Mapping genomic loci implicates genes and synaptic biology in schizophrenia

SUMMARY

Schizophrenia has a heritability of 60-80%, much of which is attributable to common risk alleles. Here, in a 2-stage genome-wide association study of up to 76,755 people with schizophrenia and 243,649 controls, we report common variant associations at 287 distinct genomic loci. Associations were concentrated in genes expressed in CNS neurons, excitatory and inhibitory, but not other tissues or cell types. Using fine-mapping and functional genomic data, we identify 120 genes (106 protein-coding) as likely to underpin associations at some of these loci, including 16 genes with credible causal non-synonymous or UTR variation. We also implicate fundamental processes related to neuronal function, including synaptic organisation, differentiation, and transmission. 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 neurodevelopmental disorders. We identify biological processes relevant to schizophrenia pathophysiology, show convergence of common and rare variant associations in schizophrenia and neurodevelopmental disorders, and provide a rich resource of prioritised genes and variants to advance mechanistic studies.

MAIN TEXT

INTRODUCTION

Schizophrenia typically manifests in late adolescence or early adulthood[1] and is associated with reduced life expectancy, elevated risk of suicide[2], serious physical illnesses[3], 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.

Much of the between-individual variation in risk is genetic, involving large numbers of common alleles,[5] 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 of the disorder to date and analysed the findings to prioritise variants, genes and biological processes that contribute to pathogenesis.

RESULTS

Association Meta-Analysis

We carried out a **primary GWAS** in 74,776 cases and 101,023 controls followed by an **Extended GWAS** which included additional data for the most significant SNPs **(Methods)**. In the primary GWAS, we combined by meta-analysis i) individual genotypes from a **core PGC dataset** of 90 cohorts of European (EUR) and East Asian (ASN) ancestry from the Psychiatric Genomics Consortium (PGC) totalling 67,390 cases and 94,015 controls. ii) summary-level data from 7,386 cases and 7,008 controls from 9 cohorts of African-American (AA) and Latino (LAT) ancestry[10]. We analysed up to 7,585,078 SNPs with MAF ≥ 1% in 175,799 individuals of whom 74.3% were EUR, 17.5% ASN, 5.7% AA, and 2.5% LAT (**Supplementary Cohort Descriptions**). This **primary GWAS** identified 313 independent SNPs (linkage disequilibrium (LD) r2 < 0.1) that exceeded genome-wide significance ($p<5x10^{-8}$) **(Extended Data Figure 1; Supplementary Table 1**), spanning 263 distinct loci.

In the **Extended GWAS,** we meta-analysed the primary GWAS results with summary statistics from deCODE Genetics (1,979 cases, 142,626 controls) for index SNPs with $P<10^{-5}$ and identified 342 LD-independent significant SNPS (**Supplementary Table 2**) located in 287 loci (**Supplementary Table 3; Supplementary Figures 1-2).** Comparisons with the 128 associations (108 loci) we reported in 2014 are provided (**Supplementary Note**); one association (rs3768644; chr2:72.3Mb) is no longer supported[11].

Separate GWAS for males and females had a genetic correlation statistically indistinguishable 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 reported sex differences in age at onset, symptom profile, course, and outcome[12].

SNP-based heritability and Polygenic Prediction

In the EUR sample, the SNP-based heritability (h^2_{SNP}) (i.e. proportion of variance in liability attributable to all measured SNPs) was estimated[13] to be 0.24 (SE 0.007). Using the all ancestry primary GWAS as the discovery sample, polygenic risk score (PRS) analysis explained a median of 0.073 of variance in liability (SNPs with GWAS $p<0.05$), and 0.024 when restricted to genome-wide significant SNPs. For almost all cohorts, PRS had more explanatory power based on risk alleles derived from the larger combined ancestry GWAS than from the matched ancestry GWAS; given the ancestry specific sample sizes, unsurprisingly[9], this effect was strongest for the non-EUR samples (**Extended Data Figure 2 Supplementary Table 5**).

