Advancing Gene Discovery for Substance Use Disorders Using Additional Traits Related to Behavioral Disinhibition

Holly E. Poore, PhD[1], Chris Chatzinakos, PhD[2], Travis T. Mallard, PhD[3,4], Sandra Sanchez-Roige, PhD[5,6,7], Fazil Aliev, PhD[1], Alexander Hatoum, PhD[8], COGA Collaborators*, Irwin D. Waldman, PhD[9], Abraham A. Palme, PhD r[5,7], K. Paige Harden, PhD[10,11], Peter B. Barr, PhD[2†], and Danielle M. Dick, PhD[1†]

1. Department of Psychiatry, Robert Wood Johnson Medical School, Rutgers University
2. Department of Psychiatry and Behavioral Science, SUNY Downstate Health Sciences University
3. Center for Precision Psychiatry, Department of Psychiatry, Massachusetts General Hospital
4. Department of Psychiatry, Harvard Medical School
5. Department of Psychiatry, University of California San Diego
6. Department of Medicine, Vanderbilt University Medical Center
7. Institute for Genomic Medicine, University of California San Diego
8. Department of Psychiatry, Washington University School of Medicine
9. Department of Psychology, Emory University
10. Department of Psychology, University of Texas at Austin
11. Population Research Center, University of Texas at Austin

* A list of authors and their affiliations appears at the end of the paper

[†] These authors jointly supervised this work.

Correspondence should be addressed to Holly E. Poore, 671 Hoes Lane W, Piscataway, NJ 08854. Email: holly.poore@rutgers.edu and Danielle M. Dick 671 Hoes Lane W, Piscataway, NJ 08854. Email: danielle.m.dick@rutgers.edu

Key Points

Question: Can we advance gene discovery for substance use disorders (SUDs) by incorporating information about correlated outcomes related to behavioral disinhibition?

Findings: Combining genetic effects common to SUDs and behavioral disinhibition in a multivariate genome-wide association study of > 2.2 million individuals identified 708 genomic risk loci and accounted for more variance in SUD phenotypes compared with modeling each set of phenotypes separately.

Meaning: SUDs and behavioral dysregulation are influenced by a common set of common variants; modeling their joint contributions improves power for genetic discovery and polygenic prediction.

Abstract

Importance: Substance use disorders (SUDs) frequently co-occur with each other and with other traits related to behavioral disinhibition, a spectrum of outcomes referred to as externalizing. Nevertheless, genome-wide association studies (GWAS) typically study individual SUDs separately. This single-disorder approach ignores genetic covariance between SUDs and other traits and may contribute to the relatively limited genetic discoveries to date.

Objective: To identify the most effective model for capturing genetic relationships between SUDs and externalizing phenotypes, optimizing the detection of genetic influences on SUDs while maintaining specificity.

Design: We used Genomic SEM to estimate SNP effects on a broad factor representing liability to externalizing and SUDs, on factors representing liability to behavioral disinhibition and SUDs separately, and on residualized SUDs. Subsequent gene-based, tissue expression, and polygenic score (PGS) analyses were used to compare the ability of these alternative approaches to identify genetic influences on SUDs.

Setting: This study was carried out from May 2023 - September 2024.

Participants: We used GWAS summary statistics based on samples of European ancestry from previous studies of externalizing and SUD phenotypes in the main multivariate GWAS (N > 2.2 million). We used two independent samples to estimate polygenic associations, a family-based sample enriched for substance use problems (COGA; N = 7,530) and a population-based sample representative of the United States, (All of Us; N = 77,442).

Exposures: N/A

Main Outcomes and Measures: Across the three factors (Externalizing; SUDs; Behavioral Disinhibition) and four residualized SUDs (alcohol, tobacco, opioid, and cannabis), we compared the number, putative function, previous associations of significant genomic risk loci and genes, and variance explained by polygenic scores in substance use outcomes.

Results: We identified genomic risk loci and genes uniquely associated with Externalizing that are relevant to the neurobiology of substance use. Genes identified for residual SUDs were involved in substance-specific processes (e.g., metabolism). The Externalizing PGS accounted for the most variance in substance outcomes relative to the PGS for the other factors and residual PGS appeared to capture substance specific signals.

Conclusions and Relevance: Our findings suggest that modeling both a broad genetic liability to externalizing behaviors and substance-specific liabilities enhances the detection of genetic effects related to SUDs and explains more variance in substance use outcomes.

SUDs frequently co-occur, and twin studies indicate that this overlap is due in large part to shared genetic influences. This shared genetic influence, which broadly impacts SUD risk, accounts for up to 74-80% of the genetic influences on alcohol use disorders and 62-74% of genetic influences on other substance use disorders[1,2], with the remaining genetic risk being substance specific. Despite this, gene-identification efforts most commonly study SUDs individually, potentially hampering our ability to identify genes involved in SUDs. More recently, multivariate genome-wide association studies[3,4] (GWAS) have been applied to SUDs, providing further evidence that most genomic risk is shared[5].

SUDs also frequently co-occur with other psychiatric disorders and behavioral traits characterized by behavioral disinhibition, such as childhood conduct disorder, adult antisocial behavior, and personality traits related to impulsivity[6,7]. Here too, twin data suggest that shared genetic influences contribute to this overlap, with these traits and disorders loading together with SUDs on a common underlying shared genetic factor[8,9]. In twin studies, this underlying latent factor is highly heritable (~80%), moreso than any of the disorders or traits studied individually. In the psychological literature, this spectrum of behaviors and disorders, including SUDs, is typically referred to as *Externalizing*[7,10].

In the current study, we leverage genetic correlations among SUDs and related behavioral disinhibition phenotypes to improve power to detect genetic effects on SUDs[11]. This builds our previous work[12], in which we modeled the relationship between SUDs and behavioral disinhibition traits. The best fitting models were (1) a common factor in which all SUDs and traits related to behavioral disinhibition loaded onto a single factor, and (2) a two-factor model in which the factor representing shared SUD risk correlated .9 with a latent factor underlying other disorders and traits related to behavioral disinhibition. Here, we extend this work by conducting multivariate genomic analyses from data on > 2.2 million individuals of these three factors, representing broad externalizing, behavioral disinhibition, and SUDs.

The value of jointly analyzing genetically correlated traits to improve power is evidenced by previous multivariate genomic investigations, including from our group, in which we analyzed seven externalizing related traits and identified over 550 independent genetic loci that were significantly associated with broad externalizing risk. Nevertheless, use of an overly broad model may result in identification of genetic variants that are related to psychopathology more broadly, rather than specific to SUDs. Thus, in the current study, we carry forward both the common and two-factor models from our previous work and evaluate the ability of each to increase gene-identification for without sacrificing specificity of SUDs. To evaluate power, we compared the magnitude of genetic signal, number of significant genomic risk loci and genes, and variance accounted for by resulting polygenic scores. To evaluate specificity, we explored the functions and previous associations of identified risk loci and genes and differential tissue expression and polygenic prediction of specific phenotypes in two independent samples.

Methods

Results are reported using STREGA reporting guidelines[13].

