3-16-2021

Genome-Wide Association Studies of Schizophrenia and Bipolar Disorder in a Diverse Cohort of US Veterans

Tim B. Bigdeli
Ayman H. Fanous
Yuli Li
Nallakkandi Rajeevan
Frederick Sayward

See next page for additional authors

Follow this and additional works at: https://scholarworks.utrgv.edu/som_pub

Part of the Medicine and Health Sciences Commons

Recommended Citation

This Article is brought to you for free and open access by the School of Medicine at ScholarWorks @ UTRGV. It has been accepted for inclusion in School of Medicine Publications and Presentations by an authorized administrator of ScholarWorks @ UTRGV. For more information, please contact justin.white@utrgv.edu, william.flores01@utrgv.edu.
Genome-Wide Association Studies of Schizophrenia and Bipolar Disorder in a Diverse Cohort of US Veterans

Tim B. Bigdeli,1,2 Ayman H. Fanous,1,2 Yuli Li,3,4 Nallakkandi Rajeevan,1,4 Frederick Sayward,1,4 Giulio Genovese,5,6 Rishab Gupta,1,4 Krishnan Radhakrishnan,5,7 Anil K. Malhotra,8–10 Ning Sun,5,7 Qiongshi Lu,9 Yiming Hu,9 Boyang Li,10 Quan Chen,10 Shrikant Mane,10 Perry Miller,1,4 Kei-Hoi Cheung,11 Raquel E. Gur,12 Tiffany A. Greenwood,13 David L. Braff,13,14 Consortium on the Genetics of Schizophrenia (COGS)9, Eric D. Achtyes,15 Peter F. Buckley,16 Michael A. Escamilla17, Douglas Lehrer18, Dolores P. Malaspina19, Steven A. McCarroll5,6, Mark H. Rapaport20, Marquis P. Vawter21, Michele T. Pato,2 Carlos N. Pato,2 Genomic Psychiatry Cohort (GPC) Investigators, Hongyu Zhao,22, Thomas R. Kosten,21, Mary Brophy,20,24 Saiju Pyarajan,21 Yunling Shi,21 Timothy J. O’Leary,25 Theresa Gleason,22, Ronald Przygodzki,25, Sumitra Muralidhar,22 J. Michael Gaziano,23,26 Million Veteran Program (MVP),2 Grant D. Huang,25, John Concato,1,2 Larry J. Siever,19,27, Mihaela Aslan,5,6, and Philip D. Harvey*,28,29

1VA New York Harbor Healthcare System, Brooklyn, NY; 2Department of Psychiatry and Behavioral Sciences, SUNY Downstate Medical Center, Brooklyn, NY; 3Cooperative Studies Program Clinical Epidemiology Research Center (CSP-CERC), VA Connecticut Healthcare System, West Haven, CT; 4Department of Medicine, Yale School of Medicine, New Haven, CT; 5Stanley Center for Psychiatric Research, Broad Institute of MIT and Harvard, Cambridge, MA; 6Department of Genetics, Harvard Medical School, Boston, MA; 7College of Medicine, University of Kentucky, Lexington, KY; 8Center for Psychiatric Neuroscience, Feinstein Institute for Medical Research, Manhasset, NY; 9Division of Psychiatry Research, The Zucker Hillside Hospital, Northwell Health, Glen Oaks, NY; 10Department of Psychiatry, Hofstra Northwell School of Medicine, Hempstead, NY; 11Department of Biostatistics & Medical Informatics, University of Wisconsin-Madison, Madison, WI; 12Departments of Psychiatry and Child & Adolescent Psychiatry and Lifespan Brain Institute, University of Pennsylvania Perelman School of Medicine and Children’s Hospital of Philadelphia, Philadelphia, PA; 13Department of Psychiatry, University of California, Los Angeles, Los Angeles, CA; 14VISN-22 Mental Illness, Research, Education and Clinical Center (MIRECC), VA San Diego Healthcare System, San Diego, CA; 15Cherry Health and Michigan State University College of Human Medicine, Grand Rapids, MI; 16School of Medicine, Virginia Commonwealth University, Richmond, VA; 17Department of Psychiatry, School of Medicine, University of Texas Rio Grande Valley, Harlingen, TX; 18Department of Psychiatry, Wright State University, Dayton, OH; 19Department of Psychiatry, Icahn School of Medicine at Mount Sinai, New York, NY; 20Department of Psychiatry and behavioral Sciences, Emory University, Atlanta, GA; 21Department of Psychiatry and the Behavioral Science, University of California, Irvine, CA; 22Department of Psychiatry, Neuroscience, Pharmacology, and Immunology and Rheumatology, Baylor College of Medicine, Houston, TX; 23Massachusetts Area Veterans Epidemiology, Research, and Information Center (MAVERIC), Jamaica Plain, MA; 24Section of Hematology and Medical Oncology, Boston University School of Medicine, Boston, MA; 25Office of Research and Development, Veterans Health Administration, Washington, DC; 26Department of Medicine, Brigham and Women’s Hospital, Harvard Medical School, Boston, MA; 27University of Miami Miller School of Medicine, Miami, FL; 28Departments of Psychiatry, University of Miami Miller School of Medicine, Miami, FL

†Full consortium authorship is listed in supplementary acknowledgments.

