Evidence of shared genetic influences underlying schizophrenia and alcohol use disorder, but not alcohol consumption

Emma C. Johnson[1], Manav Kapoor[2], Alexander S. Hatoum[1], Hang Zhou[3,4], Renato Polimanti[3,4], Frank R. Wendt[3,4], Raymond K. Walters[5,6], Dongbing Lai[7], Rachel L. Kember[8,9], Sarah Hartz[1], Jacquelyn L. Meyers[10,11], Roseann E. Peterson[12], Stephan Ripke[5,6,13], Tim B. Bigdeli[10], Ayman H. Fanous[10], Carlos N. Pato[10], Michele T. Pato[10], Alison M. Goate[2], Henry R. Kranzler[8,9], Michael C. O'Donovan[14], James T.R. Walters[14], Joel Gelernter[3,4,15,16], Howard J. Edenberg[7,17], Arpana Agrawal[1]

[1]Department of Psychiatry, Washington University School of Medicine, Saint Louis, MO, USA
[2]Department of Neuroscience, Icahn School of Medicine at Mount Sinai, New York, NY, USA
[3]Department of Psychiatry, Division of Human Genetics, Yale University School of Medicine, New Haven, CT, USA
[4]Department of Psychiatry, Veterans Affairs Connecticut Healthcare System, West Haven, CT, USA
[5]Analytic and Translational Genetics Unit, Department of Medicine, Massachusetts General Hospital and Harvard Medical School, Boston, MA, USA
[6]Stanley Center for Psychiatric Research, Broad Institute of MIT and Harvard, Cambridge, MA, USA
[7]Department of Medical and Molecular Genetics, Indiana University School of Medicine, Indianapolis, IN, USA
[8]Department of Psychiatry, University of Pennsylvania Perelman School of Medicine, Philadelphia, PA, USA
[9]VISN 4 MIRECC, Crescenz VAMC, Philadelphia, PA, USA
[10]Department of Psychiatry and Behavioral Sciences, SUNY Downstate Health Sciences University, Brooklyn, NY, USA
[11]Henri Begleiter Neurodynamics Laboratory, SUNY Downstate Health Sciences University, Brooklyn, NY, USA
[12]Department of Psychiatry, Virginia Institute for Psychiatric and Behavioral Genetics, Virginia Commonwealth University, Richmond, VA, USA
[13]Department of Psychiatry and Psychotherapy, Charité - Universitätsmedizin Berlin, Campus Mitte, Berlin, Germany
[14]MRC Centre for Neuropsychiatric Genetics and Genomics, Division of Psychological Medicine and Clinical Neurosciences, Cardiff University School of Medicine, Cardiff, UK
[15]Department of Genetics, Yale University School of Medicine, New Haven, CT, USA
[16]Department of Neuroscience, Yale University School of Medicine, New Haven, CT, USA
[17]Department of Biochemistry and Molecular Biology, Indiana University School of Medicine, Indianapolis, IN, USA

Abstract: 286 words
Manuscript: 4,714 words
Figures: 4

Corresponding author: Emma C Johnson
Washington University School of Medicine
Dept. of Psychiatry, 660 S. Euclid, CB 8134
Saint Louis, MO 63110, USA
emma.c.johnson@wustl.edu

Abstract

Background: Alcohol use disorder (AUD) and schizophrenia (SCZ) frequently co-occur, and recent genome-wide association studies (GWAS) have identified significant genetic correlations between them. In parallel, mounting evidence from GWAS suggests that alcohol *consumption* is only weakly genetically correlated with SCZ, but this has not yet been systematically investigated.

Methods: We used the largest published GWAS for AUD (total cases = 77,822) and SCZ (total cases = 46,827) to systematically identify genetic variants that influence both disorders (in either the same or opposite direction of effect) as well as disorder-specific loci, and contrast our findings with GWAS data for drinks per week (DPW; N = 537,349) as a measure of alcohol consumption.

Results: We identified 55 independent genome-wide significant SNPs with the same direction of effect on AUD and SCZ, 9 with robust opposite effects, and 99 with disorder-specific effects. We also found evidence for 12 genes whose pleiotropic associations with AUD and SCZ are consistent with mediation via gene expression in the prefrontal cortex. The genetic covariance between AUD and SCZ was concentrated in genomic regions functional in brain tissues (p = 0.001). The genetic correlation between DPW and SCZ (r_g = 0.102, SE = 0.022) was significantly lower than that for AUD and SCZ (r_g = 0.392, SE = 0.029; p-value of the difference = 9.3e-18), and the genetic covariance between DPW and SCZ was not enriched for any meaningful tissue-specific categories.

Conclusions: Our findings provide a detailed view of genetic loci that influence risk of both AUD and SCZ, suggest that biological commonalities underlying genetic variants with an effect on both disorders are manifested in brain tissues, and provide further evidence that SCZ shares meaningful genetic overlap with AUD and not merely alcohol consumption.

Introduction

Schizophrenia (SCZ) and alcohol use disorder (AUD) are serious psychiatric disorders[1]. AUD is more common in individuals with SCZ (prevalence of 20-30%[2,3], compared to ~6% in the general population[4]) and a diagnosis of both is associated with greater psychiatric comorbidity[5], more clinical complications[6], and a lower likelihood of sustained medication adherence[7] than either disorder alone. Both SCZ and AUD are moderately to highly heritable (twin-h^2 for SCZ = 81%[8], AUD = 49%[9]), and genome-wide association studies (GWAS) have consistently found positive genetic correlations of AUD with SCZ (e.g., r_g = 0.34, p = 3.7e-21)[10]. Further, polygenic risk scores (PRS) for AUD are significantly associated with SCZ risk[10] and vice versa[11]. In contrast, the genetic correlation between SCZ and measures of typical alcohol consumption is weak (e.g., drinks/week: r_g = 0.01, p = 0.67[12]), suggesting that SCZ might share substantial genetic liability only with the psychopathological aspects of disordered drinking[13] and not alcohol consumption *per se*.

Despite the substantial genetic correlation between AUD and SCZ, there is little known regarding the underlying pleiotropic mechanisms in terms of the specific risk alleles, genes, and molecular pathways involved. Although recent studies have begun to elucidate the contributions of pleiotropic loci to shared genetic variance amongst disorders and complex traits[14,15], these efforts have not included AUD, one of the most common psychiatric disorders. In addition, traditional cross-disorder GWAS methods can obscure the effects of divergent variants, i.e., those that are pleiotropic for two disorders but whose effect alleles operate in opposite directions, conferring risk for one disorder and protective effects for the other (e.g., potassium ion response genes that distinguish SCZ from bipolar disorder[16]). The identification of such variants is fundamental to identifying the pathways that contribute to diagnostic boundaries.

The current study outlines the nature of the shared genetic underpinnings of AUD and SCZ by conducting cross-disorder analyses of large genome-wide datasets of both European- and African-ancestry individuals (see **Figure 1** for overview). We conducted ancestry-specific cross-disorder meta-analyses to systematically identify pleiotropic loci with significant convergent and divergent effects on both SCZ and AUD, and loci specific to each disorder. We also linked pleiotropic variants to gene expression data from the frontal cortex in an effort to prioritize the genes that are more likely to be causal. Because the extent to which the correlation between AUD and SCZ is attributable to salient functional categories remains unknown, we partitioned the genetic covariance between AUD and SCZ into relevant annotations. Finally, given recent evidence that AUD might be more polygenic than SCZ[17], with a greater proportion of risk variants with small effect sizes[18], we examined the genetic effect size distributions of

AUD and SCZ. Across all approaches, we contrasted AUD with the largest GWAS of alcohol consumption[12] to assess whether the genetic overlap between SCZ and alcohol-related phenotypes is restricted to risk for pathological drinking or is also evident for a more widely assessed continuous measure of typical alcohol intake.

Methods and Materials

Samples

Alcohol Use Disorder

For the European-ancestry subset, we utilized GWAS data from the largest available AUD meta-analysis[19] (N = 313,959; N_{cases} = 57,564). This study meta-analyzed AUD GWAS data from the Million Veteran Program (MVP), where case status was derived from International Classification of Diseases (ICD) codes of alcohol-related diagnoses from electronic health records (EHR) data, and the Psychiatric Genomics Consortium (PGC) alcohol dependence GWAS[17] with cases based on DSM-IV diagnoses. We meta-analyzed two published GWAS to create the African-ancestry subset: the MVP Phase 1 GWAS of AUD[10] (N = 56,648; N_{cases} = 17,267), and the PGC GWAS[17] (N = 5,799; N_{cases} = 2,991). We used METAL[20] (which combines genome-wide summary statistics across multiple samples) to generate the African-ancestry summary statistics by meta-analyzing the GWAS data from the MVP Phase 1 AUD and PGC alcohol dependence GWAS using an inverse variance-weighted fixed-effects model, excluding SNPs with INFO score < 0.8 and/or minor allele frequency < 0.01 within each sample.