PRS explained most variance in liability in cohorts of European ancestry (again a result of the ancestry composition of the GWAS[9]) and in samples which by ascertainment likely include the most severe cases (hospitalised patients or those treated with clozapine) **(Supplementary Note).** However, even in EUR cohorts, the median Area Under the Receiver Operating Characteristic Curve (AUC) is only 0.72, meaning the liability explained is insufficient for predicting diagnosis in the general population. Nevertheless, as a quantitative estimate of liability to schizophrenia, PRS has applications in research, and in those contexts, PRS can index substantial differences in liability between individuals in the primary GWAS. Compared to the lowest centile of PRS, the highest centile of PRS has an OR for schizophrenia of 39 (95% CI=29-53), and 5.6 (CI 4.9-6.5) when the top centile is compared with the remaining 99% of individuals (**Supplementary Table**

6). An extended discussion of heritability and polygenic prediction is provided in the **Supplementary Note.**

Post-GWAS processing

We next performed a number of secondary analyses in the core PGC dataset in which individual genotypes were available based on fully aligned QC and imputation procedures, and where the data in the HRC reference dataset allowed us to account for LD.

Gene Set Enrichments

Tissue and cell types

Genes with relatively high specificity for bulk expression in every tested region of human brain[14] were significantly enriched for associations **(Extended Data Figure 3**. Comparison with our earlier studies[11,15] shows increasingly clear contrast between the enrichments in brain and non-brain tissues. More strongly than in prior studies[16], from human single cell expression data[17], we found associations were enriched in genes with high expression in excitatory glutamatergic neurons from cerebral cortex and hippocampus (pyramidal CA1 and CA3 cells, and granule cells of dentate gyrus) and also human cortical inhibitory interneurons **(Figure 4a)**. In mouse single-cell RNA-seq data[16], we found similar patterns of enrichments in genes with high expression in excitatory glutamatergic pyramidal neurons from the cortex and hippocampus **(Figure 4b)**, and inhibitory cortical interneurons. We also found associations were enriched in inhibitory medium spiny neurons, the predominant cells of the striatum.

Supportive results were also obtained using a different dataset of 265 cell types in the mouse central and peripheral nervous system[18]. Very strong enrichments were again seen for genes expressed in excitatory glutamatergic neurons of the cortex (especially the deep layers) and hippocampus but also the amygdala **(Supplementary Figure 3)**. Highly significant enrichments were also seen for other neuronal populations, including as above, inhibitory medium spiny neurones in striatum, but also both excitatory and inhibitory neurons from the midbrain, thalamus and hindbrain, and inhibitory cells from the hippocampus. There was little evidence for enrichment of genes with highly specific expression in glia or microglia. Overall, the findings across all the datasets are consistent with the hypothesis that schizophrenia is primarily a

disorder of neuronal function, but do not suggest that pathology is restricted to a circumscribed brain region.

Associations enriched in Neuronal Ontologies

Of 7,315 gene ontology (GO) classifications 24 were associated with schizophrenia (**Supplementary Table 7**). All were relevant to neuronal function including development, differentiation, and synaptic transmission, and involved multiple cellular components including ion channels, synapses, and both axon and dendritic annotations. Using the expert-curated ontology of the SynGO consortium[19], we further examined the synaptic signal and found that conditionally significant annotations were mainly within postsynaptic terms (**Supplementary Tables 8, 9**), although enrichment was also found for genes involved in synaptic organisation and signalling.

Gene Prioritisation

To facilitate biological interpretation and laboratory follow up, we sought to prioritise specific variants and genes most likely to explain associations using a combination of fine-mapping, transcriptomic analysis, and functional genomic annotations. The initial steps in these procedures were necessarily based on 293 index SNPs (255 loci) that attained significance in the core PGC dataset (**Methods, Supplementary Table 10**), we then focussed on the loci that remained significant in the full Extended GWAS to maximise robustness (**Figure 1**).

Fine-mapping

We performed stepwise analyses **(Supplementary Note)**, conditioning associations in loci on their index SNP (and any subsequent conditionally independent associations) to identify regions

that contained independent signals (conditional $p<10^{-6}$). This analysis supported the existence of independent associations in ~10% of loci **(Supplementary Table 10b)**.

