medRxiv preprint doi: https://doi.org/10.1101/2022.08.09.22278128; this version posted August 11, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity.
It is made available under a CC-BY-NC 4.0 International license .

Variant-risk-exon interplay impacts on cortical *cell-cell interactions* in severe psychiatric disorders

Karolina Worf[1,2†] & Natalie Matosin[3,4,5†], Nathalie Gerstner[1,3,6], Anna S. Fröhlich[3,6], Anna C. Koller[7,8], Franziska Degenhardt[7,8], Holger Thiele[9], Marcella Rietschel[10], Madhara Udawela[11], Elizabeth Scarr[13], Brian Dean[11,12], Fabian J. Theis[1,2,14], Janine Knauer-Arloth[1,3‡*] & Nikola S. Mueller[1‡*]

[1] Institute of Computational Biology, Helmholtz Zentrum München, Neuherberg 85764, Germany
[2] TUM School of Life Sciences Weihenstephan, Technical University of Munich, Freising 85354, Germany
[3] Department of Translational Research in Psychiatry, Max Planck Institute of Psychiatry, Munich 80804, Germany
[4] Illawarra Health and Medical Research Institute, Wollongong, NSW 2522, Australia
[5] Molecular Horizons, School of Chemistry and Molecular Bioscience, Faculty of Science, Medicine and Health, University of Wollongong, NSW 2522, Australia
[6] International Max Planck Research School for Translational Psychiatry, Max Planck Institute of Psychiatry, 80804, Munich, Germany
[7] Institute of Human Genetics, University of Bonn, Bonn 53127, Germany
[8] Department of Genomics, Life & Brain Center, University of Bonn, Bonn 53127, Germany
[9] Cologne Center for Genomics, University of Cologne, Cologne 50923, Germany
[10] Department of Genetic Epidemiology in Psychiatry, Central Institute of Mental Health, University Medical Center Mannheim/University of Heidelberg, Mannheim 68159, Germany
[11] The Molecular Psychiatry Laboratory, The Florey Institute of Neuroscience and Mental Health, Parkville, Victoria 3052, Australia
[12] The Department of Florey, The University of Melbourne, Parkville, Victoria, 3050, Australia.
[13] The Department of Psychiatry, The University of Melbourne, Parkville, Victoria, 3050, Australia.
[14] Department of Mathematics, Technical University of Munich, Garching 85748, Germany

† these first-authors authors contributed equally
‡ these senior-authors contributed equally

*** Lead contacts:**

- Nikola S. Mueller, E-mail: nikola.mueller@helmholtz-muenchen.de
- Janine Knauer-Arloth, Tel.: +49 89 30622-554, E-mail: arloth@psych.mpg.de

ABSTRACT

In psychiatric disorders, common and rare genetic variants cause widespread dysfunction of cells and their interactions, especially in the prefrontal cortex, giving rise to psychiatric symptoms. To better understand these processes, we traced the effects of common and rare genetics, and cumulative disease risk scores, to their molecular footprints in human cortical single-cell types. We demonstrated that examining gene expression at single-exon resolution is crucial for understanding the cortical dysregulation associated with diagnosis and genetic risk derived from common variants. We then used disease risk scores to identify a core set of genes that serve as a footprint of common and rare variants in the cortex. Pathways enriched in these genes indicated impaired cell-cell interactions and neuronal activity in psychopathology. With single-nuclei-RNA-sequencing, we pinpointed these effects to inhibitory cortical neurons and oligodendrocyte progenitors, two cell-types that closely interact. This constitutes a clear cellular target for new treatments for psychiatric disorders.

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

medRxiv preprint doi: https://doi.org/10.1101/2022.08.09.22278128; this version posted August 11, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity.
It is made available under a CC-BY-NC 4.0 International license .

INTRODUCTION

Schizophrenia (SCZ) is a complex psychiatric disorder with a life-time risk < 1% (Saha et al., 2005) and a heritability of approximately 60-80% (Wray and Gottesman, 2012). SCZ shares common pathophysiological, clinical and biological characteristics with bipolar disorder (BD) and major depressive disorder (MDD) (Cross-Disorder Group of the Psychiatric Genomics Consortium, 2013; Schulze et al., 2014; Zhu et al., 2019). One origin of psychiatric disorders including SCZ is altered connectivity and plasticity in neural circuits in the brain. The dorsolateral prefrontal cortex (DLPFC) brain region is known to play an important role in working memory and cognitive dysfunction, which are common characteristics of many psychiatric disorders (Huang et al., 2017) including SCZ, BD and MDD. Acknowledging these shared features, cross-disorder psychiatric studies enable the examination of patients based on their common biological processes, rather than solely based on phenotypic features, which constitutes a priority research approach in psychiatry (Cardno and Owen, 2014; Cross-Disorder Group of the Psychiatric Genomics Consortium, 2013, 2019; Docherty et al., 2016). While alterations in the living human brain can only be studied with neuroimaging methods, postmortem specimens provide the resolution to examine disease-related changes in the neural circuits and transcriptome-wide molecular processes.

Genome-wide association studies (GWAS) are an effective tool to investigate the genetic predisposition of complex psychiatric disorders. In the last years, large SCZ, BP and MDD GWAS have identified hundreds of genetic loci associated with these disorders from hundreds of thousands of samples at the level of single variants. For example, Fromer et al. (Consortium et al.) performed a GWAS on SCZ with 69,369 cases and 236,642 controls and identified 270 genetic risk loci that were strongly enriched in gene expression of the Brodmann area 9 (BA9) PFC region. Based on the single variant GWAS results, a new approach to calculate polygenic risk scores (PRSs) was established (Consortium and The International Schizophrenia Consortium, 2009). The PRS for an individual is determined on the basis of their cumulative genetic risk, as identified by GWAS, to be affected by a disorder. In psychiatric research, PRSs still need improvement before they may be used as predictive diagnosis tools; for example, the PRS for SCZ explained only 7.7% of the variance in genetic liability (Consortium et al.). With improved PRS methods and larger sample sizes in GWAS, PRSs can be refined to provide insight into the genetic architecture of psychiatric disorders, and computer-aided diagnosis of psychiatric disorders could become a trend in the future (Jonas et al., 2019; Zheutlin et al., 2019).

In parallel to the GWAS approach studying common single nucleotide polymorphisms (SNPs), examination of rare or de novo, highly penetrant variants may also clarify the genetic factors that contribute to psychiatric disorders. For example, Gulsuner et al. (Gulsuner et al., 2013) showed that SCZ de novo mutations imply a disruption of neurogenesis and transcription. Wainschtein et al. (Wainschtein et al., 2019) hypothesised that the lower heritability estimate of common SNPs compared to population-based estimates is due to rare

medRxiv preprint doi: https://doi.org/10.1101/2022.08.09.22278128; this version posted August 11, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity.
It is made available under a CC-BY-NC 4.0 International license .

variants, which are often located in regions of low linkage disequilibrium (LD) and are more likely protein altering. This is also consistent with previous observations that rarer and evolutionarily younger SNPs exhibit higher SNP heritability in many complex traits including MDD, BP and SCZ (Gazal et al., 2017).

Complementary studies are aiming to pinpoint molecular changes in the brain on the level of gene expression. This can be achieved using postmortem brain samples, well established for large cohorts like Genotype-Tissue Expression (GTEx) (Aguet et al.), CommonMind Consortium (CMC) (Fromer et al., 2016; Hoffman et al., 2019) and PsychENCODE (Wang et al., 2018). CMC is a partnership of experts from various research fields to collect and analyse large-scale genomic data from brain tissue of individuals with and without neuropsychiatric disorders. The PsychENCODE Consortium focuses on the investigation of non-coding genomic elements and their implications in psychiatric disorders. More recent developments in single-cell/nucleus RNA-sequencing (sc/snRNA-seq) technologies have enabled the study of cell-type-specific gene expression (Grindberg et al., 2013; Lake et al., 2017). The first single-cell atlas for SCZ was recently published and identified differentially expressed genes mainly in inhibitory and excitatory neurons (Ruzicka et al.).

The majority of GWAS significant variants are located in non-coding regions of the genome, thus not directly affecting the protein sequence (Edwards et al., 2013; Maurano et al., 2012). However, some of these variants might function by influencing gene expression, also called expression quantitative trait loci (eQTL) (Consortium and GTEx Consortium, 2017). Previous studies have shown that SNPs with evidence of association with SCZ are more likely to be eQTLs (Jaffe et al., 2018). The majority of these studies focus on eQTL analysis at the gene-level (Bhalala et al., 2018; Kim et al., 2014; Niu et al., 2019), with some recently looking at the sub-gene-level, like splicingQTL (sQTL): SNPs associated with differential splicing (Aguet et al.; Walker et al., 2019). Alternative splicing (AS) affects genes with multiple exons and provides a mechanism by which genes can produce diverse ranges of gene products. AS effects have been previously reported for SCZ (Chung et al., 2016; Clinton et al., 2003; Glatt et al., 2009; Martin et al., 2009; Takata et al., 2017). The first comprehensive study of brain sQTLs found evidence for a strong enrichment among SCZ GWAS loci (Takata et al., 2017). Since most of the known disease-causing mutations occur in exons, genetic methods that focus on protein-coding regions in the genome are efficient in identifying potentially pathogenic mutations (Xue et al., 2019). Therefore, the examination of exons may help us to understand the genetic architecture of SCZ beyond the gene-level (Guan et al., 2014; Ramasamy et al., 2014).

Careful tracing of genetic effects up to affected cell types in postmortem dorsolateral prefrontal cortex (DLPFC) is unmet and represents a key gap in knowledge. In our study, we performed a multi-factorial investigation that dissects the effects of disease-status, common variants and polygenic risk scores on gene expression in the DLPFC of cross-diagnostic psychiatric subjects. We demonstrate that examination of gene expression at the exon-level is essential to fully capture molecular changes. In parallel, we complement the

medRxiv preprint doi: https://doi.org/10.1101/2022.08.09.22278128; this version posted August 11, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity.
It is made available under a CC-BY-NC 4.0 International license .

genotype-/risk-modulated exon-level differential expression with rare and damaging variants. By examining both common genetic effects on gene expression and rare variant damaged genes, we found that pathways involved in molecular signalling and particularly membrane transportation, as well as inhibitory and oligodendrocyte precursor cells, are enriched in severe psychiatric disorders. Our results highlight how studying sub-gene-level genetic architecture can elucidate new knowledge regarding the development of severe psychiatric disorders and assist in classifying these disorders based on biology rather than symptomatology (Stephan et al., 2016). It also highlights how novel cellular targets for improved drug development can be identified.

medRxiv preprint doi: https://doi.org/10.1101/2022.08.09.22278128; this version posted August 11, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity.
It is made available under a CC-BY-NC 4.0 International license .

RESULTS

Cortical gene expression at the resolution of exons associates with diagnosis of psychiatric disorders

Gene expression analyses are typically performed at the level of gene models, which may level out effects of higher granularity ultimately resulting in AS. Especially in highly complex diseases, subtle but globally molecular changes may occur at this lower resolution, such as differentially expressed exons. To address this potential drawback of gene model analysis in the context of highly complex psychiatric cross-disorder, we profiled expression of postmortem DLPFC tissue of BA9 from 169 individuals aged 18-87 years (68 SCZ, 24 BD and 15 MDD and 62 controls, Table 1) using Affymetrix exon microarrays. We designed an exon array data pre-processing, analysis and testing strategy that allowed us to compare differential expression results across gene, transcript and exon levels (Figure S1). 6.5 million array probes were summarised into 242,443 exons, 100,750 transcripts and 17,447 gene models, while only one exon or transcript per gene model was subjected to multiple testing correction, so as not to confound the correction of genes with hundreds of exons (Figure S1 and methods). Using linear regression modelling for differential expression analysis, we included age, cause of death (CoD), postmortem interval (PMI) and genetic ancestry covariates (Dim1 - Dim4, Table 1 and Figure S2) as covariates to account for effects of clinical and demographic variables. The BA9 expression data was complemented by matching genotype data for all individuals, and following quality control and imputation, 9,572,139 common variants were yielded.

To increase statistical power and cut across traditional nosological boundaries, samples with SCZ, MDD and BP were grouped to one cross-disorder diagnosis, with SCZ being the dominant diagnosis. We found that zero, six and 2,223 genes (FDR < 10%) respectively at the level of gene models, transcripts and exons, respectively, were associated with cross-disorder cases in BA9 of human DLPFC (Figure 1A-D and Figure S3A-C). Tissue enrichment analysis with FUMA (Watanabe et al., 2017) confirmed that genes of the differentially expressed exons were particularly enriched for GTEx v8 brain regions (Figure S3D). KEGG pathway analysis of exon-level hits identified six pathways involved in cell-cell interactions, cell motility and various organismal systems (Figure S3E). Extracellular matrix (ECM)-receptor interaction (FDR = $7x10^{-5}$) and complement and coagulation cascade (FDR = 0.027) pathways contained 69% and 91% up-regulated genes, respectively. Conversely, 65% of the genes in the axon guidance (FDR = 0.046) pathway were down-regulated (Figure S3E).

As anticipated for the cohort size, gene-level analysis did not identify any differentially expressed genes. The six genes differentially expressed at the transcript-level were also differentially expressed at the exon-level. To better understand the gene-, transcript-, and exon-level difference, we analysed the fold changes (FCs) of all exon hits at the respective transcript- and gene-level and observed a 2.1- and 4.9-fold decrease in the median of the absolute log_2FCs, respectively (Figure 1E). Next, we selected the fibronectin type III domain

medRxiv preprint doi: https://doi.org/10.1101/2022.08.09.22278128; this version posted August 11, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity.
It is made available under a CC-BY-NC 4.0 International license .

containing 3A (FNDC3A), an ECM- glycoprotein that plays vital roles during tissue repair (Bierbaum et al., 2017), as a multi-exon gene regulated only at the exon-level. Only one of 26 exons was significantly associated with diagnosis (ENSE00003488149, FDR = 0.0539), while most exons showed no expression differences (Figure 1F). This suggests that the median summarization of many probes per gene compensates for the pronounced regulation of the few sub-gene alterations.

In addition, we explored the level difference for other covariates to understand if sub-gene resolution implies an information gain for other factors (Figure S3A,G and Table S2 to Table S6). Genetic ancestry, captured by one ancestry dimension (Dim2) of the genotype data, showed prominent differences at the gene-level. Age-related effects exerted similar impacts at each level (Figure S3F). Notably, age and pH strongly influenced gene expression. These findings together suggest that exon-level differences are related to diagnosis, rather than a global effect of brain expression.

Common variants associated with changes in cortical gene expression are enriched for psychiatric cross-disorder GWAS traits

In parallel to the association of cross-disorder diagnosis labeling alone with DLPFC gene expression, we set out to determine the common variants that influence DLPFC expression independently of diagnosis. To that end, we computed *cis*-eQTL on all three levels with MatrixEQTL(Shabalin, 2012), yielding 44,040 full gene-eQTLs, 203,069 transcript-eQTLs and 477,352 exon-eQTLs at FDR < 5% (Figure 2A, Figure S4 and Table S8 to Table S10). Again, the number of uniquely identified SNPs increased from gene-, transcript- to exon-eQTL analysis. In contrast, 6% genes (n = 72 out of 1,114) of the gene-eQTL were unique to this level, while 65% genes (n = 4,133 out of 6,401) were exclusively detected at the exon-eQTL level (Figure 2C). To illustrate the exon-level-specific effect, we used the calcium-binding gene neurocalcin delta (NCALD) with SNP rs505460, where only one exon out of 13 exons (ENSE00001231633, FDR = $4x10^{-45}$) was differentially expressed in a variant-dependent manner (Figure 2D). Notably, 49% of the differentially expressed exon-level genes (n = 1,090 out of 2,223 genes) were in common with exon-eQTL genes (eGenes).

