SCISPACE
formerly Typeset

Open access • Posted Content • DOI:10.1101/2020.09.12.20192922

Mapping genomic loci prioritises genes and implicates synaptic biology in schizophrenia — Source link

Stephan Ripke, James T.R. Walters, Michael Conlon O'Donovan

Institutions: Charité, Cardiff University

Published on: 13 Sep 2020 - medRxiv (Cold Spring Harbor Laboratory Press)

Topics: Developmental disorder and Schizophrenia

Related papers:

- Biological insights from 108 schizophrenia-associated genetic loci
- Genome-wide association study identifies 30 loci associated with bipolar disorder
- Identification of common genetic risk variants for autism spectrum disorder
- Discovery of the first genome-wide significant risk loci for attention deficit/hyperactivity disorder
- Genome-wide association analyses identify 44 risk variants and refine the genetic architecture of major depression

Share this paper:

View more about this paper here: https://typeset.io/papers/mapping-genomic-loci-prioritises-genes-and-implicates-1klan2wiz1

medRxiv preprint doi: https://doi.org/10.1101/2020.09.12.20192922; this version posted September 13, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. All rights reserved. No reuse allowed without permission.

Mapping genomic loci prioritises genes and implicates synaptic biology in schizophrenia

Authors

The Schizophrenia Working Group of the Psychiatric Genomics Consortium

Corresponding Authors

Stephan Ripke[1,2], James TR Walters[3], Michael C O'Donovan[3]

Affiliations

1. Dept. of Psychiatry and Psychotherapy, Charité - Universitätsmedizin, Berlin 10117, Germany
2. Analytic and Translational Genetics Unit, Massachusetts General Hospital, Boston MA 02114, USA
3. MRC Centre for Neuropsychiatric Genetics and Genomics, Division of Psychiatry and Clinical Neurosciences, Cardiff University, Hadyn Ellis Building, Maindy Road, Cardiff, CF24 4HQ

Contact

sripke@broadinstitute.org, waltersjt@cf.ac.uk, odonovanmc@cf.ac.uk

NOTE: This preprint reports new research that has not been certified by peer review and should not be used to guide clinical practice.

medRxiv preprint doi: https://doi.org/10.1101/2020.09.12.20192922; this version posted September 13, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. All rights reserved. No reuse allowed without permission.

SUMMARY

Schizophrenia is a psychiatric disorder whose pathophysiology is largely unknown. It has a heritability of 60-80%, much of which is attributable to common risk alleles, suggesting genome-wide association studies can inform our understanding of aetiology[1]. Here, in 69,369 people with schizophrenia and 236,642 controls, we report common variant associations at 270 distinct loci. Using fine-mapping and functional genomic data, we prioritise 19 genes based on protein-coding or UTR variation, and 130 genes in total as likely to explain these associations. Fine-mapped candidates were enriched for genes associated with rare disruptive coding variants in people with schizophrenia, including the glutamate receptor subunit *GRIN2A* and transcription factor *SP4*, and were also enriched for genes implicated by such variants in autism and developmental disorder. Associations were concentrated in genes expressed in CNS neurons, both excitatory and inhibitory, but not other tissues or cell types, and implicated fundamental processes related to neuronal function, particularly synaptic organisation, differentiation and transmission. We identify biological processes of pathophysiological relevance to schizophrenia, show convergence of common and rare variant associations in schizophrenia and neurodevelopmental disorders, and provide a rich resource of priority genes and variants to advance mechanistic studies.

INTRODUCTION

Schizophrenia is a clinically heterogeneous disorder, which typically manifests in late adolescence or early adulthood[1]. It is associated with elevated risk of suicide[2] and serious physical illnesses[3], reduced life expectancy, and substantial health and social costs. Treatments are at least partially effective in most people, but many have chronic symptoms, and adverse treatment effects are common[4]. There is a need for novel therapeutic target discovery, a process impeded by our limited understanding of pathophysiology.

medRxiv preprint doi: https://doi.org/10.1101/2020.09.12.20192922; this version posted September 13, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. All rights reserved. No reuse allowed without permission.

Much of the variation in risk between individuals is genetic involving alleles that span the full range of frequencies, including large numbers of common alleles[5] as well as rare copy number variants (CNVs)[6] and rare coding variants (RCVs)[7,8]. A recent genome-wide association study (GWAS) reported 176 genomic loci containing common alleles associated with schizophrenia[9] but the causal variants driving these associations and the biological consequences of these variants are largely unknown. To increase our understanding of the common variant contribution to schizophrenia, we performed the largest GWAS to date and analysed the findings to prioritise variants, genes and biological processes that contribute to pathogenesis.

RESULTS

Association Meta-Analysis

The primary GWAS was performed on 90 cohorts including 67,390 cases and 94,015 controls (161,405 individuals; equivalent in power to 73,189 each of cases and controls). This was a trans-ancestry analysis with ~80% of the sample of European ancestry and 20% of East Asian ancestry (**Sample Supplementary Note**). After identical data processing protocols, we conducted within cohort association analyses followed by meta-analysis of 7,585,078 SNPs with MAF ≥ 1%. We identified 294 independent SNPs (linkage disequilibrium (LD) r2 < 0.1)

medRxiv preprint doi: https://doi.org/10.1101/2020.09.12.20192922; this version posted September 13, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity.
All rights reserved. No reuse allowed without permission.

that exceeded genome-wide significance ($p<5x10^{-8}$) **(Figure 1; Supplementary Table 1**).

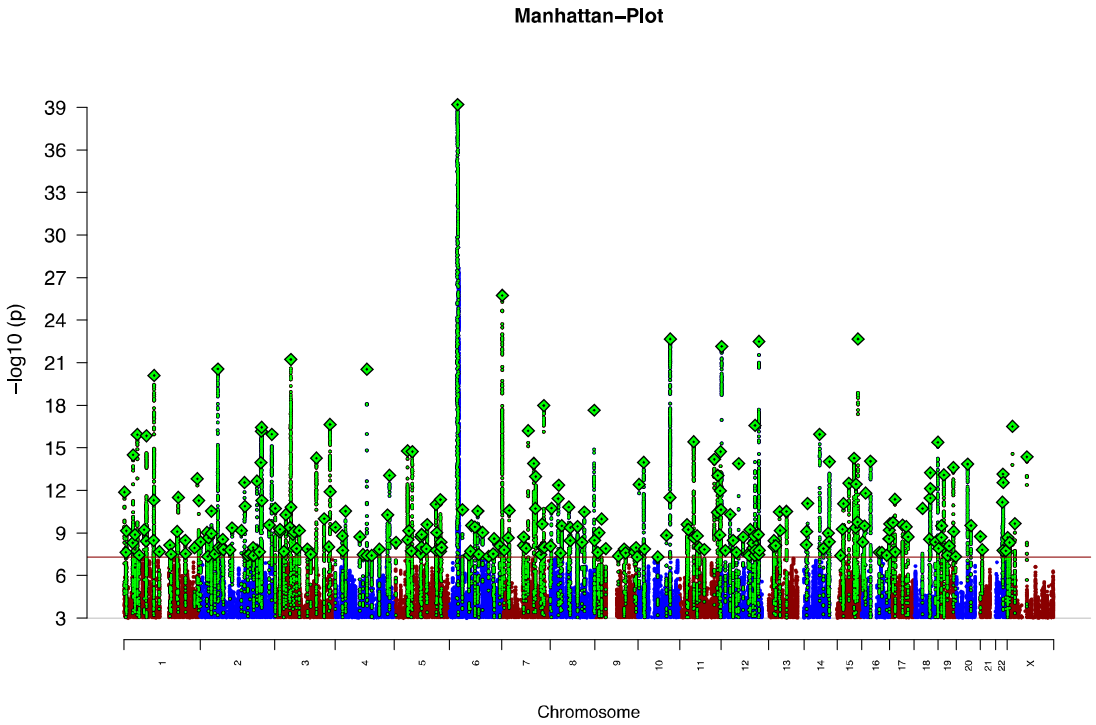
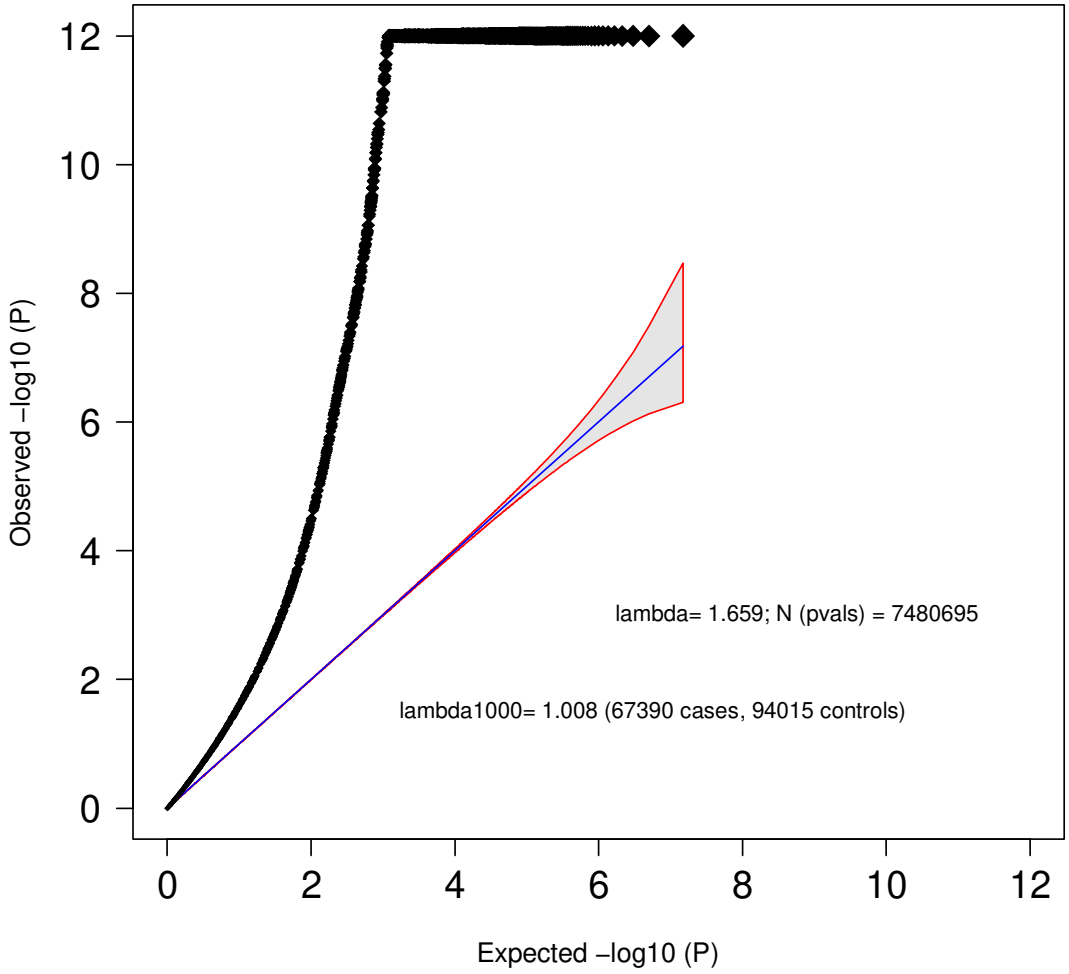


Figure 1: ***Discovery GWAS Manhattan plot****. The x-axis indicates chromosomal position and the y-axis is the significance of association (−log10(P)). The red line represents genome-wide significance level (5×10^{-8}). SNPs in green are in linkage disequilibrium (LD; R2 >0.1) with index SNPs (diamonds) which represent independent genome-wide significant associations.*

As expected[10], we observed substantial genome-wide test-statistic inflation above the null **(Extended Data Figure 1);** at least 90% of this is due to polygenicity **(Supplementary**

medRxiv preprint doi: https://doi.org/10.1101/2020.09.12.20192922; this version posted September 13, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. All rights reserved. No reuse allowed without permission.

Note) rather than confounders.



Extended Data Figure 1: ***Discovery GWAS Quantile-quantile plot****. The x-axis shows the expected −log10(P) values for association under the null distribution given number of independent tests. The y-axis denotes the observed −log10(P). We truncate the Y-axis at −log10(P)=12. The shaded area surrounded by a red line indicates the 95% confidence interval under the null. Lambda is the observed median χ2 test statistic divided by the median expected χ2 test statistic under the null.*

For index SNPs with $P<10^{-5}$, we obtained summary association statistics from deCODE Genetics (1,979 cases, 142,627 controls), the results of which *en masse* were consistent with those from our primary GWAS (**Supplementary Note**). Meta-analysis with deCODE identified 329 LD-independent significant SNPs **(Supplementary Table 2)**. These were located in 270 loci (i.e. distinct regions of the genome; **Supplementary Table 3; Supplementary Figures 1-2**). Comparisons with the 128 associations (108 loci) we reported

medRxiv preprint doi: https://doi.org/10.1101/2020.09.12.20192922; this version posted September 13, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity.
All rights reserved. No reuse allowed without permission.

