

This is an Open Access document downloaded from ORCA, Cardiff University's institutional repository: <https://orca.cardiff.ac.uk/id/eprint/146689/>

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

Citation for final published version:

Trubetskoy, Vassily, Pardinias, Antonio , Qi, Ting, Panagiotaropoulou, Georgia, Awasthi, Swapnil, Bigdeli, Tim B., Bryois, Julien, Chen, Chia-Yen, Dennison, Charlotte A., Hall, Lynsey, Lam, Max, Watanabe, Kyoto, Frei, Oleksandr, Ge, Tian, Harwood, Janet C., Koopmans, Frank, Magnusson, Sigurdur, Richards, Alexander L., Sidorenko, Julia, Wu, Yang, Zeng, Jian, Grove, Jakob, Kim, Minsoo, Li, Zhiqiang, Voloudakis, Georgios, Zhang, Wen, Adams, Mark, Agartz, Ingrid, Atkinson, Elizabeth G., Agerbo, Esben, Al Eissa, Mariam, Albus, Margot, Alexander, Madeline, Alizadeh, Behrooz Z., Alptekin, Köksal, Als, Thomas D., Amin, Farooq, Arolt, Volker, Arrojo, Manuel, Athanasiu, Lavinia, Azevedo, Maria Helena, Bacanu, Silviu A., Bass, Nicholas J., Begemann, Martin, Belliveau, Richard A., Bene, Judit, Benyamin, Beben, Bergen, Sarah E., Blasi, Giuseppe, Bobes, Julio, Bonassi, Stefano, Braun, Alice, Affonseca Bressan, Rodrigo, Bromet, Evelyn J., Bruggeman, Richard, Buckley, Peter F., Buckner, Randy L., Bybjerg-Grauholm, Jonas, Cahn, Wiepke, Cairns, Murray J., Calkins, Monica E., Carr, Vaughan J., Castle, David, Catts, Stanley V., Chambert, Kimberley D., Chan, Raymond C. K., Chaumette, Boris, Cheng, Wei, Cheung, Eric F. C., Chong, Siow Ann, Cohen, David, Consoli, Angele, Cordeiro, Quirino, Costas, Javier, Curtis, Charles, Davidson, Michael, Davis, Kenneth L., De Haan, Lieuwe, Degenhardt, Franziska, DeLisi, Lynn E., Demontis, Ditte, Dickerson, Faith, Dikeos, Dimitris, Dinan, Timothy, Djurovic, Srdjan, Duan, Jubao, Ducci, Giuseppe, Dudbridge, Frank, Eriksson, Johan G., Fananas, Lourdes, Faraone, Stephen V., Fiorentino, Alessia, Forstner, Andreas, Frank, Josef, Freimer, Nelson B., Fromer, Menachem, Frustaci, Alessandra, Gadelha, Ary, Genovese, Giulio, Gershon, Elliot S., Giannitelli, Marianna, Giegling, Ina, Giusti-Rodríguez, Paola, Godard, Stephanie, Goldstein, Jacqueline I., González Peñas, Javier, González-Pinto, Ana, Gopa, Srihari, Gratten, Jacob, Green, Michael F., Greenwood, Tiffany A., Guillin, Olivier, Gülöksüz, Sinan, Gur, Raquel E., Gur, Ruben C., Gutiérrez, Blanca, Hahn, Eric, Hakonarson, Hakon, Haroutunian, Vahram, Hartmann, Annette M., Harvey, Carol, Hayward, Caroline, Henskens, Frans A., Herms, Stefan, Hoffmann, Per, Howrigan, Daniel P., Ikeda, Masashi, Iyegbe, Conrad, Joa, Inge, Julià, Antonio, Kähler, Anna K., Kam-Thong, Tony, Kamatani, Yoichiro, Karachanak-Yankova, Sena, Kebir, Oussama, Keller, Matthew C., Kelly, Brian J., Khrunin, Andrey, Kim, Sung-Wan, Klovins, Janis, Kondratiev, Nikolay, Konte, Bettina, Kraft, Julia, Kubo, Michiaki, Kučinskias, Vaidutis, Ausrele Kučinskiene, Zita, Kusumawardhani, Agung, Kuzelov a-Ptackova, Hana, Land, Stefano, Lazzeroni, Laura C., Lee, Phil H., Legge, Sophie E., Lehrer, Douglas S., Lencer, Rebecca, Lerer, Bernard, Li, Miaoxin, Lieberman, Jeffrey, Light, Gregory A., Limborska, Svetlana, Liu, Chih-Min, Lönnqvist, Jouko, Loughland, Carmel M., Lubiski, Jan, Luykx, Jurjen J., Lydam, Amy , Macel Jr., Milan, Mackinnon, Andrew, Magnusson, Patrik K. E., Marder, Eyal, Brion S., Maier, Wolfgang, Malaspina, Dolores, Mallet, Jacques, Marder, Stephen R., Marsal, Sara, Martin, Alicia R., Martorell, Lourdes, Mattheisen, Manuel, McCarley, Robert W., McDonald, Colm, McGrath, John J., Medeiros, Helena, Meier,