Multivariate GWAS

We used Genomic SEM[11] to estimate SNP effects in two models (Figure 1), identified in our recent paper[12], which were drawn from previous separate GWAS of externalizing[14] and addiction risk[3] with a total sample size of 2,219,357 individuals (Supplementary Table 1). In our two-factor model, the SUD factor included problematic alcohol use PAUD)[15], problematic tobacco use (PTU)[4], opioid use disorder (OUD)[16], and cannabis use disorder (CUD)[17]. The remaining non-SUD indicators from the original externalizing model (attention deficit hyperactivity disorder [ADHD][18], risk taking [RISK][19], number of sexual partners [NSEX][19], age at first sex [FSEX][19], smoking initiation [SMOK][20], and cannabis initiation [CANN][21]) loaded onto a behavioral disinhibition factor (boxes B and C in Figure 1). In our common factor model, all behavioral disinhibition phenotypes and SUDs loaded onto a common externalizing factor (Box

A in Figure 1). We retained the same GWAS as were used in the respective original multivariate GWAS except in the case of OUD, for which a better powered GWAS was available.

We performed multivariate genome-wide association analyses by estimating SNP effects on each of the latent factors (boxes A, F, and G in Figure 1) and the residual SUDs in the single factor model (boxes B-E in Figure 1). These residual GWAS capture genetic influences unique to that disorder after accounting for what it shared with other behavioral disinhibition phenotypes and SUDs. All GWAS included only individuals whose genomes were most similar to those from reference panels sampled from Europe (hereto referred to as “EUR”). Clumping was performed in Functional Mapping and Annotation of Genome-Wide Association Studies (FUMA) version 1.6.1[22] using an r^2 threshold of ≥ 0.6 to define independent significant SNPs, a second threshold of $r^2 \geq 0.1$ to define lead SNPs, and a maximum distance between LD blocks of 250kb to merge into a locus.

To characterize differences in statistical power among the multivariate GWAS, we examine the mean χ^2, λ_{GC}, and number of genome-wide significant risk loci for each factor and residual SUD. We evaluated the novelty of our findings by comparing the genomic risk loci identified in our analyses to 1) loci identified in the original externalizing and addiction risk GWAS and 2) loci previously identified for substance use phenotypes in the GWAS literature. This latter test was performed by comparing the genomic risk loci for our factors and correlated SNPs ($r^2 > 0.1$) to those in the NHGRI-EBI GWAS Catalog[23] (version e111_r2024-05-05). Finally, we assessed the relative performance of our two models by comparing the degree of heterogeneity of SNP effects. We did so by calculating Q_{SNP} heterogeneity statistics, which can be used to identify SNPs that have an effect on one or more indicator phenotypes that is better explained by pathways independent of the factor. If a factor is truly capturing the majority of genetic variance shared among its indicator phenotypes, there will be few Q_{SNP} loci relative to the number of factor loci.

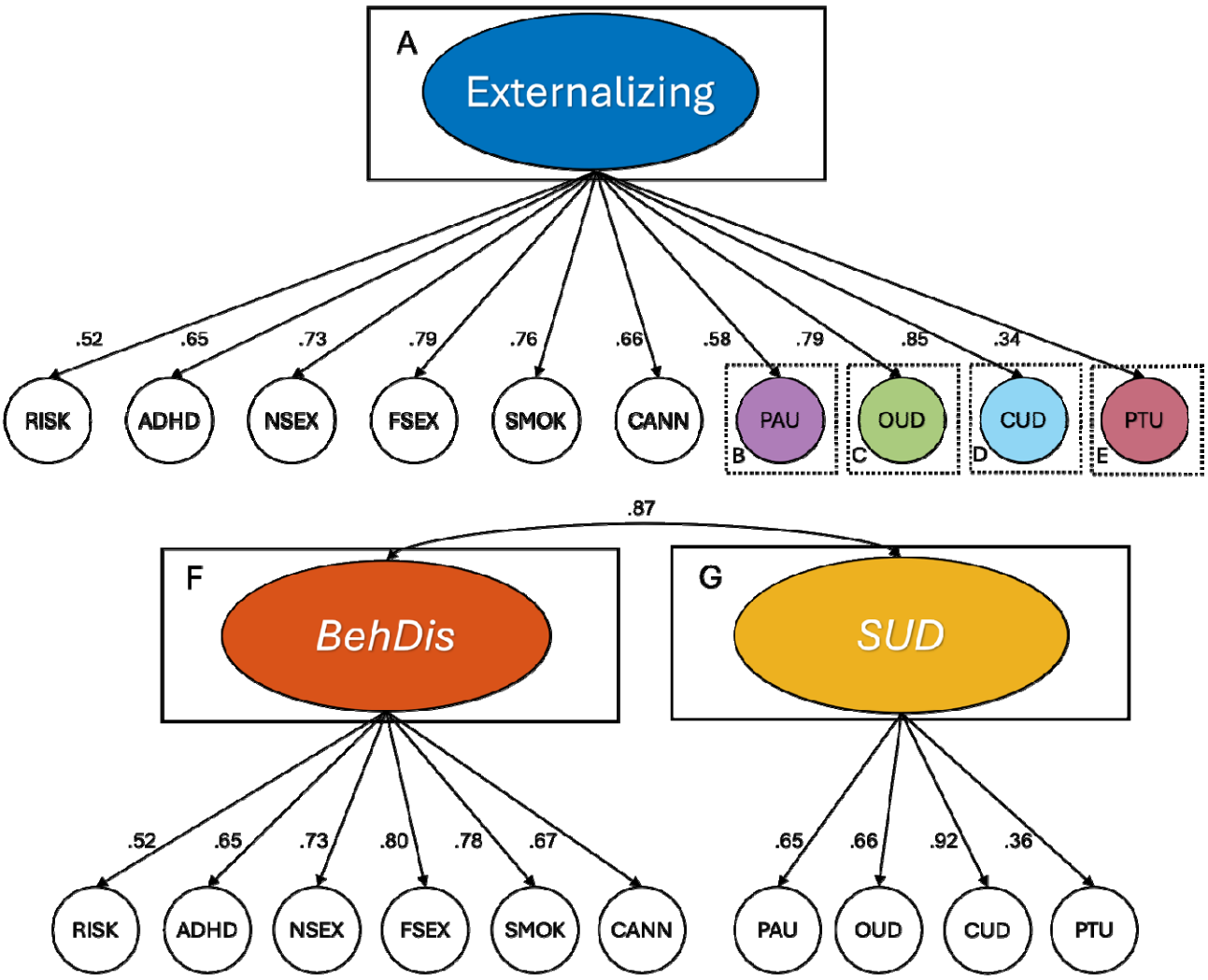

Figure 1. Path diagrams of models used in the current analyses. Box A represents the broad Externalizing factor onto which all behavioral disinhibition and SUD phenotypes load. Boxes B-E represent residual SUD phenotypes. Boxes F and G represent narrower factors reflecting Behavioral Disinhibition and SUD phenotypes, respectively. Single headed arrows indicate factor loadings whereas the double headed arrow indicates a correlation between the two factors.