Schizophrenia

For the European-ancestry sample, we used the PGC Phase 2 + CLOZUK (a sample of individuals with schizophrenia who were treated with clozapine) SCZ GWAS meta-analysis[21] (total N = 105,318; N_{cases} = 40,675). We used the summary statistics from the Genomic Psychiatry Cohort (GPC) SCZ GWAS[22] (N = 10,070; N_{cases} = 6,152) for the African-ancestry cross-disorder analysis.

Drinks per week

As a contrast with AUD, we used genome-wide association data for the European-ancestry drinks per week (DPW) GWAS (N = 537,349) from GSCAN[12], making this the largest sample in our analysis. GSCAN is a meta-GWAS that included many different cohorts, including 23andMe; however, we did not include the 23andMe data in our analysis due to data sharing restrictions.

Analysis

Cross-disorder association analysis

We used "Association analysis based on SubSETs" (ASSET)[23] to combine the genome-wide association data for AUD and SCZ, using the two-tailed meta-analysis approach to obtain a single cross-disorder association statistic, correcting for sample overlap. Unlike traditional meta-analysis approaches, ASSET takes into account SNPs with significant effects on multiple disorders even if the effects on the traits are in opposite directions. Default parameters were applied using the "h.traits" function, and we used LD Score Regression[24,25] (LDSC) to obtain a rough estimate of sample overlap, which we accounted for in the additional covariance term. We then separated the ASSET results into four subsets: a "convergent" subset (effect allele with the same direction of effect for both disorders), a "divergent" subset (effect allele with opposite directions of effect on the disorders), a subset of SNPs with AUD-specific effects, and a subset of SNPs with effects only on SCZ (**Supplemental Figure 1**).

We then uploaded the subset results (i.e., convergent, divergent, AUD-only, and SCZ-only SNPs from ASSET output) to FUMA[26] for annotation and identification of genome-wide significant risk loci and independent lead SNPs. We further subset these convergent and divergent loci to exclude top lead SNPs with $p > 0.05$ in either individual disorder GWAS, to create a more conservative set of cross-disorder variants with at least nominal significance in both disorders (although the full subsets were used for gene-set and pathway analyses).

In the African-ancestry samples, which were smaller, we focused on the overall set of pleiotropic cross-disorder variants, rather than parsing the pleiotropic variants into subsets with convergent and divergent effects. In addition, as much of the data required for follow-up analyses described below were restricted to European-ancestry individuals, these follow-up analyses were not conducted on the African-ancestry cross-disorder summary statistics.

Gene, gene-set, and pathway analyses

Subset results (i.e., convergent and divergent SNPs from ASSET output) were uploaded to FUMA[26] for gene-based analyses in MAGMA[27]. These analyses included gene-set analyses using curated gene sets and GO terms from MsigDB[28] (an online collection of annotated gene-sets), gene-property analyses based on tissue expression data from GTEx[29] v8 (a repository of expression data by brain region from autopsies of 960 donors) and BrainSpan[30] (a resource detailing gene expression patterns across specific brain regions from 42 brain specimens, spanning prenatal and postnatal time points in both males and females), and data visualization (more details in **Supplemental Materials**).

SMR eQTL analyses and differential gene expression analyses

To examine whether the effects of pleiotropic variants with convergent effects on AUD and SCZ may be mediated by gene expression patterns, we conducted a summary data-based Mendelian randomization[32] (SMR) analysis on a set of meta-analyzed expression quantitative trait loci (eQTL) data in the prefrontal cortex (combining data from ROSMAP[33], PsychENCODE[34], and COGA-INIA datasets[35]; total N = 1,986). SMR is a Mendelian randomization-based analysis that integrates GWAS summary statistics with eQTL data to test whether the effect size of a SNP on the phenotype of interest is mediated by gene expression. SMR does not require raw eQTL data to build the weights. We excluded variants with potential pleiotropic effects (which can confound SMR analyses) using the HEIDI outlier method[36] (excluding SNPs with HEIDI outlier $p < 0.05$).

We also examined whether genes mapped to pleiotropic loci by MAGMA were significantly enriched for genes showing differential expression in the prefrontal cortex using two comparisons in independent samples: we compared gene expression in 65 individuals with alcohol dependence and 73 healthy controls[35], and 258 individuals with SCZ and 279 controls[37] (details in Supplemental Note) using Fisher's exact test[38].

Genetic correlations and partitioned covariance

We used LDSC[24,25] to estimate the genetic correlations (r_g) between AUD, SCZ, DPW, and two negative control traits (height and chronic ischemic heart disease). We compared the genetic correlations between AUD and SCZ, and DPW and SCZ using a block jackknife method implemented through LDSC to test the null hypothesis that r_g(AUD, SCZ) minus r_g(DPW, SCZ) = 0.

We also used LDSC applied to Specifically Expressed Genes (LDSC-SEG) to estimate the enrichment of AUD and SCZ across 13 specific brain regions (annotations defined using GTEx[39] gene expression data).

We used GeNetic cOVariance Analyzer (GNOVA)[40] to partition the genetic covariance (ρ_g) between AUD and SCZ into salient annotation categories. These included 1) functional vs. non-functional areas of the genome (GenoCanyon[41] annotations, defined by integrating genomic conservation measures and biochemical annotation data to generate a functional potential score for each genetic variant), 2) tissue- and regional-specific functionality (GenoSkyline[42,43] annotations, which are tissue-specific functional regions defined by integrating high-throughput epigenetic annotations, 3) GTEx v6[39] brain region annotations from LDSC cell-type specific analyses), and 4) minor allele frequency quartiles. We excluded the MHC region (chr6:26000885 - chr6:33999991) from both the LDSC-SEG and GNOVA analyses due to the long-range and complex LD in this region.

In sensitivity analyses conducted on individuals of European ancestry in the UK Biobank (from the Neale lab GWAS: https://www.nealelab.is/uk-biobank), we calculated the genetic correlation and partitioned genetic covariance between AUD, SCZ, and two negative control traits: height (N = 360,388) and chronic ischemic heart disease (CHD; N = 361,194; N_{cases} = 12,769),.

Latent causal variable (LCV) analysis

We used the latent causal variable (LCV) approach[44] to test for evidence of bi-directional partial genetic causality. LCV, an alternative to traditional Mendelian Randomization approaches, is based on full genome-wide data, is not biased by sample overlap, and has fewer false positives in the presence of pleiotropy and for highly polygenic traits[44].

Effect size distributions for AUD, DPW, and SCZ

Given recent evidence that AUD and SCZ have differing levels of polygenicity[17,18], we used GENetic Effect-Size distribution Inference from Summary-level data[45] (GENESIS) to estimate the effect size distribution parameters for each disorder/trait (details given in **Supplemental Material**). Effect size distributions are characterized by three descriptive statistics per GWAS: π_c describes the proportion of susceptibility SNPs, σ^2 describes the variance in effect size for non-null SNPs, and α describes residual effects not captured by the variance of effect sizes such as population stratification, underestimated effects of extremely small-effect-size SNPs, and/or genomic deflation. We calculated GWAS risk locus discovery and phenotypic variance explained using three projected sample sizes (N = 500,000, 1 million, and 5 million), two-component model parameter estimates, 1000 bootstrap samplings for confidence interval calculation, and a genome-wide significance threshold of 5e-8.

Results

Identifying pleiotropic variants, genes, and pathways in individuals of European ancestry

The cross-disorder analysis of AUD and SCZ in European-ancestry individuals identified numerous significant pleiotropic loci: after genome-wide clumping via FUMA[26], there were 55 independent risk loci (with 60 lead SNPs) with convergent effects and 47 risk loci (56 lead SNPs) with divergent effects (i.e., effect allele with opposite effects on AUD and SCZ; **Supplemental Figure 1, Supplemental Tables 1-3)**. We also identified disorder-specific loci through ASSET: 91 with SCZ-only effects, and 8 with AUD-only effects (**Supplemental Tables 4-5**). MAGMA gene-based analyses identified 142 significant genes from the convergent subset and 115 genes from the divergent subset (**Supplemental Figure 2, Supplemental Tables 6-7**).

