Rare coding variants in schizophrenia-associated genes affect generalised cognition in the UK Biobank

Author list and affiliations:
Eilidh Fenner[1], Peter Holmans[1], Michael C O'Donovan[1], Michael J Owen[1], James T R Walters[1+], Elliott Rees[1+]

1. Centre for Neuropsychiatric Genetics and Genomics, Division of Psychological Medicine and Clinical Neurosciences, Cardiff University, Cardiff, UK

[+] these authors contributed equally to this work.

Corresponding authors:
Dr Elliott Rees
Hadyn Ellis Building, Maindy Road, Cardiff, CF24 4HQ
Phone: +44 (0)29 2068 8375
Email: reeseg@cardiff.ac.uk

Professor James T R Walters
Hadyn Ellis Building, Maindy Road, Cardiff, CF24 4HQ
Phone: +44 29206 88434
Email: WaltersJT@cardiff.ac.uk

Abstract

Impairments in cognitive function are a feature of schizophrenia that strongly predict functional outcome and are generally not improved by current medications. However, the nature of the relationship between cognitive impairment and schizophrenia risk, and particularly the extent to which this reflects shared underlying biology, remains uncertain. We analysed exome-sequencing data from the UK Biobank to test for association between generalised cognition and damaging rare coding variation in genes and loci associated with schizophrenia in 30,487 people without the disorder. Rare protein-truncating variants (PTVs) and damaging missense variants in loss-of-function intolerant (LoFi) genes were associated with lower generalised cognition. Moreover, we found significantly stronger effects for damaging missense variants in credible causal genes at schizophrenia GWAS loci and for rare PTVs affecting LoFi genes in regions defined by schizophrenia-enriched CNVs. This suggests shared underlying biology between schizophrenia risk and general cognitive function in the population, and that exploiting large population sequencing datasets to identify genes with shared effects on cognition and schizophrenia can provide a route towards determining biological processes underlying cognitive impairment in schizophrenia.

Introduction

Schizophrenia is a severe psychiatric disorder, with a lifetime risk of around 1%[1] which shows substantial clinical heterogeneity. Impaired cognitive function is an important feature of schizophrenia that often precedes the onset of psychosis[2–4] and is a strong predictor of functional outcome[5,6]. Relative to controls, those with schizophrenia demonstrate impairments across multiple cognitive domains, suggesting the disorder is associated with a generalised, rather than a domain specific, cognitive impairment. Indeed, generalised cognition, or *g*, explains the majority of variance in cognition between people with schizophrenia and healthy controls and is also a better predictor of functional outcome in schizophrenia compared with any individual cognitive test[7,8].

Schizophrenia is highly heritable and polygenic, with liability conferred by rare and common alleles in many genes, particularly those under selective constraint against mutations predicted to result in a loss of protein function, also known as loss-of-function intolerant (LoFi) genes[9–11]. The largest genome-wide association study (GWAS) of common alleles in schizophrenia to date identified associations at 287 distinct loci[12], and among these associations fine-mapping approaches prioritised 106 protein-coding genes as credibly causal. The genome-wide burden of rare copy number variants (CNVs) is also increased in schizophrenia cases compared with controls, with a set of 63 CNVs reported to be associated with developmental disorders showing particular enrichment in cases[13,14]. Moreover, 12 specific CNVs have been robustly associated with schizophrenia liability[13,14]. Finally, sequencing studies have identified 12 genes that are individually enriched in schizophrenia for damaging types of rare coding variants (RCVs)[15,16]. There is also evidence for a convergence between genes implicated in schizophrenia by common and rare alleles[12,15].

Genetic liability for schizophrenia is pleiotropic, with shared effects on liability to several other psychiatric and developmental disorders as well as on variation in cognitive function in the general population. For example, the 12 CNVs associated with schizophrenia are all risk factors for developmental disorders[13], and individuals without a psychiatric or developmental disorder who have one of these CNVs perform worse on a range of cognitive tests compared to those who do not have such CNVs[17,18]. Similarly, damaging RCVs in developmental disorder associated genes, some of which have shared effects for schizophrenia[10], are associated with poorer cognition in individuals without a psychiatric or developmental disorder who have these variants[19,20]. Finally, a higher burden of schizophrenia common alleles is associated with lower cognitive function in the population[21].

Less is known about the effects of damaging RCVs in schizophrenia genes and loci on cognition in individuals without a psychiatric or developmental disorder. Recent studies using data from the UK Biobank (UKBB) have shown damaging RCVs in constrained genes are associated with longer reaction-time, lower fluid intelligence scores, and lower levels of educational attainment[19,22]. One study also reported that damaging RCVs in genes mapping to schizophrenia loci identified by GWAS are significantly associated with lower educational attainment, and also nominally significantly associated with longer reaction-time and lower fluid intelligence[19]. However, the study did not control for the effects of these mutations in LoFi or brain-expressed genes, or specifically test the small subset of the genes within those loci that have been reported to be credibly causal. Thus, it is unclear whether the results of that study are relevant to RCVs in schizophrenia-associated genes, or if they simply reflect the known enrichment of LoFi genes, or brain-expressed genes, within the schizophrenia-associated common allele loci[11]. Furthermore, there have been no studies examining the relationship between cognition in individuals without a psychiatric or developmental disorder

and RCVs in sets of genes implicated in schizophrenia by rare CNVs or RCVs. It is important to establish whether RCVs in schizophrenia genes and loci are associated with impaired cognition in those who do not have the disorder, as this would provide evidence that the association of impaired cognition with schizophrenia is, in part, explained by shared biology between schizophrenia risk and cognitive function in the population. Moreover, identifying specific genes with shared effects on schizophrenia and on cognition, and distinguishing them from those with specific effects, could inform our understanding of the biological processes underlying cognitive impairment in schizophrenia and lead to the development of stratified treatment approaches.

In the current study, we used exome sequencing data from 200,619 individuals in UKBB to examine whether RCVs in genes implicated in schizophrenia have effects on cognition in people without a psychiatric or developmental disorder. Whereas previous RCV studies of cognition in UKBB analysed individual tests of cognition[19], we analysed a measure of generalised cognition (*g*), which we formed from a principal component analysis of four cognitive tests. We focussed our study on *g*, instead of individual cognitive tests, as it is a more robust measure of general cognitive ability, and also because schizophrenia is strongly associated with a generalised rather than specific cognitive deficit[8,23,24]. Our findings demonstrate that rare protein-truncating variants (PTVs) and damaging missense variants in LoFi genes are associated with lower *g*, which to our knowledge has not been shown before. We also show that rare deleterious missense variants in credible causal genes underlying schizophrenia common allele loci are associated with lower *g*, and that these effects are significantly stronger than those observed for such variants in all genes mapping to the schizophrenia common allele loci, and critically, stronger than associations in LoFi genes or in brain-expressed genes in general. Additionally, rare PTVs in LoFi genes within

schizophrenia-enriched CNVs have significantly stronger effects on *g* than rare PTVs in LoFi genes in general. Our findings suggest that RCVs in schizophrenia-associated genes impact cognition in the general population, and that this is not confounded by the facts that schizophrenia-associated genes are enriched for LoFi genes and for genes expressed in brain. These findings suggest shared underlying biology between schizophrenia risk and general cognitive function in the population. They also validate the approaches used in the study by Trubetskoy et al[12] to prioritise credible causal genes.

