RESEARCH

Partitioning the Genomic Components of Behavioral Disinhibition and Substance Use (Disorder) Using Genomic Structural Equation Modeling

Tanya B. Horwitz[1] · Katerina Zorina-Lichtenwalter[1] · Daniel E. Gustavson[1] · Andrew D. Grotzinger[1,2] ·
Michael C. Stallings[1,2]

Abstract

Externalizing behaviors encompass manifestations of risk-taking, self-regulation, aggression, sensation-/reward-seeking, and impulsivity. Externalizing research often includes substance use (SUB), substance use disorder (SUD), and other (non-SUB/SUD) "behavioral disinhibition" (BD) traits. Genome-wide and twin research have pointed to overlapping genetic architecture within and across SUB, SUD, and BD. We created single-factor measurement models—each describing SUB, SUD, or BD traits—based on mutually exclusive sets of European ancestry genome-wide association study (GWAS) statistics exploring externalizing variables. We then assessed the partitioning of genetic covariance among the three facets using correlated factors models and Cholesky decomposition. Even when the residuals for indicators relating to the same substance were correlated across the SUB and SUD factors, the two factors yielded a large correlation ($r_g = 0.803$). BD correlated strongly with the SUD ($r_g = 0.774$) and SUB ($r_g = 0.778$) factors. In our initial decompositions, 33% of total BD variance remained after partialing out SUD and SUB. The majority of covariance between BD and SUB and between BD and SUD was shared across all factors, and, within these models, only a small fraction of the total variation in BD operated via an independent pathway with SUD or SUB outside of the other factor. When only nicotine/tobacco, cannabis, and alcohol were included for the SUB/SUD factors, their correlation increased to $r_g = 0.861$; in corresponding decompositions, BD-specific variance decreased to 27%. Further research can better elucidate the properties of BD-specific variation by exploring its genetic/molecular correlates.

Keywords Substance use · Externalizing · Behavioral genetics · Genomic SEM

✉ Tanya B. Horwitz
  tanya.horwitz@colorado.edu

  Katerina Zorina-Lichtenwalter
  katerina.zorina@colorado.edu

  Daniel E. Gustavson
  daniel.gustavson@colorado.edu

  Andrew D. Grotzinger
  andrew.grotzinger@colorado.edu

  Michael C. Stallings
  michael.stallings@colorado.edu

1 Institute for Behavioral Genetics, University of Colorado Boulder, 1480 30th St., Boulder, CO 80303, USA

2 Psychology and Neuroscience, University of Colorado Boulder, Meunzinger D244, 345 UCB, Boulder, CO 80309, USA

Introduction

Background

Behavioral researchers have traditionally placed phenotypes that relate to risk-taking, lack of self-regulation, aggression, and/or impulsivity on the "externalizing" psychopathology spectrum. Related behaviors that exemplify sensation- or reward-seeking without meeting diagnostic psychopathology criteria have likewise been grouped under the externalizing framework (Karlsson Linnér et al. 2021). Studies that probe externalizing often consider substance use disorders (SUDs), which are characterized by obsessive/compulsive substance use-seeking and -intake patterns, negative emotional responses to cessation of the substance (Koob and Volkow 2016), and other related symptoms (Krueger et al. 2005; Kendler and Myers 2014). However, substance consumption

behaviors not associated with these life-complicating factors (i.e., substance initiation and intake quantity/frequency; hereafter, SUB), have also been considered in externalizing research (Karlsson Linnér et al. 2021). Externalizing behavior additionally encompasses a combination of several pathological and non-pathological traits—i.e., risk tolerance, number of sexual partners, age at first sex (Karlsson Linnér et al. 2021), antisocial personality disorder, and conduct disorder (Krueger et al. 2005; Kendler and Myers 2014)—that do not directly measure substance use or substance use disorders, sometimes referred to as "behavioral disinhibition" (BD) (Poore et al. 2023).

Genome-wide association studies (GWAS) and twin studies have shown traits and latent constructs related to SUB, SUD, and BD to be genetically correlated, suggesting that these constructs have overlapping genetic architectures (Kendler and Myers 2014; Karlsson Linnér et al. 2021; Poore et al. 2023). There is evidence for "general" genetic liability underlying SUD and for additional genetic liability specific to certain individual SUDs (Hatoum et al. 2022). Though SUDs examined in large GWAS are, unsurprisingly, usually genetically correlated with use measures for the corresponding substance, the degree of genetic overlap varies across substances. For example, while the point estimate of the genetic correlation between the Fagerström Test for Nicotine Dependence (FTND) and number of cigarettes smoked per day is around unity ($r_g = 0.97$, s.e. $= 0.12$) (Hatoum et al. 2022), the correlation between cannabis use and cannabis use disorder is lower, at $r_g = 0.50$ (s.e. $= 0.05$) (Johnson et al. 2020). Meanwhile, the literature on genome-wide correlations between drinking quantity/frequency and problem drinking behavior(s) paints a more complex picture, having yielded point estimates that vary enormously depending on the alcohol behavior metrics, cohorts, sample size, statistical methodology, genetic ancestry group, and degree of trait heterogeneity being examined (Sanchez-Roige et al. 2019b; Mallard et al. 2022; White and Bierut 2023; Kember et al. 2023). In addition to twin and family studies that have explored the genetics of externalizing based on phenotypic resemblance among relatives, a 2021 multivariate analysis estimated a general genomic externalizing factor using summary association data from seven high-powered GWAS (Karlsson Linnér et al. 2021), a design that evaluates the relationship between measured genotypes and a phenotype of interest. Of these seven GWAS, one examined alcohol use disorder, two examined non-pathological substance use traits (cannabis initiation and smoking initiation), and four examined traits that could be classified as forms of BD (age at first sex, number of sexual partners, attention-deficit/hyperactivity disorder, and risk tolerance); a more recent externalizing factor(s) publication included additional SUD constructs (Poore et al. 2023). The 2021 study identified 579 independent genomic loci associated with general externalizing. Factor-level analysis of general genomic SUD liability (Hatoum et al. 2023) yielded a substantially narrower set of associated

genome-wide significant loci than did the seven-indicator externalizing factor (though the externalizing factor was likely better-powered).

Because externalizing encompasses a multidimensional umbrella of behaviors, a sophisticated understanding of its genetic architecture may require a closer examination of its substructures. Further research is required to determine the extent to which genomic variation associated with facets of externalizing overlaps and the extent to which it operates through independent pathways. The current paper will address this topic by filling two gaps in the literature. First, there is strong theoretical and empirical motivation behind exploring a delineation between subthreshold and above-threshold manifestations of pathologically-implicated constructs. One recent genetic analysis found evidence that utilizing separate factors for internalizing psychopathology and internalizing-related traits improved model fit (Gustavson et al. 2024). There has also been some investigation into the propensity of genetic loci to predict differences between pathological use, non-pathological use, and/or abstinence for some specific/single substances (Sanchez-Roige et al. 2019b; Polimanti et al. 2020; Mallard et al. 2022; Kember et al. 2023). However, few publications in the genome-wide space have explored how common liability underlying normative use/non-use across substances may differ from common liability underlying pathological substance use across substances. Second, little is known about potential genome-wide pathways relating to externalizing domains like risk-taking, excitement-seeking, attention, and reward processing outside the purview of drug use.

