Original Article

CAMBRIDGE
UNIVERSITY PRESS

Polygenic overlap of substance use behaviors and disorders with externalizing and internalizing problems independent of genetic correlations

Laila Al-Soufi[1,2] , Guy Hindley[3], Linn Rødevand[3], Alexey A. Shadrin[3], Piotr Jaholkowski[3], Vera Fominykh[3], Romain Icick[3,4], Markos Tesfaye[3,5], Javier Costas[1,6] and Ole A. Andreassen[3]

[1]Psychiatric Genetics group, Instituto de Investigación Sanitaria de Santiago de Compostela (IDIS), Santiago de Compostela, Galicia, Spain. Red de Investigación en Atención Primaria de Adicciones (RIAPAd); [2]Department of Zoology, Genetics and Physical Anthropology, Universidade de Santiago de Compostela (USC), Santiago de Compostela, Galicia, Spain; [3]Centre for Precision Psychiatry, Division of Mental Health and Addiction, Institute of Clinical Medicine, University of Oslo, Oslo, Norway; [4]Université Paris-Cité, INSERM, Optimisation thérapeutique en neuropsychopharmacologie OPTEN U1144, 75006, Paris, France; [5]Department of Clinical Science, University of Bergen, Bergen, Norway and [6]Complexo Hospitalario Universitario de Santiago de Compostela (CHUS), Servizo Galego de Saúde (SERGAS), Santiago de Compostela, Galicia, Spain

Abstract

Background. Externalizing and internalizing pathways may lead to the development of substance use behaviors (SUBs) and substance use disorders (SUDs), which are all heritable phenotypes. Genetic correlation studies have indicated differences in the genetic susceptibility between SUBs and SUDs. We investigated whether these substance use phenotypes are differently related to externalizing and internalizing problems at a genetic level.
Methods. We analyzed data from genome-wide association studies (GWAS) of four SUBs and SUDs, five externalizing traits, and five internalizing traits using the bivariate causal mixture model (MiXeR) to estimate genetic overlap beyond genetic correlation.
Results. Two distinct patterns were found. SUBs demonstrated high genetic overlap but low genetic correlation of shared variants with internalizing traits, suggesting a pattern of mixed effect directions of shared genetic variants. Conversely, SUDs and externalizing traits exhibited considerable genetic overlap with moderate to high positive genetic correlation of shared variants, suggesting concordant effect direction of shared risk variants.
Conclusions. These results highlight the importance of the externalizing pathway in SUDs as well as the limited role of the internalizing pathway in SUBs. As MiXeR is not intended for the identification of specific genes, further studies are needed to reveal the underlying shared mechanisms of these traits.

Introduction

Substance use stands as a major risk factor contributing significantly to the global burden of disease (Degenhardt et al., 2018). To improve prevention and treatment, it is important to increase our understanding of the pathogenesis of substance use behaviors (SUBs) (e.g. initiation, lifetime use, and severity of consumption) and substance use disorders (SUDs). This includes identifying factors influencing different aspects of substance use, from severity of substance consumption to pathological substance use.

Evidence from developmental psychology suggests two main pathways for substance use initiation and escalation towards SUDs: an *externalizing* pathway characterized by high levels of impulsivity, aggression, and sensation-seeking, and an *internalizing* pathway characterized by using substances to cope with negative affect, depression, and anxiety (Hussong, Jones, Stein, Baucom, & Boeding, 2011; Krueger et al., 2002). This is supported by prospective evidence of distinct externalizing and internalizing pathways leading to future SUBs and SUDs (Meque, Dachew, Maravilla, Salom, & Alati, 2019; Ning, Gondek, Patalay, & Ploubidis, 2020). However, evidence is weaker for the internalizing pathway. This may be due to comorbidity between internalizing and externalizing problems and differences regarding gender, developmental timing, or type of substance use (Heitzeg, Hardee, & Beltz, 2018; Ning et al., 2020). For instance, internalizing traits seem to be related to reduced alcohol consumption in middle adulthood subjects, but to increased consumption in more severe outcomes (i.e. heavy drinking and alcohol use disorder [AUD]) during transition to adulthood (Ning et al., 2020).

Additional support for a causal effect of both externalizing and internalizing problems on SUBs and SUDs comes from the Mendelian randomization studies. Mendelian randomization leverages information from molecular genetic studies to infer causal relationships between phenotypes (Chen, Tubbs, Liu, Thach, & Sham, 2024). This approach found evidence for a causal effect of attention-deficit/hyperactivity disorder (ADHD) on tobacco and cannabis use behaviors, but also for reverse causation (Treur et al., 2019; Vilar-Ribó et al., 2020). It also indicated that depression and neuroticism are risk factors for tobacco use and alcohol dependence and discarded reverse causation (Polimanti et al., 2019; Sallis, Davey Smith, & Munafò, 2019; Yao et al., 2021). However, there is also some evidence for a bidirectional causal relationship between depression and smoking behavior (Wootton et al., 2020).

Twin and family studies have estimated heritability of around 40–60% for different SUDs, partly specific and partly shared between substances (Deak & Johnson, 2021). Genome-wide association studies (GWAS) revealed that a considerable portion of the heritability is due to the additive effect of many common variants of low individual effect, that is SUDs are highly polygenic traits. Many of these genetic variants are shared between SUDs and among SUDs and related phenotypes (Deak & Johnson, 2021; Gelernter & Polimanti, 2021). Polygenic scores (PGS) may be used to estimate the risk of SUD as the sum of susceptibility alleles carried by an individual weighted by their effect. Cross-trait PGS confirmed the existence of shared genetic susceptibility between related phenotypes. For instance, PGS of the average number of alcoholic drinks consumed per week (DrnkWk), major depressive disorder (MDD), and ADHD are associated with AUD (Facal et al., 2021), PGS of ADHD is associated with tobacco use (Leppert et al., 2020), and PGS of smoking initiation with DrnKWk, nicotine dependence and conduct disorder (Chang et al., 2019).