Sandra, Melegh, Bela, Melle, Ingrid, Mesholam-Gately, Raquelle I., Metspalu, Andres, Michie, Patricia T., Milani, Lili, Milanova, Vihara, Mitjans, Marina, Molden, Espen, Molina, Esther, Molto, María Dolores, Mondelli, Valeria, Moreno, Carmen, Morley, Christopher P., Muntané, Gerard, Murphy, Kieran C., Myin-Germeys, Inez, Nenadić, Igor, Nestadt, Gerald, Nikitina-Zake, Liene, Noto, Cristiano, Neuchterlein, Keith H., O'Brien, Niamh Louise, O'Neill, F. Anthony., Oh, Sang-Yun, Olincy, Ann, Ota, Vanessa Kiyomi, Pantelis, Christos, Papadimitrou, George N., Paredella, Mara, Paurio, Tiina, Pellegrino, Renata, Periyasamy, Sathish, Perkins, Diana O., Pfuhlmann, Bruno, Pietilainen, Olli, Pimm, Jonathan, Porteous, David, Powell, John, Quattrone, Diego, Quested, Digby, Radant, Allen D., Rampino, Antonio, Rapaport, Mark H., Rautanen, Anna, Reichenberg, Abraham, Roe, Cheryl, Roffman, Joshua L., Roth, Julian, Rothermundt, Matthias, Rutten, Bart P. F., Saker-Delye, Safaa, Salomaa, Veiko, Sanjuan, Julio, Santoro, Marcos Leite, Savitz, Adam, Schall, Ulrich, Scott, Rodney J., Seidman, Larry J., Sharp, Sally Isabel, Shi, Jianxin, Sigurdsson, Engilbert, Sim, Kang, Skarabis, Nora, Slominsky, Petr, So, Hon-Cheong, Sobell, Janet L., Soderman, Erik, Stain, Helen J., Steen, Nils Eiel, Steixner-Kumar, Agnes, Stogmann, Elisabeth, Stone, William S., Straub, Richard E., Streit, Fabian, Stregman, Eric, Stroup, T. Scott, Subramaniam, Mythily, Sugar, Catherine A., Suvisaari, Jaana, Svrakic, Dragan M., Swerdlow, Neal R., Szatkiewicz, Jin P., Ta, Thi Minh Tam, Takahashi, Atsushi, Terao, Chikashi, Thibaut, Florence, Toncheva, Draga, Tooney, Paul A., Torretta, Silvia, Tosato, Sarah, Tura, Gian Battista, Turetsky, Bruce I., Ucock, Alp, Vaaler, Arne, Van Amelsvoort, Therese, Van Winkel, Ruud, Veijola, Juha, Waddington, John, Walter, Henrik, Waterreus, Anna, Webb, Bradley T., Weiser, Mark, Williams, Nigel M. , Witt, Stephanie H., Wormley, Brandon K., Wu, Jing Qin, Xu, Zhida, Yolken, Robert, Zai, Clement C., Zhou, Wei, Zhu, Feng, Zimprich, Fritz, Atbaşoğlu, Eşref Cem, Ayub, Muhammad, Benner, Christian, Bertolino, Alessandro, Black, Donald W., Bray, Nicholas J. , Breen, Gerome, Buccola, Nancy G., Byerley, William F., Chen, Wei J., Cloninger, C. Robert, Crespo-Facorro, Benedicto, Donohoe, Gary, Freedman, Robert, Galletly, Cherrie, Ganda, Michael J., Gennarelli, Massimo, Hougaard, David M., Hwu, Hai-Gwo, Jablensky, Assen V., McCarroll, Steven A., Moran, Jennifer L., Mors, Ole, Mortensen, Preben, Müller-Myhsok, Bertram, Neil, Amanda L., Nordentoft, Merete, Pato, Michele T., Petryshen, Tracey L., Pirinen, Matti, Pulver, Ann E., Schulze, Thomas G., Silverman, Jeremy M., Smoller, Jordan W., Stahl, Eli A., Tsuang, Debby W., Vilella, Elisabet, Wang, Shi-Heng, Xu, Shuhua, Indonesia Schizophrenia Consortium, PsychENCODE, Psychosis Endophenotypes International Consortium, The SynGO Consortium, Adolfsson, Rolf, Arango, Celso, Baune, Bernhard T., Belangero, Sintia Iole, Borglum, Anders D., Braff, David, Bramon, Elvira, Buxbaum, Joseph D., Champion, Dominique, Cervilla, Jorge A., Cichon, Sven, Collier, David A., Corvin, Aidan, Curtis, David, Di Forti, Mara, Domenici, Enrico, Ehrenreich, Hannelore, Escott-Price, Valentina , Esko, Tonu, Fanous, Ayman H., Gareeva, Anna, Gawlik, Micha, Gejman, Pablo V., Gill, Michael, Glatt, Stephen J., Golimbet, Vera, Hong, Kyung Sue, Hultman, Christina M., Hyman, Steven E., Iwata, Nakao, Jonsson, Erik G., Kahn, Rene S., Kennedy, James L., Khusnutdinova, Elza, Kirov, George , Knowles, James A., Krebs, Marie-Odile, Laurent-Levinson, Claudine, Lee, Jimmy, Lencz, Todd, Levinson, Douglas F., Li, Qingqin S, Liu, Jianjun, Malhotra, Anil K., Malhotra, Dheeraj, McIntosh, Andrew, McQillin, Andrew, Menezes, Paolo R., Morgan, Vera A., Morris, Derek W., Mowry, Bryan J., Murray, Robin M., Nimgaonkar, Vishwajit, Nothen, Markus M., Ophoff, Roel A., Paciga, Sara A., Palotie, Aarno, Pato, Carlos N., Qin, Shengying, Rietschel, Marcella, Riley, Brien P., Rivera, Margarita, Rujescu, Dan, Saka, Meram C., Sanders, Alan R., Schwab, Sibylle G., Serretti, Alessandro, Sham, Pak C., Shi, Yongyong, St Clair, David, Stefansson, Hreinn, Stefansson, Kari, Tsuang, Ming T., Van Os, Jim, Vawter, Marquis P., Weinberger, Daniel R., Werge, Thomas, Wildenauer, Dieter B., Yu, Xin, Yue, Weihua, Holmans, Peter A. , Pocklington, Andrew , Roussos, Panos, Vassos, Evangelos, Verhage,