As ASSET[23] searches for and determines the most likely subset for each SNP (i.e., classifying SNPs as having an effect only on AUD, only on SCZ, or on both disorders), some of the pleiotropic SNPs identified by ASSET were only significant in one of the single-disorder GWAS. For a more conservative description of pleiotropic loci of divergent effect, we further considered only the 9 loci where the top lead SNPs had opposite effects but $p < 0.05$ for both AUD and SCZ (**Supplemental Table 3**; the convergent lead SNPs already had $p < 0.05$ for both disorders). In the convergent subset, the strongest association was on chromosome 1 (lead SNP rs2340405, cross-disorder p = 2.4e-17; AUD p = 5.8e-8, SCZ p = 2.2e-13). The strongest divergent signal was on chromosome 4 (lead SNP rs13135092, cross-disorder p = 2.9e-31; AUD p = 4.9e-18, SCZ p = 7.9e-16), located in an intron of *SLC39A8.* Thus, although rs13135092 is pleiotropic for both AUD and SCZ, the same effect allele (G) *increases risk for* AUD and *decreases risk* (i.e., is protective) for SCZ. Of note, the effect allele at this SNP is also negatively associated with drinks per week (beta = -0.03, p = 2.13e-19), an effect opposite that for AUD.

MAGMA competitive gene-set analyses identified 2 significant GO terms in the convergent subset of variants, one related to DNA binding (GO: 0043565, p = 5.3e-7) and one related to neuronal differentiation (GO: 0045664, p = 5.8e-7; **Supplemental Table 8**). There was one significant gene-set identified for the divergent subset, a curated gene-set consisting of genes down-regulated in the cortex of mice exposed to an enriched learning environment (rampon_enriched_learning_environment_early_dn, p = 2.2e-6; **Supplemental Table 9**).

Both the convergent and divergent subsets of variants showed enrichment in all 13 brain tissues in MAGMA gene-property analyses, and the divergent subset also showed enrichment in pituitary tissues. Using BrainSpan data, the convergent subset of variants showed some enrichment at early mid-prenatal and late mid-prenatal stages, while the divergent subset showed enrichment in brain tissues of individuals at middle adulthood (i.e., 36 years of age) (**Supplemental Figure 3**).

Cross-disorder variants in African ancestry individuals

In the African-ancestry cross-disorder analysis, we identified one genome-wide significant locus of divergent effect (**Supplemental Table 10**), although the lead SNP at this locus had $p > 0.05$ for SCZ. Based on genome-wide pleiotropic variants, there were two significant genes (*ADH1B* and *ADH4*) identified in the MAGMA gene-based analysis; however, these are likely attributable to the AUD GWAS, with little support in the SCZ GWAS (**Supplemental Table 11**).

Integration of eQTL data and differential gene expression in EUR-ancestry samples

Of the 22 genes that survived Bonferroni correction (for the number of genes tested; p_{SMR} < 1.16e-5), SMR analyses identified 12 convergent genes whose cross-disorder associations with AUD and SCZ were consistent with mediation via gene expression in the prefrontal cortex (PFC; the remaining 10 genes with p_{SMR} < 1.16e-5 had p_{HEIDI} < 0.05, indicating pleiotropy outlier status; **Figure 2**; significant results in **Supplemental Table 12;** full results in **Supplemental Table 13**).

Neither the genes with variants of convergent nor divergent effect were enriched in specific differential gene expression analyses of postmortem PFC tissue of individuals with SCZ (N_{cases} = 258) vs. controls or of AUD (N_{cases} = 65) vs. controls.

Genetic covariance and correlation

There were significant positive genetic correlations (r_g) between AUD and SCZ (r_g = 0.392, p = 1.2e-42), and between AUD and DPW (r_g = 0.680, p = 9.2e-181). The r_g between DPW and SCZ (r_g = 0.102) was significantly lower than that between AUD and SCZ (p = 2.9e-6; Z_{diff} for difference in r_gs = 8.58, p = 9.3e-18).

In sensitivity tests, the negative control measure of chronic ischemic heart disease was not significantly correlated with AUD but height showed a small negative correlation (r_g = -0.093, SE = 0.021, p = 7.54e-6). SCZ showed a nominal genetic correlation with heart disease (r_g = -0.063, SE = 0.029, p = 0.032), but none with height. There were no significant findings when partitioning the genetic covariance between AUD and height or SCZ and heart disease into tissue-specific categories with GNOVA.

LDSC-SEG analyses that partitioned heritability enrichment by tissue-specific gene expression revealed significant enrichment for SCZ in three of 13 brain regions: the cortex, frontal cortex, and anterior cingulate cortex (FDR q-values < 2e-5), and for AUD in the anterior cingulate cortex (FDR q-value = 0.007; **Supplemental Figure 4**).

The genetic covariance (ρ_g) of AUD and SCZ was significantly attributable both to functional regions of the genome (p = 1.0e-8), including those specifically functional in brain tissues (p = 0.001; see **Figure 3a, Supplemental Table 15**) and non-functional regions (p = 1.5e-19; **Supplemental Table 14**), as well as all minor allele frequency quartiles except the lowest frequency quartile (p = 2.9e-5 - 3.9e-9; **Supplemental Table 16**). While the point estimate of genetic covariance was highest in genes functional in immune tissues (ρ_g = 0.017), this estimate had a relatively large standard error (0.007) and did not reach significance after multiple testing corrections. In contrast, the only GNOVA annotation category enriched for ρ_g between DPW and SCZ was non-functional regions of the genome (p = 4.8e-4; **Supplemental Tables 17-19**).

Because the genetic covariance of AUD and SCZ was significantly concentrated in brain tissues, we partitioned it further into specific brain regions, using the 13 GTEx brain region annotations provided by LDSC-SEG[46]. We found significant concentrations of genetic covariance in all 13 brain regions tested, with the greatest concentrations in the anterior cingulate cortex (p = 7.6e-12), frontal cortex (p = 3.3e-9), cortex (p = 1.2e-7), and amygdala (p = 3.1e-9; **Supplemental Table 20, Figure 3b**). In contrast, no specific brain regions showed statistically significant genetic covariance between DPW and SCZ after multiple testing (**Supplemental Table 21**).

Testing for putative causal relationships using LCV analysis

We found no evidence for a genetic causal relationship, either full or partial, between the liability for AUD and for SCZ, in either direction (Z-score for partial genetic causality = -0.032; p-value for null hypothesis that genetic causality proportion (GCP) = 0) = 0.975.)

Comparison of effect size distributions of AUD, DPW, and SCZ

Assuming a single normal distribution of variant effect sizes for SCZ, AUD, and DPW, we observed evidence of (1) a significantly greater number of susceptibility loci (π_c p_{diff} = 4.2e-6) and (2) a significantly greater effect size variance (σ^2 p_{diff} = 1.5e-4) in SCZ than AUD, indicating that AUD is more polygenic than SCZ. Although the proportion of susceptibility loci (π_c) for DPW, the GWAS with the largest sample size, was comparable to that for SCZ, the effect size of these risk loci was significantly smaller than both AUD (σ^2 p_{diff} = 0.005) and SCZ (σ^2 p_{diff} = 3.8e-28) (**Figure 4a & 4b**).

In GWAS sample-size projections, SCZ risk locus discovery and explained phenotypic variance were predicted to increase rapidly with larger samples (9,546 risk loci at *N*=5,000,000). However, at comparable sample sizes, AUD (2,381 risk loci) and DPW (3,173 risk loci) were predicted to show more modest improvements in projected GWAS metrics (**Figure 4c & 4d**).

Discussion

Utilizing a subset-based meta-analytic approach, our cross-disorder analysis of AUD and SCZ identified loci with either the same or opposite direction of effects on AUD and SCZ; 55 convergent loci and 9 divergent loci had lead SNPs that were p < 0.05 in both individual disorders. The genetic covariance between AUD and SCZ was concentrated in genes functional in the brain, and eQTL analyses of the convergent subset of variants (i.e., those with the same direction of effect on both disorders) identified 12 genes whose association with AUD and SCZ may be mediated via gene expression in brain tissues. Extending prior studies[10,13], we found that even though the genetic correlation between SCZ and DPW was significant, it was lower

than the SNP-rg between AUD and SCZ and with little evidence of biologically meaningful genetic overlap between SCZ and DPW, suggesting that SCZ shares genetic variation primarily with the psychopathological aspects of problem drinking and AUD.

The top divergent SNP, rs13135092, showed strong associations with both AUD (p = 4.9e-18) and SCZ (7.9e-16), with the G allele exerting a risk-increasing effect on AUD and a protective effect on SCZ. This SNP has also shown associations with cognitive performance, body mass index, and risky behaviors, amongst other complex traits[47]. There are fewer previous trait associations for the strongest SNP with convergent effects on AUD (p = 5.8e-8) and SCZ (p = 2.2e-13), rs2340405, but it has been previously linked to other psychiatric disorders including major depression (though not genome-wide significant: p = 6.5e-7), as well as the contrast between schizophrenia and bipolar disorder[47,48].

