medRxiv preprint doi: https://doi.org/10.1101/2021.05.26.21257794; this version posted May 29, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity.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 4.0 International license .

Title: Ultra-rare, rare, and common genetic variant analysis converge to implicate negative selection and neuronal processes in the aetiology of schizophrenia

Authors: Wonuola A Akingbuwa[1,2], Anke R Hammerschlag[1,2,3], Meike Bartels[1,2], Michel G Nivard[1*], Christel M Middeldorp[1,3,4*]

[1]Department of Biological Psychology, Vrije Universiteit Amsterdam, Amsterdam, the Netherlands, [2]Amsterdam Public Health Research Institute, Amsterdam University Medical Centres, Amsterdam, the Netherlands, [3]Child Health Research Centre, the University of Queensland, Brisbane, Queensland, Australia, [4]Child and Youth Mental Health Service, Children's Health Queensland Hospital and Health Services, Brisbane, Queensland, Australia

*** Jointly supervised the work**

Corresponding Author: Wonuola Akingbuwa, MSc, Department of Biological Psychology, Vrije Universiteit Amsterdam, Van der Boechorststraat 7, 1081 BT Amsterdam, the Netherlands (o.a.akingbuwa@vu.nl)

NOTE: This preprint reports new research that has not been certified by peer review and should not be used to guide clinical practice.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/10.1101/2021.05.26.21257794; this version posted May 29, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity.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 4.0 International license .

ABSTRACT

Both common and rare genetic variants (minor allele frequency > 1% and < 0.1% respectively) have been implicated in the aetiology of schizophrenia. In this study, we integrate single-cell gene expression data with publicly available Genome-Wide Association Study (GWAS) and exome sequenced data in order to investigate in parallel, the enrichment of common and (ultra-)rare variants related to schizophrenia in several functionally relevant gene sets. Four types of gene sets were constructed 1) protein-truncating variant (PTV)-intolerant (PI) genes 2) genes expressed in brain cell types and neurons ascertained from mouse and human brain tissue 3) genes defined by synaptic function and location and 4) intersection genes, i.e., PI genes that are expressed in the human and mouse brain cell gene sets. We show that common as well as (ultra-)rare schizophrenia-associated variants are overrepresented in PI genes, in excitatory neurons from the prefrontal cortex and hippocampus, medium spiny neurons, and genes enriched for synaptic processes. We also observed stronger enrichment in the intersection genes. Our findings suggest that across the allele frequency spectrum, genes and genetic variants likely to be under stringent selection, and those expressed in particular brain cell types, are involved in the same biological pathways influencing the risk for schizophrenia.

medRxiv preprint doi: https://doi.org/10.1101/2021.05.26.21257794; this version posted May 29, 2021. 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 4.0 International license .

INTRODUCTION

Schizophrenia is a severe and highly heritable psychiatric disorder with onset in late adolescence or early adulthood. It is associated with early mortality and greatly reduced fertility[1, 2], which put selective pressure on genetic variants related to schizophrenia. Despite apparent negative selection, schizophrenia remains highly heritable, with a relatively high prevalence[3]. Its severe clinical presentation and persistence despite negative selection make understanding the nature of the genetic effects on schizophrenia essential.

There is abundant evidence that both common (minor allele frequency > 1%) and rare genetic variants are related to schizophrenia[4-6], with the effects of the variants inversely correlated to their frequency. Rare novel variants can have larger effect, while common variants can only persist in the presence of negative selection if their effects are small[7, 8]. However, common variants, despite smaller individual effects, collectively explain a substantial proportion of the total genetic variance in schizophrenia[6, 9]. The presence of negative selection results in extreme polygenicity where substantial portions of the genome carry variants with tiny individual effects on schizophrenia, yet only critical (core) genes when perturbed by an influential mutation would strongly impact the disorder[10].

The genetic architecture of schizophrenia has forced two separate lines of enquiry into its genetic aetiology. Genome wide associations studies (GWAS) have successfully targeted common variant with small individual effects[6, 11, 12]. Importantly, they have resulted in valuable leads for functional follow-up studies of individual loci. For example, in-depth study of the lead genome wide hit for SCZ in the Major Histocompatibility Complex (MHC) locus has implicated complement component 4 (C4) gene expression and possibly synaptic pruning in puberty in the aetiology of schizophrenia[13]. While further analysis of the schizophrenia locus in the *SLC39A8* gene implicated manganese (Mn) related brain phenotypes in the aetiology of schizophrenia[14].

medRxiv preprint doi: https://doi.org/10.1101/2021.05.26.21257794; this version posted May 29, 2021. 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 4.0 International license .

At the same time, whole exome sequencing (WES) has been used to identify rare mutations with larger effects. Since the variants of interest are rare, these studies require equally large samples. To reduce the multiple testing burden, early research has focused on specific classes of variants for which one can assume a deleterious (risk increasing) effect a-priori. Specifically, researchers have leveraged modest sample sizes by focusing on *singleton* (i.e. only observed once) variants in genes intolerant to mutations, that are predicted to be protein-truncating variants (PTVs) (i.e. disruptive and likely lead to loss of gene function). These variants and genes are the ones most likely to increase the risk of schizophrenia when perturbed by a mutation of consequence[4, 15-17]. This has been borne out by results from these studies, with a recent study implicating 10 genes in which ultra-rare variants are significantly associated with schizophrenia[4].

These parallel lines of genetic inquiry based on different analytical strategies share a common goal: increased understanding of the neuro-biology of schizophrenia. Subsequently, functional genomic analyses are crucial for understanding the pathways and mechanisms via which these disorder-associated genetic variants may act. Functional analyses of common variants from GWAS have implicated the brain and brain-expressed genes in the aetiology of schizophrenia[6, 11]. WES studies, which make it possible to test for a burden of rare variants across shared functional units such as genes or gene sets, have similarly implicated the brain in schizophrenia aetiology[15, 16]. More recently, high throughput single-cell RNA sequencing techniques, which are able to provide expression profiles of individual brain cells at greater resolution, have been developed. They allow prioritization of specific brain cell types associated with disorders or traits. Single-cell expression data of mouse and human brain cells reveal that disorder-associated common and rare variants are enriched in genes expressed in (excitatory) neurons more than in other (non-neuronal) brain cells[4, 18, 19].