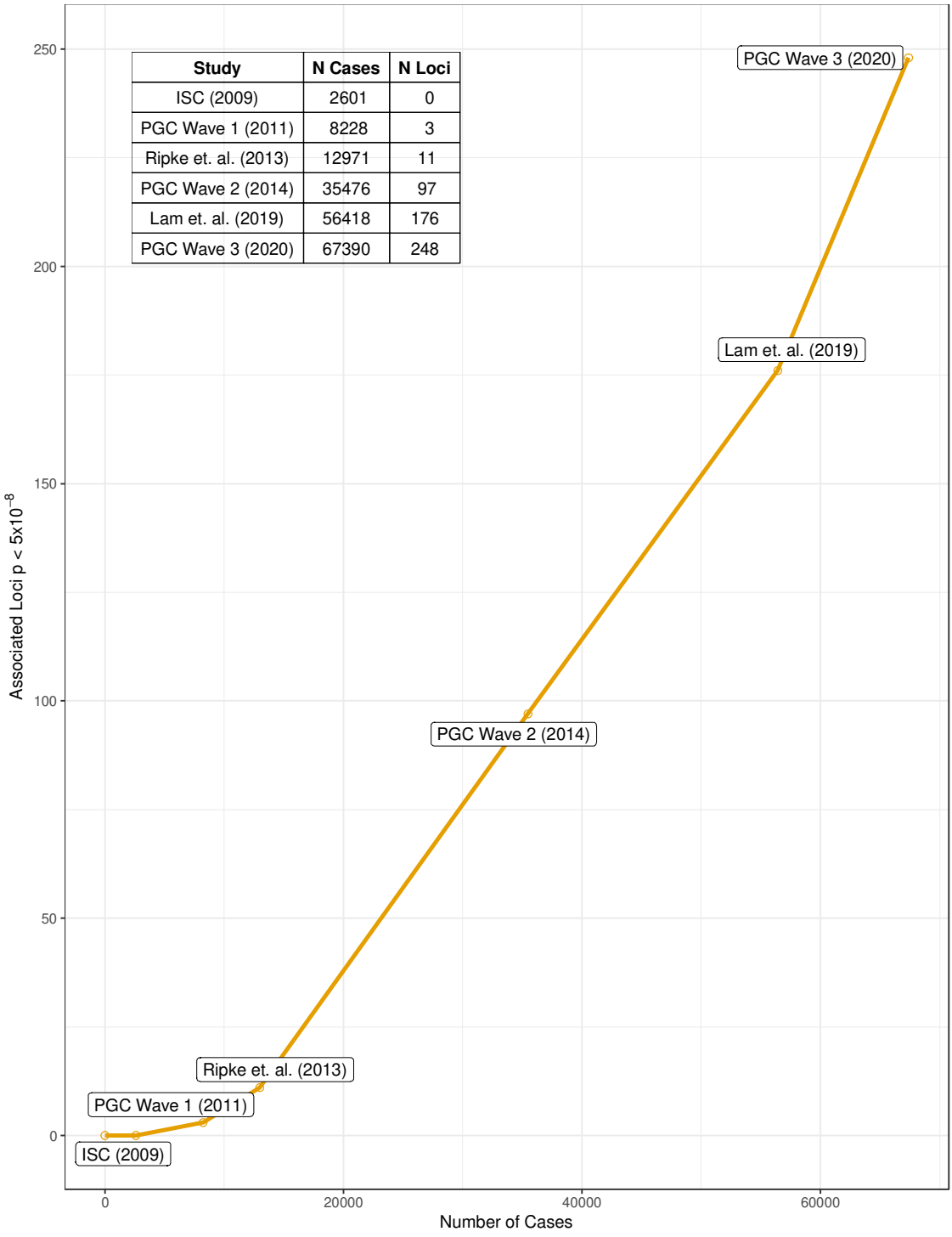
89 in 2014[11] are provided in the supplementary note, and the trajectory of associated loci by

90 sample size is depicted in **Extended Data Figure 2;** of note, one previously reported

91 association at rs3768644 (chr2:72.3Mb) is no longer supported[11].

92

medRxiv preprint doi: https://doi.org/10.1101/2020.09.12.20192922; this version posted September 13, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity.
All rights reserved. No reuse allowed without permission.



Extended Data Figure 2: ***GWAS progress over time****. The relationship of GWAS associations to sample-size is shown in this plot with selected SCZ GWAS meta analyses of the past 11 years. The x-axis shows number of cases. The y-axis shows the number of independent loci discovered with at least one genome-wide significant index SNP in the discovery meta-analysis (e.g. without replication data). The publications listed are ISC (2009)[5], PGC Wave 1 (2011)[21], Ripke et. al. (2013)[52], PGC Wave 2 (2014)[11], Lam et. al. (2019)[9] PGC Wave 3 (2020) - this manuscript. The slope of ~4 newly discovered loci per 1000 cases between 2013 and 2019 increased to a slope of ~6 with the latest sample-size increase.*

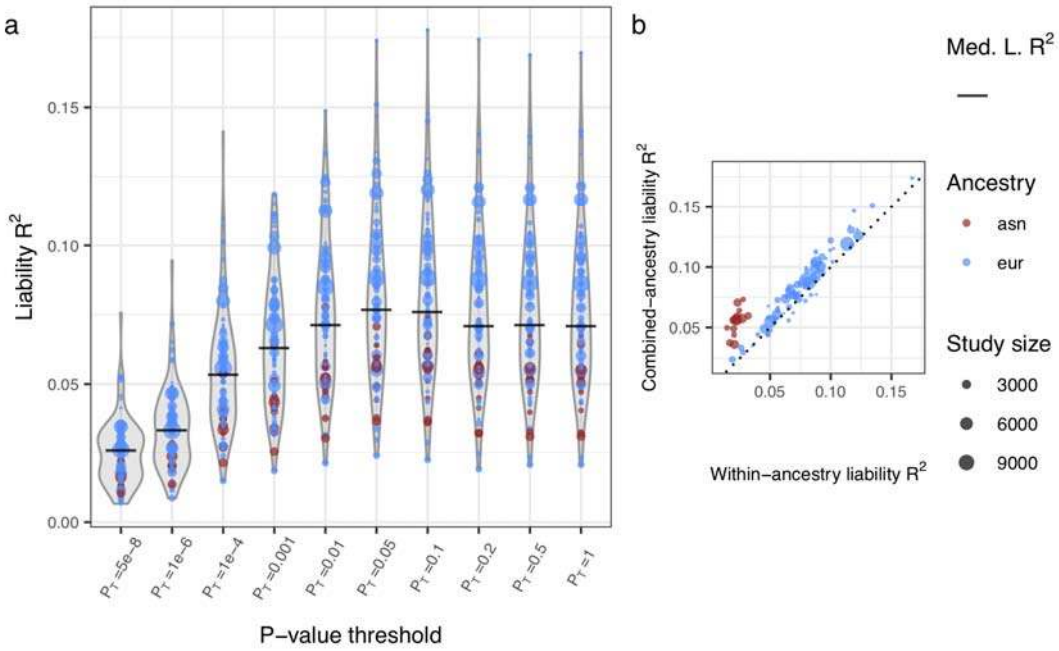
medRxiv preprint doi: https://doi.org/10.1101/2020.09.12.20192922; this version posted September 13, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. All rights reserved. No reuse allowed without permission.

The genome-wide significant index SNPs from the primary analyses showed no significant evidence for heterogeneity by sex (**Supplementary Table 4**), and separate GWAS for males and females had a genetic correlation not significantly different from 1 (r_g=0.992, SE 0.024). These and other analyses (**Supplementary Note**) show that common variant genetic liability to schizophrenia is essentially identical in males and females despite well-established sex differences in age at onset, symptom profile, course, and outcome[12].

Heritability and Polygenic Prediction

The proportion of variance in liability attributable to all measured SNPs, the SNP-based heritability (h^2_{SNP}), was estimated[13] to be 0.24 (SE 0.007). Polygenic risk score (PRS) analysis **(Supplementary Note)** explained up to 0.077 of variance in liability (using SNPs with GWAS p-value less than 0.05), and 0.026 when restricted to genome-wide significant SNPs (**Extended Data Figure 5**).

medRxiv preprint doi: https://doi.org/10.1101/2020.09.12.20192922; this version posted September 13, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. All rights reserved. No reuse allowed without permission.



Extended Data Figure 5: ***Polygenic risk prediction performance****. A) Distributions of liability R^2 across 89 left-out-cohorts for polygenic risk scores built from SNPs with different p-value thresholds. Distributions of liability R^2 (assuming a prevalence of 1%) are shown for each p-value threshold, with point size representing size of the left-out cohort and colour representing ancestry. The sample-size-weighted-mean liability R^2 is represented as a bar for each method. B) Liability R^2 of predicted and observed phenotypes in left-out cohorts using threshold p=0.05. The polygenic risk scores that give the x-axis R^2 are derived from two separate leave-one-out GWAS meta-analyses: one with cohorts of Asian ancestry (with results combined across left-out cohorts), and another with cohorts of European ancestry. The polygenic risk scores that give the y-axis R^2 are those from A.*

For almost all European and Asian cohorts, PRS had more explanatory power using risk alleles derived from the full combined ancestry GWAS than from the matched ancestry GWAS. PRS explained more variance in liability in cohorts of European ancestry (likely a result of the ancestry composition of the GWAS) but also in samples which by ascertainment are likely to include the most severe cases (i.e. hospitalized patients including those treated with clozapine) **(Supplementary Figure 4, Supplementary Note).**

A fuller discussion of heritability and polygenic prediction is provided in the **Supplementary Note**; importantly, the liability captured by PRS is insufficient for predicting diagnosis in the general population, with an average Area Under the Receiver Operating Characteristic Curve

medRxiv preprint doi: https://doi.org/10.1101/2020.09.12.20192922; this version posted September 13, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. All rights reserved. No reuse allowed without permission.

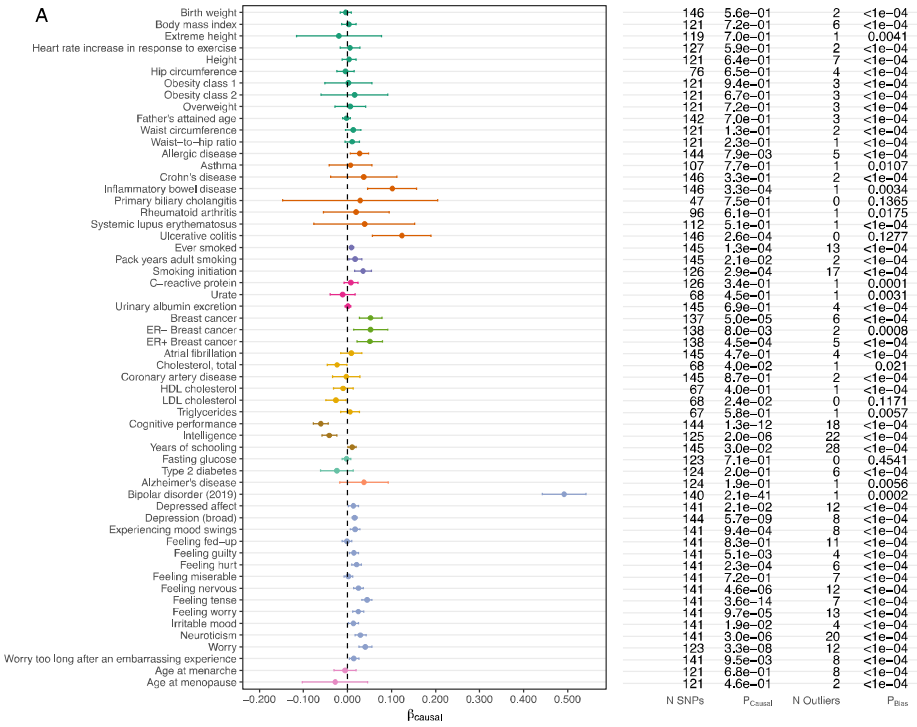
(AUC) of 0.71. Nevertheless, as a quantitative estimate of liability to schizophrenia, PRS has many applications in research settings, for example for patient stratification, or for identifying correlates of liability in population samples[14]. In those contexts, PRS indexes substantial differences in liability between individuals; in European samples, compared to the lowest centile of PRS, the highest centile of PRS has an OR for schizophrenia of 44 (95% CI=31-63), and 7.0 (CI 5.8-8.3) when the top centile is compared with the remaining 99% of individuals (**Supplementary Table 5**).