Methods

UK Biobank

Between 2006 and 2010, UKBB recruited around 500,000 participants from the UK through NHS registers, with no exclusion criteria apart from reasonable access to an assessment centre. UKBB participants were genotyped using the Affymetrix Axiom UKBB array (~450,000 individuals) and the UK BiLEVE array (~50,000 individuals). Exome sequencing data released for 200,619 UKBB participants in October 2020 were used for the current study and were generated using the IDT xGen Exome Research Panel v1.0 exome capture kit and Illumina NovaSeq 6000 instruments; S2 flow cells were used for the initial 50k sequencing release, and S4 flow cells used for the following 150k samples. Further details on sequencing protocol can be found in the primary UKBB publication[25]. Participants were aged 40-69 at recruitment (mean age = 56.4) and 54.9% were female. All participants gave consent for their data to be used by UKBB projects and agreed to being followed up (Supplementary Materials 1.1). The research presented in this study has been conducted using the UKBB resource under Application Number 13310.

Whole exome sequencing data

Data processing and quality control

We downloaded and analysed a joint-genotyped variant call set in pVCF format, which was centrally generated by the UKBB with Deep Variant 0.010[26] using CRAMs that were processed using the 'OQFE protocol'[27–30], with reads aligned to GRCh38 reference genome[31]. For further details, see [32,33].

We performed genotype, sample, and variant quality control (QC) using Hail[34]. Genotypes were excluded if they met any of the following criteria: depth ≤ 10X; genotype quality ≤ 30; homozygous genotypes for the reference allele with an allele balance > 0.1; homozygous genotypes for the alternate allele with an allele balance < 0.9; heterozygous genotypes with an allele balance < 0.25 or > 0.75. Variant sites were then excluded if they met any of the following criteria: labelled as 'mono-allelic'[35]; call rate ≤ 0.9; overlapped a low-complexity region; site has > 6 alternative alleles (including indels and SNVs); mean genotype quality across all samples ≤ 40; no alternative alleles after applying genotype QC; and non-autosomal sites. In total, 2,530,021 from 15,922,704 variants failed QC (Supplementary Table 1).

Samples were excluded if they met any of the following criteria: sex predicted from genetic data did not match reported documented sex (n=104); sequencing call rate < 0.8 (n=88, Supplementary Figure 1); diagnoses of autism, schizophrenia, or intellectual disability from primary care data, hospital inpatient data, death register records or self-report data (n=503). No sample was excluded for low mean sequencing depth or mean genotype quality, most likely because this dataset had already received central QC prior to public release. To identify

related individuals, we used array data to estimate pair-wise kinship coefficients between all samples. The R package ukbtools[36] was used to remove 14,699 samples ensuring that no pairs of retained participants were third degree relations or closer, defined as a kinship co-efficient > 0.0442[37,38].

Sample ancestries were imputed by applying a principal component (PC) analysis to a pruned set of high-quality common variants, and then using the means and standard deviations of PCs 1 and 2 for each self-reported ethnic background category to define ancestry groups (Supplementary Materials 1.2 and Supplementary Figure 2). Based on PC-defined ancestry groups, 170,893 samples were of European ancestries; 4,453 samples were of South Asian ancestries; 3,264 samples were of African ancestries; and 1,060 samples were of East Asian ancestries. 5,555 participants who fell into two, or zero, PC-defined ancestry groups were excluded.

In total, 20,955 individuals failed sample QC, retaining 179,670 individuals for analysis.

Variant annotation

Variants were annotated in Hail using Ensembl's VEP[39]. Protein-truncating variants (PTVs) were defined as splice acceptor, splice donor, stop-gain or frameshift variants that were annotated as high confidence for causing loss of protein function by LoFTEE[40]. Deleterious missense variants were defined as missense variants with a REVEL[41] score > 0.75.

Schizophrenia-associated gene sets

Schizophrenia-associated protein-coding, autosomal genes were collated from GWAS[12], CNV[13] and RCV[15] studies (Supplementary Table 2).

Constrained genes: Genes under selective constraint for PTVs, known as LoFi genes, are enriched for damaging RCVs, rare CNVs, and common alleles in schizophrenia[9–11]. We analysed RCVs in 2,821 LoFi genes (defined as genes with a GnomAD pLI score[42] ≥ 0.9) and in the remaining LoF-tolerant genes (those with a GnomAD pLI score < 0.9). Exploratory analyses were performed using the GnomAD LoF observed/expected upper bound fraction (LOEUF) metric of constraint against PTVs[43].

Common allele loci: Genes implicated in schizophrenia by common alleles were taken from the largest schizophrenia GWAS[12]. This study identified 282 autosomal loci associated with schizophrenia and then prioritised 102 credible causal protein-coding genes for some of these loci using fine-mapping and summary-based mendelian randomisation (SMR) supplemented by chromatin conformation analysis. We analysed three schizophrenia common allele loci gene sets: **1)** all 1,463 genes that overlapped the 282 implicated loci; **2)** 181 genes closest to the index SNP for each associated locus (loci where the nearest gene to the index SNP was non-coding were not included in this analysis); **3**) 102 credible causal protein-coding genes that were prioritised in the original paper[12].

Schizophrenia CNV loci: A set of 63 CNVs for which there is at least some evidence for association with developmental disorders have been shown to be collectively enriched in people with schizophrenia[13]. We focussed our CNV analysis on 756 genes in regions defined by these 63 CNVs, hereafter referred to as the schizophrenia-enriched CNV set. 12 of these

63 CNVs are individually associated with schizophrenia with robust statistical significance[13], hereafter referred to as the schizophrenia-associated CNV set (n = 146 genes). Additional exploratory analyses were performed by testing LoFi and LoF-tolerant genes in the schizophrenia-enriched and schizophrenia-associated CNV loci separately.

RCV enriched genes: Genes enriched for RCVs in schizophrenia were taken from the SCHEMA sequencing study[15]. We analysed 29 autosomal genes enriched for RCVs in schizophrenia at a false discovery rate (FDR) of < 5%.

Phenotypic data

Phenotypic data were collected from UKBB participants at multiple timepoints, in person and online. All participants attended an initial assessment centre where baseline data were collected by: touchscreen questionnaires, including cognitive function measures which took around 5 minutes; a face-to-face interview; and blood sample collection. Participants were later invited to attend additional visits (instances) and to complete online questionnaires, including an online cognitive test battery. Supplementary Table 3 provides an overview of the cognitive tests and the number of participants included in our study who completed them. Cognitive test scores were taken from the first time a test was completed for participants who completed the same test multiple times. We did not include a test instance if less than 5,000 participants had completed the given cognitive test at this instance for the first time. For all tests, scores were converted to a normal distribution if not already normally distributed and standardised through conversion to z-scores.