Biological annotation

Gene-based methods. We used three methods to identify genes associated with the three latent genomic factors. First, we used multi-marker analysis of genomic annotation (MAGMA; version 1.08)[24], in which genome-wide SNPs were mapped to 18,235 protein-coding genes from Ensembl v102, and SNPs within each gene were jointly tested for association with each factor. We evaluated Bonferroni corrected significance adjusted for the number of genes (one sided $p < 2.74 \times 10^{-6}$). Next, we used MetaXcan[25] to conduct a Transcriptome-Wide Association Study (TWAS) using genetically regulated expression models from GTEx v8[26]. This analysis leveraged GWAS summary statistics to estimate gene-trait associations. Within-tissue Bonferroni correction was applied to identify statistically significant TWAS genes. Finally, we used summary-data-based Mendelian randomization (SMR)[27] to 1) test the extent to which gene expression mediated the relationship between SNPs and the phenotype and 2) distinguish

causality and pleiotropy models from the linkage model using the heterogeneity in dependent instruments (HEIDI) test. The latter test was used to identify genes that are more likely to be functionally relevant to the phenotype and should therefore be prioritized for follow up. We identified genes of interest as those that met SMR test Bonferroni correction significance threshold and had a HEIDI test p-value > .05.

To account for differences in gene-based mapping methods and ensure higher confidence in the genes associated with each factor, we further evaluated the intersection of genes that were identified in all three gene-based analyses. As our primary goal is to improve gene-identification for SUDs, we explored the specificity of genes associated with each factor as well as evidence that these genes are relevant to the neurobiological pathways implicated in SUDs. To do this, we quantified the number of genes that were unique to each factor (i.e., associated with one factor and not the other two factors) and explored the traits with which these genes have been previously associated. We also used MAGMA to identify genes associated with each residual SUD phenotype.

Tissue expression. We also used MAGMA tissue expression analysis to test the relationship between expressed genes in brain, liver, and lung tissues and genetic associations across the three factors. We used weights from GTEx v8 for 15 brain, liver, and lung tissue types[26]. We report standardized beta coefficients of the associations to highlight any differences in direction or magnitude of associations across the three factors.

Polygenic scores

We calculated polygenic scores from GWAS of each latent genomic factor (Externalizing, Behavioral Disinhibition, and SUD) among EUR individuals from the Collaborative Study on the Genetics of Alcoholism (COGA; *N* =7,530), a multi-site family-based study[28] and All of Us (N_{max} = 77,442), a national cohort study[29,30] (see Supplementary Material for addition sample description). We also calculated PGS for the residual SUD phenotypes (alcohol, opioid, tobacco, and cannabis use disorders) in COGA. These analyses allowed us to compare the total variance explained by each of the PGS and identify the specificity, or loss thereof, in polygenic prediction of substance use phenotypes. In COGA we included phenotypes related to substance initiation, consumption, and use disorders. In All of Us we analyzed diagnoses of alcohol use disorder, tobacco use disorder, other drug disorder (DUD; i.e., use disorder for all other drugs, including opioids or cannabis), major depressive disorder, bipolar disorder, and schizophrenia using phecodes in participants' linked electronic health records (EHR). We included data from release 7 (May 6, 2018 to February 23, 2023) on those with available EHR and whole genome sequence data.

We used PRS-CS[31] to adjust original GWAS beta weights for linkage disequilibrium and Plink2[32] to construct each PGS from these weights. We evaluated the incremental R^2/pseudo-R^2 (ΔR^2) attained by adding the polygenic score to a regression with baseline covariates (e.g., age, sex, and ancestry PCs). We used least squares regression for continuous outcomes and logistic regression for categorical ones and adjusted the standard errors in COGA to account for the family structure. We estimated 95% CIs for ΔR^2 using bootstrapping (1,000 iterations).

Results

Multivariate GWAS

Phenotype	Mean χ^2	λ_{GC}	GWS Risk Loci
Externalizing	3.45	2.58	708
Behavioral Disinhibition	3.26	2.49	631
SUD	1.42	1.34	48
Residual AUD	1.21	1.18	14
Residual OUD	1.10	1.09	0
Residual PTU	1.25	1.18	23
Residual CUD	1.03	1.04	1

After merging across the ten sets of summary statistics, 5,963,905 SNPs were available for analysis. Table 1 shows the mean χ^2, λ_{GC}, LD score intercept, and number of genome-wide significant risk loci for each factor and residual phenotype. Relative to the Behavioral Disinhibition and SUD factors, the Externalizing factor GWAS produced higher mean χ^2 and λ_{GC} values, suggesting that it is better powered to detect genetic influences.

We identified 708, 631, and 48 genomic risk loci for the Externalizing, Behavioral Disinhibition, and SUD factors, respectively (Figure 2; Supplementary Table 2). Of the 708 Externalizing loci and their correlates within LD regions (r^2 > .1), 187 (26%) were not identified in the previous Externalizing[14] or Addiction Risk[3] GWAS and 424 (60%) were not previously associated with a substance use trait. Although we identified the greatest number of risk loci using the broad Externalizing factor, we observed an increased number of hits for Behavioral Disinhibition and SUD relative to their original multivariate GWAS as well. Of the 631 Behavioral Disinhibition loci and their correlates, 94 (15%) were not identified in the previous Externalizing GWAS. Similarly, of the 48 SUD loci and their correlates, 33 (69%) were not identified in the previous Addiction Risk GWAS, and 6 (13%) were not previously associated with any substance use trait in the GWAS Catalog. Finally, we identified 14, 23, and 1 genomic risk loci for the residual AUD, PTU, and CUD phenotypes, respectively, whereas the residual OUD GWAS did not identify any genome-wide significant hits (Supplementary Table 3). Many of these loci were mapped to genes involved in metabolism or receptor activity of specific substances. For example, residual AUD loci were mapped to genes in the alcohol dehydrogenase family (*ADH1B, ADH4, ADH5, ADH6*), which is involved in metabolism of alcohol and other related substances, and loci for residual PTU were mapped to genes in the nicotinic acetylcholine receptor family of proteins (*CHRNA3, CHRNA5, CHRNA6, CHRNB3, CHRNB4)*.

We also performed SNP-level tests of heterogeneity (Q_{SNP}) to investigate the extent to which SNPs identified with the factors had consistent, pleiotropic effects on the constituent indicators of that factor. We identified 134 and 112 Q_{SNP} for the Externalizing and two-factor Behavioral Disinhibition and SUD model, respectively. Only one Q_{SNP} loci from the Externalizing model overlapped with any of the 708 genomic risk loci identified for that factor. This was rs4702, a noncoding transcript variant located in the gene *FURIN* that has previously associated with OUD[33]. Similarly, only three significant loci from the Behavioral Disinhibition and SUD model were Q_{SNP} loci. These loci were rs4702 and rs1229984, a missense variant in *ADH1B* associated with alcohol phenotypes[34,35], and rs13135092, an intronic variant located in *SLC39A8* also previously associated with AUD[36]. Of the 38 loci identified across all four residual SUDs, 14 overlapped with Q_{SNP} loci from the Externalizing and/or SUD factor, highlighting the pattern that Q_{SNP} loci represent substance-specific genetic effects.

