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

Tanya B. Horwitz[a,b,c*], Katerina Zorina-Lichtenwalter[a,d], Daniel E. Gustavson[a,e], Andrew D. Grotzinger[a,b,f] & Michael C. Stallings[a,b,g]

*Corresponding author

[a] Institute for Behavioral Genetics, University of Colorado Boulder, 1480 30th St, Boulder, CO, United States of America 80303.

[b] Psychology and Neuroscience, University of Colorado Boulder, Meunzinger D244, 345 UCB, Boulder, CO, United States of America 80303.

Emails:

[c]tanya.horwitz@colorado.edu; ORCID: 0000-0001-8031-3944

[d]katerina.zorina@colorado.edu; ORCID: 000-0003-4329-4961

[e]daniel.gustavson@colorado.edu; ORCID: 0000-0002-1470-4928

[f]andrew.grotzinger@colorado.edu; ORCID: 0000-0001-7852-9244

[g]michael.stallings@colorado.edu; ORCID: 0000-0002-8231-9159

1

Abstract

Externalizing behaviors encompass manifestations of risk-taking, self-regulation, aggression, sensation-/reward-seeking, and impulsivity. Externalizing research often includes substance use (SU), substance use disorder (SUD), and other (non-SU/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 applied trivariate Cholesky decomposition to these factors in order to identify BD-specific genomic variation and assess the partitioning of BD's genetic covariance with each of the other facets. 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 zero-order correlation (r_g=.803). BD correlated strongly with the SUD (r_g=.774) and SUB factors (r_g=.778). In our initial decompositions, 33% of total BD variance remained after removing variance associated with SUD and SUB. The majority of covariance between BD and SU and between BD and SUD was shared across all factors. When only nicotine/tobacco, cannabis, and alcohol were included for the SUB/SUD factors, their zero-order correlation increased to r_g=.861; in corresponding decompositions, BD-specific variance decreased to 27%. In summary, BD, SU, and SUD were highly genetically correlated at the latent factor level, and a significant minority of genomic BD variation was not shared with SU and/or SUD. Further research can better elucidate the properties of BD-specific variation by exploring its genetic/molecular correlates.

Keywords: Substance use, externalizing, behavioral genetics, GenomicSEM

Introduction

1.1 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 (Kendler and Myers, 2014; Krueger et al., 2005). 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 (Kendler and Myers, 2014; Krueger et al., 2005)--that do not directly measure substance use or substance use disorders, which we will refer 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 (Karlsson Linnér et al., 2021; Kendler and Myers, 2014; 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.,

3

2022). Though SUDs examined in large GWAS are unsurprisingly 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 = .97, s.e.= .12) (Hatoum et al., 2022), the correlation between cannabis use and cannabis use disorder is lower, at r_g = .50 (s.e. = .05) (Johnson et al., 2020). In addition to twin and family studies that have explored the genetics of externalizing based on phenotypic resemblance among relatives, a recent 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). This study identified 579 independent genomic loci associated with general externalizing. A more recent publication reported a broader externalizing factor that included additional SUD constructs (Poore et al., 2023). Factor-level analysis of general genomic SUD liability (Hatoum et al., 2023) has 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). 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.

1.2 Broad Analysis Plan

In the present study, we used genome-wide association data and Genomic Structural

Equation Modeling software (Grotzinger et al., 2019) to estimate genetic correlations between

externalizing traits, extract latent factors underlying three externalizing facets (SUD, SUB, and

BD), and decompose their shared and unshared genetic variance via hierarchical trivariate

Cholesky models. The findings from this analysis will shed light on whether phenotypic

subcategories of externalizing behavior constitute separable genomic entities.

Material and Methods

2.1 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

(S_{LDSC}) and an associated sampling covariance matrix (V_{S_LDSC}). 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 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 loading and correlation parameters that minimize discrepancy between the model-implied genetic covariance matrix and the empirical covariance matrix.

2.2 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^2SNP) 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 non-European ancestry groups are under-powered for some of the indicators of interest and may have different patterns of linkage disequilibrium and allele frequencies (Peterson et al., 2019). 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, as well as their associated publications.