Among the different methods to measure shared genetic susceptibility, cross-trait Linkage Disequilibrium (LD) score regression has reached enormous popularity in recent years due to the main advantage that it relies only on GWAS summary statistics. LD score regression calculates the correlation coefficient of additive genetic effects for two traits across the genome (genetic correlation, r_g) assuming a simple infinitesimal model, that is all SNPs are causal with normally distributed additive genetic effects (Bulik-Sullivan et al., 2015). Recent studies have examined the genetic correlation between different SUBs and SUDs using LD score regression. They have shown that smoking severity, defined as the number of cigarettes smoked per day (Cigsperday), and nicotine dependence are almost genetically identical ($r_g = 0.95$) (Quach et al., 2020). In fact, Cigsperday was proposed as a 'genetic proxy' for nicotine dependence, considering their strong genetic correlation and their similar pattern of associations with multiple diseases (Sanchez-Roige, Cox, Johnson, Hancock, & Davis, 2021). By contrast, the genetic correlation between alcohol consumption, measured as the number of drinks per week (DrnkWk), and alcohol dependence was weaker ($r_g = 0.37$) (Walters et al., 2018). Similarly, there is a moderate genetic correlation between smoking initiation and nicotine dependence ($r_g = 0.40$) (Quach et al., 2020). Low to moderate genetic correlation also exists between SUBs involving different substances, such as smoking initiation and DrnkWk ($r_g = 0.27$) (Saunders et al., 2022).

Significant genetic correlations have also been identified between both externalizing and internalizing traits and both SUBs and SUDs across different substances. For example, ADHD was genetically correlated with SUDs such as nicotine dependence, alcohol dependence and cannabis use disorder (CUD) (Abdellaoui, Smit, van den Brink, Denys, & Verweij, 2021; Vink, Treur, Pasman, & Schellekens, 2020; Walters et al., 2018); as well as with SUBs such as tobacco and cannabis use (Liu et al., 2019; Pasman et al., 2018; Soler Artigas et al., 2020). Other risky-behavior traits considered as externalizing were associated with SUBs. For example, age at first sexual intercourse (AFS) and age at first birth (AFB) showed a significant genetic correlation with age of smoking initiation and cannabis use (Mills et al., 2021). Further, other studies reported a significant genetic correlation of depression with both SUBs and SUDs, including smoking initiation, smoking cessation, nicotine dependence, and alcohol dependence (Abdellaoui et al., 2021; Liu et al., 2019; Vink et al., 2020; Walters et al., 2018).

Despite the insight provided by genome-wide genetic correlation to understand shared genetic susceptibility among related phenotypes, it has a clear limitation. Genome-wide genetic correlation, as estimated by LD score regression, can capture genetic overlap between two phenotypes only when the majority of shared variants exhibit a strong net correlation in the direction of effects relative to each other (i.e. effects are correlated in the same or opposite directions in the two phenotypes). However, it does not allow for the assessment of genetic overlap if the shared variants reveal a balanced mixture of concordant and discordant directions of effects. The bivariate causal mixture model (MiXeR) overcomes this limitation, characterizing the pattern of genetic overlap between two traits beyond genome-wide genetic correlation (Frei et al., 2019). First, MiXeR estimates the number of causal variants involved in each trait, instead of assuming an infinitesimal model. Next, MiXeR estimates the number of shared variants between the two traits and the genetic correlation of shared variants. Therefore, in contrast to genome-wide genetic correlation studies, MiXeR enables the discovery of genetic overlap even if the shared variants possess a mixture of concordant and discordant effect directions.

In the current study, we aimed to characterize the genetic relationship of SUBs and SUDs with externalizing and internalizing traits using MiXeR. Specifically, we investigated if: (i) SUBs and SUDs differ in their pattern of polygenic overlap with externalizing and internalizing traits, including the proportion of genetic overlap and genetic correlation of shared variants, and (ii) these differences are present regardless of the particular substance involved. This approach can improve our understanding of the contribution of the externalizing and internalizing pathways to SUBs and SUDs.

Methods

Genome-wide association studies (GWAS)

We included summary statistics for SUBs and SUDs and externalizing and internalizing traits from large GWAS (Table 1 and Supplementary Table 1). These GWAS were selected according to their large sample size and public availability. All GWAS summary statistics were based on samples of European ancestry to share similar linkage disequilibrium patterns, a requirement for MiXeR analysis.

For the three most studied drugs, tobacco, alcohol, and cannabis, we selected at least one SUB and one SUD or SUD-proxy. As tobacco use behavior, we chose ever be a regular smoker (EverSmk) (Saunders et al., 2022), which reflects smoking initiation. To assess tobacco use disorder (TUD), we generated our own GWAS of TUD using UK Biobank data (further information in Supplementary Methods). UK Biobank data was obtained under accession number 27412. We also used Cigsperday (Saunders et al., 2022), which is a good 'genetic proxy' of nicotine dependence (Sanchez-Roige et al.,

