Evaluating the role of common risk variation in the recurrence risk of schizophrenia in multiplex schizophrenia families

Mohammad Ahangari[1,2], Amanda Elswick Gentry[1], Tan-Hoang Nguyen[1,3], Brian Verrelli[4], Silviu-Alin Bacanu[1,3,5], Robert Kirkpatrick[1,3], Kenneth Kendler[1,3,5], Bradley Todd Webb[6], Brien Riley[1,3,5]

[1]Virginia Institute for Psychiatric and Behavioral Genetics, Virginia Commonwealth University, Richmond, VA, USA. [2]Integrative Life Sciences PhD Program, Virginia Commonwealth University, Richmond, VA, USA. [3]Department of Psychiatry, Virginia Commonwealth University, Richmond, VA, USA. [4]Center for the Study of Biological Complexity, Virginia Commonwealth University, Richmond, Virginia, USA. [5]Department of Human and Molecular Genetics, Virginia Commonwealth University, Richmond, VA, USA. [6]GenOmics, Bioinformatics, and Translational Research Center, Biostatistics and Epidemiology Division, RTI

Abstract

Importance: Multiplex schizophrenia families have higher recurrence risk of schizophrenia compared to the families of singleton cases in the population, but the source of increased familial recurrence risk is unknown. Determining the source of this observation is essential, as it will define the relative focus on common versus rare genetic variation in case-control and family studies of schizophrenia.

Objective: To evaluate the role of common risk variation in the recurrence risk of schizophrenia, by comparing the polygenic risk scores in familial versus ancestry matched singleton cases of schizophrenia.

Design: We used the latest genome-wide association study data of schizophrenia (N=166,464) to construct polygenic risk scores in multiplex family members, singleton cases and controls. To account for the high degree of relatedness in the samples, analyses were carried out using a mixed effects logistic regression model with the family structure modeled as a random effect.

Setting: Population and family based.

Participants: We used a large, homogenous sample of 1,005 individuals from 257 families from the Irish Study of High-Density Schizophrenia Families, 2,224 singleton cases and 2,284 population controls all from the population of the island of Ireland.

Exposures: Polygenic risk scores, diagnostic categories, familial or singleton case status.

Main outcomes and measures: The primary outcomes were the mixed effects logistic regression results generated from comparison between different groups.

Results: Polygenic risk scores in singleton cases did not differ significantly from familial cases (p=0.49), rejecting the hypothesis that an increased burden of common risk variation can account for the higher recurrence risk of schizophrenia in multiplex families.

Conclusions and relevance: This study suggests that a higher burden of common schizophrenia risk variation cannot account for the increased familial recurrence risk of schizophrenia in multiplex families. In the absence of elevated polygenic risk scores in multiplex schizophrenia families, segregation of rare variation in the genome and environmental exposures unique to the families may explain the increased multiplex familial recurrence risk of schizophrenia. These findings also further validate the concept of a genetically influenced psychosis spectrum in multiplex schizophrenia families as shown by a continuous increase of common risk variation burden from unaffected relatives to familial cases of schizophrenia in the families.

Introduction

Schizophrenia (SCZ) is a severe, clinically heterogeneous disorder with a prevalence of ~1%[1], and heritability (h^2) estimates of ~80%[2–6]. Family history is the strongest risk factor for developing SCZ[7]. Despite high heritability, ~2/3 of SCZ cases report no family history of psychotic illness, and most subjects with a positive family history report only a single affected relative[8,9], concordant with the rates of 29% family history positive and 71% family history negative observed in the sample of singleton SCZ cases studied here[10].

Many linkage studies of SCZ were undertaken in samples of families with multiple cases of SCZ or related psychosis-spectrum disorders, like the Irish Study of High-Density Schizophrenia Families (ISHDSF)[11–14]. Such multiplex families display substantially higher recurrence risk of SCZ than reported in singleton cases[8,9]. This discrepancy in recurrence risk suggests that there may be important differences in the genetic or environmental risk architecture between familial and singleton SCZ cases that warrant further investigation.

One explanation of this difference is that SCZ cases from multiplex families may carry a higher burden of common risk variation, and higher SCZ polygenic risk scores (PRS), than ancestry matched singleton cases. Another explanation is that the increased recurrence risk in multiplex families may be attributable to segregation of rarer, higher risk variation, identified through exome or whole-genome sequencing. Sequencing studies suggest strongly that rare, deleterious variation in the genome is involved in the genetic etiology of SCZ and other psychiatric disorders[15–20], but the extent to which rare variation contributes to SCZ risk in multiplex families is currently unknown. A third hypothesis, not addressed here, is that familial cases may have increased exposure to environmental risks.