2.2.1 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;**

  $\frac{h^2 SNP}{s.e.(h^2 SNP)}$ **= 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)** (Karlsson Linnér et al., 2019)

  (**N=370,711; $\frac{h^2 SNP}{s.e.(h^2 SNP)}$ = 28.77**)

- **Age at first sexual intercourse (FSEX)** (reverse-coded) (Mills et al., 2021) (**N

  =397,388; $\frac{h^2 SNP}{s.e.(h^2 SNP)}$ = 31.96**)

- **General risk tolerance (RISK) (Effective N = 268,876; $\frac{h^2 SNP}{s.e.(h^2 SNP)}$ = 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)

2.2.2 SUD Factor

The lower-order SUD factor described the indicators in Hatoum *et al.* (2022) and Hatoum

et al. (2023):

- **Problem alcohol use (PAU) (Effective N = 300,790; $\frac{h^2 SNP}{s.e.(h^2 SNP)}$ = 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 Identification Test problem items
(AUDIT-P), a continuous metric, or by the ICD-9 or -10, depending on the cohort
(Zhou et al., 2020b).

- **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).

- **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).

- **Problem tobacco use (PTU) (N = 15,988; $\frac{h^2 SNP}{s.e.(h^2 SNP)}$ = 4.22)**, based on the
  Fagerström Test for Nicotine Dependence (FTND) in ever-smokers (Hancock et
  al., 2018). 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 restrictions.

2.2.3   SUB Factor

We initially explored a total of seven pre-registered non-pathological substance use measures, three of which were not ultimately included. 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*≈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).

2.3 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 .01

9

and a minimum imputation quality (INFO) score of .90 (except for opioid use disorder, for which

we used a minimum INFO threshold of .70) (Hatoum et al., 2023, 2022; 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 (Bulik-Sullivan et al., 2015a; Bulik-Sullivan et

al., 2015b; The 1000 Genomes Project Consortium, 2015). 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

Figure 1.

2.4 Factor Analysis

Next, we used Genomic SEM to specify models with the three externalizing factors

(SUD, SUB, and BD) as lower-order factors 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

10

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; Neale, 1992). 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.

2.5 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.

2.6 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 .90 or greater and good if at least

.95. Models with an SRMR under .10 suggest acceptable fit, while an SRMR under .05 indicates

good fit. Similarly, lower AIC's indicate better fit (Grotzinger et al., 2019).

Results

3.1 Lower-Order BD, SUD, and SUB Factors

As expected, the measurement models for the BD and SUD latent constructs had good model

fit (CFI = .949, SRMR = .065, AIC = 181.793; CFI = 1.000, SRMR = .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 Section 2.2.3) satisfied our pre-registered criteria and achieved

good model fit (CFI = .994, SRMR = .063, AIC = 26.145). Supplementary Figures 2A-2C show

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 (Figure 2) and

when they were integrated into the Cholesky models (Figures 3A-3B).

3.2 Correlated Factors

Before performing Cholesky decomposition, we fit a correlated factors model, which

showed substantial zero-order genetic correlations between the SUD factor, the BD factor, and

the SUB factor (Figure 2). The correlated factors model showed decent fit (CFI =.921, SRMR =

.106, AIC =1638.065). All indicators were associated with significant moderate or high loadings,

ranging from a standardized loading of .367 (s.e.=.022) for DPW to a standardized loading of

.961 (s.e.=.039) for CUD. The SUB factor correlated with SUD at r_g = .803 (s.e. = .027) and with

BD at r_g = .778 (s.e.= .016), and the latter two correlated at r_g = .774 (s.e.= .027). Of the residual

correlations, only the correlation between DPW's and PAU's error terms was significant (at r_g =

.666, p = 1.066e-34). Thus, cross-factor associations for the same substance were not a primary

driver of the inter-factor correlations. Rather, the variance captured in the latent SUD and SUB

factors likely reflected common genomic variation across substance classes.

3.3 Cholesky Decomposition

Models 1 and 2, the main Cholesky decompositions, are depicted in Figures 3A and 3B,

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

13

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.

3.3.1 Fit Statistics and Model Comparison

Models 1 and 2 (CFI =.921, SRMR=.106, AIC =1638.063 for Model 1, 1638.067 for Model 2) were associated with a significant improvement in fit ($\Delta\chi^2(3) = 782.469$ for Model 1, 782.465 for Model 2, $p < .001$) when compared to a single twelve-indicator common factor solution (CFI =.881, SRMR = .121, AIC =2414.532), the latter of which produced standardized loadings ranging from .333 (s.e.=.021) for DPW to .868 (s.e.=.063) for DRUG (see Supplementary Table S8). However, 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

14

for the correlated factors model is in Supplementary Table S9, while the full parameter output for

Models 1 and 2 are in Supplementary Tables S10 and S11, respectively.

3.3.2 Partitioning of Variance

In Model 1, 64% of SUB_{G0a} and 60% of BD_{G0a} was 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}.

