427 (2014).
9. Genovese, G. *et al.* Increased burden of ultra-rare protein-altering variants among 4,877 individuals with schizophrenia. *Nat. Neurosci.* **19**, 1433–1441 (2016).
10. Marshall, C. R. *et al.* Contribution of copy number variants to schizophrenia from a genome-wide study of 41,321 subjects. *Nat. Genet.* **49**, 27–35 (2017).
11. Singh, T. *et al.* The contribution of rare variants to risk of schizophrenia in individuals with

and without intellectual disability. *Nat. Genet.* 1–10 (2017).

12. Gottesman, I. I. & Shields, J. A polygenic theory of schizophrenia. *Proc. Natl. Acad. Sci. U. S. A.* **58**, 199–205 (1967).
13. Schizophrenia Working Group of the Psychiatric Genomics Consortium. Mapping genomic loci prioritises genes and implicates synaptic biology in schizophrenia. *Submitted* (2020).
14. Loh, P.-R. *et al.* Contrasting genetic architectures of schizophrenia and other complex diseases using fast variance-components analysis. *Nat. Genet.* **47**, 1385–1392 (2015).
15. Karayiorgou, M. *et al.* Schizophrenia susceptibility associated with interstitial deletions of chromosome 22q11. *Proc. Natl. Acad. Sci. U. S. A.* **92**, 7612–7616 (1995).
16. Zuk, O. *et al.* Searching for missing heritability: Designing rare variant association studies. *Proceedings of the National Academy of Sciences* **111**, E455–E464 (2014).
17. Lee, S., Abecasis, G. R., Boehnke, M. & Lin, X. Rare-variant association analysis: study designs and statistical tests. *Am. J. Hum. Genet.* **95**, 5–23 (2014).
18. Rivas, M. A. *et al.* Effect of predicted protein-truncating genetic variants on the human transcriptome. *Science* **348**, 666–669 (2015).
19. Fromer, M. *et al.* De novo mutations in schizophrenia implicate synaptic networks. *Nature* **506**, 179–184 (2014).
20. Purcell, S. M. *et al.* A polygenic burden of rare disruptive mutations in schizophrenia. *Nature* **506**, 185–190 (2014).
21. Singh, T. *et al.* Rare loss-of-function variants in SETD1A are associated with schizophrenia and developmental disorders. *Nat. Neurosci.* **19**, 571–577 (2016).
22. Howrigan, D. P. *et al.* Exome sequencing in schizophrenia-affected parent-offspring trios reveals risk conferred by protein-coding de novo mutations. *Nat. Neurosci.* **23**, 185–193 (2020).
23. Gulsuner, S. *et al.* Genetics of schizophrenia in the South African Xhosa. *Science* **367**, 569–573 (2020).

24. Karczewski, K. J. *et al.* The mutational constraint spectrum quantified from variation in 141,456 humans. *Nature* **581**, 434–443 (2020).
25. Lek, M. *et al.* Analysis of protein-coding genetic variation in 60,706 humans. *Nature* **536**, 285–291 (2016).
26. Samocha, K. E. *et al.* Regional missense constraint improves variant deleteriousness prediction. *bioRxiv* 148353 (2017) doi:10.1101/148353.
27. Samocha, K. E. *et al.* A framework for the interpretation of de novo mutation in human disease. *Nat. Genet.* **46**, 944–950 (2014).
28. Hu, W., MacDonald, M. L., Elswick, D. E. & Sweet, R. A. The glutamate hypothesis of schizophrenia: evidence from human brain tissue studies. *Ann. N. Y. Acad. Sci.* **1338**, 38–57 (2015).
29. Kang, H. J. *et al.* Spatio-temporal transcriptome of the human brain. *Nature* **478**, 483–489 (2011).
30. Psychiatric Genetics Consortium. Psychiatric genome-wide association study analyses implicate neuronal, immune and histone pathways. *Nat. Neurosci.* (2015) doi:10.1038/nn.3922.
31. Finucane, H. K. *et al.* Heritability enrichment of specifically expressed genes identifies disease-relevant tissues and cell types. *Nat. Genet.* **50**, 621–629 (2018).
32. Koopmans, F. *et al.* SynGO: An Evidence-Based, Expert-Curated Knowledge Base for the Synapse. *Neuron* **103**, 217–234.e4 (2019).
33. Zeisel, A. *et al.* Molecular Architecture of the Mouse Nervous System. *Cell* **174**, 999–1014.e22 (2018).
34. Skene, N. G. *et al.* Genetic identification of brain cell types underlying schizophrenia. *Nat. Genet.* **50**, 825–833 (2018).
35. De Rubeis, S. *et al.* Synaptic, transcriptional and chromatin genes disrupted in autism. *Nature* **515**, 209–215 (2014).