Matthijs, Visscher, Peter M., Yang, Jian, Posthuma, Danielle, Andreassen, Ole A., Kendler, Kenneth S., Owen, Michael J., Wray, Naomi R., Daly, Mark J., Huang, Hailiang, Neale, Benjamin M., Sullivan, Patrick F., Ripke, Stephan, Walters, James T. R., O'Donovan, Michael C. and Schizophrenia Working Group of the Psychiatric Genomics Consorti 2022. Mapping genomic loci prioritises genes and implicates synaptic biology in schizophrenia. *Nature* 604, pp. 502-508. 10.1038/s41586-022-04434-5

Publishers page: <http://doi.org/10.1038/s41586-022-04434-5>

Please note:

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

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

**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<sup>1</sup> and is associated with reduced life expectancy, elevated risk of suicide<sup>2</sup>, serious physical illnesses<sup>3</sup>, 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<sup>4</sup>. 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,<sup>5</sup> rare copy number variants (CNVs)<sup>6</sup>, and rare coding variants (RCVs)<sup>7,8</sup>. A recent genome-wide association study (GWAS) reported 176 genomic loci containing common alleles associated with schizophrenia<sup>9</sup> 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<sup>10</sup>. We analysed up to 7,585,078 SNPs with MAF  $\geq$  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)  $r^2 < 0.1$ ) that exceeded genome-wide significance ( $p < 5 \times 10^{-8}$ ) (**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<sup>11</sup>.

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<sup>12</sup>.

### *SNP-based heritability and Polygenic Prediction*

In the EUR sample, the SNP-based heritability ( $h^2_{\text{SNP}}$ ) (i.e. proportion of variance in liability attributable to all measured SNPs) was estimated<sup>13</sup> 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<sup>9</sup>, 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<sup>9</sup>) 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<sup>14</sup> were significantly enriched for associations (**Extended Data Figure 3**). Comparison with our earlier studies<sup>11,15</sup> shows increasingly clear contrast between the enrichments in brain and non-brain tissues. More strongly than in prior studies<sup>16</sup>, from human single cell expression data<sup>17</sup>, 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<sup>16</sup>, 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<sup>18</sup>. 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<sup>19</sup>, 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<sup>20</sup> 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<sup>21</sup>, which co-segregates with bipolar disorder in several multiplex pedigrees and is associated with bipolar disorder and schizophrenia at a population level<sup>22</sup>. *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**). 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<sup>23</sup>, 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)<sup>24</sup> to find evidence that GWAS signals co-localise with eQTLs (from adult brain<sup>25</sup>, fetal brain<sup>26</sup> or whole blood<sup>27</sup>) and the HEIDI test<sup>24</sup> 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<sup>28</sup>.

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<sup>29</sup> (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<sup>19</sup> (**Figure 3**). Of the 106 proteins encoded, 15 have synaptic annotations (**Supplementary Table 19**); 7 postsynaptic, 5 both pre- and postsynaptic, 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*, *LRR4B*, *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)<sup>8,30,31</sup>. We tested for and found that genes in which rare variants increase risk of ASD and DD<sup>32,33</sup> 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.,<sup>34,35</sup>, 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<sup>36</sup>, 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<sup>37</sup>. 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*<sup>38</sup>, 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<sup>1</sup>.

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<sup>39</sup> 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 Ca<sup>2+</sup> 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. Vösa, 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.A.C.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 –VZFN00064203 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 Programme; 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, MGj, SJG, KSH, HHu, NI, RSK, KSK, JAK, JLe, TL, DFL, JLi, AMcl, 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 Sanofi; 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](http://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<sup>33</sup> and developmental disorders<sup>32</sup>. Sets representing genes that are intolerant to loss-of function mutations<sup>40</sup> (LoF-intolerant) and brain-expressed genes<sup>41</sup> 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  $K \leq 3.5$  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.