*To whom correspondence should be addressed: Department of Psychiatry and Behavioral Sciences, University of Miami Miller School of Medicine, 1120 NW 14th Street, Suite 1450 Miami, FL 33136, USA; tel: (305)-243-4094, fax: (305)-243-1619, e-mail: PHarvey@med.miami.edu

Background: Schizophrenia (SCZ) and bipolar disorder (BIP) are debilitating neuropsychiatric disorders, collectively affecting 2% of the world’s population. Recognizing the major impact of these psychiatric disorders on the psychosocial function of more than 200 000 US Veterans, the Department of Veterans Affairs (VA) recently completed genotyping of more than 8000 veterans with SCZ and BIP in the Cooperative Studies Program (CSP) #572. Methods: We performed genome-wide association studies (GWAS) in CSP #572 and benchmarked the predictive value of polygenic risk scores (PRS) constructed from published findings. We combined our results with available summary statistics from several recent GWAS, realizing the largest and most diverse studies of these disorders to date. Results: Our primary GWAS uncovered new associations between CHD7 variants and SCZ, and novel BIP associations with variants in Sortilin Related VPS10 Domain Containing Receptor 3 (SORCS3) and downstream of PCDH11X. Combining our results with published summary statistics for SCZ yielded 39 novel susceptibility loci including CRH1, and we identified 10 additional findings for BIP (28 326 cases and 90 570 controls). PRS trained on published GWAS were
significantly associated with case-control status among European American (P < 10^-9) and African American (P < 0.0005) participants in CSP #572. Conclusions: We have demonstrated that published findings for SCZ and BIP are robustly generalizable to a diverse cohort of US veterans. Leveraging available summary statistics from GWAS of global populations, we report 52 new susceptibility loci and improved fine-mapping resolution for dozens of previously reported associations.

Key words: schizophrenia/bipolar disorder/genome-wide association studies (GWAS)/US veterans

Introduction

In the United States, the combined societal costs of treatment and loss of productivity associated with schizophrenia (SCZ) were estimated to be $62.7 billion in 20021 and more than $155 billion in 2013.2 For bipolar disorder (BIP), which is phenomenologically related and has a partially shared genetic basis, these costs exceeded $200 billion in 2015.5 Even during psychosis-free periods, cognitive deficits are thought to contribute to deteriorating functional skills in the areas of self-care, social and occupational function, and medication adherence, as well as other aspects of illness management, all needed for independent living.8 Recognizing the major impact of these disorders on the psychosocial function of more than 200 000 affected veterans, the US Department of Veterans Affairs (VA) funded and recently completed recruitment, assessment, and genotyping of 3690 veterans with SCZ and 5095 with BIP as part of the Cooperative Studies Program (CSP) #572, Genetics of Functional Disability in Schizophrenia and Bipolar Illness.7 This study included in-person assessment of a range of phenotypic variables, including extensive assessments for neurocognitive function and disability, and diagnoses were confirmed by structured clinical interviews.

Over the last decade, the genome-wide association study (GWAS) has emerged as a powerful research paradigm, and its success as applied to neuropsychiatric disorders exemplified by the trajectory of discovery in SCZ.3,9–12 Since the landmark Psychiatric Genomics Consortium (PGC) study that identified 108 distinct loci associated with SCZ,13 meta-analyses with new data have seen the number of susceptibility loci increase to 128,14 145,15 and 176.16 Progress for BIP risk has been comparatively modest16–18 likely owing to its clinical heterogeneity and smaller sample sizes. A critical realization is that common genetic risk variants have small effects and likely number in the thousands.19 These hard-won advances underscore the importance of ever-increasing sample sizes, well-phenotyped case-control cohorts, and the widespread accessibility of GWAS summary statistics.

We describe primary GWAS of SCZ and BIP in CSP #572 cases and screened controls from the Million Veteran Program (MVP)20 and employ polygenic risk scores (PRS) to benchmark the generalizability of published findings to our VA patient population. Noting the appreciable representation of African ancestry (AA) participants in CSP #572 and drawing upon genotype-level data from the Genomic Psychiatry Cohort (GPC),13,21 Molecular Genetics of Schizophrenia (MGS) study,9 Consortium on the Genetics of Schizophrenia (COGS),22 and Bipolar Genome Study (BiGS),23 these represent the largest such studies in AA populations to date. We combine our results with available summary statistics from several recent GWAS,13–15,18 yielding additional findings and highlighting challenges and opportunities for studies of these complex disorders.