In order to evaluate the robustness of our eQTLs, we compared the eGenes, to previously reported PFC (BA9) eQTLs from GTEx v8 (425 donors) and the CMC (467 donors with SCZ and BP; see Figure 2E). Overall, overlap of the general cohort with GTEx data was expectedly lower than to the disease-matched CMC cohort. The largest overlap > 75% was yielded for our exon-eQTLs and CMC eQTLs/sQTLs (Figure 2E), indicating reliable identification of prefrontal cortex eQTLs in our study when compared to CMC data.

To better understand the properties of our *cis*-eQTLs, we characterized the eQTL (e)SNPs according to genomic locations and regulatory features. Most eSNPs across all three levels were localised within intronic

medRxiv preprint doi: https://doi.org/10.1101/2022.08.09.22278128; this version posted August 11, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity.
It is made available under a CC-BY-NC 4.0 International license .

regions (Figure 2F) and were significantly enriched for synonymous variants, splice regions, promoters, missense, 5' and 3' UTR variants. Interestingly, eSNPs of exon-eQTLs were specifically enriched for stop gains. Integrative analysis of chromatin states of the DLPFC using ChromHMM (Ernst and Kellis, 2012) mainly showed that eSNPs of all three levels are associated to both active and repressive chromatin marks (see Figure 2G). The eSNPs of gene-eQTLs showed a specific enrichment to weak repressed polycomb elements.

Next, we explored how the identified eSNPs map onto genetic risk for psychiatric disorders. To do this, we tested whether eSNPs were overrepresented amongst variants associated with psychiatric disorders from large-scale GWAS of the Psychiatric Genomics Consortium (PGC) (Sullivan et al., 2018) and non-psychiatric phenotypes as negative controls. We detected significant enrichment of eSNPs on all three levels at a nominal GWAS p-value cut-off associated with SCZ, educational attainment and cross-disorder analysis, Figure 2H. For transcript- and exon-level eSNPs, we found significant enrichment in MDD-2018, BD and attention deficit hyperactivity disorder. Only exon-level eSNPs were enriched for MDD-2019. In summary, the gene-, transcript- and exon-eQTL hits may be significantly attributed to the genetic architecture of psychiatric diseases, with exon-level data specifically capturing the largest overlap.

Multi-SNP effects are associated with cross-diagnostic disease risk influence cortical exon expression

We next followed up on the exciting link between exons differentially regulated in BA9 of psychiatric subjects, and their encoding on genomic loci that were previously identified in large psychiatric GWAS. We specifically sought to determine how to effectively encode genetic disease risk in a variable, to associate it again with exon expression in the prefrontal cortex. To that end, we implemented PRS analysis, which allows us to compress the full genetic risk load to a cumulative sum of risk variants per individual. We used PRSice2 (Choi and O'Reilly, 2019) with a uniform p-value threshold of 0.01 to calculate risk for the major GWAS that were significantly enriched before (Figure 2H), namely SCZ, MDD, BD, ADHD and cross-disorder GWAS from PGC for all 169 samples. Notably, the PRSs did not correlate with the available covariates (Figure S5). Further, we focused on exon-level expression as it provides the strongest signal in our cohort. Using the identical analysis approach as for differential gene expression analysis, we computed the exon expression quantitative trait (eQT)-Score using the individual PRS scores of each GWAS together with exon expression (Figure 3A).

The exon eQT-Score analysis identified multiple genes for SCZ, MDD and CDG (Figure 3A), while the absolute number of associated genes (3,454-8,359, corresponding to a total of 9667 unique genes) is larger than for the cross-disorder diagnosis (2,223, c.f. Figure 1B). This suggests that the full load of the genetic disease architecture provides more information than the binary diagnosis. As negative control GWAS we used type 2 diabetes (T2D-2017), which yielded zero hits.

medRxiv preprint doi: https://doi.org/10.1101/2022.08.09.22278128; this version posted August 11, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity.
It is made available under a CC-BY-NC 4.0 International license .

Most exon eQT-Score inherent genes are in common between MDD-2019 and CDG (n = 3,349 with 28.1% shared effect direction), followed by MDD-2018 and MDD-2019 (n = 2,452, with 77.7% shared effect direction), SCZ and CDG (n = 1,745 with 27.3% shared effect direction), and SCZ and MDD-2019 (n = 2,277 with 58.6% shared effect direction; Figure 3C. The union of exon eQT-Scores of the four single GWAS (ADHD, BD, MDD-2019 and SCZ; n = 9667 genes and 11,697 exons, Table S14) hereafter is referred to as joint exon eQT-Score set.

The joint exon eQT-Score set covered 70.8% of the exon-level genes associated with cross-disorder diagnosis (Figure 3D), thus a large fraction of regulated genes is attributable to genetic disease risk. Asking whether the multi-SNP approach of the exon eQT-Score set provides more information, we compared the set to single SNP eQTL results. In total, 69.3% of the single-SNP exon-eQTLs overlap with the multi-SNP joint exon eQT-Score set. The effect sizes of the multi-SNP approach were overall larger than for single-SNP (Figure 3B).

We further investigated if the identified joint exon eQT-Scores were also differentially expressed for diagnosis (Table S1iii, 16.3% and n = 1,574 DE genes) and show single SNP-effects in *cis* (Table S10, 45.9% and n = 4,434 exon eQTL genes), see Figure 3D.

Rare and common variants share risk for cross-diagnostic psychiatric disorders

Since common variants only account for 23% of heritability in SCZ (Shi et al., 2016), we also examined rare variants (ExAC MAF < 0.1%) using whole-exome sequencing (WES) in an independent cohort of 142 patients diagnosed with different subtypes of SCZ (data set 2, see Methods). After filtering the rare variants for protein damaging effects (see Methods), 2,870 rare mutations, residing in 2,378 genes with 7 to 39 variants per patient, were detected. Twenty-five percent of the genes impacted by rare variants (n = 595 genes) were found exclusively in this dataset. Most genes (n = 1,984) affected by rare variants were found only once in the dataset, with only a small amount found two or more times (see Figure 4A). Titin (TTN), a gene described to be frequently mutated also in large WES cancer cohorts (Oh et al., 2020), is the only gene that was disrupted by rare variants in eight different patients.

To explore how genes might be affected by rare and common variants, we compared common single *cis*-eSNP effects, common multi-SNP effects and rare harmful effects and identified a core set of 832 genes, which were present in all three data types (see Figure 4B). We performed pathway enrichment analysis of these 832 core set genes to identify if they share the same function. This analysis indicated that one-fifth of the genes (n = 159, 19.1%) were significantly enriched in eleven KEGG pathways such as ECM-receptor interaction (FDR = 8^{-8}), focal adhesion (FDR = 2^{-4}) and ABC transporters (FDR = 9^{-3}), indicating a similar functional impact of rare and common variants in psychiatric disorders (see Figure 4C).

medRxiv preprint doi: https://doi.org/10.1101/2022.08.09.22278128; this version posted August 11, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity.
It is made available under a CC-BY-NC 4.0 International license .

To characterise the cell-type specificity of these core genes, we used snRNA-seq data from the human orbitofrontal cortex (Data set 3, see Methods). Among the 18 delineated cell-type clusters, the core genes were substantially enriched in oligodendrocyte precursors (OR = 5.65, bonferroni p = $4x10^{-5}$) and subtypes of inhibitory neurons (ORs > 4.22, bonferroni p < $1.1x10^{-3}$; Figure 4D). We next aimed to gain deeper mechanistic insight into the subset of core genes found in one of the five significantly enriched cell types (n = 29 genes). An overview of all annotated categories of these genes is given in Figure 5A. The majority (90-97%) of these genes overlapped with public cortex gene eQTL genes and isoform eQTL genes (CMC data set), 16 of the genes overlapped with transcript eQTL genes and only two with gene eQTL genes and at least ten overlapped with the differentially expressed exon level genes. 93% of the genes were an MDD exon eQT-Score gene hit, followed by 72% CDG and 48% SCZ exon eQT-Score genes.

To determine whether genes were closely connected, we used BioGRID (Oughtred et al., 2021) interactions with standard scores and selected only genes that were linked to at least one of the 29 genes. Fifty-two percent of the genes were represented in a network with 86 nodes and 182 undirected edges (Figure 5B). Eleven genes formed a strongly interconnected cluster.

One gene of this cluster was Glutamate Decarboxylase 1 (*GAD1*). *GAD1* showed enrichment in inhibitory neurons, GWAS SNPs for SCZ, BD, MDD-2019 and ADHD, eight ChromHMM states measured in DLPFC and four genomic features (VEP), cortex isoform eGenes (CMC) and is down-regulated in CDG (t-stats = -2.29, FDR = 0.1) and up-regulated in SCZ exon eQT-Score genes (t-stats = 2.39, FDR = 0.1) (Figure 5A-C) with controls showing a slightly lower SCZ PRS than cases (Figure 5D). Alterations in the GABA system are widely reported in schizophrenia (Jonge et al., 2017), and accordingly we found significantly down-regulation GAD1 exon expression (ENSE00003699191.1) for SNP rs1420385 (Figure 5E). The Ubiquitin Specific Peptidase 24 (*USP24*) and Protocadherin Related 15 (*PCDH15*) genes were also examined and the results can be found in the supplements (Figure S6).

medRxiv preprint doi: https://doi.org/10.1101/2022.08.09.22278128; this version posted August 11, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity.
It is made available under a CC-BY-NC 4.0 International license .

DISCUSSION

More than 95% of genes with multiple exons are alternatively spliced (Pan et al., 2008). Alternative splicing in the brain is highly complex compared to other tissues (Yeo et al., 2004) and suggested to play an important role in neuronal development and function (Raj and Blencowe, 2015). Splicing defects have been suggested to be involved in the pathogenesis of psychiatric disorders, including schizophrenia (Melé et al., 2015; Su et al., 2018; Zhang et al., 2022). In this study, we characterised differential exon expression, which is an important but relatively underexplored mechanism in psychiatry research, directly in the human cortex using two postmortem brain data sets. Even though our study was limited by only being able to study exon expression derived from exon arrays, we were still able to show large and significant findings. Most studies summarise expression by averaging across transcripts or genes and exclude expression changes in specific exons. Our results provide the first profiling of exon-specific expressions in the human brain. We found no significant main effects of psychiatric cross-disorder diagnosis at the gene-level and only a few at the transcript-level (n = 6 genes). However, the exons of 2,223 genes were differentially expressed between cases and controls (see Figure 1B). This highlights how the subtle effects of exons are easily overlooked, but hugely important to uncover disease pathogenesis. Our findings are in line with previous work that has shown an exon-based strategy improves the detection of small but effective expression changes that might be missed by gene-based approaches (Gandal et al., 2018; Laiho and Elo, 2014). Most of the transcript- and exon-level differentially expressed genes we identified are supported by other studies. For example, our top hit at the transcript-level, *Cyclic Nucleotide Gated Channel Subunit Beta* 1 (*CNGB1*, exon-level p-value: 9.2×10^{-6}, transcript-level p-value: 6×10^{-6}) was found to be differentially expressed in exome sequencing studies of various psychiatric disorders including schizophrenia and bipolar disorder (Ganesh et al., 2019) . Our top hit at exon-level, *Cystatin B* (*CSTB*, exon-level p-value: 8.8×10^{-6}, transcript-level p-value: 1×10^{-5}) was also differentially expressed in a schizophrenia case-control study (Gardiner et al., 2013). *Dipeptidyl peptidase 4* (*DPP4,* exon-level p-value: 0.0017, transcript-level p-value: 2.9×10^{-5}) found at both levels is a hit in the GWAS of SCZ from 2014 (Schizophrenia Working Group of the Psychiatric Genomics Consortium, 2014) and 19 of the 130 loci (14.62%) found in the recent GWAS of SCZ (Consortium et al.) intersect with our exon-level differentially expressed genes.

By integrating the transcriptome and genetic variation data, we found that the fraction of significant eQTLs at the exon-level was two times higher than at the transcript-level and even ten times higher than at the gene-level (Figure 2A). An exon-level specific eQTL is *Neurocalcin Delta* (*NCALD)*, visualised in Figure 2D. The effect of the differentially expressed exon ENSE00001231633 was not detectable at the transcript and gene-level, likely because of the compensating effect of the other non-differentially expressed exons at this locus. *NCALD* is a brain-enriched highly conserved neuronal calcium sensor protein (Di Sole et al., 2012) and has been associated with multiple neurological disorders. In a genetic rat model of

medRxiv preprint doi: https://doi.org/10.1101/2022.08.09.22278128; this version posted August 11, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity.
It is made available under a CC-BY-NC 4.0 International license .

schizophrenia, *NCALD* was downregulated (Vercauteren et al., 2007), and SNPs in *NCALD* have been associated with autism and BP (Ben-David et al., 2011; Xu et al., 2014). In a recent study *NCALD* has been linked to adult neurogenesis (Upadhyay et al., 2019). Interestingly, 75-77% of our exon-level eQTL genes overlap with previously published postmortem eQTL genes and 65-75% with isoQTL genes, demonstrating the validity of our results as well as the added value of examining exon-specific expression. The analysis of variant effects coming from exon-eQTLs leads to a high accumulation of gene flanking regions as well as synonymous and missense variants. Only exon eSNPs are enriched in stop-gained variants, which could imply a higher protein-damaging impact than seen on the transcript- or gene-level. Furthermore, since common variants have a very small effect size and rare variants are sought to contribute more to genetic risk of psychiatric disorders, but are hardly detected in GWAS (Gibson, 2012; Manolio et al., 2009), our results model the joint effects of common and rare variants. Our investigation of rare variants from whole-exome sequencing (WES) of schizophrenia index patients detected mutations that affect the same genes as those found for exon-eQTL (44.5% overlap), suggesting a common basis of loci for SCZ. The SNPs of our exon-level eQTLs were enriched for GWAS SNPs of SCZ, MDD, BD, ADHD and CDG, confirming that SCZ is a complex and polygenic disorder that shares many common genetic and phenotypic features with other psychiatric disorders. Despite the complexity and shared genetic basis of psychiatric diseases, we show that our results are stable and can be supported by other studies.

Symptoms and disease progression of patients diagnosed with psychiatric disorders are highly heterogeneous within the same disorder. In parallel, different psychiatric diseases are highly comorbid, supporting the model of a transdiagnostic phenotype. In addition, psychiatric diseases are highly polygenic and a recent study showed that GWAS and eQTL studies cover systematically different variants (Mostafavi et al., 2022). With our study, we highlight an alternative way to integrate expression with genetic variation by using eQT-Scores on exon-level expression. This approach provided more information than binary case-control diagnosis, with 1.5-3.7 fold more genes differentially expressed (Figure 3A) and 22.3-47.5 times higher effect sizes (Figure 3B) than single exon-level eQTL genes. Our findings contradict a study by Curtis et al (Curtis), who could not find any associations between PRS and gene-level gene expression using the post-mortem dorsolateral prefrontal cortex data (BA9/46) from the CommonMind Consortium. This is likely because gene-level approaches underestimate the gene expression level changes, and we gain information by estimating gene expression at certain exons. Together, our study shows that a more sensitive approach examining exon level expression combined with polygenic continuous phenotypes (in the form of a PRS) improves the results towards risk prediction. In the future, larger sample sizes and advanced polygenic risk score calculations that include functional annotation information (including cell type or epigenetic marks) could lead to even better exon eQT-Score results, and provide more precision for identifying novel diagnostic tools.

medRxiv preprint doi: https://doi.org/10.1101/2022.08.09.22278128; this version posted August 11, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity.
It is made available under a CC-BY-NC 4.0 International license .