36. The Deciphering Developmental Disorders Study. Large-scale discovery of novel genetic causes of developmental disorders. *Nature* **519**, 223–228 (2015).
37. Kaplanis, J. *et al.* Integrating healthcare and research genetic data empowers the discovery of 49 novel developmental disorders. *bioRxiv* 797787 (2019) doi:10.1101/797787.
38. Barbosa, S. *et al.* Opposite Modulation of RAC1 by Mutations in TRIO Is Associated with Distinct, Domain-Specific Neurodevelopmental Disorders. *Am. J. Hum. Genet.* (2020) doi:10.1016/j.ajhg.2020.01.018.
39. Haukka, J., Suvisaari, J. & Lönnqvist, J. Fertility of patients with schizophrenia, their siblings, and the general population: a cohort study from 1950 to 1959 in Finland. *Am. J. Psychiatry* **160**, 460–463 (2003).
40. Laursen, T. M. & Munk-Olsen, T. Reproductive patterns in psychotic patients. *Schizophr. Res.* **121**, 234–240 (2010).
41. Power, R. A. *et al.* Fecundity of patients with schizophrenia, autism, bipolar disorder, depression, anorexia nervosa, or substance abuse vs their unaffected siblings. *Arch. Gen. Psychiatry* **70**, 22–30 (2013).
42. Endele, S. *et al.* Mutations in GRIN2A and GRIN2B encoding regulatory subunits of NMDA receptors cause variable neurodevelopmental phenotypes. *Nat. Genet.* **42**, 1021–1026 (2010).
43. Mukai, J. *et al.* Recapitulation and Reversal of Schizophrenia-Related Phenotypes in Setd1a-Deficient Mice. *Neuron* **104**, 471–487.e12 (2019).
44. Firth, H. V. *et al.* DECIPHER: Database of Chromosomal Imbalance and Phenotype in Humans Using Ensembl Resources. *Am. J. Hum. Genet.* **84**, 524–533 (2009).
45. Wang, Y. *et al.* Deletion of the Cul1 gene in mice causes arrest in early embryogenesis and accumulation of cyclin E. *Curr. Biol.* **9**, 1191–1194 (1999).
46. O'Hagan, R. C. *et al.* Myc-enhanced expression of Cul1 promotes ubiquitin-dependent

proteolysis and cell cycle progression. *Genes Dev.* **14**, 2185–2191 (2000).

47. Dealy, M. J. *et al.* Loss of Cul1 results in early embryonic lethality and dysregulation of cyclin E. *Nat. Genet.* **23**, 245–248 (1999).
48. Aksu, M. *et al.* Xpo7 is a broad-spectrum exportin and a nuclear import receptor. *J. Cell Biol.* **217**, 2329–2340 (2018).
49. Katrancha, S. M. *et al.* Neurodevelopmental disease-associated de novo mutations and rare sequence variants affect TRIO GDP/GTP exchange factor activity. *Hum. Mol. Genet.* **26**, 4728–4740 (2017).
50. Sadybekov, A., Tian, C., Arnesano, C., Katritch, V. & Herring, B. E. An autism spectrum disorder-related de novo mutation hotspot discovered in the GEF1 domain of Trio. *Nat. Commun.* **8**, 601 (2017).
51. Zhao, Y. *et al.* Cryo-EM structures of apo and antagonist-bound human Cav3.1. *Nature* **576**, 492–497 (2019).
52. Mangoni Matteo E. *et al.* Bradycardia and Slowing of the Atrioventricular Conduction in Mice Lacking CaV3.1/α1G T-Type Calcium Channels. *Circ. Res.* **98**, 1422–1430 (2006).
53. Anderson, M. P. *et al.* Thalamic Cav3.1 T-type Ca2+ channel plays a crucial role in stabilizing sleep. *Proc. Natl. Acad. Sci. U. S. A.* **102**, 1743–1748 (2005).
54. Coutelier, M. *et al.* A Recurrent Mutation in CACNA1G Alters Cav3.1 T-Type Calcium-Channel Conduction and Causes Autosomal-Dominant Cerebellar Ataxia. *Am. J. Hum. Genet.* **97**, 726–737 (2015).
55. Göllner, H. *et al.* Complex phenotype of mice homozygous for a null mutation in the Sp4 transcription factor gene. *Genes Cells* **6**, 689–697 (2001).
56. Supp, D. M., Witte, D. P., Branford, W. W., Smith, E. P. & Potter, S. S. Sp4, a member of the Sp1-family of zinc finger transcription factors, is required for normal murine growth, viability, and male fertility. *Dev. Biol.* **176**, 284–299 (1996).
57. Zhou, X. *et al.* Reduced expression of the Sp4 gene in mice causes deficits in

sensorimotor gating and memory associated with hippocampal vacuolization. *Mol. Psychiatry* **10**, 393–406 (2005).