In this preregistered study (https://osf.io/uyv2s), we integrated single-cell gene expression data with results from GWAS and exome sequenced data in order to investigate, in parallel, the enrichment of common and (ultra-)rare variants related to schizophrenia in specific brain cell types. We

medRxiv preprint doi: https://doi.org/10.1101/2021.05.26.21257794; this version posted May 29, 2021. 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 4.0 International license .

investigated whether trait-associated common and (ultra-)rare variants were enriched in classes of genes that are functionally relevant for schizophrenia. These included sets of genes expressed in different brain cell types and neurons, as well as PTV-intolerant (PI) genes, i.e. genes under stringent selection. As synaptic functions have previously been implicated in the aetiology of schizophrenia, we included gene sets based on synaptic processes and composition. Finally, we investigated gene sets made up of the intersection of PI genes and the brain expressed genes (i.e. PI genes that are expressed in the brain) as these are potentially smaller gene sets rich in genes related to the biology of schizophrenia and therefore of considerable value in follow-up analysis, if proven relevant. By synchronizing the functional analyses across common and rare variants, the current study attempts to answer two questions: 1) Do common and (ultra-)rare variant gene set and cell type enrichment analyses converge to similar results for schizophrenia and if so, 2) What gene sets are implicated across both (ultra-)rare and common variant analyses? The current analyses may shed light on whether common and (ultra-)rare variants reveal unique aspects of the aetiology of schizophrenia, or implicate the same pathways.

medRxiv preprint doi: https://doi.org/10.1101/2021.05.26.21257794; this version posted May 29, 2021. 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 4.0 International license .

METHODS

Data and sources

GWAS data

To identify gene sets enriched for common variants, we obtained summary data from large publicly available GWAS results of schizophrenia in individuals of European ancestry[11], as well as those of East Asian ancestry[9].

Exome sequencing data

To identify gene sets enriched for (ultra-)rare variants, we obtained genotype and phenotype data from the Swedish Schizophrenia Exome Sequencing Project[15], a case-control sample of 12380 unrelated Swedish individuals. Cases primarily had diagnoses of schizophrenia, although a small proportion of individuals were diagnosed with bipolar disorder. See data availability for more information on this data set.

Gene sets

Protein-truncating variant (PTV)-intolerant (PI) genes were obtained from the Genome Aggregation Database (gnomAD), and ascertained using the probability of loss-of-function intolerance (pLI) metric. We selected genes with pLI > 0.9, producing a list of 3063 genes[20].

Human brain cell gene sets were based on single-nucleus RNA-sequence (snRNA-seq) data generated on the Genotype-Tissue Expression (GTEx) project brain tissues[21]. We included a total of 14 cell types as ascertained in the study referenced. Sources and processing of expression data are described there. Excluding sporadic genes and genes with low expression, for the 14 cell types we selected the top 1600 (roughly 15%) differentially expressed genes in each cell type, which likely cover all, or most, genes that have a vital function in a specific cell type.

medRxiv preprint doi: https://doi.org/10.1101/2021.05.26.21257794; this version posted May 29, 2021. 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 4.0 International license .

Mouse brain cell gene sets were based on data obtained from a previous study[19]. Extensive description of sources and processing of expression data are described there. In that study, cells were assigned to level 1 classification, with subtypes of level 1 assigned as level 2 on the basis of single-cell RNA-sequence (scRNA-seq) data and clustering analyses. We focused our analyses on the 24 level 1 gene sets. As the scRNA-seq data were from mouse brains, we mapped the gene homologs using the human-mouse homolog reference from Mouse Genome Informatics. Similar to the human brain gene sets, we selected the top 1600 differentially expressed genes in each cell type.

Synaptic gene sets were selected based on synaptic gene ontology from the SynGo database[22], including gene sets defined by cellular component, i.e. the location in which the genes are active, or by biological process, i.e. the synaptic processes/functions they influence. In order to ensure that gene sets were powered enough to detect significant effects, we selected only gene sets containing 50 or more genes, resulting in a total of 35 gene sets.

PI x brain cell gene sets contained the intersection genes that are PTV-intolerant and are present in each human and mouse brain cell gene set. Although the PI x brain cell type are smaller than either the PI-gene set or the brain cell type specific gene set, they are potentially more strongly enriched given the genes are (1) relevant to brain function and (2) under strong negative selection.

In total 112 different gene sets were included in the analyses. All genes included in each gene set are available in Supplementary Table 1.

Common variant analyses

Common variant enrichment was evaluated using competitive analyses in MAGMA (v1.08b)[23]. MAGMA is a commonly used program for gene and gene set analysis that applies the principles of linear regression. It works by first computing gene-level associations in which *p*-values for individuals SNPs around a gene are averaged, while taking linkage disequilibrium (LD) structure into account. LD is estimated using ancestry-appropriate reference panels for each population investigated. We used

medRxiv preprint doi: https://doi.org/10.1101/2021.05.26.21257794; this version posted May 29, 2021. 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 4.0 International license .

the European panel of 1000 Genomes Project phase 3 for LD estimation in the GWAS based on individuals of European ancestry, and the East Asian panel in the GWAS based on individuals of East Asian ancestry. Subsequently, gene-based *p*-values were converted to z scores to test associations between each gene set and schizophrenia diagnosis. For each GWAS summary dataset, we excluded SNPs with INFO <0.8, as well as duplicate SNPs. Gene location information with start and stop sites were obtained from the MAGMA website, with no windows specified around the genes.

Rare variant analyses

Analyses of exome sequenced data including QC was mainly performed using Hail 0.2 (Hail Team. Hail 0.2.62-84fa81b9ea3d. https://github.com/hail-is/hail/commit/84fa81b9ea3d). QC generally followed those described in a previous study[16], as well as those detailed here: https://astheeggeggs.github.io/BipEx/index.html. Full description, including variant annotation and (ultra-)rare variant definitions are provided in the Supplementary text.

Gene set (association) analyses

We assessed (ultra-)rare variant enrichment in each gene set using logistic regression, testing the association between the burden of (ultra-)rare variants, and schizophrenia diagnosis. Sex and 10 genetic PCs were included as covariates in each set of analyses. We excluded individuals who were more than 4 median absolute deviations from the study specific median number of synonymous (ultra-)rare variants. The significance threshold was set to 5% false discovery rate in analyses across all gene sets per variant allele frequency.

We deviated from the pre-registration outlined by omitting parallel analyses for autism spectrum disorder. As analyses of autism exome sequence data unexpectedly required considerably more processing than the schizophrenia dataset, publication of the current valuable results for schizophrenia would be unreasonably delayed.

medRxiv preprint doi: https://doi.org/10.1101/2021.05.26.21257794; this version posted May 29, 2021. 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 4.0 International license .