Mega-analyses of SCZ genome-wide association study (GWAS) data by the Psychiatric Genomics Consortium Schizophrenia Working Group (PGC-SCZ) have identified common risk variants associated with SCZ[21–23]. In the most recent PGC mega-analysis (PGC3-SCZ), 270 independent risk loci were found to be robustly associated with SCZ, with single nucleotide polymorphism (SNP) based h^2 of ~24%. GWAS data from such analyses are frequently used to construct PRS to index an individual's common variant genetic risk for a disorder. Although PRS currently lack power to predict SCZ in the general population, they have been shown to index meaningful differences in SCZ liability between individuals. In the European PGC3-SCZ samples the highest PRS centile has an OR of 44 (95% CI=31-63) for SCZ compared to the lowest centile of PRS, and OR of 7 (95% CI=5.8-8.3) when the top centile is compared with the remaining 99% of the individuals in the sample[23].

We have previously used the summary statistics from the first wave of PGC-SCZ mega-analysis[21] to investigate whether the concept of the psychosis spectrum is supported by empirical data in the ISHDSF[27]. PRS analyses have been performed for other psychiatric phenotypes in multiplex family samples smaller than the ISHDSF[24–26]. Here, we extend our previous work by using PRS profiling in a large, homogeneous sample of multiplex SCZ families, singleton SCZ cases and population controls from the island of Ireland to directly test the hypothesis that common risk variation in the genome may explain the increased recurrence risk of SCZ in multiplex SCZ families compared to families of singleton cases. Furthermore, to demonstrate the specificity of the constructed PRS from PGC3-SCZ GWAS for SCZ, we also constructed a PRS for low-density lipoprotein (LDL) as a negative control in our analysis. The source of the increased familial recurrence risk of SCZ is important for future research into the genetic

etiology of familial SCZ and potentially for both diagnosis and treatment of SCZ with different familial backgrounds.

Methods

Sample Description

Irish Study of High-Density Schizophrenia Families

Probands in the ISHDSF sample were ascertained from psychiatric hospitals in the Republic of Ireland and Northern Ireland[28]. Inclusion criteria were two or more first-degree relatives meeting DSM-III-R criteria for SCZ or poor-outcome schizoaffective disorder (PO-SAD) and all four grandparents being born in Ireland or the United Kingdom. Relatives of probands were interviewed by trained field staff. Hospital and out-patient records were obtained and abstracted in > 98% of cases with SCZ or PO-SAD diagnoses.

The concentric diagnostic schema of the ISHDSF includes 4 case definitions: *narrow* (SCZ, PO-SAD), *intermediate* (adding schizotypal personality, schizophreniform, and delusional disorders, atypical psychosis and good-outcome schizoaffective disorder), *broad* (adding psychotic affective illness, paranoid, avoidant and schizoid personality disorders and other disorders that significantly aggregate in relatives of Irish probands) and *very broad* (adding any other psychiatric illness). The ISHDSF sample also includes *unaffected* family members with no diagnosis of any psychiatric illness (Table 1). The ISHDSF diagnostic schema is described extensively elsewhere[29].

Irish Schizophrenia Genomics Consortium Case/Control Sample (ISGC)

The ISGC sample was assembled for a GWAS of SCZ in Ireland. Details of recruitment, screening and quality control (QC) methods used for the ISGC sample have been previously described in detail elsewhere[30]. Briefly, the case sample was recruited through community mental health service and inpatient units in the Republic of Ireland and Northern Ireland following protocols with local ethics approval. All participants were interviewed using a structured clinical interview for DSM-III-R or DSM-IV, were over 18 years of age and reported all four grandparents born either in Ireland or the United Kingdom. Cases were screened to exclude substance-induced psychotic disorder or psychosis due to a general medical condition.

Controls from the Irish Biobank used in ISGC were blood donors from the Irish Blood Transfusion Service recruited in the Republic of Ireland. Inclusion criteria were all four grandparents born in Ireland or the United Kingdom and no reported history of psychotic illness. Due to the relatively low lifetime prevalence of SCZ, misclassification of controls should have minimal impact on power[31].

Genotyping and QC

Samples were genotyped using 3 different arrays. 830 individuals representing 237 families from ISHDSF were genotyped on the Illumina 610-Quad Array. An additional 175 ISHDSF individuals were later genotyped on the Infinium PsychArray V.1.13 Array. For the case-control sample, 3,240 individuals (1,509 cases, 1,733 controls), were genotyped using the Affymetrix V.6.0 Array. An additional 716 cases and 580 controls were later genotyped on the PsychArray with the additional ISHDSF individuals described above. The same QC protocols were applied to

all three datasets and full details are described elsewhere for ISHDSF[29] and the case-control sample[30]. In brief, exclusion criteria for samples were a call rate of <95%, more than one Mendelian error in the ISHDF, and difference between reported and genotypic sex. Exclusion criteria for SNPs were MAF <1%, call rate <98%, and $p<0.0001$ for deviation from Hardy-Weinberg expectation. The final ISHDSF sample included 1,005 individuals from 257 pedigrees, and the final case-control sample included 4,508 individuals (2,224 cases and 2,284 controls), whose SNP data passed all QC filters.