58. Zhou, X. *et al.* Reduced NMDAR1 expression in the Sp4 hypomorphic mouse may contribute to endophenotypes of human psychiatric disorders. *Hum. Mol. Genet.* **19**, 3797–3805 (2010).
59. Wenthold, R. J., Petralia, R. S., Blahos J, I. I. & Niedzielski, A. S. Evidence for multiple AMPA receptor complexes in hippocampal CA1/CA2 neurons. *J. Neurosci.* **16**, 1982–1989 (1996).
60. Malinow, R. & Malenka, R. C. AMPA receptor trafficking and synaptic plasticity. *Annu. Rev. Neurosci.* **25**, 103–126 (2002).
61. Bredt, D. S. & Nicoll, R. A. AMPA receptor trafficking at excitatory synapses. *Neuron* **40**, 361–379 (2003).
62. Wu, Y. *et al.* Mutations in ionotropic AMPA receptor 3 alter channel properties and are associated with moderate cognitive impairment in humans. *Proc. Natl. Acad. Sci. U. S. A.* **104**, 18163–18168 (2007).
63. Salpietro, V. *et al.* AMPA receptor GluA2 subunit defects are a cause of neurodevelopmental disorders. *Nat. Commun.* **10**, 3094 (2019).
64. Balu, D. T. The NMDA Receptor and Schizophrenia: From Pathophysiology to Treatment. *Adv. Pharmacol.* **76**, 351–382 (2016).
65. Javitt, D. C. & Zukin, S. R. Recent advances in the phencyclidine model of schizophrenia. *Am. J. Psychiatry* **148**, 1301–1308 (1991).
66. Krystal, J. H. *et al.* Subanesthetic effects of the noncompetitive NMDA antagonist, ketamine, in humans. Psychotomimetic, perceptual, cognitive, and neuroendocrine responses. *Arch. Gen. Psychiatry* **51**, 199–214 (1994).
67. Dalmau, J. & Graus, F. Antibody-Mediated Encephalitis. *N. Engl. J. Med.* **378**, 840–851 (2018).

68. Carvill, G. L. *et al.* GRIN2A mutations cause epilepsy-aphasia spectrum disorders. *Nat. Genet.* **45**, 1073–1076 (2013).

69. Strehlow, V. *et al.* GRIN2A-related disorders: genotype and functional consequence predict phenotype. *Brain* **142**, 80–92 (2019).

70. Garcia-Gonzalo, F. R. *et al.* Interaction between HERC1 and M2-type pyruvate kinase. *FEBS Lett.* **539**, 78–84 (2003).

71. Chong-Kopera, H. *et al.* TSC1 stabilizes TSC2 by inhibiting the interaction between TSC2 and the HERC1 ubiquitin ligase. *J. Biol. Chem.* **281**, 8313–8316 (2006).

72. Mashimo, T. *et al.* Progressive Purkinje cell degeneration in tambaleante mutant mice is a consequence of a missense mutation in HERC1 E3 ubiquitin ligase. *PLoS Genet.* **5**, e1000784 (2009).

73. Ikebuchi, K. *et al.* RB1CC1 activates the promoter and expression of RB1 in human cancer. *Int. J. Cancer* **125**, 861–867 (2009).

74. Chano, T. *et al.* RB1CC1 activates RB1 pathway and inhibits proliferation and cologenic survival in human cancer. *PLoS One* **5**, e11404 (2010).

75. Finn, R. D. *et al.* Pfam: the protein families database. *Nucleic Acids Res.* **42**, D222–30 (2014).

Figure and table legends

Figure 1. Study design and analytic approach.

A: Study design. Case-control and parent-proband trio sample sizes, variant classes, and analytical methods are described. The case-control stage is shown on the left, and the *de novo* mutation stage is shown on the right.
B: Principal components analysis of SCHEMA samples. 1000 Genomes samples with reported ancestry are plotted in the background, and SCHEMA samples are displayed in the foreground. For each global ancestry group, we report the number of cases and controls in the discovery data set in red and blue respectively, and the number of external controls in black. AFR: African, ASJ: Ashkenazi Jewish, AMR: Latin American, EAS: East Asian, EST: Estonian, FIN: Finnish, EUR: non-Finnish European, SAS: South Asian.
C: Case-control enrichment of ultra-rare protein-coding variants in genes intolerant of protein-truncating variants. *P* values displayed are from comparing the burden of variants of the labeled consequence in cases compared to controls. By definition, MPC enrichment is only shown for pLI > 0.9 genes. Bars represent the 95% CIs of the point estimates. pLI: probability of loss-of-function intolerant in the gnomAD database.
D: Enrichment of schizophrenia *de novo* mutations in *P* value bins derived from the Stage 1 (case-control) gene burden analysis. The enrichment *P* values displayed are calculated as a Poisson probability having equal or greater than the observed number of mutations given the baseline mutation rate in 3,402 schizophrenia trios. The relative rate is given by the ratio of observed to expected rate of *de novo* mutations. Bars represent the 95% CIs of the point estimates.