RESULTS

Common variants

We used MAGMA to evaluate the enrichment of common schizophrenia-associated variants in PI genes, as well as genes expressed in brain cells and synapses. In the European sample, schizophrenia-associated common variants were strongly enriched in PI genes, as well as genes highly expressed in human brain cells. Enrichment in genes expressed in mouse brain cells was also present across the different cell types, but not to the same extent as the human brain cells. Across brain cell types, we observed higher association betas at the intersection of PI and brain cell expressed genes suggesting stronger enrichment, although confidence intervals were largely overlapping (Figure 1). Wider standard errors in these analyses likely reflect the loss of power from selecting only intersecting genes. Across the human brain cell types analysed, enrichment was strongest for genes expressed in excitatory neurons from the hippocampus (pyramidal and granule) and prefrontal cortex, as well as GABAergic interneurons and oligodendrocyte cells. In mouse brain cells enrichment was strongest in pyramidal cortical and hippocampal neurons, as well as medium spiny neurons (a type of GABAergic neuron). Again we observed more significantly enriched gene sets in the PI and mouse brain interaction gene sets, additionally implicating dopaminergic, and serotonergic neurons as well as interneurons. Finally, we observed significant enrichment in genes associated with synaptic processes. Common variants were most enriched in postsynaptic cellular components and biological processes. The standard errors for these associations were wide, likely due to the lower number of genes per gene set (Supplementary Figure 1).

In analyses of individuals of East-Asian ancestry we observed a similar pattern of results as the in the analyses of European individuals, in that there was stronger enrichment at the intersection of PI and brain cell expressed genes. However, we only observed significant enrichment in four gene sets, including the PI gene set (Supplementary Figures 2 and 3). This may have been related to power as this GWAS had a smaller sample size.

medRxiv preprint doi: https://doi.org/10.1101/2021.05.26.21257794; this version posted May 29, 2021. 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 4.0 International license .

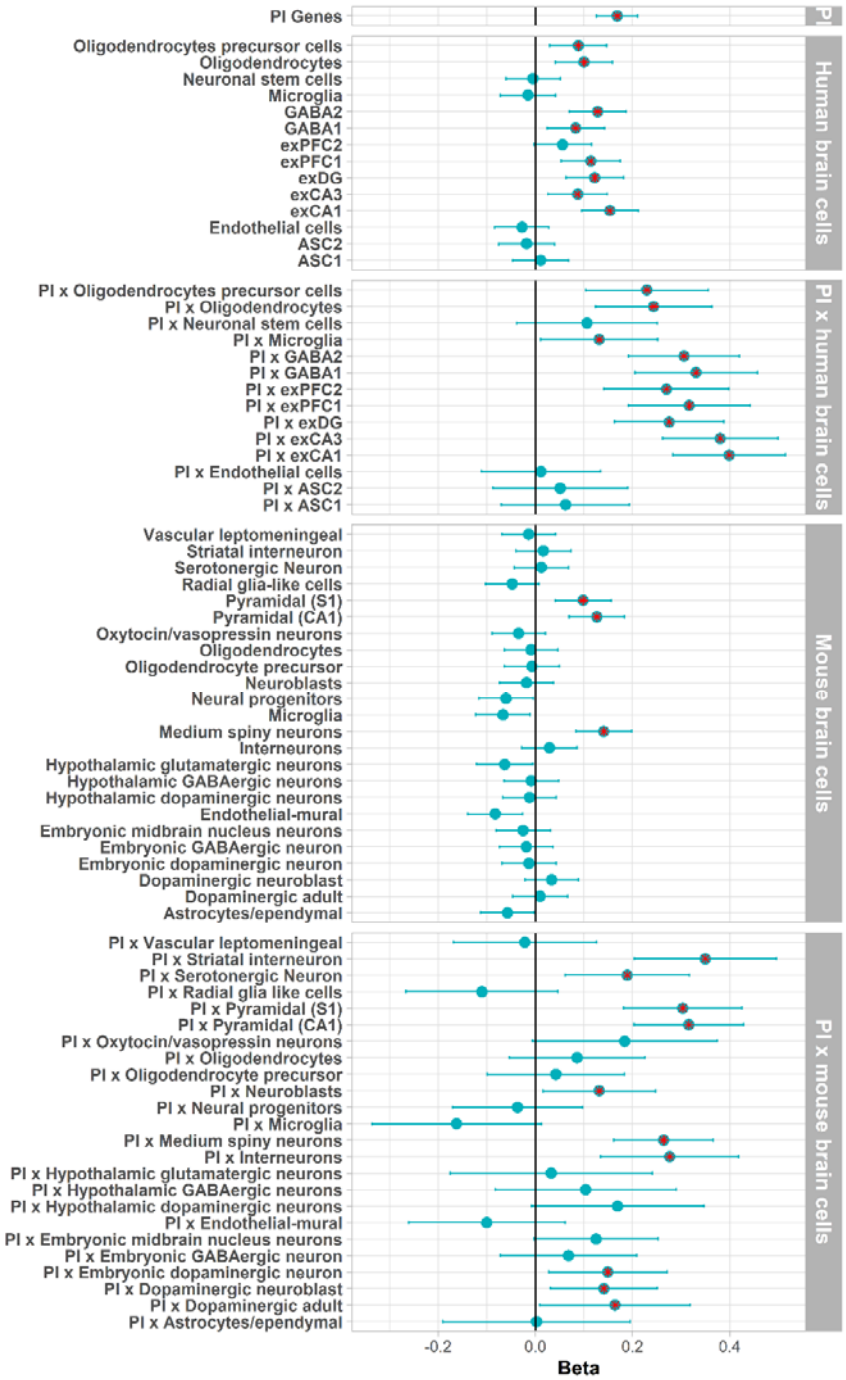

Figure 1. **Common variant enrichment in PI and brain cell gene sets**. Red stars denote significant gene sets after multiple testing correction. ASC=astrocytes, exCA1/exCA3=pyramidal neurons from the Hippocampal *Cornu Ammonis* regions, exDG=granule neurons from the Hippocampal dentate gyrus region, exPFC1/exPFC2=pyramidal neurons from the prefrontal cortex, GABA1/GABA2=GABAergic interneurons. Although not included in the figure, the synaptic gene sets were included in multiple testing correction.

medRxiv preprint doi: https://doi.org/10.1101/2021.05.26.21257794; this version posted May 29, 2021. 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 4.0 International license .

(Ultra-)rare variants

After QC, 10592 individuals (4623 cases and 5969 controls) were included in the ultra-rare variants analyses, while 10553 individuals (4603 cases and 5950 controls) were included in the rare variant analyses. The sample sizes were different across both sets of analyses due to the exclusion of individuals who were more than 4 median absolute deviations from the study specific median number of synonymous variants. We tested the association between the burden of rare (AF < 0.1%) or ultra-rare PTVs and schizophrenia diagnosis. Ultra-rare variants were those observed in 1 out of 188,023 individuals (our sample + gnomAD + DiscovEHR). Schizophrenia cases had a significantly higher burden of both ultra-rare ($\beta = 0.082$, SE = 0.017, $P = 2.79 \times 10^{-6}$) and rare ($\beta = 0.026$, SE = 0.007, $P = 0.0004$) PTVs compared to controls. As a negative control, burden scores for synonymous variants were computed, which were not significantly different between cases and controls for ultra-rare ($\beta = 0.013$, SE = 0.009, $P = 0.151$) or rare ($\beta = 0.0008$, SE = 0.002, $P = 0.718$) variants.

