Title page

Title: Polygenic Risk Scores Provide Strong Evidence for the Biologic Relevance of Molecular Subtypes of Schizophrenia

Authors: C. Harker Rhodes, MD, PhD[1, 2]
Richard Granger, PhD[3]

Author Affiliations: [1] Henry M. Jackson Foundation for the Advancement of Military Medicine
[2] Department of Defense Brain Tissue Repository
[3] Department of Psychological and Brain Sciences, Dartmouth College

Contact Information:
C. Harker Rhodes, MD, PhD 144
Sunset Rock Rd Lebanon, NH 03766

Tel: 603-443-3360 FAX:
E-mail: CHarkerRhodes@gmail.com

Molecular Subtypes of Schizophrenia

Abstract

We have previously described how the post-mortem DLPFC transcriptomes of schizophrenia patients can be used to define two patient groups: "Type 1" patients with a relatively normal DLPFC transcriptome and "Type 2" patients with hundreds of genes differentially expressed in the DLPFC at a statistical significance which survives Bonferroni correction. The biological significance of the distinction between Type 1 and Type 2 patients remains an open question. In the present report we examine the schizophrenia polygenic risk scores (PRS) in those two groups of schizophrenic patients and observe that PRS account for twice the variance in schizophrenia risk in the patients with a relatively normal DLPFC transcriptome ("Type 1") compared to those with dramatically abnormal DLPFC transcriptomes ("Type 2"). That result supports the hypothesis that the DLPFC transcriptome-defined schizophrenia subtypes are biologically meaningful.

As important as that result is, a serendipitous observation involving patient ancestry is potentially more important because it suggests an approach which future studies might use to investigate the pathogenic pathways leading to these two distinct forms of schizophrenia. Summary statistics from the Psychiatric Genetics Consortium (PGC) wave 3 core cohort GWAS data were used to calculate schizophrenia PRS. That cohort is composed primarily of Caucasians with about 13% Asians. It contains no Africans or African Americans. However, the cohort we studied here (the NIMH Human Brain Collection Core, or HBCC cohort) is composed of roughly 50% Caucasians and 50% African Americans. When the HBCC cohort is stratified based on ancestry it becomes obvious that in African Americans the polygenic risk scores based on summary statistics from the (Caucasian) PGC core cohort capture the genetic risk associated with Type 1, but not Type 2 schizophrenia. Future work comparing summary statistics from schizophrenia GWAS studies of Caucasian and African American populations should reveal the genes and genetic pathways which are specifically involved in the pathogenesis of the two types of schizophrenia

Disclosures: The opinions expressed herein are those of the authors and not necessarily representative of those of the Uniformed Services University of the Health Sciences (USU), the Department of Defense (DOD), the United States Army, Navy, Air Force, VA, NIH or any other US federal agency.

Molecular Subtypes of Schizophrenia

Introduction

The popular “neurodevelopmental” hypothesis for the pathogenesis of schizophrenia postulates that schizophrenia is the result of some early pathogenic event occurring around the time of birth which predisposes the patient to the disease without producing clinical symptoms, and that the early event is followed by later event(s) during adolescence which trigger the appearance of clinical symptoms (see, for example, [1]). While the early and late events are presumably both due to a combination of environmental and genetic factors, the specific environmental and genetic factors are almost certainly different for the two events.

An alternate formulation of the neurodevelopmental hypothesis views schizophrenia as a syndrome, not a single disease. According to this alternate hypothesis there are two distinct diseases which can produce the schizophrenic phenotype, one of which is a result of perinatal event(s) while a second results from pathogenic event(s) in adolescence. Just as dementia often results from a combination of Alzheimer's disease and vascular dementia, so the schizophrenic phenotype often results from the comorbid presence of both forms of schizophrenia. This formulation of the neurodevelopmental hypothesis predicts that schizophrenic patients lie on a biologic spectrum with some patients for whom the schizophrenic phenotype is due primarily to the perinatal event, while for other patients it is the primarily the adolescent event which is driving the disease pathogenesis.

These two distinct diseases are predicted to respond differently to disease-specific therapies; a failure to appropriately stratify patients may produce negative results in therapeutic trials of what otherwise might be effective new drugs.

This alternate formulation of the neurodevelopmental hypothesis builds on three previous observations:

1) The observation that schizophrenic patients can be divided into two groups based on the extent to which their DLPFC transcriptomes are abnormal [2],

2) The observation that schizophrenic patients can be divided into two groups based on inflammatory markers in the brain [3-5], and

3) The observation that the schizophrenic patients with a relatively normal DLPFC transcriptome are also those with low levels of CNS inflammatory markers which those with hundreds of differentially expressed genes in the DLPFC are those with high levels of inflammatory markers (manuscript in preparation).