Figure 2. Results from the meta-analysis of ultra-rare coding variants in 3,402 trios, 24,248 cases, and 97,322 controls.

A: Manhattan plot. $-\log_{10}$ *P*-values are plotted against the chromosomal location of each gene. Genes reaching exome-wide significance are in red, and genes significant at FDR < 5% are in orange. Red dashed line: $P = 2.14 \times 10^{-6}$; Blue dashed line: FDR < 5%, or $P = 8.23 \times 10^{-5}$.
B: *Q-Q* plot. Observed $-\log_{10}$ *P*-values are plotted against expectation given a uniform distribution. Genes reaching exome-wide significance are plotted with a larger size. The direction of effect is indicated by the color of each point. Dark blue dashed line: $P = 2.14 \times 10^{-6}$; Light blue dashed line: FDR < 5%.

Figure 3. Biological insights from exome sequence data.

A**:** Common and rare allelic series at NMDA receptor subunit *GRIN2A*. The Locus Zoom plot (top) displays the common variant association of the gene. The color of each dot corresponds to the LD with the index SNP, and the properties of the index SNP are displayed. The gene plot (bottom) displays the protein-coding variants that contribute to the exome signal in *GRIN2A*. Variants discovered in cases are plotted above the gene, and those from control are plotted below. Each variant is colored based on inferred consequence, and the protein domains and missense constrained regions of the gene are also labelled[26,75]. The frequencies and counts in cases and controls are displayed for each variant class. AF: allele frequency, AC: allele count.
B: Temporal expression of *GRIN2A* in the human brain. We show *GRIN2A* expression in four prenatal and four postnatal periods derived from whole-brain tissue in BrainSpan[29]. The expression values plotted are in transcript-per-million (TPM).

Figure 4. Shared genetic signal with schizophrenia GWAS.

A: Case-control enrichment of ultra-rare protein-coding variants in genes prioritized from fine-mapping of the PGC schizophrenia GWAS[13]. The reported *P* value is the Fisher combined *P* value of Class I and Class II variants. Bars represent the 95% CIs of the point estimates.
B, C, D: Prioritization of GWAS loci using exome data. The Locus Zoom plot of three GWAS loci is displayed. For each gene in or adjacent to the region, we show the case-control counts of PTVs in the exome data, along with the meta-analysis *P*-value. *SP4, STAG1* and *FAM120A* are highlighted as the only genes with notable signals in the exome data within each locus.

Figure 5. Shared genetic signal between schizophrenia and other neurodevelopmental disorders.

A: Case-control enrichment of ultra-rare protein-coding variants in DD/ID and ASD-associated genes. We test for the burden of schizophrenia URVs in genes identified in the most recent exome sequencing studies of ASD and DD/ID[1,37]. The reported *P* value is the Fisher combined *P* value of Class I and Class II variants. Bars represent the 95% CIs of the point estimates.

B: Heatmap displaying the strength of association for schizophrenia-associated genes in our discovery data set and in genes implicated by *de novo* mutations in trios diagnosed with DD/ID. We display three groups of genes: Bonferroni significant in schizophrenia and DD/ID, Bonferroni significant only in schizophrenia, and FDR < 5% in schizophrenia and Bonferroni significant in DD/ID. The degree of association from each sequencing study is displayed as the color corresponding to $-\log_{10}$ *P*-values in that study. The case-control *P*-value is reported for schizophrenia, while *de novo* enrichment from a Poisson rate test is reported for DD/ID. Results are further stratified to tests of Class I (PTV and MPC > 3) and Class II (missense [MPC 2 - 3]) variants.

C: Allelic series in *TRIO* between schizophrenia and DD/ID risk variants. The gene plot displays the protein-coding variants that contribute to the exome signal in *TRIO*. Variants discovered in schizophrenia cases are plotted above the gene, and missense *de novo* mutations from DD/ID probands are plotted below. Each variant is colored based on inferred consequence, and the protein domains of the gene are also labelled. The variant counts are displayed for each variant class.