# ONLINE METHODS

## *Table of Contents*

ONLINE METHODS .....	18
Table of Contents .....	18
Overview of Samples .....	18
Association Analysis .....	19
<i>Association / Meta-analysis</i> .....	20
Polygenic Prediction .....	20
Secondary analyses in core PGC dataset .....	21
Gene Set Enrichments .....	21
<i>Tissue and cell types</i> .....	21
Conditional SNP Association Analyses .....	22
Fine-mapping .....	22
Summary-data-based Mendelian Randomization (SMR) analysis, FUSION and EpiXcan .....	22
DATA AVAILABILITY .....	24
CODE AVAILABILITY .....	24
METHODS REFERENCES .....	25
EXTENDED DATA ITEM LEGENDS .....	28

## 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 in **Supplementary Cohort Descriptions**. The core PGC dataset included 90 cohorts for which we had individual level genotype data fully processed under a uniform pipeline. This core dataset contains genotypes on 161,405 unrelated subjects; 67,390 schizophrenia/schizoaffective disorder cases and 94,015 controls, equivalent in power to 73,189 of each. 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<sup>1</sup>. Around 80% of the probands (53,386 cases and 77,258 controls) were of European Ancestry, and the remainder (14,004 cases and 16757 controls) were of East Asian ancestry<sup>2</sup>. We additionally included in the Primary GWAS summary statistics from 9 cohorts comprising African-American (AA; 6152 cases 3918 controls) and Latino (1234 cases, 3090 controls) participants; the combined sample is equivalent in power to 6,551 each of cases and controls. 1249 LD – independent ( $r^2 > 0.1$ ) Variants showing evidence for association ( $P < 1 \times 10^{-5}$ ) were further meta-analysed with an additional dataset of 1,979 cases and 142,626 controls of European ancestry obtained from deCODE genetics, thus the final analysis represents 320,404 diploid genomes.

## Association Analysis

*Technical Quality Control of the 90 cohorts comprising the primary PGC sample.*

Technical Quality control was performed on the core PGC cohorts separately according to standards developed by the PGC<sup>3</sup> 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.

*Genomic Quality Control: Principal Component Analysis (PCA) and Relatedness Checking in the core PGC dataset*

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 ( $r^2 < 0.2$ , 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<sup>4</sup>; 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.

To control for false positive associations due to inflated test statistics we evaluated the effectiveness of the primary technical and genomic quality control parameters on the genome-wide inflation of test statistics using the lambda GC (median)<sup>5</sup> and as necessary made the QC parameters more stringent until this value was between 1.0 and 1.4 (before inclusion of principal components as covariates) and/or between 1.0 and 1.15 after inclusion of PCA covariates. Additionally, we applied loose PCA filters for strongly stratified datasets even if we did not observe strong inflation of test statistics in order to retrieve reliable test statistics (see **Supplementary Figure 4**). Since the core PGC cohorts came from many distinct centres, countries, and continents, various measures (e.g., tightening of the technical QC parameters and/or genomic quality control) had to be taken in an iterative process to achieve this goal.

**Supplementary Table 22** lists detailed per cohort exclusion numbers for individuals in the non-Asian samples. The Asian cohorts were sufficiently homogeneous as they did not show marked population structure in principal component analyses. The exclusion numbers for individuals during **technical QC** are in most cohorts low. For six cohorts (marked in yellow in **Supplementary Table 22**) it was necessary to exclude more than 100 cases during **genomic QC** so that Lambda GC fell within the window mentioned above. **Supplementary Figure 4** gives details about this process and explains why the excluded cases could not be used with the presently available control cohorts for this manuscript.

*Imputation of the core PGC dataset*

Genotype imputation of case-control cohorts was performed using the pre-phasing/imputation stepwise approach implemented in EAGLE 2<sup>6</sup> / MINIMAC3<sup>7</sup> (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<sup>8</sup> 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<sup>9</sup>. 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<sup>10</sup> into the above HRC reference panel.

### *Association / Meta-analysis*

In each individual cohort, association testing was based on an additive logistic regression model using PLINK<sup>11</sup>. 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 extreme ancestry outliers. We conducted a meta-analysis of the results (including the 9 cohorts comprising African-American and Latino participants) 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  $r^2 < 0.1$  within 3 Mb windows. We recorded the left and rightmost variant with  $r^2 < 0.1$  to an index SNP to define an associated clump. To define loci, we added a 50kb window on each side of the LD clump and combined overlapping LD-clumps into a single locus.

Due to the strong signal and high linkage disequilibrium in the MHC, only one SNP was kept from the extended MHC region (chr6:25-35Mb).

We additionally examined the X chromosome for evidence of heterogeneity between the sexes and X chromosome dosage compensation using the methods described by Lee and colleagues<sup>12,13</sup>

**(Supplementary Note)**. To minimise possible confounding effects of ancestry on effect sizes by sex, we restricted this analysis to those of European ancestry.