We also employed the Bayesian fine-mapping method implemented in FINEMAP[20] to infer the most likely number of distinct causal variants driving our GWAS results. FINEMAP was based on 255 regions determined by the LD clumping procedure (**Supplementary Table 11e**), after merging clumps if their boundaries physically overlapped and excluding the extended MHC region (**Methods**). For regions predicted to contain 3 or fewer causal variants (N=249; **Figure 1**; **Supplementary Tables 11a, 11b**), we extracted from FINEMAP the posterior probabilities (PP) of being causal for every SNP across the region, and constructed credible sets of SNPs that cumulatively capture 95% of the regional PP (**Supplementary Note**).

For 33 regions, the 95% credible set contained 5 or fewer SNPs (**Supplementary Table 11c)** and for 9, only a single SNP. We highlight rs4766428 (PP>0.99) which is the only credible SNP in a locus that contains 25 genes and is located within *ATP2A2*. Mutations in *ATP2A2* cause Darier Disease[21], which co-segregates with bipolar disorder in several multiplex pedigrees and is associated with bipolar disorder and schizophrenia at a population level[22]. *ATP2A2* encodes a sarcoplasmic/endoplasmic reticulum calcium pump, suggesting that its role in schizophrenia pathogenesis may be through regulating neuronal cytoplasmic calcium levels. The likely relevance of calcium metabolism is also suggested by enrichment for associations in and around voltage-gated calcium channels (**Supplementary Tables 3 and 7**)**.**

We denote as our broad fine-map set 628 genes (435 protein coding) that contained at least one credible SNP (**Figure 1a**). At a genome-wide level, genes that are expressed in brain, and that are relatively intolerant to loss-of-function mutations, are known to be enriched for schizophrenia associations and this was confirmed here (**Figure 2a**). Protein-coding genes in the broad fine-map set were enriched for these properties compared to the other protein-coding genes within the associated regions (**Figure 2b**), consistent with genes in this set having an increased probability of influencing liability to schizophrenia. To identify the most credible causal genes, we prioritised those mapping to the 287 loci that were genome-wide significant in our Extended GWAS that also contained a) at least one nonsynonymous (NS) or untranslated region (UTR) variant with a PP> 0.1 b) the entire credible set (**Supplementary Tables 13, 14**). These protein-coding genes had a greater than 3-fold enrichment for loss of function intolerance compared with other protein-coding genes within the loci that were not tagged by credible SNPs (**Supplementary Table 15; Supplementary Note**), supporting our strategy to delimit credible causal genes.

Among the 70 FINEMAP prioritised genes (64 protein-coding) were 16 genes (protein-coding by definition) based on NS or UTR variants (**Supplementary Table 13**). These include *SLC39A8* in which rs13107325, previously a moderately high credible SNP[23], is now strongly supported as causal (PP > 0.99). Other non-synonymous variants with high PP were found in genes with minimal functional characterization including *THAP8*, *WSCD2*, and in two E3 ubiquitin ligases *PJA1* and *CUL9*. Missense and UTR variants prioritised *interferon regulatory factor 3* (*IRF3* while *KLF6*, a transcription factor, was highlighted by three variants in the 3' UTR. Finally, we identified 61 genes (55 protein-coding) in which the 95% credible set is restricted to a single gene **(Supplementary Table 14)**.

Prioritisation by Gene Expression

To detect GWAS associations that are credibly explained by eQTLs, that is, variants that influence gene expression, we used summary-based Mendelian randomisation (SMR)[24] to find evidence that GWAS signals co-localise with eQTLs (from adult brain[25], fetal brain[26] or whole blood[27]) and the HEIDI test[24] to then reject co-localisations due to LD between distinct schizophrenia-associated and eQTL variants (**Supplementary Table 16**). To retain brain relevance, we considered only findings from blood that replicated in brain. After removing duplicates identified in multiple tissues (**Supplementary Tables 17a-c**), we identified 101 SMR-implicated genes (**Supplementary Table 17d**)**;** the use of alternative methodologies supported the robustness of the SMR findings (**Supplementary Note and Supplementary Table 17e**).