3.4 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. In a correlated factors model, there was a somewhat higher

correlation between the modified SUD and SUB factors (r_g = .861, s.e. = .033). The point

estimate of the correlation between BD and the modified SUD factor, r_g = .849 (s.e. = .035), was

higher than that of the correlation between BD and the modified SUB factor (r_g = .773, s.e. =

15

.017) (see Supplementary Figure 3 and Supplementary Table S12). 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). Additionally, Modified Model 1 produced a non-significant cross-loading ($p = .383$) 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} (see Supplementary Figures 4A-4B and see Supplementary Tables S13-S14 for depictions and full parameter outputs of Modified Models 1 and 2).

When we dropped the residual correlations, the zero-order correlation between SUB and SUD increased somewhat (from .803 up to .852) (Supplementary Figure 5, Supplementary Table S15) when compared to the original correlated factors model. Correlations with BD remained essentially unchanged (from .778 to .780 for SUB; from .774 to .770 for SUD). Meanwhile, dropping the residual correlations in the modified model increased the point estimate of the correlation between MOD_SUB and MOD_SUD (from .861 up to .932), while BD's zero-order correlations with MOD_SUB and MOD_SUD did not change notably (from .773 to .780 and from .849 to .844, respectively) (Supplementary Figure 6, Supplementary Table S16). Supplementary Figures 7A-7B and Supplementary Tables S17-S18 correspond to results for Cholesky Decompositions utilizing the original factors from Models 1 and 2 with the residual correlations dropped. Supplementary Figures 8A-8B and Supplementary Tables S19-S20 correspond to results for Cholesky Decompositions utilizing the modified factors with the residual correlations dropped.

Discussion

4.1 Summary

16

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 zero-order intercorrelations, even when the residuals for indicators relating to the same substance were correlated across the SUB and SUD factors. Using hierarchical trivariate Cholesky decomposition, we then 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 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. Finally, across all models, between approximately a quarter and approximately a third of BD-associated genetic variance was independent of both SUB and SUD.

4.2 Relevance to Past Research and Future Directions

Depending on the factor definition for SUB and SUD, the point estimates of their zero-order genetic correlation ranged from .803-.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

17

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. 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 (Gelernter and Polimanti, 2021; Mallard et al., 2022; Sanchez-Roige et al., 2019b). 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 SU/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 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.

Finally, 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) 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

19

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.

4.3 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 participants included were taken from 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 the relationships between these

constructs, will require evaluation of environment 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 exhaustively capture 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 biased results due to being underpowered. 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.

4.4 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. Using hierarchical Cholesky decomposition in Genomic SEM, this study attempted to isolate the extent to which genetic variability underlying BD exists beyond shared variance with the prior two constructs and examined the nature of the genetic covariation between all three constructs.

BD was highly correlated with SUD and SUB, the latter two of which showed some evidence of being more correlated with one another still. 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.

Trait	N	$Neff$	$h^2SNP/$ $s.e.(h^2SNP)$	Publication(s)	Summary Statistics Source
ADHD	225,534	141,035.20	21.12	https://www.nature.com/articles/s41588-022-01285-81	https://figshare.com/articles/dataset/adhd2022/22564390
NSEX	370,711	N/A	28.77	https://www.nature.com/articles/s41588-018-0309-32	https://thessgac.com/papers/2/5
FSEX	397,338	N/A	31.96	https://www.nature.com/articles/s41562-021-01135-33	http://ftp.ebi.ac.uk/pub/databases/gwas/summary_statistics/GCST90000001-GCST90001000/GCST90000047/
RISK	466,571	268,876	22.17	https://www.nature.com/articles/s41588-018-0309-32	https://thessgac.com/papers/2/1
PTU	15,988	N/A	4.22	https://www.ncbi.nlm.nih.gov/pmc/articles/PMC5882602/4	https://ftp-ncbi-nlm-nih-gov.colorado.idm.oclc.org/dbgap/studies/phs001532/analyses/
PAU	435,563	300,789.60	18.47	https://www.nature.com/articles/s41593-020-0643-55	https://figshare.com/articles/dataset/sud2019-alcuse/14672193 https://figshare.com/articles/dataset/sud2018-alc/14672187
CUD	357,806	47,952.51	10.71	https://www.ncbi.nlm.nih.gov/pmc/articles/PMC7674631/6	https://figshare.com/articles/dataset/sud2020-cud/14842692
OUD	82,707	32,703.49	6.14	https://www.ncbi.nlm.nih.gov/pmc/articles/PMC7270886/7	https://figshare.com/articles/dataset/sud2020-op/14672211
CI	184,765	144,698.95	15.96	https://www.ncbi.nlm.nih.gov/pmc/articles/PMC6386176/8	Available Upon Request
SI	~1,300,000	1,363,137.10	29.93	https://www.nature.com/articles/s41586-022-05477-49 (GSCAN) https://www.nature.com/articles/s41588-018-0307-510 (23andMe)	https://conservancy.umn.edu/handle/11299/241912