Table 1. GWAS summary statistics used in MiXeR analysis

Trait	Description	Source	Sample size (N cases)
SUBs and SUDs or SUD-proxies			
EverSmk	Ever been a regular smoker, i.e. ever vs. never regular smokers	(Saunders et al., 2022)	805,431 (393,707)
Cigsperday	Average number of cigarettes smoked per day, either as a current or as a former smoker	(Saunders et al., 2022)	326,497
DrnkWk	Average number of drinks drunk each week	(Saunders et al., 2022)	666,978
AUD	Alcohol use disorder	(Icick et al., 2023)	981,384 (75,635)
Externalizing traits			
ADHD	Attention deficit hyperactivity disorder	(Demontis et al., 2023)	225,534 (38,691)
RISK	General risk tolerance	(Karlsson Linnér et al., 2019)	466,571
NSEX	Lifetime number of sexual partners	(Karlsson Linnér et al., 2019)	370,711
AFB	Age at first birth	(Mills et al., 2021)	542,901
AFS	Age at first sexual intercourse	(Mills et al., 2021)	397,338
Internalizing traits			
DEP	Lifetime diagnosis of major depressive disorder or broad self-reported depression	(Howard et al., 2019)	500,199 (170,756)
NEURO	Weighted neuroticism score based on the response to 12 items	(Nagel, Jansen, et al., 2018)	390,278
WORRY-NEURO	Worry subcluster of NEURO	(Nagel, Jansen, et al., 2018)	348,219
DEP-NEURO	Depressed affect subcluster of NEURO	(Nagel, Jansen, et al., 2018)	357,957
MOOD	Experiencing mood swings	(Nagel, Watanabe, et al., 2018)	373,733 (168,180)

2021). As alcohol use behaviors, we selected DrnkWk (Saunders et al., 2022), a score comprising the items related to alcohol consumption of the Alcohol Use Disorders Identification Test (AUDIT-C) (Sanchez-Roige et al., 2019) and AUDIT-P (Sanchez-Roige et al., 2019). AUD was assessed by an AUD GWAS (Icick et al., 2023). For cannabis, we selected lifetime cannabis use (CANN) (Pasman et al., 2018) and CUD (Johnson et al., 2020).

As externalizing traits, we chose those four traits that were previously used to construct an externalizing genetic factor and do not involve substance use (Karlsson Linnér et al., 2021): ADHD (Demontis et al., 2023), AFS (Mills et al., 2021), number of sexual partners (NSEX) (Karlsson Linnér et al., 2019), and general risk tolerance (RISK) (Karlsson Linnér et al., 2019). In addition, AFB (Mills et al., 2021) and childhood aggressive behavior (AGG) (Ip et al., 2021), related to externalizing behavior (Baselmans et al., 2021; Mills et al., 2021), were also included.

For internalizing traits, different phenotypes of depression and anxiety were selected: a broader diagnosis of depression (DEP) (Howard et al., 2019), MDD (Wray et al., 2018), a continuous trait for anxiety based on the Generalized Anxiety Disorder 2-item scale (GAD) (Levey et al., 2020) and self-reported anxiety (ANX) (http://www.nealelab.is/uk-biobank/). Moreover, neuroticism, a core feature of internalizing behavior (Griffith et al., 2010), measured by a neuroticism score based on 12 items (NEURO) was also included (Nagel, Jansen, et al., 2018). Specific subclusters and items of NEURO were also analyzed: depressed affect subcluster of NEURO (DEP-NEURO) (Nagel, Jansen, et al., 2018), worry subcluster of NEURO (WORRY-NEURO) (Nagel, Jansen, et al., 2018), mood instability (MOOD) (Nagel, Watanabe, Stringer, Posthuma, & Van Der Sluis, 2018), and often feeling lonely (LONE) (Nagel, Watanabe, et al., 2018). The two subclusters of NEURO show distinct patterns of genetic correlation with behavioral, neuropsychiatric, anthropometric, and health-related phenotypes, while a genetic correlation of NEURO is a mix of both subclusters (Nagel, Jansen, et al., 2018; Nagel, Watanabe, et al., 2018). Subjective well-being (SWB) (Kim et al., 2022) was also selected as it is highly correlated with anxious/depressive behavior (Bartels, Cacioppo, Van Beijsterveldt, & Boomsma, 2013). Other aspects of loneliness were also included: the frequency of being able to confide in someone close (CONF) (http://www.nealelab.is/uk-biobank/) and a combined multi-trait GWAS of loneliness and social isolation (LONE-MTAG) (Day, Ong, & Perry, 2018), which is strongly genetically correlated with neuroticism and depressive symptoms.

Data analysis

We applied MiXeR, version 1.2.0, to characterize the pattern of genetic overlap between SUBs and SUDs and externalizing and internalizing traits using GWAS summary statistics (Frei et al., 2019). First, univariate MiXeR was applied to each trait to estimate their polygenicity, that is the fraction of causal variants in the genome, and discoverability, that is the causal effect size variance, maximizing likelihood estimation (Holland et al., 2020). SNP-based heritability (h^2_{SNP}) is derived from the product of these estimates, taking heterozygosity into account (Frei et al., 2019). Polygenicity is presented as the number of causal variants required to explain 90% of h^2_{SNP}. This threshold of 90% is applied to exclude those variants with infinitesimally small effects. Next, bivariate MiXeR was applied to assess the pattern of genetic overlap between each pair of traits. Using the parameters estimated in univariate MiXeR for each trait, bivariate MiXeR estimates the number of shared causal variants, the number of unique causal variants specific to each trait and the correlation of effect sizes of shared variants (r_{gs}). To enhance result interpretation, we assessed the degree of genetic overlap between each pair of traits by quantifying the percentage of shared variants with respect to the SUB/SUD. To assess model fit, the difference between the Akaike information criterion (AIC) for the reference model and the MiXeR-fitted model was calculated. Positive AIC differences indicate that the MiXeR-fitted estimates are more informative than the reference model. The minimum and maximum possible overlap are used as the reference

models in bivariate MiXeR. The minimum possible overlap between two traits is restricted by their genome-wide genetic correlation and the maximum possible overlap is restricted by the polygenicity of the least polygenic trait. Further information about MiXeR is provided in Supplementary Methods and the original publication (Frei et al., 2019).