In other words, there appear to be two biologically distinct groups of schizophrenic patients; the extent to which this distinction represents a dichotomy or a continuum is unknown.

Genome-wide studies of large cohorts have identified many genetic polymorphisms associated with an increased risk of schizophrenia, and polygenic risk scores (PRS) based on those studies have been shown to account for about 10% of the variance in risk for schizophrenia (see, for example the Psychiatric Genetics Consortium wave 3 results [6]). Presumably those PRS include both polymorphisms affecting the risk of the perinatal event (possibly polymorphisms affecting placental function) and polymorphisms affecting the neurologic functioning of the adult brain. However, the relative extent to which the PRS as currently formulated predict the risk of the perinatal vs the risk of the adolescent events is unknown.

In this study we use schizophrenia PRS based on the Psychiatric Genetics Consortium wave 3 summary statistics to characterize a cohort in which the schizophrenic patients were previously stratified as either "Type 1" (patients with a relatively normal DLPFC transcriptome) or "Type 2" (patients with thousands of differentially expressed genes in their DLPFC). If the extent to which PRS explains the variance in risk for schizophrenia differs in those two groups of patients, it will be strong evidence for the biologic relevance of this alternate formulation of the neurodevelopmental hypothesis.

Methods

Sources of data

The patient cohort for this study is the autopsy cohort originally collected by the former Clinical Brain Disorders Branch (CBDB) of the NIMH intramural program under the direction of Dr. Weinberger and subsequently and overlapping cohorts subsequently expanded by investigators at both the NIMH Human Brain Collection Core (HBCC) and the Lieber Institute.

Expression array data from the DLPFC of those patients collected by the CBDB is publicly available from dbGaP (study accession phs000979.) and RNAseq data from the DLPFC and caudate is available from investigators at the Lieber Institute. The separation of the schizophrenic patients in this cohort into two groups, "Type 1" patients with a relatively normal DLPFC transcriptome and "Type 2" patients with hundreds of genes differentially expressed in the DLPFC has been previously described [2]. Similar results can be obtained using the RNAseq data from the DLPFC, but because those results are unpublished, this study is done using the published array data based stratification of the cohort.

The genetically imputed ancestry of the subjects in this cohort was published recently [7] and is publicly available from the HBCC.

The calculation of PRS data for this cohort has been published recently [7] and is publicly available from the HBCC. Those PRS scores were calculated based on summary statistics from the Psychiatric Genetics Consortium wave 3 results [6] which are publicly available at https://pgc.unc.edu/for-researchers/download-results/.

Data analysis

This study was done using R version 4.3.2 (2023-10-31 ucrt; "Eye Holes") running under Windows 10 x64 (build 19045). Recognizing the overwhelming importance of ancestry as a covariate, the PRS data was adjusted for ancestry (Caucasian vs African American) using robust linear regression. In everything which follows, "PRS data" refers to that ancestry-adjusted PRS data. The fraction of the schizophrenia risk variance attributable to PRS was calculated using rlm(){MASS} to train a robust logistic linear regression model followed by NagelkirkeR2(){fmbs} to calculate Nagelkirke's pseudo-R^2 [8]. The statistical significance of the difference between Nagelkirke's pseudo-R^2 values calculated for different groups of patients was estimated by a Monte Carlo method. The R code for the data analysis reported here is on request.

Results

Cohort demographics

This cohort is a convenience sample based on Medical Examiner cases for whom the next of kin consented to the use of post-mortem tissue for this purpose. It is, therefore, not necessarily representative of the general population and this study limitation needs to be kept in mind when interpreting these

results.

There were six self-reported Caucasians and four self-reported African Americans whose genetically imputed ancestry differed from their self-reported ancestry; those subjects were excluded from further analysis. Those individuals were excluded from further analysis. Because there were so few individuals with self-reported Hispanic or Asian ancestry they were also excluded from further analysis. The final cohort is composed of individuals whose self-reported and genetically imputed ancestry is concordant and is either Caucasian or African American. The final cohort is well balanced for diagnosis and race (table 1). Additional cohort demographics are described in [2].

	Controls	Schizophrenics	
		Type 1	Type 2
Caucasians	82 (47%)	42 (52%)	48 (58%)
African Americans	92 (53%)	38 (48%)	35 (42%)

Table 1: Distribution of cohort subjects by ancestry and diagnosis. Most importantly, the percentages of Type 1 and Type 2 patients is the roughly same in the Caucasians and in the African Americans.