DPW	~1,070,917	N/A	21.32	https://www.nature.com/articles/s41586-022-05477-49 (GSCAN) https://www.nature.com/articles/s41588-018-0307-510 (23andMe)	https://conservancy.umn.edu/handle/11299/241912
DRUG	22,572	N/A	4.69	https://pubmed.ncbi.nlm.nih.gov/30718321/11	N/A
CPD	~360,808	N/A	20.16	https://www.nature.com/articles/s41586-022-05477-49 (GSCAN) https://www.nature.com/articles/s41588-018-0307-510 (23andMe)	https://conservancy.umn.edu/handle/11299/241912
DRINK STAT	398,853	376,693.97	21.22	https://www.nature.com/articles/s41588-018-0307-510	N/A
FREQ	462,016	N/A	19.98	https://onlinelibrary.wiley.com/doi/abs/10.1002/ajmg.b.3287412	N/A

Table 1. Summary data, associated publications, 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. 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 access 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 sections 2.2.1-2.2.3 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, DRINK STAT = Drinking Status (drinker vs. non-drinker), FREQ = Drinking Frequency.

Acknowledgements

Author Contributions

TBH: Conceptualization, Methodology, Software, Validation, Formal analysis, Resources, Data Curation, Writing – Original Draft, Visualization, Project administration **KZ:** Methodology, Writing – Review & Editing, Visualization **DEG:** Software, Writing – Review & Editing, Supervision **ADG:** Methodology, Software, Writing – Review & Editing, Supervision **MCS:** Conceptualization, Methodology, Resources, Writing – Review & Editing, Supervision, Project administration, Funding acquisition

Data Availability Statement

Links to available (unrestricted) summary statistics are provided in Table 1.

Declarations

Funding

Conflict of interest

The authors declare that they have no conflict of interest.

References

Brick, L.A.D., Benca-Bachman, C., Johnson, E.C., Gustavson, D., Carper, M., Palmer, R.H.C., 2023. Genetic associations among internalizing and externalizing traits with polysubstance use among young adults. https://doi.org/10.1101/2023.04.04.23287779

Bulik-Sullivan, B., Finucane, H.K., Anttila, V., Gusev, A., Day, F.R., Loh, P.-R., Duncan, L., Perry, J.R.B., Patterson, N., Robinson, E.B., Daly, M.J., Price, A.L., Neale, B.M., 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, B.K., Loh, P.-R., Finucane, H.K., Ripke, S., Yang, J., Patterson, N., Daly, M.J., Price, A.L., Neale, B.M., 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, P.A., Malanchini, M., Mallard, T.T., Biroli, P., Cox, S.R., Grotzinger, A.D., Tucker-Drob, E.M., Abdellaoui, A., Arseneault, L., Van Bergen, E., 2021. Investigating the genetic architecture of noncognitive skills using GWAS-by-subtraction. Nat. Genet. 53, 35–44.

Demontis, D., Walters, G.B., Athanasiadis, G., Walters, R., Therrien, K., Nielsen, T.T., Farajzadeh, L., Voloudakis, G., Bendl, J., Zeng, B., Zhang, W., Grove, J., Als, T.D., Duan, J., Satterstrom, F.K., Bybjerg-Grauholm, J., Bækved-Hansen, M., Gudmundsson, O.O., Magnusson, S.H., Baldursson, G., Davidsdottir, K., Haraldsdottir, G.S., Agerbo, E., Hoffman, G.E., Dalsgaard, S., Martin, J., Ribasés, M., Boomsma, D.I., Soler Artigas, M., Roth Mota, N., Howrigan, D., Medland, S.E., Zayats, T., Rajagopal, V.M., Nordentoft, M., Mors, O., Hougaard, D.M., Mortensen, P.B., Daly, M.J., Faraone, S.V., Stefansson, H., Roussos, P., Franke, B., Werge, T., Neale, B.M., Stefansson, K., Børglum, A.D., 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, A.D., de la Fuente, J., Privé, F., Nivard, M.G., Tucker-Drob, E.M., 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

Grotzinger, A.D., Rhemtulla, M., de Vlaming, R., Ritchie, S.J., Mallard, T.T., Hill, W.D., Ip, H.F., Marioni, R.E., McIntosh, A.M., Deary, I.J., Koellinger, P.D., Harden, K.P., Nivard, M.G., Tucker-Drob, E.M., 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