By integrating all possible genetic modelling of the SCZ phenotype (eQTL, eQT-Score and rare variants), we found a core set of 832 genes, which combined rare and common risk for psychopathology. This core set revealed a network of genes involved in extracellular matrix (ECM)-receptor interaction, focal adhesion and ABC transporters. These pathways have already been reported in several SCZ postmortem brain studies. For example, Beretta et al. (Berretta, 2012) showed that the brain ECM plays an important role in SCZ. Abnormalities affecting several ECM components have been described in subjects with SCZ and Alzheimer's disease, leading to further GABA-related abnormalities and focal adhesion (Sethi and Zaia, 2017). Besides the genetic heterogeneity of psychiatric disorders, the cellular heterogeneity of the brain plays a role in disease pathogenesis. We show that the core gene set is enriched in oligodendrocyte precursors (OPCs) and inhibitory neuron subtypes (parvalbumin [PVALB], somatostatin [SST] and vasoactive intestinal peptide-expressing [VIP] interneurons). These findings are also consistent with results from genetic studies that suggest that inhibitory neuron pathology has an impact on the normal development of cortical circuits (Uhlhaas and Singer, 2012) and that dysfunction of OPCs might contribute to a disturbed connectivity between brain regions seen in various psychiatric disorders (Edgar and Sibille, 2012; Raabe et al., 2018). Also, PVALB, SST and VIP interneurons were identified as part of the transcriptomic taxonomy of the cortical GABAergic system, highlighting that genetics-driven molecular changes within these cell types can impact cortical inhibitory signalling and entire brain circuitry (Huang and Paul, 2019). Future studies focusing on single nucleic omics will continue to uncover salient cell types involved in psychiatric disease, such as Ruzicka et al. (Ruzicka et al., 2021) who also found neuronal cell types to be the most affected in SCZ postmortem brains.

To demonstrate how to leverage the multi-level information from genetics (common/rare), and tissue transcriptomics (bulk exon array expression and snRNA-seq), we chose to zoom into a small set of our core genes (n = 29 genes). The justification for this focus was based on the collected evidence from all our functional annotations, including the enrichment in single cell types to explore how genetics impacts expression of specific cells in the brain, and how this is related to disease. We showed that a large proportion (11 genes, 38%) interact closely with each other, indicating enrolment in the same pathways. Several publications have already demonstrated that different data sets and findings in psychiatric disorders often complement each other and also implicate the same pathways (Ganapathiraju et al., 2016; Schizophrenia Working Group of the Psychiatric Genomics Consortium, 2014). One promising candidate was the *Glutamic Acid Decarboxylase 1* (*GAD1*) gene, in which rare SCZ risk variants were also discovered by Magri et al. (Magri et al., 2018). *GAD1* encodes two proteins, one of them is *GAD67*, for which a reduction in expression seems to be one of the consistent findings in post-mortem studies of schizophrenia (Jonge et al., 2017). *GAD67* is involved in the production of gamma-aminobutyric acid (GABA) from glutamic acid and is an important inhibitory neurotransmitter in the brain (Smith, 2018). This also explains why we found *GAD1* to

medRxiv preprint doi: https://doi.org/10.1101/2022.08.09.22278128; this version posted August 11, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity.
It is made available under a CC-BY-NC 4.0 International license .

be particularly enriched in GABA-ergic and metabolic pathways. Further analysis of the core gene set could lead to new insights into schizophrenia and underpin the finding of a common basis in psychiatric disorders.

Taken together, the presented data reveals the importance of studying gene expression at the exon level directly in the human brain, and incorporating multi-level holistic datasets to identify important, combined risk gene sets with pathological relevance. This approach enabled us to identify inhibitory cortical neurons and oligodendrocyte progenitor interactions, presenting a clear cellular target for further study and treatment development for psychiatric disorders.

METHODS

Study samples, tissue collection and processing

Data set 1: Human post-mortem brain samples (exon array data)

The primary cohort (data set 1) has been previously described in Scarr et al. (Scarr et al., 2018) and Dean et al. (Dean and Scarr, 2016). Briefly, postmortem DLPFC tissue from 169 adult subjects aged 18-87 years with either SCZ (n = 68), MDD (n = 24), BD (n = 15) and matched controls (n = 62) were included in the study (Table 1). Demographic, clinical and pharmacological data were obtained during a case history review conducted using the Diagnostic Instrument for Brain Studies (DIBS), as described previously (Scarr et al., 2018). Tissue collection and processing was performed as described previously (Scarr et al., 2018). Postmortem brains were collected with approval from the Ethics Committee of the Victorian Institute of Forensic Medicine and all tissue was collected by this Institute after gaining written consent of the next of kin. This study was approved by the Human Ethics Committee of Melbourne Health (Scarr et al., 2018). All tissue was obtained from the Victorian Brain Bank at the Florey Institute for Neuroscience and Mental Health. Brodmann area 9 (BA 9) was taken from the lateral surface of the frontal lobe from an area comprising the middle frontal superior gyrus to the inferior frontal sulcus of the left hemisphere.

Data set 2: Human whole blood samples (SCZ index patients data)

Data set 2 consists of a total of 142 index patients (see Table 2) with different subtypes of SCZ. The diagnosis was made by an experienced clinician based on the medical record, information from the Structured Clinical Interview for DSM Disorders (SCID) and/or the Operational Criteria Checklist for Psychotic Illness (OPCRIT) of a patient. All patients met the DSM-IV (Diagnostic and Statistical Manual of Mental Disorders, 4th Edition) criteria for the diagnosis of SCZ and stated that they were of German origin. 26 patients were diagnosed with disorganized, 13 with catatonic, 85 with paranoid, 10 with residual and 8 with the undifferentiated subtype of SCZ according to the DSM-IV. The 142 index patients were selected for

medRxiv preprint doi: https://doi.org/10.1101/2022.08.09.22278128; this version posted August 11, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity.
It is made available under a CC-BY-NC 4.0 International license .

whole exome sequencing at the Central Institute of Mental Health (Mannheim, Germany) and the University Hospital Bonn (Bonn, Germany). The study was approved by the respective ethics committees of the corresponding centers (Votum 0223.4). Each patient gave a written declaration of consent and the respective ethics committees of the two institutions approved the study.

Data set 3: Human post-mortem brain samples (snRNA-seq data)

Postmortem orbitofrontal cortex tissues from Brodmann Area 11 (BA11) from two neurotypical individuals (see Table 3) were used for single nuclei RNA-sequencing (snRNA-seq). These brain tissues were fresh-frozen and obtained from the NSW Brain Tissue Resource Centre in Sydney, Australia. Informed consent was given by both donors or their next of kin for brain autopsy. Ethical approval for this study was obtained from the Human Research Ethics Committees at the University of Wollongong (HE2018/351) and the Ludwig-Maximilians-Universität Munich (17-085 and 18-393). BA 11 was dissected from the 3rd 8-10mm coronal slice from each fresh hemisphere for each subject.

Gene expression data

Data set 1: Exon arrays

RNA preparation and expression array processing were carried out as described previously (Scarr et al., 2018). Briefly, total RNA was isolated from ~100 mg frozen grey matter using 1.0 ml TRIzol reagent (Life Technologies, Scoresby, VIC, Australia). After homogenization and phase separation, the aqueous phase was added to an equal volume of 70% ethanol. RNA isolation was performed with RNeasy minikits (Qiagen, Cat. No. 74104, Chadstone Centre, VIC, Australia), with all samples treated with DNase using column digestion. DNA contamination was ruled out by PCR using specific primers for genomic DNA. RNA quantity and quality were analysed by spectrophotometry (NanoDrop; Thermo Fisher Scientific Australia, Scoresby, VIC, Australia) and by determining RNA Integrity Numbers (RINs) using an Agilent 2100 bioanalyzer (Agilent Technologies, Santa Clara, CA, USA) and all samples with RIN $\geq$ 6.00 were used for further analyses with Affymetrix Human Exon 1.0 ST v2 Arrays according to the manufacturer's instructions (Affymetrix, Santa Clara, CA, USA). Following hybridization, chips were scanned and the fluorescent signals converted into a DAT file for quality control. Finally, cell intensity (CEL) files were generated for further analyses.

Reading the raw data CEL files and storing the 6,553,600 microarray probes was performed using the *oligo* version 1.50.0 package (Carvalho and Irizarry, 2010) in R 3.6.1. For background adjustment, quantile normalisation and summarization (using median-polish) to probeset level *oligos* Robust Multichip Average (RMA) algorithm performing all three steps at once was used. Probesets without any start or stop information or labelled as controls in the current NetAffx annotation file for the Human Exon 1.0 ST v2 Array,

medRxiv preprint doi: https://doi.org/10.1101/2022.08.09.22278128; this version posted August 11, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity.
It is made available under a CC-BY-NC 4.0 International license .

downloaded from the Affymetrix support website, were removed from the dataset. Cross-hybridized probesets (*number_cross_hyb_probes* ≥ 1), composed of RNA target sequences binding to short DNA probes that are not exactly their complement, and probesets lying on the X, Y and M chromosome were also excluded from the dataset.

SVA version 3.34.0 (Johnson et al., 2007; Leek and Storey, 2007) package in R 3.6.1 was used for batch correction of known and hidden batches. Batch correction was conducted at probeset level before annotation and summarization to different genetic levels. We first removed the five known batches with *SVA's ComBat* function and then applied surrogate variable (SV) analysis to remove hidden effects. One SV was found, which was the only one explaining high variance in the expression dataset (Figure S2C-E). Gene annotations were downloaded from GENCODE (Frankish et al., 2019) release 19, which is the current version for the human hg19 (GRCh37) genome built on Ensembl version 74. Only protein-coding genes manually annotated by the Human and Vertebrate Analysis and Annotation (HAVANA) team were considered.

Summarization to gene, transcript, and exon-level was performed as follows: First, an annotation dataset was generated, where the GENCODE and the Affymetrix probeset information were merged by overlapping the locations in both datasets using the mergeByOverlaps function of the *GenomicAlignments* version 1.22.1 package in R 3.6.1. This annotation dataset was filtered for probesets, where at least one gene symbol correlated in both merged files and no multiple annotations were found. Second, for each sample the median of all expression values across all probes containing the same Ensembl ID was calculated. Multiple gene mappings were removed from all datasets. Some genes, transcripts and exons shared the exact same expression values in all samples leading to identical rows with different assigned IDs. This happens due to overlapping gene, transcript or exon locations. Instead of completely removing all identical rows, one row was kept and the different gene, transcript or exon IDs were merged into one ID. This led to expression values of 17,447 genes, 100,750 transcripts and 242,443 exons (all based on Ensembl IDs).

Data set 3: snRNA-seq

Nuclei extraction and library preparation:

Nuclei were isolated using an adapted version of a previously published protocol (Matevossian and Akbarian, 2008). Briefly, frozen, dissected brain samples (50 - 60 mg) were dounce-homogenised in 1 ml nuclei extraction buffer (0.32 M Sucrose, 3 mM $Mg(Ac)_2$, 5 mM $CaCl_2$, 0.1 mM EDTA, 10 mM TrisHCl pH 8.1, 0.1% IGEPAL CA-630, 40 U/ml RiboLock RNase-Inhibitor (ThermoScientific)) on ice. Homogenate was layered onto 1.8 ml of sucrose cushion (1.8 M Sucrose, 3 mM $Mg(Ac)_2$, 10 mM TrisHCl pH 8.1) and ultra-centrifuged at 28100 rpm for 2.5 hours at 4°C. Supernatant was removed and nuclei pellet was resuspended in 100 µl resuspension buffer (1X PBS, 3 mM $Mg(Ac)_2$, 5 mM $CaCl_2$, 1% BSA, 40 U/ml

medRxiv preprint doi: https://doi.org/10.1101/2022.08.09.22278128; this version posted August 11, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity.
It is made available under a CC-BY-NC 4.0 International license .

RiboLock RNase-Inhibitor). Nuclei suspension was filtered through a cell strainer cap. Nuclei were stained with DAPI 1:1000 and counted using a hemocytometer. Libraries for snRNA-seq were prepared following the user guide of 10X Genomics (Chromium Single Cell 3' Reagents kit v3) with a target recovery of 10,000 nuclei per sample. Libraries were pooled equimolarily and were treated with Illumina Free Adapter Blocking Reagent before sequencing on the NovaSeq 6000 System (Illumina, San Diego, California, USA).

Sequence Alignment, Filtering, Normalisation, Clustering and Cell type assignment

Sequence reads were demultiplexed using the sample index, aligned to a pre-mRNA reference and UMI were counted after demultiplexing of nuclei barcodes using Cell Ranger v3.1.0. Count matrices were further processed using Scanpy v1.4.4 (Wolf et al., 2018). Count matrices of the two individuals were combined. Nuclei were filtered according to counts, minimum genes expressed and % of mitochondrial genes (Max counts > 50 000, Min counts < 1000, Min genes > 400, Mito % ≥ 10). Genes expressed in < 20 nuclei were removed. Data were normalised and log-transformed using Scran (Lun et al., 2016). Embeddings were created using BBKNN (Polański et al., 2020) and Louvain clustering (Blondel et al., 2008) using highly variable genes was applied for clustering. One cluster was excluded from the analysis due to highly variable genes being driven by or containing several MT-genes. Cell types were assigned to clusters based on marker gene expression as follows (Nagy et al. (Nagy et al., 2020) and Velmeshev et al. (Velmeshev et al., 2019)): Excitatory neurons: *SATB2, SLC17A7*, Layers: L2-4: *CUX2, THSD7A, L4-6: RORB, POU6F2, TSHZ2, RXFP1, L5-6: ETV1, KCNK2, PCP4*, Inhibitory neurons: *GAD1, GAD2*, Inhibitory neuron subtypes: In_PVALB: *PVALB*, In_SST: *SST*, In_VIP: *VIP, CALB2*, In_SV2C: *SV2C*, fibrous astrocytes (Astro_FB): high *GFAP, TNC, AQP4, GJA1*, protoplasmic astrocytes (Astro_PP): high *SLC1A2, AQP4, GJA1,* Microglia: *CD74, P2RY12, C3, CX3CR1,* Oligodendrocyte precursors (OPCs): *PCDH15, PDGFRA, OLIG1*, Oligodendrocytes (Oligo): *PLP1, MBP, MOBP, MOG,* Endothelial cells (Endo): *CLDN5, FN1, FLT1*.

Genotype data, imputation and PRS in data set 1

Firstly, 100ng of genomic DNA was extracted from blocks of cerebellum from 169 subjects composed of 107 cases and 72 controls (see Table 1). Extractions were performed as previously described (Scarr et al., 2009; Strauss, 2001).

Genotyping of the samples was performed using Illumina Infinium Global Screening Arrays according to the manufacturer's standard protocols. Quality control (QC) was conducted in PLINK 1.90b6.6(Chang et al., 2015). QC steps on samples included removal of individuals with a missing rate > 2%, cryptic relatives (PI-HAT > 0.0125), an autosomal heterozygosity deviation ($|F_{het}|$ > 4 SD) and genetic outliers (distance in the ancestry components from the mean > 4 SD). QC steps on variants included removal of variants with a call rate < 98%, a MAF < 1%, and HWE test p-values ≤ 10^{-6}. Furthermore, variants on non-autosomal

medRxiv preprint doi: https://doi.org/10.1101/2022.08.09.22278128; this version posted August 11, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity.
It is made available under a CC-BY-NC 4.0 International license .

chromosomes were excluded. Imputation was performed with IMPUTE2, following phasing in SHAPEIT, using the 1,000 genomes phase III reference panel. QC of imputed probabilities was conducted in QCTOOL v1.5 64-bit. Imputed SNPs were excluded if MAF < 1%, HWE test p-values ≤ 10^{-6}, or an INFO metric < 0.6. SNP coordinates are given according to hg19 (n = 9,164,462 SNPS).

PRS for all 169 individuals were calculated using PRSice-2 v2.2.11.b (14th Oct 2019) (Choi and O'Reilly, 2019) using a P-value threshold of 0.01. As input, we used the hard-called imputed and quality controlled genotypes of the data set 1 as target, imputed genotypes from an independent cohort (recMDD (Muglia et al., 2010), n = 1,774 Caucasian individuals) as LD reference and 9 different publicly available GWAS summary statistics as base dataset. Polygenic risk was measured for different psychiatric disorders using following GWAS summary statistics of the Psychiatric Genomics Consortium (PGC): SCZ (SCZ-2014) (Schizophrenia Working Group of the Psychiatric Genomics Consortium, 2014), BD (BD-2019) (Stahl et al., 2019), MDD (MDD-2018 and MDD-2019) (Howard et al., 2019; Wray et al., 2018), autism spectrum disorder (ASD-2019) (Consortium and The Autism Spectrum Disorders Working Group of The Psychiatric Genomics Consortium, 2017), attention-deficit/ hyperactivity disorder (ADHD-2019) (Demontis et al., 2019) and cross disorder (CDG-2019) (Cross-Disorder Group of the Psychiatric Genomics Consortium, 2019). We additionally generated PRS for two non-psychiatric phenotypes as negative controls: the Social Science Genetic Association Consortium (SSGAC) for educational attainment (EA-2018) (Lee et al., 2018) and the DIAbetes Genetics Replication And Meta-analysis (DIAGRAM) Consortium for type 2 diabetes (T2D-2017) (Scott et al., 2017) (Table S13i). P-value, pseudo-R^2 based on Cox & Snell (Cox and Snell, 1989) and Cragg & Uhler (Nagelkerke) (Nagelkerke, 1991) approach were calculated using a fitted linear model approach and the Nagelkerke function of the *rcompanion* version 2.3.25 package in R (Table S13ii).