It is well known that results for methods based on summary statistics are dependent on the power of the GWAS. Underpowered GWAS may lead to unreliable results. To limit bivariate MiXeR analyses to the most well-powered GWAS, we selected GWAS according to their univariate MiXeR model fit. Specifically, we selected those traits with substantial improvement in model fit by the causal mixture model in comparison to the infinitesimal model, which assumes all variants as causal. An arbitrary threshold of 50 as the difference in AIC between the infinitesimal model and the causal mixture model was applied. The difference in AIC between these two models is strongly correlated with the power of a GWAS, roughly estimated as the product of the effective sample size and h^2_{SNP} in our data (Pearson's correlation coefficient = 0.92, 95% CI: 0.83–0.96) (Supplementary Methods and Supplementary Figure 1). Using BIC as an alternative measure of model fit would lead to a similar selection (Supplementary Table 2). For Cigsperday GWAS, we excluded the cluster of neuronal nicotinic acetylcholine receptor (nAchR) genes at chromosome 15 because its strong effect on this phenotype violates MiXeR's assumption that the effect sizes of causal variants are normally distributed. Further details about the exclusion of the cluster of nAchR genes at chromosome 15 for Cigsperday are shown in Supplementary Methods.

Results

After excluding GWAS with poor MiXeR model fit, we included the following tobacco and alcohol use behaviors, respectively: EverSmk and DrnkWk; and a SUD or SUD-proxy for each of these substances: Cigsperday and AUD (Figure 1). We included five externalizing traits: AFS, NSEX, AFB, RISK, and ADHD; and five internalizing traits: NEURO, DEP-NEURO, WORRY-NEURO, DEP, and MOOD.

Univariate MiXeR results

Univariate MiXeR estimated the polygenicity and discoverability of each of the selected traits (Supplementary Table 2). Among SUBs and SUDs, EverSmk was the most polygenic trait (N = 11.6 K causal variants explaining 90% of the h^2_{SNP}, SE = 0.7 K). DrnkWk and AUD had similar polygenicity (DrnkWk: N = 7.7 K, SE = 0.6 K; AUD: N = 7.7 K, SE = 0.3 K). Cigsperday was the least polygenic trait with 4.3 K (SE = 0.6 K) variants involved. All selected externalizing and internalizing traits were highly polygenic with 9.7 K–12.2 K variants explaining 90% of the h^2_{SNP}, except for ADHD with 7.7 K (SE = 0.4 K) causal variants.

Bivariate MiXeR results

Bivariate MiXeR was used to estimate the pattern of genetic overlap between each of the four selected SUBs and SUDs and the 10 selected externalizing and internalizing traits. Specifically, the percentage of shared variants with respect to the SUB/SUD and
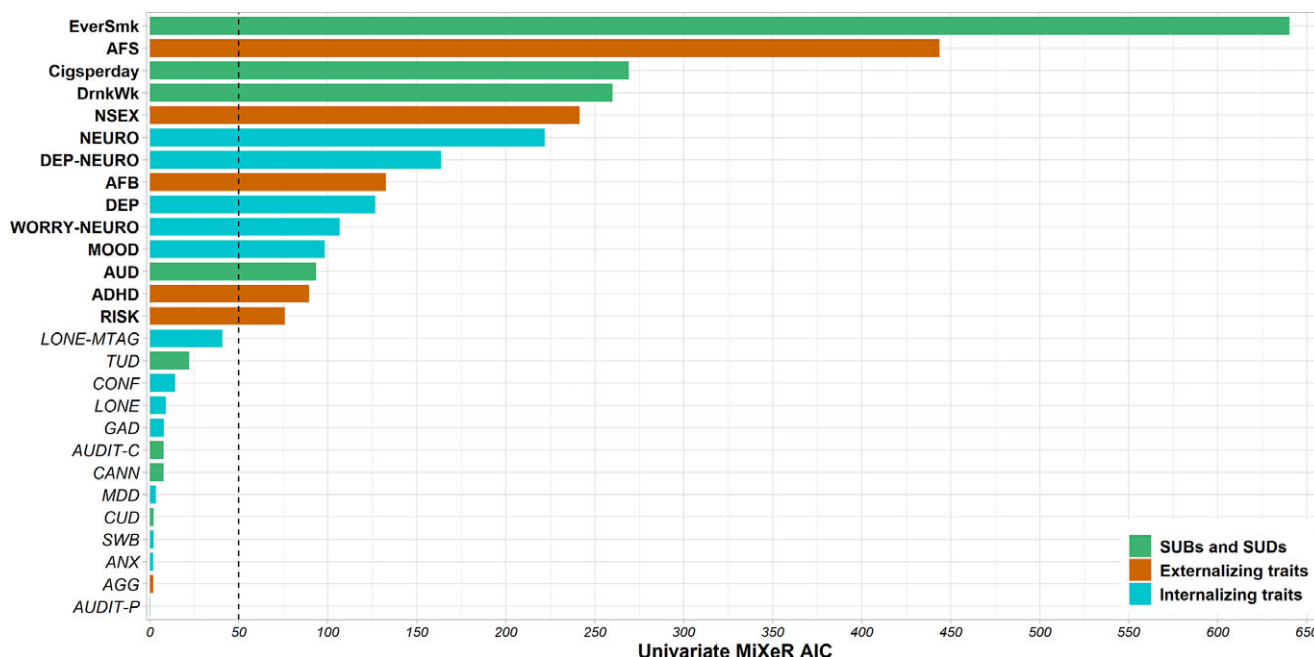