Generalised cognition (*g*)

Cognitive assessments were brief and unsupervised. Individual cognitive tests are known to vary in their reliability and stability[44,45], but performance between different cognitive tests is correlated[44,46] (Supplementary Table 4). The latter property enables a measure of generalised cognition (*g)* to be derived from a PC analysis of multiple cognitive tests[45]. We derived *g* as the first PC from a PC analysis of the following: **1)** Numeric memory (online), **2)** Reaction time, **3)** Pairs matching, **4)** Trail making test B. Supplementary Materials 1.3 describes the derivation of *g* in detail. In total, we had sufficient data to calculate *g* for 31,147 participants.

Statistical analysis

Linear regression was used to test for association between *g*, the dependent variable, and RCV burden in different schizophrenia-associated gene sets. We included as covariates burden of synonymous variants (apart from when investigating synonymous variant burden), sequencing batch (the first 41,940 sequenced samples were sequenced using a different protocol to the remaining samples), assessment centre, and PCs 1-10. Given cognitive ability is non-linearly associated with age, and that this can differ by sex, we included covariates for sex and standardised age, standardised age^2, sex*standardised age and sex* standardised age^2.

We explored whether associations of RCV burden and *g* were explained by variation in specific cognitive tests by including individual cognitive test scores as covariates in the analyses of LoFi genes. When the instance at which a test was first taken differed across individuals (e.g. some participants had completed the cognitive test for the first time at baseline, whereas others had first completed that same test online), we added the test instance

as an additional covariate. In the analysis of *g*, we did not explore the effects of controlling for cognitive tests that were taken by < 100 samples, due to insufficient power.

To compare associations of RCV burden and *g* between gene sets, we performed z-tests using the equation $\frac{(\beta_1 - \beta_2)}{\sqrt{(\sigma_1)^2 - (\sigma_2)^2}}$ where: β_1 is the beta and σ_1 is the standard error of association from the RCV burden in gene set 1; and β_2 is the beta and σ_2 is the standard error of association from the RCV burden in gene set 2. Only independent gene sets were compared using z-tests.

Results

Distribution of protein-coding variation across ancestries

13,181,267 high-quality autosomal coding variants were observed in 185,225 unrelated samples that passed QC prior to ancestry annotation and filtering. 6,075,677 (46.09%) of these were singleton variants and 12,542,466 (95.15%) had an allele count ≤50 (equivalent to a minor allele frequency ≤ 1.39×10^{-4}).

Substantial variation was observed in the number of singletons per individual across ancestry groups, with participants of European ancestries carrying fewer singletons than any other group (Supplementary Table 5). These differences are expected given 95.11% of the sample was of European ancestries, and a higher proportion of alleles are expected to be observed two or more times in larger samples. Excluding variants observed in GnomAD[43] and TopMED[47], which contain individuals from diverse populations, did not significantly reduce the ancestry differences in the average number of singletons observed per individual (Supplementary Table 5). Given our inability to define ultra-rare variants in non-European

ancestry groups, and insufficient power to detect associations of RCVs and phenotypes in these groups, we, like others[19], focussed our analysis only on samples with genetically defined European ancestries (n=170,893).

Estimating generalised cognition in UKBB

g was estimated in 30,487 samples that passed QC and had genetically defined European ancestries. In these samples, *g* explained 38.93% of the variance of the tests included in its formation. It was also moderately correlated with each of the individual tests used to derive *g* (weakest correlation coefficient = -0.52), as well as additional cognitive tests not used to derive *g* (weakest correlation co-efficient = 0.42, Supplementary Table 4). All correlations of *g* and individual cognitive tests were in the expected direction, where higher *g* correlated with higher score on each test (we note that negative correlation coefficients are expected when lower scores for a given test indicate higher cognitive ability). These results suggest *g* forms a cross-domain measure of generalised cognition. We performed a sensitivity analysis of *g*, by modifying some of the individual cognitive tests included in its formation (see Supplementary Materials 1.4 for details) and found alternative estimates of *g* were highly correlated with the original estimate (weakest correlation co-efficient = 0.83, Supplementary Table 6), suggesting *g* is robust to different input measures.

Rare variant burden and generalised cognition

Constrained genes

Both singleton PTVs and singleton deleterious missense variants in LoFi genes were associated with lower *g* (PTVs: $\beta = -0.097$, $p = 7.92 \times 10^{-11}$; deleterious missense: $\beta = -0.063$, $p = 2.98 \times 10^{-5}$; Figure 1). PTVs and deleterious missense variants with an allele count of 2 or

higher in LoFi genes were not associated with *g* (Supplementary Table 7). In LoF-tolerant genes, no allele frequency category of PTV or deleterious missense variant was associated with g (Supplementary Table 7). Moreover, when genes were defined by LOEUF deciles, singleton PTVs in the two most constrained deciles, and singleton deleterious missense variants in the three most constrained deciles were associated with lower *g* (Supplementary Figure 3). Given the strongest associations between *g* and PTV and deleterious missense variants in LoFi genes were observed for singletons, we focussed subsequent analyses only on these classes of allele.