Broad analysis plan

We will use genome-wide association data and Genomic Structural Equation Modeling software (Grotzinger et al. 2019) to 1) estimate genetic correlations between externalizing traits, 2) construct a genomically-informed latent factor underlying SUB, 3) measure the covariance between SUB and validated BD and SUD factors, and 4) decompose the shared and unshared genetic variance across the three factors via hierarchical trivariate Cholesky models; here, we will particularly focus on how SUD and SUB may differentially relate to BD. The findings from these analyses will shed light on whether phenotypic subcategories of externalizing behavior constitute separable genomic entities.

Material and methods

Description of genomic structural equation modeling

In accordance with our preregistered plan (https://osf.io/fjy5h), we used the software Genomic Structural Equation Modeling (Genomic SEM) (Grotzinger et al. 2019) to model the genomic factor-level structure of general and domain-specific externalizing. Genomic SEM leverages LD Score regression (LDSC) (Bulik-Sullivan et al. 2015b), lavaan (Rosseel et al. 2023), and GWAS summary statistics to fit confirmatory models based on genetic (rather than phenotypic) correlations.

In the first stage of Genomic SEM, LDSC creates an empirical genetic covariance matrix (the S matrix) and an associated sampling covariance matrix (the V_S matrix). The unstandardized S matrix contains single nucleotide polymorphism (SNP) heritabilities—the proportion of phenotypic variance that can be statistically explained by SNPs—of each phenotype on the diagonals and genetic covariances between each phenotype on the off-diagonals. Meanwhile, the V_S matrix is a sampling covariance matrix used to account for study sample overlap among the source GWAS summary statistics. In the second stage, Genomic SEM fits a user-specified structural equation model, estimating factor loadings and correlation parameters that minimize discrepancy between the model-implied genetic covariance matrix and the empirical covariance matrix.

Measures

We began by defining three lower-order general factors (the measurement models): Behavioral Disinhibition (BD), (non-pathological) substance use (SUB), and Substance Use Disorder (SUD)—each representing a dimension of externalizing and each comprising mutually exclusive groups of reflective indicators—with the intention of investigating how SUB and SUD may relate to BD differently. Per our pre-registration, we only included GWAS for traits with sample sizes of at least 10,000 and SNP heritability (h^2_{SNP}) Z-statistics > 4 (Bulik-Sullivan et al. 2015a). For traits that were meta-analyzed but whose meta-analyzed association statistics were not available, we ran inverse-variance weighted meta-analysis ourselves in METAL (Willer et al. 2010). All samples were comprised of individuals of European ancestry, as different genetic ancestry groups may have different patterns of linkage disequilibrium and allele frequencies (Peterson et al. 2019) and analyses looking at non-European ancestry groups are under-powered for some of the indicators of interest. Table 1 displays a summary of the four GWAS used for the BD factor, the four GWAS used for the SUD

factor, and the seven pre-registered GWAS initially explored for the SUB factor.

BD factor

The lower-order BD factor utilized the four (non-substance-associated) BD indicators established in Poore et al. (2023):

- **Attention-deficit/hyperactivity disorder (ADHD) (Effective $N = 141,035$; $\frac{h^2SNP}{s.e.(h^2SNP)} = 21.12$)**, as defined by the International Classification of Diseases (ICD)-10 (with some samples taken from an inpatient or outpatient psychiatric setting), or prescription for an ADHD medication, depending on the cohort; ADHD for the Psychiatric Genomics Consortium was diagnosed using multiple different measures (case/control) (Demontis et al. 2023).
- **Number of lifetime sexual partners (NSEX) ($N = 370,711$; $\frac{h^2SNP}{s.e.(h^2SNP)} = 28.77$)** (Karlsson Linnér et al. 2019).
- **Age at first sexual intercourse (FSEX)** (reverse-coded) **($N = 397,338$; $\frac{h^2SNP}{s.e.(h^2SNP)} = 31.96$)** (Mills et al. 2021).
- **General risk tolerance (RISK) (Effective $N = 268,876$; $\frac{h^2SNP}{s.e.(h^2SNP)} = 22.17$)**, which was based on several items that varied depending on the cohort and that were similar to the question: "Would you describe yourself as someone who takes risks?" (Karlsson Linnér et al. 2019).

SUD factor

The lower-order SUD factor described the indicators in Hatoum et al. (2022) and Hatoum et al. (2023) (with the exception of cigarettes per day, which we instead explored as a potential SUB Factor indicator):

- **Problem tobacco use (PTU) ($N = 15,988$; $\frac{h^2SNP}{s.e.(h^2SNP)} = 4.22$)**, based on the Fagerström Test for Nicotine Dependence (FTND) in ever-smokers. Unlike in the Hatoum et al. study, we only used the FTND—which is used to determine degree of nicotine dependence (mild, moderate, or severe)—as a measure of PTU and did not integrate cigarettes per day into the indicator. We also had a smaller sample size for the FTND GWAS because of data sharing restrictions (Hancock et al. 2018).
- **Problem alcohol use (PAU) (Effective $N = 300,790$; $\frac{h^2SNP}{s.e.(h^2SNP)} = 18.47$)**, defined by alcohol dependence in the fourth edition of the Diagnostic and Statistical Manual of Mental Disorders (DSM-IV) or alcohol use disorder, as determined by the Alcohol Use Disorders

Table 1 Summary data and data sources for the four behavioral disinhibition traits, four substance use disorder traits, and seven pre-registered substance use traits initially considered for the substance use factor

Trait	N	N_{eff}	$h^2_{SNP}/s.e.(h^2_{SNP})$	Summary statistics source
ADHD	225,534	141,035.20	21.12	https://figshare.com/articles/dataset/adhd2022/22564390
NSEX	370,711	N/A	28.77	https://thessgac.com/papers/2/5
FSEX	397,338	N/A	31.96	http://ftp.ebi.ac.uk/pub/databases/gwas/summary_statistics/GCST90000001-GCST90001000/GCST90000047/
RISK	466,571	268,876	22.17	https://thessgac.com/papers/2/1
PTU	15,988	N/A	4.22	https://ftp-ncbi-nlm-nih-gov.colorado.idm.oclc.org/dbgap/studies/phs001532/analyses/
PAU	435,563	300,789.60	18.47	https://figshare.com/articles/dataset/sud2019-alcuse/14672193 https://figshare.com/articles/dataset/sud2018-alc/14672187 https://www.ncbi.nlm.nih.gov/projects/gap/cgi-bin/study.cgi?study_id=phs001672.v1.p1
CUD	357,806	47,952.51	10.71	https://figshare.com/articles/dataset/sud2020-cud/14842692
OUD	82,707	32,703.49	6.14	https://figshare.com/articles/dataset/sud2020-op/14672211
CI	184,765	144,698.95	15.96	Available Upon Request
SI	~ 1,300,000	1,363,137.10	29.93	https://conservancy.umn.edu/handle/11299/241912
DPW	~ 1,070,917	N/A	21.32	https://conservancy.umn.edu/handle/11299/241912
DRUG	22,572	N/A	4.69	N/A
CPD	~ 360,808	N/A	20.16	https://conservancy.umn.edu/handle/11299/241912
DS	398,853	376,693.97	21.22	N/A
FREQ	462,016	N/A	19.98	Available upon request

N = raw sample size used

Note that, in some cases, sample sizes in publicly available data are lower than total sample size reported in the corresponding publication(s) or in the table because of data sharing restrictions on some cohorts

N_{eff} = effective sample size (for dichotomous traits), as provided in the summary statistics or calculated with respect to prevalence and sample size across cohorts

h^2_{SNP} = the proportion of trait variation in the sample(s) accounted for by variation in single nucleotide polymorphisms (SNPs)

See the BD factor, SUD factor, and SUB factor Sections for further descriptions of phenotypes

ADHD Attention-deficit hyperactivity disorder, *NSEX* number of lifetime sexual partners, *FSEX* age at first sexual intercourse (reverse-coded), *RISK* general risk tolerance, *PTU* problem tobacco use, *PAU* problem alcohol use, *CUD* cannabis use disorder, *OUD* opioid use disorder, *CI* cannabis initiation, *SI* smoking initiation, *DPW* drinks per week, *DRUG* drug experimentation (the number of different classes of drugs an individual has used out of eleven), *CPD* cigarettes per day, *DS* drinking status (drinker vs. non-drinker), *FREQ* drinking frequency

Identification Test problem items (AUDIT-P), a continuous metric, or by the ICD-9 or -10, depending on the cohort (Zhou et al. 2020b).