Mendelian randomisation analyses

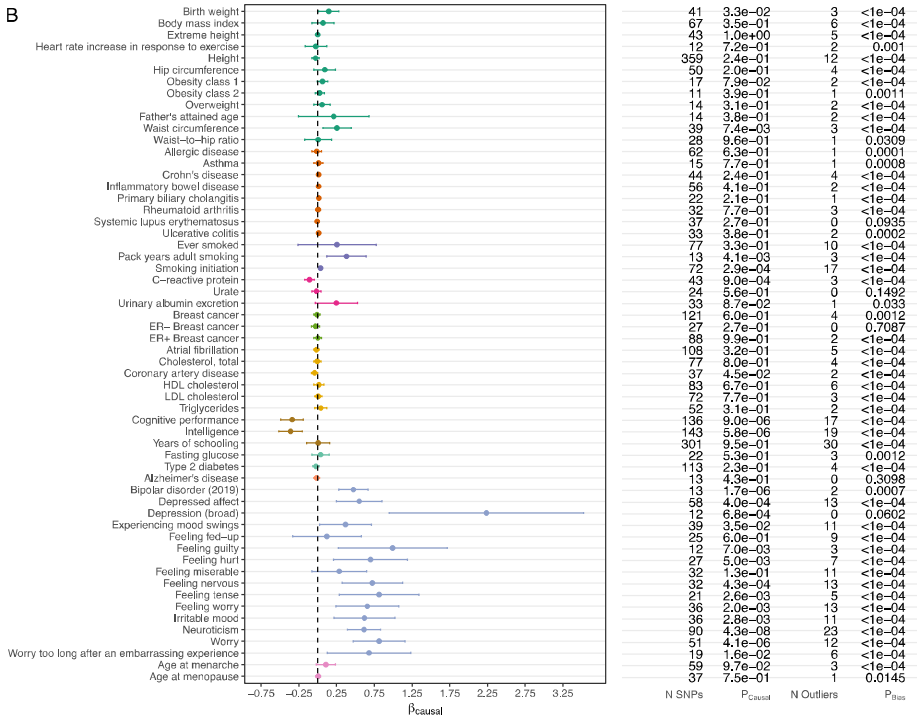
Schizophrenia is genetically correlated with several psychiatric, cognitive, and behavioural phenotypes[15]. It is also associated with several non-CNS related traits of clinical importance, including cardiovascular disorders, metabolic syndrome[3,16], and some autoimmune disorders[17]. The reasons for these associations are unclear, but they may point to traits with causal influences on SZ which could provide clues to prevention. We used 2-sample Mendelian randomisation[18] using summary-level genomic data for 58 traits to test hypotheses of causal relationships in both directions, that is, where schizophrenia could be a cause or a consequence of the other trait. **(Methods, Supplementary Table 6, Supplementary Note)**.

As reported for complex disorders[16] we found evidence that alleles with effects on schizophrenia liability frequently increase liability to other traits but that generally these effects are independent of each other (a phenomenon called horizontal pleiotropy) indicated by P_{Bias} in **Extended Data Figure 3**.

medRxiv preprint doi: https://doi.org/10.1101/2020.09.12.20192922; this version posted September 13, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity.
All rights reserved. No reuse allowed without permission.



159



160
161 *Extended Data Figure 3:* ***Relationships between schizophrenia and various complex traits and diseases estimated with bi-***
162 ***directional Mendelian randomization****. Panel (a): schizophrenia as the exposure. Panel (b): schizophrenia as the outcome.*
163 *N SNPs: number of SNPs used as instrumental variables.* P_{causal}*: p-value for causal effects from MR-PRESSO*[16] *after*
164 *removing SNPs with potential pleiotropic bias. N Outliers: number of SNPs with potential pleiotropic bias identified by MR-*
165 *PRESSO.* P_{bias}*: p-value for MR-PRESSO global test for pleiotropic bias. All tests required a minimum of 10 SNPs for the*
166 *exposure. All GWAS summary statistics include at least 1,000,000 SNPs and 10,000 samples. Traits of similar types are*
167 *given the same colour.*

medRxiv preprint doi: https://doi.org/10.1101/2020.09.12.20192922; this version posted September 13, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. All rights reserved. No reuse allowed without permission.

After accounting for pleiotropic bias, 22 of the exposure/outcome tests showed evidence of correlated SNP effect sizes (corrected for 116 tests, $P<4.31x10^{-4}$) (**Extended Data Figure 3**), between schizophrenia and a range of other psychiatric, cognitive or behavioural traits, including bipolar disorder, depression phenotypes, smoking initiation and cognition. However, analyses suggest that correlated genetic effect sizes are not consistent with clear causal relationships between any pair of these traits, and hence are more likely to represent shared aspects of biology (consistent with horizontal pleiotropy). Moreover, as many of these phenotypes also have overlapping features, partial sharing of genetic effects suggests that there may be underlying dimensional traits that are common features of, or increase liability to, these traits rather than one being necessarily causal for another. We did, however, find evidence suggesting schizophrenia increases liability to ulcerative colitis and, as reported before[19], breast cancer, suggesting the need for studies capable of identifying the potential mechanisms behind these relationships which potentially could be preventable.

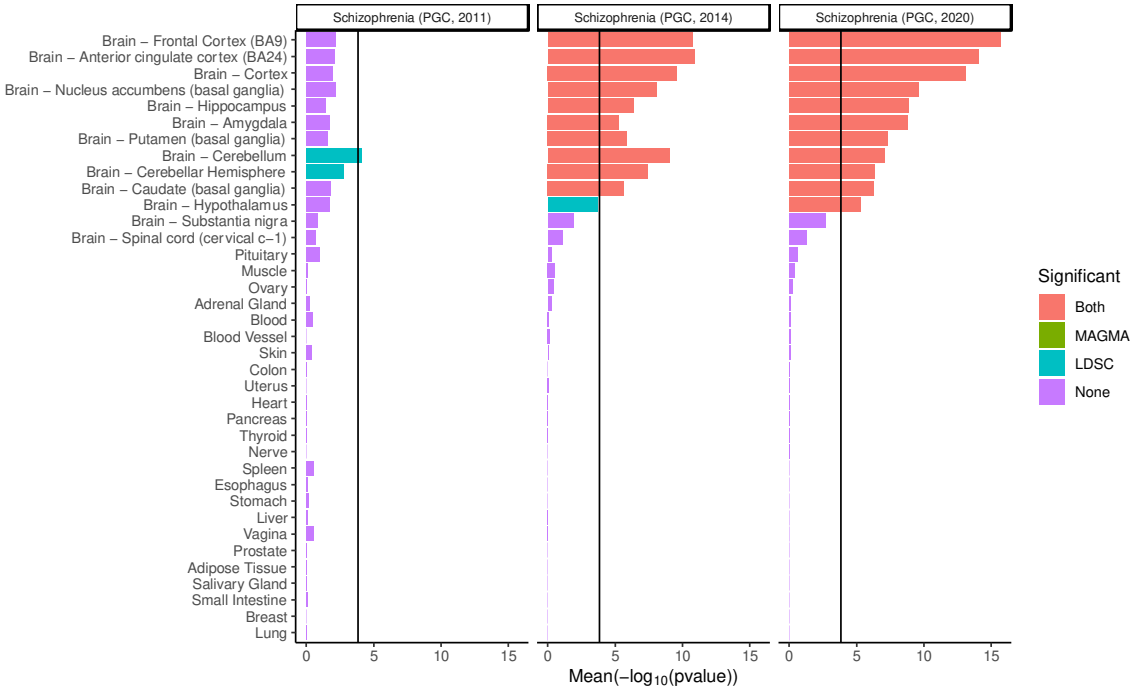
Gene Set Enrichments

To inform hypotheses about aetiopathogenesis we tested for enrichment of associations in sets of genes defined by their expression in tissues, cell types, cell compartments, or annotation to cellular compartments. These analyses test the importance of the characteristics used to define gene sets for disease aetiology, but they do not speak to the involvement or not of individual genes.

medRxiv preprint doi: https://doi.org/10.1101/2020.09.12.20192922; this version posted September 13, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. All rights reserved. No reuse allowed without permission.

Tissue and cell types

Genes with high relative specificity for bulk expression in many regions of human brain[20] were strongly enriched for schizophrenia associations **(Extended Data Figure 4)**.



Extended Data Figure 4: ***Association between 37 human tissues and schizophrenia.*** *The mean of strength of association evidence using two enrichment methods ($-log_{10}P_{MAGMA}$, $-log_{10}P_{LDSC}$) derived from bulk RNA-seq[20], is shown for each tissue. The bar colour indicates whether the cell type is significantly associated with both methods (i.e. MAGMA and LDSR), one method or none. The black vertical bar represents the significance threshold corrected for the total number of tissues/cell types tested in this study (P = 0.05/341). We also analysed previous waves of PGC schizophrenia GWAS[11,21] for comparison.*

Comparison with our earlier studies[11,21] shows increasingly clear contrast between the enrichments in brain and non-brain tissues as sample size increases. Consistent with, but stronger than, previous studies[22], in single cell expression data[23], we found associations were enriched in genes with high expression in human cortical inhibitory interneurons and excitatory neurons from cerebral cortex and hippocampus (pyramidal and granule cells) **(Figure 2a)**. From mouse single-cell RNA-seq data[22], genes highly expressed in excitatory pyramidal neurons from the cortex and hippocampus were strongly associated with schizophrenia **(Figure 2b)** as were those in inhibitory cortical and inhibitory medium spiny

medRxiv preprint doi: https://doi.org/10.1101/2020.09.12.20192922; this version posted September 13, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity.
All rights reserved. No reuse allowed without permission.

neurons, the latter being the predominant cells of the striatum.

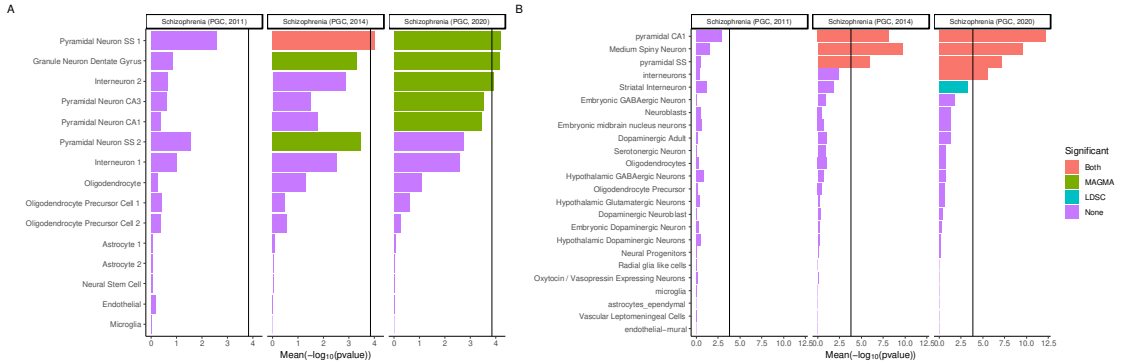


Figure 2: ***Associations between schizophrenia and cell types from multiple brain regions in human and mouse.*** *The mean of strength of association evidence using two enrichment methods ($-log_{10}P_{MAGMA}$, $-log_{10}P_{LDSC}$) between gene expression specificity and schizophrenia is shown (**A**) for 15 human cell types (derived from single nuclei) from the cortex and hippocampus (**B**) and for 24 cell types (derived from single cell RNA-seq) from five different brain regions in mouse (cortex, hippocampus, striatum, midbrain and hypothalamus) and from specific enrichments for oligodendrocytes, serotonergic neurons, dopaminergic neurons and cortical parvalbuminergic interneurons. The bar colour indicates whether the cell type is significantly associated with both methods (MAGMA and LDSC), one method or none. The black vertical bar represents the significance threshold corrected for the total number of tissues/cell types tested in this study (P = 0.05/341). Results obtained for previous iterations of schizophrenia GWAS[11,21] are shown for comparison. Pyramidal SS: pyramidal neurons from the somato-sensory cortex, Pyramidal CA1: pyramidal neurons from the CA1 region of the hippocampus, Pyramidal CA3: pyramidal neurons from the CA3 region of the hippocampus. Where types of cell (e.g. interneuron) formed sub-clusters in the source data, these are designated by the suffix 1 or 2.*

Across 265 cell types in the mouse central and peripheral nervous system[24], the strongest enrichments were for genes expressed in glutamatergic neurons located in the deep layers of cortex, amygdala, and hippocampus **(Supplementary Figure 5),** although enrichments were seen for inhibitory and excitatory neuronal populations more widely (e.g., excitatory neurons from the midbrain, thalamus and hindbrain). Significant enrichments for association signals were not observed for genes with highly specific expression in glia or microglia. Overall, the findings are consistent with the hypothesis that schizophrenia is primarily a disorder of neuronal function, but do not suggest pathology is likely restricted to a circumscribed brain region.

Biological Processes

Of 7,083 gene ontology (GO) classifications curated to include only annotations with experimental or phylogenetic support, 27 were associated (corrected for multiple testing)

medRxiv preprint doi: https://doi.org/10.1101/2020.09.12.20192922; this version posted September 13, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. All rights reserved. No reuse allowed without permission.

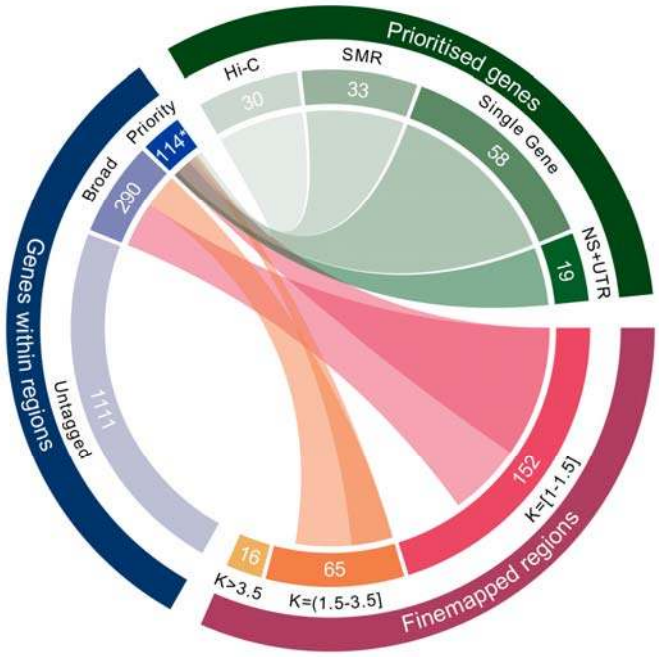
with schizophrenia (**Supplementary Table 7**). Stepwise conditional analysis to minimise redundancy due to genes present in multiple classifications identified 9 associations, all related to neuronal excitability, development, and structure, with prominent enrichment at the synapse. Using the expert-curated synaptic ontology from the SynGO consortium[25], we found conditionally significant annotations were mainly to postsynaptic terms (**Supplementary Tables 8, 9**), although enrichments were also found for trans-synaptic signalling and synaptic organisation.