Figure 1. GWAS filtering according to their univariate MiXeR AIC. GWAS selected for MiXeR analyses, that is, those with a difference between the AIC calculated for the infinitesimal model and the AIC calculated for the causal mixture model higher than 50, are highlighted in bold. The dashed line represents univariate MiXeR AIC difference = 50. Traits are described in Table 1 and Supplementary Table 1. ADHD, attention deficit hyperactivity disorder; AFB, age at first birth; AFS, age at first sexual intercourse; AGG, childhood aggressive behavior; ANX, self-response to ever being diagnosed by anxiety, nerves or generalized anxiety disorder; AUD, alcohol use disorder; AUDIT-C, items related to alcohol consumption of the alcohol use disorders identification test; AUDIT-P, items related to problematic consequences of drinking of the alcohol use disorders identification test; CANN, lifetime cannabis use; Cigsperday, average number of cigarettes smoked per day; CONF, frequency of being able to confide in someone close; CUD, cannabis use disorder; DEP, lifetime diagnosis of major depressive disorder or broad self-reported depression; DEP-NEURO, depressed affect subcluster of neuroticsm; DrnkWk, average number of drinks drunk each week; EverSmk, ever been a regular smoker; GAD, generalized anxiety disorder; LONE, often feeling lonely; LONE-MTAG, combined multi-trait GWAS of loneliness and social isolation; MDD, major derepssior disorder; MOOD, experiencing mood swings; NEURO, neuroticism; NSEX, lifetime number of sexual partners; RISK, general risk tolerance; SWB, subjectrive well-being; TUD, tobacco use disorder; WORRY-NEURO, worry subcluster of neuroticism.

the genetic correlation of shared variants were estimated for each pair of traits. Full results can be found in Supplementary Table 3.

Bivariate MiXeR model fit
For almost all bivariate analyses, both AIC compared to minimum and maximum overlap were positive, indicating a good model fit (Supplementary Table 3). Minimum AIC was negative for AUD with NSEX and DEP, indicating that the overlap modeled by MiXeR was not distinguishable from the minimum overlap taking into account the genetic correlation between both traits. Maximum AIC was negative for some analyses for which there was a high overlap between both traits, indicating that the overlap estimated by MiXeR was not distinguishable from complete genetic overlap.

Pattern of genetic overlap between externalizing/internalizing traits and SUBs
The percentage of SUB causal variants shared with externalizing traits ranged from 52.0% to 95.9% (Figure 2 and Supplementary Figure 2). The genetic correlation of shared variants was moderate to high with almost all externalizing traits (absolute values of r_{gs} = 0.28–0.87), except between DrnkWk and AFB (absolute value of r_{gs} = 0.05) (Figure 3).

In the case of internalizing traits, genetic overlap with SUBs was high (Figure 2 and Supplementary Figure 2). EverSmk and DrnkWk shared more than 82% and 93% of their causal variants with internalizing traits, respectively (Figure 2). The genetic correlation of shared variants with internalizing traits was clearly lower than with externalizing traits (absolute values of r_{gs} = 0.02–0.35) (Figure 3).

Despite following the same general pattern of genetic overlap, some differences emerged between tobacco use behavior (EverSmk) and alcohol use behavior (DrnkWk). EverSmk exhibited a slightly higher genetic correlation of shared variants (absolute values of r_{gs} = 0.40–0.87) with externalizing traits compared to DrnkWk (absolute values of r_{gs} = 0.05–0.81) (Figure 3). Similarly, internalizing traits exhibited a higher genetic correlation of shared variants with EverSmk (absolute values of r_{gs} = 0.06–0.35) compared to DrnkWk (absolute values of r_{gs} = 0.02–0.11). Notably, DrnkWk showed a negative genetic correlation of shared variants, although close to 0, with all internalizing traits except for DEP, while EverSmk only displayed a negative genetic correlation of shared variants with WORRY-NEURO among internalizing traits.

Pattern of genetic overlap between externalizing/internalizing traits and SUDs/SUD-proxies
The percentage of causal variants for SUDs and SUD-proxies shared with externalizing traits ranged from 56.5% to 99.4% (Figure 2 and Supplementary Figure 2). The genetic correlation of shared variants for SUDs and SUD-proxies with externalizing traits was relatively high (absolute values of r_{gs} = 0.25–0.95) (Figure 3). In all pairs, a higher value for externalizing behavior was positively correlated with both SUBs and SUDs among the shared variants (Figure 3).