We used three approaches to prioritise genes from these 101 candidates (**Supplementary Note; Supplementary Tables 17f, 17g, 18**). We identified (i) 32 genes as the single SMR-implicated gene at the locus or through conditional analysis of a locus containing multiple candidates: (ii) 16 genes where the putatively causal eQTLs captured 50% or more of the FINEMAP posterior probability (iii) 29 genes where chromatin conformation analysis (Hi-C analysis of adult and fetal brain) suggested that a promoter of that gene interacted with a putative regulatory element containing a FINEMAP credible SNP[28].

After removing duplicates, there were 55 SMR/SMR-Hi-C prioritised genes **(Supplementary Table 12)** of which 46 were protein-coding. Genes where putatively causal eQTLs captured a particularly high FINEMAP PP (>95%) **(Supplementary Table 17g)** include *ACE* encoding angiotensin converting enzyme, the target of a major class of anti-hypertensive drugs

(schizophrenia under-expression), *DCLK3* encoding a neuroprotective kinase[29](schizophrenia under-expression) and *SNAP91* (discussed below; schizophrenia over-expression).

Combining all approaches, FINEMAP and SMR, we prioritised 120 genes of which 106 are protein-coding (**Figure 1; Extended Data Table 1**).

Synaptic Location and Function of Prioritised Genes

Following the findings from the genome-wide enrichment tests, we examined prioritised genes in the context of synaptic location and function in the SynGO database[19] **(Figure 3.** Of the 106 proteins encoded, 15 have synaptic annotations **(Supplementary Table 19)**;
7 postsynaptic, 5 both pre- and post- synaptic, 2 presynaptic, and 1 gene is not mapped to any specific compartment.

The results are consistent with the genome-wide enrichment tests pointing to postsynaptic pathology. However, many prioritised genes had additional locations suggesting that presynaptic pathology may also be involved. The encoded proteins map to 16 unique biological terms in the hierarchy (**Supplementary Table 19**), but there are specific themes. Multiple genes encode receptors and ion channels, including voltage-gated calcium and chloride channels (*CACNA1C*, *CLCN3*), metabotropic receptors (glutamate (*GRM1*) and GABA (*GABBR2*)), and the ligand-gated NMDA receptor subunit (*GRIN2A*). Others involve proteins playing a role in endocytosis (*SNAP91*), synaptic organisation and differentiation (*DLGAP2*, *LRRC4B*, *GPM6A*, *PAK6*), including *PTPRD* a receptor protein tyrosine phosphatase presynaptic organizer that trans-synaptically interacts with multiple postsynaptic cell adhesion molecules (e.g. *IL1RAPL1*), and modulation of chemical transmission (*MAPK3*, *DCC*, *CLCN3*, *DLGAP2*). 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 interactions occur at a limited set of specific synapse types, or whether the diversity points to multiple types in different brain regions.

Convergence of Common and Rare Variant Associations

The Schizophrenia Exome Sequencing Meta-Analysis (SCHEMA) consortium (companion paper) identified 32 genes with damaging ultra-rare mutations associated with schizophrenia (FDR<0.05), including 10 at exome-wide significance. We found both sets of genes were enriched for common variant associations, as were more weakly associated SCHEMA genes down to uncorrected P<0.001 (**Figure 2a, Supplementary Tables 20, 21**). Moreover, within associated loci, protein coding genes containing one or more FINEMAP credible SNPs were enriched for SCHEMA genes relative to other protein-coding genes (**Figure 2b; Supplementary Table 21**). There are rare variant overlaps in liability to schizophrenia, autism spectrum disorder (ASD) and developmental disorder (DD)[8,30,31]. We tested for and found that genes in which rare variants increase risk of ASD and DD[32,33] are also enriched for schizophrenia common variant associations. Moreover, they are also enriched among genes containing FINEMAP credible SNPs **(Figure 2 Supplementary Tables 20, 21)**.