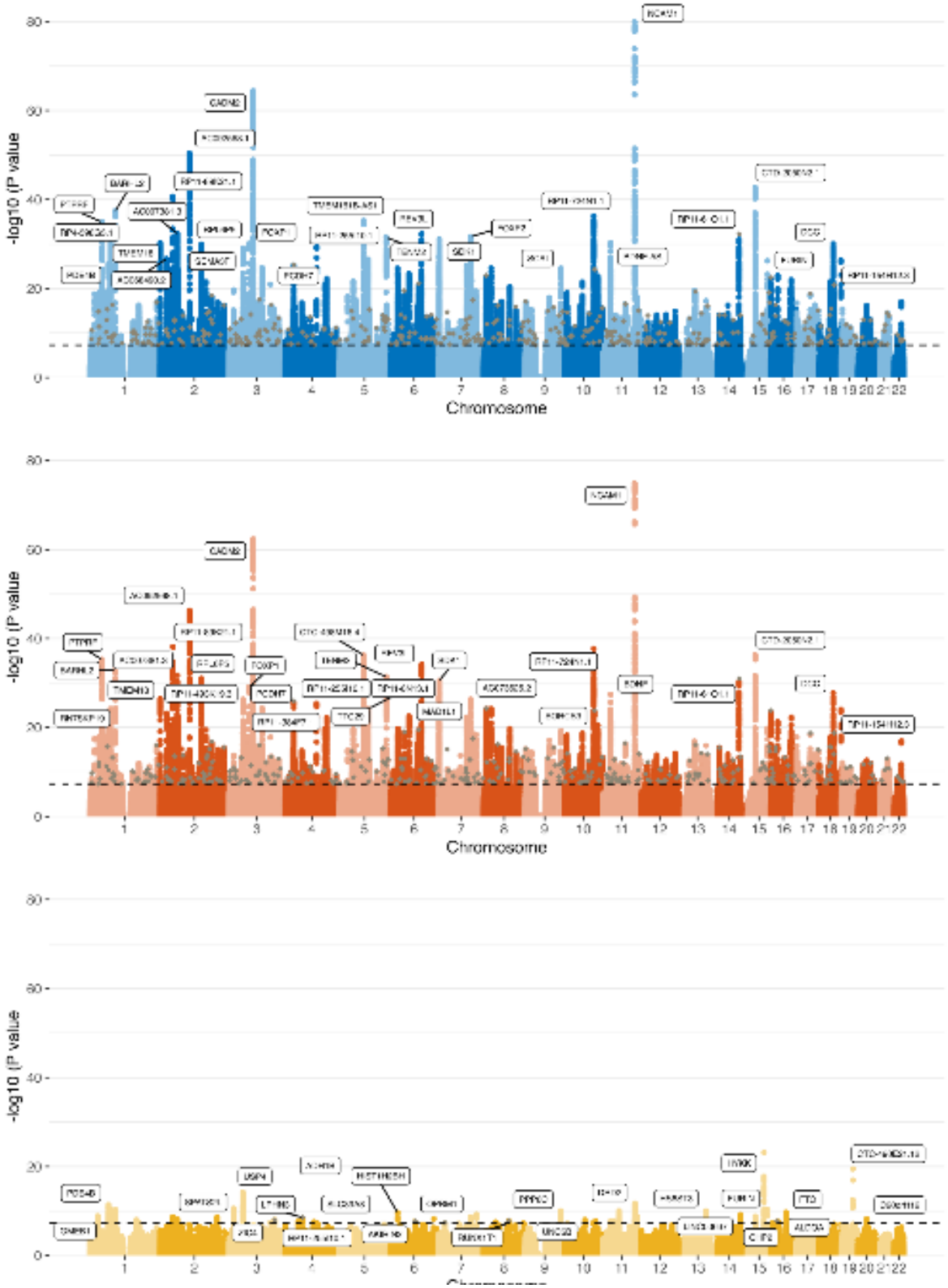

Figure 2. Manhattan plots of (top to bottom) Externalizing, Behavioral Disinhibition, and SUD. Brown points represent novel SUD loci (loci not previously associated with a substance use trait). Top loci are mapped to the nearest gene using ANNOVAR[37] annotation.

Biological annotation

Identifying genes for latent genomic factors. MAGMA identified 1325, 1125, and 69 genes associated with Externalizing, Behavioral Disinhibition, and SUD, respectively. Metaxcan identified 3463, 3078, and 207 genes. SMR and HEIDI tests prioritized 977, 868, and 20 genes for Externalizing, Behavioral Disinhibition, and SUD, respectively. The intersection of genes identified by these two methods resulted in 142, 127, and 5 high confidence genes significantly associated with Externalizing, Behavioral Disinhibition, and SUD, respectively (Figure 3a; Supplementary Table 4). We found 38 unique genes for Externalizing, 23 for Behavioral Disinhibition, and 3 for SUDs (Figure 3b). Of the 38 genes unique Externalizing genes, 34 (89%) have been previously associated with a substance use phenotype (e.g., initiation, consumption, or use disorder),[23] including *CDH12*, which has been associated with nicotine dependence, and *KLHL29*, which has been associated with cannabis, nicotine, and alcohol use. Of the three unique SUD genes, *PPP6C* has been associated with CUD, AUD, and OUD. In addition, we identified 44, 26, and 3 genes that were not identified in their respective original GWAS.