- **Cannabis use disorder (CUD) (Effective $N = 47,953$; $\frac{h^2 SNP}{s.e.(h^2 SNP)} = 10.71$),** defined by cannabis abuse or dependence according to the DSM-III or DSM-IV, cannabis abuse or dependence according to the ICD-10, or cannabis use disorder according to the DSM-5, depending on the cohort (case vs. exposed *or* unexposed control) (Johnson et al. 2020).
- **Opioid use disorder (OUD) (Effective $N = 32,703$; $\frac{h^2 SNP}{s.e.(h^2 SNP)} = 6.14$),** defined by at least one inpatient or two outpatient ICD-9 or ICD-10 codes, or by opioid dependence, as defined by the DSM-IV, depending on the cohort (case vs. exposed control) (Zhou et al. 2020a).

SUB factor

We initially explored a total of seven pre-registered non-pathological substance use measures (presented in Table 1), three of which were not ultimately included (see the Supplementary Note and the Lower-order BD, SUD, and SUB factors Section). Thus, the lower-order SUB factor was based around four indicators:

- **Lifetime cannabis initiation (CI) (Effective $N = 144,699$; $\frac{h^2 SNP}{s.e.(h^2 SNP)} = 15.96$),** which was based on several items that varied depending on the cohort. For example: "Have you ever in your life used the following: Marijuana?"; "Have you taken CANNABIS (marijuana, grass, hash, ganja, blow, draw, skunk, weed, spliff, dope), even if it was a long time ago?" (Pasman et al. 2018).

- **Smoking initiation (SI) (Effective $N = 1,363,137$; $\frac{h^2 SNP}{s.e.(h^2 SNP)} = 29.93$)**, as defined by having ever or never been a regular smoker, or having ever smoked 100 or more cigarettes in one's lifetime, depending on the cohort (Liu et al. 2019; Saunders et al. 2022).
- **Drinks per week (DPW) ($N \approx 1,070,917$; $\frac{h^2 SNP}{s.e.(h^2 SNP)} = 21.32$)** (Liu et al. 2019; Saunders et al. 2022).
- **Drug experimentation (DRUG) ($N = 22,572$; $\frac{h^2 SNP}{s.e.(h^2 SNP)} = 4.69$)**, the number of different classes of drugs an individual has used out of eleven (Sanchez-Roige et al. 2019a).

Preparation of summary statistics for Genomic SEM

We formatted downloaded summary data from relevant GWAS using Genomic SEM's munge() function. We filtered summary statistics for SNPs available in the reference dataset, HapMap3, and further restricted to SNPs with a minimum minor allele frequency (MAF) of 0.01 and a minimum imputation quality (INFO) score of 0.90 (except for opioid use disorder, for which we used a minimum INFO threshold of 0.70) (Hatoum et al. 2022, 2023; Poore et al. 2023). Additionally, for each dichotomous trait, we used cohort-specific

prevalence rates to calculate the sum of effective sample sizes (N); effective N corresponds to the N for a GWAS with equal power to that of the cohort's raw sample size in a 1:1 case:control cohort design and is calculated as $4v_k(1 - v_k)n_k$, where v is equivalent to cohort-specific prevalence and n is equivalent to the cohort's raw sample size (Grotzinger et al. 2023). Thus, effective N corrects for ascertainment in studies that disproportionately sample for cases. Using the ldsc() function, we then calculated pairwise genetic covariances for all traits using the 1000 Genomes Phase 3 European ancestry linkage disequilibrium (LD) scores as a reference (The 1000 Genomes Project Consortium 2015 2022, 2023). Figure 1 shows the unstandardized genetic covariances (in the lower triangle) and genetic correlations (in the upper triangle) among the twelve traits we included in our main models, along with SNP-based heritabilities on the diagonal. An expanded genetic covariance/correlation matrix that additionally includes the pre-registered SUB indicators that were excluded from the final models is shown in Supplementary Fig. 1.

Factor analysis

Next, we used Genomic SEM to specify the three externalizing factors (SUD, SUB, and BD) as lower-order factors

Fig. 1 The genetic covariance/correlation matrix depicting unstandardized genetic covariances (in the lower left triangle) and genetic correlations (in the upper right triangle) for the twelve externalizing traits used in the main models. Diagonals show the SNP (single nucleotide polymorphism)-based heritabilities. *CI* Cannabis initiation, *SI* smoking initiation, *DRUG* drug experimentation (the number of different classes of drugs an individual has used out of eleven), *DPW* drinks per week, *FSEX* age at first sexual intercourse (reverse-coded), *ADHD* attention-deficit hyperactivity disorder, *NSEX* number of lifetime sexual partners, *RISK* general risk tolerance, *OUD* opioid use disorder, *CUD* cannabis use disorder, *PAU* problem alcohol use, *PTU* problem tobacco use