Imputation

Genotypes passing QC were phased using Eagle V.2.4[32] and phased genotypes were then imputed to the Haplotype Reference Consortium (HRC) reference panel[33] on the Michigan Imputation Server using Minimac4[34]. The HRC reference includes 64,975 samples from 20 different studies that are predominantly of European ancestry, suitable for imputation in our sample from Ireland. Each of the three genotype sets were imputed separately and the imputed genotype probabilities in VCF format were downloaded from the Michigan Imputation Server, and genotype dosages in the VCF files were extracted and used for PRS construction and analyses. As part of the post-imputation QC for this analysis, variants with MAF <1% and r^2 score of <0.3 were excluded from the analyses. After imputation and all QC, 9,298,012 SNPs in ISHDSF, 11,080,279 SNPs in the case-control sample and 11,081,999 SNPs in the combined PsychArray sample remained for analysis; 9,008,825 SNPs were shared across all three datasets and were used for PRS construction and downstream analyses.

Construction of Polygenic Risk Score

The GWAS summary statistics for PGC3-SCZ (N= 306,011) were first QC'd by excluding variants with MAF < 0.01 and imputation quality score of < 0.9, as well as removing strand ambiguous variants and insertion deletion polymorphisms. We then constructed PRS for all subjects using a Bayesian regression framework by placing a continuous shrinkage prior on SNP effect sizes using PRS-CS[35]. PRS-CS uses LD information from an external reference panel (1000 Genomes European Phase 3 European sample[36]) to estimate the posterior effect sizes for each SNP in a given set of summary statistics. Although *p*-value thresholding and clumping methods have been used for PRS construction[37], PRS-CS has shown substantial improvement in predictive power compared to those methods for PRS construction[38]. Similar to Linkage Disequilibrium Score regression (LDSC)[39], PRS-CS limits the SNPs for PRS construction to approximately 1.2 million high quality variants from HapMap3. By restricting the PRS construction to HapMap3 variants, the partitioning provides ~ 500 SNPs per LD block which substantially reduces memory and computational costs. The constructed PRS using PRS-CS method were normalized against the score distribution in the population control for all subsequent analyses.

In order to show the specificity of the PRS constructed from PGC-SCZ in our analysis, an additional PRS for low density lipoprotein (LDL, N=87,048) from the ENGAGE Consortium[40] was also constructed using the same protocol described for PGC-SCZ above. Genetic correlation studies show that there is no significant correlation between SCZ and LDL, making LDL an appropriate phenotype as a negative control for this analysis[41,42].

Genomic Relationship Matrix and Statistical Analyses

Statistical analyses on PRS were carried out using a mixed effects logistic regression model with in the GMMAT package[43] in R[44]. To account for the high degree of relatedness among individuals in the study, we used a series of mixed effects logistic regression models and modeled the family structure as a random effect with genetic relationship matrix (GRM) calculated using LDAK with LD correction parameters suited for families[45]. In addition to adjusting for GRM as a random effect in the mixed model, we also included sex as a fixed effect. The final results were adjusted for multiple testing using the Holm method by utilizing *p.adjust()* command in R.

Results

The mean SCZ PRS across the diagnostic categories for SCZ and LDL are displayed in Figures 1a and 1b. As expected, no significant differences in LDL PRS were observed between any of the diagnostic categories versus population controls, indicating the specificity of PGC3-SCZ PRS in this study.

PGC3-SCZ PRS results show that *Narrow* spectrum category in the families, which includes familial cases of SCZ, had the highest mean PRS (Z=1.74, SE=0.11) followed by singleton cases from the population (Z=1.71, SE=0.06), *intermediate* spectrum cases (Z=1.39, SE=0.14), *broad* spectrum cases (Z=1.26, SE=0.19), *very-broad* spectrum cases (Z=1.09, SE=0.13), *unaffected* family members (Z=0.99, SE=0.10) and population controls (Z=0.004, SE=0.05).

No significant PRS difference between familial and singleton cases of SCZ

Logistic mixed models for SCZ PRS (Table 2, Supplementary Table 2 for full comparison results) showed no significant difference between *familial* SCZ cases from ISHDF and singleton SCZ cases from the population, ($p = 0.49$), rejecting the hypothesis that an increased burden of common risk variants can account for the higher recurrence risk of SCZ in ISHDSF.

All family members carry a high burden of common SCZ risk variants

All ISHDSF diagnostic categories, including the *unaffected* family members show a significantly higher SCZ PRS than those observed in the population controls (Table 2). PRS comparison within the families (Supplementary Table 2) shows that the PRS from the *very-broad* category, which includes individuals with any psychiatric disorders, is not significantly different from *unaffected* members of the families, indicating a similar burden of common SCZ risk variants in these two groups in ISHDSF. This observation is consistent with SCZ transmission through unaffected family members observed in some ISHDSF pedigrees. No significant difference was found between the PRS for *intermediate* and *broad* categories after multiple testing comparison in the families, indicating that individuals in both categories have a similar burden of common SCZ risk variants despite the presence of a range of diagnoses on the psychosis spectrum such as atypical psychosis and delusional disorder in the *intermediate* category, and non-psychotic spectrum disorders such as major depressive disorder in the *broad* category. Within the families, both familial and singleton cases of SCZ, show a significantly higher PRS compared to all other categories in the ISHDSF (Supplementary Table 2), underlining the important role of common risk variation in the genetic architecture of SCZ in familial as well as singleton cases. Finally, we observe a significantly higher PRS in *unaffected* family members compared to the population

controls ($P < 1.07e^{-29}$), indicating the presence of a higher burden of common schizophrenia risk variants across all family members, regardless of their diagnostic status.