Whole exome sequencing in data set 2

DNA was obtained from whole blood by salting-out with saturated sodium chloride solution (Miller et al., 1988). The SureSelectXT Human All Exon V5 capture library from Agilent Technologies was used for target enrichment. Exome sequencing of the 142 SCZ patients was performed on an Illumina HiSeq2500 v4 system using 2 x 125 bp paired-end sequencing. The data was annotated using GRCh37 (hg19) reference genome build of 2009 from NCBI.

For data analyses the VARBANK v2.10-v2.14 graphical user interface of the Cologne Center for Genomics (Cologne, Germany) was used. Filter criteria were defined in such a way that heterozygous and homozygous variants (allele reading frequency above 25%) were detected. In addition, the sequencing reads were filtered for a coverage of ≥ 10 and a base quality ≥ 20. The focus of the dataset lies on single-nucleotide variants (SNV), insertions (INS), deletions (DEL) and their combination (InDel), which either 1) have a functional

medRxiv preprint doi: https://doi.org/10.1101/2022.08.09.22278128; this version posted August 11, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity.
It is made available under a CC-BY-NC 4.0 International license .

effect on the protein structure, e.g. by alternating the protein structure, or 2) have a medium to strong splice site effect.

Identified variants were filtered for gene damaging mutations using a Combined Annotation Dependent Depletion (CADD) (Kircher et al., 2014) v1.3 score ≥ 20. 346 mutations are insertions, deletions or indels without any CADD score assignment and therefore kept in the dataset. Furthermore, the variants were filtered for a minor allele frequency (MAF) ≤ 0.1% using the data of the Exome Aggregation Consortium (ExAC) (Lek et al., 2016) excluding psychiatric cohorts resulting in ~45,000 samples. Approximately 60% of the ExAC dataset is of European descent, so the data was considered appropriate for MAF calculations of the index patient cohort. Filtering the dataset led to 2,870 rare mutations on 2,378 genes (Ensembl gene IDs).

Phenotype data in data set 1

For three samples, no pH value was given, so we assigned them the mean value of all pH values from all other samples. We calculated the first ten ancestry dimensions (Dim1-10) using multidimensional scaling (MDS) based on raw Hamming distances in PLINK. Samples did not group into different subpopulations, although some samples were known to be of Asian (n = 5) and mixed, Asian and Caucasian (n = 1), origin, while the majority (n = 123) were of Caucasian descent (Figure S2A). Only the Eigenvalue of the first four genetic ancestry dimensions (Dim1-4) changed verifiably, so only these were used in further analyses (Figure S2B). Based on Stenbacka et al. (Stenbacka and Jokinen, 2015), the cause of death was divided into three categories: 1) "natural" for all natural causes, e.g. ischemic heart disease or pneumonia, 2) "violent" for violent suicides or accidents, such as hanging, drowning or a car accident, and 3) "non-violent" for all types of poisoning. A visual overview of all phenotypes is given in Figure S2D-E.

Differential expression analysis (DEA)

For the differential expression analyses, we collapsed all cases (SCZ, BD, MDD) and *limma* version 3.42.2 package in R 3.6.1 was used. We calculated the effect of case-control status on the difference in gene expression controlling for Age, Sex, pH, PMI, RIN, RIN^2, Suicide and Type of Death (Natural, Violent and Non-Violent) as well as the first four dimensions of the genotype defined ancestry (Dim1-Dim4) and one found surrogate variable (SV1), which was not correlated to any other covariate of the model (Figure S2C,E). Expression matrices of 17,447 genes, 100,750 transcripts and 242,443 exons coming from 169 samples (107 cases and 62 controls) after preprocessing were used for the analysis. The differential expression analysis was performed for each genetic level separately and only significant (FDR < 0.1) hits were considered for further analysis. For transcript- and exon-level expression the FDR was calculated as follows: 1) p-values were generated using *limma*, correcting for all covariates, 2) for each Ensembl gene ID only the transcript or exon with the lowest p-value was taken generating a new expression matrix with only one transcript or exon per

medRxiv preprint doi: https://doi.org/10.1101/2022.08.09.22278128; this version posted August 11, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity.
It is made available under a CC-BY-NC 4.0 International license .

gene, and 3) *limma* was used to recalculate the FDR on the new expression matrix. We performed the differential expression analysis not only for diagnosis but all given covariates correcting for all other covariates of the model (Figure S2**,** Figure S3A and Table S1 - Table S6).

eQTL analysis and eSNP clumping

For the eQTL analysis, all imputed SNPs with a minor allele frequency (MAF) < 5% were removed (n = 6,830,577 SNPs). These SNPs were fitted to the expression of gene (n = 17,447), transcript (n = 100,750) and exon-level (n = 242,443) using the additive linear model of MatrixEQTL(Shabalin, 2012), adjusting for the same covariates (age, sex, pH, PMI, RIN, RIN2, suicide, ToD, Dim1-4, SV1) as in the DEA. The *cis*-eQTL output threshold was set to 0.05, the physical distance to 1Mb, no FDR memory was saved and no trans-eQTLs were calculated (threshold set to 0). For identity, the error covariance matrix was set to numeric and the minimal p-value by gene SNP was not calculated. Frequently, eQTL SNPs (eSNPs) are in LD and thus more often associated with each other than by chance. To find these linked eSNPs, we performed clumping using PLINK v1.90b6.7 64-bit (2nd Dec 2018) (Chang et al., 2015). For the clumping we set an r-squared LD threshold of 0.2, a significance threshold for index SNPs at 0.05, a secondary significance threshold for clumped SNPs at 1 and a physical distance threshold of 1Mb.

Exon eQT-Score generation

To generate expression quantitative trait scores (exon eQT-Score) we used the same procedure as for the DEA replacing diagnosis with PRS (n = 9), i.e., PRS was fitted to the expression using a linear model adjusting for age, sex, pH, PMI, RIN, RIN2, suicide, ToD, Dim1-4, SV1. Each PRS was used once, resulting in exon eQT-Score for the different phenotypes of the underlying 9 GWAS used for the PRS calculation.

Enrichment analysis

GWAS and epigenetic enrichment analysis

We performed an enrichment analysis using public data from Ensembl Variant Effect Predictor (VEP) (McLaren et al., 2016), the core 15-state model (ChromHMM v1.10) of chromatin in DLPFC from the Roadmap Epigenomics Project (Ernst and Kellis, 2012) and the GWAS summary statistics for ADHD-2019, ASD-2019, BD-2019, CDG-2013, CDG-2019, EA-2018, MDD-2012, MDD-2018, MDD-2019, SCZ-2014 and T2D-2017. As a background dataset, we used all eQTL SNPs overlapping with the GWAS datasets. To reduce bias, we generated 11 MAF bins for each eSNP set (gene-eSNP, transcript-eSNP and exon-eSNP) and the background SNP set by using 0.05 steps from 0 to 1. We then performed an enrichment analysis for 10,000 permutations, where we calculated the overlap of randomly selected background SNPs in the size of an eSNP set and its MAF bin distributions with one of the public datasets (VEP, ChromHMM or the GWAS

medRxiv preprint doi: https://doi.org/10.1101/2022.08.09.22278128; this version posted August 11, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity.
It is made available under a CC-BY-NC 4.0 International license .

summary statistics). To generate an empirical p-value, we calculated how often one of the 10,000 overlaps was higher than the actual overlap of the eSNP set to the public dataset and divided it through the number of permutations. An odds ratio (OR) was calculated by the actual overlap divided by the mean of the 10,000 overlaps of the resampling (Table S12).

KEGG pathway and tissue enrichment analysis

For the KEGG pathway and tissue enrichment analysis of exon level differentially expressed genes, we used the GENE2FUNC function of the online platform Functional Mapping and Annotation of Genome-wide Association Studies (FUMA GWAS) (Watanabe et al., 2017) excluding disease and drug treatment relations. The Ensembl version was set to v85 to be as near as possible to v74, which was used for the annotation of the expression datasets. Maximal adjusted p-value was set to 1 and all other parameters were kept in default. The background was set to the exon based gene set (n = 17,496). Tissue specificity was performed on differentially expressed genes defined for each label of each GTEx v8 (Aguet et al.) postmortem expression data set using a hypergeometric test, see Watanabe et al. (Watanabe et al., 2017). Significant tissue enrichment was obtained with Bonferroni corrected p-values ($P_{bon} \leq 0.05$). For the KEGG pathway enrichment FUMA uses hypergeometric tests to evaluate if genes of interest are overrepresented in any of the given pathways. Here, only significant pathways with an FDR ≤ 5% were considered.

Cell-type enrichment analysis

We used single-nuclei gene expression data (Data set 3) to identify cell types enriched among the cells with the highest mean expression of the identified core genes (n = 832). Core genes and the background of all tested genes (n = 17,496) were mapped to the single-nuclei expression data using Ensembl gene IDs (Ruffier et al., 2017). 813 of the core genes and 14,902 of the background genes were also detected in the single-nuclei data set. The cell-type distribution of the 25% cells with the highest mean expression value in the core genes were compared to the same distribution taking all background genes into account. Statistical significance of the enrichment in each cell-type was evaluated using Fisher's exact test (Fisher, 1992).

Core gene set generation and analysis

To generate the core gene set (n = 832), we overlapped the found exon-eQTL genes (n = 6,401) and joint exon eQT-Score genes (n = 9,667) from data set 1 with the rare variant damaged genes (n = 2,378) from data set 2 using Ensembl gene IDs. Filtering for the five most significant cell types (In_PVALB_1, In_PVALB_2, In_SST_1, In_VIP, OPCs), a subset of 29 genes from the core gene set was generated and visualised with a heat plot using *ggplot2* version 3.3.2 package in R 3.6.1. BioGRID (Oughtred et al., 2021) version 4.4.201 was used to determine the interactions between the subset genes, and the interactions were visualised using Cytoscape (Shannon et al., 2003) version 3.8.2. Individual genes were examined by cell type dimension

medRxiv preprint doi: https://doi.org/10.1101/2022.08.09.22278128; this version posted August 11, 2022. The copyright holder for this preprint **(which was not certified by peer review)** is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity.
It is made available under a CC-BY-NC 4.0 International license .

reduction plot using Uniform Manifold Approximation and Projection (UMAP) (McInnes et al., 2018) of *Scanpy* (Wolf et al., 2018) version 1.4.5 package in python 3.8.1, scatter plot and box plot using *ggplot2*.

medRxiv preprint doi: https://doi.org/10.1101/2022.08.09.22278128; this version posted August 11, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity.
It is made available under a CC-BY-NC 4.0 International license .

ACKNOWLEDGEMENTS

The Victorian Brain Bank tissue collection was supported by the National Health and Medical Research Council (NHMRC; Australia; grant number 566967) and the Cooperative Research Centre (CRC) for Mental Health and the Operational Infrastructure Support from the Victorian State Government. We would additionally like to thank Geoff Pavey for his efforts in curating the Australian Victorian Brain Bank. The New South Wales Brain Tissue Resource Centre at the University of Sydney was supported by the University of Sydney. Research reported in this publication was supported by the National Institute of Alcohol Abuse and Alcoholism of the National Institutes of Health under Award Number NIAAA012725-15. The content is solely the responsibility of the authors and does not represent the official views of the National Institutes of Health. Dr Knauer-Arloth was supported by the Brain Behaviour Research Foundation (NARSAD Young Investigator Grant, #28063). Dr Matosin was supported by an Al and Val Rosenstrauss Fellowship from the Rebecca L Cooper Medical Research Foundation and grants from the Brain Behaviour Research Foundation (NARSAD Young Investigator Grant, #26486) and the Rebecca L. Cooper Medical Research Foundation (#PG2020645). We thank Elisabeth B. Binder for the useful discussions during the project implementation and Vanessa Schmoll for her support with genotypes of data set 1. We thank André Heimbach, Markus Nöthen, Frederick Neukirch, Lisa E. Winkle for their support and generation of data set 2. We thank Miriam Gagliardi and Simone Röh for assistance with dataset 3. We thank Ghalia Rehawi for feedback on the enrichment analysis. We thank all participants for their contribution to this study.

CONFLICT OF INTEREST

All authors do not have conflicts to declare.

DATA AND CODE AVAILABILITY

Code for the analysis is available at GitHub: https://github.com/cellmapslab/PostmortemBrainAnalysis. Supplementary tables are available at Google Drive:
https://drive.google.com/drive/folders/1JRyThjCwb7JFE6LdQQiAHwXQZFIO2o0V? usp=sharing.
Expression data of data set 1 are available on GEO (GSE208338) and the genotypes are available upon approved request. snRNA-seq data of data set 3 are available on GEO (GSE205642). Exome-sequencing data of data set 2 can be made available upon reasonable request by the authors. The web links for the publicly available datasets used in the study are as follows: 15-core state model of chromatin from Roadmap Epigenomics Roadmap (ChromHMM v1.10) for functional annotation of eSNPs:
https://egg2.wustl.edu/roadmap/web_portal/chr_state_learning.html#core_15state, BioGRID, a database of protein, genetic and chemical interactions: https://thebiogrid.org/, Combined Annotation Dependent Depletion (CADD): https://cadd.gs.washington.edu/, CommonMind Consortium for result comparison: http://CommonMind.org, Current NetAffx probeset annotation file for the Affymetrix HuEx 1.0 ST v2

medRxiv preprint doi: https://doi.org/10.1101/2022.08.09.22278128; this version posted August 11, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity.
It is made available under a CC-BY-NC 4.0 International license .

microarray: http://www.affymetrix.com/Auth/analysis/downloads/na36/wtexon/HuEx-1_0-st-v2.na36.hg19.probeset.csv.zip, DIAbetes Genetics Replication And Meta-analysis (DIAGRAM) Consortium for the type 2 diabetes (T2D) GWAS summary statistic: https://diagram-consortium.org/downloads.html, GENCODE release 19 (GRCh37.p13) annotation file: http://ftp.ebi.ac.uk/pub/databases/gencode/Gencode_human/release_19/gencode.v19.annotation.gtf.gz, Genotype-Tissue Expression (GTEx) V8 for annotation/enrichment analysis and result comparison: https://www.gtexportal.org/home/datasets, Human Ageing Genomic Resources (HAGR) for known human ageing genes: https://genomics.senescence.info/genes/human_genes.zip, National Center for Biotechnology Information (NCBI) for GRCh37/hg19 reference genome of the index-patient rare variant dataset: http://www.ncbi.nlm.nih.gov/assembly/2758/, Psychiatric Genomics Consortium (PGC) for GWAS summary statistics of psychiatric disorders: https://www.med.unc.edu/pgc/, Social Science Genetic Association Consortium (SSGAC) for the educational attainment (EA) GWAS summary statistic: https://www.thessgac.org/data

AUTHOR CONTRIBUTION STATEMENTS

K.W. performed main computational analysis, data visualizations and wrote the first draft of the manuscript. N.M. provided intellectual input and performed critical revisions of the manuscript, led the collaboration for exon array resources (data set 1), and carried out the design and tissue acquisition for snRNA-seq analysis (data set 3). J.K.-A. and N.S.M. (Lead Contacts) jointly conceived, designed and supervised the study. K.W., J.K.-A. and N.S.M. equally contributed to the analysis strategy design, result interpretation and writing the manuscript. A.S.F. performed snRNA-seq library preparation and conducted snRNA-seq analysis with N.G. (data set 3). M.R. recruited SCZ index patients (data set 2). F.D., H.T. and A.K. generated, processed and analysed SCZ index patient data (data set 2). B.D. was involved in the collection and selection of appropriate post-mortem tissue samples (data set 1). N.S.M., J.K.-A., N.M., B.D., F.D. and F.J.T. acquired funding for the study. All other authors saw, had the opportunity to comment on, and approved the final manuscript.

medRxiv preprint doi: https://doi.org/10.1101/2022.08.09.22278128; this version posted August 11, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity.
It is made available under a CC-BY-NC 4.0 International license .