D: Allelic series in *GRIN2A.* See **D** for description.

Figure 6. The contributions of ultra-rare PTVs to schizophrenia risk.

A: Genetic architecture of schizophrenia. Significant genetic associations for schizophrenia from the most recent GWAS, CNV, and sequencing studies are displayed. The in-sample odds ratio is plotted against the minor allele frequency in the general population. The color of each dot corresponds to the source of the association, and the size of the dot to the odds ratio. The shaded area represented the loess-smoothed lines of the upper and lower bounds of the point estimates.

B: Case-control enrichment of ultra-rare protein-coding variants in genes intolerant of protein-truncating variants after excluding schizophrenia-associate genes. We perform the test with all constrained genes (pLI > 0.9) and after excluding all schizophrenia-associated genes with FDR < 5%. *P* values displayed are from comparing the burden of variants of the labelled consequence in cases compared to controls. Bars represent the 95% CIs of the point estimates.

Extended Figure Table 1. Case-control and *de novo* counts of the ten Bonferroni significant genes in the main analysis. Case-control counts displayed are the total counts for variants with minor allele count <= 5. PTV: protein-truncating variant, mis3: missense variants with MPC > 3, mis2: missense variants with MPC 2 - 3; Q value: adjusted *P* value after FDR adjustment; Class I: PTV and missense variants (MPC > 3); Class II: missense variants (MPC 2 - 3).

Acknowledgements

We would like to thank the patients and families who participated in our studies in the past two decades, without whom our research and findings would not be possible. Research reported in this publication was supported by the National Institute of Mental Health, and the National Human Genome Research Institute of the National Institutes of Health under award numbers: U01MH10564, U01MH105578, U01MH105666, U01MH109539, R01MH085548, R01MH085521, and U54HG003067. We would also like to acknowledge the generous support from the Stanley Family Foundation, Kent and Elizabeth Dauten, and The Dalio Foundation who have enabled us to rapidly expand our data generation collections with the goal of moving towards better treatments for schizophrenia and other psychiatric disorders. We wish to acknowledge all of the research participants in the BRIDGES cohort. This work was supported by NIMH (R01 MH094145; Michael Boehnke and Richard M.Myers, PIs and U01 MH105653 (Michael Boehnke, PI). The collection and storage of cases and controls from the Centre for Addiction and Mental Health (CAMH) in Toronto and from the Institute of Psychiatry, Psychology and Neuroscience (IoPPN), King's College London in London, U.K. was supported by funding from GlaxoSmithKline. CAMH was supported by the Canadian Institutes of Health Research (MOP-172013, PI John B. Vincent, CAMH). IoPPN was supported by funding from the National Institute for Health Research(NIHR) Biomedical Research Centre at South London and Maudsley NHS Foundation Trust and King's College London (IoPPN). The views expressed are those of the author(s) and not necessarily those of the UK NHS, the NIHR or the UK Department of Health. Case and control collection was supported by Heinz C. Prechter Bipolar Research Fund at the University of Michigan Depression Center to Melvin G. McInnis. Data and biomaterials were collected for the Systematic Treatment Enhancement Program for Bipolar Disorder (STEP-BD), a multi-center, longitudinal project selected from responses to RFP #NIMH-98-DS-0001, "Treatment for Bipolar Disorder" which was led by Gary Sachs and coordinated by Massachusetts General Hospital in Boston, MA with support from 2N01 MH080001-001. The Genomic Psychiatric Cohort (GPC) was supported by NIMH (U01 MH105641 (PI Carlos Pato), R01 MH085548 (PIs Carlos Pato and Michele Pato), R01MH104964 (PIs Carlos Pato and Michele Pato). The MCTFR study was supported through grants from the National Institutes of Health DA037904, DA024417, DA036216, DA05147, AA09367, DA024417,HG007022, and HL117626. The work at Cardiff University was supported by Medical Research Council Centre Grant No. MR/L010305/1 and Program Grant No. G0800509.