Next, we assessed (ultra-)rare variant enrichment in each gene set using logistic regression, testing the association between the burden of (ultra-)rare variants and schizophrenia diagnosis. Both rare and ultra-rare variants were significantly enriched in PI genes. Overall, there was greater enrichment of ultra-rare PTVs compared to rare PTVs across gene sets analysed. Similar to common variants, enrichment of rare and ultra-rare PTVs was greater at the intersection of PI genes and genes expressed in mouse and human brain cells (Figure 2), though in this case there was more enrichment of genes expressed in mouse brain cells compared to human brain cells. Across the brain cell types, enrichment was again strongest for genes expressed in excitatory neurons from the hippocampus and prefrontal cortex, as well as GABAergic interneurons, oligodendrocyte cells, and medium spiny neurons. We also observed enrichment in synaptic genes (Supplementary Figure 4). Complete regression results for PTVs and synonymous variants are described in Supplementary Tables 2 – 5.

medRxiv preprint doi: https://doi.org/10.1101/2021.05.26.21257794; this version posted May 29, 2021. 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 4.0 International license .

Figure 2. **(Ultra-)rare variant enrichment in PI and brain cell gene sets**. Red stars denote significant gene sets after multiple testing correction. ASC=astrocytes, exCA1/exCA3=pyramidal neurons from the Hippocampal *Cornu Ammonis* regions, exDG=granule neurons from the Hippocampal dentate gyrus region, exPFC1/exPFC2=pyramidal neurons from the prefrontal cortex, GABA1/GABA2=GABAergic interneurons. Although not included in the figure, the synaptic gene sets were included in multiple testing correction.

medRxiv preprint doi: https://doi.org/10.1101/2021.05.26.21257794; this version posted May 29, 2021. 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 4.0 International license .

Convergence of enrichment across allele frequency spectrum

We investigated whether common and (ultra-)rare variants converged to similar results by rank correlating the association betas across the three allele frequencies for each gene set. We also evaluated overlapping gene set enrichment across the allele frequency spectrum.

We observed moderate to high correlations of effect sizes across the three pairs of comparisons. Correlation estimates ranged from 0.515 to 0.740, and from 0.561 to 0.802 when weighted by the inverse of standard errors from each association estimate (weighting precise estimates more heavily than imprecise estimates when comparing across common and rare enrichment results) (Figures 3 – 5). Additionally, 16 gene sets were significantly enriched across all three allele frequencies, implicating PI genes, medium spiny neurons, and pyramidal neurons from the mouse brain, as well as multiple intersection gene sets, across the allele frequency spectrum. 28 gene sets were significantly enriched across common and ultra-rare variants, 17 across rare and ultra-rare variants, and 16 across common and rare variants.

medRxiv preprint doi: https://doi.org/10.1101/2021.05.26.21257794; this version posted May 29, 2021. 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 4.0 International license .

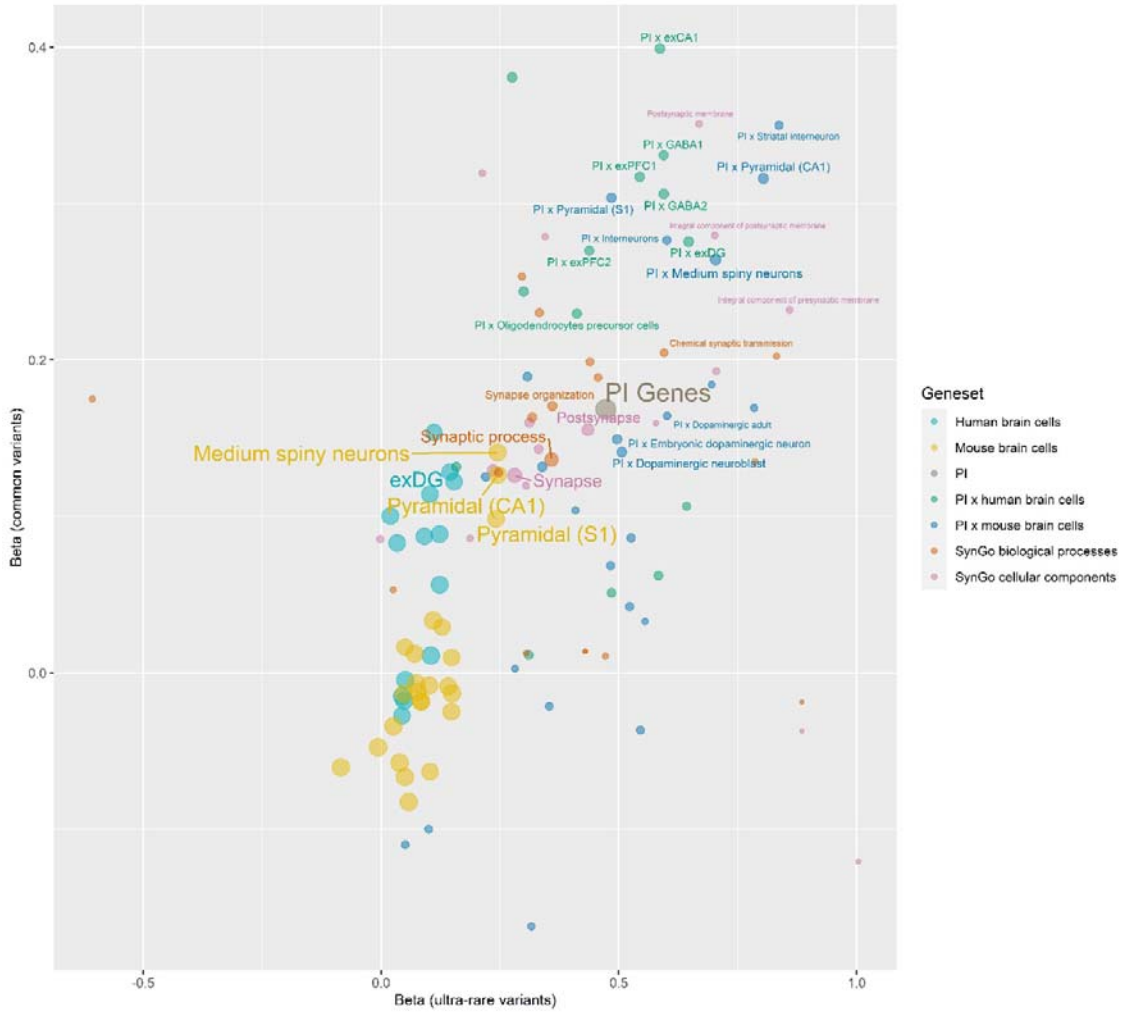

Figure 3. **Correlation between gene set enrichment in common vs ultra-rare variants**. Point sizes represent the weight assigned to each correlation estimate, obtained by calculating the inverse of the product of both standard errors. Correlation estimate is 0.537, while weighted correlation is 0.708. Labelled gene sets are significantly enriched across both common and rare variants

medRxiv preprint doi: https://doi.org/10.1101/2021.05.26.21257794; this version posted May 29, 2021. 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 4.0 International license .