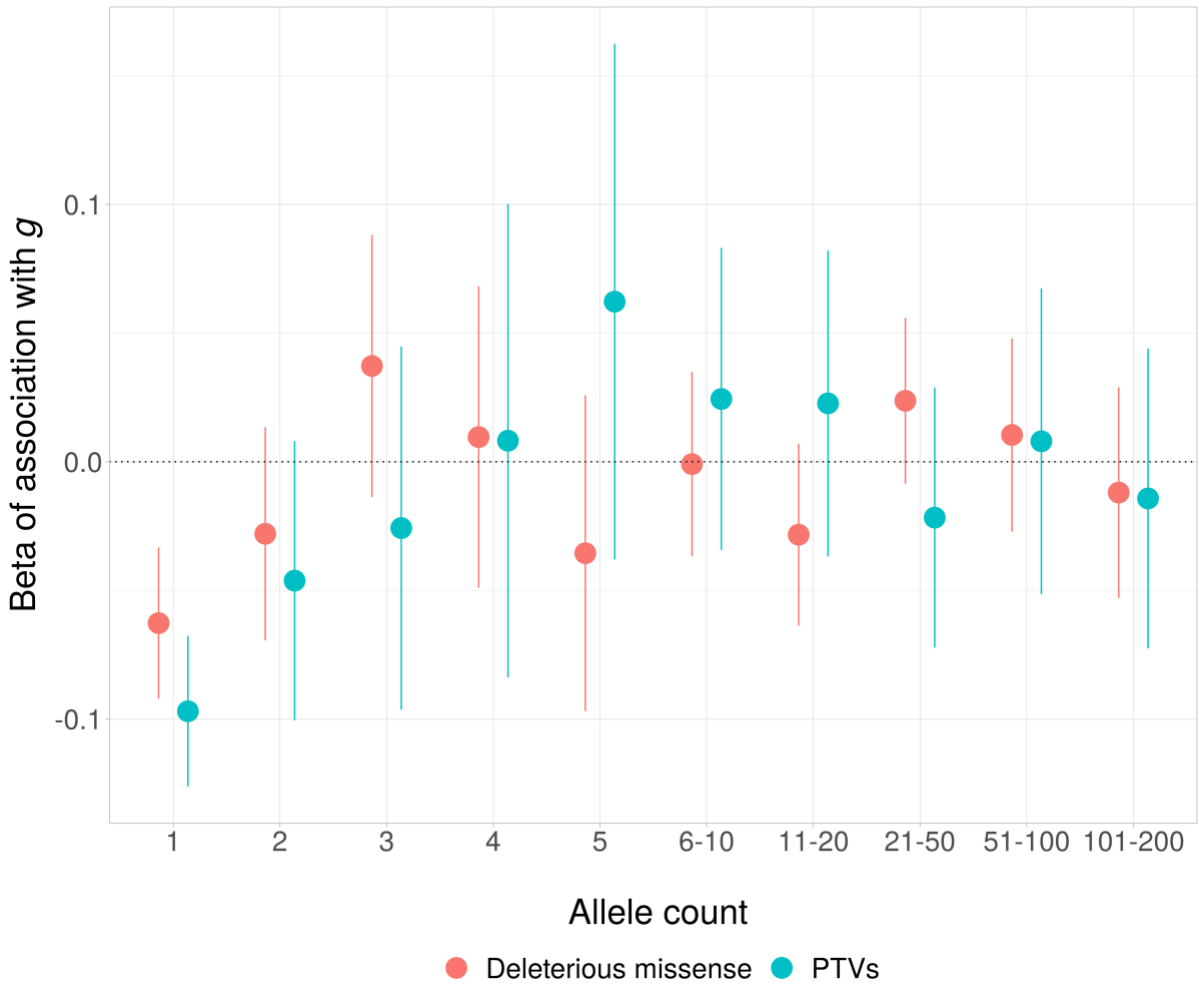

Figure 1 – Association between *g* and deleterious missense variants (red) and PTVs (blue) in LoF-intolerant genes. PTV = protein-truncating variant; *g* = generalised intelligence

Controlling for individual cognitive tests that were not included in the primary formation of *g* (trail-making test A, fluid intelligence, and symbol digit substitution), as well as those individual tests included in its formation (numeric memory, reaction time, pairs matching, and trail making test B), only partly attenuated the association between singleton PTVs in LoFi genes and *g* (Supplementary Table 8). Similar findings were observed for singleton deleterious missense variants in LoFi genes, although these variants were no longer associated with *g* when controlling for trail making test B (Supplementary Table 8). Controlling for *g* significantly attenuated the strength of association between singleton RCV burden and performance on individual cognitive tests (Supplementary Table 8). Collectively, these findings suggest that damaging rare coding variants in LoFi genes have broad effects on cognition that extend beyond those captured by any individual cognitive test.

Schizophrenia common allele loci

Neither singleton PTVs nor singleton deleterious missense variants in protein-coding genes mapping to broad schizophrenia common allele loci were significantly associated with *g* (Figure 2). Singleton deleterious missense variants in genes nearest the locus index SNPs were significantly associated with *g* ($\beta = -0.11$, $p = 0.038$; Figure 2). Singleton deleterious missense variants in credible causal genes were also significantly associated with *g* ($\beta = -0.23$, $p = 0.00096$; Figure 2), while those in genes that were not prioritised in the schizophrenia GWAS were not (Supplementary Table 9). No significant associations were found for singleton PTVs in any of the common allele gene sets (Figure 2, Supplementary Table 9). The effects on *g* of singleton deleterious missense variants in credible causal genes were significantly larger than in genes not prioritised in the schizophrenia GWAS (z-test $p = 0.0020$), and, importantly, they were also larger than for LoFi genes in general (z-test $p = 0.0051$).

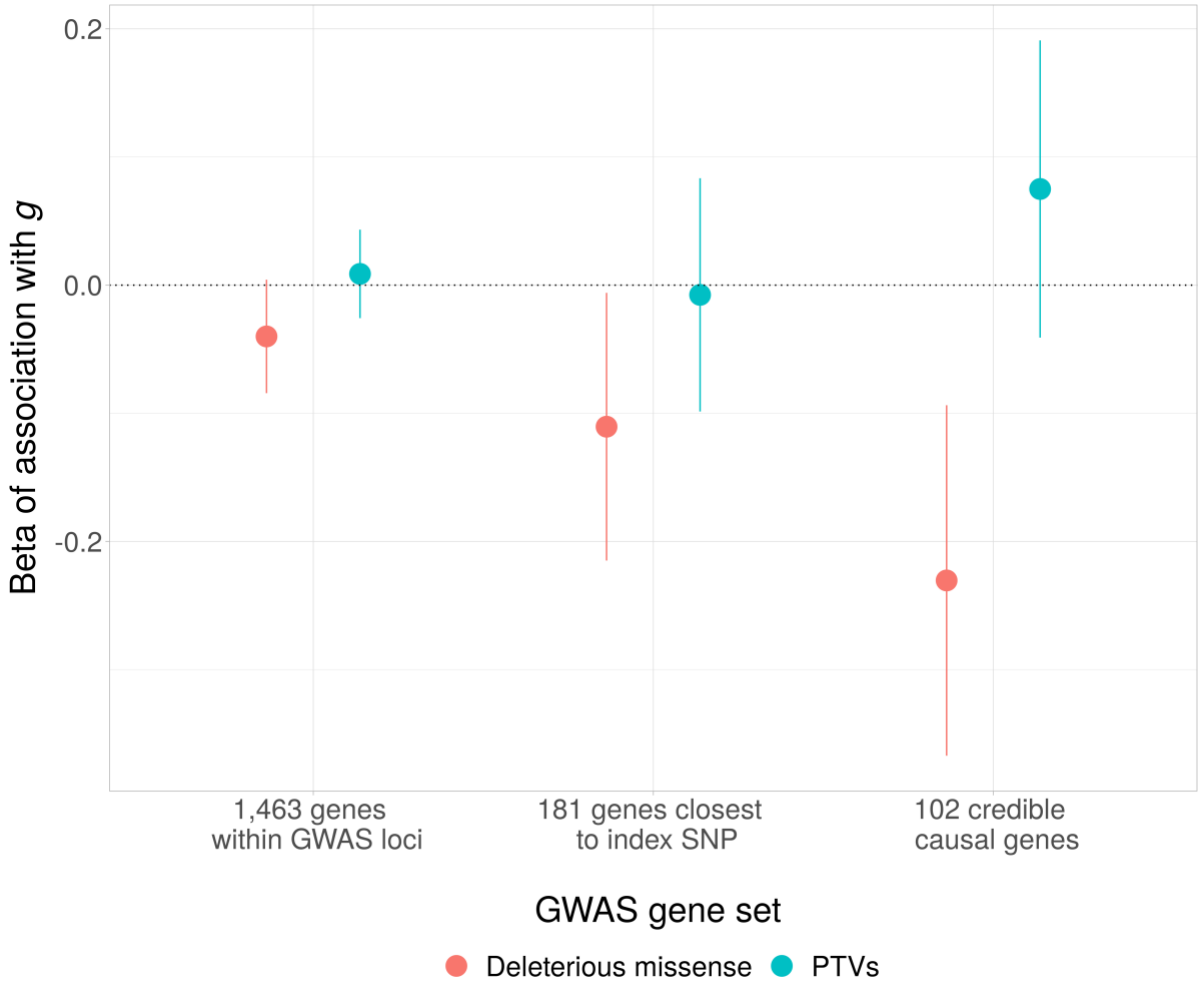

Figure 2 - Association between *g* and singleton deleterious missense variants (red) and singleton PTVs (blue) in genes within schizophrenia common allele loci. PTV = protein-truncating variant; *g* = generalised intelligence.