Genetic overlap with internalizing traits for SUDs and SUD-proxies (59.7%–97.4%) was more variable than for SUBs (82.0%–97.8%) (Figure 2 and Supplementary Figure 2). Cigsperday and AUD demonstrated a higher genetic correlation of shared variants with internalizing traits, except WORRY-NEURO, (absolute values of r_{gs} = 0.23–0.98) than SUBs (absolute values of r_{gs} = 0.02–0.35) (Figure 3). In contrast to the remaining internalizing traits, WORRY-NEURO showed a similar pattern of genetic overlap with both SUBs and SUDs. WORRY-NEURO shared a large number of variants (>95%) with all SUBs and SUDs, despite possessing low genetic correlation of shared variants (Figure 2 and Figure 3). The direction of the genetic correlation of shared variants was positive for pairs involving AUD and internalizing traits, contrary to
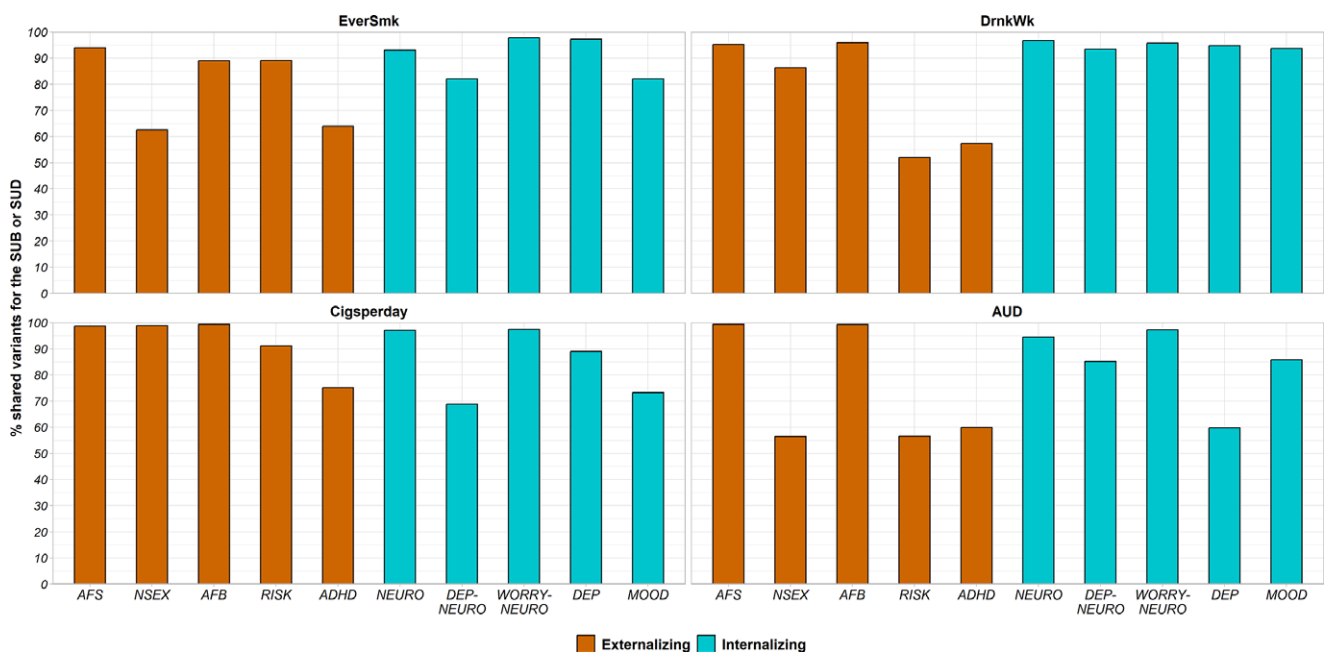

Figure 2. Percentage of variants of each SUB (EverSmk and DrnkWk) or SUD/SUD-proxy (Cigsperday and AUD) shared with each externalizing (colored in orange) or internalizing (colored in green) trait. Traits are described in Table 1. ADHD, attention deficit hyperactivity disorder; AFB, age at first birth; AFS, age at first sexual intercourse; AUD, alcohol use disorder; Cigsperday, average number of cigarettes smoked per day; DEP, lifetime diagnosis of major depressive disorder or broad self-reported depression; DEP-NEURO, depressed affect subcluster of neuroticism; DrnkWk, average number of drinks drunk each week; EverSmk, ever been a regular smoker; MOOD, experiencing mood swings; NEURO, neuroticism; NSEX, lifetime number of sexual partners; RISK, general risk tolerance; WORRY-NEURO, worry subcluster of neuroticism.
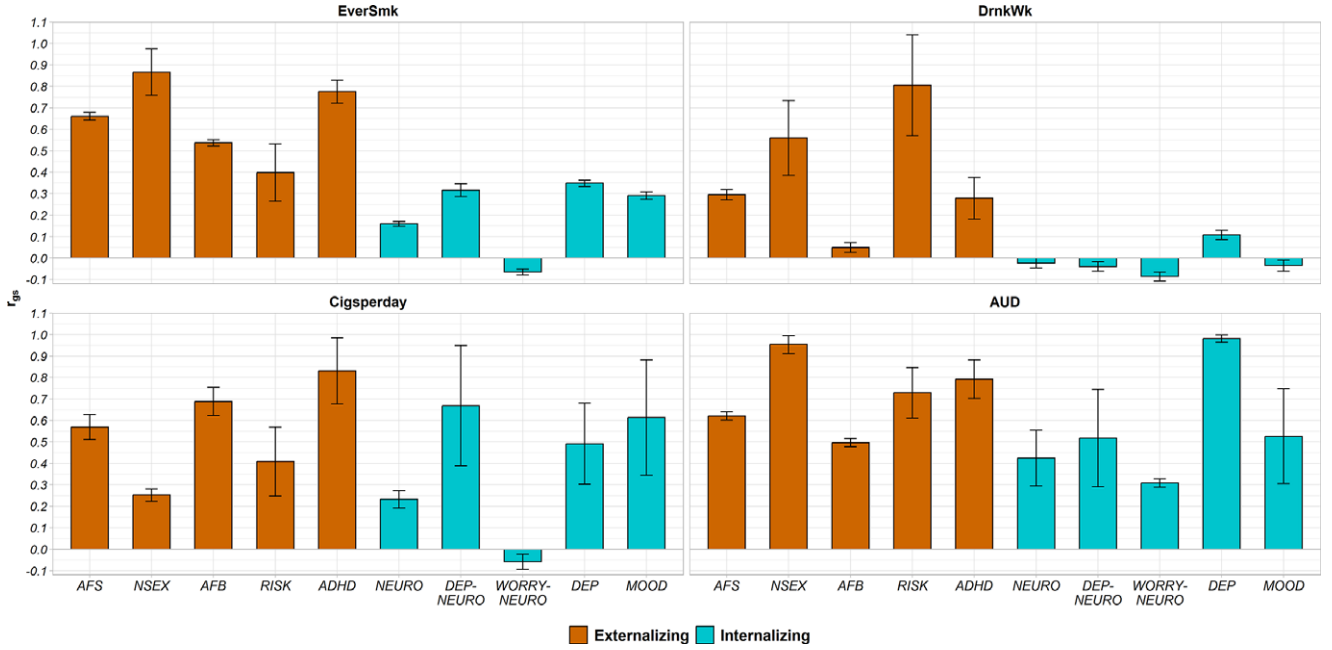
Figure 3. Genetic correlation of the shared variants (r_{gs}) between each SUB (EverSmk and DrnkWk) or SUD/SUD-proxy (Cigsperday and AUD) and each externalizing (colored in orange) or internalizing (colored in green) trait. Traits are described in Table 1. Error bars represent ±1 SD. The direction was reversed for AFB and AFS to indicate the genetic correlation with a higher externalizing behavior. ADHD, attention deficit hyperactivity disorder; AFB, age at first birth; AFS, age at first sexual intercourse; AUD, alcohol use disorder; Cigsperday, average number of cigarettes smoked per day; DEP, lifetime diagnosis of major depressive disorder or broad self-reported depression; DEP-NEURO, depressed affect subcluster of neuroticism; DrnkWk, average number of drinks drunk each week; EverSmk, ever been a regular smoker; MOOD, experiencing mood swings; NEURO, neuroticism; NSEX, lifetime number of sexual partners; RISK, general risk tolerance; WORRY-NEURO, worry subcluster of neuroticism.