	CI	SI	DRUG	DPW	FSEX	ADHD	NSEX	RISK	OUD	CUD	PAU	PTU
CI	0.13	0.57	0.62	0.37	0.26	0.21	0.74	0.42	0.28	0.52	0.43	0.01
SI	0.07	0.1	0.85	0.35	0.64	0.6	0.54	0.32	0.54	0.71	0.55	0.39
DRUG	0.08	0.09	0.11	0.6	0.6	0.31	0.77	0.42	0.53	0.81	0.61	−0.02
DPW	0.02	0.02	0.03	0.03	0.19	0.01	0.26	0.2	0.34	0.23	0.65	0.02
FSEX	0.04	0.08	0.08	0.01	0.15	0.65	0.62	0.42	0.51	0.68	0.39	0.39
ADHD	0.03	0.09	0.05	0	0.11	0.2	0.42	0.35	0.43	0.59	0.35	0.48
NSEX	0.08	0.06	0.08	0.01	0.08	0.06	0.1	0.59	0.36	0.58	0.46	0.1
RISK	0.05	0.03	0.04	0.01	0.05	0.05	0.06	0.09	0.23	0.41	0.32	0.22
OUD	0.04	0.06	0.06	0.02	0.07	0.07	0.04	0.02	0.12	0.78	0.71	0.6
CUD	0.06	0.08	0.09	0.01	0.09	0.09	0.06	0.04	0.09	0.11	0.55	0.48
PAU	0.04	0.05	0.05	0.03	0.04	0.04	0.04	0.02	0.06	0.05	0.06	0.38
PTU	0	0.05	0	0	0.06	0.08	0.01	0.02	0.08	0.06	0.03	0.13
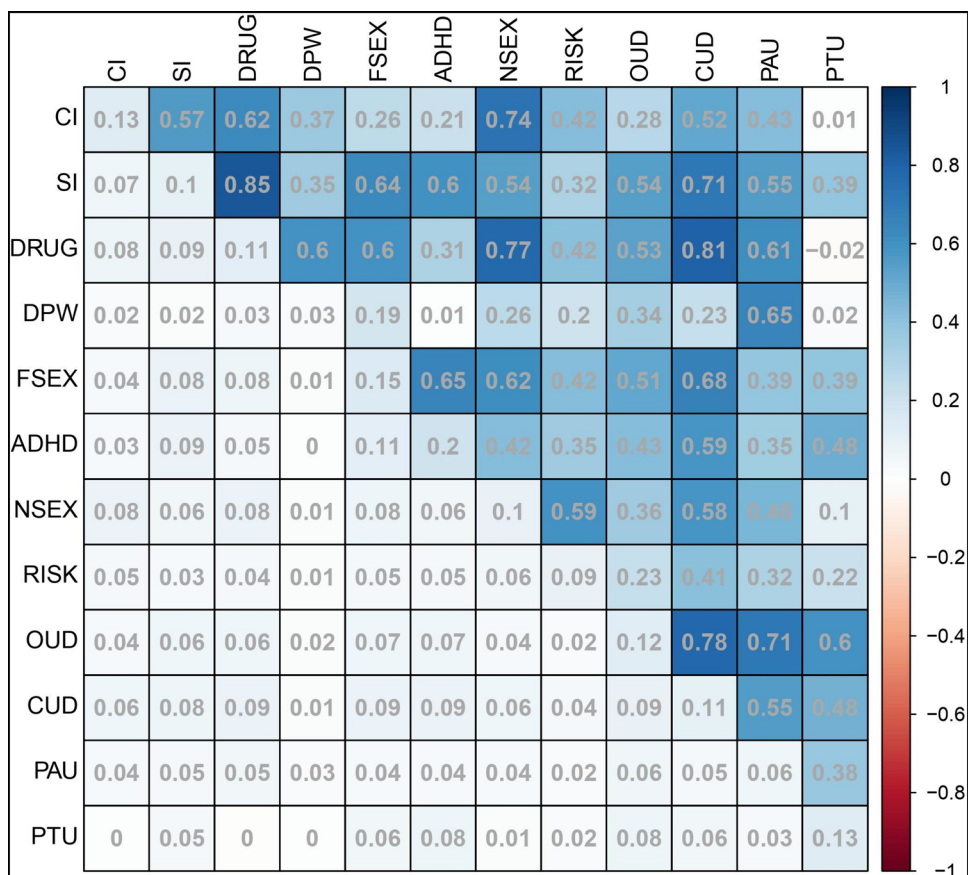
in the ensuing structural models. In order to extract influences of within-substance intercorrelations, we correlated the residuals of PTU and SI, CUD and CI, and PAU and DPW across the SUD and SUB factors. We constructed a correlated factors model including all three factors. Then, we integrated the lower-order factors into a hierarchical trivariate Cholesky decomposition framework. The orthogonal higher-order latent factors in the Cholesky decomposition completely broke down the genetic variance and covariance among the three lower-order latent factors into shared and independent sources of variance (Demange et al. 2021). The leftmost higher-order factor captured all genetic variance of the leftmost lower-order factor as well as its shared variance with the other lower-order factors. The middle higher-order factor captured all of the leftover genetic variance for the middle lower-order factor and its shared variance with the rightmost lower-order factor. Finally, the rightmost higher-order factor captured the genetic variance unique (in the model) to the rightmost lower-order factor. Model 1 had the following order of lower-order factors, from left-to-right: SUD, SUB, BD; and Model 2 had the order: SUB, SUD, BD. Thus, both models had BD in the final lower-order factor position. We utilized the usermodel() function in Genomic SEM to run both models using diagonally weighted least squares (DWLS) estimation. We scaled the higher-order factors using unit variance identification.

Model notation

In the model descriptions, the letters in the factors' subscripts denote the model number, where *a* corresponds to Model 1 and *b* corresponds to Model 2 in the two models that followed the pre-registered plan (additionally, *c* corresponds to Modified Model 1 and *d* to Modified Model 2 in the post-hoc models; see below). Meanwhile, the numerals in the factors' subscripts denote the level of the factor, where 1 denotes higher-order and 0 denotes lower-order.

Description of fit metrics

Though not the primary metric of interest for the decompositions, we assessed model fit for the measurement models, the main correlated factor models, and the main Cholesky decompositions using three indices: the comparative fit index (CFI), the Akaike Information Criterion (AIC), and the Standardized Root Mean Square Residual (SRMR). In Genomic SEM, the CFI measures the improvement in fit of the specified model compared to a model that estimates heritability of phenotypes but assumes no genetic covariances between them. The AIC is a relative fit index that balances fit with parsimony, incorporating χ^2 and the number of free parameters in the model. The SRMR, an index of approximate model fit, is the standardized root mean square difference between the model-implied and observed correlation matrices. CFI values, which can range from 0–1, are considered acceptable at 0.90 or greater and good if at least 0.95. Models with an SRMR under 0.10 suggest acceptable fit, while an SRMR under 0.05 indicates good fit. Similarly, lower AIC's indicate better fit.

Results

Lower-order BD, SUD, and SUB factors

As expected, the measurement models for the BD and SUD latent constructs had good model fit (CFI = 0.949, SRMR = 0.065, AIC = 181.793; CFI = 1.000, SRMR = 0.013, AIC = 16.277, respectively; Supplementary Tables S1, S2). We tested a series of models for our SUB factor (described further in the Supplementary Note; the associated parameters for these models are displayed in Supplementary Tables S3–S7). Our final four-indicator SUB measurement model (see Supplementary Table S7 and SUB Factor Section) satisfied our pre-registered criteria and achieved good model fit (CFI = 0.994, SRMR = 0.063, AIC = 26.145). Supplementary Fig. 2A–C how the measurement models for the BD, SUD, and SUB factors separately with their standardized loadings. The SUD and SUB factors each yielded one standardized loading (for OUD and DRUG, respectively) that was slightly greater than 1.00. However, these loadings dropped below 1.00 when the three latent factors were allowed to correlate with one another (Fig. 2) and when they were integrated into the Cholesky models (Supplementary Fig. 3A and B).