Variance in risk of schizophrenia accounted for by PRS

(The primary result of this study)

As is customary in studies of PRS, Duncan et al. (2023) [7] calculated the PRS data for this cohort using polymorphisms selected based on a range of p-value thresholds varying from $P \geq 1.0$ (all polymorphisms nominally associated with risk of schizophrenia) all the way down to $P < 10^{-8}$ (genome wide significance). That data is shown in Figure 1.

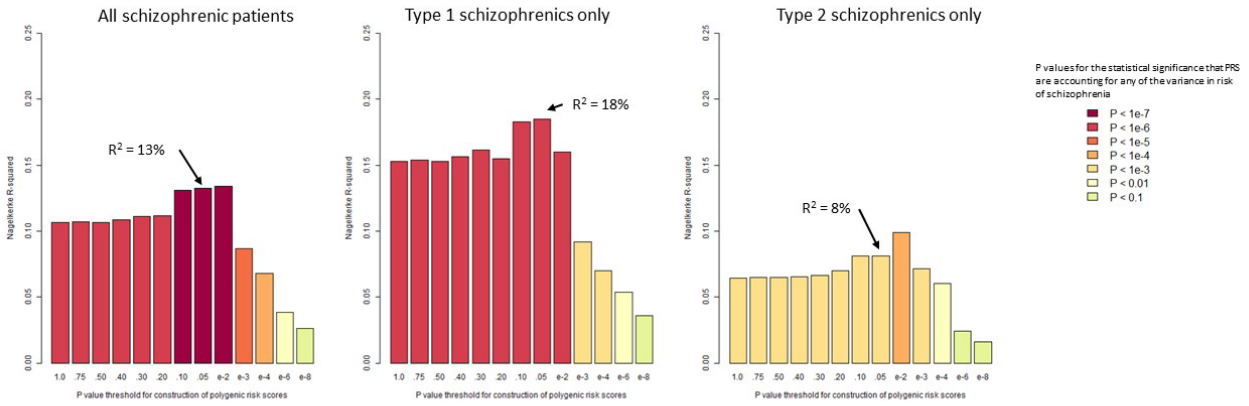

Figure 1: The fraction of variance in schizophrenia risk accounted for by PRS.
The ordinate (the bar heights) is Nagelkirke's pseudo-R^2, a measure of the fraction of the variance accounted for by PRS. The statistial significancs of the logisitic regression on which the Nagelkirke's pseudo-R^2 is based is indicated by the bar colors.

As is customary in studies of this sort and to be consistent with Duncan et al. (2023) [7] in their interpretation of their own data, we will use the PRS data calculated using a $P_{threshold}$ of 0.05 (arrows) as a summary measure of the Nagelkirke's pseudo-R^2.

All three values of R^2 calculated using a $P_{threshold}$ of 0.05 are statistically different from the other two ($P < 0.05$).

The effect of stratification of cohort based on ancestry

(The serendipitous observation)

The PRS used here is based on the best available schizophrenia risk GWAS data, the PGC wave 3 data [6]. The cohort on which that study was based is composed of 81% Causasians, 19% East Asians, and no African Americans. When the subjects in the present study were stratified by ancestry separately (figure 2) we made the serendipitous observation that this PRS, based on primarily Caucasians, captures the genetic risk of Type 1 schizophrenia in African Americans very well, but captures virtually none of the risk of Type 2 schizophrenia in African Americans.