Linking cross-disorder associations to eQTLs in the prefrontal cortex (PFC) suggested that the effects of pleiotropic variants on AUD and SCZ may be mediated by expression levels of several genes (**Figure 2**). Several of the identified genes were previously implicated in GWAS of metabolic traits (including *NAT8*, *TRPS1* (up-regulated), *BAG5*, *PPP1R13B* (down-regulated)), immunological traits (e.g., *RERE, TRPS1,* both up-regulated) and psychiatric phenotypes (e.g., up-regulated: *LRP8,* down-regulated: *PPP1R13B*). Although these findings may not be limited to gene expression in the PFC, the limited data currently available do not permit examination of whether these findings extend to other brain regions.

While there are limited samples available with measured gene expression in post mortem brain tissue, particularly for substance use disorders, we used two independent datasets (alcohol dependent and control samples from New South Wales Tissue Resource Centre at the University of Sydney, schizophrenia and control samples from the CommonMind consortium) to link pleiotropic variants to differential gene expression. We found no evidence for enrichment of genes pleiotropically associated with both AUD and SCZ when considering genes differentially expressed in the brains of individuals with AUD or SCZ (compared to respective controls). However, our lack of findings could be due to limited statistical power, particularly for the AUD dataset (N_{cases} = 65).

While the genetic contributions to SCZ are known to be concentrated in genes expressed in the brain and central nervous system, this has not been thoroughly examined for alcohol dependence phenotypes. When we partitioned the genetic covariance between AUD and SCZ, it was concentrated in brain-specific functional regions of the genome (**Figure 3**). The significant LDSC-SEQ enrichment in the anterior cingulate cortex (**Supplemental Figure 4**)

lends further support to the idea that AUD, for which excessive alcohol consumption is a necessary but not sufficient condition, is influenced by genetic variants expressed in the brain.

In contrast to the strong presence of brain-expressed genes in the AUD-SCZ analyses, we found little evidence of genetic overlap between SCZ and DPW, in line with other recent reports[12,17]. The genetic correlation between SCZ and AUD was significantly larger than the genetic correlation between SCZ and DPW, providing additional evidence that, despite the robust genetic correlation between AUD and DPW (r_g = 0.68, p = 9.2e-181), alcohol use disorder has a distinct underlying genetic etiology and displays stronger patterns of genetic overlap with other psychiatric disorders (e.g., major depression, bipolar disorder)[10] than alcohol consumption. In addition, the very small genetic correlation between alcohol consumption and SCZ argues against our results being due simply to an excess of heavy drinkers in the SCZ sample; rather, our findings are specific to the overlap of SCZ and AUD. We conclude that continuous measures of alcohol intake, particularly in predominantly population-based cohorts, are unlikely to capture the heaviest aspects of habitual drinking consistent with AUD and do not show the same patterns of genetic and biological overlap with psychiatric disorders. Although this finding may seem intuitive, no prior study has statistically contrasted these genetic correlations nor examined whether the genetic correlation between SCZ and DPW is biologically meaningful. The divergence in the relationship between SCZ and AUD versus DPW is in contrast to GWAS of cannabis and nicotine wherein measures of lifetime[49] and regular use[12] as well as heavier (e.g., cigarettes per day[12]) and problem use (e.g., cannabis use disorder[50]) show significant genetic correlations with SCZ of roughly equivalent magnitude.

In our sensitivity analysis, we found a significant negative genetic correlation between AUD and height (r_g = -0.093, SE = 0.021, p = 7.54e-6), which is likely driven in part by the very large sample sizes and polygenic nature of both traits, though there is some evidence for this relationship in the literature. One previous study found a negative genetic correlation between alcohol consumption and childhood height in males[51] (r_g = -0.23, p = 0.002) and another study found a negative correlation between problematic alcohol use and comparative height at age 10 (r_g = -0.09, p = 7.3e-5), though this was not significant when a Bonferroni correction was applied for the 715 traits tested[19].

We did not find evidence supporting any genetically causal relationship between AUD and SCZ. This suggests that the genetic overlap we see for AUD and SCZ is most likely due to horizontal pleiotropy (genetic variants directly influencing both AUD and SCZ) rather than vertical pleiotropy (whereby genetic variants have an indirect effect on one trait *only* through their direct effect on the other trait and the causal relationship between the traits)[52]. The lack of

causal effects is unlikely to be due to inadequate statistical power, as Zhou et al have previously reported causal effects of other psychiatric phenotypes (e.g., major depression), but not SCZ, on AUD[19].

Our findings also highlighted the distinct genetic architecture of AUD relative to SCZ (**Figure 4**), including evidence of higher polygenicity for AUD. This is in line with previous findings from Walters et al.[17], who concluded that alcohol dependence has a genetic architecture similar (according to power calculations) to major depression, which required nearly half a million samples (and more than 135,000 cases) to identify 44 genomic risk loci[53]. Assessment of effect size distributions highlighted a greater proportion of relatively large effect size SNPs underlying SCZ than either AUD or DPW (**Figure 4a & b**). This means that larger samples may be required to achieve comparable power for AUD as has been reached for SCZ. (**Figure 4c**). These differences are likely due to the co-occurrence of different mechanisms (*e.g.,* evolutionary pressures and phenotypic heterogeneity) shaping the genetic architecture of psychiatric disorders and traits[18].

There are several limitations of the current study. First, nearly all gene expression reference datasets and other bioinformatic resources have been trained on European-ancestry samples. In addition, our African-ancestry samples were under-powered relative to the European-ancestry samples, and the SCZ sample more so than the AUD sample. Thus, the African-ancestry results appear to be driven almost entirely by the larger AUD GWAS. Nonetheless, we conducted analyses in the African-ancestry samples because most cross-disorder GWAS rely on European-ancestry samples, generating inequity in downstream analyses of non-European populations. For instance, summary statistics from this cross-disorder GWAS could be used to generate polygenic risk scores in independent cohorts of African-American participants. Second, given the available data, we were limited to studying common genetic variation (minor allele frequencies > 1%); thus, there may be rare variants of importance underlying the comorbidity between AUD and SCZ that are outside the scope of the current study. Further, misreports of drinking behavior may have biased our estimates of the genetic correlation between DPW and SCZ: if individuals under-report their drinking behavior, or were heavy drinkers in the past but abstinent at the time of the study, this could downwardly bias correlations between DPW and SCZ[54]. Future studies should investigate this further. Finally, while most of the AUD samples were screened for SCZ, some individuals among the SCZ samples very likely also had AUD. Unfortunately, there are no data currently available on which SCZ samples were screened for substance use disorders, to permit an estimation of such a bias.

Cross-disorder GWAS provide insights into genetic sources of comorbidity. While prior cross-disorder studies have provided foundational results for common psychiatric disorders, ours is amongst the first studies to focus on identifying pleiotropic loci for a substance use disorder, AUD, and SCZ. The high degree of polygenicity underlying AUD underscores the need for more large, collaborative efforts that gather genotyped samples of problematic alcohol use. To understand the shared biology between substance use disorders and other psychiatric disorders, such efforts must include large numbers of individuals with problematic alcohol use rather than simple measures of consumption, despite the increased burden of diagnostic assessment compared with a one- or two-question survey on consumption. With increasing availability of additional 'omics data (e.g., single-cell RNAseq, epigenome-wide data, larger case-control samples with both genetic and transcriptomic data) and GWAS in diverse populations, these findings provide a firm foundation on which our understanding of the shared and unique biological mechanisms that contribute to both AUD and SCZ risk can be based.

FUNDING

ECJ was supported by grant F32AA027435. RP was supported by R21DA047527 and R21DC018098. FRW was supported by F32MH122058. AA is supported by K02DA032573.

This work was conducted using the summary statistics from the Psychiatric Genomics Consortium's Substance Use Disorders Working group. The Psychiatric Genomics Consortium's Substance Use Disorders (PGC-SUD) working group is supported by MH109532 with funding from NIMH and NIDA.

The GPC was supported by grants R01 MH085548 and R01 MH104964 from the National Institute of Mental Health (NIMH), and genotyping of samples was provided by the Stanley Center for Psychiatric Research at Broad Institute.

COGS was supported by grants R01 MH065571, R01 MH065588, R01 MH065562, R01 MH065707, R01 MH065554, R01 MH065578, R01 MH065558, R01 MH86135, and R01 MH094320 from the National Institute of Mental Health.