Correlated factors

Before performing Cholesky decomposition, we fit a correlated factors model, which showed substantial genetic correlations between the SUD factor, the BD factor, and the SUB factor (Fig. 2). The correlated factors model showed decent fit (CFI = 0.921, SRMR = 0.106, AIC = 1638.065). All indicators were associated with significant moderate-to-high loadings, ranging from a standardized loading of 0.367 (s.e. = 0.022) for DPW to a standardized loading of 0.961 (s.e. = 0.039) for CUD. The SUB factor correlated with SUD at $r_g = 0.803$ (s.e. = 0.027) and with BD at $r_g = 0.778$ (s.e. = 0.016), and the latter two correlated at $r_g = 0.774$ (s.e. = 0.027). Of the residual correlations, only the correlation between DPW's and PAU's error terms was significant (at $r_g = 0.666$, $p = 1.066\text{e-}34$). Thus, cross-factor associations for the same substance were unlikely to have been a primary driver of the inter-factor correlations. Rather, the variance captured in the latent SUD and SUB factors likely reflected
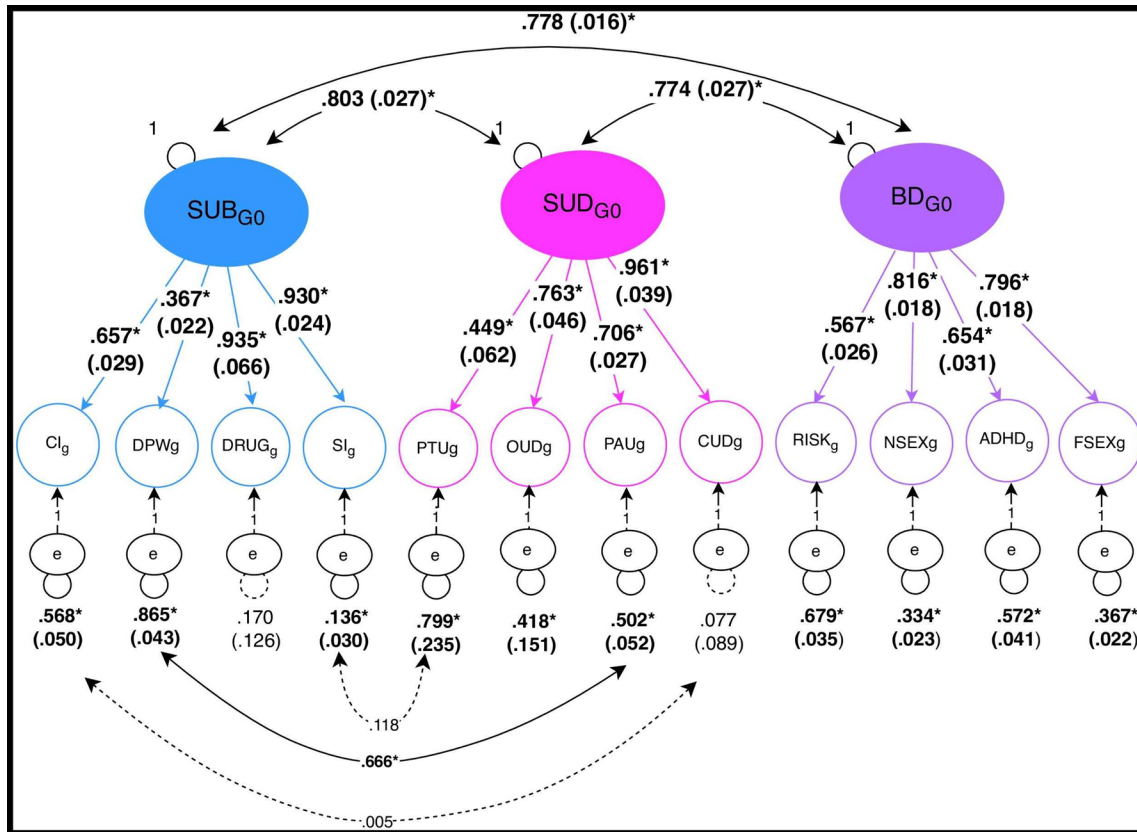
Fig. 2 A standardized correlated factors depiction of the three lower-order externalizing common factors depicted in the measurement models. In order to ensure that within-substance, cross-factor inter-correlations were not the primary drivers of the relationship between the factors, we correlated the residuals of PTU and SI, CUD and CI, and PAU and DPW across SUD and SUB. SUB_{G0} is the lower-order Substance Use factor, SUD_{G0} is the lower-order Substance Use Disorder factor, and BD_{G0} is the lower-order Behavioral Disinhibition factor. Each factor describes four indicators, the latter of which are defined by the summary results from genome-wide association studies for the trait in question. The *g* subscripts denote genomic variance. In addition to SUB_{G0}, SUD_{G0}, and BD_{G0}, the single trait

indicators are represented as latent factors to demonstrate that their genomic properties are not directly observed. *CI* Cannabis initiation, *SI* smoking initiation, *DRUG* drug experimentation (the number of different classes of drugs an individual has used out of eleven), *DPW* drinks per week, *FSEX* age at first sexual intercourse (reverse-coded), *ADHD* attention-deficit hyperactivity disorder, *NSEX* number of lifetime sexual partners, *RISK* general risk tolerance, *OUD* opioid use disorder, *CUD* cannabis use disorder, *PAU* problem alcohol use, *PTU* problem tobacco use. *$p < .05$. Standard errors are in parentheses, and dotted paths denote non-significance

common genomic variation across substance classes. The full parameter output for the three correlated factors model is in Supplementary Table S8.

Though our paper's goals centered around a trivariate structure, we also compared the fit of the three correlated factors model to that of a single, twelve-indicator common factor solution and to that of a two-factor model with a correlated BD factor and a SUB/SUD factor that combined indicators from the SUB and SUD factors, where both the common factor and two-factor models retained the same residual correlations used for the original correlated factors model. Both the common factor and two factor models fit significantly worse than the three factor model ($\Delta\chi^2(3) = 782.467, p < 0.001$ and $\Delta\chi^2(2) = 217.306, p < 0.001$, respectively). Still, it is

important to note that the chi-squared statistic is highly sensitive to small differences in Genomic SEM due to the well-powered nature of the GWAS used as input (Grotzinger et al. 2019). The full parameter output for the common factor model (CFI = 0.881, SRMR = 0.121, AIC = 2414.532) is in Supplementary Table S9, and the full parameter output for the two factor model (CFI = 0.910, SRMR = 0.115, AIC = 1851.371) is in Supplementary Table S10.

Cholesky decomposition

Models 1 and 2, the main Cholesky decompositions, are depicted in Supplementary Fig. 3A and B, respectively, along with loadings standardized with respect to the full model, including endogenous latent factors. Model 1

sought to capture the proportion of BD variance explained by a higher-order construction of SUD, as well as genetic covariance specific to lower-order SUB and lower-order BD independent of lower-order SUD. In Model 1, the first-position higher-order factor (SUD_{G1a}) captured all of the variance in the (left-most) lower-order SUD factor (SUD_{G0a}) and also captured covariance between SUD_{G0a} and the other lower-order factors in the model (SUB_{G0a} and BD_{G0a}) as well as covariance between the latter two that was shared with SUD_{G0a} variance. $SUBRes_{G1a}$ accounted for the remaining SUB_{G0a} factor variation and, by virtue of its partialing out covariance shared with SUD, also captured the independent influence of (non-pathological) substance-related indicators on BD_{G0a} in the model. Finally, $BDRes_{G1a}$ captured the BD-specific variation remaining after both pathological and non-pathological substance use were accounted for. In Model 2, the higher-order substance use factor (SUB_{G1b}) was the left-most factor such that SUB_{G1b} captured all of the variation in the corresponding lower-order factor (SUB_{G0b}), while residual higher-order substance use disorder ($SUDRes_{G1b}$) was in the middle position.