In the GWAS of Trubetskoy et al.[12], credible causal protein-coding genes at common allele loci were prioritised using two primary methods: statistical fine-mapping (n = 61 genes) and SMR (n = 45 genes, 4 of which were also prioritised by fine-mapping). Singleton deleterious missense variants in both sets were independently associated with lower *g* (fine-map prioritised genes: $\beta = -0.20$, $p = 0.014$; SMR prioritised genes: $\beta = -0.35$, $p = 0.0078$; Supplementary Table 9).

Prioritisation as a credible causal candidate by SMR[12] required a gene to have an eQTL and therefore be expressed in brain. As cognition is presumably (primarily) a brain related

phenotype, brain expression could potentially act as a confounder. Singleton PTVs and singleton deleterious missense variants in all brain-expressed genes (see Supplementary Table 2 for a definition of this set) were indeed significantly associated with *g* (PTVs: $\beta = -0.033$, $p = 2.02 \times 10^{-6}$; deleterious missense variants: $\beta = -0.028$, $p = 0.0010$). However, the effects on *g* for singleton deleterious missense variants in credible causal genes, and also in the subset of SMR prioritised genes, were significantly stronger than in all brain-expressed genes that were not part of these respective sets (credible causal genes vs. brain-expressed genes: z-test $p = 0.0017$; SMR prioritised genes vs. brain-expressed genes: z-test $p = 0.0072$). Thus, the associations we observe here do not simply reflect a background association between *g* and singleton deleterious missense variants in brain-expressed genes.

Schizophrenia CNV loci

Singleton PTVs in genes within regions defined by the 63 schizophrenia-enriched CNVs were significantly associated with *g* ($\beta = -0.077$, $p = 0.0026$), and this association was concentrated within LoFi genes within these loci (LoFi genes: $\beta = -0.35$, $p = 1.95 \times 10^{-6}$; LoF-tolerant genes: $\beta = -0.037$, $p = 0.18$; Figure 3, Supplementary Table 10). The effect on *g* from PTV singletons in LoFi genes within schizophrenia-enriched CNVs was stronger than the effect in all LoFi genes excluding those within the CNV loci (z-test $p = 0.00021$). This finding was not explained by LoFi genes in schizophrenia-enriched CNVs being more mutation intolerant than LoFi genes outside of these loci (Supplementary Materials Section 1.5). Singleton deleterious missense variants in genes within schizophrenia-enriched CNV loci, or within LoFi or LoF-tolerant genes in these loci, were not associated with *g* (Figure 3).

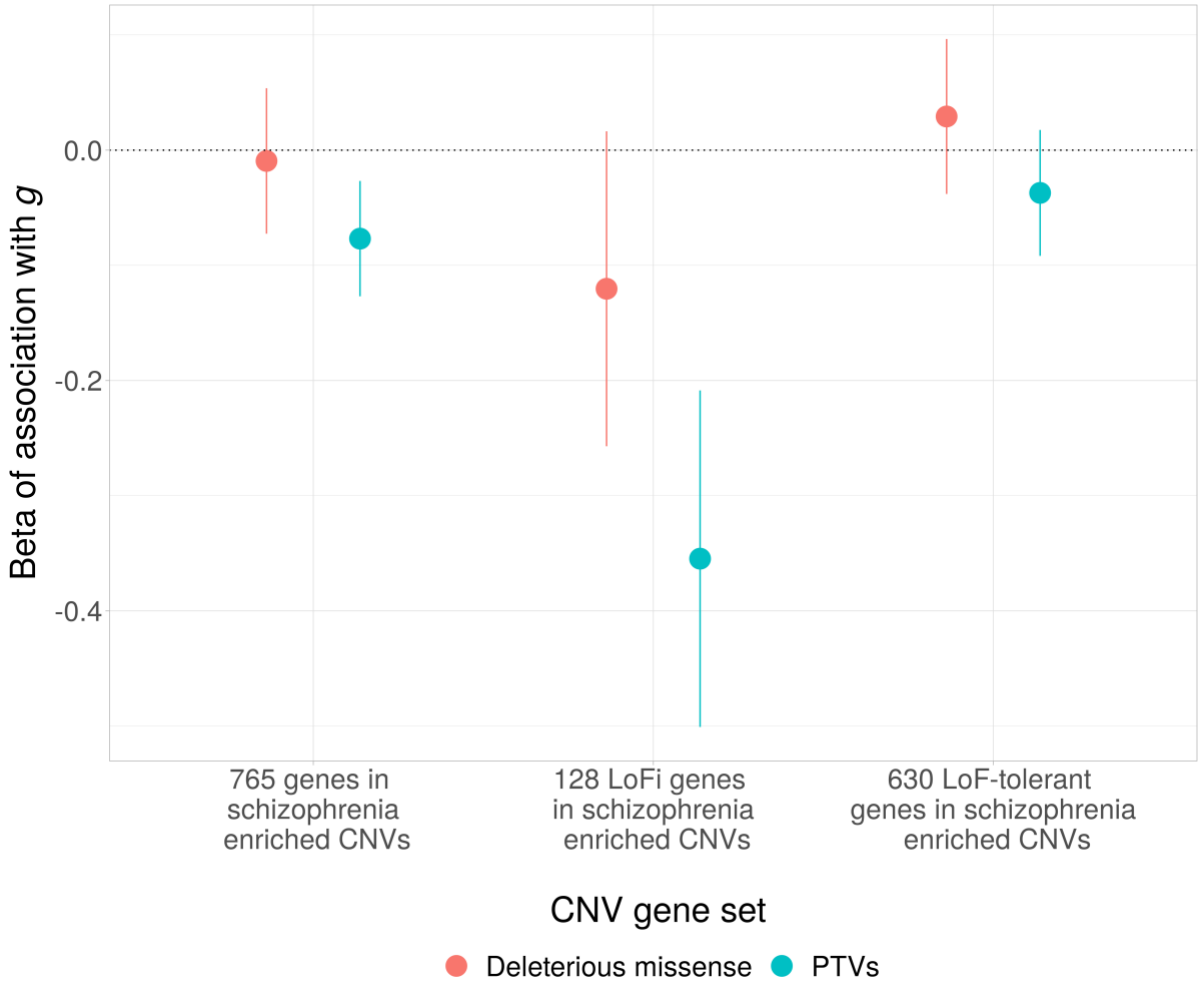

Figure 3 – Association between *g* and singleton deleterious missense variants (red) and singleton PTVs (blue) in genes within schizophrenia-enriched CNV loci. LoFi and LoF-tolerant genes in schizophrenia-enriched CNV loci were tested separately. 7 genes in the schizophrenia-enriched CNV loci did not have pLI metrics, and therefore were not included in either the LoFi or LoF-tolerant set. LoFi = LoF-intolerant; *g* = generalised intelligence.