Methods

Ascertainment and Diagnosis

All patients were required to meet lifetime (DSM-IV)24 criteria for SCZ (any subtype) or bipolar I disorder (any current phase). Participants received the Structured Clinical Interview for the DSM-IV (SCID),25 administered in-person by a trained rater. If needed, information from medical charts, patients’ clinicians, and other informants were utilized to confirm diagnoses. A diagnosis of schizoaffective disorder (SAD) based on the SCID was an exclusion criterion, in that, we did not anticipate being able to recruit a comparable number of participants with confirmed diagnoses of SAD. Patients with bipolar II disorder were not selected for participation given concerns about the reliability of this diagnosis.

Other exclusion criteria included major neurologic illnesses or systemic medical illnesses that could interfere with the central nervous system (CNS) function and test performance. We did not exclude patients with a diagnosis of substance abuse, given common co-occurrence (ie, comorbidity) and issues of representativeness: such patients were not enrolled if they appeared to be intoxicated, but could be reassessed at a later date.

Table 1 displays the demographic and clinical characteristics of CSP #572 participants with confirmed diagnoses of SCZ or BIP.

We selected control participants from among MVP enrollees by leveraging the VA electronic health record (EHR). We excluded individuals with any history of psychotic or affective disorders or past treatment with neuroleptic, mood-stabilizing, or antidepressant medications (see supplementary note).

This study was approved by the VA Central Institutional Review Board, and all participants provided written informed consent; no one who required the permission of a guardian to participate was enrolled in this study.

Genotyping, Imputation, and Ancestry Assignment

Genomic data processing was performed together for 8021 CSP #572 and >350 000 MVP participants (release 2.1;
At participating VA sites, blood samples were collected on the same day as in-person assessments. The VA Central Biorepository (VACB) provided the necessary supplies directly to the sites and designed the processes by which biospecimens were collected, packaged, shipped, and logged into a computerized tracking system and by which DNA was extracted. Staff were trained at the study “kickoff” and retrained periodically.

Genotyping was performed using a custom Affymetrix Axiom Biobank array with probes for 723,305 single-nucleotide polymorphisms (SNPs). Details of quality control procedures are described elsewhere. We excluded duplicates and samples with >2.5% missing genotype calls, excess heterozygosity, or discordance between genetic sex and phenotypic gender. We excluded SNPs with missingness >5% or minor allele frequency (MAF) that deviated by >10% from the 1000 Genomes Project Phase 3 (KGP3) data. Cleaned genotype data were imputed using Minimac3 and the KGP3 data, giving predicted genotype dosages for 49,134,253 SNPs and insertions/deletions.

Pairwise genetic relatedness was estimated using KING. Individuals demonstrating excess relatedness were excluded. We removed one individual at random from each pair of first-degree relatives, preferentially retaining cases from case-control pairs.

Genome-wide average proportions of GBR (UK Caucasians), PEL (Peruvians), YRI (Yoruba from Nigeria), CHB (Han Chinese), and LWK (Luhya from Kenya) ancestries were obtained from admixture analysis, and we assigned individuals to broadly defined European ancestry (EA) (GBR > .8) and AA (YRI > .6) groups. Among CSP #572 participants, 1021 (488 SCZ and 533 BIP) did not receive an ancestry assignment.

We used flashpca to extract ancestry principal components (PCs) from 22,766 independent autosomal SNPs.