Fine-mapping

We performed stepwise analyses, conditioning associations in loci on their index SNP (and any subsequent conditionally independent associations) to identify regions that contained independent signals (conditional $P<10^{-6}$) **(Supplementary Note and Supplementary Table 10)**. Where there were independent associations to non-overlapping regions in a locus, we fine-mapped these separately. We used FINEMAP[26] to assign posterior probabilities (PP) of being causal to SNPs, and constructed credible sets of SNPs that cumulatively capture 95% of the regional PP. We then focussed on loci (**Supplementary Table 3)** that were genome-wide significant in the primary dataset and that were predicted to contain 3 or fewer causal

medRxiv preprint doi: https://doi.org/10.1101/2020.09.12.20192922; this version posted September 13, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. All rights reserved. No reuse allowed without permission.

252 variants **(N=217; Figure 3; Supplementary Tables 11a, 11b)**.



253

254 *Figure 3:* ***FINEMAP results and overview of the gene prioritisation analysis.*** *Fine-mapped GWAS regions and number of*
255 *expected causal variants (K) inferred by FINEMAP. Links are displayed to indicate how the results of FINEMAP were used*
256 *to select sets of protein-coding genes within associated GWAS loci for further analyses. Two gene sets are shown: "Broad"*
257 *genes contain at least one FINEMAP credible SNP, while "Priority" genes meet the prioritisation criteria described in the*
258 *text. The combined total of 114 protein-coding prioritised genes includes 3 that are outside of the 233 fine-mapped regions,*
259 *but which have expression QTLs within those regions (FOXN2) and Hi-C interactions containing FINEMAP credible SNPs*
260 *(ZCCHC7, ALG12). Criteria for defining prioritised genes are also shown. "NS+UTR": Genes containing a FINEMAP*
261 *credible nonsynonymous or UTR SNP with posterior probability greater than 10%. "Single Gene": Genes in which the*
262 *entire FINEMAP credible SNP set was contained within one gene. "SMR": protein coding genes in which SMR analysis*
263 *points towards a single gene at a locus, or where the credible eQTL causal SNPs capture >50% of the FINEMAP posterior*
264 *probability. "Hi-C": protein coding genes with SMR evidence, and where enhancer-promoter or promoter-promoter*
265 *interaction exists between segments of DNA containing FINEMAP credible SNPs. Note these criteria are not mutually*
266 *exclusive and thus individual genes might fulfil more than one criterion (**Supplementary Table 20**). For clarity, links to*
267 *genes within associated GWAS loci that were not tagged by FINEMAP credible SNPs are not depicted.*
268

269 For 32 loci, the 95% credible set contained 5 or fewer SNPs (**Supplementary Table 11c)** and

270 for 8 loci, only a single SNP. We draw particular attention to intronic SNP rs4766428

271 (PP>0.97) at *ATP2A2* which encodes a Sarcoplasmic/Endoplasmic Reticulum Calcium pump.

272 Mutations in *ATP2A2* cause Darier Disease[27], which co-segregates with bipolar disorder in

273 several multiplex pedigrees, and is also associated with bipolar disorder and schizophrenia at

274 a population level[28]. Given enrichment for associations in and around voltage gated calcium

medRxiv preprint doi: https://doi.org/10.1101/2020.09.12.20192922; this version posted September 13, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. All rights reserved. No reuse allowed without permission.

channels (**Supplementary Tables 3 and 7**), *ATP2A2* may be involved in pathogenesis through its role in regulating cytoplasmic calcium in neurons.

We identified a fine-mapped set comprising genes with at least one credible SNP. We *prioritised* those containing a) at least one nonsynonymous (NS) or untranslated region (UTR) variant with a PP≥0.1 **(Supplementary Table 12)** or (b) the entire credible set was annotated to that gene **(Supplementary Table 13)**. Genes that are expressed in brain, and that are relatively intolerant to loss-of-function mutations, are known to be enriched for schizophrenia associations ([29] and **Figure 4A**). Genes prioritised by FINEMAP were further enriched for these properties compared to other genes in the loci **(Figure 4B; Supplementary Table 14; Supplementary Note)** supporting our prioritisation strategy.

medRxiv preprint doi: https://doi.org/10.1101/2020.09.12.20192922; this version posted September 13, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. All rights reserved. No reuse allowed without permission.

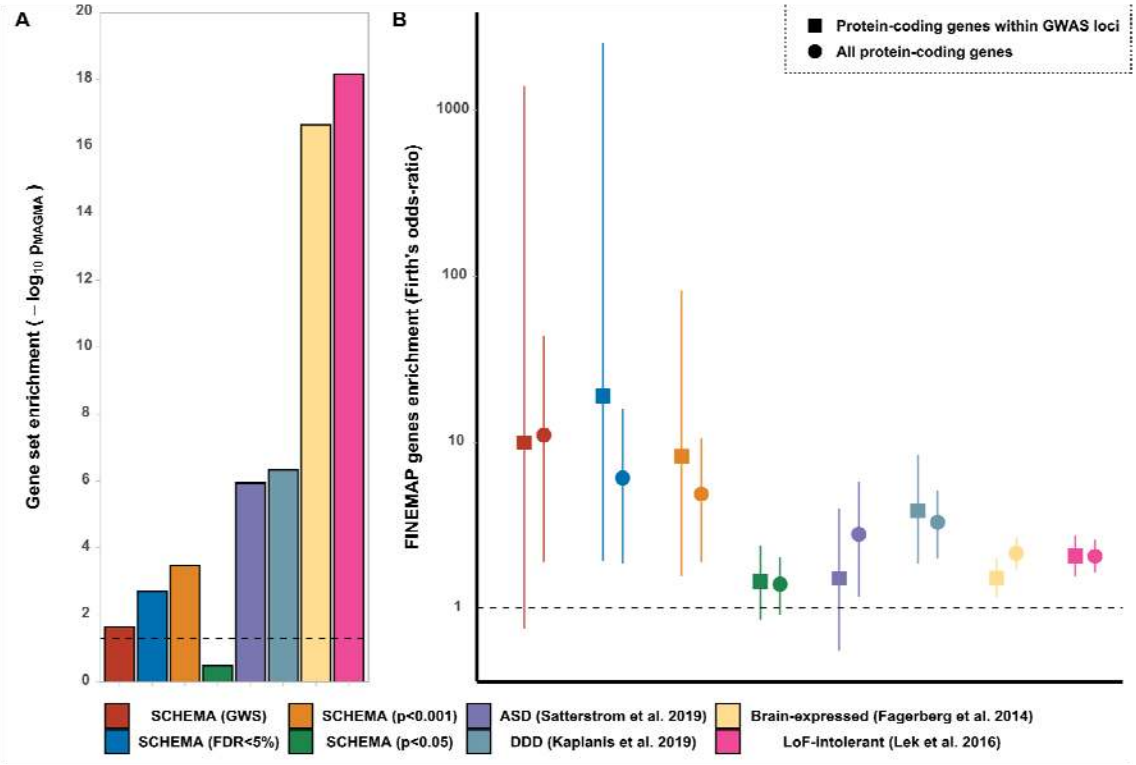


Figure 4: ***Gene set tests.*** *Gene sets tested were retrieved from genome and exome-wide sequencing studies of schizophrenia (SCHEMA; companion paper), autism-spectrum disorder[37] and developmental disorders[36]. For reference, additional sets representing LoF-intolerant genes[53] and brain-expressed genes[54] are also shown. A: MAGMA gene set enrichment analysis, dotted line indicates nominal significance (p=0.05). B: Logistic regression (with Firth's bias reduction method) showing the odds-ratio (and 95% confidence interval) of the association between genes containing at least 1 credible FINEMAP SNP and genes from the sets indicated. Odds-ratios are relative to the remaining protein-coding genes within GWAS loci (squares) or across the genome (circles). Dotted line indicates no enrichment. Analyses are adjusted for gene size.*

We prioritised 19 genes (all protein coding by definition) based on NS or UTR variants (**Supplementary Table 12**), including *SLC39A8*, which mediates zinc and manganese uptake, in which rs13107325, previously a moderately high credible SNP[29], is now strongly supported as responsible (PP > 0.99). The associated T allele has pleiotropic associations of potential relevance to schizophrenia[30]. Other non-synonymous variants with high PP were found in genes with minimal functional characterization including *THAP8* (2 alleles in perfect LD), *WSCD2*, and in 2 E3 ubiquitin ligases *PJA1* and *CUL9*. Also notable for reasons discussed above (see *ATP2A2*), a nonsynonymous and a UTR variant prioritised the voltage gated calcium channel subunit *CACNA1I*. A mixture of missense and UTR variants prioritised *interferon regulatory factor 3* (*IRF3*; cumulative PP=0.49) which promotes

medRxiv preprint doi: https://doi.org/10.1101/2020.09.12.20192922; this version posted September 13, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. All rights reserved. No reuse allowed without permission.

306 interferon expression[31] while *KLF6*, a transcription factor, was highlighted by three variants
307 in the 3' UTR (cumulative PP > 0.99). Interestingly, IRF deficiency has been functionally
308 and genetically linked to herpes simplex encephalitis[32], which can present with
309 schizophrenia-like features. Finally, we identified 65 genes (58 protein coding) in which the
310 95% credible set is restricted to a single gene **(Supplementary Table 13)**.
311

312 ***Common and Rare Variant Associations with Schizophrenia***

313 The Schizophrenia Exome Sequencing Meta-Analysis (SCHEMA) consortium (companion
314 paper) identified 32 genes with damaging ultra-rare mutations associated with schizophrenia
315 (FDR<0.05), including 10 at exome-wide significance. Using MAGMA[33], we found both sets
316 of genes were enriched for common variant associations, as were more weakly associated
317 SCHEMA genes down to uncorrected P<0.001 (**Figure 4A, Supplementary Tables 15, 16**).
318 The fine-mapped set was also enriched for SCHEMA genes relative to other genes at
319 associated loci (**Figure 4B; Supplementary Table 16**). Given established rare variant
320 overlaps between schizophrenia, autism spectrum disorder (ASD) and developmental
321 disorder (DD) [8,34,35], we tested for and found the fine-mapped set was also enriched for genes
322 in which rare variants increase risk of these disorders[36,37] **(Figure 4, Supplementary Tables**
323 **15, 16)**. These convergences suggest rare variant data can inform gene prioritisation at
324 GWAS loci.
325

326 Of the 10 exome-wide significant genes identified by SCHEMA, two were prioritised
327 candidates from fine-mapping; *GRIN2A* encoding a glutamatergic NMDA receptor subunit,
328 and *SP4*, a transcription factor highly expressed in brain and which is regulated by NMDA
329 transmission, and also regulates, NMDA receptor abundance[38]. Two other genes implicated
330 by SCHEMA at FDR<0.05 had support from fine-mapping: *STAG1* which is involved in

medRxiv preprint doi: https://doi.org/10.1101/2020.09.12.20192922; this version posted September 13, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. All rights reserved. No reuse allowed without permission.

controlling chromosome segregation and regulating expression, and *FAM120A*, which encodes an RNA binding protein. SNPs mapping to these genes captured 85% and 70% of the FINEMAP PP respectively (**Supplementary Table 11b**). The prioritised fine-mapped set also contained 8 genes implicated in ASD and/or DD; 5 transcriptional regulators (*FOXP1*, *MYT1L*, *RERE*, *BCL11B, KANSL1*), the well-known candidate *CACNA1C*[52], and genes mentioned elsewhere in this paper (*GRIN2A* and *SLC39A8*).

Prioritisation by Gene Expression

It is thought that common variant associations in schizophrenia are frequently mediated by eQTLs, that is, variants that influence gene expression[39]. We used summary-based Mendelian randomisation (SMR)[40] to detect GWAS associations that co-localise with eQTLs (from adult brain[41], fetal brain[42] or whole blood[43]) and the HEIDI test[40] to reject co-localisations due to LD between distinct schizophrenia-associated and eQTL variants **(Supplementary Table 17)**. To retain brain relevance, we considered only findings from blood that replicated in brain. After removing duplicates identified in multiple tissues **(Supplementary 18a-c)**, we identified 116 SMR-implicated genes **(Supplementary Table 18d);** the use of alternative methodologies supported the robustness of the SMR findings **(Supplementary Note and Supplementary Table 18e)**.