Authors

Tarjinder Singh[1,2], Timothy Poterba[1,2], David Curtis[3,4], Huda Akil[5], Mariam Al Eissa[6], Jack D. Barchas[5], Nicholas Bass[6], Tim B. Bigdeli[7], Gerome Breen[8], Evelyn J. Bromet[9], Peter F. Buckley[10], William E. Bunney[11,12], Jonas Bybjerg-Grauholm[13], William F. Byerley[5,14], Sinéad B. Chapman[2], Wei J. Chen[15], Claire Churchhouse[1,2], Nicholas Craddock[16], Charles Curtis[8], Caroline M. Cusick[2], Lynn DeLisi[5], Sheila Dodge[17], Michael A. Escamilla[18], Saana Eskelinen[19,20], Ayman H. Fanous[21], Stephen V. Faraone[22], Alessia Fiorentino[6], Laurent Francioli[1,23], Stacey B. Gabriel[17], Diane Gage[2], Sarah A. Gagliano Taliun[24], Andrea Ganna[1,25], Giulio Genovese[2], David C. Glahn[26], Jakob Grove[27], Mei-Hua Hall[28], Eija Hämäläinen[25], Henrike O. Heyne[1,2,29], Matti Holi[30], David M. Hougaard[13], Daniel P. Howrigan[1,2], Hailiang Huang[1,2], Hai-Gwo Hwu[31], René S. Kahn[32,33], Hyun Min Kang[24], Konrad J. Karczewski[1,2], George Kirov[34], James A. Knowles[35], Francis S. Lee[11,36], Douglas S. Lehrer[37], Francesco Lescai[38,39], Dolores Malaspina[32], Stephen R. Marder[32], Steven A. McCarroll[2,40], Helena Medeiros[21], Lili Milani[41], Christopher P. Morley[42], Derek W. Morris[43], Preben Bo Mortensen[44], Richard M. Myers[11,45], Merete Nordentoft[46,47], Niamh L. O'Brien[6], Ana Maria Olivares[2], Dost Ongur[28], Willem H. Ouwehand[48], Duncan S. Palmer[1,2], Tiina Paunio[49], Digby Quested[50], Mark H. Rapaport[51], Elliott Rees[34], Brandi Rollins[11,12], F. Kyle Satterstrom[2,52], Alan Schatzberg[5], Edward Scolnick[2], Laura J. Scott[24], Sally I. Sharp[6], Pamela Sklar, Jordan W. Smoller[53,54], Janet l. Sobell[55], Matthew Solomonson[23], Christine R. Stevens[2,23], Jaana Suvisaari[56], Grace Tiao[23], Stanley J. Watson[57], Nicholas A. Watts[23], Douglas H Blackwood[58], Anders D. Børglum[59,60], Bruce M. Cohen[28], Aiden P. Corvin[61], Tõnu Esko[41], Nelson B. Freimer[62], Stephen J. Glatt[22], Christina M. Hultman[63], Andrew McQuillin[6], Aarno Palotie[23,25], Carlos N. Pato[7], Michele T. Pato[7], Ann E. Pulver[64], David St. Clair[65], Ming T. Tsuang[66], Marquis P. Vawter[11,67], James T. Walters[34], Thomas M. Werge[68,69,70,71], Roel A. Ophoff[62,72], Patrick F. Sullivan[73,74], Michael J. Owen[34], Michael Boehnke[24], Michael C. O'Donovan[34], Benjamin M. Neale[1,2,23]*, Mark J. Daly[1,2,23,25]*

[1]Analytic and Translational Genetics Unit, Department of Medicine, Massachusetts General Hospital, Boston, Massachusetts, USA [2]Stanley Center for Psychiatric Research, Broad Institute of MIT and Harvard, Cambridge, Massachusetts, USA [3]UCL Genetics Institute, University College London, UK [4]Centre for Psychiatry, Queen Mary University London, UK [5]Pritzker Neuropsychiatric Disorders Research Consortium [6]Division of Psychiatry, University College London, UK [7]Department of Psychiatry & Behavioral Sciences, SUNY Downstate College of Medicine, Brooklyn, New York, USA [8]Social Genetic and Developmental Psychiatry; Institute of Psychiatry, Psychology and Neuroscience, King's College London, SE5 8AF, UK [9]Department of Psychiatry and Behavioural Health, Stony Brook University, HSC, Level T-10, Stony Brook, New York, USA [10]Department of Psychiatry, Virginia Commonwealth University, 1201 E Marshall St., Richmond, Virginia, USA [11]Pritzker Neuropsychiatric Disorders Research Consortium [12]University of California, Irvine, Department of Psychiatry and Human Behavior [13]iPSYCH, The Lundbeck Foundation Initiative for Integrative Psychiatric Research, Denmark; Center for Neonatal Screening, Department for Congenital Disorders, Statens Serum Institut, Copenhagen, Denmark. [14]Department of Psychiatry, University of California, SF, California, USA [15]College of Public Health; National Taiwan University; Taipei, Taiwan [16]National Centre for Mental Health, Cardiff University, UK [17]Genomics Platform, Broad Institute of MIT and Harvard, Cambridge, Massachusetts, USA [18]Texas Tech University Health Sciences Center El Paso, Texas, USA [19]Psychiatry, University of Helsinki and Helsinki University Hospital, Helsinki, Finland [20]Department of Public Health Solutions, Mental Health Unit, National Institute for Health and Welfare, Helsinki, Finland [21]Department of Psychiatry and Behavioral Sciences, SUNY Downstate Medical Center, Brooklyn, New York, USA [22]Department of Psychiatry and Behavioral Sciences; SUNY Upstate Medical University; Syracuse, New York, USA [23]Program in Medical and Population Genetics, Broad Institute of Harvard and MIT, Cambridge, Massachusetts, USA [24]Department of Biostatistics and Center for Statistical Genetics, University of Michigan, Ann Arbor, Michigan, USA [25]Institute for Molecular Medicine Finland, University of Helsinki, Helsinki, Finland [26]Department of Psychiatry, Boston Children's Hospital, Boston, Massachusetts, USA [27]Department of Biomedicine and iSEQ, Center for Integrative Sequencing, Aarhus University, Denmark; iPSYCH, The Lundbeck Foundation Initiative for Integrative Psychiatric Research, Denmark; Center for Genomics and Personalized Medicine, CGPM, Aarhus, Denmark; Bioinformatics Research Centre, Aarhus University, Aarhus, Denmark [28]McLean Hospital/Harvard Medical School, Belmont, Massachusetts, USA [29]Institute for Molecular Medicine Finland, University of Helsinki, Helsinki, Finland" [30]Department of Psychiatry,