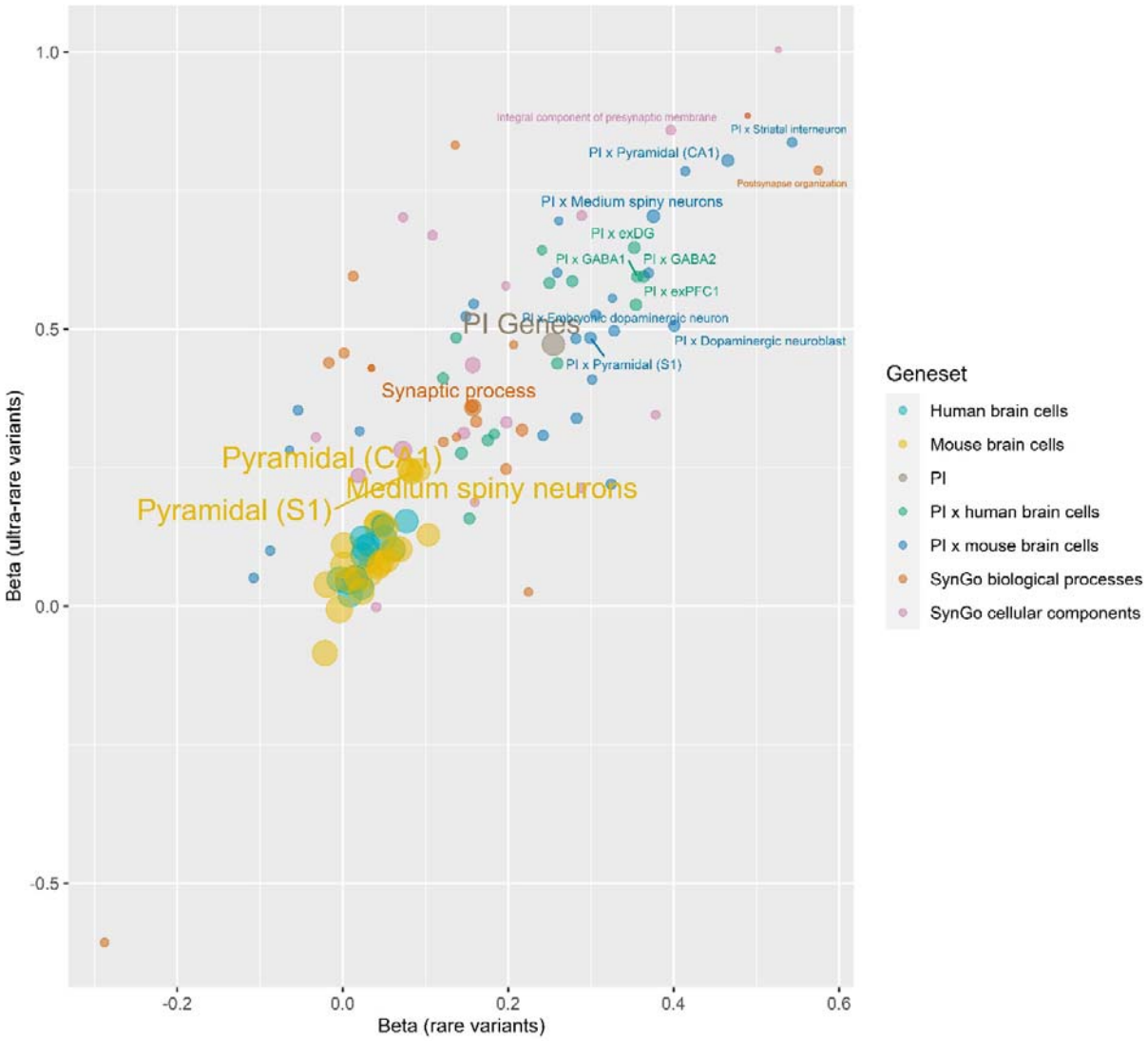

Figure 4. **Correlation between gene set enrichment in rare vs ultra-rare variants**. Point sizes represent the weight assigned to each correlation estimate, obtained by calculating the inverse of the product of both standard errors. Correlation estimate is 0.740, while weighted correlation is 0.802. Labelled gene sets are significantly enriched across both common and rare variants

medRxiv preprint doi: https://doi.org/10.1101/2021.05.26.21257794; this version posted May 29, 2021. 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 4.0 International license .