We used three approaches to prioritise genes from these 116 candidates **(Supplementary Note)**. We identified (i) 33 genes as the single SMR-implicated gene at the locus or through conditional analysis of a locus containing multiple candidates **(Supplementary Table 18f; Supplementary Note)**: (ii) 18 genes where the putatively causal eQTLs captured 50% or more of the FINEMAP posterior probability (**Supplementary Table 18g**): (iii) 30 genes where chromatin conformation analysis (Hi-C analysis of adult and fetal brain) suggested a

medRxiv preprint doi: https://doi.org/10.1101/2020.09.12.20192922; this version posted September 13, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. All rights reserved. No reuse allowed without permission.

promoter of that gene interacted with a putative regulatory element containing a FINEMAP credible SNP (**Supplementary Table 19**)[44].

After removing duplicates, there were 60 SMR-prioritised genes **(Supplementary Table 20)**. The neurodevelopmental disorder gene *RERE* (over-expression in schizophrenia) was prioritised by all 3 strategies and separately by fine-mapping (see above). Others with strong evidence **(Supplementary Table 18g)** include *ACE* encoding angiotensin converting enzyme, the target of a major class of anti-hypertensive drugs (schizophrenia under-expression), *DCLK3* encoding a neuroprotective kinase[45](schizophrenia under-expression) and *SNAP91* (discussed below; schizophrenia over-expression).

Combining all approaches, FINEMAP and SMR, we prioritised 130 genes of which 114 are protein coding (**Figure 3; Extended Data Table 1**).

Synaptic Location and Function of Prioritised Genes

Given the implication of synaptic pathology from the enrichment tests, we examined prioritised genes in the context of synaptic location and function in the SynGO database[25]. Of the 114 proteins encoded, 13 have synaptic annotations **(Supplementary Table 21)**; 6 are postsynaptic, 5 are both pre- and post- synaptic, 1 is presynaptic, and 1 gene (MAPK3) is not

medRxiv preprint doi: https://doi.org/10.1101/2020.09.12.20192922; this version posted September 13, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity.
All rights reserved. No reuse allowed without permission.

mapped to any specific compartment (**Figure 5B**).

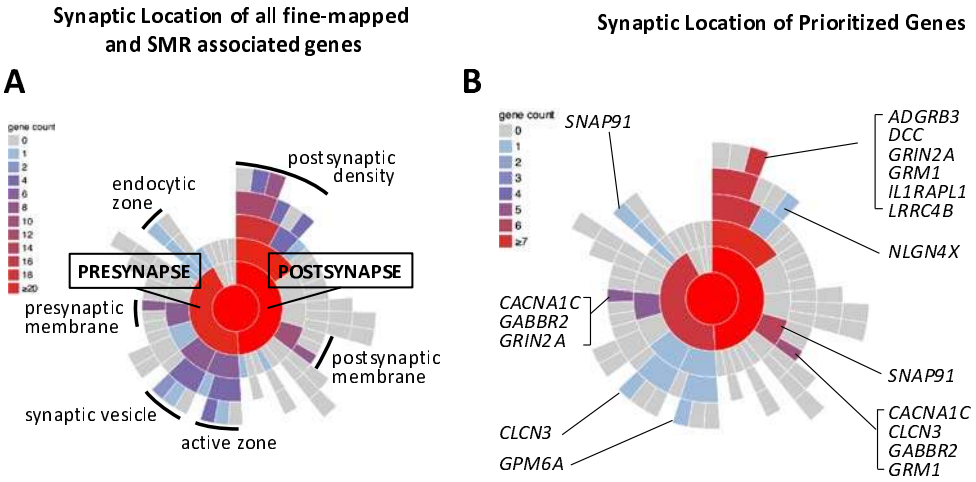


Figure 5: ***Mapping of all FINEMAP/SMR genes (A) and prioritized genes (B) to synaptic locations using SYNGO****. Sunburst plots depict synaptic locations with child terms in concentric rings, starting with the synapse (center), pre- and postsynaptic locations in the first ring and child terms in subsequent ring. The number of genes in each term is indicated by the colour scheme in the legend. FINEMAP/SMR genes are those tagged by at least one credible SNP identified by FINEMAP or which are associated using SMR (N=437 protein coding genes) of which N=50 are SynGO annotated, 44 to cellular components. Prioritized genes are the subset of protein coding genes identified as likely to be causal (**Extended Data Table 1**; N=114) using the prioritization criteria described in the main text of which 13 are SynGO annotated, 12 to cellular components. Genes were annotated using the SYNGO synaptic location database (www.syngoportal.org) which is exclusively based on expert annotation of published evidence (see Koopmans et al[25]).*

The results are consistent with the genome-wide enrichment tests pointing to postsynaptic pathology, but the fact that many prioritised genes have additional locations suggests presynaptic pathology may also be involved. The encoded proteins map to 17 unique biological terms in the hierarchy **(Supplementary Table 21**), but there are specific themes. Multiple genes encode receptors and ion channels, including voltage-gated calcium and chloride channels (*CACNA1C*, *CLCN3*), metabotropic receptors (glutamate and GABA), and the ligand-gated NMDA receptor subunit (*GRIN2A*). Others involve proteins playing a role in synaptic organisation and differentiation (*ADGRB3*, *LRRC4B*, *GPM6A*, *IL1RAPL1*), the trans-synaptic signalling complex (*IL1RAPL1*, *SNAP91)* and modulation of chemical transmission (*MAPK3*, *DCC*, *CLCN3*). *ADGRB3*, encoding adhesion G protein-coupled receptor B3, is a receptor for ligands of the secreted complement factor family and together

medRxiv preprint doi: https://doi.org/10.1101/2020.09.12.20192922; this version posted September 13, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. All rights reserved. No reuse allowed without permission.

they control the synaptic connectivity and refinement[46,47]. This is notable given that the complement component 4 genes, which have been shown to account for some of the association signal across the MHC locus[49] and are part of a classical complement pathway in which multiple complement proteins regulate synaptic refinement[50]. The diversity of synaptic proteins identified in this study suggests multiple functional interactions of schizophrenia risk converging on synapses. It remains to be determined whether these proteins co-act together at specific synapse types, or whether the diversity points to multiple types in different brain regions.

DISCUSSION

We have performed the largest GWAS of schizophrenia to date and in doing so, identify a substantial increase in the number of associated loci. We show that genes we prioritise within associated loci by fine-mapping are highly enriched for those with an increased burden of rare deleterious mutations in schizophrenia, and identify *GRIN2A*, *SP4*, *STAG1*, and *FAM120A* as specific genes where the convergence of rare and common variant associations strongly supports their pathogenic relevance for the disorder. Importantly, this convergence also implies that the pathogenic relevance of altered function of these genes extends beyond the small proportion of cases who carry rare mutations in them. We also demonstrate that common variant schizophrenia associations, and more importantly, 8 prioritised associated genes that likely explain associations, overlap with genes implicated by rare coding variants in neurodevelopmental disorders, opening the door for exploiting the increasing power of rare variant studies of those disorders to further prioritise genes from GWAS studies.

The power of the present study allowed us to show enrichment of common variant associations restricted to genes expressed in CNS neurons, both excitatory and inhibitory, and

medRxiv preprint doi: https://doi.org/10.1101/2020.09.12.20192922; this version posted September 13, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. All rights reserved. No reuse allowed without permission.

fundamental biological processes related to neuronal function. This points to neurons as the most important source of pathology in the disorder. We also show that genes with high relative specificity for expression in almost all tested brain regions are enriched for genetic association. This suggests that abnormal neuronal function in schizophrenia is not confined to a small number of brain structures, which in turn might explain its diverse psychopathology, its association with a broad range of cognitive impairments, and the lack of regional specificity in neuroimaging measures associated with the disorder[1].

Schizophrenia pathophysiology is likely to extend beyond the synapse. Nevertheless, the concentration of associations we show in genes with pre- and post-synaptic locations (including glutamate and GABA receptors and their associated proteins) and with functions related to synaptic organisation, differentiation and transmission, point to these neuronal compartments and their attendant functions as being of central importance. This is further supported by studies showing that CNVs[51] and rare damaging coding variants in genes with similar functions, and even some of the same genes (SCHEMA; companion paper) substantially increase risk of schizophrenia. Genomic studies of all designs now converge in suggesting research aiming for a mechanistic understanding of the disorder should be targeted towards these areas of biology; the large number of candidates identified here offers an unprecedented empirically-supported resource for that endeavour.

Ethics

The study protocols were approved by the institutional review board at each centre involved with recruitment. Informed consent and permission to share the data were obtained from all subjects, in compliance with the guidelines specified by the recruiting centres' institutional

medRxiv preprint doi: https://doi.org/10.1101/2020.09.12.20192922; this version posted September 13, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. All rights reserved. No reuse allowed without permission.

review boards. Genotyping of samples recruited in mainland China were processed and analysed by Chinese groups on Chinese local servers, to comply with the Human Genetic Resources Administrative Regulations. Only summary statistics, with no individual-level data, were included in the final study from samples recruited from mainland China.

Acknowledgements

The National Institute of Mental Health (USA) provides core funding for the Psychiatric Genomics Consortium was Award Number U01MH109514. The content is the responsibility of the authors and does not necessarily represent the official views of the National Institutes of Health. We acknowledge Pamela Sklar (deceased); as one of the PIs of that award, her contribution was substantial. The work of the contributing groups was supported by numerous grants from governmental and charitable bodies as well as philanthropic donation (details to be provided). We also acknowledge Ed Scolnick, Chief Scientist Emeritus, Stanley Center of the Broad Institute, whose support for this study was vital to its success. We also acknowledge the Wellcome Trust Case Control Consortium for the provision of control genotype information. Membership of the Psychosis Endophenotype International Consortium,
the SynGO consortium, the PsychENCODE Consortium, the eQTLGen consortium, the BIOS Consortium and the Indonesia Consortium are provided in the accompanying author and consortium XL file.
S.X. also gratefully acknowledges the support of the National Natural Science Foundation of China (NSFC) grant (31525014, 91731303, 31771388, 31961130380, and 32041008), the UK Royal Society-Newton Advanced Fellowship (NAF\R1\191094), Key Research Program of Frontier Sciences (QYZDJ-SSW-SYS009) and the Strategic Priority Research Program

medRxiv preprint doi: https://doi.org/10.1101/2020.09.12.20192922; this version posted September 13, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. All rights reserved. No reuse allowed without permission.

(XDB38000000) of the Chinese Academy of Sciences, and the Shanghai Municipal Science and Technology Major Project (2017SHZDZX01).

This study is supported by Research Council of Norway (283798, 262656, 248980, 273291, 248828, 248778, 223273); KG Jebsen Stiftelsen, South-East Norway Health Authority, EU H2020 # 847776.

Dr. Faraone is supported by the European Union's Seventh Framework Programme for research, technological development and demonstration under grant agreement no 602805, the European Union's Horizon 2020 research and innovation programme under grant agreements No 667302 & 728018 and NIMH grants 5R01MH101519 and U01 MH109536-01.

This study was supported by FAPESP - Fundação de Amparo à Pesquisa do Estado de São Paulo (Brazil) - Grant numbers: 2010/08968-6 (S.I.B.); 2014/07280-1 (S.I.B.); 2007/58736-1 (M.AC.S.); 2011/50740-5 (R.A.B.); 2016/04983-7 (J.J.M.); 10/19176-3 (V.K.O. & S.I.B.); 12/12686-1 (M.L.S. & S.I.B.); CAPES - Coordenação de Aperfeiçoamento de Pessoal de Nível Superior Code 001.

The Singapore team (Lee Jimmy, Liu Jianjun, Sim Kang, Chong Siow Chong, Mythily Subramanian) acknowledges the National Medical Research Council Translational and Clinical Research Flagship Programme (grant number: NMRC/TCR/003/2008).

Supported by LM2018132, CZ.02.1.01/0.0/0.0/18_046/0015515 and IP6003 – VZFNM00064203 to MM Jr.

Dr. Arango has been funded by the Spanish Ministry of Science and Innovation. Instituto de Salud Carlos III (SAM16PE07CP1, PI16/02012, PI19/024), co-financed by ERDF Funds from the European Commission, "A way of making Europe", CIBERSAM. Madrid Regional Government (B2017/BMD-3740 AGES-CM-2), European Union Structural Funds. European Union Seventh Framework Program; and European Union H2020 Program under the

medRxiv preprint doi: https://doi.org/10.1101/2020.09.12.20192922; this version posted September 13, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. All rights reserved. No reuse allowed without permission.

496 Innovative Medicines Initiative 2 Joint Undertaking (grant agreement No 115916, Project
497 PRISM, and grant agreement No 777394, Project AIMS-2-TRIALS), Fundación Familia
498 Alonso and Fundación Alicia Koplowitz .
499 E. Bramon was supported by the Medical Research Council (G0901310 and G1100583),
500 the Wellcome Trust (085475/B/08/Z, 085475/Z/08/Z), Mental Health Research UK, British
501 Medical Association's Margaret Temple Fellowship, and the NIHR Biomedical Research
502 Centre at University College London Hospitals NHS Foundation Trust and University
503 College London.
504 M.Dolores Moltó is funded by the European Regional Development Fund (ERDF)-Valencian
505 Community 2014-2020, Spain.
506

medRxiv preprint doi: https://doi.org/10.1101/2020.09.12.20192922; this version posted September 13, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. All rights reserved. No reuse allowed without permission.