FIGURES

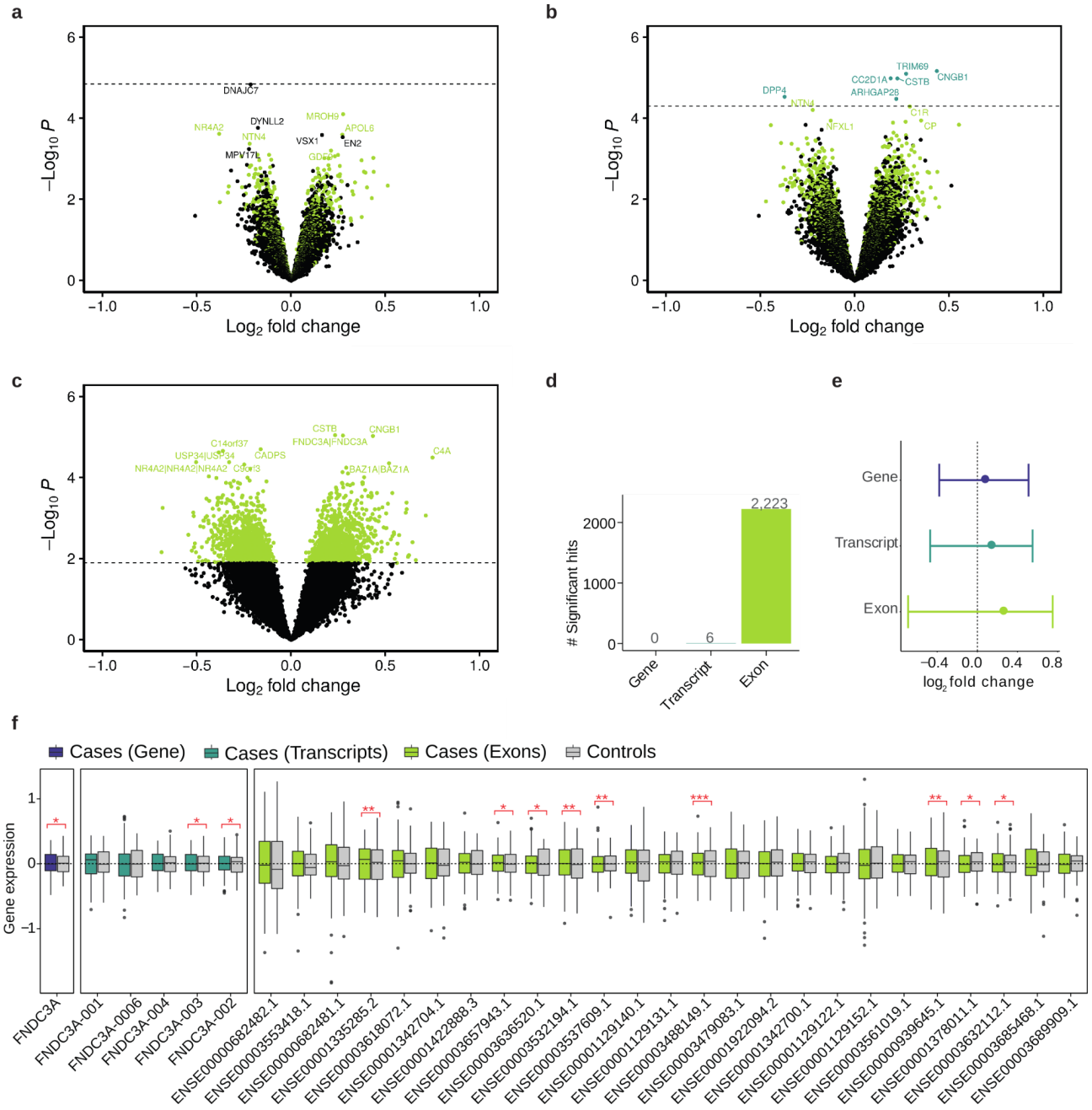
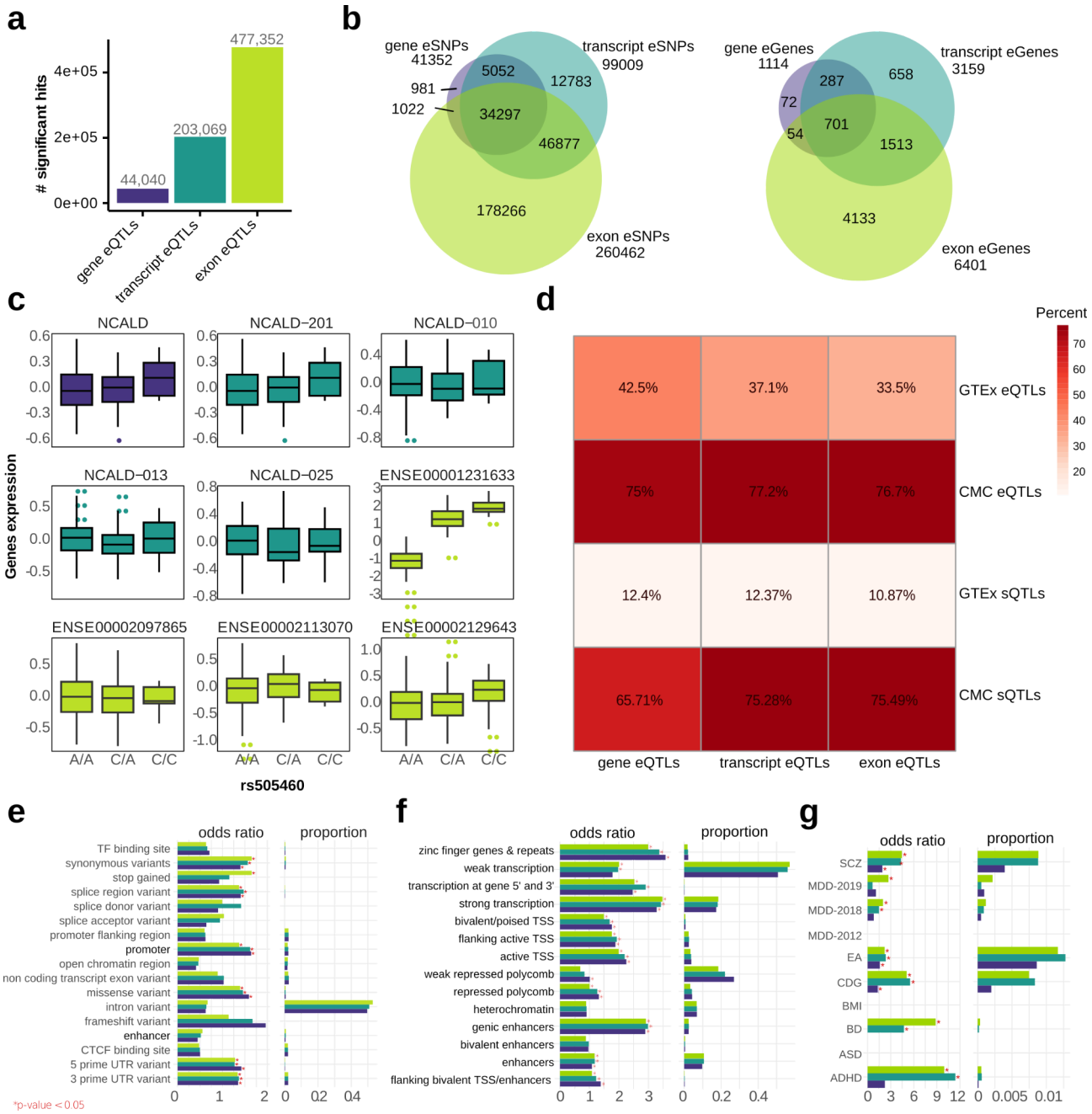


Figure 1: Differential expression results. Volcano plots of the (a) gene-, (b) transcript- and (c) exon-level differential expression analysis results. On the x-axis the $\log_2$ fold changes and on the y-axis the $-\log_{10}$(P-values) are given. The dashed horizontal line represents the significance threshold (FDR < 0.1) with all significant hits colored in turquoise for transcript-level and light green for exon-level differentially expressed gene hits. 51% of the exon-level (1,139 of 2,223 genes) and 83% transcript-level differentially expressed genes (5 of 6 genes) are upregulated. (d) Bar plot of significant differentially expressed genes for gene-, transcript- and exon-level. (e) Forest plot showing the $\log_2$FC range and median of the absolute $\log_2$FC (dot) of the 2,223 exon-level differentially expressed genes for all three levels. The 2,223 genes exhibit a larger magnitude of changes on the exon-level (median absolute $\log_2$FC = 0.23, range of -0.69 to 0.75)

medRxiv preprint doi: https://doi.org/10.1101/2022.08.09.22278128; this version posted August 11, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity.
It is made available under a CC-BY-NC 4.0 International license .

compared to the two other levels (transcript median absolute $\log_2$FC = 0.11, range of -0.47 to 0.55, and gene median absolute $\log_2$FC = 0.047, range of -0.38 to 0.51). (f) Boxplots show the effect of cross-disorder diagnosis on *FNDC3A* expression, separately for each level: full gene, five of its six transcripts and 25 of its 46 exons with available expression values for cases (purple, turquoise or light green) and control subjects (grey). The x axis indicates expression residuals. Transcripts and exons were sorted by the difference in median expression from cases to controls. Asterisks represent the significance level for the corresponding p-values.



medRxiv preprint doi: https://doi.org/10.1101/2022.08.09.22278128; this version posted August 11, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity.
It is made available under a CC-BY-NC 4.0 International license .

Figure 2: eQTL results and downstream their genomic features enrichment analysis. (a) Bar plot of significant eQTLs for gene, transcript and exon-level. (b) Venn diagrams visualising the eSNPs and eGenes overlap all three levels. (c) An example association plot for rs505460-NCALD expression on all three levels. We selected the full gene, four out of 30 transcripts and four out of 68 exons. The SNP shows only effect on the expression of exon ENSE00001231633. The x axis indicates expression residuals. (d) Overlap between genes from the three eQTL levels and eQTL and sQTL genes from GTEx or CMC data sets in the human cortex. (e) Genomic overlap between variant effect predictor (VEP). (f) 15- state model of the Roadmap Epigenomics Project measured in DLPFC. (g) GWAS traits and eSNPs from all three levels. Results are presented as bar plots showing the odds ratios with Asterisk indicating significant p-values for the enrichment test, and proportions of overlap with the original data sets. ADHD attention-deficit/hyperactivity disorder, ASD autism spectrum disorder, BMI body mass index, SCZ schizophrenia, MDD major depressive disorder, BP bipolar disorder, CDG cross disorder meta-analysis, T2D type 2 diabetes and EA educational attainment.

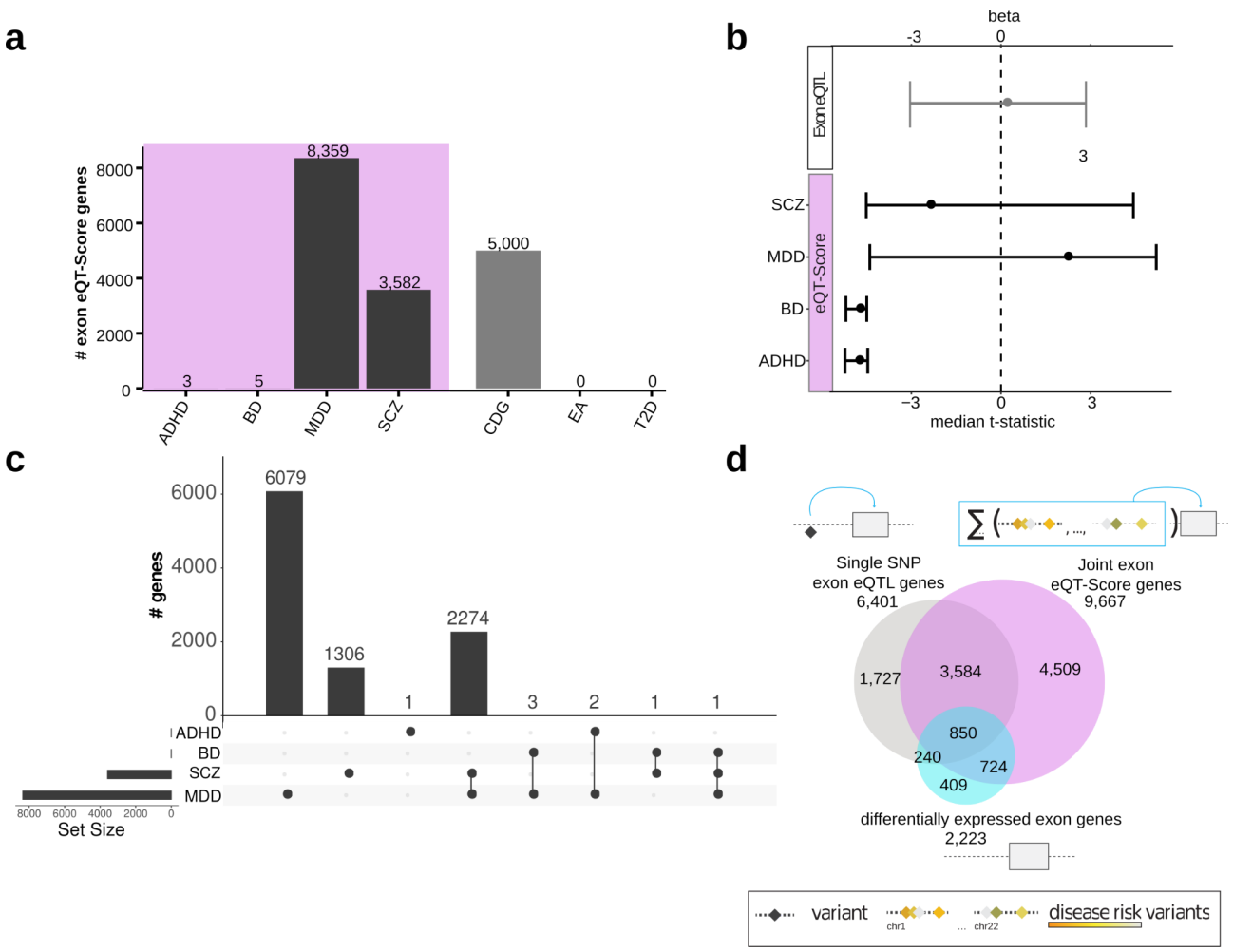


Figure 3: Multi-SNP effects and their influence on cortical expression. (a) Bar plot shows the significant number of exon expression - polygenic risk score associations, also called exon expression quantitative trait score (exon eQT-Score). The y-axis shows the counts of exon eQT-Score genes and the x-axis displays the

medRxiv preprint doi: https://doi.org/10.1101/2022.08.09.22278128; this version posted August 11, 2022. The copyright holder for this preprint **(which was not certified by peer review)** is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity.
It is made available under a CC-BY-NC 4.0 International license .