Restricting to genes within the 12 schizophrenia-associated CNV loci, PTV singletons were again associated with lower *g* (β = -0.19, p=0.005; Supplementary Table 10; Supplementary Figure 4). This effect was significantly larger than that observed for PTVs in the schizophrenia-enriched CNV set after excluding the 12 schizophrenia-associated CNVs (z test p = 0.035). The effect size point estimate for PTVs in LoFi genes in schizophrenia-associated CNVs (β = -0.36) was similar to that observed for PTVs affecting LoFi genes in all 63 schizophrenia-enriched CNVs, but this was not significant (p = 0.052), most likely due to the small number of variants tested (24 singleton PTVs; Supplementary Table 10;

Supplementary Figure 4). For LoF-tolerant genes in the 12 schizophrenia-associated CNVs, PTVs were weakly associated with lower *g* (β = - 0.16, p = 0.041; Supplementary Table 10; Supplementary Figure 4). The effects on *g* for PTVs affecting LoFi genes in schizophrenia-associated CNV loci did not significantly differ from those affecting LoF-tolerant genes within these loci (z test p = 0.16), but this test is likely to be underpowered.

Genes associated with rare coding variants in schizophrenia

Singleton PTVs and singleton deleterious missense variants in SCHEMA FDR < 5% genes were not significantly associated with *g* (Supplementary Table 11). There was a trend for singleton PTVs in SCHEMA FDR < 5% genes to be associated with lower *g* (β = -0.25; p = 0.082), but this test is likely underpowered to identify true effects on cognition due to the small number of genes (n = 29) and variants analysed (n = 40 singleton PTVs).

Single gene burden tests

We performed single-gene association tests for damaging singleton coding variants and *g*, testing singleton PTVs, singleton deleterious missense variants, and both of these variants combined, for 883 unique genes taken from the credible causal schizophrenia GWAS gene set, schizophrenia-enriched CNV loci, and the SCHEMA FDR < 5% gene set (number of single gene tests = 2,649). We did not test genes with fewer than five qualifying variants as we lacked power to produce reliable results on the impact of variants within these genes on cognition. None of these genes showed significant evidence for association after correction for multiple testing (Bonferroni corrected p value 0.05/2,649 = 1.89×10^{-5}, Supplementary Table 12). The strongest finding was observed for singleton PTVs and singleton deleterious missense variants in *CACNA1B* (β = -0.63, p = 0.00073; Supplementary Table 12).

Discussion

In the UKBB, we show that damaging RCVs in genes and loci implicated in schizophrenia by common variant and CNV studies are associated with lower generalised cognition in individuals without a psychiatric or developmental disorder. To our knowledge, this is the first study to investigate the relationship between RCVs in schizophrenia genes and generalised cognition. We focussed on *g*, as opposed to individual cognitive tests, for two main reasons. First, a large body of evidence suggests schizophrenia is associated with global cognitive impairment[8,23,24], and compared to controls, people with schizophrenia show greater impairment in generalised cognition than for tests of specific cognitive domains[48]. Second, compared to individual cognitive tests, *g* reduces noise by limiting test-specific error and is a robust measure of general cognitive ability[45,49,50].

Genes depleted for loss-of-function mutations in healthy populations have consistently been shown to be enriched for common alleles, rare CNVs and RCVs in schizophrenia[9–11]. Here, we provide two novel lines of evidence demonstrating that damaging RCVs in these genes also contribute to variation in generalised cognition in individuals without a psychiatric or developmental disorder. We first show that singleton PTVs and deleterious missense variants in LoFi genes are significantly associated with lower *g*. We then show that this association is not explained by individual cognitive tests, including reaction time and fluid intelligence, both of which have recently been associated with damaging types of RCVs in LoFi genes[19,22]. These variants therefore have effects on cognition that are not captured by tests of single domains of cognitive function. We also found that only the rarest alleles (those occurring in only one sample in our dataset, equating to a MAF of $< 2.7\text{x}10^{-6}$) in LoFi genes were significantly associated with *g*, thus supporting findings from other studies of cognition in

UKBB[22], as well as studies of cognition in schizophrenia[51–53], that are consistent with the notion that damaging RCVs impacting generalised cognition are under strong selective constraint.

Our study also provides novel insights into the effects on cognition of RCVs in genes implicated in schizophrenia by common alleles. A recent study reported a significant association between damaging RCVs in genes overlapping schizophrenia common allele loci and lower educational attainment, and also nominally significant associations with lower fluid intelligence and longer reaction time, although these findings did not control for the effects of these variants in LoFi or brain-expressed genes[19]. Here, we show for the first time singleton damaging missense variants in the schizophrenia credible causal genes within the loci have significant effects on generalised cognition that are not confounded by the enrichments for brain expression and loss of function intolerance that have been reported for genes at associated loci. Our findings also provide orthogonal support for the SMR and fine-mapping approaches used to prioritise credible causal genes for schizophrenia in the study by Trubetskoy et al.[12], and suggest that data from population studies of cognition and RCVs could assist gene prioritization within CNVs and GWAS loci.

Association between a set of developmental disorder CNVs enriched in schizophrenia and cognition in unaffected individuals is established[17,18]. We now show that singleton PTVs in genes within these schizophrenia-enriched CNVs also contribute to lower generalised cognition at the population level. We found that PTVs in LoFi genes within the schizophrenia-enriched CNV loci have significantly stronger effects on *g* than PTVs in LoFi genes outside of the CNV loci, a novel finding that is not explained by greater levels of constraint acting on LoFi genes within the CNV loci. In our analysis of the 12 schizophrenia-associated CNV loci, we found suggestive evidence that PTVs in both LoFi and LoF-tolerant

genes contribute to lower *g*. We were underpowered to compare the effects on *g* for PTVs in LoFi and LoF-tolerant genes within the 12 schizophrenia-associated CNVs, however, a larger effect size point estimate was observed for PTVs in LoFi genes, which is consistent with findings from our analysis of all 63 schizophrenia-enriched CNV loci.

We did not find significant association between generalised cognition and RCVs in sets of genes enriched for similar types of mutation in schizophrenia, but these tests were based on a small number of genes (n = 29) and consequently are not well powered. We did observe suggestive evidence for association between PTVs in SCHEMA FDR < 5% genes and lower *g*, but as this finding was not significant, it requires replication in larger samples. Similarly, no significant associations after correction for multiple testing were found between *g* and damaging singleton coding variants in individual genes. The most significant single-gene association involved singleton PTVs and deleterious missense variants in *CACNA1B* (β = -0.63, p = 0.00073), which encodes a voltage-gated calcium channel alpha sub-unit involved in the synaptic release of neurotransmitters. Although voltage-gated calcium channels have previously been shown to have important roles in cognition[54,55], and in psychiatric disorders[56–58], further genetic evidence is required before damaging RCVs in *CACNA1B* can be considered to be convincingly associated with generalised cognition.