MAIN REFERENCES

1. Owen, M. J., Sawa, A. & Mortensen, P. B. Schizophrenia. *The Lancet* **388**, 86–97 (2016).
2. Plana-Ripoll, O. *et al.* A comprehensive analysis of mortality-related health metrics associated with mental disorders: a nationwide, register-based cohort study. *Lancet* **394**, 1827–1835 (2019).
3. Momen, N. C. *et al.* Association between mental disorders and subsequent medical conditions. *N. Engl. J. Med.* **382**, 1721–1731 (2020).
4. Jääskeläinen, E. *et al.* A systematic review and meta-analysis of recovery in schizophrenia. *Schizophr Bull.* **39**, 1296–1306 (2013).
5. International Schizophrenia Consortium. Common polygenic variation contributes to risk of schizophrenia and bipolar disorder. *Nature* **460**, 748–752 (2009).
6. Pocklington, A. J. *et al.* Novel Findings from CNVs Implicate Inhibitory and Excitatory Signaling Complexes in Schizophrenia. *Neuron* **86**, 1203–1214 (2015).
7. Singh, T. *et al.* The contribution of rare variants to risk of schizophrenia in individuals with and without intellectual disability. *Nat. Genet.* **49**, 1167–1173 (2017).
8. Rees, E. *et al*. De novo mutations identified by exome sequencing implicate rare missense variants in SLC6A1 in schizophrenia. *Nat. Neurosci.* **23**, 179–184 (2020).
9. Lam, M. *et al.* Comparative genetic architectures of schizophrenia in East Asian and European populations. *Nat. Genet.* **51**, 1670–1678 (2019).
10. Yang, J. *et al.* Genomic inflation factors under polygenic inheritance. *Eur. J. Hum. Genet.* **19**, 807–812 (2011).
11. Schizophrenia Working Group of the Psychiatric Genomics Consortium. Biological insights from 108 schizophrenia-associated genetic loci. *Nature* **511**, 421–427 (2014).
12. Räsänen, S., Pakaslahti, A., Syvälahti, E., Jones, P. B. & Isohanni, M. Sex differences

medRxiv preprint doi: https://doi.org/10.1101/2020.09.12.20192922; this version posted September 13, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. All rights reserved. No reuse allowed without permission.

in schizophrenia: A review. *Nord. J. Psychiatry* **54**, 37–45 (2000).

13. Zeng, J. *et al.* Bayesian analysis of GWAS summary data reveals differential signatures of natural selection across human complex traits and functional genomic categories. *bioRxiv* 752527 (2019). doi:10.1101/752527

14. Lancaster, T. M. *et al.* Structural and Functional Neuroimaging of Polygenic Risk for Schizophrenia: A Recall-by-Genotype-Based Approach. *Schizophr. Bull.* **45**, 405–414 (2019).

15. Brainstorm Consortium *et al.* Analysis of shared heritability in common disorders of the brain. *Science (80-.).* **360**, eaap8757 (2018).

16. Verbanck, M., Chen, C. Y., Neale, B. & Do, R. Detection of widespread horizontal pleiotropy in causal relationships inferred from Mendelian randomization between complex traits and diseases. *Nat. Genet.* **50**, 693–698 (2018).

17. Jeppesen, R. & Benros, M. E. Autoimmune diseases and psychotic disorders. *Front. Psychiatry* **10**, (2019).

18. Lawlor, D. A., Tilling, K. & Davey Smith, G. Triangulation in aetiological epidemiology. *Int. J. Epidemiol.* **45**, 1866–1886 (2016).

19. Byrne, E. M. *et al.* Is Schizophrenia a Risk Factor for Breast Cancer?—Evidence From Genetic Data | Schizophrenia Bulletin | Oxford Academic. *Schizophr. Bull.* **45**, 1251–1256 (2019).

20. GTEx Consortium. Genetic effects on gene expression across human tissues. *Nature* **550**, 204–213 (2017).

21. The Schizophrenia Psychiatric Genome-Wide Association Study (GWAS) Consortium. Genome-wide association study identifies five new schizophrenia loci. *Nat. Genet.* **43**, 969–978 (2011).

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

medRxiv preprint doi: https://doi.org/10.1101/2020.09.12.20192922; this version posted September 13, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. All rights reserved. No reuse allowed without permission.

Nat. Genet. **50**, 825–833 (2018).

23. Habib, N. *et al.* Massively parallel single-nucleus RNA-seq with DroNc-seq. *Nat. Methods* **14**, 955–958 (2017).

24. Zeisel, A. *et al.* Molecular Architecture of the Mouse Nervous System. *Cell* **174**, 999-1014.e22 (2018).

25. Koopmans, F. *et al.* SynGO: An Evidence-Based, Expert-Curated Knowledge Base for the Synapse. *Neuron* **103**, 217-234.e4 (2019).

26. Benner, C. *et al.* FINEMAP: Efficient variable selection using summary data from genome-wide association studies. *Bioinformatics* **32**, 1493–1501 (2016).

27. Sakuntabhai, A. *et al.* Mutations in ATP2A2, encoding a Ca2+ pump, cause Darier disease. *Nat. Genet.* **21**, 271–277 (1999).

28. Cederlöf, M. *et al.* The association between Darier disease, bipolar disorder, and schizophrenia revisited: A population-based family study. *Bipolar Disord.* **17**, 340–344 (2015).

29. Pardiñas, A. F. *et al.* Common schizophrenia alleles are enriched in mutation-intolerant genes and in regions under strong background selection. *Nat. Genet.* **50**, 381–389 (2018).

30. Costas, J. The highly pleiotropic gene *SLC39A8* as an opportunity to gain insight into the molecular pathogenesis of schizophrenia. *Am. J. Med. Genet. Part B Neuropsychiatr. Genet.* **177**, 274–283 (2018).

31. Au, W. C., Moore, P. A., Lowther, W., Juang, Y. T. & Pitha, P. M. Identification of a member of the interferon regulatory factor family that binds to the interferon-stimulated response element and activates expression of interferon-induced genes. *Proc. Natl. Acad. Sci. U. S. A.* **92**, 11657–11661 (1995).

32. Mogensen, T. H. IRF and STAT transcription factors - From basic biology to roles in

medRxiv preprint doi: https://doi.org/10.1101/2020.09.12.20192922; this version posted September 13, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. All rights reserved. No reuse allowed without permission.

582 infection, protective immunity, and primary immunodeficiencies. *Frontiers in*
583 *Immunology* **10**, (2019).
584 33. de Leeuw, C. A., Mooij, J. M., Heskes, T. & Posthuma, D. MAGMA: Generalized
585 Gene-Set Analysis of GWAS Data. *PLOS Comput. Biol.* **11**, e1004219 (2015).
586 34. Rees, E. *et al.* Analysis of intellectual disability copy number variants for association
587 with schizophrenia. *JAMA Psychiatry* **73**, 963–969 (2016).
588 35. Fromer, M. *et al.* De novo mutations in schizophrenia implicate synaptic networks.
589 *Nature* **506**, 179–184 (2014).
590 36. Kaplanis, J. *et al.* Integrating healthcare and research genetic data empowers the
591 discovery of 49 novel developmental disorders. *bioRxiv* (2019). doi:10.1101/797787
592 37. Satterstrom, F. K. *et al.* Large-Scale Exome Sequencing Study Implicates Both
593 Developmental and Functional Changes in the Neurobiology of Autism. *Cell* **180**, 568-
594 584.e23 (2020).
595 38. Priya, A., Johar, K. & Wong-Riley, M. T. T. Specificity protein 4 functionally
596 regulates the transcription of NMDA receptor subunits GluN1, GluN2A, and GluN2B.
597 *Biochim. Biophys. Acta - Mol. Cell Res.* **1833**, 2745–2756 (2013).
598 39. Richards, A. L. *et al.* Schizophrenia susceptibility alleles are enriched for alleles that
599 affect gene expression in adult human brain. *Mol. Psychiatry* **17**, 193–201 (2012).
600 40. Zhu, Z. *et al.* Integration of summary data from GWAS and eQTL studies predicts
601 complex trait gene targets. *Nat. Genet.* **48**, 481–487 (2016).
602 41. Gandal, M. J. *et al.* Transcriptome-wide isoform-level dysregulation in ASD,
603 schizophrenia, and bipolar disorder. *Science (80-.).* **362**, (2018).
604 42. O'Brien, H. E. *et al.* Expression quantitative trait loci in the developing human brain
605 and their enrichment in neuropsychiatric disorders. *Genome Biol.* **19**, 194 (2018).
606 43. Võsa, U. *et al.* Unraveling the polygenic architecture of complex traits using blood

medRxiv preprint doi: https://doi.org/10.1101/2020.09.12.20192922; this version posted September 13, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. All rights reserved. No reuse allowed without permission.

eQTL metaanalysis. *bioRxiv* 447367 (2018). doi:10.1101/447367

44. Wang, D. *et al.* Comprehensive functional genomic resource and integrative model for the human brain. *Science (80-.).* **362**, eaat8464 (2018).

45. Galvan, L. *et al.* The Striatal Kinase DCLK3 Produces Neuroprotection Against Mutant Huntingtin - PubMed. *Brain* **141**, 1434–1454 (2018).

46. Sigoillot, S. M. *et al.* The secreted protein C1QL1 and its receptor BAI3 control the synaptic connectivity of excitatory inputs converging on cerebellar purkinje cells. *Cell Rep.* **10**, 820–832 (2015).

47. Kakegawa, W. *et al.* Anterograde C1ql1 signaling is required in order to determine and maintain a single-winner climbing fiber in the mouse cerebellum. *Neuron* **85**, 316–329 (2015).

48. Kamitaki, N. *et al.* Complement genes contribute sex-biased vulnerability in diverse disorders. *Nature* **582**, 577–581 (2020).

49. Sekar, A. *et al.* Schizophrenia risk from complex variation of complement component 4. *Nature* **530**, 177–183 (2016).

50. Schafer, D. P. *et al.* Microglia Sculpt Postnatal Neural Circuits in an Activity and Complement-Dependent Manner. *Neuron* **74**, 691–705 (2012).

51. Kirov, G. *et al.* De novo CNV analysis implicates specific abnormalities of postsynaptic signalling complexes in the pathogenesis of schizophrenia. *Mol. Psychiatry* **17**, 142–153 (2012).

52. Ripke, S. *et al.* Genome-wide association analysis identifies 13 new risk loci for schizophrenia. *Nat. Genet.* **45**, 1150–1159 (2013).

53. Lek, M. *et al.* Analysis of protein-coding genetic variation in 60,706 humans. *Nature* **536**, 285–291 (2016).

medRxiv preprint doi: https://doi.org/10.1101/2020.09.12.20192922; this version posted September 13, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity.
All rights reserved. No reuse allowed without permission.

631 54. Fagerberg, L. *et al.* Analysis of the human tissue-specific expression by genome-wide
632 integration of transcriptomics and antibody-based proteomics. *Mol. Cell. Proteomics*
633 **13**, 397–406 (2014).

634

medRxiv preprint doi: https://doi.org/10.1101/2020.09.12.20192922; this version posted September 13, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. All rights reserved. No reuse allowed without permission.

METHODS

Table of Contents

Overview of Samples

Details of each of the samples (including sample size, ancestry, and whether included in the previous publication by the PGC) are given as a **Sample Supplementary Note**. From 90 cohorts, we performed GWAS on 161,405 unrelated subjects; 67,390 schizophrenia/schizoaffective disorder cases and 94,015 controls. A parent-proband trio is considered to comprise one case and one control. Approximately half (31,914 cases and 47,176 controls) of the samples were not included in the previous GWAS of the PGC[1]. Around 80% of the cases were of European Ancestry, and the remainder were of East Asian ancestry[2]. Variants showing evidence for association ($P< 1x10^{-5}$) were further meta-analysed with an additional dataset of 1,979 cases and 142,627 controls of European ancestry obtained from deCODE genetics, thus the final analysis represents 306,011 diploid genomes.

medRxiv preprint doi: https://doi.org/10.1101/2020.09.12.20192922; this version posted September 13, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. All rights reserved. No reuse allowed without permission.

Association Analysis

Quality Control

Quality control was performed on the cohorts separately according to standards developed by the Psychiatric Genomics Consortium (PGC)[3] including SNP missingness < 0.05 (before sample removal); subject missingness < 0.02; autosomal heterozygosity deviation ($|F_{het}| < 0.2$); SNP missingness < 0.02 (after sample removal); difference in SNP missingness between cases and controls < 0.02; and SNP Hardy-Weinberg equilibrium (HWE: $P > 10^{-6}$ in controls or $P > 10^{-10}$ in cases). For family-based cohorts we excluded individuals with more than 10,000 Mendelian errors and SNPs with more than 4 Mendelian errors. For X-Chromosomal genotypes we applied an additional round of the above QC to the male and female subgroups separately.