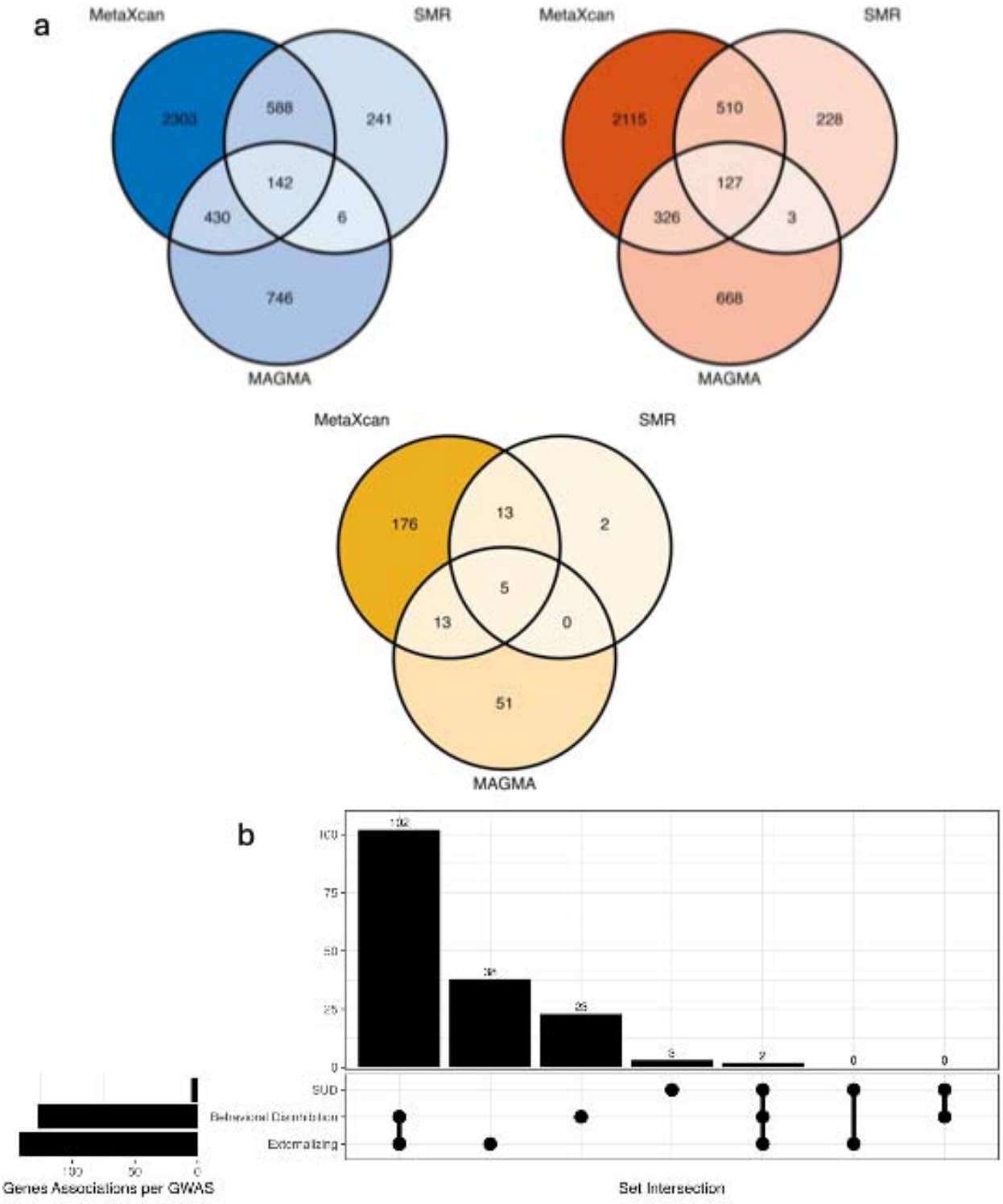

Figure 3. (a) Venn diagrams showing the overlap of genes identified by the MetaXcan, SMR, and MAGMA analyses for Externalizing (blue), Behavioral Disinhibition (orange), and SUD (yellow). (b) upset plot showing intersecting sets of high confidence genes (those identified by all three gene-based methods) across Externalizing, Behavioral Disinhibition, and SUD.

Identifying genes for residual SUDs. We also used MAGMA to identify genes associated with the four residual SUDs. We found 34, 38, and 1 genes associated with residual AUD, PTU, and OUD, respectively (Supplementary Table 5). We did not identify any significant genes associated with residual CUD. The genes associated with residual SUDs were unique to that phenotype, meaning that there were no overlapping genes identified across residual SUDs. We saw substance-specificity in the genes associated with the residual SUD phenotypes. The majority of genes identified for the residual SUDs were previously associated with a substance use phenotype and many were associated *only* with a phenotype for that specific substance. For example, of the 34 genes associated with residual AUD, 22 were previously associated with any substance use phenotype and 19 of those were previously associated *only* with an alcohol phenotype.

Tissue expression. We found little evidence of differential associations across the three factors with gene expression in brain, liver, and lung tissues (Supplementary Figure 1). The direction of effects was the same across factors, but the magnitude of the associations for SUDs was significantly smaller than those for Behavioral Disinhibition and Externalizing, highlighting the differences in power, but not evidence of differences in magnitude of association.

Polygenic scores

Factor PGS. Figure 4a and 4b compare the variance accounted for by each of the factor PGS in relevant substance use outcomes in COGA and All of Us (full results in Supplementary Tables 8 and 9). EXT_{PGS} explained the most variance in all phenotypes except alcohol initiation in COGA and AUD diagnosis in All of Us, although the confidence intervals for the R^2 estimates overlapped in many instances. BD_{PGS} typically predicted the second most variance, except in the case of AUD symptoms and FTND scores in COGA, and AUD in All of Us. In COGA, EXT_{PGS} explained between 1.6% (CUD) and 3.3% (AUD) of the variance and SUD_{PGS} explained .39% (OUD) and 2.9% (FTND symptoms). In All of Us, EXT_{PGS} explained between 3.1 (AUD) and 7.1% (PTU) of the variance in SUD diagnosis and SUD_{PGS} accounted for 3.1% (AUD) and 4.4% (DUD).

Residual SUD PGS. We next compared the variance explained by the residual SUD PGS in substance initiation and use disorders in COGA (Figure 4c). Only the $resPTU_{PGS}$ accounted for a significant portion of the variance in any variable, explaining 3.2% of the variance in problematic tobacco use. Nevertheless, the R^2 estimates for each use disorder showed substance specific prediction in use disorders, such that the residual PGS predicted the most variance in the use disorder phenotype corresponding to that substance (e.g., $resAUD_{PGS}$ predicted the most variance in AUD symptoms etc).