DrnkWk (Figure 3). By contrast, the direction of the genetic correlation of shared variants with internalizing traits was positive for both tobacco traits.

Relationship between the genetic correlation of shared variants and the number of shared causal variants

An inverse relationship was observed between the genetic correlation of shared variants and the number of shared causal variants for each pair of traits, with a significant negative correlation between these variables (r = −0.66, P = 4.32x10^{-6}) (Figure 4). Those pairs of traits exhibiting a lower genetic correlation of shared variants than expected given the number of shared variants comprised predominantly SUBs and internalizing traits (small open triangles and squares in Figure 4). In contrast, pairs showing a genetic correlation of shared variants higher than expected based on the number of shared variants involved SUDs or SUD-proxies with externalizing traits (large filled triangles and squares in Figure 4).

Among the analyzed trait pairs, a subgroup exhibited a comparable degree of genetic overlap (>90% of shared causal variants) (Figure 4). Some pairs involving SUDs and SUD-proxies with externalizing traits demonstrated a relatively high genetic correlation of shared variants (absolute values of r_{gs} = 0.50–0.69), despite their high degree of genetic overlap, such as Cigsperday and AFB or AUD and AFS. Conversely, pairs with a similar degree of genetic overlap comprising SUDs or SUD-proxies with internalizing traits demonstrated a lower genetic correlation of shared variants (absolute values of r_{gs} = 0.06–0.42), such as Cigsperday and WORRY-NEURO or Cigsperday and NEURO. However, it was much lower for SUBs, notably DrnkWk, and internalizing traits (absolute values of r_{gs} = 0.02–0.16, except for EverSmk and DEP (absolute value of r_{gs} = 0.35)).

Discussion

Using MiXeR, we detected a distinct pattern of genetic overlap between SUBs and SUDs with externalizing and internalizing traits. SUBs (i.e. EverSmk and DrnkWk) and internalizing traits demonstrated a notably high genetic overlap but a particularly low genetic correlation of shared variants. This finding may be explained by a divergence in the effect distribution of the numerous shared risk variants between SUBs and internalizing traits, suggesting a limited role in the genetics of internalizing problems at SUBs. By contrast, SUBs showed a higher genetic correlation of shared variants with externalizing traits, suggesting that externalizing problems may have a stronger genetic influence on SUBs, albeit with variations depending on the substance. With regards to SUDs or SUD-proxies (i.e. Cigsperday and AUD), we observed a notably high genetic correlation of shared variants with externalizing traits, also in cases of high genetic overlap. This observation suggests that most risk variants for SUDs and SUD-proxies are shared with externalizing problems, and these variants show a similar effect distribution on both traits. A similar pattern, although not as marked, was found for pairs involving SUDs and SUD-proxies with internalizing problems. These findings suggest that while both externalizing and internalizing pathways may be involved in SUDs, externalizing pathway seems to play a more consistent role.