Hancock, D.B., Guo, Y., Reginsson, G.W., Gaddis, N.C., Lutz, S.M., Sherva, R., Loukola, A., Minica, C.C., Markunas, C.A., Han, Y., 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, A.S., Colbert, S.M.C., Johnson, E.C., Huggett, S.B., Deak, J.D., Pathak, G.A., Jennings, M.V., Paul, S.E., Karcher, N.R., Hansen, I., Baranger, D.A.A., Edwards, A., Grotzinger, A.D., Tucker-Drob, E.M., Kranzler, H.R., Davis, L.K., Sanchez-Roige, S., Polimanti, R., Gelernter, J., Edenberg, H.J., Bogdan, R., Agrawal, A., 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

Hatoum, A.S., Johnson, E.C., Colbert, S.M.C., Polimanti, R., Zhou, H., Walters, R.K., Gelernter, J., Edenberg, H.J., Bogdan, R., Agrawal, A., 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

Johnson, E.C., Demontis, D., Thorgeirsson, T.E., Walters, R.K., Polimanti, R., Hatoum, A.S., Sanchez-Roige, S., Paul, S.E., Wendt, F.R., Clarke, T.-K., Lai, D., Reginsson, G.W., Zhou, H., He, J., Baranger, D.A.A., Gudbjartsson, D.F., Wedow, R., Adkins, D.E., Adkins, A.E., Alexander, J., Bacanu, S.-A., Bigdeli, T.B., Boden, J., Brown, S.A., Bucholz, K.K., Bybjerg-Grauholm, J., Corley, R.P., Degenhardt, L., Dick, D.M., Domingue, B.W., Fox, L., Goate, A.M., Gordon, S.D., Hack, L.M., Hancock, D.B., Hartz, S.M., Hickie, I.B., Hougaard, D.M., Krauter, K., Lind, P.A., McClintick, J.N., McQueen, M.B., Meyers, J.L., Montgomery, G.W., Mors, O., Mortensen, P.B., Nordentoft, M., Pearson, J.F., Peterson, R.E., Reynolds, M.D., Rice, J.P., Runarsdottir, V., Saccone, N.L., Sherva, R., Silberg, J.L., Tarter, R.E., Tyrfingsson, T., Wall, T.L., Webb, B.T., Werge, T., Wetherill, L., Wright, M.J., Zellers, S., Adams, M.J., Bierut, L.J., Boardman, J.D., Copeland, W.E., Farrer, L.A., Foroud, T.M., Gillespie, N.A., Grucza, R.A., Harris, K.M., Heath, A.C., Hesselbrock, V., Hewitt, J.K., Hopfer, C.J., Horwood, J., Iacono, W.G., Johnson, E.O., Kendler, K.S., Kennedy, M.A., Kranzler, H.R., Madden, P.A.F., Maes, H.H., Maher, B.S., Martin, N.G., McGue, M., McIntosh, A.M., Medland, S.E., Nelson, E.C., Porjesz, B., Riley, B.P., Stallings, M.C., Vanyukov, M.M., Vrieze, S., Davis, L.K., Bogdan, R., Gelernter, J., Edenberg, H.J., Stefansson, K., Børglum, A.D., Agrawal, A., 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., Meddens, S.F.W., Wedow, R., Fontana, M.A., Lebreton, M., Tino, S.P., Abdellaoui, A., Hammerschlag, A.R., Nivard, M.G., Okbay, A., Rietveld, C.A., Timshel, P.N., Trzaskowski, M., Vlaming, R. de, Zünd, C.L., Bao, Y., Buzdugan, L., Caplin, A.H., Chen, C.-Y., Eibich, P., Fontanillas, P., Gonzalez, J.R., Joshi, P.K., Karhunen, V., Kleinman, A., Levin, R.Z., Lill, C.M., Meddens, G.A., Muntané, G., Sanchez-Roige, S., Rooij, F.J. van, Taskesen, E., Wu, Y., Zhang, F., Auton, A., Boardman, J.D., Clark, D.W., Conlin, A., Dolan, C.C., Fischbacher, U., Groenen, P.J.F., Harris, K.M., Hasler, G., Hofman, A., Ikram, M.A., Jain, S., Karlsson, R., Kessler, R.C., Kooyman, M., MacKillop, J., Männikkö, M., Morcillo-Suarez, C., McQueen, M.B., Schmidt, K.M., Smart, M.C., Sutter, M., Thurik, A.R., Uitterlinden, A.G., White, J., Wit, H. de, Yang, J., Bertram, L., Jan Bonder, M., Boomsma, D.I., Esko, T., Fehr, E., Hinds, D.A., Johannesson, M., Kumari, M., Laibson, D., Magnusson, P.K.E., Meyer, M.N., Navarro, A., Palmer, A.A., Pers, T.H., Posthuma, D., Schunk, D., Stein, M.B., Svento, R., Tiemeier, H., Timmers, P.R.H.J., Turley, P., Ursano, R.J., Wagner, G.G., Wilson,