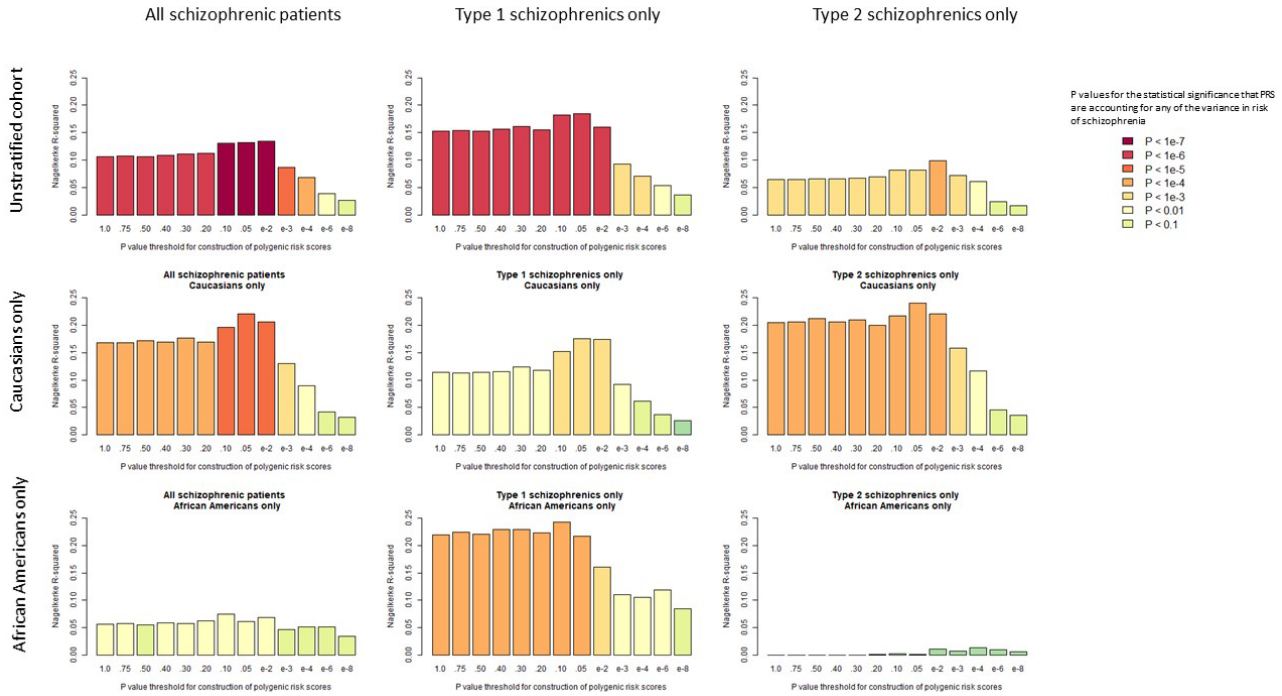

Figure 2: PRS contribution to the risk of schizophrenia when the cohort is stratified by ancestry.

We call this unexpected observation serendipitous because the fact that any PRS, no matter how it was derived, can so successfully differentiate Type 1 from Type 2 schizophrenia is very strong evidence that these are two distinct diseases. But perhaps more importantly, it suggests the possibility that PRS studies based on a variety of different summary GWAS statistics might reveal genes or genetic pathways responsible for the two different types of schizophrenia.

Discussion

We previously described the use of gene expression array data from the dorsolateral prefrontal cortex of schizophrenic patients can be used to identify two groups of patients, those with a relatively normal DLPFC transcriptome ("Type 1") and those with hundreds of genes differentially expressed at statistical significance levels which survive Bonferroni correction ("Type 2) [2]. This is not to say that the Type 1 patients have a completely normal DLPFC transcriptome - the Bonferroni criterion is much more stringent than the criteria usually used in studies of this sort. But at any level of statistical significance, the differences between the Type 1 and Type 2 patients are striking. Subsequent work has shown that similar results are obtained using RNAseq data rather than expression array data and examining tissue from either the DLPFC or the caudate nucleus [9]. Those results rule out the possibility of the earlier results being due to some artifact of expression array data, some systematic difference in the sampling of the DLPFC, or some other unrecognized confound. It is therefore essential that the distinction between Type 1 and Type 2 schizophrenic patients be validated by some other, unrelated dataset.

The most important result of this study is that without stratification for ancestry the best polygenic risk scores for schizophrenia currently available explain 18% of the variance in risk for schizophrenia for type 1 patients, but only 6% for type 2 patients. That result is statistically significant at a level of $P < 0.002$ and strongly supports the hypothesis that the distinction between Type 1 and Type 2 schizophrenia is biologically important.

The observation was also made that when the PRS is based on data from a cohort composed primarily of Caucasians and containing no African Americans the PRS captures the risk of Type 1, but not Type 2 schizophrenia in African Americans. That further observation does NOT alter the conclusion that this PRS data supports the hypothesis that there are important biologic differences between Type 1 and Type 2 schizophrenia. In fact the amazing difference in the genetic risk between Type 1 and Type 2 disease seen in African Americans is pretty conclusive evidence of biologic differences between these two types of schizophrenia.

It is likely that differences in allele frequencies and linkage distribution structure between the Caucasians and African Americans accounts for at least some of the difference seen when the HBCC cohort is stratified based on ancestry. However, it is unlikely that this is accounting for all of the dramatic effect observed. In this context it is important to remember that in the United States ancestry is to a certain extent a proxy for socioeconomic-associated environmental exposures. There is an enormous literature which we are not able to review or critically evaluate here demonstrating differences in obstetric care and pregnancy complications between Caucasian and Black Americans. If there are specific obstetric complications that in the US population occur primarily in African Americans, there may be entire genetic pathways that are relevant in one group and not the other. This is entirely speculative at this point, but it leads to testable hypotheses which will be the subject of future study.