Funding support for the Whole Genome Association Study of Bipolar Disorder and the Genome-Wide Association of Schizophrenia Study was provided by the NIMH (R01 MH67257, R01 MH59588, R01 MH59571, R01 MH59565, R01 MH59587, R01 MH60870, R01 MH59566, R01

MH59586, R01 MH61675, R01 MH60879, R01 MH81800, U01 MH46276, U01 MH46289, U01 MH46318, U01 MH79469, and U01 MH79470) and the genotyping of samples was provided through the Genetic Association Information Network (GAIN).

This research is based, in part, on data from the Million Veteran Program, Office of Research and Development, Veterans Health Administration, with support from award #I01 BX003341 and the VISN 4 Mental Illness Research, Education and Clinical Center. This publication does not represent the views of the Department of Veterans Affairs or the United States Government.

FINANCIAL DISCLOSURES

Dr. Kranzler is a member of an advisory board for Dicerna Pharmaceuticals and a member of the American Society of Clinical Psychopharmacology's Alcohol Clinical Trials Initiative, which was supported in the last three years by AbbVie, Alkermes, Dicerna, Ethypharm, Indivior, Lilly, Lundbeck, Otsuka, Pfizer, Arbor, and Amygdala Neurosciences. Dr. Kranzler and Dr. Gelernter are named as inventors on PCT patent application #15/878,640 entitled: "Genotype-guided dosing of opioid agonists," filed January 24, 2018. Drs. Johnson, Kapoor, Hatoum, Zhou, Polimanti, Wendt, R.K. Walters, Lai, Kember, Hartz, Meyers, Peterson, Ripke, Bigdeli, Fanous, C.N. Pato, M.T. Pato, Goate, O'Donovan, J.T.R. Walters, Edenberg, and Agrawal have no disclosures or conflicts of interest to report.

ACKNOWLEDGEMENTS

The Psychiatric Genomics Consortium's Substance Use Disorders (PGC-SUD) working group gratefully acknowledges prior support from NIAAA and thanks all our contributing investigators and study participants who make this research possible. We acknowledge collaborations with the Million Veteran Program, Psychiatric Genomics Consortium Substance Use Disorders Working Group, and Genomic Psychiatry Cohort Investigators. A list of Genomic Psychiatry Cohort investigators and the MVP Core Acknowledgement are provided in the Supplemental Materials. We are grateful to Luke Evans for helpful discussion of the manuscript.

References

1. Whiteford, H. A. *et al.* Global burden of disease attributable to mental and substance use disorders: Findings from the Global Burden of Disease Study 2010. *Lancet* **382**, 1575–1586 (2013).
2. Hunt, G. E., Large, M. M., Cleary, M., Lai, H. M. X. & Saunders, J. B. Prevalence of comorbid substance use in schizophrenia spectrum disorders in community and clinical settings, 1990–2017: Systematic review and meta-analysis. *Drug Alcohol Depend.* **191**, 234–258 (2018).

3. Castillo-Carniglia, A., Keyes, K. M., Hasin, D. S. & Cerdá, M. Psychiatric comorbidities in alcohol use disorder. *The Lancet Psychiatry* **6**, 1068–1080 (2019).
4. SAMHSA. *2018 National Survey on Drug Use and Health (NSDUH)*. (2019).
5. Brady, K. T., Killeen, T. & Jarrell, P. Depression in alcoholic schizophrenic patients. *Am. J. Psychiatry* **150**, 1255–1256 (1993).
6. Duke, P. J., Pantelis, C. & Barnes, T. R. E. South Westminster Schizophrenia Survey. Alcohol use and its relationship to symptoms, tardive dyskinesia and illness onset. *Br. J. Psychiatry* **164**, 630–636 (1994).
7. Drake, R. E. & Wallach, M. A. Substance Abuse Among the Chronic Mentally Ill. *Psychiatr. Serv.* **40**, 1041–1046 (1989).
8. PF, S., KS, K. & MC, N. Schizophrenia as a complex trait: Evidence from a meta-analysis of twin studies. *Arch. Gen. Psychiatry* **60**, 1187–1192 (2003).
9. Verhulst, B., Neale, M. C. & Kendler, K. S. The heritability of alcohol use disorders: a meta-analysis of twin and adoption studies. *Psychol. Med.* **45**, 1061–1072 (2015).
10. Kranzler, H. R. *et al.* Genome-wide association study of alcohol consumption and use disorder in 274,424 individuals from multiple populations. *Nat. Commun.* **10**, 1499 (2019).
11. Reginsson, G. W. *et al.* Polygenic risk scores for schizophrenia and bipolar disorder associate with addiction. *Addiction Biology* (2017). doi:10.1111/adb.12496
12. Liu, M. *et al.* Association studies of up to 1.2 million individuals yield new insights into the genetic etiology of tobacco and alcohol use. *Nat. Genet.* **51**, 237–244 (2019).
13. Sanchez-Roige, S. *et al.* Genome-Wide Association Study Meta-Analysis of the Alcohol Use Disorders Identification Test (AUDIT) in Two Population-Based Cohorts. *Am. J. Psychiatry* (2018). doi:https://doi.org/10.1176/appi.ajp.2018.18040369
14. Lam, M. *et al.* Pleiotropic Meta-Analysis of Cognition, Education, and Schizophrenia Differentiates Roles of Early Neurodevelopmental and Adult Synaptic Pathways. *Am. J. Hum. Genet.* **105**, 334–350 (2019).
15. Lee, P. H. *et al.* Genomic Relationships, Novel Loci, and Pleiotropic Mechanisms across Eight Psychiatric Disorders. *Cell* **179**, 1469-1482.e11 (2019).
16. Ruderfer, D. M. *et al.* Genomic Dissection of Bipolar Disorder and Schizophrenia, Including 28 Subphenotypes. *Cell* **173**, 1705-1715.e16 (2018).
17. Walters, R. K. *et al.* Transancestral GWAS of alcohol dependence reveals common genetic underpinnings with psychiatric disorders. *Nat. Neurosci.* **21**, 1656–1669 (2018).
18. Wendt, F. R. *et al.* Natural selection influenced the genetic architecture of brain structure, behavioral and neuropsychiatric traits. *bioRxiv* 2020.02.26.966531 (2020). doi:10.1101/2020.02.26.966531
19. Zhou, H. *et al.* Genome-wide meta-analysis of problematic alcohol use in 435,563 individuals yields insights into biology and relationships with other traits. *Nat. Neurosci.* (2020). doi:10.1038/s41593-020-0643-5
20. Willer, C. J., Li, Y. & Abecasis, G. R. METAL: fast and efficient meta-analysis of genomewide association scans. *Bioinformatics* **26**, 2190–2191 (2010).
21. Pardiñas, A. F. *et al.* Common schizophrenia alleles are enriched in mutation-intolerant genes and in regions under strong background selection. *Nat. Genet.* **50**, 381 (2018).
22. Bigdeli, T. B. *et al.* Contributions of common genetic variants to risk of schizophrenia among individuals of African and Latino ancestry. *Mol. Psychiatry* 1–13 (2019).
23. Bhattacharjee, S. *et al.* A Subset-Based Approach Improves Power and Interpretation for the Combined Analysis of Genetic Association Studies of Heterogeneous Traits. *Am. J. Hum. Genet.* **90**, 821–835 (2012).
24. Bulik-Sullivan, B. *et al.* An atlas of genetic correlations across human diseases and traits. *Nat. Genet.* **47**, 1236 (2015).
25. Bulik-Sullivan, B. K. *et al.* LD Score regression distinguishes confounding from polygenicity in genome-wide association studies. *Nat. Genet.* **47**, 291–295 (2015).