Convergences between rare variants and fine-mapped GWAS signals have been previously observed in other traits e.g.,[34,35], suggesting that genes most strongly implicated by fine-mapping and which have additional support from rare variant data are compelling candidates. Of the 10 exome-wide significant genes identified by SCHEMA[36], two were prioritised candidates from fine-mapping; *GRIN2A* encoding a glutamatergic NMDA receptor subunit, and *SP4*, a

transcription factor highly expressed in brain and which is regulated by NMDA transmission, and also regulates NMDA receptor abundance[37]. Two other genes supported by SCHEMA at FDR<0.05 had strong support from fine-mapping: *STAG1*, which is involved in controlling chromosome segregation and regulating gene expression, and *FAM120A*, which encodes an RNA binding protein. SNPs mapping to these genes had cumulative FINEMAP PP of 0.88 and 0.72 respectively (**Supplementary Table 11b**). The prioritised fine-mapped set also contained 4 genes implicated in DD; a transcriptional regulator (*BCL11B*), the well-known *CACNA1C*[38], and genes mentioned elsewhere in this paper (*GRIN2A* and *SLC39A8*). Genes encoding additional transcriptional regulators are also of note; *RERE, FOXP1* and *MYT1L*. *RERE* was prioritised by SMR and is associated with *DD*. *FOXP1* and *MYT1L* are associated with both DD and ASD and met our fine-mapping prioritisation criteria in the core PGC dataset (**Supplementary Table 12**).

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 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 role in the disorder. Importantly, this convergence also implies that the pathogenic relevance of altered function of these genes extends beyond the small proportion of cases carrying rare mutations. We also demonstrate that common variant schizophrenia associations are enriched at genes implicated in neurodevelopmental disorders, opening the door for using the increasing power of rare variant studies of those disorders to further prioritise genes

from GWAS studies. Exploiting this, in addition to *GRIN2A* we identify *BCL11B, CACNA1C, RERE, FOXP1, MYT1L* and *SLC39A8* as genes with strong support.

Enrichment of common variant associations was restricted to genes expressed in CNS neurons, both excitatory and inhibitory, and fundamental biological processes related to neuronal function. This points to neurons as the most important site 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, association with a broad range of cognitive impairments, and lack of regional specificity in neuroimaging measures[1].

Disrupted neuronal function in schizophrenia is unlikely to be restricted to the synapse, but the concentration of associations in genes with pre- and post-synaptic locations, and with functions related to synaptic organisation, differentiation and transmission, point to the pathophysiological importance of these neuronal compartments and their attendant functions. This is further supported by studies showing substantial effects on schizophrenia risk of CNVs[39] and rare damaging coding variants in genes with similar functions, including some of the same genes (SCHEMA; companion paper). Genomic studies, therefore, converge in highlighting these areas of biology as targets for research aiming for a mechanistic understanding of the disorder; the large number of prioritised genes and variants identified here offer 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 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.

MAIN TEXT REFERENCES

1. Owen, M. J., Sawa, A. & Mortensen, P. B. Schizophrenia. *The Lancet* vol. 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. 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.* (2020) doi:10.1038/s41593-019-0565-2.

9. 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.

10. Bigdeli, T. B. *et al.* Contributions of common genetic variants to risk of schizophrenia among individuals of African and Latino ancestry. *Mol. Psychiatry* **25**, 2455–2467 (2020).

11. 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 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. Aguet, F. *et al.* Genetic effects on gene expression across human tissues. *Nature* **550**, 204–213 (2017).

15. Genome-wide association study identifies five new schizophrenia loci. *Nat. Genet.* **43**, 969–978 (2011).

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

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

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

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

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

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

22. 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).

23. 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).

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

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

26. 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).

27. Võsa, U. *et al.* Large-scale cis- and trans-eQTL analyses identify thousands of genetic loci and polygenic scores that regulate blood gene expression. *Nat. Genet.* **53**, 1300–1310 (2021).

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

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

30. Rees, E. *et al.* Analysis of intellectual disability copy number variants for association with schizophrenia. *JAMA Psychiatry* **73**, 963–969 (2016).

31. Fromer, M. *et al.* De novo mutations in schizophrenia implicate synaptic networks. *Nature* **506**, 179–184 (2014).

32. Kaplanis, J. *et al.* Integrating healthcare and research genetic data empowers the discovery of 49 novel developmental disorders. *bioRxiv* (2019) doi:10.1101/797787.

33. Satterstrom, F. K. *et al.* Large-Scale Exome Sequencing Study Implicates Both Developmental and Functional Changes in the Neurobiology of Autism. *Cell* **180**, 568-584.e23 (2020).