Partitioning of variance

The full parameter output for Models 1 and 2 (CFI = 0.921, SRMR = 0.106, AIC = 1638.066 for both models) are in Supplementary Tables S11 and S12, respectively. In Model 1, 64% of SUB_{G0a} and 60% of BD_{G0a} were shared with the higher-order SUD_{G1a} factor. Genetic variance associated with substance use but not substance use disorder ($SUBRes_{G1a}$) explained only 7% of the total genetic variance in BD_{G0a}. Approximately 33% of BD_{G0a} was not explained by either SUD_{G1a} or $SUBRes_{G1a}$ and was therefore BD-specific. Additionally, 20% of the genetic covariance between SUB_{G0a} and BD_{G0a} was independent of SUD_{G1a}. In Model 2, SUB_{G1b} explained 64% of SUD_{G0b} and 61% of BD_{G0b}, while $SUDRes_{G1b}$ accounted for 6% of BD_{G0b} variance. As in Model 1, 33% of BD_{G0b} was BD-specific. Finally, 19% of the genetic covariance between SUD_{G0b} and BD_{G0b} in Model 2 was independent of SUB_{G1b}.

Post-hoc models

Because the polysubstance use indicator (DRUG) in our SUB factor did not have a corresponding polysubstance indicator in the SUD factor and because the OUD indicator for SUD did not have a corresponding recreational opioid use indicator for SUB, we also tested a post-hoc iteration of our correlated factors and two initial Cholesky models that dropped the DRUG and OUD indicators. It is also important to note that, because of their relatively low sample sizes and high loadings, it is possible that these

indicators—particularly DRUG, with a standardized loading of 0.935 (and a large standard error (0.066) in the original correlated factors model)—disproportionately influenced their respective factors.

In a correlated factors model, there was a somewhat higher correlation between the modified SUD and modified SUB factors (r_g = 0.861, s.e. = 0.033). The point estimate of the correlation between BD and the modified SUD factor, r_g = 0.849 (s.e. = 0.035), was higher than that of the correlation between BD and the modified SUB factor (r_g = 0.773, s.e. = 0.017) (see Supplementary Fig. 4 and Supplementary Table S13). As a result, Model 1 and Model 2 with the Modified SUB and SUD factors, when compared to the original models, yielded somewhat less BD-specific variance (27% of total BD; see Supplementary Figs. 5A and 5B and Supplementary Tables S14 and S15 for depictions and full parameter outputs pertaining to Modified Models 1 and 2). Additionally, Modified Model 1 produced a non-significant cross-loading (p = 0.383) of BD_{G0c} on MOD_$SUBRes_{G1c}$. Thus, in this model, effectively all of the covariance between MOD_SUB_{G0c} and BD_{G0c} was shared with MOD_SUD_{G0c}. The correlation between SUB and modified SUD was 0.880 (s.e. = 0.033) and the correlation between SUD and modified SUB was 0.788 (se = 0.028). Thus, modifying SUB alone produced a factor structure that was closer to that observed in the original correlated factors model and modifying SUD alone produced a structure that more closely resembled the fully modified correlated factors model.

When we dropped the residual correlations from the original correlated factors model, the correlation between SUB and SUD increased somewhat (from 0.803 up to 0.852) (Supplementary Fig. 6, Supplementary Table S16). Correlations with BD remained essentially unchanged (from 0.778 to 0.780 for SUB; from 0.774 to 0.770 for SUD). Meanwhile, dropping the residual correlations in the modified model increased the point estimate of the correlation between (modified) SUB and (modified) SUD (from 0.861 up to 0.932), while BD's correlations with (modified) SUB and (modified) SUD did not change notably (from 0.773 to 0.780 and from 0.849 to 0.844, respectively) (Supplementary Fig. 7, Supplementary Table S17). Supplementary Figs. 8A and 8B and Supplementary Tables S18 and S19 correspond to results for Cholesky Decompositions utilizing the original factors from Models 1 and 2 with the residual correlations dropped. Supplementary Figs. 9A and 9B and Supplementary Tables S20-S21 correspond to results for Cholesky Decompositions utilizing the modified factors with the residual correlations dropped.

Discussion

Summary

We used Genomic SEM to partition the genome-wide components of externalizing behaviors into a four-indicator behavioral disinhibition (BD) factor, a four-indicator substance use disorder (SUD) factor, and a four-indicator (non-pathological) substance use (SUB) factor. The three factors showed high intercorrelations, even when the residuals for indicators relating to the same substance were correlated across the SUB and SUD factors. Of the covariances specified between the error terms, only the residual correlation between problem alcohol use (PAU) and drinks per week (DPW) was significant.

Using hierarchical trivariate Cholesky decomposition, we analyzed the residual genomic variance components that remained for externalizing above and beyond 1) SUD, 2) SUB, and 3) both SUD and SUB. The majority of the genetic variation in BD intersected with the joint variation shared across the other two domains, which reinforces findings from past research examining genomic links between externalizing constructs (Karlsson Linnér et al. 2021; Poore et al. 2023). The covariance between BD and each substance factor independent of the other was considerably smaller than covariance shared among all three factors.

The modified SUB and modified SUD factors, which included only the three most commonly-used recreational substance categories—nicotine/tobacco, alcohol, and cannabis—had a larger genetic correlation point estimate than did the original substance factors that included polysubstance use and opioid use disorder. Compared to the original model, the correlation between (modified) SUD and BD changed more drastically than did the correlation between (modified) SUB and BD. Additionally, dropping only OUD resulted in the point estimates of the correlation between SUB and SUD changing more drastically than dropping only DRUG. This implies that OUD more heavily drove the nature of the factor structure than DRUG. Dropping the residual correlations across SUB and SUD and across modified SUB and modified SUD increased the correlations between both the first and the second pair of factors only slightly. This was due to the fact that cross-substance correlations were often commensurate with—and in some cases higher than—within-substance correlations.

Finally, across all models, between approximately a quarter and approximately a third of BD-associated genetic variance was independent of both SUB and SUD.

Relevance to past research and future directions

Depending on the factor definition for SUB and SUD, the point estimates of their genetic correlation ranged from 0.803 to 0.932. In light of the large genomic overlap between the two, combining genomic information associated with pathological and non-pathological substance use phenotypes has the potential to boost power for substance-related research questions. For example, it is possible that combining such pathological and non-pathologic genetic information is a more powerful approach to predicting substance use disorder liability. However, it is important to note that the degree of overlapping genetic signal across univariate GWAS does not always directly translate to magnitude of overlapping genetic signal for latent factors. For example, Mallard et al. (2022) demonstrated that, when compared to standard GWAS meta-analysis, applying a multivariate framework to GWAS of AUDIT items changed alcohol consumption's genetic associations with problem use, psychiatric disorders, health outcomes, and socioeconomic outcomes substantially. In a similar vein, the genetic correlations between our latent SUB and SUD constructs were higher than what might be expected based on the bivariate genetic correlations between the SUD and SUB indicators, with only one of these pairwise correlations ($r_{gCUD,DRUG} = 0.81$) surpassing the lowest genetic correlation calculated between SUD and SUB ($r_g = 0.803$) across all of the correlated factors models.