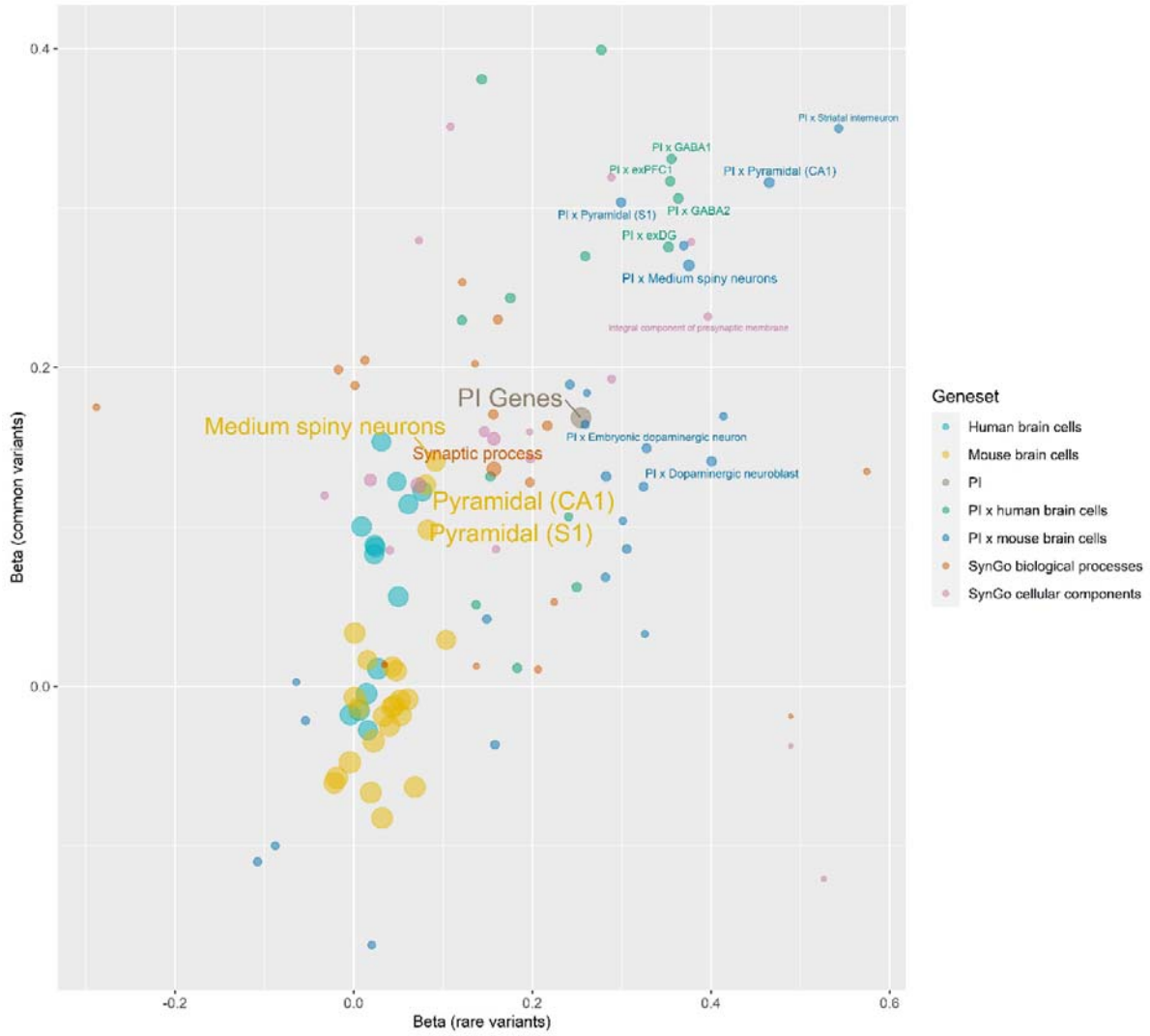

Figure 5. **Correlation between gene set enrichment in common vs rare variants**. Point sizes represent the weight assigned to each correlation estimate, obtained by calculating the inverse of the product of both standard errors. Correlation estimate is 0.515, while weighted correlation is 0.561. Labelled gene sets are significantly enriched across both common and rare variants

medRxiv preprint doi: https://doi.org/10.1101/2021.05.26.21257794; this version posted May 29, 2021. 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 4.0 International license .

DISCUSSION

We used GWAS and whole exome analyses of schizophrenia to investigate whether common and (ultra-)rare PTV enrichment converge to similar results in terms of what gene sets are implicated across analyses. We observed partial convergence across the gene sets significantly enriched for common and (ultra-)rare variants, in that multiple gene sets were significantly enriched across all three variant classes. Enrichment analyses implicated mainly excitatory neurons from the prefrontal cortex and hippocampus, medium spiny neurons, and GABAergic neurons, as well as PI genes, synaptic components, and processes. Moreover, across all three allele frequencies, enrichment was stronger in the gene sets containing the intersection of brain cell types and PI genes compared to the brain cell gene sets.

Brain cell enrichment findings are consistent with findings from previous analyses of both common and (ultra-)rare variants associated with schizophrenia[4, 6, 18, 19]. Overlapping significant enrichment between ultra-rare and common variants provides additional evidence of some convergence in genes and biological mechanisms implicated by genetic variants across the allelic spectrum. Recent analyses showed significant enrichment of ultra-rare variants in genes implicated by schizophrenia GWAS, and that two genes implicated in rare variant analyses also showed associations in the schizophrenia GWAS[4]. Additionally, we showed that genes likely to be under stringent selection (PI genes) are implicated in both common and (ultra-)rare variants, while stronger enrichment in the intersection of brain cell types and PI genes suggests that PI genes are generally important, but even more so in these cell types. These gene sets, particularly the intersection gene sets, potentially provide a manageable set of genes and biological processes to target for follow-up analyses.

An important factor to consider in light of our and similar findings, is how much progress the results represent with regards to disease biology. Analyses of common variant cell-type enrichment support current distinctions between neurological disorders like Parkinson's disease and Alzheimer's disease, versus psychiatric disorders, as they have shown different association patterns. Parkinson's disease

medRxiv preprint doi: https://doi.org/10.1101/2021.05.26.21257794; this version posted May 29, 2021. 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 4.0 International license .

has implicated cholinergic and monoaminergic neurons, Alzheimer's disease has implicated microglial cells, while psychiatric disorders like schizophrenia have implicated excitatory neurons[18, 24, 25]. Findings across psychiatric disorders are less clear. Analyses of individual psychiatric disorders suggest similar patterns of cell-type associations for disorders including schizophrenia, bipolar disorder, MDD, and anorexia nervosa, although enrichment was generally strongest for schizophrenia[18, 26]. However, analyses of psychiatric disorder factors have shown that common variant enrichment in excitatory and GABAergic genes from human brain cells is limited to a psychotic disorder factor (comprised of schizophrenia and bipolar disorder), with no enrichment observed in compulsive (anorexia nervosa, obsessive compulsive disorder, Tourette syndrome), neurodevelopmental (attention deficit/hyperactivity disorder, autism spectrum disorder, post-traumatic stress disorder, and MDD), or internalizing factors (post-traumatic stress disorder, MDD and anxiety disorders)[27]. Importantly, similar neuronal enrichment has been observed for non-psychiatric cognitive traits including intelligence, educational attainment, neuroticism, and body mass index (BMI), which show modest but robust genetic correlations with psychiatric disorders including schizophrenia[18, 28-30].

These findings suggest that at the current resolution of analyses (expression differences between all cell types), common variant enrichment in genes predominantly expressed in neurons is non-specific and pervasive across various behavioural traits, although the lack of findings for most psychiatric disorders may also point to differences in statistical power across the GWASs. Future analyses, for example based on more comprehensive single-cell sequencing of all neuron subtypes, could identify genes that are specific to certain developmental stages in order to seek out cell types, cell functions or developmental phases that are specific to schizophrenia. Such analyses using whole-brain developmental expression profiles have shown enrichment of schizophrenia-associated common variants in the prefrontal cortex during early midfetal development[31]. Limits on data availability currently make it difficult to investigate whether these similarities across various traits in common variant cell type enrichment translate to (ultra-)rare variant enrichment, and is a significant avenue

medRxiv preprint doi: https://doi.org/10.1101/2021.05.26.21257794; this version posted May 29, 2021. 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 4.0 International license .

for future research, although brain-expressed genes have also been found to be enriched for ultra-rare variants associated with educational attainment[32]. This is vital as the importance/contribution of variant classes along the allelic spectrum may vary depending on phenotype.