Imputation

Genotype imputation of case-control cohorts was performed using the pre-phasing/imputation stepwise approach implemented in EAGLE 2[4] / MINIMAC3[5] (with 132 genomic windows of variable size and default parameters). The imputation reference consisted of 54,330 phased haplotypes with 36,678,882 variants from the publicly available HRC reference, release 1.1 (URL: https://ega-archive.org/datasets/EGAD00001002729). Chromosome X imputation was conducted using individuals passing quality control for the autosomal analysis. ChrX imputation and association analysis was performed separately for males and females. For trio-based cohorts, families with multiple (N) affected offspring were split into N parent-offspring trios, duplicating the parental genotype information. Trios were phased with SHAPEIT 3[6]. We created pseudo-controls based on the non-transmitted alleles from the parents. Phased case-pseudo-control genotypes were then taken forward to the IMPUTE4 algorithm[7] into the above HRC reference panel.

medRxiv preprint doi: https://doi.org/10.1101/2020.09.12.20192922; this version posted September 13, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. All rights reserved. No reuse allowed without permission.

Principal Component Analysis (PCA) and Relatedness Checking

We performed PCA for all 90 cohorts separately using SNPs with high imputation quality (INFO >0.8), low missingness (<1%), MAF>0.05 and in relative linkage equilibrium (LD) after 2 iterations of LD pruning (r2 < 0.2, 200 SNP windows). We removed well known long-range-LD areas (MHC and chr8 inversion). Thus, we retained between 57K and 95K autosomal SNPs in each cohort. SNPs present in all 90 cohorts (N=7,561) were used for robust relatedness testing using PLINK v1.9[8]; pairs of subjects with PIHAT > 0.2 were identified and one member of each pair removed at random, preferentially retaining cases and trio members over case-control members.

Association / Meta-analysis

In each individual cohort, association testing was based on an additive logistic regression model using PLINK[8]. As covariates we used a subset of the first 20 principal components (PCA), derived within each cohort. By default, we included the first 4 PCAs and thereafter every PCA that was nominally significantly associated (p<0.05) to case-control status. PCAs in trios were only used to remove ancestry outliers. We conducted a meta-analysis of the results using a standard error inverse-weighted fixed effects model. For chrX, gene dosages in males were scored 0 or 2, in females, 0/1/2. We summarised the associations as number of independently associated index SNPs. Index SNPs were LD independent and had r2 < 0.1 within 3 Mb windows. We recorded the left and rightmost variant with r2<0.1 to an index SNP to define an associated clump. After adding a 50kb window on each side of the LD clump we merged overlapping LD-clumps to a total of 248 distinct genomic loci (5 on the X-chromosome) with at least one genome-wide significant signal.

medRxiv preprint doi: https://doi.org/10.1101/2020.09.12.20192922; this version posted September 13, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. All rights reserved. No reuse allowed without permission.

712 Due to the strong signal and high linkage disequilibrium in the MHC, only one SNP was kept
713 from the extended MHC region (chr6:25-35Mb).
714
715 We additionally examined the X chromosome for evidence of heterogeneity between the
716 sexes and X chromosome dosage compensation using the methods described by Lee and
717 colleagues [9,10] (**Supplementary Note**). To minimise possible confounding effects of ancestry
718 on effect sizes by sex, we restricted this analysis to those of European ancestry.
719
720 We obtained summary association results from deCODE genetics for 1,187 index SNPs ($P < 1x10^{-5}$) based on 1,979 cases and 142,627 controls of European ancestry. Genotyping was
722 carried out at deCODE Genetics. X chromosome summary statistics were only present for the
723 Icelandic component. We used this sample to establish that SNP associations from the
724 primary GWAS replicated *en masse* in an independent sample (see **Supplementary Note**) by
725 showing the directions of effect of index SNPs differed from the null hypothesis of randomly
726 oriented effects and also comparing the expected number of same direction effects with those
727 if all associations were true, taking into account the discovery magnitude of effect, and the
728 replication effect-estimate precision (**Supplementary Note**).
729
730 The summary statistics from deCODE were combined with those from our primary GWAS
731 dataset using an inverse variance-weighted fixed effects model. Similarily to the discovery
732 meta-analysis (see above) we merged overlapping LD-clumps to a total of 270 distinct
733 genomic regions (5 on the X-chromosome) with at least one genome-wide significant signal.
734

medRxiv preprint doi: https://doi.org/10.1101/2020.09.12.20192922; this version posted September 13, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. All rights reserved. No reuse allowed without permission.

Polygenic Prediction

We estimated the cumulative contribution of SNPs to polygenic risk of schizophrenia using a series of 89 leave-one-out polygenic prediction analyses based on LD-clumping and P-value thresholding (P+T) [11] using PLINK[8]. For calculating polygenic scores, we included the most significant SNP for any pair of SNPs within <500kb and with LD R^2 >0.1. We included only those with minor allele frequency >1%. We considered a range of P-value thresholds; 5×10^{-8}, 1×10^{-6}, 1×10^{-4}, 1×10^{-3}, 1×10^{-2}, 5×10^{-2}, 1×10^{-1}, 2×10^{-1}, 5×10^{-1} and 1.0. We performed logistic regression analysis within each case-control sample, to assess the relationship between case status and PRS (P+T) quantiles. The same principal components used for each GWAS were used as covariates for this analysis. Whenever the number of controls at a quantile was fewer than 5 times the number of covariates[12], or if the higher bound for the PRS Odds Ratio (OR) became infinity, Firth's penalised likelihood method was used to compute regression statistics, as implemented in the R package "logistf"[13]. ORs from these calculations were then meta-analysed using a fixed-effects model in the R package "metafor"[14]. To ensure stability of the estimates, meta-analysis was conservatively restricted to the 51 case-control samples which contained more than 10 individuals in the top 1% PRS, with at least one of them being a control. Analogous analyses were conducted to assess the ORs between individuals at the top and bottom quantiles. To assess the performance of PRS as a predictor of schizophrenia case status, a combined area under the receiver operating characteristic curve (AUROC) was estimated using the non-parametric meta-analysis implemented in the R package "nsROC"[15].

Mendelian Randomisation

We identified 58 publicly available sets of GWAS summary statistics suitable for bi-directional 2-sample Mendelian randomisation[16]. We used the IEU GWAS summary database (https://gwas.mrcieu.ac.uk/datasets/) for complex traits and diseases, with additional manual

medRxiv preprint doi: https://doi.org/10.1101/2020.09.12.20192922; this version posted September 13, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. All rights reserved. No reuse allowed without permission.

curation (SNP rsid, chromosome, position, reference and alternate alleles, effect size, standard errors, and P values) (see **Supplementary Table 6** for traits and references). We included only GWAS with over 10,000 samples, >1,000,000 SNPs. We performed QC including matching reference and alternate alleles and flipped stands to ensure all effect sizes in the GWAS studies are relative to the same allele (for strand ambiguous SNPs, if MAF ≤0.42, allele frequencies were used determine the strand, otherwise they were discarded). All GWAS were based on European ancestry samples (including schizophrenia).

To extract independent genetic instruments, we performed LD-clumping using 1000 Genomes European super-population subjects as the LD reference, with an R^2 t hreshold of 0.001 and a window of 10Mb. We excluded the extended MHC region (chromosome 6:25-35 Mb) to avoid bias by pseudo-pleiotropy due to extensive LD in this region. We extracted 153 independent genome-wide significant SNPs for schizophrenia (EUR sample only). For each exposure-outcome pair, we used independent genome-wide significant SNPs for the exposure as instrumental variables (IV) and extracted the effects of these SNPs on the exposure and the outcome. We excluded exposure-outcome pairs with less than 10 IVs. Due to the differences in the source GWAS summary statistics, especially the imputation reference and sample sizes, IV for each exposure-outcome pair varied from 11 to 359 (**Supplementary Table 6**).

We used several methods to estimate causal effects while detecting and correcting for horizontal pleiotropy. We used inverse variance weighted meta-analysis to incorporate causal effect estimates from multiple IVs for each exposure-outcome pair (and Maximum likelihood estimates as a supplement) [17]. IVW meta-analysis in MR is prone to bias due to horizontal pleiotropy so we used MR-Egger regression[18] to detect average pleotropic bias, modified Q test and Q' test[19] and MR-PRESSO global test[20] to detect overall pleiotropic bias through

medRxiv preprint doi: https://doi.org/10.1101/2020.09.12.20192922; this version posted September 13, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. All rights reserved. No reuse allowed without permission.

outlier detection, and MR-PRESSO outlier test to identify specific genetic variants with horizontal pleiotropy. Furthermore, we used MR-PRESSO to correct for pleiotropic bias by removing outlier IVs in IVW meta-analysis. We adjusted MR-PRESSO results for multiple testing by Bonferroni correction with the total number of exposure-outcome pairs tested (N=116). As sensitivity analyses, we also performed MR-Egger regression to estimate causal effects while correcting for average pleiotropic bias and weighted median estimation to provide consistent causal effect estimates when at least 50% of IVs are valid. Causal effect estimates and bias detection test results are presented. Scatter plots were used to inspect the distribution of genetic effects of the SNP instruments used in each MR analysis (**Supplementary Figure 11**). We used MR-PRESSO as our primary method to assess causal effects between exposure-outcome pairs and potential pleiotropic biases. Further comments on the application of MR are provided in the **Supplementary Note**.

Gene Set Enrichments

Tissue and cell types

We collected bulk RNA-seq data across 53 human tissues (GTEx v8, median across samples)[21]; from a study of 19,550 nuclei from frozen adult human post-mortem hippocampus and prefrontal cortex representing 16 different cell types[22]; from a study of ~10,000 single cells from 5 mouse brain regions (cortex, hippocampus, hypothalamus, midbrain and striatum, in addition to specific enrichments for oligodendrocytes, dopaminergic neurons, serotonergic neurons and cortical parvalbuminergic interneurons) that identified 24 cell types[23]; from a study of~500,000 single cells from the mouse nervous system (19 regions) that identified 265 cell types[24].

medRxiv preprint doi: https://doi.org/10.1101/2020.09.12.20192922; this version posted September 13, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. All rights reserved. No reuse allowed without permission.

809 Datasets were processed uniformly[25]. First, we calculated the mean expression for each gene
810 for each type of data if these statistics were not provided by the authors. We used the pre-
811 computed median expression (transcript per million (TPM)) across individuals for the GTEx
812 tissues (v8). For the GTEx dataset, we excluded tissues with less than 100 samples, merged
813 tissues by organ (with the exception of brain tissues), excluded non-natural tissues (e.g. EBV-
814 transformed lymphocytes) and testis (outlier in hierarchical clustering), resulting in 37
815 tissues. Genes without unique names and genes not expressed in any cell types were
816 excluded. We scaled the expression data to 1M Unique Molecular Identifiers (UMIs) or TPM
817 for each cell type/tissue. After scaling, we excluded non-protein coding genes, and, for mouse
818 datasets, genes that had no expert curated 1:1 orthologs between mouse and human (Mouse
819 Genome Informatics, The Jackson laboratory, version 11/22/2016). We then calculated a
820 metric of gene expression specificity by dividing the expression of each gene in each cell
821 type/tissue by the total expression of that gene in all cell types/tissue, leading to values
822 ranging from 0 to 1 for each gene (0: meaning that the gene is not expressed in that cell
823 type/tissue, 1 that 100% of the expression of that gene is performed in that cell type/tissue).
824 We selected the 10% most specific genes per cell type (or tissue) with an expression level of
825 at least 1TPM, or 1 UMI per million, for downstream analyses and used MAGMA[26] to test
826 whether they were enriched for genetic associations. We performed a one-sided test as we
827 were only interested in enrichments for genetic associations (in contrast with depletions). We
828 also applied partitioned LD score regression (LDSC) as described[27] to the top 10% genes for
829 each cell type for heritability enrichment. We selected the one-sided coefficient z-score p-
830 value as a measure of the association of the cell type/tissue with schizophrenia.
831

medRxiv preprint doi: https://doi.org/10.1101/2020.09.12.20192922; this version posted September 13, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. All rights reserved. No reuse allowed without permission.

Ontology Gene sets

Gene set analyses were performed using MAGMA v1.06[26]. Gene boundaries were retrieved from Ensembl release 92 (GRCh37) using the "biomaRt" R package [28] and expanded by 35 kb upstream and 10 kb downstream to include likely regulatory regions[29]. Gene-wide p-values were calculated from European and Asian summary statistics separately, and meta-analysed within the software. LD reference data files were from the European and East Asian populations of the Haplotype Reference Consortium[30]. We performed 100,000 permutations to estimate p-values corrected for multiple testing under the family-wise error rate (FWER) approach[31]. Within each collection of gene sets, we carried out stepwise conditional analyses as described [32], selecting the most significant set in that collection, which was then added as a covariate to a regression with all other sets. This was repeated until none of the remaining sets showed a nominally significant uncorrected p-value. Specifically, we tested the following gene sets:

(i) Gene ontology: 7,083 sets extracted from the GO database (http://geneontology.org/, accession date: 04/07/2018) curated to include only annotations with experimental or phylogenetic supporting evidence. Gene sets and automated curation pipeline are provided in https://github.com/janetcharwood/pgc3-scz_wg-genesets

(ii) SynGO ontology: Described [33], this collection was analysed as two subsets; "biological process" (135 gene sets) and "cellular component" (60 gene sets). We controlled for a set of 10,360 genes with detectable expression in brain tissue measured as Fragments Per Kilobase of transcript per Million mapped reads (FPKM)[34] to detect synaptic signals above signals simply reflecting the property of brain expression. Gene sets were reconstructed using a "roll-up" method, in which

medRxiv preprint doi: https://doi.org/10.1101/2020.09.12.20192922; this version posted September 13, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. All rights reserved. No reuse allowed without permission.

parent categories contained all genes annotated to child categories. For conditional testing, as the regression covariate, we prioritised more specific child annotations[35].