Helsinki University Hospital, Helsinki University, Helsinki, Finland [31]Department of Psychiatry; National Taiwan University; Taipei, Taiwan [32]Department of Psychiatry, Icahn School of Medicine at Mount Sinai, New York, USA [33]MIRECC, JP Peters VA Hospital, Bronx, NY [34]MRC Centre for Neuropsychiatric Genetics and Genomics, Division of Psychological Medicine and Clinical Neurosciences, Cardiff University, UK. [35]Department of Cell Biology, SUNY Downstate Medical Center, Brooklyn, New York, USA. [36]Weill Cornell Medical College, New York, New York, USA [37]Department of Psychiatry, Wright State University, 3640 Colonel Gleen Hwy, Dayton, Ohio, USA [38]Department of Biomedicine, Aarhus University, Denmark; [39]iPSYCH, The Lundbeck Foundation Initiative for Integrative Psychiatric Research, Denmark [40]Department of Genetics, Harvard Medical School, Boston, Massachusetts, USA [41]Institute of Genomics, University of Tartu, Estonia [42]Departments of Public Health and Preventive Medicine, Family Medicine, and Psychiatry and Behavioral Sciences, State University of New York, Upstate Medical University, Syracuse, New York, USA [43]National University Ireland, Galway, Ireland [44]Aarhus University, Denmark [45]HudsonAlpha Institute for Biotechnology, Huntsville, AL 35806 [46]Copenhagen Research Center for Mental Health, Mental Health Services, Copenhagen University Hospital, Copenhagen, Denmark [47]Department of Clinical Medicine, University of Copenhagen, Copenhagen, Denmark. The Lundbeck Foundation Initiative for Integrative Psychiatric Research, iPSYCH, Copenhagen, Denmark. [48]University of Cambridge, Cambridge, UK [49]Department of Psychiatry, University of Helsinki, Helsinki, Finland [50]Oxford Health NHS Foundation Trust, Oxford, UK [51]Department of Psychiatry and Behavioral Sciences, Emory University, Atlanta, Georgia 30322, USA [52]Analytic and Translational Genetics Unit, Massachusetts General Hospital, Boston, Massachusetts, USA [53]Psychiatric and Neurodevelopmental Genetics Unit, Massachusetts General Hospital, Boston, Massachusetts, USA [54]Department of Psychiatry, Harvard Medical School, Boston, Massachusetts, USA [55]Department of Psychiatry and the Behavioral Sciences, Keck School of Medicine, University of Southern California, Los Angeles, California, USA [56]Finnish Institute for Health and Welfare, Helsinki, Finland [57]Department of Psychiatry, Michigan Neuroscience Institute, Medical School, University of Michigan, Michigan, USA [58]University of Edinburgh, Edinburgh, UK [59]Department of Biomedicine and iSEQ, Center for Integrative Sequencing, Aarhus University, Denmark [60]iPSYCH, The Lundbeck Foundation Initiative for Integrative Psychiatric Research, Denmark; Center for Genomics and Personalized Medicine, CGPM, Aarhus, Denmark [61]Trinity College Dublin, Dublin, Ireland [62]Center for Neurobehavioral Genetics, University of California, Los Angeles, California, USA [63]Department of Medical Epidemiology and Biostatistics, Karolinska Institutet, Stockholm, Sweden [64]School of Medicine, Johns Hopkins University, Baltimore, Maryland, USA [65]University of Aberdeen, Aberdeen, UK [66]Center for Behavioral Genomics, Department of Psychiatry; University of California, San Diego; La Jolla, California, USA [67]University of California, Irvine, Department of Psychiatry and Human Behavior, Irvine, California, USA [68]Institute of Biological Psychiatry, Mental Health Services, Copenhagen University Hospital, Copenhagen, Denmark [69]Department of Clinical Medicine, University of Copenhagen, Copenhagen, Denmark [70]Center for GeoGenetics, GLOBE Institute, University of Copenhagen, Copenhagen, Denmark [71]The Lundbeck Foundation Initiative for Integrative Psychiatric Research, iPSYCH, Copenhagen, Denmark. [72]Erasmus Medical Center, Erasmus University, Rotterdam, The Netherlands [73]Karolinska Institute, Sweden [74]Univeristy of North Carolina Hospitals, North Carolina, USA