As outlined above, our rationale to focus our analysis on *g* was, in part, informed by previous studies showing g to be a robust measure of generalised cognition[49,50]. While the favourable psychometric properties of *g* compared to individual cognitive tests is a strength of our study, our focus on *g* may have induced a participation bias towards individuals with better cognitive function, as individuals with lower cognitive functioning may be less likely to complete all the cognitive tests included in our measure of *g*. Indeed, it has recently been shown that UKBB participants with PTVs in genes associated with autism spectrum

disorders, or in LoFi genes, are less likely to have completed the fluid intelligence test[59]. In addition to the known UKBB volunteer bias, whereby participants have a higher average socio-economic status and are generally healthier than the UK general population[60], such participation biases may reduce power to detect effects from RCVs that are strongly associated with lower cognition, should these RCVs be depleted in our sample compared with an unbiased sample. Another limitation is that we could not determine whether our findings are generalisable to people from non-European ancestries, due to the limited number of people with these ancestries in UKBB.

In conclusion, we show that rare damaging RCVs in constrained genes contribute to variation in generalised cognitive function in individuals without a psychiatric or developmental disorder from UKBB, with significantly stronger effects on cognition observed for damaging RCVs in genes previously associated with liability to schizophrenia. Our findings strengthen and extend the evidence for an overlap between genetic liability for schizophrenia and that for lower cognition in individuals without a psychiatric disorder. As such, they point to shared underlying biology between schizophrenia risk and general cognitive function in the population that is not explained by general gene properties such as loss of function intolerance and brain expression. This study demonstrates the utility of exploiting large sequencing datasets of unaffected individuals, such as UKBB, to identify genes with shared effects on cognition and schizophrenia and provides a route towards determining the biological processes underlying cognitive impairment in schizophrenia.

Acknowledgments

This work was supported by a Wellcome Trust Integrative Neuroscience PhD Studentship to EF (108891/B/15/Z /WT), and a UKRI Future Leaders Fellowship Grant to ER (MR/T018712/1).

This work uses data provided by patients and collected by the NHS as part of their care and support: Copyright © (2023), NHS England; Re-used with the permission of the NHS England and UK Biobank; All rights reserved.

This research used data assets made available by National Safe Haven as part of the Data and Connectivity National Core Study, led by Health Data Research UK in partnership with the Office for National Statistics and funded by UK Research and Innovation (research which commenced between 1st October 2020 – 31st March 2021 (MC_PC_20029); 1st April 2021 - 30th September 2022 (MC_PC_20058).

Competing interests

ER, JTRW, MCO and MJO reported receiving grants from Akrivia Health outside the submitted work. JTRW, MJO and MCO reported receiving grants from Takeda Pharmaceutical Company Ltd outside the submitted work. Takeda and Akrivia played no part in the conception, design, implementation, or interpretation of this study.

Data availability

This study used data from the UK Biobank, which is available for health-related research upon registration and application through the UK Biobank Access Management System (https://www.ukbiobank.ac.uk/enable-your-research/register). The code required to reproduce our analyses is publicly available (https://github.com/eilidhfenner/UKBB_WES_data_processing).

References

1. Saha, S., Chant, D., Welham, J. & McGrath, J. A Systematic Review of the Prevalence of Schizophrenia. *PLOS Med.* **2**, e141 (2005).

2. Sørensen, H. J., Mortensen, E. L., Parnas, J. & Mednick, S. A. Premorbid Neurocognitive Functioning in Schizophrenia Spectrum Disorder. *Schizophr. Bull.* **32**, 578–583 (2006).

3. Reichenberg, A. *et al.* Static and dynamic cognitive deficits in childhood preceding adult schizophrenia: a 30-year study. *Am. J. Psychiatry* **167**, 160–169 (2010).

4. Fusar-Poli, P. *et al.* Cognitive Functioning in Prodromal Psychosis: A Meta-analysis. *Arch. Gen. Psychiatry* **69**, 562–571 (2012).

5. Bowie, C. R., Reichenberg, A., Patterson, T. L., Heaton, R. K. & Harvey, P. D. Determinants of Real-World Functional Performance in Schizophrenia Subjects: Correlations With Cognition, Functional Capacity, and Symptoms. *Am. J. Psychiatry* **163**, 418–425 (2006).

6. Green, M. F., Horan, W. P. & Lee, J. Nonsocial and social cognition in schizophrenia: current evidence and future directions. *World Psychiatry* **18**, 146–161 (2019).

7. Green, M. F., Kern, R. S., Braff, D. L. & Mintz, J. Neurocognitive Deficits and Functional Outcome in Schizophrenia: Are We Measuring the "Right Stuff"? *Schizophr. Bull.* **26**, 119–136 (2000).

8. Dickinson, D., Iannone, V. N., Wilk, C. M. & Gold, J. M. General and specific cognitive deficits in schizophrenia. *Biol. Psychiatry* **55**, 826–833 (2004).

9. Singh, T. *et al.* The contribution of rare variants to risk of schizophrenia in individuals with and without intellectual disability. *Nat. Genet.* **49**, 1167–1173 (2017).

10. Rees, E. *et al.* Schizophrenia, autism spectrum disorders and developmental disorders share specific disruptive coding mutations. *Nat. Commun.* **12**, 5353 (2021).

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

12. Trubetskoy, V. *et al.* Mapping genomic loci implicates genes and synaptic biology in schizophrenia. *Nature* **604**, 502–508 (2022).

13. Rees, E. *et al.* Analysis of Intellectual Disability Copy Number Variants for Association With Schizophrenia. *JAMA Psychiatry* **73**, 963–969 (2016).

14. Marshall, C. R. *et al.* Contribution of copy number variants to schizophrenia from a genome-wide study of 41,321 subjects. *Nat. Genet.* **49**, 27–35 (2017).

15. Singh, T. *et al.* Rare coding variants in ten genes confer substantial risk for schizophrenia. *Nature* 1–9 (2022) doi:10.1038/s41586-022-04556-w.

16. Liu, D. *et al.* Schizophrenia risk conferred by rare protein-truncating variants is conserved across diverse human populations. *Nat. Genet.* **55**, 369–376 (2023).

17. Kendall, K. M. *et al.* Cognitive Performance Among Carriers of Pathogenic Copy Number Variants: Analysis of 152,000 UK Biobank Subjects. *Biol. Psychiatry* **82**, 103–110 (2017).

18. Stefansson, H. *et al.* CNVs conferring risk of autism or schizophrenia affect cognition in controls. *Nature* **505**, 361–366 (2014).

19. Chen, C.-Y. *et al.* The impact of rare protein coding genetic variation on adult cognitive function. *Nat. Genet.* 1–12 (2023) doi:10.1038/s41588-023-01398-8.