### Table 1. Demographic and Clinical Characteristics of the CSP #572 Participants

<table>
<thead>
<tr>
<th>Category</th>
<th>Factor</th>
<th>Schizophrenia</th>
<th>Bipolar I</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>N (%)</td>
<td>N (%)</td>
<td></td>
</tr>
<tr>
<td>Demographic</td>
<td>Age (mean ± SD years)</td>
<td>55.1 ± 10.1</td>
<td>52.6 ± 11.5</td>
</tr>
<tr>
<td></td>
<td>Female</td>
<td>286 (7.3%)</td>
<td>1004 (18.6%)</td>
</tr>
<tr>
<td></td>
<td>Ancestry:</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>European American</td>
<td>1553 (39.8%)</td>
<td>3786 (70.2%)</td>
</tr>
<tr>
<td></td>
<td>African American</td>
<td>2122 (54.3%)</td>
<td>1320 (24.5%)</td>
</tr>
<tr>
<td></td>
<td>Other</td>
<td>231 (5.9%)</td>
<td>285 (5.3%)</td>
</tr>
<tr>
<td>Marital status:</td>
<td>Never married</td>
<td>1569 (40.0%)</td>
<td>980 (18.1%)</td>
</tr>
<tr>
<td></td>
<td>Currently married</td>
<td>768 (19.5%)</td>
<td>1795 (33.2%)</td>
</tr>
<tr>
<td></td>
<td>Previously married</td>
<td>1586 (40.2%)</td>
<td>2628 (48.5%)</td>
</tr>
<tr>
<td>Education:</td>
<td>Less than high school</td>
<td>1709 (43.5%)</td>
<td>1333 (24.7%)</td>
</tr>
<tr>
<td></td>
<td>High school</td>
<td>1833 (46.7%)</td>
<td>2886 (53.4%)</td>
</tr>
<tr>
<td></td>
<td>More than high school</td>
<td>384 (9.8%)</td>
<td>1184 (21.9%)</td>
</tr>
<tr>
<td>Annual income:</td>
<td>≤$20,000</td>
<td>1766 (51.2%)</td>
<td>2071 (40.8%)</td>
</tr>
<tr>
<td></td>
<td>&gt;$20,000</td>
<td>1682 (48.8%)</td>
<td>3006 (59.2%)</td>
</tr>
<tr>
<td>Clinical</td>
<td>Psychiatric comorbidity:</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Single</td>
<td>1213 (30.8%)</td>
<td>1555 (28.7%)</td>
</tr>
<tr>
<td></td>
<td>Multiple</td>
<td>1259 (31.9%)</td>
<td>2799 (51.7%)</td>
</tr>
<tr>
<td></td>
<td>Medical comorbidity</td>
<td>2026 (51.4%)</td>
<td>2809 (51.9%)</td>
</tr>
<tr>
<td>Lifetime suicide status:</td>
<td>Ideation</td>
<td>941 (23.9%)</td>
<td>1504 (27.8%)</td>
</tr>
<tr>
<td></td>
<td>Attempt or behavior</td>
<td>1816 (46.1%)</td>
<td>2953 (54.5%)</td>
</tr>
<tr>
<td></td>
<td>Current PTSD</td>
<td>867 (22.0%)</td>
<td>1787 (33.0%)</td>
</tr>
<tr>
<td>Lifetime major depression</td>
<td></td>
<td>1577 (40.0%)</td>
<td>2953 (54.5%)</td>
</tr>
<tr>
<td>Current negative symptoms</td>
<td></td>
<td>1380 (35.0%)</td>
<td>2953 (54.5%)</td>
</tr>
<tr>
<td>Impairment:</td>
<td>Everyday functioning UCSD</td>
<td>1533 (38.9%)</td>
<td>838 (15.5%)</td>
</tr>
<tr>
<td>Performance-based Skills Assessment (UPSA)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Any neuropsychological test</td>
<td></td>
<td>3683 (93.4%)</td>
<td>4750 (87.7%)</td>
</tr>
<tr>
<td>Electronic health record (EHR):</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Inpatient treatment</td>
<td></td>
<td>2978 (75.3%)</td>
<td>2978 (75.3%)</td>
</tr>
<tr>
<td>Medications:</td>
<td>Antipsychotics</td>
<td>3784 (95.7%)</td>
<td>4793 (88.4%)</td>
</tr>
<tr>
<td></td>
<td>Mood-stabilizers</td>
<td>2567 (64.9%)</td>
<td>5104 (94.1%)</td>
</tr>
<tr>
<td></td>
<td>Antidepressants</td>
<td>2928 (74.1%)</td>
<td>4529 (83.5%)</td>
</tr>
</tbody>
</table>
after excluding regions harboring long-range linkage disequilibrium (LD)\textsuperscript{33} (supplementary figures S1 and S2).

Following these checks, corrective actions, and assignments, 2,883 SCZ and 4,117 BIP cases and 50,436 controls were available for analysis.

**GWAS and Meta-Analyses With Published Findings**

We tested for association between imputed genotype dosages and case-control status by logistic regression using PLINK,\textsuperscript{34,35} including the first 10 ancestry PCs as covariates, and retaining variants with imputation quality (INFO) $\geq .9$ and MAF $\geq .01$. We combined AA and EA results using Han and Eskin's\textsuperscript{36} random effects (RE2) model implemented in METASOFT, which is optimized to detect associations in the presence of heterogeneity. We used this method to combine sex-specific X-chromosome results.

We defined statistically independent genome-wide significant (GWS) by LD-based “clumping” of SNPs with $P < 10^{-4}$ and LD $r^2 \geq .1$ with the lead variant within a 3 megabase (Mb) window.\textsuperscript{12} We annealed clumped intervals within 250 kilobases (kb) into physically distinct loci.

We obtained genotype-level data for the GPC, including MGS and COGS samples, and for the BiGS study\textsuperscript{33,37,38} via the database of Genotypes and Phenotypes (dbGaP; phs000017.v3.p1). We reanalyzed these data as previously described,\textsuperscript{13} applying analogous ancestry assignments. We classified as Latino ancestry (LA) any participant with $\geq .25$ admixed American (AMR) ancestry and $< .25$ AA, East Asian (EAS), or South Asian ancestry.

We combined SCZ results for CSP #572, GPC, and the meta-analysis of PGC-SCZ2 and CLOZUK (named for a series of UK-based cases registered for clozapine treatment)\textsuperscript{14,39} (https://walters.psycm.cf.ac.uk/). We also considered a recent meta-analysis of PGC-SCZ2 and multiple EAS cohorts\textsuperscript{15} (https://www.med.unc.edu/pgc/) but did not combine EAS-only results with CLOZUK + PGC because of sample overlap with 3 EAS cohorts in PGC-SCZ2,\textsuperscript{12} and because CLOZUK-only summary statistics were not available to us.