GWAS in the eQT-Score calculation. (b) Forest plot of the effect size of the exon eQTL genes and exon eQT-Score genes. The y-axis shows median beta or t-statistic and the x-axis displays the GWAS used in the eQT-Score calculation. Exon eQT-Score genes exhibit an overall larger effect size (median absolute SCZ t-stats = -2.37, range from -4.51 to 4.43, median absolute MDD t-stats = 2.23, range from -4.39 to 5.2, median absolute BD t-stats = -4.37, range from -5.19 to -4.49, median absolute ADHD t-stats = -4.75, range from -5.22 to -4.46) compared with the minimal effects for single exon eQTL genes (median absolute beta = -0.1, range from -3.77 to 2.48, p-value Wilcoxon test $< 2.2 \times 10^{-16}$). (c) Upset plot shows the overlap between ADHD, BD, MDD and SCZ GWAS exon eQT-Score genes. Each row corresponds to a GWAS study and black dots indicate an intersection. Bar chart on the top indicates the size of the intersection. Bar chart on the left shows the number of exon eQT-Score genes. (d) Venn diagram shows the overlap between the exon eQTL genes (grey), the joint genes of the ADHD, BD, MDD and SCZ GWAS exon eQT-Scores (purple) and the differentially expressed exon genes (blue). ADHD attention-deficit/hyperactivity disorder, SCZ schizophrenia, MDD major depressive disorder, BP bipolar disorder, CDG cross disorder meta-analysis, T2D type 2 diabetes and EA educational attainment.

medRxiv preprint doi: https://doi.org/10.1101/2022.08.09.22278128; this version posted August 11, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity.
It is made available under a CC-BY-NC 4.0 International license .

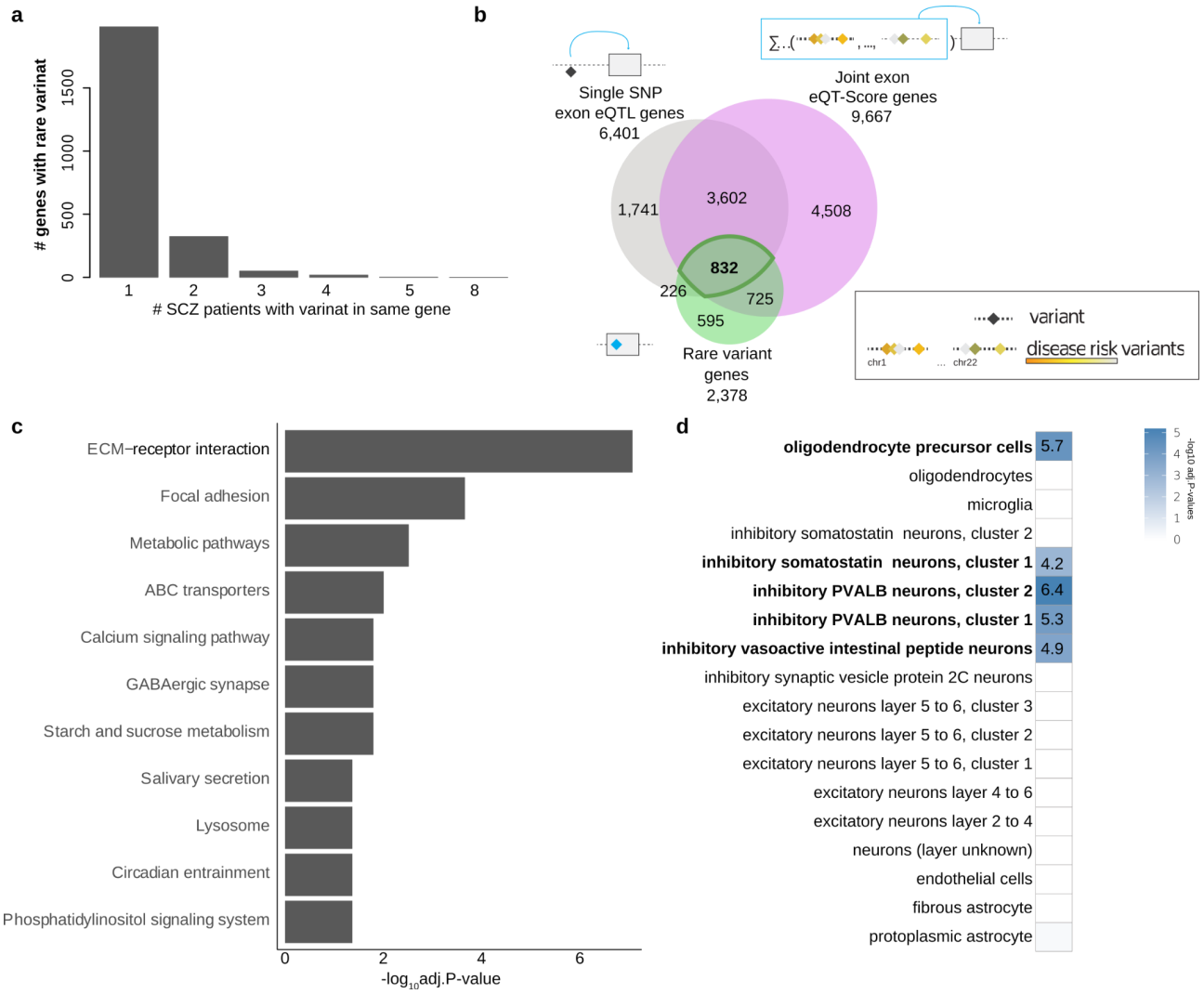


Figure 4: Genes disrupted by rare and common variants. (a) Bar plot shows the frequency of rare variants in index patients of SCZ (n=142 patients). The x-axis shows the number of SCZ index patients with rare variants that disrupt the same gene, and the y-axis shows the number of genes found for that number of patients. (b) Venn diagram of the overlap between single SNP exon eQTL genes (grey), joint exon eQT-Score genes (purple) and genes altered by rare variants in SCZ index patients (green). (c) Bar plot shows the KEGG pathway enrichment for the core set of genes (N = 832). Only significant gene sets are shown. The y-axis shows the enriched KEGG pathways/gene sets and the x-axis shows the -$\log_{10}$ of the adjusted p-value (FDR). (d) Heatmap depicting cell type specificity for the core gene set of enrichment signals defined at gene-level from snRNA-seq data. Odds ratio is given only for significantly enriched cell types (Bonferroni adjusted $P < 0.05$).

medRxiv preprint doi: https://doi.org/10.1101/2022.08.09.22278128; this version posted August 11, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity.
It is made available under a CC-BY-NC 4.0 International license .

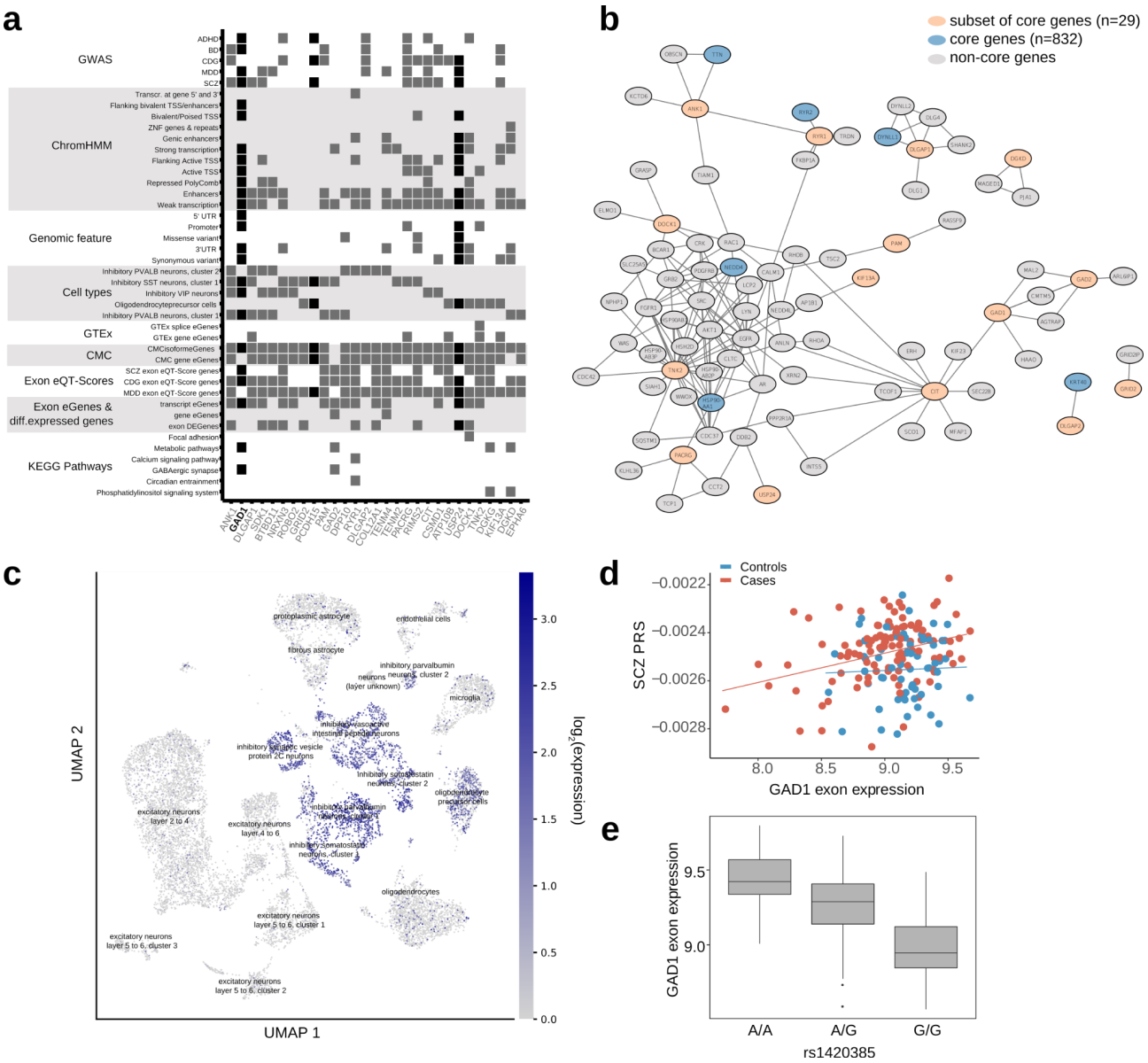


Figure 5: Missense gene *GAD1*, a core set gene enriched in inhibitory neurons. (a) Heatmap of the 29 genes from the core gene set that were significantly enriched (FDR < 0.05) in at least one of the overrepresented five cell types presented in Figure 4E (inhibitory parvalbumin neurons cluster 1 & 2, oligodendrocyte precursor cells, inhibitory somatostatin neurons cluster 1 and inhibitory vasoactive intestinal peptide neurons). The plot shows in which datasets the genes are significant. (b) Protein-protein interaction network (86 nodes; 182 undirected edges) of the 29 subset core genes. The network illustrates connections between 15 of the 29 genes (orange nodes). (c) UMAP of the Glutamate Decarboxylase 1 (GAD1) expression, where grey denotes minimal expression and blue high expression. (d) Association plot for SCZ PRS- ENSE00003538005.1 exon expression of *GAD1*, separated into cases and controls (SCZ exon eQT-Score). The x axis indicates expression residuals; the y axis shows SCZ PRS values. (e) The box plot indicates the top exon-eQTL of *GAD1*: ENSE00003699191.1 exon expression - rs1420385. The y axis indicates expression values; the x axis shows genotypes.

medRxiv preprint doi: https://doi.org/10.1101/2022.08.09.22278128; this version posted August 11, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity.
It is made available under a CC-BY-NC 4.0 International license .

TABLES

Table 1: Summary of the human postmortem brain samples and their demographic and clinical variability. For continuous data the mean ± standard error and for categorical data the categories separated by dashes are given for control and case subjects as well as the different diagnosis types (SCZ, BD, MDD) of the cases, N = number of samples, Age = years, sex = males (M) and females (F), postmortem interval (PMI) = hours (h), potential of hydrogen (pH), RNA integrity number (RIN), suicide = yes (Y) and no (N), cause of death (CoD) = grouped into natural (N), non-violent (NV) and violent (V) death and Dim 1 - 4 = the first four dimensions of the ancestry information calculated with the sample genotypes.

Diagnosis	N	Age (years)	Sex (M/F)	PMI (h)	pH	RIN	Suicide (Y/N)	CoD (N/NV/V)	Dim 1	Dim 2	Dim 3	Dim 4
Controls	62	49.33 ± 2.0	49/13	41.89 ± 1.9	6.37 ± 0.03	7.69 ± 0.1	0/62	56/1/5	-8e-5 ± 0.002	-0.0008 ± 0.001	-0.0019 ± 0.002	0.0009 ± 0.002
Cases	107	50.06 ±1.7	74/33	42 ± 1.3	6.36 ± 0.02	7.6 ± 0.08	48/59	39/24/44	-0.0005 ± 0.002	-0.0011 ± 0.001	-9e-5 ± 0.001	-0.0005 ± 0.001
SCZ	68	46.88 ± 2.1	54/14	43.59 ± 1.6	6.31 ± 0.03	7.82 ± 0.1	33/35	30/10/28	0.0002 ± 0.002	-0.0016 ± 0.002	-0.0008 ± 0.002	-0.0004 ± 0.001
BD	15	57.61 ± 3.6	8/7	39.57 ± 3.9	6.27 ± 0.04	7.07 ± 0.1	5/10	6/4/5	0.0013 ± 0.005	0.0005 ± 0.003	0.0032 ± 0.003	-0.0038 ± 0.004
MDD	24	54.35 ± 3.7	12/12	38.98 ± 2.9	6.54 ± 0.04	7.42 ± 0.1	21/3	3/10/11	-0.003 ± 0.002	-0.0008 ± 0.003	-6e-5 ± 0.003	0.0009 ± 0.002

Table 2: Summary and demographic of the SCZ index patients. For continuous data the mean ± standard error and for categorical data the categories separated by dashes are given for all schizophrenia subjects as well as the different schizophrenia subtypes according to DSM-IV. Given are the diagnosis, number of samples (N), sex differentiated into males (M) and females (F), age at the interview in years and the ethnicity.

Diagnosis	N	Sex (M/F)	Age	Ethnicity
SCZ	142	107/35	34.72 ± 0.97	European
Disorganised	26	19/7	30.81 ± 1.86	European
Catatonic	13	10/3	35.00 ± 2.98	European
Paranoid	85	55/30	35.35 ± 1.23	European
Residual	10	4/6	41.6 ± 3.9	European

medRxiv preprint doi: https://doi.org/10.1101/2022.08.09.22278128; this version posted August 11, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity.
It is made available under a CC-BY-NC 4.0 International license .

Undifferentiated	8	6/2	36.43 ± 5.97	European

Table 3: Demographic of the two samples used for snRNA-seq. For continuous data the mean ± standard error except for age and for categorical data the categories separated by dashes are given for both subjects. Given are the diagnosis, number of samples (N), sex differentiated between male (M) and female (F), age range in years, postmortem interval (PMI) in hours (h), RNA integrity number (RIN), cause of death (CoD) grouped into natural (N), non-violent (NV) and violent (V) death and ethnicity.

Diagnosis	N	Sex (M/F)	Age (years)	PMI (h)	RIN	CoD (N/NV/V)	Ethnicity
Control	2	1/1	52 - 70	35.75 ± 14.25	7.2 ± 0.5	2/0/0	European

medRxiv preprint doi: https://doi.org/10.1101/2022.08.09.22278128; this version posted August 11, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity.
It is made available under a CC-BY-NC 4.0 International license .

REFERENCES

Aguet, F., Barbeira, A.N., Bonazzola, R., Brown, A., Castel, S.E., Jo, B., Kasela, S., Kim-Hellmuth, S., Liang, Y., Oliva, M., et al. The GTEx Consortium atlas of genetic regulatory effects across human tissues. https://doi.org/10.1101/787903.

Ben-David, E., Granot-Hershkovitz, E., Monderer-Rothkoff, G., Lerer, E., Levi, S., Yaari, M., Ebstein, R.P., Yirmiya, N., and Shifman, S. (2011). Identification of a functional rare variant in autism using genome-wide screen for monoallelic expression. Hum. Mol. Genet. *20*, 3632–3641. .

Berretta, S. (2012). Extracellular matrix abnormalities in schizophrenia. Neuropharmacology *62*, 1584–1597. https://doi.org/10.1016/j.neuropharm.2011.08.010.