20. Kingdom, R., Beaumont, R. N., Wood, A. R., Weedon, M. N. & Wright, C. F. Genetic modifiers of rare variants in monogenic developmental disorder loci. 2022.12.15.22283523 Preprint at https://doi.org/10.1101/2022.12.15.22283523 (2022).

21. Lencz, T. *et al.* Molecular genetic evidence for overlap between general cognitive ability and risk for schizophrenia: a report from the Cognitive Genomics consorTium (COGENT). *Mol. Psychiatry* **19**, 168–174 (2014).

22. Gardner, E. J. *et al.* Reduced reproductive success is associated with selective constraint on human genes. *Nature* **603**, 858–863 (2022).

23. Nuechterlein, K. H. *et al.* Identification of separable cognitive factors in schizophrenia. *Schizophr. Res.* **72**, 29–39 (2004).

24. Schaefer, J., Giangrande, E., Weinberger, D. R. & Dickinson, D. The global cognitive impairment in schizophrenia: Consistent over decades and around the world. *Schizophr. Res.* **150**, 42–50 (2013).

25. Van Hout, C. V. *et al.* Exome sequencing and characterization of 49,960 individuals in the UK Biobank. *Nature* **586**, 749–756 (2020).

26. Poplin, R. *et al.* A universal SNP and small-indel variant caller using deep neural networks. *Nat. Biotechnol.* **36**, 983–987 (2018).

27. Szustakowski, J. D. *et al.* Advancing human genetics research and drug discovery through exome sequencing of the UK Biobank. *Nat. Genet.* **53**, 942–948 (2021).

28. Regier, A. A. *et al.* Functional equivalence of genome sequencing analysis pipelines enables harmonized variant calling across human genetics projects. *Nat. Commun.* **9**, 4038 (2018).

29. Krasheninina, O. *et al.* Open-source mapping and variant calling for large-scale NGS data from original base-quality scores. 2020.12.15.356360 Preprint at https://doi.org/10.1101/2020.12.15.356360 (2020).

30. dnanexus/oqfe. https://hub.docker.com/r/dnanexus/oqfe.

31. GRCh38 - hg38 - Genome - Assembly - NCBI. https://www.ncbi.nlm.nih.gov/assembly/GCF_000001405.26/.

32. UKB Showcase: WES protocol. https://biobank.ndph.ox.ac.uk/showcase/refer.cgi?id=915.

33. UKB Showcase: Exome Sequences. https://biobank.ndph.ox.ac.uk/showcase/label.cgi?id=170.

34. Hail Team. Hail. https://github.com/hail-is/hail.

35. Lin, M. F. *et al.* GLnexus: joint variant calling for large cohort sequencing. *bioRxiv* 343970 (2018) doi:10.1101/343970.

36. Hanscombe, K. ukbtools: Manipulate and Explore UK Biobank Data. (2019).

37. Relationship Inference in KING. https://www.kingrelatedness.com/manual.shtml.

38. Manichaikul, A. *et al.* Robust relationship inference in genome-wide association studies. *Bioinformatics* **26**, 2867–2873 (2010).

39. McLaren, W. *et al.* The Ensembl Variant Effect Predictor. *Genome Biol.* **17**, 122 (2016).

40. Karczewski, K. konradjk/loftee. (2020).

41. Ioannidis, N. M. *et al.* REVEL: An Ensemble Method for Predicting the Pathogenicity of Rare Missense Variants. *Am. J. Hum. Genet.* **99**, 877–885 (2016).

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

43. Karczewski, K. J. *et al.* The mutational constraint spectrum quantified from variation in 141,456 humans. *Nature* **581**, 434–443 (2020).

44. Lyall, D. M. *et al.* Cognitive Test Scores in UK Biobank: Data Reduction in 480,416 Participants and Longitudinal Stability in 20,346 Participants. *PLoS ONE* **11**, e0154222 (2016).

45. Fawns-Ritchie, C. & Deary, I. J. Reliability and validity of the UK Biobank cognitive tests. *PLoS ONE* **15**, (2020).

46. Spearman, C. *'General Intelligence' Objectively Determined and Measured*. 73 (Appleton-Century-Crofts, 1961). doi:10.1037/11491-006.

47. Taliun, D. *et al.* Sequencing of 53,831 diverse genomes from the NHLBI TOPMed Program. *Nature* **590**, 290–299 (2021).

48. McCutcheon, R. A., Keefe, R. S. E. & McGuire, P. K. Cognitive impairment in schizophrenia: aetiology, pathophysiology, and treatment. *Mol. Psychiatry* 1–17 (2023) doi:10.1038/s41380-023-01949-9.

49. Johnson, W., Nijenhuis, J. te & Bouchard, T. J. Still just 1 g: Consistent results from five test batteries. *Intelligence* **36**, 81–95 (2008).

50. Johnson, W., Bouchard, T. J., Krueger, R. F., McGue, M. & Gottesman, I. I. Just one g: consistent results from three test batteries. *Intelligence* **32**, 95–107 (2004).

51. Rammos, A. *et al.* Family-based analysis of the contribution of rare and common genetic variants to school performance in schizophrenia. *Mol. Psychiatry* 1–7 (2023) doi:10.1038/s41380-023-02013-2.

52. Creeth, H. D. J. *et al.* Ultrarare Coding Variants and Cognitive Function in Schizophrenia. *JAMA Psychiatry* **79**, 963–970 (2022).

53. Song, J. *et al.* The impact of educational attainment, intelligence and intellectual disability on schizophrenia: a Swedish population-based register and genetic study. *Mol. Psychiatry* **27**, 2439–2447 (2022).

54. Cosgrove, D. *et al.* Cognitive Characterization of Schizophrenia Risk Variants Involved in Synaptic Transmission: Evidence of CACNA1C's Role in Working Memory. *Neuropsychopharmacol. Off. Publ. Am. Coll. Neuropsychopharmacol.* **42**, 2612–2622 (2017).

55. Jones, E. G. Calcium channels in higher-level brain function. *Proc. Natl. Acad. Sci. U. S. A.* **104**, 17903–17904 (2007).

56. Zamponi, G. W. Targeting voltage-gated calcium channels in neurological and psychiatric diseases. *Nat. Rev. Drug Discov.* **15**, 19–34 (2016).

57. Andrade, A. *et al.* Genetic Associations between Voltage-Gated Calcium Channels and Psychiatric Disorders. *Int. J. Mol. Sci.* **20**, 3537 (2019).

58. Heyes, S. *et al.* Genetic disruption of voltage-gated calcium channels in psychiatric and neurological disorders. *Prog. Neurobiol.* **134**, 36–54 (2015).

59. Rolland, T. *et al.* Phenotypic effects of genetic variants associated with autism. *Nat. Med.* 1–10 (2023) doi:10.1038/s41591-023-02408-2.

60. Fry, A. *et al.* Comparison of Sociodemographic and Health-Related Characteristics of UK Biobank Participants With Those of the General Population. *Am. J. Epidemiol.* **186**, 1026–1034 (2017).