J.F., Gratten, J., Kutalik, Z., Lee, J.J., Cesarini, D., Benjamin, D.J., Koellinger, P.D., Beauchamp, J.P., 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, T.T., Barr, P.B., Sanchez-Roige, S., Madole, J.W., Driver, M.N., Poore, H.E., de Vlaming, R., Grotzinger, A.D., Tielbeek, J.J., Johnson, E.C., Liu, M., Rosenthal, S.B., Ideker, T., Zhou, H., Kember, R.L., Pasman, J.A., Verweij, K.J.H., Liu, D.J., Vrieze, S., COGA Collaborators, Kranzler, H.R., Gelernter, J., Harris, K.M., Tucker-Drob, E.M., Waldman, I.D., Palmer, A.A., Harden, K.P., Koellinger, P.D., Dick, D.M., 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

Kendler, K.S., 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, G.F., Volkow, N.D., 2016. Neurobiology of addiction: a neurocircuitry analysis. Lancet Psychiatry 3, 760–773. https://doi.org/10.1016/S2215-0366(16)00104-8

Krueger, R.F., Markon, K.E., Patrick, C.J., Iacono, W.G., 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., Li, Y., Brazel, D.M., Chen, F., Datta, G., Davila-Velderrain, J., McGuire, D., Tian, C., Zhan, X., Choquet, H., Docherty, A.R., Faul, J.D., Foerster, J.R., Fritsche, L.G., Gabrielsen, M.E., Gordon, S.D., Haessler, J., Hottenga, J.-J., Huang, H., Jang, S.-K., Jansen, P.R., Ling, Y., Mägi, R., Matoba, N., McMahon, G., Mulas, A., Orrù, V., Palviainen, T., Pandit, A., Reginsson, G.W., Skogholt, A.H., Smith, J.A., Taylor, A.E., Turman, C., Willemsen, G., Young, H., Young, K.A., Zajac, G.J.M., Zhao, W., Zhou, W., Bjornsdottir, G., Boardman, J.D., Boehnke, M., Boomsma, D.I., Chen, C., Cucca, F., Davies, G.E., Eaton, C.B., Ehringer, M.A., Esko, T., Fiorillo, E., Gillespie, N.A., Gudbjartsson, D.F., Haller, T., Harris, K.M., Heath, A.C., Hewitt, J.K., Hickie, I.B., Hokanson, J.E., Hopfer, C.J., Hunter, D.J., Iacono, W.G., Johnson, E.O., Kamatani, Y., Kardia, S.L.R., Keller, M.C., Kellis, M., Kooperberg, C., Kraft, P., Krauter, K.S., Laakso, M., Lind, P.A., Loukola, A., Lutz, S.M., Madden, P.A.F., Martin, N.G., McGue, M., McQueen, M.B., Medland, S.E., Metspalu, A., Mohlke, K.L., Nielsen, J.B., Okada, Y., Peters, U., Polderman, T.J.C., Posthuma, D., Reiner, A.P., Rice, J.P., Rimm, E., Rose, R.J., Runarsdottir, V., Stallings, M.C., Stančáková, A., Stefansson, H., Thai, K.K., Tindle, H.A., Tyrfingsson, T., Wall, T.L., Weir, D.R., Weisner, C., Whitfield, J.B., Winsvold, B.S., Yin, J., Zuccolo, L., Bierut, L.J., Hveem, K., Lee, J.J., Munafò, M.R., Saccone, N.L., Willer, C.J., Cornelis, M.C., David, S.P., Hinds, D.A., Jorgenson, E., Kaprio, J., Stitzel, J.A., Stefansson, K., Thorgeirsson, T.E., Abecasis, G., Liu, D.J., Vrieze, S., 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, T.T., Savage, J.E., Johnson, E.C., Huang, Y., Edwards, A.C., Hottenga, J.J., Grotzinger, A.D., Gustavson, D.E., Jennings, M.V., Anokhin, A., Dick, D.M., Edenberg, H.J., Kramer, J.R., Lai, D., Meyers, J.L., Pandey, A.K., Harden, K.P., Nivard, M.G., de Geus, E.J.C., Boomsma, D.I., Agrawal, A., Davis, L.K., Clarke, T.-K., Palmer, A.A., Sanchez-

29

Roige, S., 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, M.C., Tropf, F.C., Brazel, D.M., van Zuydam, N., Vaez, A., eQTLGen Consortium, BIOS Consortium, Human Reproductive Behaviour Consortium, Pers, T.H., Snieder, H., Perry, J.R.B., Ong, K.K., den Hoed, M., Barban, N., Day, F.R., 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

Neale, M.C., 1992. Methodology for genetic studies of twins and families. Dordrecth: Kluwer Academic Publishers.