We obtained summary association results from deCODE genetics for 1,228 index SNPs ( $P < 1 \times 10^{-5}$ ) based on 1,979 cases and 142,626 controls of European ancestry. Genotyping was carried out at deCODE Genetics. We used this sample to establish that SNP associations from the primary GWAS replicated *en masse* in an independent sample (see **Supplementary Note**) by showing the directions of effect of index SNPs differed from the null hypothesis of randomly oriented effects and also comparing the expected number of same direction effects with those if all associations were true, taking into account the discovery magnitude of effect, and the replication effect-estimate precision (**Supplementary Note**).

The summary statistics from deCODE were combined with those from our primary GWAS dataset using an inverse variance-weighted fixed effects model. Similarly to the discovery meta-analysis (see above) we merged overlapping LD-clumps to a total of 287 distinct genomic regions (5 on the X-chromosome) with at least one genome-wide significant signal.

## **Polygenic Prediction**

We estimated the cumulative contribution of SNPs to polygenic risk of schizophrenia using a series of leave-one-out polygenic prediction analyses based on LD-clumping and P-value thresholding (P+T)<sup>14</sup> (also known as C+T) using PLINK<sup>11</sup>. 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;  $5 \times 10^{-8}$ ,  $1 \times 10^{-6}$ ,  $1 \times 10^{-4}$ ,  $1 \times 10^{-3}$ ,  $1 \times 10^{-2}$ ,  $5 \times 10^{-2}$ ,  $1 \times 10^{-1}$ ,  $2 \times 10^{-1}$ ,  $5 \times 10^{-1}$  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<sup>15</sup>, 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"<sup>16</sup>. ORs from these calculations were then

meta-analysed using a fixed-effects model in the R package “metafor”<sup>17</sup>. To ensure stability of the estimates, meta-analysis was conservatively restricted to 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, we calculated liability  $R^2$ , Nagelkerke’s  $R^2$  following Lee et. al. 2012<sup>18</sup> and a combined area under the receiver operating characteristic curve (AUROC). Both liability  $R^2$  and Nagelkerke’s  $R^2$  included any principal components marginally associated with the outcome within each cohort, in the baseline model. AUROC was estimated using the non-parametric meta-analysis implemented in the R package “nsROC”<sup>19</sup>. Polygenic score analysis of the African-American and Latino cohorts were conducted by the authors of the study reporting those datasets<sup>20</sup>.

## Secondary analyses in core PGC dataset

Some of the secondary analyses (Gene-set enrichments, conditional SNP association analyses, fine-mapping) necessitate access to individual level data, require identical QC and imputation procedures, and/or an accurate LD reference panel meaning these analyses could only be reliably performed in a subset of the dataset. The following analyses focussed on the core PGC dataset for which these conditions are met.

### *Gene Set Enrichments*

#### *Tissue and cell types*

We collected bulk RNA-seq data across 53 human tissues (GTEx v8, median across samples)<sup>21</sup>; from a study of 19,550 nuclei from frozen adult human post-mortem hippocampus and prefrontal cortex representing 16 different cell types<sup>22</sup>; 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<sup>23</sup>; from a study of ~500,000 single cells from the mouse nervous system (19 regions) that identified 265 cell types<sup>24</sup>.

Datasets were processed uniformly<sup>25</sup>. First, we calculated the mean expression for each gene for each type of data if these statistics were not provided by the authors. We used the pre-computed median expression (transcript per million (TPM)) across individuals for the GTEx tissues (v8). For the GTEx dataset, we excluded tissues with less than 100 samples, merged tissues by organ (with the exception of brain tissues), excluded non-natural tissues (e.g. EBV-transformed lymphocytes) and testis (outlier in hierarchical clustering), resulting in 37 tissues. Genes without unique names and genes not expressed in any cell types were excluded. We scaled the expression data to 1M Unique Molecular Identifiers (UMIs) or TPM for each cell type/tissue. After scaling, we excluded non-protein coding genes, and, for mouse datasets, genes that had no expert curated 1:1 orthologs between mouse and human (Mouse Genome Informatics, The Jackson laboratory, version 11/22/2016). We then calculated a metric of gene expression specificity by dividing the expression of each gene in each cell type/tissue by the total expression of that gene in all cell types/tissue, leading to values ranging from 0 to 1 for each gene (0: meaning that the gene is not expressed in that cell type/tissue, 1 that 100% of the expression of that gene is performed in that cell type/tissue). We selected the 10% most specific genes per cell type (or tissue) with an expression level of at least 1TPM, or 1 UMI per million, for downstream analyses and used MAGMA v1.08<sup>26</sup> to test whether they were enriched for genetic associations. We performed a one-sided test as we were only interested in enrichments for genetic associations (in contrast with depletions). We also applied partitioned LD score regression (LDSC) as described<sup>27</sup> to the top 10% genes



for each cell type for heritability enrichment. We selected the one-sided coefficient z-score p-value as a measure of the association of the cell type/tissue with schizophrenia.