Discussion

This study aimed to investigate the source of increased recurrence risk of SCZ in multiplex SCZ families compared to the families of singleton cases in the population. Multiplex SCZ families represent the upper bounds of the distribution of recurrence risk for SCZ, and our results provide empirical evidence that increased recurrence risk of SCZ in multiplex SCZ families is unlikely to be attributable to an increased burden of common SCZ risk variation in the genome. Therefore, the hypothesis that high familial recurrence risk of SCZ in multiplex families may be attributable to excess rare variation specific to schizophrenia warrants further investigation. These findings also further validate the concept of a genetically influenced psychosis spectrum in multiplex SCZ families as shown by a continuous increase of common risk variation burden across all members of ISHDSF, from *unaffected* family members, to *narrow* SCZ cases in ISHDSF.

Analyses of multiplex BIP families have shown that affected and unaffected individuals in in these families have significantly higher PRS for both BIP and SCZ compared to ancestry matched population controls, and familial cases have a significantly higher BIP PRS compared to ancestry matched, singleton cases. These results in addition to sparse evidence for the involvement of rare risk variation in the genetic architecture of BIP, demonstrates the importance of common risk variation in the genetic architecture of familial BIP[25].

In the last decade, whole exome sequencing studies (WES) of SCZ in family and case-control samples have demonstrated that in addition to common risk variation, rare variation also plays an important role in SCZ risk[46–49]. Although WES studies are only now reaching sample sizes

sufficiently powered to detect individually associated rare variation and rare variant enriched genes[49], earlier WES and rare variation studies observe consistent enrichment of rare variation in certain gene-sets and functional categories related to SCZ[46], indicating the contribution of rare variation in the genetic architecture of SCZ. In addition to rare variation identified through WES studies, SNP signals from PGC3-SCZ GWAS are shown to be enriched in noncoding functional sequences in the genome[23], further underscoring the importance of conducting large scale whole-genome sequencing in SCZ to identify rare variation in non-coding regions of the genome. Results from the 1000 Genomes Project demonstrates that rare functional variation is frequent in the genome[50] and shows strong population specificity[51]. For example, using GWAS probe intensity data in our Irish case-control sample, we have previously detected a rare, novel 149kb duplication overlapping the PAK7 gene only found in the Irish population[52]. This duplication is associated with SCZ in our CC sample (p=0.007) and a replication sample of Irish and UK case-controls (including bipolar disorder) with 22 carriers in 11,707 cases and 10 carriers in 21,204 controls (p=0.0004, OR=11.3). This duplication is in strong linkage disequilibrium (LD) with local haplotypes ($p=2.5\times10^{-21}$), indicating a single ancestral event and inheritance identical by descent in carriers. We are currently in the process of sequencing the whole genome of all ISHDSF members as well as the full set of ancestry matched cases and controls to investigate whether rare variation identified from whole genome sequencing could account for the high recurrence risk in multiplex SCZ families.

The results presented in this study should be interpreted in the context of several limitations. First, the liability that is captured by PRS constructed from PGC3-SCZ is currently insufficient for predicting a diagnosis of SCZ (AUC=0.71)[23], meaning that PRS alone cannot be used as a diagnostic tool. Despite that, PRS is shown to be a reliable measurement of common risk

variation in the genome, making it a suitable for indexing an individual's risk for SCZ in this study. Second, Diagnostic ascertainment of ISHDSF required two first-degree relatives with a diagnosis of SCZ or PO-SAD. Therefore, the various diagnostic categories in the ISHDSF diagnostic schema are not equally represented across all families[28]. For example, the lower number of individuals satisfying *broad* and *very broad* diagnostic schema in the families, means that the power of analysis in those subgroups is lower. However, the *narrow* category which includes familial SCZ cases in the ISHDSF sample, has the highest number of individuals across all the diagnostic categories in ISHDSF, making the sample suitable for the main hypothesis being tested in this work. Third, the SCZ PRS from PGC3-SCZ GWAS accounts for only ~2.6% of the total variance in SCZ[23], and genetic risks from rare and structural variation are not represented in the PRS. As a result, some known genetic risk factors for SCZ such as the 22-q11 deletion[53] are not included in PRS construction, and such genetic risk factors are best measured through whole genome sequencing studies. Fourth, this analysis did not assess the common risk variant burden of each family separately, and the degree to which common risk variants may impact each family could vary between different families. Since the environmental factors for families have also not been systematically assessed here, integrating rare genetic variation from whole sequencing studies with environmental influences in future analyses could further elucidate the role of rare variation and environmental influences on the differences in recurrence risk of SCZ in multiplex SCZ families versus families of singleton SCZ cases from the population. Finally, as more samples from under-represented populations are collected, it is essential to replicate these findings in other more diverse populations.