Pasman, J.A., Verweij, K.J.H., Gerring, Z., Stringer, S., Sanchez-Roige, S., Treur, J.L., Abdellaoui, A., Nivard, M.G., Baselmans, B.M.L., Ong, J.-S., Ip, H.F., van der Zee, M.D., Bartels, M., Day, F.R., Fontanillas, P., Elson, S.L., de Wit, H., Davis, L.K., MacKillop, J., Derringer, J.L., Branje, S.J.T., Hartman, C.A., Heath, A.C., van Lier, P.A.C., Madden, P.A.F., Mägi, R., Meeus, W., Montgomery, G.W., Oldehinkel, A.J., Pausova, Z., Ramos-Quiroga, J.A., Paus, T., Ribases, M., Kaprio, J., Boks, M.P.M., Bell, J.T., Spector, T.D., Gelernter, J., Boomsma, D.I., Martin, N.G., MacGregor, S., Perry, J.R.B., Palmer, A.A., Posthuma, D., Munafò, M.R., Gillespie, N.A., Derks, E.M., Vink, J.M., 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, R.E., Kuchenbaecker, K., Walters, R.K., Chen, C.-Y., Popejoy, A.B., Periyasamy, S., Lam, M., Iyegbe, C., Strawbridge, R.J., Brick, L., Carey, C.E., Martin, A.R., Meyers, J.L., Su, J., Chen, J., Edwards, A.C., Kalungi, A., Koen, N., Majara, L., Schwarz, E., Smoller, J.W., Stahl, E.A., Sullivan, P.F., Vassos, E., Mowry, B., Prieto, M.L., Cuellar-Barboza, A., Bigdeli, T.B., Edenberg, H.J., Huang, H., Duncan, L.E., 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

Poore, H.E., Hatoum, A., Mallard, T.T., Sandra Sanchez-Roige, Waldman, I.D., Palmer, A.A., Harden, K.P., Barr, P.B., Dick, D.M., 2023. A Multivariate Approach to Understanding the Genetic Overlap between Externalizing Phenotypes and Substance Use Disorders. Addict. Biol. 28. https://doi.org/10.1111/adb.13319

Rosseel, Y., Jorgensen, T.D., Rockwood, N., Oberski, D., Byrnes, J., Vanbrabant, L., Savalei, V., Merkle, E., Hallquist, M., Rhemtulla, M., Katsikatsou, M., Barendse, M., Scharf, F., Du, H., 2023. lavaan: Latent Variable Analysis.

Samuel, C.A., 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, S.L., Gray, J.C., de Wit, H., MacKillop, J., Palmer, A.A., 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, A.A., Fontanillas, P., Elson, S.L., 23andMe Research Team, the Substance Use Disorder Working Group of the Psychiatric Genomics Consortium,