Conditional Analyses

We performed stepwise conditional analyses of the 270 loci looking for independent associations. We performed association testing and meta-analysis across each locus, adding the allele dosages of the index SNP as a covariate. Where a second SNP had a conditional p-value of less than $1x10^{-6}$, we considered this as evidence for a second signal and repeated the process adding this as an additional covariate. We repeated this until no additional SNPs in the region achieved $p<1x10^{-6}$. We also searched for long range dependencies. Here we tested the all pairs of independent signals for conditional independence (**Supplementary Note).**

Fine-mapping

We used FINEMAP[36] to fine-map loci, excluding the xMHC region due to its complex LD structure. To simplify the causal SNP content of regions [37], for loci formed from non-overlapping clumps which conditional analysis indicated contained independent signals, we fine-mapped those clumps separately. A region in chromosome 5 with long-range LD dependencies (chr5:46-50Mb) was analysed as a single locus. As fine-mapping requires data from all markers at the locus[38] we used only included loci/clumps that attained genome-wide significance (GWS) in the discovery sample. To be conservative, we excluded 13 GWS clumps in the discovery sample that were not GWS in the replication data, although these are most likely true positives. In total, we attempted to fine-map 233 regions. Further details about the fine-mapping process are given in the **Supplementary Note.**

medRxiv preprint doi: https://doi.org/10.1101/2020.09.12.20192922; this version posted September 13, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. All rights reserved. No reuse allowed without permission.

Summary-data-based Mendelian Randomization (SMR) analysis, FUSION and EpiXcan

We used SMR[39] as our primary method to identify SNPs which might mediate association with schizophrenia through effects on gene expression. The significance for SMR is set at the Bonferroni corrected threshold of 0.05/M where M is the number of genes with significant eQTLs tested for a given tissue. Significant SMR associations imply colocalization of the schizophrenia associations with eQTL. We applied the HEIDI test[39] to filter out SMR associations ($P_{HEIDI} < 0.01$) due to linkage disequilibrium between SCZ-associated variants and eQTLs. *cis*-eQTL summary data were from three studies: fetal brain (N=120)[40], adult brain (n = ~1,500)[41] and blood(n = ~32,000)[42]. Linkage disequilibrium (LD) data required for the HEIDI test[39] were estimated from the Health and Retirement Study (HRS)[43] (n = 8,557). We included only genes with at least one *cis*-eQTL at $P_{eQTL} < 5\times10^{-8}$, excluding those in MHC regions due to the complexity of this region. For blood, we included only genes with eQTLs in brain. This left 7,803 genes in blood, 10,890 genes in prefrontal cortex and 754 genes in fetal brain for analysis (see **Supplementary Note** for further details).

For genomic regions where there were multiple genes showing significant SMR associations, we attempted to resolve these with conditional analysis using GCTA-COJO[44,45]. We selected the top-associated *cis*-eQTL for one gene (or a set of genes sharing the same *cis*-eQTL) ran a COJO analysis in the schizophrenia GWAS data and the eQTL data for each of the other genes conditioning on the selected top *cis*-eQTL. We then re-ran the SMR and HEIDI analyses using these conditional GWAS and eQTL results.

We used FUSION [46] and EpiXcan [47] as tests of robustness of the SMR results. Details are supplied in the **Supplementary Note** as are our approaches to prioritising SMR associated genes.

medRxiv preprint doi: https://doi.org/10.1101/2020.09.12.20192922; this version posted September 13, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. All rights reserved. No reuse allowed without permission.

METHODS REFERENCES

1. Schizophrenia Working Group of the Psychiatric Genomics Consortium. Biological insights from 108 schizophrenia-associated genetic loci. *Nature* **511**, 421–427 (2014).

2. Lam, M. *et al.* Comparative genetic architectures of schizophrenia in East Asian and European populations. *Nat. Genet.* (2019). doi:10.1038/s41588-019-0512-x

3. Lam, M. *et al.* RICOPILI: Rapid Imputation for COnsortias PIpeLIne. *Bioinformatics* (2019). doi:10.1093/bioinformatics/btz633

4. the Haplotype Reference Consortium. Reference-based phasing using the Haplotype Reference Consortium panel. *Nat. Genet.* **48**, 1443–1448 (2016).

5. Das, S. *et al.* Next-generation genotype imputation service and methods. *Nat. Genet.* **48**, 1284–1287 (2016).

6. O'Connell, J. *et al.* Haplotype estimation for biobank-scale data sets. *Nat. Genet.* **48**, 817–820 (2016).

7. Bycroft, C. *et al.* Genome-wide genetic data on ~500,000 UK Biobank participants. *bioRxiv* 166298 (2017). doi:10.1101/166298

8. Chang, C. C. *et al.* Second-generation PLINK: rising to the challenge of larger and richer datasets. *Gigascience* **4**, 7 (2015).

9. Lee, J. J. *et al.* Gene discovery and polygenic prediction from a genome-wide association study of educational attainment in 1.1 million individuals. *Nat. Genet.* **50**, 1112–1121 (2018).

medRxiv preprint doi: https://doi.org/10.1101/2020.09.12.20192922; this version posted September 13, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. All rights reserved. No reuse allowed without permission.

927 10. Tukiainen, T. *et al.* Landscape of X chromosome inactivation across human tissues.
928 *Nature* **550**, 244–248 (2017).

929 11. International Schizophrenia Consortium. Common polygenic variation contributes to
930 risk of schizophrenia and bipolar disorder. *Nature* **460**, 748–752 (2009).

931 12. Vittinghoff, E. & McCulloch, C. E. Relaxing the rule of ten events per variable in
932 logistic and cox regression. *Am. J. Epidemiol.* **165**, 710–718 (2007).

933 13. Heinze, G. & Ploner, M. *A SAS macro, S-PLUS library and R package to perform*
934 *logistic regression without convergence problems*. (2004).

935 14. Viechtbauer, W. Conducting Meta-Analyses in R with the metafor Package. *J. Stat.*
936 *Softw.* **36**, 1–48 (2010).

937 15. Martínez-Camblor, P. Fully non-parametric receiver operating characteristic curve
938 estimation for random-effects meta-analysis. *Stat. Methods Med. Res.* **26**, 5–20 (2017).

939 16. Lawlor, D. A., Tilling, K. & Davey Smith, G. Triangulation in aetiological
940 epidemiology. *Int. J. Epidemiol.* **45**, 1866–1886 (2016).

941 17. Burgess, S., Butterworth, A. & Thompson, S. G. Mendelian randomization analysis
942 with multiple genetic variants using summarized data. *Genet. Epidemiol.* **37**, 658–665
943 (2013).

944 18. Bowden, J., Smith, G. D. & Burgess, S. Mendelian randomization with invalid
945 instruments: Effect estimation and bias detection through Egger regression. *Int. J.*
946 *Epidemiol.* **44**, 512–525 (2015).

947 19. Bowden, J. *et al.* A framework for the investigation of pleiotropy in two-sample

medRxiv preprint doi: https://doi.org/10.1101/2020.09.12.20192922; this version posted September 13, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. All rights reserved. No reuse allowed without permission.

948 summary data Mendelian randomization. *Stat. Med.* **36**, 1783–1802 (2017).

949 20. Verbanck, M., Chen, C. Y., Neale, B. & Do, R. Detection of widespread horizontal
950 pleiotropy in causal relationships inferred from Mendelian randomization between
951 complex traits and diseases. *Nat. Genet.* **50**, 693–698 (2018).

952 21. GTEx Consortium. Genetic effects on gene expression across human tissues. *Nature*
953 **550**, 204–213 (2017).

954 22. Habib, N. *et al.* Massively parallel single-nucleus RNA-seq with DroNc-seq. *Nat.*
955 *Methods* **14**, 955–958 (2017).

956 23. Skene, N. G. *et al.* Genetic identification of brain cell types underlying schizophrenia.
957 *Nat. Genet.* **50**, 825–833 (2018).

958 24. Zeisel, A. *et al.* Molecular Architecture of the Mouse Nervous System. *Cell* **174**, 999-
959 1014.e22 (2018).

960 25. Bryois, J. *et al.* Genetic identification of cell types underlying brain complex traits
961 yields insights into the etiology of Parkinson's disease. *Nat. Genet.* 1–12 (2020).
962 doi:10.1038/s41588-020-0610-9

963 26. de Leeuw, C. A., Mooij, J. M., Heskes, T. & Posthuma, D. MAGMA: Generalized
964 Gene-Set Analysis of GWAS Data. *PLOS Comput. Biol.* **11**, e1004219 (2015).

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

967 28. Durinck, S., Spellman, P. T., Birney, E. & Huber, W. Mapping identifiers for the
968 integration of genomic datasets with the R/ Bioconductor package biomaRt. *Nat.*

medRxiv preprint doi: https://doi.org/10.1101/2020.09.12.20192922; this version posted September 13, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. All rights reserved. No reuse allowed without permission.

969 *Protoc.* **4**, 1184–1191 (2009).

970 29. Maston, G. A., Evans, S. K. & Green, M. R. Transcriptional Regulatory Elements in
971 the Human Genome. *Annu. Rev. Genomics Hum. Genet.* **7**, 29–59 (2006).

972 30. the Haplotype Reference Consortium. A reference panel of 64,976 haplotypes for
973 genotype imputation. *Nat. Genet.* **48**, 1279–1283 (2016).

974 31. Cox, D. D. & Lee, J. S. Pointwise testing with functional data using the Westfall-
975 Young randomization method. *Biometrika* **95**, 621–634 (2008).

976 32. Pardiñas, A. F. *et al.* Common schizophrenia alleles are enriched in mutation-
977 intolerant genes and in regions under strong background selection. *Nat. Genet.* **50**,
978 381–389 (2018).

979 33. Koopmans, F. *et al.* SynGO: An Evidence-Based, Expert-Curated Knowledge Base for
980 the Synapse. *Neuron* **103**, 217-234.e4 (2019).

981 34. Genovese, G. *et al.* Increased burden of ultra-rare protein-altering variants among
982 4,877 individuals with schizophrenia. *Nat. Neurosci.* **19**, 1433–1441 (2016).

983 35. Merico, D., Isserlin, R., Stueker, O., Emili, A. & Bader, G. D. Enrichment Map: A
984 Network-Based Method for Gene-Set Enrichment Visualization and Interpretation.
985 *PLoS One* **5**, e13984 (2010).

986 36. Benner, C. *et al.* FINEMAP: Efficient variable selection using summary data from
987 genome-wide association studies. *Bioinformatics* **32**, 1493–1501 (2016).

988 37. van de Bunt, M., Cortes, A., Brown, M. A., Morris, A. P. & McCarthy, M. I.
989 Evaluating the Performance of Fine-Mapping Strategies at Common Variant GWAS

medRxiv preprint doi: https://doi.org/10.1101/2020.09.12.20192922; this version posted September 13, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. All rights reserved. No reuse allowed without permission.

Loci. *PLOS Genet.* **11**, e1005535 (2015).

38. Benner, C. *et al.* Prospects of Fine-Mapping Trait-Associated Genomic Regions by Using Summary Statistics from Genome-wide Association Studies. *Am. J. Hum. Genet.* **101**, 539–551 (2017).

39. Zhu, Z. *et al.* Integration of summary data from GWAS and eQTL studies predicts complex trait gene targets. *Nat. Genet.* **48**, 481–487 (2016).

40. O'Brien, H. E. *et al.* Expression quantitative trait loci in the developing human brain and their enrichment in neuropsychiatric disorders. *Genome Biol.* **19**, 194 (2018).

41. Gandal, M. J. *et al.* Transcriptome-wide isoform-level dysregulation in ASD, schizophrenia, and bipolar disorder. *Science (80-.).* **362**, (2018).

42. Võsa, U. *et al.* Unraveling the polygenic architecture of complex traits using blood eQTL metaanalysis. *bioRxiv* 447367 (2018). doi:10.1101/447367

43. Sonnega, A. *et al.* Cohort profile: The Health and Retirement Study (HRS). *Int. J. Epidemiol.* **43**, 576–585 (2014).

44. Yang, J., Lee, S. H., Goddard, M. E. & Visscher, P. M. GCTA: A tool for genome-wide complex trait analysis. *Am. J. Hum. Genet.* **88**, 76–82 (2011).

45. Yang, J. *et al.* Conditional and joint multiple-SNP analysis of GWAS summary statistics identifies additional variants influencing complex traits. *Nat. Genet.* **44**, 369–375 (2012).

46. Gusev, A. *et al.* Integrative approaches for large-scale transcriptome-wide association studies. *Nat. Genet.* **48**, 245–252 (2016).

medRxiv preprint doi: https://doi.org/10.1101/2020.09.12.20192922; this version posted September 13, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity.
All rights reserved. No reuse allowed without permission.

1011 47. Zhang, W. *et al.* Integrative transcriptome imputation reveals tissue-specific and
1012 shared biological mechanisms mediating susceptibility to complex traits. *Nat.*
1013 *Commun.* **10**, (2019).