34. Luo, Y. *et al.* Exploring the genetic architecture of inflammatory bowel disease by whole-genome sequencing identifies association at ADCY7. *Nat. Genet.* **49**, 186–192 (2017).

35. Cheng, Y. *et al.* Rare genetic variants affecting urine metabolite levels link population variation to inborn errors of metabolism. *Nat. Commun.* **12**, (2021).

36. Singh, T., Neale, B. M. & Daly, M. J. Exome sequencing identifies rare coding variants in 10 genes which confer substantial risk for schizophrenia on behalf of the Schizophrenia Exome Meta-Analysis (SCHEMA) Consortium*. *medRxiv* 2020.09.18.20192815 (2020).

37. Priya, A., Johar, K. & Wong-Riley, M. T. T. Specificity protein 4 functionally regulates the transcription of NMDA receptor subunits GluN1, GluN2A, and GluN2B. *Biochim. Biophys. Acta - Mol. Cell Res.* **1833**, 2745–2756 (2013).

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

39. 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).

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

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

ACKNOWLEDGEMENTS

The National Institute of Mental Health (USA) provides core funding for the Psychiatric Genomics Consortium (PGC) under 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. The work of the contributing groups was supported by numerous grants from governmental and charitable bodies as well as philanthropic donation (details in Supplementary Note). We acknowledge a substantial contribution from Pamela Sklar (deceased) as one of the PGC PIs, and Ed Scolnick, Chief Scientist Emeritus, Stanley Center of the Broad Institute, whose support for this study was vital. We 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. We are grateful to Catrin Hopkins for illustrations.

The work at Cardiff University was additionally supported by Medical Research Council Centre Grant No. MR/L010305/1 and Program Grant No. G0800509. Dr. Shuhua Xu 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 (XDB38000000) of the Chinese Academy of Sciences, and the Shanghai Municipal Science and Technology Major Project (2017SHZDZX01). Dr. Ole Anreassen was supported by Research Council of Norway (283798, 262656, 248980, 273291, 248828, 248778, 223273); KG Jebsen Stiftelsen, South-East Norway Health Authority, EU H2020 # 847776. Béla Melegh was supported in part by the National Scientific Research Program (NKFIH) K 138669. 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. Dr. Sintia Belangero 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). Milan Macek was supported by LM2018132, CZ.02.1.01/0.0/0.0/18_046/0015515 and IP6003 –VZFNM00064203 to MM Jr. Dr. Celso 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 Innovative Medicines Initiative 2 Joint Undertaking (grant agreement No 115916, Project PRISM, and grant agreement No 777394, Project AIMS-2-TRIALS), Fundación Familia Alonso and Fundación Alicia Koplowitz. Dr. E. Bramon acknowledges support from: National Institute of Health Research UK (grant NIHR200756). Mental Health Research UK John Grace QC Scholarship 2018. An ESRC collaborative award 2020. BMA Margaret Temple Fellowship 2016. Medical Research Council New Investigator Award (G0901310) and MRC Centenary Award (G1100583), MRC project grant G1100583. National Institute of Health Research UK post-doctoral fellowship (PDA/02/06/016). NARSAD Young Investigator Awards 2005 and 2008. Wellcome Trust Research Training Fellowship, Wellcome Trust Case Control Consortium awards (085475/B/08/Z, 085475/Z/08/Z). European Commission Horizon 2020 (747429). NIHR Biomedical Research Centre for Mental Health at the South London and Maudsley NHS Foundation Trust and King's College London. NIHR Biomedical Research Centre at University College London Hospitals NHS Foundation Trust and University College London (UCLH BRC - Mental Health Theme). Dr. Dolores Moltó is funded by the European Regional Development Fund (ERDF)-Valencian Community 2014-2020, Spain. Dr. Elizabeth G. Atkinson was supported by the NIMH K01MH121659.