Our study, as well as other results[4, 6], further suggest that polygenicity, where very many genetic loci are implicated in a disorder like schizophrenia, complicates the search for individual risk loci in both common and (ultra-)rare variant analyses. As gene set analyses are strongly correlated between GWAS and (ultra-)rare variant analysis, it is not unlikely that analyses at the level of the gene show similar correlations between common and (ultra)rare variants. Strong correlations at the level of the gene may call for meta-analysis across GWAS and rare variant studies, or more subtle information integration that accounts for both the weight of evidence, as well as the regions/features/functions of the gene that are influenced by rare or common variation.

Our findings should be considered in light of some limitations. Firstly, our findings are limited by the current definitions of ultra-rare variants, and results might be subject to change if definitions change as more data becomes available. Secondly, our analyses are likely affected by the size of gene sets. Larger gene sets likely have more power to detect effects compared to the smaller gene sets. This was particularly evident in the interaction and synaptic gene sets which had wide confidence intervals. Finally, our analyses were limited to individuals of European ancestry, making it not generalizable to individuals of other ancestry. Overall similar analyses typically contain limited non-European samples, with non-European samples making up 20% of the most recent schizophrenia GWAS[6] and 26% of the most recent WES-based schizophrenia analyses[4].

In conclusion, we show that there is at least partial overlap in the genes disrupted by both common and (ultra-)rare variants associated with schizophrenia suggesting involvement of the same biological mechanisms. Genes influencing neuronal processes as well as genes likely to be under stringent selection are implicated in schizophrenia aetiology across common and (ultra-)rare

medRxiv preprint doi: https://doi.org/10.1101/2021.05.26.21257794; this version posted May 29, 2021. 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 4.0 International license .

variants. Future studies integrating information across the allele frequency spectrum might prove useful in furthering our understanding of schizophrenia aetiology.

medRxiv preprint doi: https://doi.org/10.1101/2021.05.26.21257794; this version posted May 29, 2021. 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 4.0 International license .

Acknowledgements

This project has received funding from the European Union's Horizon 2020 research and innovation programme, Marie Sklodowska Curie Actions – MSCA-ITN-2016 – Innovative Training Networks under grant agreement No [721567].

W.A.A. is supported by the European Union's Horizon 2020 research and innovation programme under the Marie Sklodowska-Curie grant agreement no. 721567, and the National Institute Of Mental Health of the National Institutes of Health under Award Number R01MH120219.

A.R.H. is supported by the Children's Hospital Foundation and University of Queensland strategic funding.

M.B. is funded by an ERC Consolidator Grant (WELL-BEING 771057).

M.G.N is supported by the National Institute Of Mental Health of the National Institutes of Health under Award Number R01MH120219, supported by ZonMw grant: 'Genetics as a research tool: a natural experiment to elucidate the causal effects of social mobility on health' (pnr: 531003014) and ZonMw project: 'Can sex- and gender-specific gene expression and epigenetics explain sex-differences in disease prevalence and etiology?' (pnr: 849200011) from The Netherlands Organisation for Health Research and Development, a VENI grant awarded by The Dutch Research Council (NWO) (VI.Veni.191G.030), and is a Jacobs Foundation Fellow.

C.M.M. has received funding from the European Union's Horizon 2020 research and innovation programme under the Marie Sklodowska-Curie grant agreement no. 721567.

Conflicts of interest

The authors declare no conflict of interest.

medRxiv preprint doi: https://doi.org/10.1101/2021.05.26.21257794; this version posted May 29, 2021. 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 4.0 International license .

REFERENCES

1. Laursen TM, Nordentoft M, Mortensen PB. Excess Early Mortality in Schizophrenia. *Annual Review of Clinical Psychology* 2014; **10**(1)**:** 425-448.

2. Power RA, Kyaga S, Uher R, MacCabe JH, Långström N, Landen M *et al.* Fecundity of Patients With Schizophrenia, Autism, Bipolar Disorder, Depression, Anorexia Nervosa, or Substance Abuse vs Their Unaffected Siblings. *JAMA Psychiatry* 2013; **70**(1)**:** 22-30.

3. Avramopoulos D. Recent Advances in the Genetics of Schizophrenia. *Complex Psychiatry* 2018; **4**(1)**:** 35-51.

4. Singh T, Poterba T, Curtis D, Akil H, Al Eissa M, Barchas JD *et al.* Exome sequencing identifies rare coding variants in 10 genes which confer substantial risk for schizophrenia. *medRxiv* 2020.

5. McGrath J, Saha S, Chant D, Welham J. Schizophrenia: A Concise Overview of Incidence, Prevalence, and Mortality. *Epidemiologic Reviews* 2008; **30**(1)**:** 67-76.

6. Ripke S, Walters JT, O'Donovan MC. Mapping genomic loci prioritises genes and implicates synaptic biology in schizophrenia. *medRxiv* 2020**:** 2020.2009.2012.20192922.

7. Gejman PV, Sanders AR, Duan J. The role of genetics in the etiology of schizophrenia. *Psychiatr Clin North Am* 2010; **33**(1)**:** 35-66.

8. Uher R. The role of genetic variation in the causation of mental illness: an evolution-informed framework. *Molecular Psychiatry* 2009; **14**(12)**:** 1072-1082.

9. Lam M, Chen C-Y, Li Z, Martin AR, Bryois J, Ma X *et al.* Comparative genetic architectures of schizophrenia in East Asian and European populations. *Nature Genetics* 2019; **51**(12)**:** 1670-1678.

10. O'Connor LJ, Schoech AP, Hormozdiari F, Gazal S, Patterson N, Price AL. Extreme Polygenicity of Complex Traits Is Explained by Negative Selection. *The American Journal of Human Genetics* 2019; **105**(3)**:** 456-476.

11. Pardiñas AF, Holmans P, Pocklington AJ, Escott-Price V, Ripke S, Carrera N *et al.* Common schizophrenia alleles are enriched in mutation-intolerant genes and in regions under strong background selection. *Nature genetics* 2018; **50**(3)**:** 381-389.