Bhalala, O.G., Nath, A.P., UK Brain Expression Consortium, Inouye, M., and Sibley, C.R. (2018). Identification of expression quantitative trait loci associated with schizophrenia and affective disorders in normal brain tissue. PLoS Genet. *14*, e1007607. .

Bierbaum, S., Hintze, V., and Scharnweber, D. (2017). 2.8 Artificial Extracellular Matrices to Functionalize Biomaterial Surfaces ☆. Comprehensive Biomaterials II 147–178. https://doi.org/10.1016/b978-0-12-803581-8.10206-1.

Blondel, V.D., Guillaume, J.-L., Lambiotte, R., and Lefebvre, E. (2008). Fast unfolding of communities in large networks. Journal of Statistical Mechanics: Theory and Experiment *2008*, P10008. https://doi.org/10.1088/1742-5468/2008/10/p10008.

Cardno, A.G., and Owen, M.J. (2014). Genetic relationships between schizophrenia, bipolar disorder, and schizoaffective disorder. Schizophr. Bull. *40*, 504–515. .

Carvalho, B.S., and Irizarry, R.A. (2010). A framework for oligonucleotide microarray preprocessing. Bioinformatics *26*, 2363–2367. .

Chang, C.C., Chow, C.C., Tellier, L.C., Vattikuti, S., Purcell, S.M., and Lee, J.J. (2015). Second-generation PLINK: rising to the challenge of larger and richer datasets. Gigascience *4*, 7. .

Choi, S.W., and O'Reilly, P.F. (2019). PRSice-2: Polygenic Risk Score software for biobank-scale data. Gigascience *8*. https://doi.org/10.1093/gigascience/giz082.

Chung, D.W., Volk, D.W., Arion, D., Zhang, Y., Sampson, A.R., and Lewis, D.A. (2016). Dysregulated ErbB4 Splicing in Schizophrenia: Selective Effects on Parvalbumin Expression. Am. J. Psychiatry *173*, 60–68. .

Clinton, S.M., Haroutunian, V., Davis, K.L., and Meador-Woodruff, J.H. (2003). Altered transcript expression of NMDA receptor-associated postsynaptic proteins in the thalamus of subjects with schizophrenia. Am. J. Psychiatry *160*, 1100–1109. .

Consortium, G., and GTEx Consortium (2017). Genetic effects on gene expression across human tissues. Nature *550*, 204–213. https://doi.org/10.1038/nature24277.

Consortium, T.A.S.D.W.G. of T.P.G., and The Autism Spectrum Disorders Working Group of The Psychiatric Genomics Consortium (2017). Meta-analysis of GWAS of over 16,000 individuals with autism spectrum disorder highlights a novel locus at 10q24.32 and a significant overlap with schizophrenia. Molecular Autism *8*. https://doi.org/10.1186/s13229-017-0137-9.

Consortium, T.I.S., and The International Schizophrenia Consortium (2009). Common polygenic variation contributes to risk of schizophrenia and bipolar disorder. Nature *460*, 748–752. https://doi.org/10.1038/nature08185.

Consortium, S.W.G. of T.P.G., Schizophrenia Working Group of the Psychiatric Genomics Consortium, Ripke, S., Walters, J.T.R., and O'Donovan, M.C. Mapping genomic loci prioritises genes and implicates synaptic biology in schizophrenia. https://doi.org/10.1101/2020.09.12.20192922.

Cox, D.R., and Snell, E.J. (1989). Analysis of Binary Data, Second Edition (CRC Press).

medRxiv preprint doi: https://doi.org/10.1101/2022.08.09.22278128; this version posted August 11, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity.
It is made available under a CC-BY-NC 4.0 International license .

Cross-Disorder Group of the Psychiatric Genomics Consortium (2013). Identification of risk loci with shared effects on five major psychiatric disorders: a genome-wide analysis. Lancet *381*, 1371–1379. .

Cross-Disorder Group of the Psychiatric Genomics Consortium (2019). Genomic Relationships, Novel Loci, and Pleiotropic Mechanisms across Eight Psychiatric Disorders. Cell *179*, 1469–1482.e11. .

Curtis, D. Polygenic risk score for schizophrenia is not strongly associated with the expression of specific genes or gene sets. https://doi.org/10.1101/205518.

Dean, B., and Scarr, E. (2016). COMT genotype is associated with differential expression of muscarinic M1 receptors in human cortex. Am. J. Med. Genet. B Neuropsychiatr. Genet. *171*, 784–789. .

Demontis, D., Walters, R.K., Martin, J., Mattheisen, M., Als, T.D., Agerbo, E., Baldursson, G., Belliveau, R., Bybjerg-Grauholm, J., Bækvad-Hansen, M., et al. (2019). Discovery of the first genome-wide significant risk loci for attention deficit/hyperactivity disorder. Nat. Genet. *51*, 63–75. .

Di Sole, F., Vadnagara, K., Moe, O.W., and Babich, V. (2012). Calcineurin homologous protein: a multifunctional Ca2+-binding protein family. Am. J. Physiol. Renal Physiol. *303*, F165–F179. .

Docherty, A.R., Moscati, A.A., and Fanous, A.H. (2016). Cross-Disorder Psychiatric Genomics. Curr Behav Neurosci Rep *3*, 256–263. .

Edgar, N., and Sibille, E. (2012). A putative functional role for oligodendrocytes in mood regulation. Transl. Psychiatry *2*, e109. .

Edwards, S.L., Beesley, J., French, J.D., and Dunning, A.M. (2013). Beyond GWASs: illuminating the dark road from association to function. Am. J. Hum. Genet. *93*, 779–797. .

Ernst, J., and Kellis, M. (2012). ChromHMM: automating chromatin-state discovery and characterization. Nat. Methods *9*, 215–216. .

Fisher, R.A. (1992). Statistical Methods for Research Workers. Springer Series in Statistics 66–70. https://doi.org/10.1007/978-1-4612-4380-9_6.

Frankish, A., Diekhans, M., Ferreira, A.-M., Johnson, R., Jungreis, I., Loveland, J., Mudge, J.M., Sisu, C., Wright, J., Armstrong, J., et al. (2019). GENCODE reference annotation for the human and mouse genomes. Nucleic Acids Res. *47*, D766–D773. .

Fromer, M., Roussos, P., Sieberts, S.K., Johnson, J.S., Kavanagh, D.H., Perumal, T.M., Ruderfer, D.M., Oh, E.C., Topol, A., Shah, H.R., et al. (2016). Gene expression elucidates functional impact of polygenic risk for schizophrenia. Nat. Neurosci. *19*, 1442–1453. .

Ganapathiraju, M.K., Thahir, M., Handen, A., Sarkar, S.N., Sweet, R.A., Nimgaonkar, V.L., Loscher, C.E., Bauer, E.M., and Chaparala, S. (2016). Schizophrenia interactome with 504 novel protein–protein interactions. Npj Schizophrenia *2*. https://doi.org/10.1038/npjschz.2016.12.

Gandal, M.J., Zhang, P., Hadjimichael, E., Walker, R.L., Chen, C., Liu, S., Won, H., van Bakel, H., Varghese, M., Wang, Y., et al. (2018). Transcriptome-wide isoform-level dysregulation in ASD, schizophrenia, and bipolar disorder. Science *362*. https://doi.org/10.1126/science.aat8127.

Ganesh, S., Ahmed P, H., Nadella, R.K., More, R.P., Seshadri, M., Viswanath, B., Rao, M., Jain, S., ADBS Consortium, and Mukherjee, O. (2019). Exome sequencing in families with severe mental illness identifies novel and rare variants in genes implicated in Mendelian neuropsychiatric syndromes. Psychiatry Clin. Neurosci. *73*, 11–19. .

Gardiner, E.J., Cairns, M.J., Liu, B., Beveridge, N.J., Carr, V., Kelly, B., Scott, R.J., and Tooney, P.A. (2013). Gene expression analysis reveals schizophrenia-associated dysregulation of immune pathways in peripheral blood mononuclear cells. J. Psychiatr. Res. *47*, 425–437. .

Gazal, S., Finucane, H.K., Furlotte, N.A., Loh, P.-R., Palamara, P.F., Liu, X., Schoech, A., Bulik-Sullivan, B., Neale, B.M., Gusev, A., et al. (2017). Linkage disequilibrium-dependent architecture of human complex traits shows action of

medRxiv preprint doi: https://doi.org/10.1101/2022.08.09.22278128; this version posted August 11, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity.
It is made available under a CC-BY-NC 4.0 International license .

negative selection. Nat. Genet. *49*, 1421–1427. .

Gibson, G. (2012). Rare and common variants: twenty arguments. Nat. Rev. Genet. *13*, 135–145. .

Glatt, S.J., Chandler, S.D., Bousman, C.A., Chana, G., Lucero, G.R., Tatro, E., May, T., Lohr, J.B., Kremen, W.S., Everall, I.P., et al. (2009). Alternatively Spliced Genes as Biomarkers for Schizophrenia, Bipolar Disorder and Psychosis: A Blood-Based Spliceome-Profiling Exploratory Study. Curr. Pharmacogenomics Person. Med. *7*, 164–188. .

Grindberg, R.V., Yee-Greenbaum, J.L., McConnell, M.J., Novotny, M., O'Shaughnessy, A.L., Lambert, G.M., Araúzo-Bravo, M.J., Lee, J., Fishman, M., Robbins, G.E., et al. (2013). RNA-sequencing from single nuclei. Proc. Natl. Acad. Sci. U. S. A. *110*, 19802–19807. .

Guan, L., Yang, Q., Gu, M., Chen, L., and Zhang, X. (2014). Exon expression QTL (eeQTL) analysis highlights distant genomic variations associated with splicing regulation. Quantitative Biology *2*, 71–79. https://doi.org/10.1007/s40484-014-0031-9.

Gulsuner, S., Walsh, T., Watts, A.C., Lee, M.K., Thornton, A.M., Casadei, S., Rippey, C., Shahin, H., Consortium on the Genetics of Schizophrenia (COGS), PAARTNERS Study Group, et al. (2013). Spatial and temporal mapping of de novo mutations in schizophrenia to a fetal prefrontal cortical network. Cell *154*, 518–529. .

Hoffman, G.E., Bendl, J., Voloudakis, G., Montgomery, K.S., Sloofman, L., Wang, Y.-C., Shah, H.R., Hauberg, M.E., Johnson, J.S., Girdhar, K., et al. (2019). CommonMind Consortium provides transcriptomic and epigenomic data for Schizophrenia and Bipolar Disorder. Sci Data *6*, 180. .

Howard, D.M., Adams, M.J., Clarke, T.-K., Hafferty, J.D., Gibson, J., Shirali, M., Coleman, J.R.I., Hagenaars, S.P., Ward, J., Wigmore, E.M., et al. (2019). Genome-wide meta-analysis of depression identifies 102 independent variants and highlights the importance of the prefrontal brain regions. Nat. Neurosci. *22*, 343–352. .

Huang, Z.J., and Paul, A. (2019). The diversity of GABAergic neurons and neural communication elements. Nat. Rev. Neurosci. *20*, 563–572. .

Huang, M.-L., Khoh, T.-T., Lu, S.-J., Pan, F., Chen, J.-K., Hu, J.-B., Hu, S.-H., Xu, W.-J., Zhou, W.-H., Wei, N., et al. (2017). Relationships between dorsolateral prefrontal cortex metabolic change and cognitive impairment in first-episode neuroleptic-naive schizophrenia patients. Medicine *96*, e7228. .

Jaffe, A.E., Straub, R.E., Shin, J.H., Tao, R., Gao, Y., Collado-Torres, L., Kam-Thong, T., Xi, H.S., Quan, J., Chen, Q., et al. (2018). Developmental and genetic regulation of the human cortex transcriptome illuminate schizophrenia pathogenesis. Nat. Neurosci. *21*, 1117–1125. .

Johnson, W.E., Li, C., and Rabinovic, A. (2007). Adjusting batch effects in microarray expression data using empirical Bayes methods. Biostatistics *8*, 118–127. .

Jonas, K.G., Lencz, T., Li, K., Malhotra, A.K., Perlman, G., Fochtmann, L.J., Bromet, E.J., and Kotov, R. (2019). Schizophrenia polygenic risk score and 20-year course of illness in psychotic disorders. Transl. Psychiatry *9*, 300. .

Jonge, J.C. de, de Jonge, J.C., Vinkers, C.H., Hulshoff Pol, H.E., and Marsman, A. (2017). GABAergic Mechanisms in Schizophrenia: Linking Postmortem and In Vivo Studies. Frontiers in Psychiatry *8*. https://doi.org/10.3389/fpsyt.2017.00118.

Kim, Y., Xia, K., Tao, R., Giusti-Rodriguez, P., Vladimirov, V., van den Oord, E., and Sullivan, P.F. (2014). A meta-analysis of gene expression quantitative trait loci in brain. Translational Psychiatry *4*, e459–e459. https://doi.org/10.1038/tp.2014.96.

Kircher, M., Witten, D.M., Jain, P., O'Roak, B.J., Cooper, G.M., and Shendure, J. (2014). A general framework for estimating the relative pathogenicity of human genetic variants. Nat. Genet. *46*, 310–315. .

Laiho, A., and Elo, L.L. (2014). A note on an exon-based strategy to identify differentially expressed genes in RNA-seq experiments. PLoS One *9*, e115964. .

Lake, B.B., Codeluppi, S., Yung, Y.C., Gao, D., Chun, J., Kharchenko, P.V., Linnarsson, S., and Zhang, K. (2017). A

medRxiv preprint doi: https://doi.org/10.1101/2022.08.09.22278128; this version posted August 11, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity.
It is made available under a CC-BY-NC 4.0 International license .

comparative strategy for single-nucleus and single-cell transcriptomes confirms accuracy in predicted cell-type expression from nuclear RNA. Sci. Rep. *7*, 6031. .

Lee, J.J., Wedow, R., Okbay, A., Kong, E., Maghzian, O., Zacher, M., Nguyen-Viet, T.A., Bowers, P., Sidorenko, J., Karlsson Linnér, R., et al. (2018). Gene discovery and polygenic prediction from a genome-wide association study of educational attainment in 1.1 million individuals. Nat. Genet. *50*, 1112–1121. .

Leek, J.T., and Storey, J.D. (2007). Capturing heterogeneity in gene expression studies by surrogate variable analysis. PLoS Genet. *3*, 1724–1735. .

Lek, M., Karczewski, K.J., Minikel, E.V., Samocha, K.E., Banks, E., Fennell, T., O'Donnell-Luria, A.H., Ware, J.S., Hill, A.J., Cummings, B.B., et al. (2016). Analysis of protein-coding genetic variation in 60,706 humans. Nature *536*, 285–291. .

Lun, A.T.L., McCarthy, D.J., and Marioni, J.C. (2016). A step-by-step workflow for low-level analysis of single-cell RNA-seq data with Bioconductor. F1000Res. *5*, 2122. .

Magri, C., Giacopuzzi, E., La Via, L., Bonini, D., Ravasio, V., Elhussiny, M.E.A., Orizio, F., Gangemi, F., Valsecchi, P., Bresciani, R., et al. (2018). A novel homozygous mutation in GAD1 gene described in a schizophrenic patient impairs activity and dimerization of GAD67 enzyme. Scientific Reports *8*. https://doi.org/10.1038/s41598-018-33924-8.

Manolio, T.A., Collins, F.S., Cox, N.J., Goldstein, D.B., Hindorff, L.A., Hunter, D.J., McCarthy, M.I., Ramos, E.M., Cardon, L.R., Chakravarti, A., et al. (2009). Finding the missing heritability of complex diseases. Nature *461*, 747–753. .

Martin, M.V., Rollins, B., Sequeira, P.A., Mesén, A., Byerley, W., Stein, R., Moon, E.A., Akil, H., Jones, E.G., Watson, S.J., et al. (2009). Exon expression in lymphoblastoid cell lines from subjects with schizophrenia before and after glucose deprivation. BMC Med. Genomics *2*, 62. .

Matevossian, A., and Akbarian, S. (2008). Neuronal nuclei isolation from human postmortem brain tissue. J. Vis. Exp. https://doi.org/10.3791/914.