*These authors contributed equally to this work

Correspondence should be addressed to: M.J.D. (mjdaly@atgu.mgh.harvard.edu), B.M.N. (bneale@broadinstitute.org), T.S. (tsingh@broadinstitute.org)

Author Contributions

T.S., L.J.S., M.B., B.M.N., M.J.D Conceived and designed the experiments
T.S., S.A.G.T., A.G., D.P.H., H.H., B.R., F.K.S., L.J.S., J.T.W., M.J.O., M.B., M.C.O., B.M.N., M.J.D Analysis design and execution
H.A., J.D.B., W.E.B., S.B.C., C.C., C.M.C., S.B.G., D.G., F.S.L., P.B.M., R.M.M., A.M.O., A.S., C.R.S., N.A.W., J.T.W., M.J.O., M.C.O., B.M.N Contributed to project management/sequencing
D.C., M.A.E., N.B., G.B., W.F.B., L.D., S.V.F., A.F., M.H., F.L., A.M.M., P.B.M., N.L.O., D.Q., E.R., S.I.S., D.H.B., A.D.B., B.M.C., T.E., S.J.G., A.M., A.P., M.P.V., J.T.W., P.F.S., M.J.O., M.C.O Recruited/assessed/contributed patient samples
T.S., T.B.B., E.J.B., P.F.B., L.F., K.J.K., A.M.M., D.S.P., M.S Contributed reagents/materials/analysis tools
T.S., D.C., A.M.M., L.J.S., J.T.W., M.J.O., M.C.O., B.M.N., M.J.D Wrote/edited the paper

Ethics declarations

IRB approvals and study consent forms from each of the sample-contributing organizations were sent to the Broad Institute before samples were sequenced and analyzed.

All ethical approvals are on file at the Massachusetts General Brigham (MGB), formerly Partners, IRB office amended to protocol #2014P001342, title: 'Molecular Profiling of Psychiatric Disease'.

Competing Interests

M.J.D. is a founder of Maze Therapeutics. B.M.N. is a member of the scientific advisory board at Deep Genomics and RBNC Therapeutics, Member of the scientific advisory committee at Milken and a consultant for Camp4 Therapeutics, Merck and Biogen. A.P. is a member of Astra Zeneca's Genomics Advisory Board. M.C.O, M.J.O, and J.T.W. are supported by a collaborative research grant from Takeda Pharmaceuticals. D.S.P. was an employee of Genomics plc, all analyses reported in this paper were performed as part of D.S.P.'s employment at the Massachusetts General Hospital and Broad Institute. The remaining authors declare no competing interests.

Data availability

We describe all datasets in the manuscript or Supplementary Information. We provide summary-level data at the variant and gene level in an online browser for viewing and download (https://schema.broadinstitute.org). There are no restrictions on the aggregated data released on the browser. For contributing data sets that are permitted to be distributed at the individual level, we have deposited the data in a public repository (the database of Genotypes and Phenotypes [dbGAP] and/or the European Genome-phenome Archive [EGA]) and provide the accessions in Table S1.

Code availability

Software and code used are described throughout the Supplementary Methods of the manuscript. In brief, for sequence data generation, we used GATK v3.4 and v3.6, Picard version 1.1431, and VerifyBamID version 1.0.0. Sample, variant QC, and analyses were performed using Hail 0.1 and 0.2 (https://hail.is/), with functions and arguments referred to in the Supplementary Methods. Wrappers and methods using Hail code can be found at https://github.com/TarjinderSingh/hailutils. Additional (basic) processing and visualization was performed using base R (v3.6) with tidyverse libraries (https://www.tidyverse.org/packages/).