AUTHOR CONTRIBUTIONS

The management group for this paper was led by MOD and JTRW with SR responsible for primary analytic matters supported by BMN and MJD. The management group was comprised of a subset of the PIs of the component studies, bioinformaticians, and analysts and were responsible for study design, conduct, management, primary and final interpretation and included OAA, BTB, SIB, ADB, DB, EB, SC, ACor, DCu, MJD, MDF, ED, HE, AHF, PVG, MGi, SJG, KSH, HHu, NI, RSK, KSK, JAK, JLe, TL, DFL, JLi, AMcI, AMcQ, VAM, DWM, BJM, BMN, MOD, RAO, MJO, AP, DPos, SQ, BPR, SR, DR, SGS, ASe, YS, EAS, PFS, MTT, MPV, JTRW, DRW, TW, NRW, XY, WY.

GWAS meta-analyses (SA, GP, SR, VT); Replication data (SMag, HS, KSt [deCODE]); African-American and Latino sample analyses (EGA, TB, GG, SR, VT); Bioinformatics (JBr, JCH, AFP, AJP, DPos, PFS, KW, SynGO consortium); Comparison of males and females (SR, JSi, VT, PMV); Heritability and Polygenic Prediction (OAA, OF, TG, HHu, BMN, MOD, AFP, ALR, SR, VT, JTRW, NRW, JZ); Phenotype stratification (CAD, EVa); Cellular and Tissue analysis (JBr, MOD, DPos, PFS, JTRW, KW); Gene Ontology (JCH, MOD, AFP, AJP, DPos, JTRW, KW); Fine-mapping (CB, MJD, HHu, MLa, MOD, GP, AFP, MP, SR, JTRW); SMR (LSH, MOD, TQ, NRW, YW, JY); Hi-C (DPos, ALR, PFS, JTRW, KW); Other TWAS (MJG, LSH, MKi, PR, GV, WZha); Integration of fine-mapping, gene expression, Hi-C informatics, rare variants (LSH, MOD, AFP, TQ, ALR, PFS, JTRW, NRW, YW, JY); SynGO (FK, MOD, AFP, ABS, MV, JTRW); Additional statistical advice (PAH). The remaining authors contributed to the recruitment, phenotyping, genotyping, or data processing for the contributing components of the meta-analysis, or provided other forms of functional annotation data. Primary drafting and editing of the manuscript was coordinated by SR, JTRW, and MOD. The primary draft sections were written by JBr, CYC, CAD, LSH, HHu, BMN, MOD, MJO, AFP, AJP, SR, ABS, PFS, VT, EVa, MV, JTRW, NRW, JY. Additional edits were from OAA, MJD, KSK. Numerous other authors provided edits, comments and suggestions, and all authors saw and approved the contents of the manuscript. The Chair of the Psychiatric Genomics Consortium is PFS and the Schizophrenia Working Group of the PGC is led MOD and JTRW.