Maurano, M.T., Humbert, R., Rynes, E., Thurman, R.E., Haugen, E., Wang, H., Reynolds, A.P., Sandstrom, R., Qu, H., Brody, J., et al. (2012). Systematic localization of common disease-associated variation in regulatory DNA. Science *337*, 1190–1195. .

McInnes, L., Healy, J., Saul, N., and Großberger, L. (2018). UMAP: Uniform Manifold Approximation and Projection. Journal of Open Source Software *3*, 861. https://doi.org/10.21105/joss.00861.

McLaren, W., Gil, L., Hunt, S.E., Riat, H.S., Ritchie, G.R.S., Thormann, A., Flicek, P., and Cunningham, F. (2016). The Ensembl Variant Effect Predictor. Genome Biol. *17*, 122. .

Melé, M., Ferreira, P.G., Reverter, F., DeLuca, D.S., Monlong, J., Sammeth, M., Young, T.R., Goldmann, J.M., Pervouchine, D.D., Sullivan, T.J., et al. (2015). The human transcriptome across tissues and individuals. Science *348*, 660–665. https://doi.org/10.1126/science.aaa0355.

Miller, S.A., Dykes, D.D., and Polesky, H.F. (1988). A simple salting out procedure for extracting DNA from human nucleated cells. Nucleic Acids Res. *16*, 1215. .

Mostafavi, H., Spence, J.P., Naqvi, S., and Pritchard, J.K. (2022). Limited overlap of eQTLs and GWAS hits due to systematic differences in discovery.

Muglia, P., Tozzi, F., Galwey, N.W., Francks, C., Upmanyu, R., Kong, X.Q., Antoniades, A., Domenici, E., Perry, J., Rothen, S., et al. (2010). Genome-wide association study of recurrent major depressive disorder in two European case-control cohorts. Mol. Psychiatry *15*, 589–601. .

Nagelkerke, N.J.D. (1991). A note on a general definition of the coefficient of determination. Biometrika *78*, 691–692. https://doi.org/10.1093/biomet/78.3.691.

Nagy, C., Maitra, M., Tanti, A., Suderman, M., Théroux, J.-F., Davoli, M.A., Perlman, K., Yerko, V., Wang, Y.C.,

medRxiv preprint doi: https://doi.org/10.1101/2022.08.09.22278128; this version posted August 11, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity.
It is made available under a CC-BY-NC 4.0 International license .

Tripathy, S.J., et al. (2020). Single-nucleus transcriptomics of the prefrontal cortex in major depressive disorder implicates oligodendrocyte precursor cells and excitatory neurons. Nat. Neurosci. *23*, 771–781. .

Niu, H.-M., Yang, P., Chen, H.-H., Hao, R.-H., Dong, S.-S., Yao, S., Chen, X.-F., Yan, H., Zhang, Y.-J., Chen, Y.-X., et al. (2019). Comprehensive functional annotation of susceptibility SNPs prioritized 10 genes for schizophrenia. Translational Psychiatry *9*. https://doi.org/10.1038/s41398-019-0398-5.

Oh, J.-H., Jang, S.J., Kim, J., Sohn, I., Lee, J.-Y., Cho, E.J., Chun, S.-M., and Sung, C.O. (2020). Spontaneous mutations in the single TTN gene represent high tumor mutation burden. NPJ Genom Med *5*, 33. .

Oughtred, R., Rust, J., Chang, C., Breitkreutz, B.-J., Stark, C., Willems, A., Boucher, L., Leung, G., Kolas, N., Zhang, F., et al. (2021). The BioGRID database: A comprehensive biomedical resource of curated protein, genetic, and chemical interactions. Protein Sci. *30*, 187–200. .

Pan, Q., Shai, O., Lee, L.J., Frey, B.J., and Blencowe, B.J. (2008). Deep surveying of alternative splicing complexity in the human transcriptome by high-throughput sequencing. Nat. Genet. *40*, 1413–1415. .

Polański, K., Young, M.D., Miao, Z., Meyer, K.B., Teichmann, S.A., and Park, J.-E. (2020). BBKNN: fast batch alignment of single cell transcriptomes. Bioinformatics *36*, 964–965. .

Raabe, F.J., Galinski, S., Papiol, S., Falkai, P.G., Schmitt, A., and Rossner, M.J. (2018). Studying and modulating schizophrenia-associated dysfunctions of oligodendrocytes with patient-specific cell systems. NPJ Schizophr *4*, 23. .

Raj, B., and Blencowe, B.J. (2015). Alternative Splicing in the Mammalian Nervous System: Recent Insights into Mechanisms and Functional Roles. Neuron *87*, 14–27. .

Ramasamy, A., Trabzuni, D., Guelfi, S., Varghese, V., Smith, C., Walker, R., De, T., UK Brain Expression Consortium, North American Brain Expression Consortium, Coin, L., et al. (2014). Genetic variability in the regulation of gene expression in ten regions of the human brain. Nat. Neurosci. *17*, 1418–1428. .

Ruffier, M., Kähäri, A., Komorowska, M., Keenan, S., Laird, M., Longden, I., Proctor, G., Searle, S., Staines, D., Taylor, K., et al. (2017). Ensembl core software resources: storage and programmatic access for DNA sequence and genome annotation. Database *2017*. https://doi.org/10.1093/database/bax020.

Ruzicka, B., Mohammadi, S., Davila-Velderrain, J., Subburaju, S., Tso, R., Hourihan, M., and Kellis, M. (2021). Single-Cell Dissection of Schizophrenia Reveals Neurodevelopmental-Synaptic Link and Transcriptional Resilience Associated Cellular State. Biological Psychiatry *89*, S106. https://doi.org/10.1016/j.biopsych.2021.02.273.

Ruzicka, W.B., Brad Ruzicka, W., Mohammadi, S., Davila-Velderrain, J., Subburaju, S., Tso, D.R., Hourihan, M., and Kellis, M. Single-cell dissection of schizophrenia reveals neurodevelopmental-synaptic axis and transcriptional resilience. https://doi.org/10.1101/2020.11.06.20225342.

Saha, S., Chant, D., Welham, J., and McGrath, J. (2005). A systematic review of the prevalence of schizophrenia. PLoS Med. *2*, e141. .

Scarr, E., Cowie, T.F., Kanellakis, S., Sundram, S., Pantelis, C., and Dean, B. (2009). Decreased cortical muscarinic receptors define a subgroup of subjects with schizophrenia. Mol. Psychiatry *14*, 1017–1023. .

Scarr, E., Udawela, M., Thomas, E.A., and Dean, B. (2018). Changed gene expression in subjects with schizophrenia and low cortical muscarinic M1 receptors predicts disrupted upstream pathways interacting with that receptor. Mol. Psychiatry *23*, 295–303. .

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

Schulze, T.G., Akula, N., Breuer, R., Steele, J., Nalls, M.A., Singleton, A.B., Degenhardt, F.A., Nöthen, M.M., Cichon, S., Rietschel, M., et al. (2014). Molecular genetic overlap in bipolar disorder, schizophrenia, and major depressive disorder. World J. Biol. Psychiatry *15*, 200–208. .

Scott, R.A., Scott, L.J., Mägi, R., Marullo, L., Gaulton, K.J., Kaakinen, M., Pervjakova, N., Pers, T.H., Johnson, A.D.,

medRxiv preprint doi: https://doi.org/10.1101/2022.08.09.22278128; this version posted August 11, 2022. The copyright holder for this preprint **(which was not certified by peer review)** is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity.
It is made available under a CC-BY-NC 4.0 International license .

Eicher, J.D., et al. (2017). An Expanded Genome-Wide Association Study of Type 2 Diabetes in Europeans. Diabetes *66*, 2888–2902. .

Sethi, M.K., and Zaia, J. (2017). Extracellular matrix proteomics in schizophrenia and Alzheimer's disease. Analytical and Bioanalytical Chemistry *409*, 379–394. https://doi.org/10.1007/s00216-016-9900-6.

Shabalin, A.A. (2012). Matrix eQTL: ultra fast eQTL analysis via large matrix operations. Bioinformatics *28*, 1353–1358. .

Shannon, P., Markiel, A., Ozier, O., Baliga, N.S., Wang, J.T., Ramage, D., Amin, N., Schwikowski, B., and Ideker, T. (2003). Cytoscape: a software environment for integrated models of biomolecular interaction networks. Genome Res. *13*, 2498–2504. .

Shi, H., Kichaev, G., and Pasaniuc, B. (2016). Contrasting the Genetic Architecture of 30 Complex Traits from Summary Association Data. Am. J. Hum. Genet. *99*, 139–153. .

Smith, K.M. (2018). Hyperactivity in mice lacking one allele of the glutamic acid decarboxylase 67 gene. Atten. Defic. Hyperact. Disord. *10*, 267–271. .

Stahl, E.A., Breen, G., Forstner, A.J., McQuillin, A., Ripke, S., Trubetskoy, V., Mattheisen, M., Wang, Y., Coleman, J.R., Gaspar, H.A., et al. (2019). Genome-wide association study identifies 30 loci associated with bipolar disorder. Nature Genetics *51*, 793–803. .

Stenbacka, M., and Jokinen, J. (2015). Violent and non-violent methods of attempted and completed suicide in Swedish young men: the role of early risk factors. BMC Psychiatry *15*, 196. .

Stephan, K.E., Bach, D.R., Fletcher, P.C., Flint, J., Frank, M.J., Friston, K.J., Heinz, A., Huys, Q.J.M., Owen, M.J., Binder, E.B., et al. (2016). Charting the landscape of priority problems in psychiatry, part 1: classification and diagnosis. Lancet Psychiatry *3*, 77–83. .

Strauss, W.M. (2001). Preparation of Genomic DNA from Mammalian Tissue. Current Protocols in Neuroscience https://doi.org/10.1002/0471142301.nsa01hs06.

Su, C.-H., D, D., and Tarn, W.-Y. (2018). Alternative Splicing in Neurogenesis and Brain Development. Front Mol Biosci *5*, 12. .

Sullivan, P.F., Agrawal, A., Bulik, C.M., Andreassen, O.A., Børglum, A.D., Breen, G., Cichon, S., Edenberg, H.J., Faraone, S.V., Gelernter, J., et al. (2018). Psychiatric Genomics: An Update and an Agenda. American Journal of Psychiatry *175*, 15–27. https://doi.org/10.1176/appi.ajp.2017.17030283.

Takata, A., Matsumoto, N., and Kato, T. (2017). Genome-wide identification of splicing QTLs in the human brain and their enrichment among schizophrenia-associated loci. Nat. Commun. *8*, 14519. .

Uhlhaas, P.J., and Singer, W. (2012). Neuronal Dynamics and Neuropsychiatric Disorders: Toward a Translational Paradigm for Dysfunctional Large-Scale Networks. Neuron *75*, 963–980. https://doi.org/10.1016/j.neuron.2012.09.004.

Upadhyay, A., Hosseinibarkooie, S., Schneider, S., Kaczmarek, A., Torres-Benito, L., Mendoza-Ferreira, N., Overhoff, M., Rombo, R., Grysko, V., Kye, M.J., et al. (2019). Neurocalcin Delta Knockout Impairs Adult Neurogenesis Whereas Half Reduction Is Not Pathological. Frontiers in Molecular Neuroscience *12*. https://doi.org/10.3389/fnmol.2019.00019.

Velmeshev, D., Schirmer, L., Jung, D., Haeussler, M., Perez, Y., Mayer, S., Bhaduri, A., Goyal, N., Rowitch, D.H., and Kriegstein, A.R. (2019). Single-cell genomics identifies cell type–specific molecular changes in autism. Science *364*, 685–689. https://doi.org/10.1126/science.aav8130.

Vercauteren, F.G.G., Flores, G., Ma, W., Chabot, J.-G., Geenen, L., Clerens, S., Fazel, A., Bergeron, J.J.M., Srivastava, L.K., Arckens, L., et al. (2007). An organelle proteomic method to study neurotransmission-related proteins, applied to a neurodevelopmental model of schizophrenia. Proteomics *7*, 3569–3579. .

Wainschtein, P., Jain, D.P., Yengo, L., Zheng, Z., Adrienne Cupples, L., Shadyab, A.H., McKnight, B., Shoemaker, B.M., Mitchell, B.D., Psaty, B.M., et al. (2019). Recovery of trait heritability from whole genome sequence data.

medRxiv preprint doi: https://doi.org/10.1101/2022.08.09.22278128; this version posted August 11, 2022. The copyright holder for this preprint **(which was not certified by peer review)** is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity.
It is made available under a CC-BY-NC 4.0 International license .

Yearbook of Paediatric Endocrinology https://doi.org/10.1530/ey.16.14.15.

Walker, R.L., Ramaswami, G., Hartl, C., Mancuso, N., Gandal, M.J., de la Torre-Ubieta, L., Pasaniuc, B., Stein, J.L., and Geschwind, D.H. (2019). Genetic Control of Expression and Splicing in Developing Human Brain Informs Disease Mechanisms. Cell *179*, 750–771.e22. .

Wang, D., Liu, S., Warrell, J., Won, H., Shi, X., Navarro, F.C.P., Clarke, D., Gu, M., Emani, P., Yang, Y.T., et al. (2018). Comprehensive functional genomic resource and integrative model for the human brain. Science *362*. https://doi.org/10.1126/science.aat8464.

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

Wolf, F.A., Angerer, P., and Theis, F.J. (2018). SCANPY: large-scale single-cell gene expression data analysis. Genome Biol. *19*, 15. .

Wray, N.R., and Gottesman, I.I. (2012). Using summary data from the danish national registers to estimate heritabilities for schizophrenia, bipolar disorder, and major depressive disorder. Front. Genet. *3*, 118. .

Wray, N.R., Ripke, S., Mattheisen, M., Trzaskowski, M., Byrne, E.M., Abdellaoui, A., Adams, M.J., Agerbo, E., Air, T.M., Andlauer, T.M.F., et al. (2018). Genome-wide association analyses identify 44 risk variants and refine the genetic architecture of major depression. Nat. Genet. *50*, 668–681. .

Xu, W., Cohen-Woods, S., Chen, Q., Noor, A., Knight, J., Hosang, G., Parikh, S.V., De Luca, V., Tozzi, F., Muglia, P., et al. (2014). Genome-wide association study of bipolar disorder in Canadian and UK populations corroborates disease loci including SYNE1 and CSMD1. BMC Med. Genet. *15*, 2. .

Xue, C.-B., Xu, Z.-H., Zhu, J., Wu, Y., Zhuang, X.-H., Chen, Q.-L., Wu, C.-R., Hu, J.-T., Zhou, H.-S., Xie, W.-H., et al. (2019). Exome Sequencing Identifies TENM4 as a Novel Candidate Gene for Schizophrenia in the SCZD2 Locus at 11q14-21. Frontiers in Genetics *9*. https://doi.org/10.3389/fgene.2018.00725.

Yeo, G., Holste, D., Kreiman, G., and Burge, C.B. (2004). Variation in alternative splicing across human tissues. Genome Biol. *5*, R74. .

Zhang, C.-Y., Xiao, X., Zhang, Z., Hu, Z., and Li, M. (2022). An alternative splicing hypothesis for neuropathology of schizophrenia: evidence from studies on historical candidate genes and multi-omics data. Mol. Psychiatry *27*, 95–112. .

Zheutlin, A.B., Dennis, J., Karlsson Linnér, R., Moscati, A., Restrepo, N., Straub, P., Ruderfer, D., Castro, V.M., Chen, C.-Y., Ge, T., et al. (2019). Penetrance and Pleiotropy of Polygenic Risk Scores for Schizophrenia in 106,160 Patients Across Four Health Care Systems. Am. J. Psychiatry *176*, 846–855. .

Zhu, Y., Womer, F.Y., Leng, H., Chang, M., Yin, Z., Wei, Y., Zhou, Q., Fu, S., Deng, X., Lv, J., et al. (2019). The Relationship Between Cognitive Dysfunction and Symptom Dimensions Across Schizophrenia, Bipolar Disorder, and Major Depressive Disorder. Front. Psychiatry *10*, 253. .