Figures:

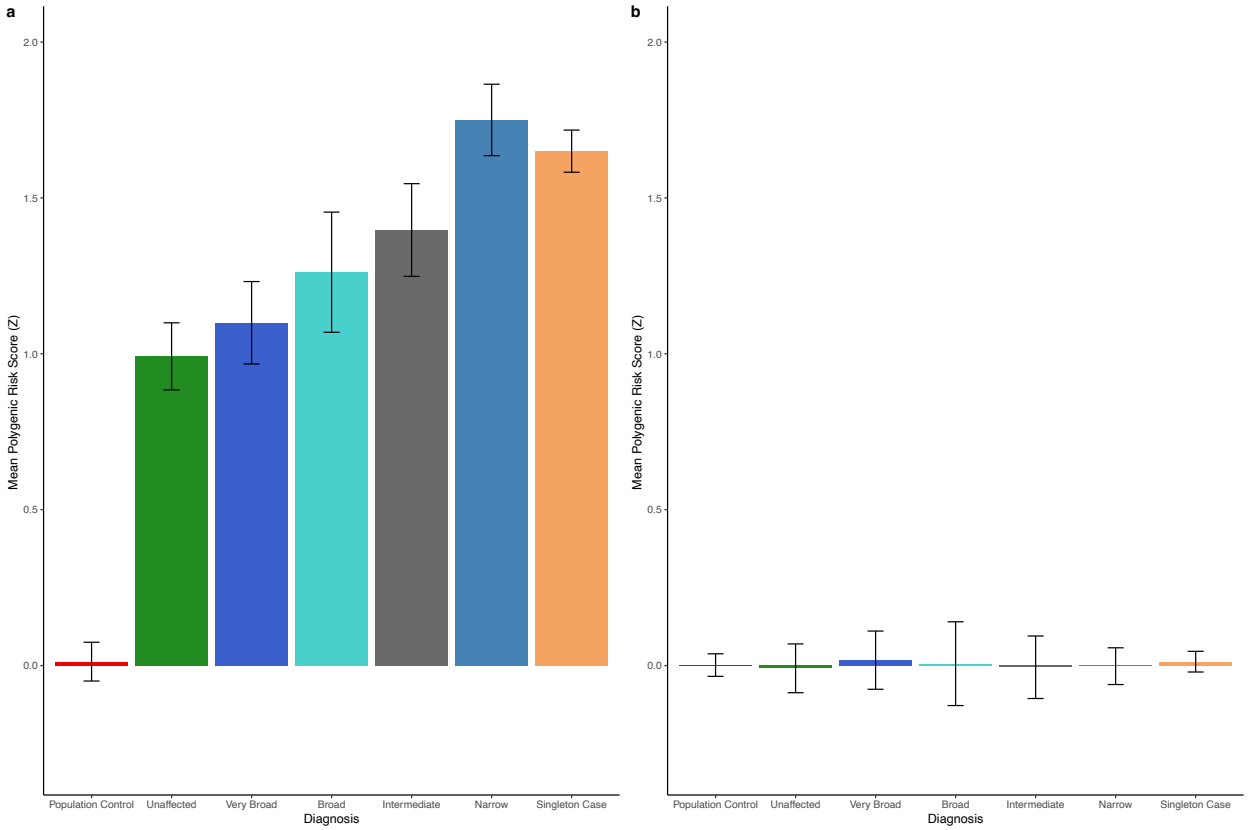

Figure 1: Mean PRS for population controls, singleton cases and each of the diagnostic categories of the individuals from the ISHDSF. Error bars represent the Standard error of the observed mean. **a**. PGC3-SCZ PRS. **b**. LDL PRS used as a negative control.

Tables:

Irish Study of High-Density Schizophrenia Families (ISHDSF)															
								Broad		Very Broad					
	Narrow		Intermediate												
Core Schizophrenia	Schizoaffecvtive Disorder	Psychosis NOS	Schizophreniform Disorder	Delusional Disorder	Schizotypal Personality Disorder	Avoidant Personality Disorder	Paranoid Personality Disorder	Schizoid Personality Disorder	Bipolar Disorder	Major Depressive Disorder	Alcohol Dependence	Generalized Anxiety Disorder	Panic Disorder	Other*	Unaffected Relatives
402	60	6													
	30	35	10	7	31										
						9	2	2	17	22					
									4	80	27	21	6	2	
															232
402	90	41	10	7	31	9	2	2	21	102	27	21	6	2	232

Unaffected Relatives: [1] There are 4 individuals with Intellectual Disability (ID) in the unaffected relatives category
Other: * Other Diagnoses includes Anorexia (1) and Cyclothymia (1)

Table 1: Sample description of ISHDSF**.** The concentric diagnostic hierarchy of ISHDSF contains 4 case definitions: *narrow*, *Intermediate, Broad* and *Very Broad.* These case definitions in ISHDSF reflect core (narrow and intermediate) and periphery (broad and very broad) of the psychosis spectrum based on previous genetic epidemiology work. Poor-outcome schizoaffective probands are placed in the narrow category (60), whereas good-outcome schizoaffective probands are placed in the intermediate category (30).