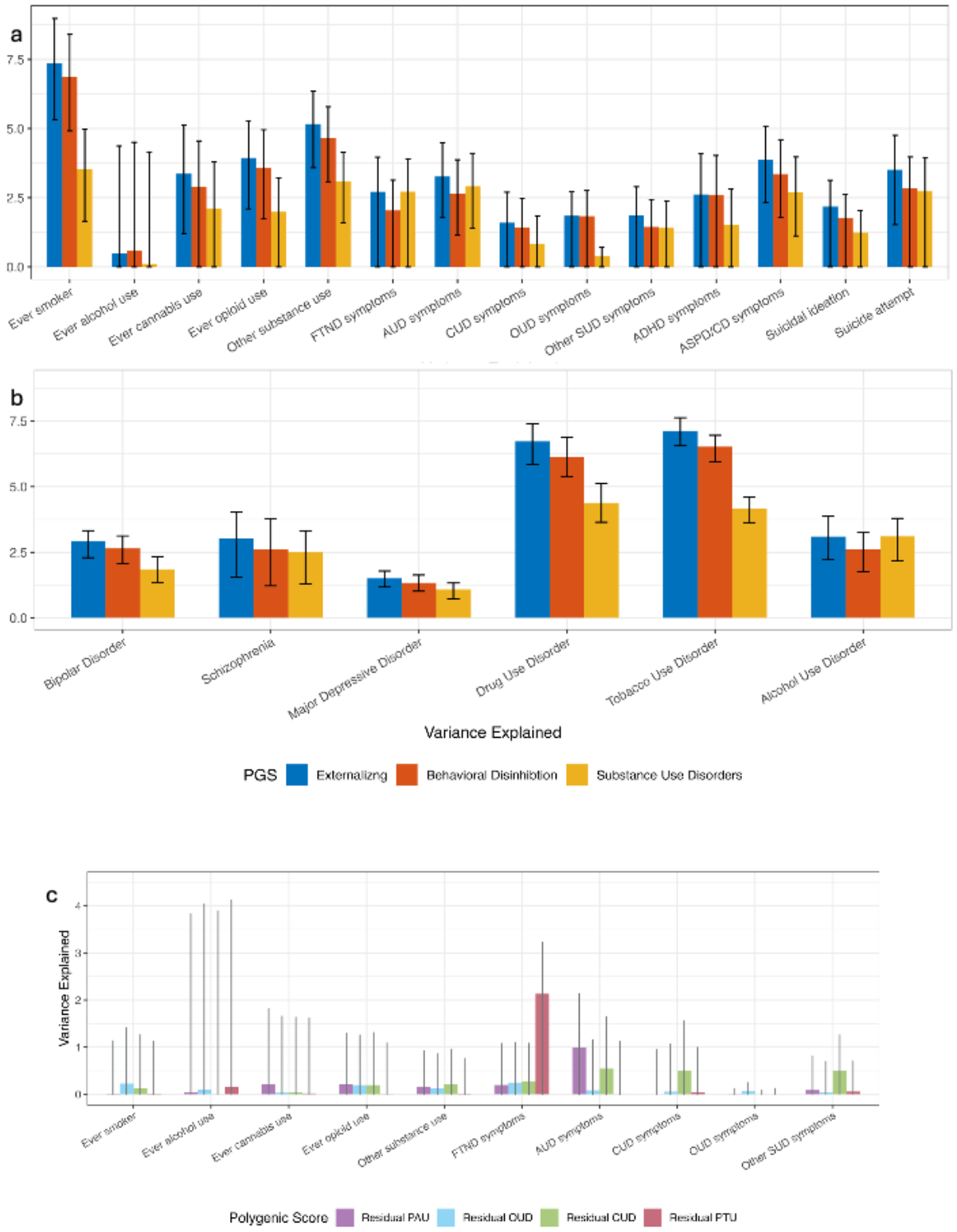

Figure 4. Variance explained by the factor PGS in (a) COGA and (b) All of Us. Panel (c) shows the variance explained by residual SUD PGS in COGA. We report R^2 estimates and 95% confidence intervals.

Discussion

Despite evidence that SUDs share phenotypic[6] and genetic[3,8,12] variance with each other and with other traits related to behavioral disinhibition, gene identification efforts typically study individual SUDs in isolation. This misses potentially important genetic variance untagged and may in part explain the slow progress in gene discovery for SUDs relative to other complex behavioral phenotypes[38]. Here, we drew from the twin literature on the nature of genetic influences on SUDs[1,2,8] and capitalized on recent advances in multivariate statistical genetic methods[11], to model the shared genetic architecture of SUDs and other externalizing phenotypes with the goal of improving gene discovery for SUDs.

Our results compellingly demonstrate that modeling genetic covariance of SUDs alongside related externalizing traits improves gene discovery for SUDs. In addition, we demonstrate that use of the externalizing factor as a gene-identification target did not reduce our ability to detect SUD-specific effects. Furthermore, our results indicate that use of a broad externalizing factor does not result in loss of specificity for SUD related signal. We base this argument on 1) the comparable number of Q_{SNP} loci between the two models, suggesting similar levels of heterogeneity; 2) the identification of substance use related genes in the Externalizing GWAS that were not identified in the GWAS for other factors; 3) lack of evidence for differential tissue expression or PGS associations across the three factors; and, in fact, 4) evidence that the Externalizing factor was best powered to capture SNP associations and that PGS derived from the externalizing factor account for variance in substance related outcomes. Finally, we demonstrate that substance-specific genetic associations (e.g., *ADH1B, CHRNA5*) are effectively captured by the residual SUD GWAS, indicating that the most insight into genetic influences on SUDs can be obtained by examining both the shared genetic liability to externalizing and the residual genetic variance in each SUD.

The finding that novel insights into the genetic influences on SUDs are best gained by studying broad genetic liability to externalizing in conjunction with residual SUD genetic variance is consistent with previous structure of psychopathology and twin literature. The importance of broad spectra *and* individual signs and symptoms is emphasized in recent structural models of psychopathology (e.g., HiTOP)[10,39]. Twin studies further support this, as externalizing factors capture a large proportion of etiological variance in their indicators, including SUDs, but individual disorders often retain a small, but statically significant, proportion of specific etiological variance[8].

Limitations & Future Directions

These findings should be interpreted in light of a few limitations. First and foremost, our analyses include only data from participants of European ancestry, thereby limiting the generalizability of our results. This was a practical choice as we relied on previously published GWAS and sufficiently powered GWAS of non-European ancestry samples are not yet available for these outcomes. Our group is working to expand to more diverse genomic populations (https://osf.io/7pfgj/). Second, our analyses are limited to disorder-level phenotypes and may mask symptom specific heterogeneity. Recent phenotypic studies exploring the comorbidity between SUDs and among SUDs and other forms of psychopathology suggest symptom-specific effects, [40-43] which we are unable to model Finally, although SUDs manifest the strongest relationships with other externalizing phenotypes, there is evidence of an internalizing pathway to problematic substance use[44,45] which is not modeled in the current study.

Conclusions

Modeling a common genetic liability to externalizing resulted in a greater number of genomic risk loci, identified novel genes, and produced a polygenic score that accounted for more variance in substance use outcomes. Our findings suggest that capitalizing on the shared genetic architecture among externalizing phenotypes can advance gene discovery for SUDs.

References

1. Kendler KS, Jacobson KC, Prescott CA, Neale MC. Specificity of genetic and environmental risk factors for use and abuse/dependence of cannabis, cocaine, hallucinogens, sedatives, stimulants, and opiates in male twins. *Am J Psychiatry*. Apr 2003;160(4):687-95. doi:10.1176/appi.ajp.160.4.687
2. Kendler KS, Myers J, Prescott CA. Specificity of genetic and environmental risk factors for symptoms of cannabis, cocaine, alcohol, caffeine, and nicotine dependence. *Archives of general psychiatry*. 2007;64(11):1313-1320.
3. Hatoum AS, Colbert SMC, Johnson EC, et al. Multivariate genome-wide association meta-analysis of over 1 million subjects identifies loci underlying multiple substance use disorders. *Nature Mental Health*. 2023;1(3):210-223. doi:10.1038/s44220-023-00034-y
4. Hatoum AS, Johnson EC, Colbert SMC, et al. The addiction risk factor: A unitary genetic vulnerability characterizes substance use disorders and their associations with common correlates. *Neuropsychopharmacol*. 2021/11/08 2021;doi:10.1038/s41386-021-01209-w
5. Miller AP, Bogdan R, Agrawal A, Hatoum AS. Generalized genetic liability to substance use disorders. *The Journal of Clinical Investigation*. 06/05/ 2024;134(11)doi:10.1172/JCI172881
6. Krueger, Markon, Patrick, Benning, Kramer. Linking Antisocial Behavior, Substance Use, and Personality: An Integrative Quantitative Model of the Adult Externalizing Spectrum. *J Abnorm Psychol*. 2007;116(4):645-666. doi:10.1037/0021-843X.116.4.645
7. Krueger RF, Hobbs KA, Conway CC, et al. Validity and utility of Hierarchical Taxonomy of Psychopathology (HiTOP): II. Externalizing superspectrum. https://doi.org/10.1002/wps.20844. *World Psychiatry*. 2021/06/01 2021;20(2):171-193. doi:https://doi.org/10.1002/wps.20844
8. Krueger, Hicks, Patrick, Carlson, Iacono, McGue. Etiological Connections among Substance Dependence, Antisocial Behavior, and Personality: Modeling the Externalizing Spectrum. *J Abnorm Psychol*. 2002;111(3):411-424. doi:10.1037/0021-843X.111.3.411
9. Kendler KS, Lönn SL, Maes HH, Lichtenstein P, Sundquist J, Sundquist K. A Swedish Population-Based Multivariate Twin Study of Externalizing Disorders. *Behavior Genetics*. 2016/03/01 2016;46(2):183-192. doi:10.1007/s10519-015-9741-7
10. Kotov R, Krueger RF, Watson D, et al. The Hierarchical Taxonomy of Psychopathology (HiTOP): A dimensional alternative to traditional nosologies. *J Abnorm Psychol*. May 2017;126(4):454-477. doi:10.1037/abn0000258
11. Grotzinger AD, Rhemtulla M, de Vlaming R, et al. Genomic structural equation modelling provides insights into the multivariate genetic architecture of complex traits. *Nat Hum Behav*. May 2019;3(5):513-525. doi:10.1038/s41562-019-0566-x
12. Poore HE, Hatoum A, Mallard TT, et al. A Multivariate Approach to Understanding the Genetic Overlap between Externalizing Phenotypes and Substance Use Disorders. *Addiction Biology*. 2023;28(9):e13319. doi:10.1111/adb.13319
13. Little J, Higgins JP, Ioannidis JP, et al. STrengthening the REporting of Genetic Association Studies (STREGA)--an extension of the STROBE statement. *Genet Epidemiol*. Nov 2009;33(7):581-98. doi:10.1002/gepi.20410
14. Karlsson Linnér R, Mallard TT, Barr PB, et al. Multivariate analysis of 1.5 million people identifies genetic associations with traits related to self-regulation and addiction. *Nat Neurosci*. 2021/10/01 2021;24(10):1367-1376. doi:10.1038/s41593-021-00908-3
15. Zhou H, Kember RL, Deak JD, et al. Multi-ancestry study of the genetics of problematic alcohol use in over 1 million individuals. *Nature Medicine*. 2023/12/01 2023;29(12):3184-3192. doi:10.1038/s41591-023-02653-5
16. Deak JD, Zhou H, Galimberti M, et al. Genome-wide association study in individuals of European and African ancestry and multi-trait analysis of opioid use disorder identifies 19 independent genome-wide significant risk loci. *Molecular Psychiatry*. 2022/10/01 2022;27(10):3970-3979. doi:10.1038/s41380-022-01709-1