26. Watanabe, K., Taskesen, E., van Bochoven, A. & Posthuma, D. Functional mapping and annotation of genetic associations with FUMA. *Nat. Commun.* **8**, 1826 (2017).
27. de Leeuw, C. A., Mooij, J. M., Heskes, T. & Posthuma, D. MAGMA: Generalized Gene-Set Analysis of GWAS Data. *PLoS Comput. Biol.* **11**, (2015).
28. Liberzon, A. *et al.* Molecular signatures database (MSigDB) 3.0. *Bioinformatics* **27**, 1739–1740 (2011).
29. Aguet, F. *et al.* Genetic effects on gene expression across human tissues. *Nature* **550**, 204–213 (2017).
30. Miller, J. A. *et al.* Transcriptional landscape of the prenatal human brain. *Nature* **508**, 199–206 (2014).
31. Lee, C. H., Shi, H., Pasaniuc, B., Eskin, E. & Han, B. A method to map and interpret pleiotropic loci using summary statistics of multiple traits. *bioRxiv* 2020.06.16.155879 (2020). doi:10.1101/2020.06.16.155879
32. Zhu, Z. *et al.* Integration of summary data from GWAS and eQTL studies predicts complex trait gene targets. *Nat. Genet.* **48**, 481–487 (2016).
33. De Jager, P. L. *et al.* A multi-omic atlas of the human frontal cortex for aging and Alzheimer's disease research. *Sci. Data* **5**, 180142 (2018).
34. Akbarian, S. *et al.* The psychencode project. *Nat. Neurosci.* **18**, 1707 (2015).
35. Kapoor, M. *et al.* Analysis of whole genome-transcriptomic organization in brain to identify genes associated with alcoholism. *Transl. Psychiatry* **9**, 89 (2019).
36. Zhu, Z. *et al.* Causal associations between risk factors and common diseases inferred from GWAS summary data. *Nat. Commun.* **9**, 224 (2018).
37. Fromer, M. *et al.* Gene expression elucidates functional impact of polygenic risk for schizophrenia. *Nature Neuroscience* **19**, 1442–1453 (2016).
38. Fisher, R. A. Statistical methods for research workers. *Stat. methods Res. Work.* (1934).
39. Lonsdale, J. *et al.* The Genotype-Tissue Expression (GTEx) project. *Nat. Genet.* **45**, 580–585 (2013).
40. Lu, Q. *et al.* A Powerful Approach to Estimating Annotation-Stratified Genetic Covariance via GWAS Summary Statistics. *Am. J. Hum. Genet.* **101**, 939–964 (2017).
41. Lu, Q. *et al.* A statistical framework to predict functional non-coding regions in the human genome through integrated analysis of annotation data. *Sci. Rep.* **5**, 10576 (2015).
42. Lu, Q. *et al.* Systematic tissue-specific functional annotation of the human genome highlights immune-related DNA elements for late-onset Alzheimer's disease. *PLoS Genet.* **13**, e1006933 (2017).
43. Lu, Q., Powles, R. L., Wang, Q., He, B. J. & Zhao, H. Integrative tissue-specific functional annotations in the human genome provide novel insights on many complex traits and improve signal prioritization in genome wide association studies. *PLoS Genet.* **12**, e1005947 (2016).
44. O'Connor, L. J. & Price, A. L. Distinguishing genetic correlation from causation across 52 diseases and complex traits. *Nat. Genet.* **50**, 1728–1734 (2018).
45. Zhang, Y., Qi, G., Park, J.-H. & Chatterjee, N. Estimation of complex effect-size distributions using summary-level statistics from genome-wide association studies across 32 complex traits. *Nat. Genet.* **50**, 1318–1326 (2018).
46. Finucane, H. K. *et al.* Heritability enrichment of specifically expressed genes identifies disease-relevant tissues and cell types. *Nat. Genet.* **50**, 621–629 (2018).
47. Watanabe, K. *et al.* A global overview of pleiotropy and genetic architecture in complex traits. *Nat. Genet.* (2019). doi:10.1038/s41588-019-0481-0
48. Ruderfer, D. M. *et al.* Polygenic dissection of diagnosis and clinical dimensions of bipolar disorder and schizophrenia. *Mol. Psychiatry* **19**, 1017–1024 (2014).
49. Pasman, J. A. *et al.* GWAS of lifetime cannabis use reveals new risk loci, genetic overlap with psychiatric traits, and a causal influence of schizophrenia. *Nat. Neurosci.* **21**, 1161–

1170 (2018).
50. Demontis, D. *et al.* Genome-wide association study implicates CHRNA2 in cannabis use disorder. *Nat. Neurosci.* 1 (2019).
51. Clarke, T.-K. *et al.* Genome-wide association study of alcohol consumption and genetic overlap with other health-related traits in UK Biobank (N=112,117). *Molecular Psychiatry* (2017). doi:10.1038/mp.2017.153
52. Hemani, G., Bowden, J. & Davey Smith, G. Evaluating the potential role of pleiotropy in Mendelian randomization studies. *Hum. Mol. Genet.* **27**, R195–R208 (2018).
53. Wray, N. R. *et al.* Genome-wide association analyses identify 44 risk variants and refine the genetic architecture of major depression. *Nat. Genet.* **50**, 668–681 (2018).
54. Xue, A. *et al.* Genome-wide analyses of behavioural traits biased by misreports and longitudinal changes. *medRxiv* 2020.06.15.20131284 (2020). doi:10.1101/2020.06.15.20131284

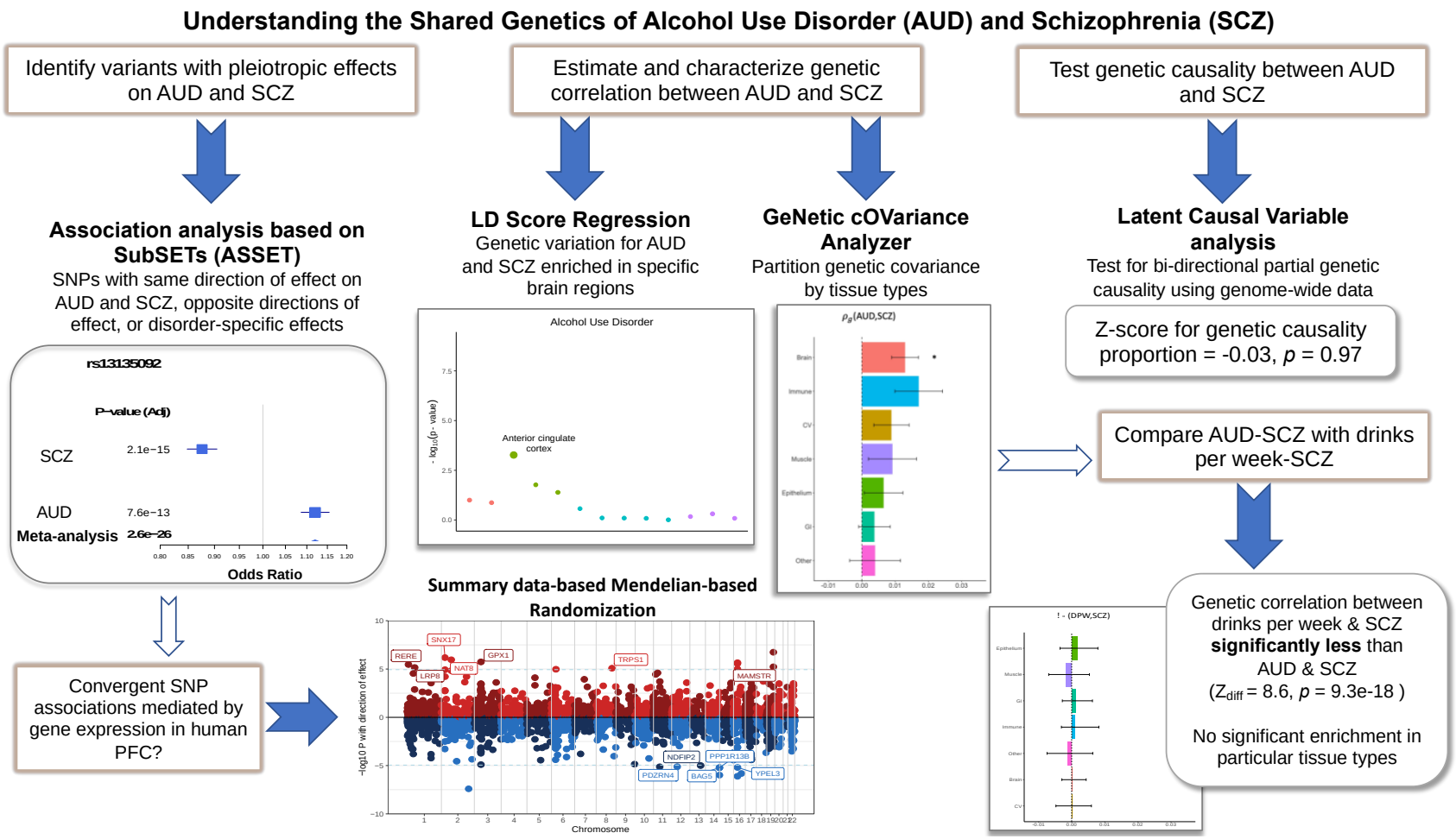


Figure 1. Conceptual overview of cross-disorder analysis of alcohol use disorder and schizophrenia.