Comparison	Unadjusted P-Value	Holm-adjusted P-value
PGC3-SCZ		
Familial cases vs Singleton cases	0.12	0.49
PGC3-SCZ		
Singleton Cases vs controls	3.20E-94	6.720E-92
Narrow vs controls	4.04E-87	8.080e-86
Intermediate vs controls	5.24E-28	9.432E-27
Broad vs controls	2.08E-13	2.912E-12
Very Broad vs controls	1.07E-22	1.819E-21
Unaffected vs controls	1.07E-29	3.040E-28
LDL		
Singleton Cases vs controls	0.88	1
Narrow vs controls	0.86	1
Intermediate vs controls	0.14	1
Broad vs controls	0.15	1
Very Broad vs controls	0.85	1
Unaffected vs controls	0.95	1

Table 2: Comparison of the PRS in singleton cases and different diagnostic categories of ISHDSF from mixed-effects logistic regression models. The first row follows the hypothesis that familial cases have higher PRS compared to singleton cases. The comparisons under **PGC3-SCZ** follow the hypothesis that singleton cases and familial categories in ISHDSF have a higher PRS for SCZ compared to population controls. The comparisons under **LDL** are used as a negative control and follow the same hypothesis that singleton cases and familial categories in ISHDSF have a higher PRS for LDL compared to population controls.

References:

1. Saha S, Chant D, McGrath J. A systematic review of mortality in schizophrenia: Is the differential mortality gap worsening over time? *Arch Gen Psychiatry*. 2007. doi:10.1001/archpsyc.64.10.1123

2. Cardno AG, Gottesman II. Twin studies of schizophrenia: From bow-and-arrow concordances to Star Wars Mx and functional genomics. *Am J Med Genet*. 2000;97(1):12-17. doi:10.1002/(sici)1096-8628(200021)97:1<12::aid-ajmg3>3.3.co;2-l

3. Heston LL. Psychiatric disorders in foster home reared children of schizophrenic mothers. *Br J Psychiatry*. 1966;112(489):819-825. doi:10.1192/bjp.112.489.819

4. Cannon TD, Kaprio J, Lönnqvist J, Huttunen M, Koskenvuo M. The genetic epidemiology of schizophrenia in a Finnish twin cohort: A population-based modeling study. *Arch Gen Psychiatry*. 1998;55(1):67-74. doi:10.1001/archpsyc.55.1.67

5. Tienari P, Wynne LC, Moring J, et al. Finnish adoptive family study: Sample selection and adoptee DSM-III-R diagnoses. *Acta Psychiatr Scand*. 2000;101(6):433-443. doi:10.1034/j.1600-0447.2000.101006433.x

6. Kendler KS, Gruenberg AM, Tsuang MT. Psychiatric Illness in First-degree Relatives of Schizophrenic and Surgical Control Patients: A Family Study Using DSM-III Criteria. *Arch Gen Psychiatry*. 1985;42(8):770-779. doi:10.1001/archpsyc.1985.01790310032004

7. Walder DJ, Faraone S V., Glatt SJ, Tsuang MT, Seidman LJ. Genetic liability, prenatal health, stress and family environment: Risk factors in the Harvard Adolescent Family High Risk for Schizophrenia Study. *Schizophr Res*. 2014;157(1-3):142-148. doi:10.1016/j.schres.2014.04.015

8. Käkelä J, Panula J, Oinas E, Hirvonen N, Jääskeläinen E, Miettunen J. Family history of psychosis and social, occupational and global outcome in schizophrenia: A meta-analysis. *Acta Psychiatr Scand*. 2014;130(4):269-278. doi:10.1111/acps.12317

9. Esterberg ML, Trotman HD, Holtzman C, Compton MT, Walker EF. The impact of a family history of psychosis on age-at-onset and positive and negative symptoms of schizophrenia: A meta-analysis. *Schizophr Res*. 2010;120(1-3):121-130. doi:10.1016/j.schres.2010.01.011

10. Riley B, Thiselton D, Maher BS, et al. Replication of association between schizophrenia and ZNF804A in the Irish Case-Control Study of Schizophrenia sample. *Mol Psychiatry*. 2010;15(1):29-37. doi:10.1038/mp.2009.109

11. Kendler KS, McGuire M, Gruenberg AM, O'Hare A, Spellman M, Walsh D. The Roscommon Family Study: I. Methods, Diagnosis of Probands, and Risk of Schizophrenia

in Relatives. *Arch Gen Psychiatry*. 1993;50(7):527-540. doi:10.1001/archpsyc.1993.01820190029004

12. Kendler KS, McGuire M, Gruenberg AM, Spellman M, O'Hare A, Walsh D. The Roscommon Family Study: II. The Risk of Nonschizophrenic Nonaffective Psychoses in Relatives. *Arch Gen Psychiatry*. 1993;50(8):645-652. doi:10.1001/archpsyc.1993.01820200059006