For BIP, we built on the most recent PGC GWAS (PGC-BIP),\textsuperscript{18} incorporating results for multiple ancestries from CSP #572, GPC, and BiGS.

Sample sizes by diagnosis and ancestry are displayed in Table 2.

**PRS Profiling**

We constructed PRS from published and current GWAS results (the “training” datasets), testing individual-level scores for association with case-control status in CSP #572 (the “target” dataset). Variants that met quality control requirements in both studies were clumped in the appropriate KGP3 population ($r^2 \geq .1$; 500 kb window). For varying $P$-value thresholds in the training dataset, scores were constructed by summing the number of tested alleles by their effect estimate. We tested for case-control differences by logistic regression using the same GWAS covariates. Predictive value is reported in terms of Nagelkerke’s pseudo-$R^2$ (fmsb package in R)\textsuperscript{40} and $R^2$-adjusted for the proportion of cases (ie, the liability scale) assuming a 1% population prevalence for each disorder.\textsuperscript{41,42}

**Heritability and Genetic Correlations**

We applied genome-based restricted maximum likelihood (GREML), as implemented in the genome-wide complex trait analysis (GCTA) software,\textsuperscript{43} to estimate the variance explained by the aggregate effects of genome-wide common variants (SNP-$f^2$), and its bivariate extension to obtain the estimates of genetic correlation ($r_g$).\textsuperscript{41} We included the same GWAS covariates and used the [−grm-cutoff .05] flag to restrict analyses to approximately unrelated individuals. Trans-ancestry $r_g$ was estimated from variants with frequency $\geq 0.05$ in both populations.\textsuperscript{44}

We applied LD score regression (LDSC)\textsuperscript{45} to estimate $r_g$ between CSP #572 and CLOZUK + PGC and PGC-BIP. We extended these analyses to 597 traits from UK Biobank (UKB) and other published studies using LD Hub (http://lhub.aikenlab.org/lhub/).\textsuperscript{46} Given the requirement of an LD reference panel and concerns about the reliability of reference LD measures for admixed populations, we did not perform LDSC for AA samples.\textsuperscript{45}

**Trans-ancestry Fine-Mapping**

For independent GWS associations in each meta-analysis, we constructed credible SNP sets from variants in LD with the lead variant ($r^2 \geq .1$ in the relevant KGP3 populations),\textsuperscript{13–15} combining their ranked posterior probabilities until the sum exceeded 0.99.\textsuperscript{47} We compared credible sets between each discovery GWAS (eg, CLOZUK + PGC) and our expanded meta-analysis on the basis of the number of credible SNPs, the length of the corresponding genomic interval, and the smallest observed $P$-value. We considered a region to be fine-mapped if the interval for the reduced credible set was smaller than the interval corresponding to SNPs with LD $r^2 \geq .6$ with the lead variant (in KGP3 EA samples).\textsuperscript{13}

**Functional Annotations and Gene-Set Enrichment**

We used the Functional Mapping and Annotation (FUMA)\textsuperscript{48} platform (http://fuma.ctglab.nl/) for follow-up bioinformatic analyses. For each GWS locus, variants were mapped to expression quantitative trait loci (eQTLs) within 1 Mb of a gene’s transcription start sites (ie, cis-eQTLs); we considered eQTLs in 13 CNS tissues from the Genotype-Tissue Expression (GTEx version 8) database that survived a 5% false discovery rate (FDR) correction. Gene-set enrichment analyses were performed with
Table 2. Sample Sizes by Study Cohort, Assigned Ancestry, and Diagnosis

<table>
<thead>
<tr>
<th>Study</th>
<th>Ancestry</th>
<th>SCZ</th>
<th>BIP</th>
<th>Controls*</th>
</tr>
</thead>
<tbody>
<tr>
<td>CSP #572</td>
<td>AA</td>
<td>1683</td>
<td>1037</td>
<td>4669</td>
</tr>
<tr>
<td></td>
<td>EA</td>
<td>1200</td>
<td>3,080</td>
<td>45,767</td>
</tr>
<tr>
<td>GPC</td>
<td>AA</td>
<td>3486</td>
<td>1641</td>
<td>2195 (2470)</td>
</tr>
<tr>
<td>GPC/COGS</td>
<td>AA</td>
<td>616</td>
<td>362</td>
<td></td>
</tr>
<tr>
<td>GPC/MGS</td>
<td>AA</td>
<td>1724</td>
<td></td>
<td>1111</td>
</tr>
<tr>
<td>GPC</td>
<td>EA</td>
<td>2919</td>
<td>835</td>
<td>2367</td>
</tr>
<tr>
<td>GPC</td>
<td>LA</td>
<td>1234</td>
<td>1032</td>
<td>3090</td>
</tr>
<tr>
<td>BiGS</td>
<td>AA</td>
<td>349</td>
<td>958</td>
<td></td>
</tr>
</tbody>
</table>