Additionally, interrogating the *non*-overlapping genetic characteristics of these two factors—as well as that of their constituent traits—could provide illuminating insights into their shared and independent associations with psychopathological, psychological, disinhibitory, personality, and medical outcomes. For example, genomic SUB variance not implicated in SUD may be less likely to include genomic loci that are associated with negative side-effects upon cessation of substances. Somewhat less intuitively: even though substance use is a prerequisite for substance use disorder, the approach used in the present study and in similar designs allows for the dissociation of genetic prediction of pathological substance use outcomes from that of non-pathological substance use outcomes. Evidence has suggested that substance use disorders tend to be more positively genetically associated with certain psychopathological traits when compared to externalizing (Poore et al. 2023) and substance consumption (Sanchez-Roige et al. 2019b; Gelernter and Polimanti 2021; Mallard et al. 2022). Though it is currently unclear to what extent the sampling procedures for SUD GWAS—which draw many cases from psychiatric populations–may be contributing to this finding (Poore et al. 2023), loci implicated in SUD but not SUB could point to a critical subset of pleiotropic effects implicated in psychopathologies. Further exploration

of this independent signal may provide valuable context for attempts to identify treatments and evaluate sources of psychiatric or medical comorbidity. Additionally, it will be important to examine non-biological phenotypes such as socioeconomic status and clinician bias in further elucidating which substance users are most likely to be diagnosed with a substance use disorder.

Research has shown that SUD and externalizing are highly genetically correlated at the common latent factor level and that putatively causal SNPs of their constituent summary statistics are largely overlapping (Poore et al. 2023). The models we tested additionally revealed that there may be important specific covariance between BD and SUB/SUD not shared between all three factors. Though this independent covariance only accounted for one fifth of the total genetic covariance between the residual higher-order factors at most, it was nonetheless non-negligible in both of the main Cholesky models. Though this covariance only explained a small proportion of total BD variance in our models, the relationship between this residual genetic overlap and prediction of outcomes outside of an externalizing framework is worthy of further study.

Evidence in the literature (Brick et al. 2023; Poore et al. 2023) suggests there is a genome-wide/polygenic signal associated with externalizing and BD independent of substance use *and* substance use disorder, further indicating that there could be significant genomic correlates specific to the current study's construction of BD. While we evaluated BD using a combination of psychiatric and non-psychiatric traits—age at first sex (FSEX, reverse-coded), general risk tolerance (RISK), number of lifetime sexual partners (NSEX), and attention-deficit/hyperactivity disorder (ADHD)—all four traits relate to reward-seeking (i.e., sexual behaviors) and/or impulsivity (i.e., certain hyperactive characteristics in ADHD). While these mechanisms are relevant to substance use and substance use disorder as well, additional domains such as attentional processes (i.e., in ADHD) and subjective opinions of one's own disinhibitory tendencies (i.e., for self-reported risk tolerance), as well as related domains not explicitly captured in this study (i.e., certain facets of antisocial behavior) are more specifically characteristic of BD. Given the substantial proportion of BD-specific residual variance we calculated, additional investigation into this construct's genomic architecture could uncover an alternative molecular context for understanding risk-taking, reward-seeking, and sensation-seeking over and above its direct common pathways with substance use and/or substance use disorder.

Finally, future research on the genetics of externalizing could consider alternative orderings and measures of the subfacets we examined in this study. While prior evidence suggests that there may not be a robust genomic factor capturing PAU, OUD, PTU, and CUD that is orthogonal to externalizing (Poore et al. 2023), other avenues of investigation may be able to better isolate genomic signal relating to substance use disorder and/or substance use from that relating to behavioral disinhibition.

Limitations

It is important to note that there are limitations to consider in the interpretation of our findings. Firstly, the samples on which we based our analyses were not representative of the general human population. All of our samples came from individuals of European ancestry, and many of the GWAS summary measures were obtained from participants in the UK Biobank, which is an older-skewing British sample, while Opioid Use Disorder and Problem Alcohol Use utilized the Million Veterans Program sample, which is comprised entirely of U.S. veterans, is only 9% female, and skews towards older participants (U.S. Department of Veterans Affairs 2024). Some cohorts also utilized a combination of childhood and adult data (i.e., in the ADHD GWAS) (Demontis et al. 2023).

While our study provided important context for understanding the genetic architecture of externalizing at the genome-wide common SNP level, a more complete picture of substance use, substance use disorders, and reward-seeking behaviors, as well as of the relationships between these constructs, will require evaluation of environmental effects and non-genomic biological phenomena—which GWAS do not directly measure—as well as potential effects of gene-environment interactions. Factors such as racial/ethnic disparities in access to analgesic drugs (Samuel et al. 2019) and consequences of trauma on externalizing and substance use behaviors would add vital context to a broader understanding of the phenomena considered in the current study. In addition, we did not include specific single nucleotide polymorphism (SNP) effects in our model, and so we did not directly identify specific genes or biological pathways associated with the constructs we measured.

Because our models' latent factors were based on a small number of indicators, the GWAS we included in our models were unlikely to have exhaustively captured the genetic components associated with the focal constructs. Because very large sample sizes are required for powerful genome-wide analysis of complex traits, smaller sample sizes—in addition to less precise heritability estimates—for traits such as opioid use disorder (OUD), polysubstance use (DRUG), and problem tobacco use (PTU) may have disproportionately influenced results. Larger sample sizes and GWAS of more variable substance use outcomes (i.e., recreational opioid use, polysubstance addiction) will be instructive to portraying a clearer picture of the genetic architecture(s) of substance phenotypes.

Conclusion

Substance use disorder (SUD), substance use (SUB), and behavioral disinhibition (BD) represent three related constructs that have been subjects of common interest in psychological and genetic research. The current study leveraged factor analysis to build latent genomic factors relating to these three constructs. Using hierarchical Cholesky decomposition in Genomic SEM, this study also attempted to isolate the extent to which genetic variability underlying BD exists beyond shared variance with SUD and SUB and examined the nature of the genetic covariation between all three constructs.

BD was highly correlated with SUD and SUB, while SUD and SUB showed some evidence of being even more correlated with one another. More than half of the variation in BD could be explained by a broad higher-order factor absorbing variation across all three constructs, while, depending on the model, between none and a small fraction of the variation in BD operated via an independent pathway with SUD or SUB outside of this broad effect. Moreover, a significant minority of residual BD variability remained after partialing out covariance with *both* SUD and SUB, indicating that there is sizable genomic BD variation that does not overlap with substance use *or* substance use disorder. Future research could demonstrate the utility of boosting power through combining data across pathological and non-pathological substance use indicators though could also shed interesting light on the *differences* in genetic and molecular correlates of SUD vs. SUB, including outside of direct implications for behavioral disinhibition/externalizing. Extending these lines of inquiry is likely to yield important insights into genetic mechanisms linked to reward response, drug metabolization, risk-taking, and psychopathology.