### *Ontology Gene sets*

Gene set analyses were performed using MAGMA v1.08<sup>26</sup>. Gene boundaries were retrieved from Ensembl release 92 (GRCh37) using the “biomaRt” R package<sup>28</sup> and expanded by 35 kb upstream and 10 kb downstream to include likely regulatory regions<sup>29</sup>. Gene-wide p-values were calculated from European and Asian summary statistics separately using the SNP-wise “mean” Imhof method, and meta-analysed within the software. LD reference data files were from the European and East Asian populations of the Haplotype Reference Consortium<sup>30</sup>. Within each gene set analysis, p-values were corrected for multiple testing using the Bonferroni procedure. Specifically, we tested the following gene sets:

- (i) Gene ontology: 7,315 sets extracted from the GO database (<http://geneontology.org/>, accession date: 09/11/2020) curated to include only annotations with experimental or phylogenetic supporting evidence.
- (ii) SynGO ontology: Described elsewhere<sup>31</sup>, 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)<sup>32</sup> to detect synaptic signals above signals simply reflecting the property of brain expression. Exploiting the hierarchical structure of SynGO, gene sets were reconstructed using a “roll-up” method, in which parent categories contained all genes annotated to child categories. For stepwise conditional testing<sup>33</sup>, we prioritised the most specific child annotations<sup>34</sup> (i.e. the lowest possible level) as regression covariates.

### *Conditional SNP Association Analyses*

We performed stepwise conditional analyses of 248 loci that were genome wide significant in the core PGC dataset 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  $1 \times 10^{-6}$ , 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  $p < 1 \times 10^{-6}$ . 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<sup>35</sup> to fine-map regions defined by LD clumps ( $r^2 > 0.1$ ), excluding the MHC locus due to its complex LD structure. Clumps which overlapped (without adding the additional 50kb used to define physically distinct loci) were combined. As fine-mapping requires data from all markers in the region<sup>36</sup> we only performed fine-mapping on regions that attained genome-wide significance (GWS) in the core PGC GWAS. In total, we attempted to fine-map 255 non-overlapping regions (**Supplementary Table 11e**). Further details about the fine-mapping process are given in the **Supplementary Note**.

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

We used SMR<sup>37</sup> 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<sup>37</sup> to filter out SMR associations ( $P_{\text{HEIDI}} < 0.01$ ) due to linkage disequilibrium between SCZ-associated variants and eQTLs. *cis*-eQTL summary data were from three studies: fetal brain ( $N=120$ )<sup>38</sup>, adult brain ( $n = \sim 1,500$ )<sup>39</sup> and blood ( $n = \sim 32,000$ )<sup>40</sup>. Linkage disequilibrium (LD) data required for the HEIDI test<sup>37</sup> were estimated from the Health and Retirement Study (HRS)<sup>41</sup> ( $n = 8,557$ ). We included only genes with at least one *cis*-eQTL at  $P_{\text{eQTL}} < 5 \times 10^{-8}$ , 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). SMR was performed using data from the primary GWAS. The results were then filtered to exclude significant SMR implicated genes where the eQTLs did not map within our definition of an associated locus in the Extended GWAS meta-analysis of our primary GWAS dataset and the dataset provided by deCODE genetics.

For genomic regions where there were multiple genes showing significant SMR associations, we attempted to resolve these with conditional analysis using GCTA-COJO<sup>42,43</sup>. 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<sup>44</sup> and EpiXcan<sup>45</sup> as tests of robustness of the SMR results. Details are supplied in the **Supplementary Note** as are our approaches to prioritising SMR associated genes.

# DATA AVAILABILITY

Summary statistics for the “Extended”, “Core”, ancestry specific and sex-stratified analyses is available at <https://www.med.unc.edu/pgc/download-results/scz/>. Genotype data are available for a subset of cohorts, including dbGAP accession numbers and/or restrictions, as described in the Supplementary Information section “Cohort Descriptions”.

# CODE AVAILABILITY

Core analysis code for RICOPIII can be found at <https://sites.google.com/a/broadinstitute.org/ricopili/>. This wraps PLINK (<https://www.cog-genomics.org/plink2/>), EIGENSOFT (<https://www.hsph.harvard.edu/alkes-price/software/>), EAGLE2 (<https://alkesgroup.broadinstitute.org/Eagle/>), MINIMAC3 (<https://genome.sph.umich.edu/wiki/Minimac3>), SHAPEIT3 ([https://mathgen.stats.ox.ac.uk/genetics\\_software/shapeit/shapeit.html](https://mathgen.stats.ox.ac.uk/genetics_software/shapeit/shapeit.html)), METAL ([https://genome.sph.umich.edu/wiki/METAL\\_Documentation](https://genome.sph.umich.edu/wiki/METAL_Documentation)), LDSR (<https://github.com/bulik/ldsc>). For downstream analyses, FINEMAP can be found at <http://christianbenner.com/>, and our utility for meta-analysing cohort-specific LD matrices can be found at <https://github.com/Pintaius/LDmergeFM>. MAGMA can be found at <https://ctg.cncr.nl/software/magma> and the GO gene sets and automated curation pipeline are provided in [https://github.com/janetcharwood/pgc3-scz\\_wg-genesets](https://github.com/janetcharwood/pgc3-scz_wg-genesets). SMR is available at <https://cnsgenomics.com/software/smr/> and SbayesS at <https://cnsgenomics.com/software/gctb/>.