**Genotyped totals**
- 12,862
- 7974
- 59,561 (59,836)

**CLOZUK + PGC**
- EA: 40,675
- EA: 64,643

**PGC-BIP**
- EA: 53,537
- EA: 28,326
- 124,204 (91,194)

**Meta-analysis totals**
- (primary)
- (secondary)

**PGC + EAS**
- EA: 33,640
- EA: 43,456
- EAS: 22,778
- EAS: 35,362

**Genotype totals**
- 69,280
- 138,379

Note: CSP, Cooperative Studies Program; GPC, Genomic Psychiatry Cohort; COGS, Consortium on the Genetics of Schizophrenia; MGS, Molecular Genetics of Schizophrenia; BiGS, Bipolar Genome Study; CLOZUK, a series of UK-based cases registered for clozapine treatment; PGC, Psychiatric Genomics Consortium; BIP, bipolar disorder; EAS, East Asian; EA, European ancestry; AA, African ancestry; LA, Latino ancestry, SCZ, schizophrenia.

Boldface indicates a summation of counts, which is implicit from "totals".

* BIP control counts are displayed parenthetically if different from SCZ analysis.

* PGC-SCZ2 included 1866 cases and 3418 controls of EAS ancestry.

MAGMA using curated gene-sets and gene ontology (GO) categories from MsigDB v6.2. The major histocompatibility complex was excluded from these analyses.

We utilized the S-PrediXcan software49,50 to predict gene expression differences from GWAS summary statistics, using elastic net prediction models for the same tissues (http://predictdb.hakyimlab.org/). We applied 5% FDR corrections within and across tissues.

Results

Primary GWAS

Our primary GWAS in CSP #572 uncovered an SCZ association with CHD7 variants at 8q12.2 and BIP associations with SORCS3 variants at 10q25.1 (figure 1) and downstream of PCDH11X at Xq21.32 (supplementary tables S1–S3; supplementary figures S3–S7).

The fraction of SNPs with identical directions of allelic effects in both ancestries was significantly greater than chance (P < .001) (supplementary table S4). Comparisons with published GWAS revealed a similar pattern, with greater concordance at more inclusive P-value thresholds (supplementary table S5; figure 2A).

PRS Profiling

Scores based on the CLOZUK + GPC meta-analysis yielded the best predictive value experiment wide, accounting for upwards of 7% of the variance on the liability scale among EA participants (P < 10−10) (figure 2B; supplementary table S6). Strikingly, as applied to AA participants, CLOZUK + PGC-trained PRS explained less variance than PRS based on the smaller AA-GPC, with scores based on GPC and PGC + EAS performing best overall.

Scores based on PGC-BIP results explained ~1.7% and 0.6% of variance in EA (P < 10−49) and AA (P < 10−25) for BIP in EA participants. Among AA participants, we saw lower estimates of 0.121 (95% CI:[0.040, 0.203]; P = .00149) and 0.094 (95% CI:[−0.014, 0.202]; P = .0367). Bivariate analyses returned significant estimates of both trans-ancestry and cross-disorder r_g (supplementary table S10).

Heritability and Genetic Correlations

Estimates of SNP-h^2 were 0.288 (95% CI:[0.208, 0.367]; P < 10−4) for SCZ and 0.188 (95% CI:[0.150, 0.225]; P < 10−25) for BIP in EA participants. Among AA participants, we saw lower estimates of 0.121 (95% CI:[0.040, 0.203]; P = .00149) and 0.094 (95% CI:[−0.014, 0.202]; P = .0367). Bivariate analyses returned significant estimates of both trans-ancestry and cross-disorder r_g (supplementary table S10).

Using LDSC and EA summary statistics, we estimated r_g between CSP #572 and CLOZUK + PGC as 0.818 (95% CI:[0.655, 0.981]; P < 10−22) and PGC-BIP as 0.721 (95% CI:[0.549, 0.894]; P < 10−15). Extension to >600 selected traits using LD Hub recapitulated previous findings44,55 and highlighted relationships between both disorders and smoking (P < 10−5), having sought treatment for anxiety or depression (P < 10−5), and deprivation index (P < 10−5) (supplementary table S11).

Joint Analyses With Published GWAS

Joint analysis of CSP #572 and GPC with CLOZUK + PGC summary statistics yielded 195 independent GWS associations in 159 physical loci. Meta-analysis of CSP #572, GPC, and PGC + EAS yielded 223 independent associations in 187 loci. Taken together, 228 distinct loci were robustly associated with SCZ and 39 are newly reported here, including 21 loci that only attained significance in joint analyses with CSP #572 (figure 1B; supplementary tables S12 and S13; supplementary figures S8 and S10).
Combining BIP results for CSP #572, GPC, and PGC-BIP yielded 23 independent associations in 23 loci, including 10 novel findings and 9 that attained significance in the CSP #572 joint analysis (supplementary table S14; supplementary figures S9 and S10).