13. Kendler KS, McGuire M, Gruenberg AM, O'Hare A, Spellman M, Walsh D. The Roscommon Family Study: III. Schizophrenia-Related Personality Disorders in Relatives. *Arch Gen Psychiatry*. 1993;50(10):781-788. doi:10.1001/archpsyc.1993.01820220033004

14. Kendler KS, McGuire M, Gruenberg AM, O'Hare A, Spellman M, Walsh D. The Roscommon Family Study: IV. Affective Illness, Anxiety Disorders, and Alcoholism in Relatives. *Arch Gen Psychiatry*. 1993;50(12):952-960. doi:10.1001/archpsyc.1993.01820240036005

15. Ament SA, Szelinger S, Glusman G, et al. Rare variants in neuronal excitability genes influence risk for bipolar disorder. *Proc Natl Acad Sci U S A*. 2015;112(11):3576-3581. doi:10.1073/pnas.1424958112

16. Cruceanu C, Schmouth JF, Torres-Platas SG, et al. Rare susceptibility variants for bipolar disorder suggest a role for G protein-coupled receptors. *Mol Psychiatry*. 2018;23(10):2050-2056. doi:10.1038/mp.2017.223

17. Goes FS, Pirooznia M, Parla JS, et al. Exome sequencing of familial bipolar disorder. *JAMA Psychiatry*. 2016;73(6):590-597. doi:10.1001/jamapsychiatry.2016.0251

18. Okayama T, Hashiguchi Y, Kikuyama H, Yoneda H, Kanazawa T. Next-generation sequencing analysis of multiplex families with atypical psychosis. *Transl Psychiatry*. 2018;8(1). doi:10.1038/s41398-018-0272-x

19. Toma C, Torrico B, Hervás A, et al. Exome sequencing in multiplex autism families suggests a major role for heterozygous truncating mutations. *Mol Psychiatry*. 2014;19(7):784-790. doi:10.1038/mp.2013.106

20. Homann OR, Misura K, Lamas E, et al. Whole-genome sequencing in multiplex families with psychoses reveals mutations in the SHANK2 and SMARCA1 genes segregating with illness. *Mol Psychiatry*. 2016;21(12):1690-1695. doi:10.1038/mp.2016.24

21. Ripke S, Sanders AR, Kendler KS, et al. Genome-wide association study identifies five new schizophrenia loci. *Nat Genet*. 2011;43(10):969-978. doi:10.1038/ng.940

22. Ripke S, Neale BM, Corvin A, et al. Biological insights from 108 schizophrenia-associated genetic loci. *Nature*. 2014. doi:10.1038/nature13595

23. Consortium SWG of the PG, Ripke S, Walters JT, O'Donovan MC. Mapping genomic loci prioritises genes and implicates synaptic biology in schizophrenia. *medRxiv*. 2020:2020.09.12.20192922.

24. de Jong S, Diniz MJA, Saloma A, et al. Applying polygenic risk scoring for psychiatric disorders to a large family with bipolar disorder and major depressive disorder. *Commun Biol*. 2018;1(1). doi:10.1038/s42003-018-0155-y

25. Andlauer TFM, Guzman-Parra J, Streit F, et al. Bipolar multiplex families have an increased burden of common risk variants for psychiatric disorders. *Mol Psychiatry*. 2019. doi:10.1038/s41380-019-0558-2

26. Szatkiewicz J, Crowley JJ, Adolfsson AN, et al. The genomics of major psychiatric disorders in a large pedigree from Northern Sweden. *Transl Psychiatry*. 2019;9(1). doi:10.1038/s41398-019-0414-9

27. Bigdeli TB, Bacanu SA, Webb BT, et al. Molecular validation of the schizophrenia spectrum. *Schizophr Bull*. 2014;40(1):60-65. doi:10.1093/schbul/sbt122

28. Kendler KS, O'Neill FA, Burke J, et al. Irish study of high-density schizophrenia families: Field methods and power to detect linkage. *Am J Med Genet - Semin Med Genet*. 1996;67(2):179-190. doi:10.1002/(SICI)1096-8628(19960409)67:2<179::AID-AJMG8>3.0.CO;2-N

29. Levinson DF, Shi J, Wang K, et al. Genome-wide association study of multiplex schizophrenia pedigrees. *Am J Psychiatry*. 2012;169(9):963-973. doi:10.1176/appi.ajp.2012.11091423

30. Donnelly P, Barroso I, Blackwell JM, et al. Genome-wide association study implicates HLA-C*01:02 as a risk factor at the major histocompatibility complex locus in schizophrenia. *Biol Psychiatry*. 2012;72(8):620-628. doi:10.1016/j.biopsych.2012.05.035

31. Colhoun HM, McKeigue PM, Smith GD. Problems of reporting genetic associations with complex outcomes. *Lancet*. 2003;361(9360):865-872. doi:10.1016/S0140-6736(03)12715-8