17. Johnson EC, Demontis D, Thorgeirsson TE, et al. A large-scale genome-wide association study meta-analysis of cannabis use disorder. *Lancet Psychiat*. Dec 2020;7(12):1032-1045. doi:10.1016/s2215-0366(20)30339-4
18. Demontis D, Walters RK, Martin J, et al. Discovery of the first genome-wide significant risk loci for attention deficit/hyperactivity disorder. *Nat Genet*. Jan 2019;51(1):63-75. doi:10.1038/s41588-018-0269-7
19. Linnér R, Karlsson, Biroli P, Kong E, et al. Genome-wide association analyses of risk tolerance and risky behaviors in over 1 million individuals identify hundreds of loci and shared genetic influences. *Nature Genetics*. 2019/02/01 2019;51(2):245-257. doi:10.1038/s41588-018-0309-3
20. Liu M, Jiang Y, Wedow R, 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*. 2019/02/01 2019;51(2):237-244. doi:10.1038/s41588-018-0307-5
21. Pasman JA, Verweij KJH, Gerring Z, et al. GWAS of lifetime cannabis use reveals new risk loci, genetic overlap with psychiatric traits, and a causal influence of schizophrenia. *Nat Neurosci*. Sep 2018;21(9):1161-1170. doi:10.1038/s41593-018-0206-1
22. Watanabe K, Taskesen E, van Bochoven A, Posthuma D. Functional mapping and annotation of genetic associations with FUMA. *Nature Communications*. 2017/11/28 2017;8(1):1826. doi:10.1038/s41467-017-01261-5
23. Sollis E, Mosaku A, Abid A, et al. The NHGRI-EBI GWAS Catalog: knowledgebase and deposition resource. *Nucleic Acids Research*. 2023;51(D1):D977-D985. doi:10.1093/nar/gkac1010
24. de Leeuw CA, Mooij JM, Heskes T, Posthuma D. MAGMA: Generalized Gene-Set Analysis of GWAS Data. *PLOS Computational Biology*. 2015;11(4):e1004219. doi:10.1371/journal.pcbi.1004219
25. Barbeira AN, Dickinson SP, Bonazzola R, et al. Exploring the phenotypic consequences of tissue specific gene expression variation inferred from GWAS summary statistics. *Nature Communications*. 2018/05/08 2018;9(1):1825. doi:10.1038/s41467-018-03621-1
26. Consortium G. The GTEx Consortium atlas of genetic regulatory effects across human tissues. *Science*. Sep 11 2020;369(6509):1318-1330. doi:10.1126/science.aaz1776
27. Zhu Z, Zhang F, Hu H, et al. Integration of summary data from GWAS and eQTL studies predicts complex trait gene targets. *Nature genetics*. 2016;48(5):481-487.
28. Dick DM, Balcke E, McCutcheon V, et al. The collaborative study on the genetics of alcoholism: Sample and clinical data. *Genes, Brain and Behavior*. 2023;22(5):e12860.
29. Investigators AoURPG. Genomic data in the All of Us Research Program. *Nature*. Mar 2024;627(8003):340-346. doi:10.1038/s41586-023-06957-x
30. Investigators TAoURP. The “All of Us” Research Program. *New England Journal of Medicine*. 2019/08/15 2019;381(7):668-676. doi:10.1056/NEJMsr1809937
31. Ge T, Chen CY, Ni Y, Feng YA, Smoller JW. Polygenic prediction via Bayesian regression and continuous shrinkage priors. *Nature Communications*. Apr 16 2019;10(1):1776. doi:10.1038/s41467-019-09718-5
32. Purcell S, Neale B, Todd-Brown K, et al. PLINK: a tool set for whole-genome association and population-based linkage analyses. *The American journal of human genetics*. 2007;81(3):559-575.
33. Kember RL, Vickers-Smith R, Xu H, et al. Cross-ancestry meta-analysis of opioid use disorder uncovers novel loci with predominant effects in brain regions associated with addiction. *Nat Neurosci*. Oct 2022;25(10):1279-1287. doi:10.1038/s41593-022-01160-z
34. Gelernter J, Sun N, Polimanti R, et al. Genome-wide Association Study of Maximum Habitual Alcohol Intake in >140,000 U.S. European and African American Veterans Yields Novel Risk Loci. *Biol Psychiatry*. Sep 1 2019;86(5):365-376. doi:10.1016/j.biopsych.2019.03.984