Fine-Mapping of New and Replicated Loci
Joint analysis of CSP #572, GPC, and CLOZUK + PGC improved fine-mapping resolution for 49 independent associations; by comparison, meta-analysis with PGC + EAS yielded an improved fine-mapping resolution of 44 independent associations (supplementary Tables S15 and S16). In both analyses, 12 associations were reduced to fewer than 10 SNPs and 1 locus was mapped to a single SNP. We fine-mapped 10 of 24 independent GWS associations for BIP, with 3 reduced to fewer than 10 SNPs and 1 mapped to a single SNP (supplementary table S17; supplementary figures S11 and S12).

Functional Annotation and Predicted Gene Expression
We mapped GWS loci from the meta-analysis of CSP #572, GPC, and CLOZUK + PGC to 1072 eQTLs (79 independent loci) influencing the expression of 284 genes in one or more CNS tissues; an additional 108 eQTLs were identified via meta-analysis with PGC + EAS, the majority being attributable to a GWS association at 17q21.31 (supplementary tables S18 and S19). Among GWS loci for BIP, we identified 142 eQTLs (14 independent loci) for 52 genes (supplementary table 20).

We explored these findings via transcriptome-wide association studies (TWAS) in the same tissues, imputing results directly from our summary statistics using S-PrediXcan (figure 3). Among 3152 and 416 gene-tissue pairs surviving FDR correction, the strongest findings were of predicted higher expression of complement C4A (C4A) in SCZ and X-Prolyl Aminopeptidase 3 (XPNPEP3) in BIP across multiple tissues. Intersecting these results with eQTL annotations for new associations highlighted 100 genes at 26 loci with predicted differential expression between SCZ cases and controls (supplementary tables S21–S23), including some previously reported findings.56,57

Gene Set Enrichment
Our primary SCZ results were significantly enriched for a single GO term, positive regulation of oligodendrocyte differentiation (supplementary tables S24 and S25). Across meta-analyses, GWAS of SCZ and BIP highlighted enrichment in 22 and 3 gene-sets (supplementary tables...
The enriched terms were related to the aspects of neuronal structure, dendrite growth, and synaptic transmission.

Discussion

We describe primary genomic analyses of SCZ and BIP in CSP #572, offering new insights into the etiologies of these disorders and demonstrating that published findings are largely transferable to this diverse cohort of US veterans. We combined our results with summary statistics from recent GWAS, realizing the largest and most inclusive GWAS of these disorders to date, and explore the functional biological relevance of 52 newly identified susceptibility loci.

Our primary GWAS of SCZ yielded a novel association in chromodomain helicase DNA-binding protein 7 (CHD7), a critical player in the epigenetic regulation of neuronal differentiation, disruptions of which are believed to be associated with SCZ and autism. Mutations in CHD7 are observed in developmental disorders, and common variant associations have been reported with a range of human traits.

We report a novel BIP association in SORCS3, encoding sortilin-related VPS10 domain-containing receptor 3, which plays key roles in glutamate homeostasis.
and establishing long-term depression in hippocampal neurons.⁷¹,⁷² SORCS3 variants have demonstrated pleiotropic effects on SCZ, major depression (MD), anorexia nervosa, attention hyperactivity disorder, Tourette’s syndrome,⁷³ and cognition.⁷⁴ This finding was not replicated in other BIP cohorts, likely reflecting heterogeneity among participants and across studies.

We observed another GWS association downstream of Protocadherin 11 X-linked (PCDH11X), one of a human-specific X/Y gene pair associated with the regulation of dendritic branching⁷⁵ and neuronal differentiation and proliferation,⁷⁶ and which has postulated evolutionary relevance to the neural correlates of language.⁷⁷ Combining results for CSP #572 with published findings, we detected an additional 39 susceptibility loci for SCZ and 10 for BIP and fine-mapped 84 associations. New loci were enriched in eQTLs for CNS-expressed genes, including a novel BIP association in Long Intergenic Non-Protein Coding RNA 1470 (LINC01470) spanning eQTLs for neuromedin U receptor 2 (NMUR2), a neuropeptide receptor that is presynaptically expressed in GABAergic neurons and enriched in the nucleus accumbens shell and has been implicated in mediating food reward and behavioral responses to cocaine.⁷⁸ Taken together, the NMUR2 and SORCS3 findings support the role of glutamatergic dysregulation in SCZ⁵²,⁷⁹,⁸⁰ and perhaps BIP.⁸¹–⁸³

We observed a novel association between SCZ and variants upstream of CRHR1 (figure 1B), encoding corticotropin-releasing hormone receptor 1 and was recently found to be associated with reexperiencing symptoms of post-traumatic stress disorder (PTSD) in the MVP.⁷⁷ Nearby SNPs are GWS for neuroticism,⁸⁴,⁸⁵ cognitive performance,⁸⁶,⁸⁷ subcortical⁸⁸ and intracranial volume,⁸⁹ Parkinson’s disease,⁹⁰–⁹³ and corticobasal degeneration,⁹⁴ sleep duration,⁹⁵ alcohol use disorder,⁹⁶–⁹⁷ various risky behaviors,⁹⁸ and a number of other traits.⁶⁷,⁹⁹–¹⁰⁴ The 17q21.31 locus harbors a ~900-kb inversion that has undergone positive selection in Europe but is rare among AA and EAS individuals;⁹⁵ the resultant pattern of long-range LD poses challenges to statistical fine-mapping and may complicate the interpretation of eQTL effects; however, we achieved a small reduction in causal credible SNPs via trans-ancestry analysis. As noted by Gelernter and colleagues,²⁷ a weaker signal in a smaller AA sample, nonetheless, lends meaningful insight into the likely underlying association at this locus.