32. Loh PR, Danecek P, Palamara PF, et al. Reference-based phasing using the Haplotype Reference Consortium panel. *Nat Genet*. 2016;48(11):1443-1448. doi:10.1038/ng.3679

33. McCarthy S, Das S, Kretzschmar W, et al. A reference panel of 64,976 haplotypes for genotype imputation. *Nat Genet*. 2016;48(10):1279-1283. doi:10.1038/ng.3643

34. Das S, Forer L, Schönherr S, et al. Next-generation genotype imputation service and methods. *Nat Genet*. 2016;48(10):1284-1287. doi:10.1038/ng.3656

35. Ge T, Chen CY, Ni Y, Feng YCA, Smoller JW. Polygenic prediction via Bayesian

regression and continuous shrinkage priors. *Nat Commun*. 2019;10(1). doi:10.1038/s41467-019-09718-5

36. Clarke L, Fairley S, Zheng-Bradley X, et al. The international Genome sample resource (IGSR): A worldwide collection of genome variation incorporating the 1000 Genomes Project data. *Nucleic Acids Res*. 2017;45(D1):D854-D859. doi:10.1093/nar/gkw829

37. Purcell SM, Wray NR, Stone JL, et al. Common polygenic variation contributes to risk of schizophrenia and bipolar disorder. *Nature*. 2009;460(7256):748-752. doi:10.1038/nature08185

38. Ni G, Zeng J, Revez JA, et al. A comparison of ten polygenic score methods for psychiatric disorders applied across multiple cohorts. *Biol Psychiatry*. 2021. doi:10.1016/j.biopsych.2021.04.018

39. Bulik-Sullivan BK, Neale BM. LD Score Regression Distinguishes Confounding from Polygenicity in GWAS. *Nat Genet*. 2015;47(3):291-295. doi:10.1038/ng.3211.LD

40. Surakka I, Horikoshi M, Mägi R, et al. The impact of low-frequency and rare variants on lipid levels. *Nat Genet*. 2015;47(6):589-597. doi:10.1038/ng.3300

41. Bulik-Sullivan B, Finucane HK, Anttila V, et al. An atlas of genetic correlations across human diseases and traits. *Nat Genet*. 2015;47(11):1236-1241. doi:10.1038/ng.3406

42. Zheng J, Erzurumluoglu AM, Elsworth BL, et al. LD Hub: A centralized database and web interface to perform LD score regression that maximizes the potential of summary level GWAS data for SNP heritability and genetic correlation analysis. *Bioinformatics*. 2017;33(2):272-279. doi:10.1093/bioinformatics/btw613

43. Chen H, Wang C, Conomos MP, et al. Control for Population Structure and Relatedness for Binary Traits in Genetic Association Studies via Logistic Mixed Models. *Am J Hum Genet*. 2016;98(4):653-666. doi:10.1016/j.ajhg.2016.02.012

44. R Development Core Team R. *R: A Language and Environment for Statistical Computing*. Vol 1.; 2011. doi:10.1007/978-3-540-74686-7

45. Speed D, Hemani G, Johnson MR, Balding DJ. Improved heritability estimation from genome-wide SNPs. *Am J Hum Genet*. 2012;91(6):1011-1021. doi:10.1016/j.ajhg.2012.10.010

46. Fromer M, Pocklington AJ, Kavanagh DH, et al. De novo mutations in schizophrenia implicate synaptic networks. *Nature*. 2014;506(7487):179-184. doi:10.1038/nature12929

47. Rees E, Han J, Morgan J, et al. De novo mutations identified by exome sequencing implicate rare missense variants in SLC6A1 in schizophrenia. *Nat Neurosci*. 2020;23(2):179-184. doi:10.1038/s41593-019-0565-2

48. Howrigan DP, Rose SA, Samocha KE, et al. Exome sequencing in schizophrenia-affected parent–offspring trios reveals risk conferred by protein-coding de novo mutations. *Nat Neurosci*. 2020;23(2):185-193. doi:10.1038/s41593-019-0564-3

49. Singh T, Poterba T, Curtis D, et al. Exome sequencing identifies rare coding variants in 10 genes which confer substantial risk for schizophrenia. *medRxiv*. 2020:2020.09.18.20192815.

50. Auton A, Abecasis GR, Altshuler DM, et al. A global reference for human genetic variation. *Nature*. 2015;526(7571):68-74. doi:10.1038/nature15393

51. MacArthur DG, Balasubramanian S, Frankish A, et al. A systematic survey of loss-of-function variants in human protein-coding genes. *Science (80-)*. 2012;335(6070):823-828. doi:10.1126/science.1215040

52. Morris DW, Pearson RD, Cormican P, et al. An inherited duplication at the gene p21 protein-activated Kinase 7 (PAK7) is a risk factor for psychosis. *Hum Mol Genet*. 2014;23(12):3316-3326. doi:10.1093/hmg/ddu025

53. Marshall CR, Howrigan DP, Merico D, et al. Contribution of copy number variants to schizophrenia from a genome-wide study of 41,321 subjects. *Nat Genet*. 2017;49(1):27-35. doi:10.1038/ng.3725