# METHODS REFERENCES

1. 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. Purcell, S., Neale, B., Todd-Brown, K., Thomas, L. & Ferreira, M. A. PLINK: a toolset for whole-genome association and population-based linkage analysis. *Am J Hum Genet* **81**, (2007).
5. Devlin, B. & Roeder, K. Genomic Control for Association Studies. *Biometrics* **55**, 997–1004 (1999).
6. Reference-based phasing using the Haplotype Reference Consortium panel. *Nat. Genet.* **48**, 1443–1448 (2016).
7. Das, S. *et al.* Next-generation genotype imputation service and methods. *Nat. Genet.* **48**, 1284–1287 (2016).
8. Haplotype Reference Consortium r1.1. <https://ega-archive.org/datasets/EGAD00001002729>.
9. O’Connell, J. *et al.* Haplotype estimation for biobank-scale data sets. *Nat. Genet.* **48**, 817–820 (2016).
10. Bycroft, C. *et al.* Genome-wide genetic data on ~500,000 UK Biobank participants. *bioRxiv* 166298 (2017) doi:10.1101/166298.
11. Chang, C. C. *et al.* Second-generation PLINK: rising to the challenge of larger and richer datasets. *Gigascience* **4**, 7 (2015).
12. 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).
13. Tukiainen, T. *et al.* Landscape of X chromosome inactivation across human tissues. *Nature* **550**, 244–248 (2017).
14. Common polygenic variation contributes to risk of schizophrenia and bipolar disorder. *Nature* **460**, 748–752 (2009).
15. Vittinghoff, E. & McCulloch, C. E. Relaxing the rule of ten events per variable in logistic and cox regression. *Am. J. Epidemiol.* **165**, 710–718 (2007).
16. Heinze, G. & Ploner, M. A SAS macro, S-PLUS library and R package to perform logistic regression without convergence problems. [https://cemsis.meduniwien.ac.at/fileadmin/user\\_upload/\\_imported/fileadmin/msi\\_akim/CeMSIS/KB/programme/tr2\\_2004.pdf](https://cemsis.meduniwien.ac.at/fileadmin/user_upload/_imported/fileadmin/msi_akim/CeMSIS/KB/programme/tr2_2004.pdf) (2004).

17. Viechtbauer, W. Conducting Meta-Analyses in R with the metafor Package. *J. Stat. Softw.* **36**, 1–48 (2010).
18. Lee, S. H., Goddard, M. E., Wray, N. R. & Visscher, P. M. A Better Coefficient of Determination for Genetic Profile Analysis. *Genet. Epidemiol.* **36**, 214–224 (2012).
19. Martínez-Camblor, P. Fully non-parametric receiver operating characteristic curve estimation for random-effects meta-analysis. *Stat. Methods Med. Res.* **26**, 5–20 (2017).
20. 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).
21. Aguet, F. *et al.* Genetic effects on gene expression across human tissues. *Nature* **550**, 204–213 (2017).
22. Habib, N. *et al.* Massively parallel single-nucleus RNA-seq with DroNc-seq. *Nat. Methods* **14**, 955–958 (2017).
23. Skene, N. G. *et al.* Genetic identification of brain cell types underlying schizophrenia. *Nat. Genet.* **50**, 825–833 (2018).
24. Zeisel, A. *et al.* Molecular Architecture of the Mouse Nervous System. *Cell* **174**, 999–1014.e22 (2018).
25. Bryois, J. *et al.* Genetic identification of cell types underlying brain complex traits yields insights into the etiology of Parkinson’s disease. *Nat. Genet.* 1–12 (2020) doi:10.1038/s41588-020-0610-9.
26. de Leeuw, C. A., Mooij, J. M., Heskes, T. & Posthuma, D. MAGMA: Generalized Gene-Set Analysis of GWAS Data. *PLOS Comput. Biol.* **11**, e1004219 (2015).
27. Finucane, H. K. *et al.* Heritability enrichment of specifically expressed genes identifies disease-relevant tissues and cell types. *Nat. Genet.* **50**, 621–629 (2018).
28. Durinck, S., Spellman, P. T., Birney, E. & Huber, W. Mapping identifiers for the integration of genomic datasets with the R/ Bioconductor package biomaRt. *Nat. Protoc.* **4**, 1184–1191 (2009).
29. Maston, G. A., Evans, S. K. & Green, M. R. Transcriptional Regulatory Elements in the Human Genome. *Annu. Rev. Genomics Hum. Genet.* **7**, 29–59 (2006).
30. A reference panel of 64,976 haplotypes for genotype imputation. *Nat. Genet.* **48**, 1279–1283 (2016).
31. Koopmans, F. *et al.* SynGO: An Evidence-Based, Expert-Curated Knowledge Base for the Synapse. *Neuron* **103**, 217–234.e4 (2019).
32. Genovese, G. *et al.* Increased burden of ultra-rare protein-altering variants among 4,877 individuals with schizophrenia. *Nat. Neurosci.* **19**, 1433–1441 (2016).