12. Liu C, Everall I, Pantelis C, Bousman C. Interrogating the Evolutionary Paradox of Schizophrenia: A Novel Framework and Evidence Supporting Recent Negative Selection of Schizophrenia Risk Alleles. *Frontiers in Genetics* 2019; **10**(389).

medRxiv preprint doi: https://doi.org/10.1101/2021.05.26.21257794; this version posted May 29, 2021. 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 4.0 International license .

13. Sekar A, Bialas AR, de Rivera H, Davis A, Hammond TR, Kamitaki N *et al.* Schizophrenia risk from complex variation of complement component 4. *Nature* 2016; **530**(7589): 177-183.

14. Mealer RG, Jenkins BG, Chen C-Y, Daly MJ, Ge T, Lehoux S *et al.* The schizophrenia risk locus in SLC39A8 alters brain metal transport and plasma glycosylation. *Scientific Reports* 2020; **10**(1): 13162.

15. Genovese G, Fromer M, Stahl EA, Ruderfer DM, Chambert K, Landén M *et al.* Increased burden of ultra-rare protein-altering variants among 4,877 individuals with schizophrenia. *Nature neuroscience* 2016; **19:** 1433.

16. Ganna A, Satterstrom FK, Zekavat SM, Das I, Kurki MI, Churchhouse C *et al.* Quantifying the Impact of Rare and Ultra-rare Coding Variation across the Phenotypic Spectrum. *American journal of human genetics* 2018; **102**(6): 1204-1211.

17. Sanders SJ, Neale BM, Huang H, Werling DM, An J-Y, Dong S *et al.* Whole genome sequencing in psychiatric disorders: the WGSPD consortium. *Nature neuroscience* 2017; **20**(12): 1661-1668.

18. Bryois J, Skene NG, Hansen TF, Kogelman LJA, Watson HJ, Liu Z *et al.* Genetic identification of cell types underlying brain complex traits yields insights into the etiology of Parkinson's disease. *Nature Genetics* 2020; **52**(5): 482-493.

19. Skene NG, Bryois J, Bakken TE, Breen G, Crowley JJ, Gaspar HA *et al.* Genetic identification of brain cell types underlying schizophrenia. *Nature Genetics* 2018; **50**(6): 825-833.

20. Karczewski KJ, Francioli LC, Tiao G, Cummings BB, Alföldi J, Wang Q *et al.* The mutational constraint spectrum quantified from variation in 141,456 humans. *Nature* 2020; **581**(7809): 434-443.

21. Habib N, Avraham-Davidi I, Basu A, Burks T, Shekhar K, Hofree M *et al.* Massively parallel single-nucleus RNA-seq with DroNc-seq. *Nature Methods* 2017; **14**(10): 955-958.

22. Koopmans F, van Nierop P, Andres-Alonso M, Byrnes A, Cijsouw T, Coba MP *et al.* SynGO: An Evidence-Based, Expert-Curated Knowledge Base for the Synapse. *Neuron* 2019; **103**(2): 217-234.e214.

23. de Leeuw CA, Mooij JM, Heskes T, Posthuma D. MAGMA: generalized gene-set analysis of GWAS data. *PLoS Comput Biol* 2015; **11**(4): e1004219.

24. Efthymiou AG, Goate AM. Late onset Alzheimer's disease genetics implicates microglial pathways in disease risk. *Mol Neurodegener* 2017; **12**(1): 43-43.

medRxiv preprint doi: https://doi.org/10.1101/2021.05.26.21257794; this version posted May 29, 2021. 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 4.0 International license .

25. Wightman DP, Jansen IE, Savage JE, Shadrin AA, Bahrami S, Rongve A *et al.* Largest GWAS (N=1,126,563) of Alzheimer's Disease Implicates Microglia and Immune Cells. *medRxiv* 2020**:** 2020.2011.2020.20235275.

26. Watson HJ, Yilmaz Z, Thornton LM, Hübel C, Coleman JRI, Gaspar HA *et al.* Genome-wide association study identifies eight risk loci and implicates metabo-psychiatric origins for anorexia nervosa. *Nature Genetics* 2019; **51**(8)**:** 1207-1214.

27. Grotzinger AD, Mallard TT, Akingbuwa WA, Ip HF, Adams MJ, Lewis CM *et al.* Genetic Architecture of 11 Major Psychiatric Disorders at Biobehavioral, Functional Genomic, and Molecular Genetic Levels of Analysis. *medRxiv* 2020**:** 2020.2009.2022.20196089.

28. Coleman JRI, Bryois J, Gaspar HA, Jansen PR, Savage JE, Skene N *et al.* Biological annotation of genetic loci associated with intelligence in a meta-analysis of 87,740 individuals. *Molecular Psychiatry* 2019; **24**(2)**:** 182-197.

29. Demange PA, Malanchini M, Mallard TT, Biroli P, Cox SR, Grotzinger AD *et al.* Investigating the genetic architecture of noncognitive skills using GWAS-by-subtraction. *Nature Genetics* 2021; **53**(1)**:** 35-44.

30. Baselmans BM, Jansen R, Ip HF, van Dongen J, Abdellaoui A, van de Weijer MP *et al.* Multivariate genome-wide analyses of the well-being spectrum. *Nature genetics* 2019; **51**(3)**:** 445.

31. Clifton NE, Hannon E, Harwood JC, Di Florio A, Thomas KL, Holmans PA *et al.* Dynamic expression of genes associated with schizophrenia and bipolar disorder across development. *Transl Psychiatry* 2019; **9**(1)**:** 74-74.

32. Ganna A, Genovese G, Howrigan DP, Byrnes A, Kurki M, Zekavat SM *et al.* Ultra-rare disruptive and damaging mutations influence educational attainment in the general population. *Nat Neurosci* 2016; **19**(12)**:** 1563-1565.

medRxiv preprint doi: https://doi.org/10.1101/2021.05.26.21257794; this version posted May 29, 2021. 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 4.0 International license .

DATA AVAILABILITY

- PI genes: https://storage.googleapis.com/gnomad-public/release/2.1.1/constraint/gnomad.v2.1.1.lof_metrics.by_gene.txt.bgz
- Human brain single nuclei RNA-Seq data: https://gtexportal.org/home/datasets
- Nuclei to cell type mapping: https://static-content.springer.com/esm/art%3A10.1038%2Fnmeth.4407/MediaObjects/41592_2017_BFnmeth4407_MOESM10_ESM.xlsx
- Mouse brain single cell RNA-Seq data: https://www.nature.com/articles/s41588-018-0129-5#Sec33
- Human-mouse homolog reference: http://www.informatics.jax.org/downloads/reports/HOM_MouseHumanSequence.rpt
- SynGo ontologies and annotations: https://www.syngoportal.org/
- Swedish Schizophrenia Exome Sequencing Project (**dbGaP Study Accession:** phs000473.v2.p2)