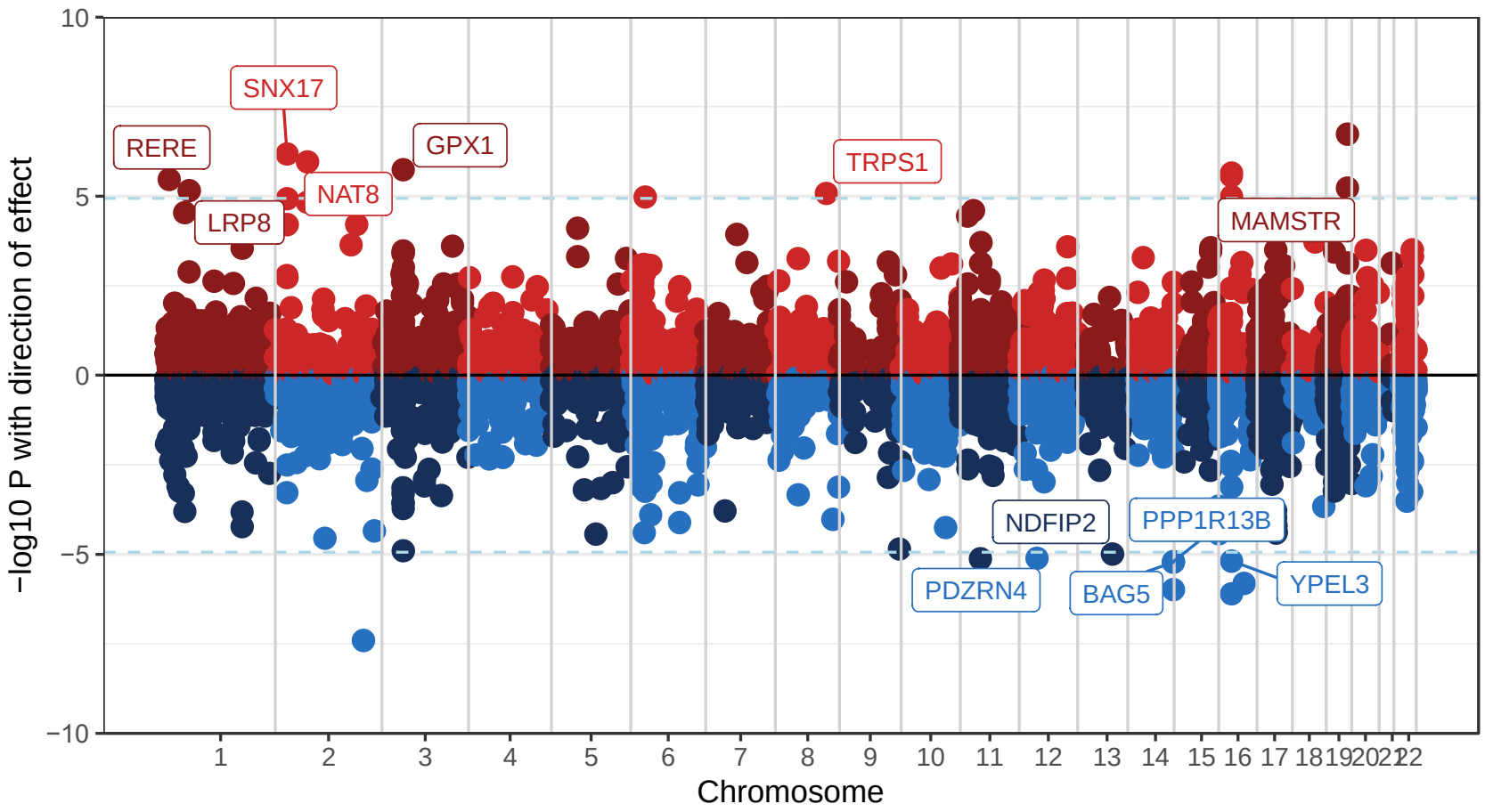
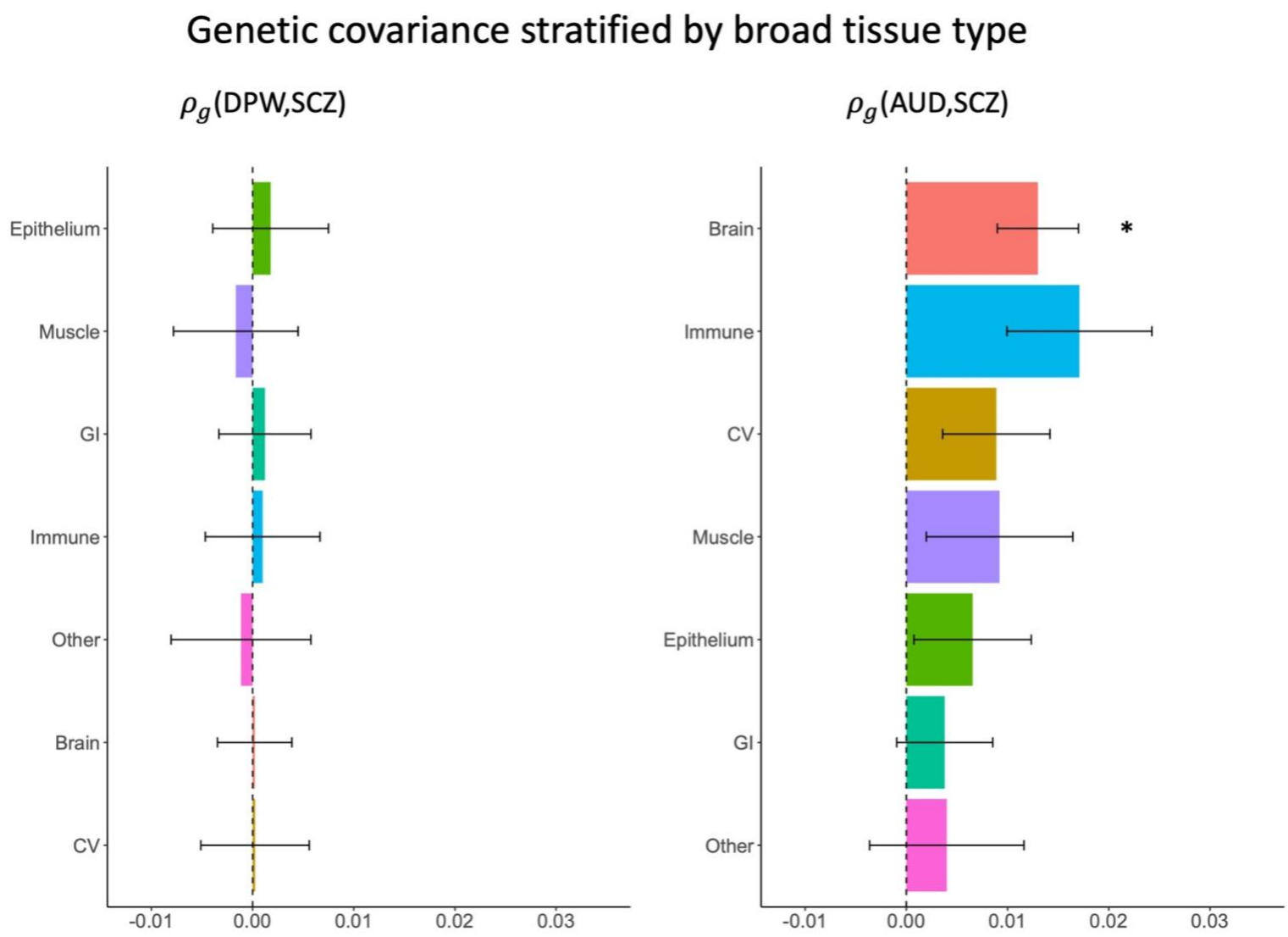
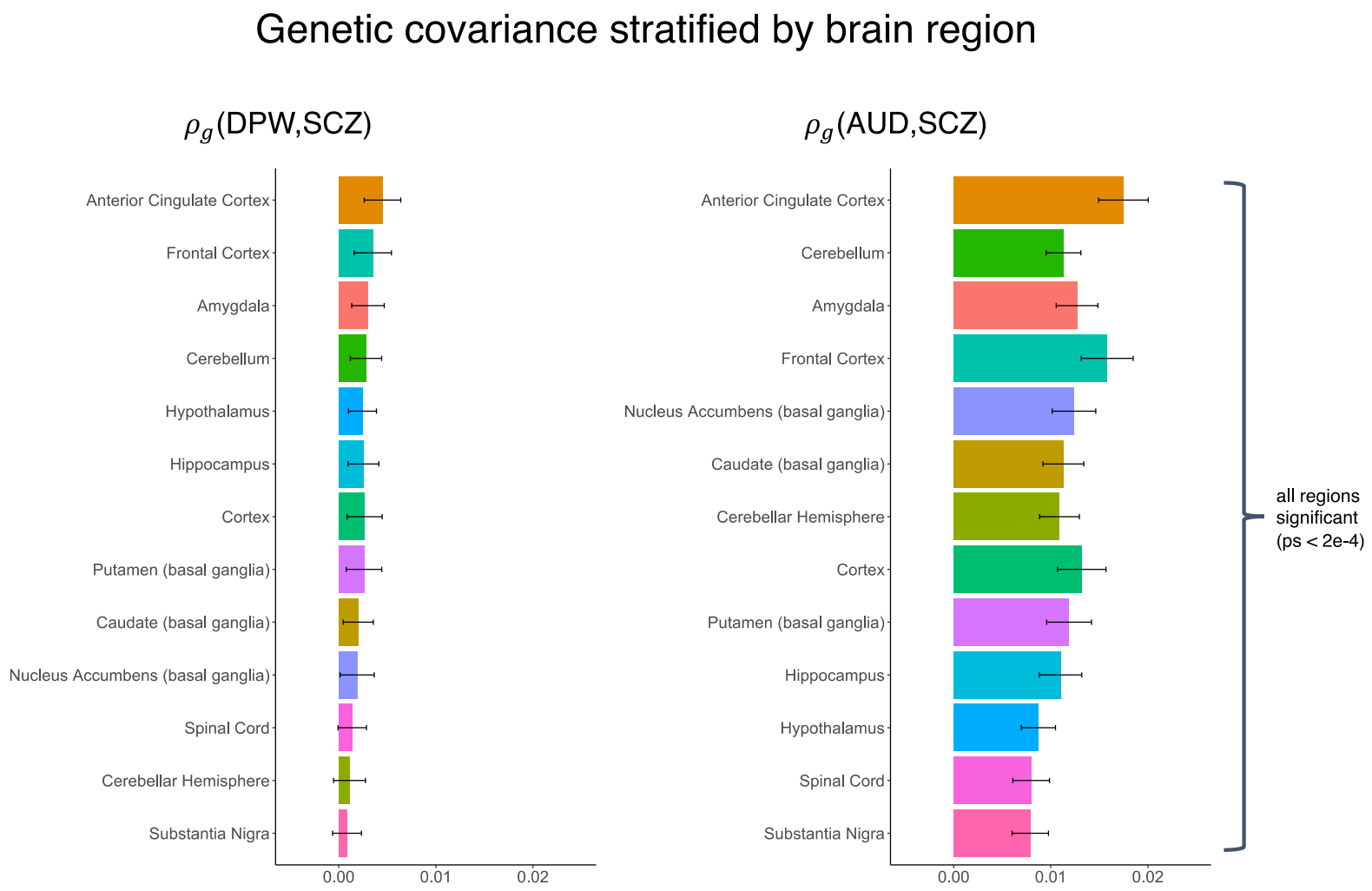