33. 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).
34. Merico, D., Isserlin, R., Stueker, O., Emili, A. & Bader, G. D. Enrichment Map: A Network-Based Method for Gene-Set Enrichment Visualization and Interpretation. *PLoS One* **5**, e13984 (2010).
35. Benner, C. *et al.* FINEMAP: Efficient variable selection using summary data from genome-wide association studies. *Bioinformatics* **32**, 1493–1501 (2016).
36. 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).
37. Zhu, Z. *et al.* Integration of summary data from GWAS and eQTL studies predicts complex trait gene targets. *Nat. Genet.* **48**, 481–487 (2016).
38. 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).
39. Gandal, M. J. *et al.* Transcriptome-wide isoform-level dysregulation in ASD, schizophrenia, and bipolar disorder. *Science (80-. )*. **362**, (2018).
40. Vösa, U. *et al.* Unraveling the polygenic architecture of complex traits using blood eQTL metaanalysis. *bioRxiv* 447367 (2018) doi:10.1101/447367.
41. Sonnega, A. *et al.* Cohort profile: The Health and Retirement Study (HRS). *Int. J. Epidemiol.* **43**, 576–585 (2014).
42. 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).
43. 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).
44. Gusev, A. *et al.* Integrative approaches for large-scale transcriptome-wide association studies. *Nat. Genet.* **48**, 245–252 (2016).
45. Zhang, W. *et al.* Integrative transcriptome imputation reveals tissue-specific and shared biological mechanisms mediating susceptibility to complex traits. *Nat. Commun.* **10**, (2019).

# EXTENDED DATA LEGENDS

## ***Extended Data Figure 1: Primary GWAS Manhattan plot***

The x-axis indicates chromosomal position and the y-axis is the significance of association ( $-\log_{10}(P)$ ). The red line represents genome-wide significance level ( $5 \times 10^{-8}$ ). SNPs in green are in linkage disequilibrium (LD;  $R^2 > 0.1$ ) with index SNPs (diamonds) which represent LD independent genome-wide significant associations.

## ***Extended Data Figure 2: Polygenic risk prediction***

**A)** Distributions of liability scale  $R^2$  across 98 left-out-cohorts for polygenic risk scores built from SNPs with different p-value thresholds. Distributions of liability  $R^2$  (assuming schizophrenia life-time risk of 1%) are shown for each p-value threshold, with point size representing size of the left-out cohort and colour representing ancestry. The median liability  $R^2$  is represented as a horizontal black line. **B)** Liability  $R^2$  of predicted and observed phenotypes in left-out cohorts using variants with p-value threshold  $p=0.05$ , from the fixed effect meta-analysis of variant effects, unadjusted for multiple comparisons. The polygenic risk scores are derived from two separate sets of leave-one-out GWAS meta-analyses: y-axis  $R^2$  based on the results of primary GWAS including all ancestries; x axis  $R^2$  based on cohorts of the same ancestry as the test samples. Circles denote core PGC samples. Triangles denote African American and Latino samples processed external to PGC by the providing author.

## ***Extended Data Figure 3: Association between 37 human tissues and schizophrenia.***

The mean of the evidence ( $-\log_{10}P$ ) obtained from two methods (MAGMA, LDSC) for testing GWAS data for enrichment of association in genes with high expression in each tissue as determined from bulk RNA-seq<sup>20</sup>. The bar colour indicates whether gene expression in the tissue is significantly associated with both methods, one method or none. The black vertical line represents the significance threshold corrected for the total number of tissues tested in this experiment. We also analysed previous waves of PGC schizophrenia GWAS<sup>11,21</sup> for comparison.

## ***Extended Data Figure 4 Legend: Associations between schizophrenia and cell types from multiple brain regions in human and mouse***

The mean of the evidence ( $-\log_{10}P$ ) obtained from two methods (MAGMA, LDSC) for testing GWAS data for enrichment of associations in genes with high expression in cell types. The 15 human cell types (derived from single nuclei) from the cortex and hippocampus.

## ***Extended Data Table 1: List of prioritized genes***

List of genes meeting prioritisation criteria summarised in Figure 1. Index SNP: index associated SNP for the locus from the GWAS. Ensembl ID: Ensembl gene identifier. Symbol ID: HGNC gene symbol. Gene Biotype: as classified by Ensembl. FINEMAP and SMR priority genes: genes meeting the prioritisation criteria described in the text. Rare priority genes: genes implicated by rare coding variants in

schizophrenia, autism spectrum disorders or developmental disorder. Full details regarding the prioritisation criteria for each gene are given in Supplementary Tables 11-18.