VT and AFP made equal contributions

Correspondence to MOD, JTRW or SR

CONFLICTS OF INTEREST

Aarno Palotie is a member of Astra Zeneca's Genomics Advisory Board.
Veikko Salomaa has consulted for Novo Nordisk and Sanofi and has ongoing research collaboration with Bayer Ltd (both unrelated to the present study).
Michael Green is a paid consultant for AiCure, Biogen, Lundbeck, and Roche, is a member of the Scientific Board of Cadent, and has received research funds from Forum.
Gregory Light has consulted to Astellas, Forum, and Neuroverse
Keith Nuechterlein has research support from Janssen, Genentech, and Brain Plasticity Inc. Also has consulted to Astellas, MedinCell, Takeda, Teva, Genentech, Otsuka, Janssen, and Brain Plasticity Inc.
David Cohen has reported past consultation for or the receipt of honoraria from Otsuka, Shire, Lundbeck, Roche and Janssen.
Mark Daly is a founder of Maze Therapeutics.
Anil K. Malhotra is a consultant to Genomind Inc, InformedDNA, and Concert Pharmaceuticals.
Rodrigo Affonseca BressanOle has received research grants from Janssen; has been a forum consultant for Janssen and Sanof; Roche; speaker bureau for Ache, Janssen, Sanofi and Torrent.
Cristiano Noto was on the speakers' bureau and/or has acted as a consultant for Janssen and Daiichi-Sankyo in the last 12 months.
Christos Pantelis has, for the last 3 years, served on an advisory board for Lundbeck and received honoraria for talks presented at educational meetings organized by Lundbeck.
David A Collier is a full-time employee and stockholder of Eli Lilly and Company.
Michael O'Donovan is supported by a collaborative research grant from Takeda Pharmaceuticals.
Michael Owen is supported by a collaborative research grant from Takeda Pharmaceuticals.
James Walters is supported by a collaborative research grant from Takeda Pharmaceuticals.
Andrew Pocklington is supported by a collaborative research grant from Takeda Pharmaceuticals.
Stephen R. Marder has consulted for the following companies: Roche, Sunovion, Lundbeck, Boeringer-Ingelheim, Acadia, and Merck.
Srihari Gopal is a full time employee and shareholder Johnson & Johnson (AMEX: JNJ).
Adam Savitz is an employee of Janssen Research & Development, LLC and own stock/stock options in the company.
Qingqin Li is an employee of Janssen Research & Development, LLC and own stock/stock options in the company.
Tony Kam-Thong is an employee of F.Hoffman-La Roche
Anna Rautanen is an employee of F.Hoffman-La Roche
Dheeraj Malhotra is an employee of F.Hoffman-La Roche
Sara Paciga is an employee of Pfizer Inc.
Ole A. Andreassen is a consultant for HealthLytix, and received speaker's honorarium from Lundbeck.
Stephen Faraone has received income, potential income, travel expenses continuing education support and/or research support from, Akili Interactive Labs, Arbor, Genomind, Ironshore, Ondosis, Otsuka, Rhodes, Shire/Takeda, Sunovion, Supernus, Tris, and Vallon. With his

institution, he has US patent US20130217707 A1 for the use of sodium-hydrogen exchange inhibitors in the treatment of ADHD. In previous years, he received support from: Alcobra, Aveksham, CogCubed, Eli Lilly, Enzymotec, Impact, Janssen, KemPharm, Lundbeck/Takeda, McNeil, Neurolifesciences, Neurovance, Novartis, Pfizer, and Vaya. He also receives royalties from books published by Guilford Press: Straight Talk about Your Child's Mental Health; Oxford University Press: Schizophrenia: The Facts; and Elsevier: ADHD: Non-Pharmacologic Interventions. He is also Program Director of www.adhdinadults.com.

Celso Arango has been a consultant to or has received honoraria or grants from Acadia, Angelini, Gedeon Richter, Janssen Cilag, Lundbeck, Minerva, Otsuka, Roche, Sage, Servier, Shire, Schering Plough, Sumitomo Dainippon Pharma, Sunovion and Takeda.

Köksal Alptekin has received grants and honoraria for consulting work, lecturing and research from Abdi İbrahim, Abdi İbrahim Otsuka, Janssen, Ali Raif and TUBITAK.

MAIN FIGURE LEGENDS

Figure 1: Overview of GWAS and gene prioritisation.

Flow diagram summarising GWAS, fine-mapping and SMR analyses and how these informed gene prioritisation.

Figure 2: Gene set enrichment tests at genome-wide level and for protein coding genes containing FINEMAP credible SNPs.

Gene sets tested were retrieved from sequencing studies of schizophrenia (SCHEMA; companion paper), autism-spectrum disorder[33] and developmental disorders[32]. Sets representing genes that are intolerant to loss-of function mutations[40] (LoF-intolerant) and brain-expressed genes[41]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% CI) for association between protein-coding genes containing at least 1 credible FINEMAP SNP (N=418 after excluding genes with no LoF-intolerance data) and genes from the sets indicated. Odds-ratios are relative to protein-coding genes within GWAS K≤3.5 loci (1,283 genes, squares) or across the genome excluding the xMHC (19,547 genes; circles). Dotted line indicates no enrichment.

Figure 3: Mapping of all FINEMAP/SMR genes (A) and prioritised 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. **A)** FINEMAP/SMR genes are protein coding genes tagged by at least one credible SNP identified by FINEMAP and/or associated using SMR (N=470) of which N=58 are SynGO annotated, 51 to cellular components. **B)** Prioritised (Extended Data Table 1; N=106) of which 15 are SynGO annotated, 14 to cellular components.