An obvious difficulty in translating this scientific observation into the clinical arena is that the DLPFC transcriptomic data on which the Type 1 / Type 2 distinction is currently based requires autopsy tissue. However the sphingosine phosphate receptor S1PR1 is differentially expressed at both the mRNA and protein levels in the brains of Type 1 vs Type 2 schizophrenic patients [10], and PET using S1PR1 is already FDA approved for human studies [11-14].

Acknowledgements

First and foremost we would like to acknowledge the families of the subjects in this study for consenting to

the study of this autopsy tissue and for providing clinical information to help establish the psychiatric diagnoses.

We would also like to acknowledge the current and former scientific staff of the Human Brain Collection Core (HBCC) and its predecessor, the former Clinical Brain Disorders Branch (CBDB) of the Intramural Program of the NIMH for collecting the clinical material and expression array data and for making it publicly available on dbGaP. Similarly, we are grateful to the members of the CMC consortium for and to the staff at the Lieber Institute (Baltimore) making their RNAseq data available. And finally, we are grateful to the members of the Duncan lab (Stanford University) for making the PRS data they calculated available.

CHR is grateful for support from the Henry M. Jackson Foundation, Inc. and the Center for Neuroscience and the Department of Defense Brain Tissue Resource.

Literature cited

1. Weinberger, D.R., *On the plausibility of "the neurodevelopmental hypothesis" of schizophrenia.* Neuropsychopharmacology, 1996. **14**(3 Suppl): p. 1S-11S.

2. Bowen, E.F.W., et al., *DLPFC transcriptome defines two molecular subtypes of schizophrenia.* Transl Psychiatry, 2019. **9**(1): p. 147.

3. Zhu, Y., et al., *Inflammation-related transcripts define "high" and "low" subgroups of individuals with schizophrenia and bipolar disorder in the midbrain.* Brain Behav Immun, 2022. **105**: p. 149-159.

4. Zhu, Y., et al., *Distinct Phenotypes of Inflammation Associated Macrophages and Microglia in the Prefrontal Cortex Schizophrenia Compared to Controls.* Front Neurosci, 2022. **16**: p. 858989.

5. Zhu, Y., et al., *Increased prefrontal cortical cells positive for macrophage/microglial marker CD163 along blood vessels characterizes a neuropathology of neuroinflammatory schizophrenia.* Brain Behav Immun, 2023. **111**: p. 46-60.

6. Trubetskoy, V., et al., *Mapping genomic loci implicates genes and synaptic biology in schizophrenia.* Nature, 2022. **604**(7906): p. 502-508.

7. Duncan, L., et al., *Polygenic scores for psychiatric disorders in a diverse postmortem brain tissue cohort.* Neuropsychopharmacology, 2023. **48**(5): p. 764-772.

8. NAGELKERKE, N.J.D., *A note on a general definition of the coefficient of determination.* Biometrika, 1991. **78**(3): p. 691-692.

9. Rhodes, C.H., et al., *Molecular Taxonomy of Schizophrenia*. 2023, Dartmouth College: 2023 Winter Conference on Brain Research.

10. Chand, G.B., et al., *Differential Sphingosine-1-Phosphate Receptor-1 Protein Expression in the Dorsolateral Prefrontal Cortex Between Schizophrenia Type 1 and Type 2.* Front Psychiatry, 2022. **13**: p. 827981.

11. Qiu, L., et al., *Metabolite Study and Structural Authentication for the First-in-Human Use Sphingosine-*

1-phosphate Receptor 1 Radiotracer. ACS Chem Neurosci, 2024. **15**(9): p. 1882-1892.

12. Qiu, L., et al., *Design, synthesis, and biological evaluation of multiple F-18 S1PR1 radiotracers in rodent and nonhuman primate.* Org Biomol Chem, 2024.

13. Qiu, L., et al., *Discovery of a Promising Fluorine-18 Positron Emission Tomography Radiotracer for Imaging Sphingosine-1-Phosphate Receptor 1 in the Brain.* J Med Chem, 2023. **66**(7): p. 4671-4688.

14. Brier, M.R., et al., *Phase 1 Evaluation of (11)C-CS1P1 to Assess Safety and Dosimetry in Human Participants.* J Nucl Med, 2022. **63**(11): p. 1775-1782.