Adams, M.J., Howard, D.M., Edenberg, H.J., Davies, G., Crist, R.C., Deary, I.J., McIntosh, A.M., Clarke, T.-K., 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, G.R.B., Wang, X., Chen, F., Jang, S.-K., Liu, M., Wang, C., Gao, S., Jiang, Y., Khunsriraksakul, C., Otto, J.M., Addison, C., Akiyama, M., Albert, C.M., Aliev, F., Alonso, A., Arnett, D.K., Ashley-Koch, A.E., Ashrani, A.A., Barnes, K.C., Barr, R.G., Bartz, T.M., Becker, D.M., Bielak, L.F., Benjamin, E.J., Bis, J.C., Bjornsdottir, G., Blangero, J., Bleecker, E.R., Boardman, J.D., Boerwinkle, E., Boomsma, D.I., Boorgula, M.P., Bowden, D.W., Brody, J.A., Cade, B.E., Chasman, D.I., Chavan, S., Chen, Y.-D.I., Chen, Z., Cheng, I., Cho, M.H., Choquet, H., Cole, J.W., Cornelis, M.C., Cucca, F., Curran, J.E., de Andrade, M., Dick, D.M., Docherty, A.R., Duggirala, R., Eaton, C.B., Ehringer, M.A., Esko, T., Faul, J.D., Fernandes Silva, L., Fiorillo, E., Fornage, M., Freedman, B.I., Gabrielsen, M.E., Garrett, M.E., Gharib, S.A., Gieger, C., Gillespie, N., Glahn, D.C., Gordon, S.D., Gu, C.C., Gu, D., Gudbjartsson, D.F., Guo, X., Haessler, J., Hall, M.E., Haller, T., Harris, K.M., He, J., Herd, P., Hewitt, J.K., Hickie, I., Hidalgo, B., Hokanson, J.E., Hopfer, C., Hottenga, J., Hou, L., Huang, H., Hung, Y.-J., Hunter, D.J., Hveem, K., Hwang, S.-J., Hwu, C.-M., Iacono, W., Irvin, M.R., Jee, Y.H., Johnson, E.O., Joo, Y.Y., Jorgenson, E., Justice, A.E., Kamatani, Y., Kaplan, R.C., Kaprio, J., Kardia, S.L.R., Keller, M.C., Kelly, T.N., Kooperberg, C., Korhonen, T., Kraft, P., Krauter, K., Kuusisto, J., Laakso, M., Lasky-Su, J., Lee, W.-J., Lee, J.J., Levy, D., Li, L., Li, K., Li, Y., Lin, K., Lind, P.A., Liu, C., Lloyd-Jones, D.M., Lutz, S.M., Ma, J., Mägi, R., Manichaikul, A., Martin, N.G., Mathur, R., Matoba, N., McArdle, P.F., McGue, M., McQueen, M.B., Medland, S.E., Metspalu, A., Meyers, D.A., Millwood, I.Y., Mitchell, B.D., Mohlke, K.L., Moll, M., Montasser, M.E., Morrison, A.C., Mulas, A., Nielsen, J.B., North, K.E., Oelsner, E.C., Okada, Y., Orrù, V., Palmer, N.D., Palviainen, T., Pandit, A., Park, S.L., Peters, U., Peters, A., Peyser, P.A., Polderman, T.J.C., Rafaels, N., Redline, S., Reed, R.M., Reiner, A.P., Rice, J.P., Rich, S.S., Richmond, N.E., Roan, C., Rotter, J.I., Rueschman, M.N., Runarsdottir, V., Saccone, N.L., Schwartz, D.A., Shadyab, A.H., Shi, J., Shringarpure, S.S., Sicinski, K., Skogholt, A.H., Smith, J.A., Smith, N.L., Sotoodehnia, N., Stallings, M.C., Stefansson, H., Stefansson, K., Stitzel, J.A., Sun, X., Syed, M., Tal-Singer, R., Taylor, A.E., Taylor, K.D., Telen, M.J., Thai, K.K., Tiwari, H., Turman, C., Tyrfingsson, T., Wall, T.L., Walters, R.G., Weir, D.R., Weiss, S.T., White, W.B., Whitfield, J.B., Wiggins, K.L., Willemsen, G., Willer, C.J., Winsvold, B.S., Xu, H., Yanek, L.R., Yin, J., Young, K.L., Young, K.A., Yu, B., Zhao, W., Zhou, W., Zöllner, S., Zuccolo, L., Batini, C., Bergen, A.W., Bierut, L.J., David, S.P., Gagliano Taliun, S.A., Hancock, D.B., Jiang, B., Munafò, M.R., Thorgeirsson, T.E., Liu, D.J., Vrieze, S., 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 [WWW Document]. URL https://www.mvp.va.gov/pwa/about (accessed 8.2.23).

Willer, C.J., Li, Y., Abecasis, G.R., 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, C.T., Cheng, Z., Kember, R.L., Nunez, Y.Z., Sherva, R.M., Tate, J.P., Dao, C., Xu, K., Polimanti, R., Farrer, L.A., Justice, A.C., Kranzler, H.R., Gelernter, J., 2020a. Association of OPRM1 Functional Coding Variant With Opioid Use Disorder. JAMA Psychiatry 77, 1072–1080. https://doi.org/10.1001/jamapsychiatry.2020.1206

Zhou, H., Sealock, J.M., Sanchez-Roige, S., Clarke, T.-K., Levey, D.F., Cheng, Z., Li, B., Polimanti, R., Kember, R.L., Smith, R.V., Thygesen, J.H., Morgan, M.Y., Atkinson, S.R., Thursz, M.R., Nyegaard, M., Mattheisen, M., Børglum, A.D., Johnson, E.C., Justice, A.C., Palmer, A.A., McQuillin, A., Davis, L.K., Edenberg, H.J., Agrawal, A., Kranzler, H.R., Gelernter, J., 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

	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