Declarations

References

Brick LAD, Benca-Bachman C, Johnson EC, et al. (2023) Genetic associations among internalizing and externalizing traits with polysubstance use among young adults. 2023.04.04.23287779

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

Bulik-Sullivan BK, Loh P-R, Finucane HK et al (2015b) LD score regression distinguishes confounding from polygenicity in genome-wide association studies. Nat Genet 47:291–295. https://doi.org/10.1038/ng.3211

Demange PA, Malanchini M, Mallard TT et al (2021) Investigating the genetic architecture of noncognitive skills using GWAS-by-subtraction. Nat Genet 53:35–44

Demontis D, Walters GB, Athanasiadis G et al (2023) Genome-wide analyses of ADHD identify 27 risk loci, refine the genetic architecture and implicate several cognitive domains. Nat Genet 55:198–208. https://doi.org/10.1038/s41588-022-01285-8

Gelernter J, Polimanti R (2021) Genetics of substance use disorders in the era of big data. Nat Rev Genet 22:712–729. https://doi.org/10.1038/s41576-021-00377-1

Grotzinger AD, Rhemtulla M, de Vlaming R et al (2019) Genomic SEM provides insights into the multivariate genetic architecture of complex traits. Nat Hum Behav 3:513–525. https://doi.org/10.1038/s41562-019-0566-x

Grotzinger AD, de la Fuente J, Privé F et al (2023) Pervasive downward bias in estimates of liability-scale heritability in genome-wide association study meta-analysis: a simple solution. Biol Psychiatry 93:29–36. https://doi.org/10.1016/j.biopsych.2022.05.029

Gustavson DE, Stern EF, Reynolds CA et al (2024) Evidence for strong genetic correlations among internalizing psychopathology and related self-reported measures using both genomic and twin/adoptive approaches. J Psychopathol Clin Sci. https://doi.org/10.1037/abn0000905

Hancock DB, Guo Y, Reginsson GW et al (2018) Genome-wide association study across European and African American ancestries identifies a SNP in DNMT3B contributing to nicotine dependence. Mol Psychiatry 23:1911–1919

Hatoum AS, Johnson EC, Colbert SMC et al (2022) The addiction risk factor: a unitary genetic vulnerability characterizes substance use disorders and their associations with common correlates. Neuropsychopharmacology 47:1739–1745. https://doi.org/10.1038/s41386-021-01209-w

Hatoum AS, Colbert SMC, Johnson EC et al (2023) Multivariate genome-wide association meta-analysis of over 1 million subjects identifies loci underlying multiple substance use disorders. Nat Ment Health 1:210–223. https://doi.org/10.1038/s44220-023-00034-y

Johnson EC, Demontis D, Thorgeirsson TE et al (2020) A large-scale genome-wide association study meta-analysis of cannabis use disorder. Lancet Psychiatry 7:1032–1045. https://doi.org/10.1016/S2215-0366(20)30339-4

Karlsson Linnér R, Biroli P, Kong E et al (2019) Genome-wide association analyses of risk tolerance and risky behaviors in over 1 million individuals identify hundreds of loci and shared genetic influences. Nat Genet 51:245–257. https://doi.org/10.1038/s41588-018-0309-3

Karlsson Linnér R, Mallard TT, Barr PB et al (2021) Multivariate analysis of 1.5 million people identifies genetic associations with traits related to self-regulation and addiction. Nat Neurosci 24:1367–1376. https://doi.org/10.1038/s41593-021-00908-3

Kember RL, Vickers-Smith R, Zhou H et al (2023) 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 180:584–593. https://doi.org/10.1176/appi.ajp.21090892

Kendler KS, Myers J (2014) The boundaries of the internalizing and externalizing genetic spectra in men and women. Psychol Med 44:647–655. https://doi.org/10.1017/S0033291713000585

Koob GF, Volkow ND (2016) Neurobiology of addiction: a neurocircuitry analysis. Lancet Psychiatry 3:760–773. https://doi.org/10.1016/S2215-0366(16)00104-8

Krueger RF, Markon KE, Patrick CJ, Iacono WG (2005) Externalizing psychopathology in adulthood: a dimensional-spectrum conceptualization and its implications for DSM-V. J Abnorm Psychol 114:537–550. https://doi.org/10.1037/0021-843X.114.4.537

Liu M, Jiang Y, Wedow R et al (2019) 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. https://doi.org/10.1038/s41588-018-0307-5

Mallard TT, Savage JE, Johnson EC et al (2022) Item-level genome-wide association study of the alcohol use disorders identification test in three population-based cohorts. Am J Psychiatry 179:58–70. https://doi.org/10.1176/appi.ajp.2020.20091390

Mills MC, Tropf FC, Brazel DM et al (2021) Identification of 371 genetic variants for age at first sex and birth linked to externalising behaviour. Nat Hum Behav 5:1717–1730. https://doi.org/10.1038/s41562-021-01135-3

Pasman JA, Verweij KJH, Gerring Z et al (2018) GWAS of lifetime cannabis use reveals new risk loci, genetic overlap with psychiatric traits, and a causal effect of schizophrenia liability. Nat Neurosci 21:1161–1170. https://doi.org/10.1038/s41593-018-0206-1

Peterson RE, Kuchenbaecker K, Walters RK et al (2019) Genome-wide association studies in ancestrally diverse populations: opportunities, methods, pitfalls, and recommendations. Cell 179:589–603. https://doi.org/10.1016/j.cell.2019.08.051

Polimanti R, Walters RK, Johnson EC et al (2020) Leveraging genome-wide data to investigate differences between opioid use vs. opioid dependence in 41,176 individuals from the psychiatric genomics consortium. Mol Psychiatry 25:1673–1687. https://doi.org/10.1038/s41380-020-0677-9

Poore HE, Hatoum A, Mallard TT et al (2023) A multivariate approach to understanding the genetic overlap between externalizing phenotypes and substance use disorders. Addict Biol. https://doi.org/10.1111/adb.13319

Rosseel Y, Jorgensen TD, Rockwood N, et al. (2023) Lavaan: Latent Variable Analysis

Samuel CA, Corbie-Smith G, Cykert S (2019) Racial/ethnic disparities in pain burden and pain management in the context of opioid overdose risk. Curr Epidemiol Rep 6:275–289. https://doi.org/10.1007/s40471-019-00202-8

Sanchez-Roige S, Fontanillas P, Elson SL et al (2019a) Genome-wide association studies of impulsive personality traits (BIS-11 and UPPS-P) and drug experimentation in up to 22,861 adult research participants identify loci in the CACNA1I and CADM2 genes. J Neurosci off J Soc Neurosci 39:2562–2572. https://doi.org/10.1523/JNEUROSCI.2662-18.2019

Sanchez-Roige S, Palmer AA, Fontanillas P et al (2019b) Genome-wide association study meta-analysis of the alcohol use disorders identification test (AUDIT) in two population-based cohorts. Am J Psychiatry 176:107–118. https://doi.org/10.1176/appi.ajp.2018.18040369

Saunders GRB, Wang X, Chen F et al (2022) Genetic diversity fuels gene discovery for tobacco and alcohol use. Nature 612:720–724. https://doi.org/10.1038/s41586-022-05477-4

The 1000 Genomes Project Consortium (2015) A global reference for human genetic variation. 526:68

U.S. Department of Veterans Affairs (2024) About MVP | Veterans Affairs. https://www.mvp.va.gov/pwa/about. Accessed 22 Jan 2024

White JD, Bierut LJ (2023) Alcohol consumption and alcohol use disorder: exposing an increasingly shared genetic architecture. Am J Psychiatry 180:530–532. https://doi.org/10.1176/appi.ajp.20230456

Willer CJ, Li Y, Abecasis GR (2010) METAL: fast and efficient meta-analysis of genomewide association scans. Bioinforma Oxf Engl 26:2190–2191. https://doi.org/10.1093/bioinformatics/btq340

Zhou H, Rentsch CT, Cheng Z et al (2020a) Association of OPRM1 functional coding variant with opioid use disorder. JAMA Psychiat 77:1072–1080. https://doi.org/10.1001/jamapsychiatry.2020.1206

Zhou H, Sealock JM, Sanchez-Roige S et al (2020b) Genome-wide meta-analysis of problematic alcohol use in 435,563 individuals yields insights into biology and relationships with other traits. Nat Neurosci 23:809–818. https://doi.org/10.1038/s41593-020-0643-5