We observed improved fine-mapping resolution at 70 independent SCZ loci, with 21 overlapping between meta-analyses of CSP #572 and GPC with PGC + CLOZUK or PGC + EAS. Integration of diverse ancestry data yielded more fine-mapped loci in the PGC + CLOZUK meta-analysis, with slightly larger reductions in the median number of credible SNPs and length of the corresponding genomic interval. Our results are comparable to those reported in the original PGC + EAS study and based on analogous methods and criteria; in both studies, 44 independent associations had improved significance and smaller credible SNP sets following trans-ancestry meta-analysis, with only 3 loci overlapping across studies. This supports distinct contributions of the AA and LA data featured herein, reflecting local LD structures that are divergent from EA and EAS populations.¹⁰⁶ In particular, the inclusion of AA study data is expected to yield the greatest improvements in fine-mapping, given a higher degree of genetic diversity and shorter haplotype blocks on average compared with out-of-Africa populations.¹⁰⁶–¹⁰⁸

Consistent with expectations, PRS constructed from published GWAS results were robustly associated with case-control status in CSP #572 but explained a small fraction of the overall variability in liability. Among AA participants, we saw markedly improved performance for

---

Fig. 3. Predicted transcriptomic differences between schizophrenia (SCZ) cases and controls. Transcriptome-wide association studies (TWAS) significance (y-axis) for novel genome-wide significant (GWS) findings in the primary meta-analysis of SCZ genome-wide association studies (GWAS). Loci with significant expression quantitative trait loci (eQTL) annotations are highlighted; the number of tissues is shown parenthetically, and symbol orientation indicates higher or lower predicted expression. Additional findings at 17q21.31 are denoted with b, and replicated GWS loci appear as watermark. *C4A* is highlighted for reference.
scores constructed from multi-ancestry GWAS findings. We observed strong associations of PRS across populations and diagnoses, highlighting the importance of studying diverse populations.  

For both SCZ and BIP, a substantive fraction of the variance was attributable to genome-wide SNPs. Estimates of SNP-$h^2$ for BIP were significantly lower than previously reported ($P < .005$), and this was also true for SCZ among AA participants in CSP #572.

It should be noted that SNP-$h^2$ represents a lower bound on the "narrow-sense" $h^2$; estimates of the total $h^2$ of SCZ range from .31 to .67.

Interpretations and Conclusions

Our findings are demonstrative of a shared genetic basis of SCZ and BIP between this highly selected and severely affected veteran cohort and previous case-control studies of both disorders. We expect that the comprehensive assessment and validity of the clinical assessments in CSP #572 represent a “gold standard” for the curation of serious mental illness (SMI) phenotypes in the US veteran population. Taken together with the completeness and scale of the VA EHR, forthcoming studies of CSP #572 and the MVP hold considerable promise to advance our understanding of how genetic and environmental/experiential factors (G × E) influence disease susceptibility, neurocognitive function, treatment response, and patient outcomes. Future analyses of this dataset will explore genomic determinants of suicidal ideation and behavior, comparative risk for PTSD with non-SMI populations, and multifactorial influences on lifetime MD, negative symptoms, and other salient symptomatic domains.

Supplementary Material

Supplementary material is available at Schizophrenia Bulletin.

Funding

This research was supported by the Department of Veterans Affairs Cooperative Studies Program (CSP #572) and the Million Veteran Program (MVP-000). The MVP is supported by the Office of Research and Development, US Department of Veterans Affairs. Dr. Bigdeli is the recipient of a 2019 NARSAD Young Investigator Grant from the Brain and Behavior Research Foundation. Other relevant funding information for GPC, COGS, MGS, and BiGS is presented in the Supplementary Materials.

Acknowledgments

The contents do not represent the views of the US Department of Veterans Affairs or the US Government. Complete acknowledgments for CSP #572, MVP, GPC, and COGS are presented in the Supplementary Materials.

Dr. Harvey has served as a consultant to multiple pharmaceutical companies and device manufacturers on phase 2 or 3 development; this consulting work has been determined to be unrelated to the content of the article. No other authors report any relevant conflicts of interest.

References

8. Wellcome Trust Case Control Consortium. Genome-wide association study of 14,000 cases of seven common diseases and 3,000 shared controls. Nature 2007;447:661–678.