Figure 2. Genes in the convergent subset whose association with AUD and SCZ may be mediated by gene expression in the prefrontal cortex, analyzed using Summary data-based Mendelian Randomization (SMR). Genes in red show up-regulated gene expression, and genes in blue show down-regulated gene expression. The 12 labeled genes are significant after Bonferroni corrections and were not excluded as pleiotropic outliers (HEIDI-outlier method p > 0.05)


Genetic covariance stratified by broad tissue type
ρ_g(DPW,SCZ)
ρ_g(AUD,SCZ)
Epithelium
Muscle
GI
Immune
Other
Brain
CV
-0.01
0.00
0.01
0.02
0.03
Brain
Immune
CV
Muscle
Epithelium
GI
Other
*
-0.01
0.00
0.01
0.02
0.03

a.



b.

Figure 3. Stratified genetic covariance between DPW and SCZ (left) and AUD and SCZ (right). a: stratified by broad tissue type. Tissue annotations were defined using the GenoSkyline-Plus annotations. Significant tissues are starred. **b:** stratified by 13 brain regions, defined using GTEx v6 gene expression data. No tissues were significant for DPW and SCZ after correction for sample overlap and multiple testing.

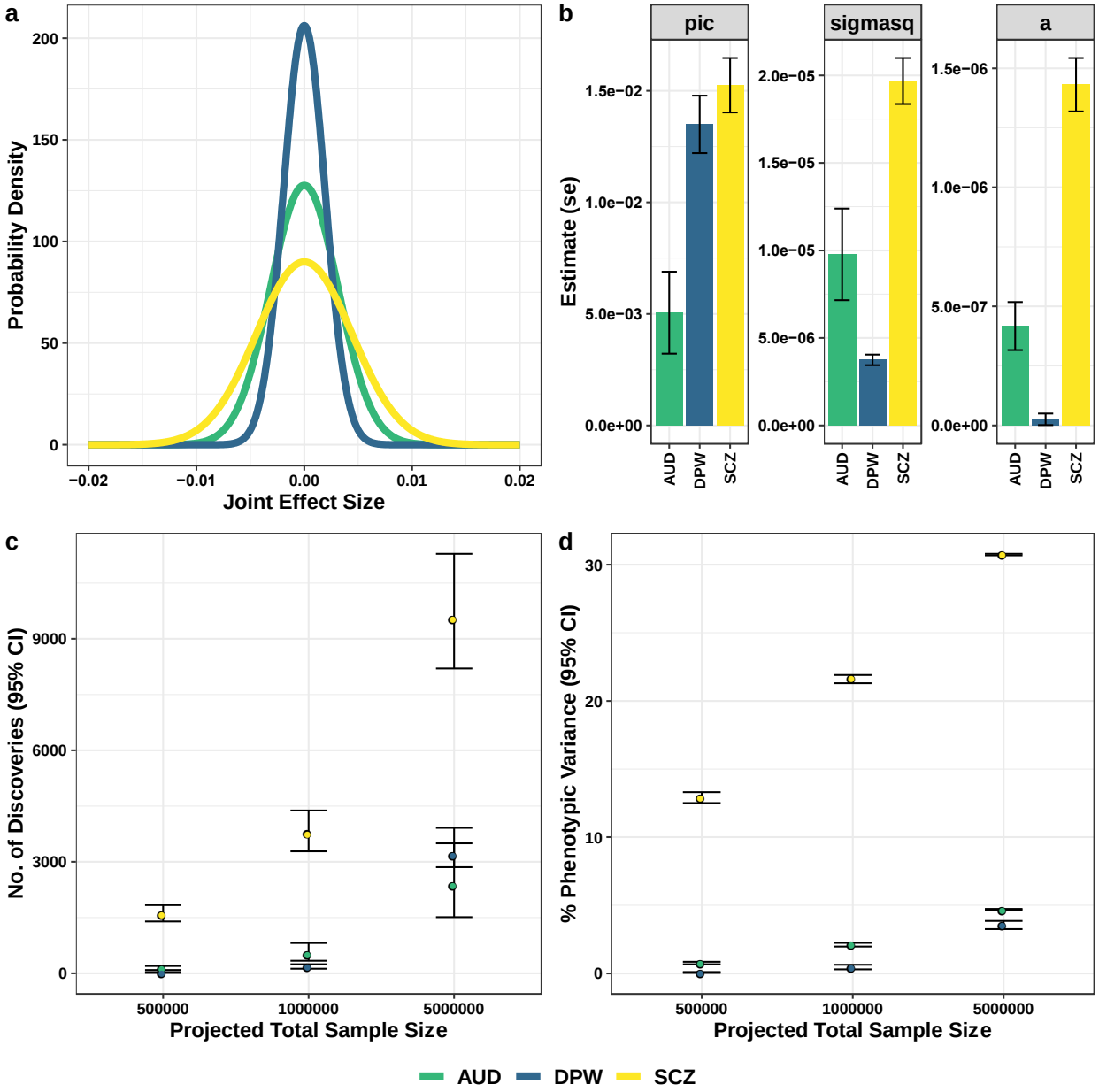


Figure 4. **Effect size distributions and future parameter projections for SCZ, AUD, and DPW GWAS. a.** Effect size distributions based on a two-component model. **b.** Effect size distribution descriptive statistics (per GWAS, π_c ("pic") describes the proportion of susceptibility SNPs, σ^2 ("sigmasq") describes the variance in effect size for non-null SNPs, and α ("a") describes residual effects not captured by the variance of effect-sizes). **c** and **d** describe projected number of risk locus discoveries and explained phenotypic variance in hypothetical future large GWAS of SCZ, AUD, and DPW. The simulations in panels **c** and **d** assume a 1:1 case-control balance, and the % variance explained in panel **d** is a reflection of genome-wide significant SNPs only (defined as 5e-8).