35. Kember RL, Vickers-Smith R, Zhou H, et al. Genetic Underpinnings of the Transition From Alcohol Consumption to Alcohol Use Disorder: Shared and Unique Genetic Architectures in a Cross-Ancestry Sample. *Am J Psychiatry*. Aug 1 2023;180(8):584-593. doi:10.1176/appi.ajp.21090892
36. Zhou H, Sealock JM, Sanchez-Roige S, 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*. Jul 2020;23(7):809-818. doi:10.1038/s41593-020-0643-5
37. Wang K, Li M, Hakonarson H. ANNOVAR: functional annotation of genetic variants from high-throughput sequencing data. *Nucleic Acids Res*. Sep 2010;38(16):e164. doi:10.1093/nar/gkq603
38. Deak JD, Johnson EC. Genetics of substance use disorders: a review. *Psychol Med*. 2021;51(13):2189-2200. doi:10.1017/S0033291721000969
39. Kotov R, Krueger RF, Watson D, et al. The Hierarchical Taxonomy of Psychopathology (HiTOP): A Quantitative Nosology Based on Consensus of Evidence. *Annual Review of Clinical Psychology*. 2021;17(Volume 17, 2021):83-108. doi:https://doi.org/10.1146/annurev-clinpsy-081219-093304
40. Forbes MK, Watts AL, Twose M, et al. A Hierarchical Model of the Symptom-Level Structure of Psychopathology in Youth. *Clinical Psychological Science*. 2024:21677026241257852. doi:10.1177/21677026241257852
41. Savage JE, Barr PB, Phung T, et al. Genetic Heterogeneity Across Dimensions of Alcohol Use Behaviors. *American Journal of Psychiatry*. 2024:appi.ajp.20231055. doi:10.1176/appi.ajp.20231055
42. Mallard TT, Savage JE, Johnson EC, et al. Item-Level Genome-Wide Association Study of the Alcohol Use Disorders Identification Test in Three Population-Based Cohorts. *Am J Psychiatry*. Jan 2022;179(1):58-70. doi:10.1176/appi.ajp.2020.20091390
43. Watts AL, Sher KJ, Heath AC, Steinley D, Brusco M. “General Addiction Liability” Revisited. *Clinical Psychological Science*. 2024:21677026241245070. doi:10.1177/21677026241245070
44. King SM, Iacono WG, McGue M. Childhood externalizing and internalizing psychopathology in the prediction of early substance use. *Addiction*. 2004;99(12):1548-1559.
45. Hussong AM, Jones DJ, Stein GL, Baucom DH, Boeding S. An internalizing pathway to alcohol use and disorder. *Psychol Addict Behav*. Sep 2011;25(3):390-404. doi:10.1037/a0024519

Acknowledgements

The authors thank **The Externalizing Consortium.** Principal Investigators: Danielle M. Dick, Philipp Koellinger, K. Paige Harden, Abraham A. Palmer. Lead Analysts: Richard Karlsson Linnér, Travis T. Mallard, Peter B. Barr, Sandra Sanchez-Roige. Significant Contributors: Irwin D. Waldman. The Externalizing Consortium has been supported by the National Institute on Alcohol Abuse and Alcoholism (R01AA015416 – administrative supplement to DMD), and the National Institute on Drug Abuse (R01DA050721 to DMD). Additional funding for investigator effort has been provided by K02AA018755, U10AA008401, P50AA022537 to DMD, as well as a European Research Council Consolidator Grant (647648 EdGe to Koellinger). The content is solely the responsibility of the authors and does not necessarily represent the official views of the above funding bodies. The Externalizing Consortium would like to thank the following groups for making the research possible: **23andMe, Add Health, Vanderbilt University Medical Center's BioVU, Collaborative Study on the Genetics of Alcoholism (COGA), the Psychiatric Genomics Consortium's (PGC) Substance Use Disorders working group, UK10K Consortium, UK Biobank,** and **Philadelphia Neurodevelopmental Cohort.** All code necessary to replicate this study is available upon request.

The authors thank **Million Veteran Program (MVP)** staff, researchers, and volunteers, who have contributed to MVP, and especially participants who previously served their country in the military and now generously agreed to enroll in the study.

We gratefully acknowledge ***All of Us*** participants for their contributions, without whom this research would not have been possible. We also thank the National Institutes of Health's *All of Us* Research Program for making available the participant data examined in this study.

Finally, we thank **The Collaborative Study on the Genetics of Alcoholism (COGA)**, Principal Investigators B. Porjesz, V. Hesselbrock, A. Agrawal; Scientific Director, A. Agrawal; Translational Director, D. Dick, includes ten different centers: University of Connecticut (V. Hesselbrock); Indiana University (H.J. Edenberg, T. Foroud, Y. Liu, M.H. Plawecki); University of Iowa Carver College of Medicine (S. Kuperman, J. Kramer); SUNY Downstate Health Sciences University (B. Porjesz, J. Meyers, C. Kamarajan, A. Pandey); Washington University in St. Louis (L. Bierut, A. Agrawal, S. Hartz); University of California at San Diego (M. Schuckit); Rutgers

University (J. Tischfield, D. Dick, R. Hart, J. Salvatore); The Children's Hospital of Philadelphia, University of Pennsylvania (L. Almasy); Icahn School of Medicine at Mount Sinai (A. Goate, P. Slesinger); and Howard University (D. Scott). Other COGA collaborators include: C. Holzhauer, M. Hesselbrock (University of Connecticut); J. Nurnberger Jr., L. Wetherill, X., Xuei, D. Lai, S. O'Connor, (Indiana University); G. Chan (University of Iowa; University of Connecticut); D.B. Chorlian, J. Zhang, P. Barr, S. Kinreich, G. Pandey, Z. Neale (SUNY Downstate); N. Mullins (Icahn School of Medicine at Mount Sinai); A. Anokhin, K. Bucholz, F. Dong, A. Hatoum, E. Johnson, V. McCutcheon, J. Rice, S. Saccone (Washington University); F. Aliev, Z. Pang, S. Kuo, S. Brislin (Rutgers University); A. Merikangas (The Children's Hospital of Philadelphia and University of Pennsylvania); H. Chin and A. Parsian are the NIAAA Staff Collaborators. We continue to be inspired by our memories of Henri Begleiter and Theodore Reich, founding PI and Co-PI of COGA, and also owe a debt of gratitude to other past organizers of COGA, including Ting- Kai Li, P. Michael Conneally, Raymond Crowe, and Wendy Reich, for their critical contributions. This national collaborative study is supported by NIH Grant U10AA008401 from the National Institute on Alcohol Abuse and Alcoholism (NIAAA) and the National Institute on Drug Abuse (NIDA).

COGA Collaborators: Bernice Porjesz, Victor Hesselbrock, Tatiana M. Foroud, Arpana Agrawal, Howard J. Edenberg, John I. Nurnberger Jr, Yunlong Liu, Samuel Kuperman, John Kramer, Jacquelyn L. Meyer, Chella Kamarajan, Ashwini K. Pandey, Laura Bierut, John Rice, Kathleen K. Bucholz, Marc A. Schuckit, Jay Tischfield, Andrew Brooks, Ronald P. Hart, Laura Almasy, Danielle M. Dick, Jessica E. Salvatore, Allison Goate, Manav Kapoor, Paul Slesinger, Denise M. Scott, Lance Bauer, Leah Wetherill, Xiaoling Xuei, Dongbing Lai, Sean J. O'Connor, Martin H. Plawecki, Spencer Lourens, Laura Acion, Grace Chan, David B. Chorlian, Jian Zhang, Sivan Kinreich, Gayathri Pandey, Michael J. Chao, Andrey P. Anokhin, Vivia V. McCutcheon, Scott Saccone, Fazil Aliev, Peter B. Barr, Hemin Chin & Abbas Parsian