Bivariate MiXeR revealed a substantial genetic overlap between SUBs and SUDs and both externalizing and internalizing traits. A similar pattern was already identified in a previous study from our group between mental disorders and other traits such as intelligence, educational attainment, neuroticism, and SWB (Hindley et al., 2022). This study concluded that a large number of pleiotropic variants are involved in all these traits with differences in the effect size distribution of these pleiotropic variants playing a more
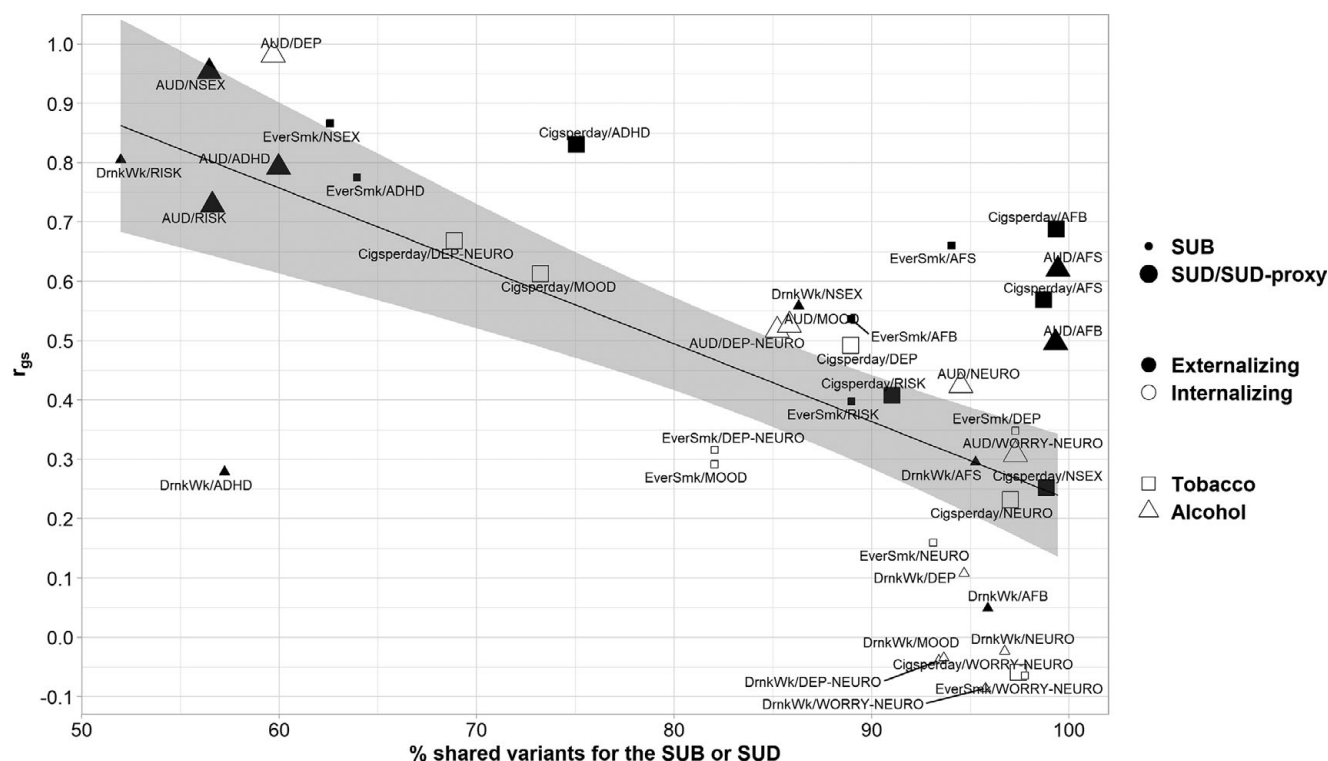
Figure 4. Scatter plot illustrating the negative relationship between the genetic correlation of shared variants (r_{gs}) and the percentage of shared variants across analyzed trait pairs. The solid line represents the linear regression, while the shaded region depicts the 95% confidence interval. On the x-axis, the percentage of shared variants between each SUB or SUD and each externalizing or internalizing trait with respect to the SUB/SUD is displayed. The y-axis shows the genetic correlation of shared variants between each SUB or SUD and each externalizing or internalizing trait. For pairs involving AFB and AFS, the direction of the genetic correlation of shared variants was reversed to indicate the genetic correlation with a higher externalizing behavior. Pairs involving SUBs (EverSmk and DrnkWk) are denoted by smaller dots, while pairs involving SUDs or SUD-proxies (Cigsperday and AUD) are represented by larger dots. Pairs involving externalizing or internalizing traits are colored in black or white, respectively. Pairs involving tobacco use behaviors or disorders (EverSmk and Cigsperday) are depicted by squares and pairs involving alcohol use behaviors or disorders (DrnkWk and AUD) are represented by triangles. ADHD, attention deficit hyperactivity disorder; AFB, age at first birth; AFS, age at first sexual intercourse; AUD, alcohol use disorder; Cigsperday, average number of cigarettes smoked per day; DEP, lifetime diagnosis of major depressive disorder or broad self-reported depression; DEP-NEURO, depressed affect subcluster of neuroticism; DrnkWk, average number of drinks drunk each week; EverSmk, ever been a regular smoker; MOOD, experiencing mood swings; NEURO, neuroticism; NSEX, lifetime number of sexual partners; RISK, general risk tolerance; SUB, substance use behaviors; SUD, substance use disorders, WORRY-NEURO, worry subcluster of neuroticism.

prominent role in the genetic predisposition for a given trait than trait-specific genetic variants. Considering our results, this conclusion can also be extended to the genetic overlap of externalizing and internalizing traits with SUBs and SUDs. Our study identified two distinct patterns within this general framework. SUBs and internalizing traits showed a high genetic overlap but a genetic correlation of shared variants close to zero, suggesting a distinctly divergent effect distribution within both traits. Conversely, certain pairs involving SUDs or SUD-proxies with externalizing traits displayed a similar distribution of effects as well as a high genetic overlap, suggesting a shared genetic basis among these traits.

Our results are consistent with previous studies suggesting that, while the internalizing pathway plays a limited role at the initial stages of substance consumption, it has a stronger impact at more severe levels of use (Hussong et al., 2011; Ning et al., 2020). A different genetic relationship of internalizing traits with SUBs and SUDs has already been suggested using MiXeR (Icick et al., 2022). In this previous study, mood instability showed a different pattern of genetic overlap with alcohol consumption, defined as AUDIT-C, and AUD. Here, we have expanded on these findings by using several internalizing traits in addition to mood instability and a different alcohol consumption trait (DrnkWk). Our findings are also consistent with previous studies suggesting a role of the externalizing pathway at all substance use stages. An effect of the

externalizing pathway on substance use initiation was suggested in a previous study which found that the shared genetic susceptibility between schizophrenia and smoking initiation is related to externalizing behavior (Al-Soufi & Costas, 2023). Importantly, our results emphasize the substantial role of the externalizing pathway in SUDs, consistent with previous studies supporting a clearer association of externalizing problems with severe substance use compared to substance consumption and internalizing problems (Edwards, Gardner, Hickman, & Kendler, 2016; Ning et al., 2020). Previous studies have implicated the genetic susceptibility to both externalizing and internalizing problems in SUDs. A recent study found that an addiction risk factor based on problematic alcohol use, problematic tobacco use, CUD, and opioid use disorder GWAS correlated genetically with externalizing behavior (absolute $r_g = 0.35$–0.52), as well as with substance use to cope with anxiety ($r_g = 0.63$) and depression ($r_g = 0.60$), suggesting a strong association of externalizing and internalizing behavior with SUDs (Hatoum et al., 2023). Another study based on a similar addiction risk factor found that the association between this addiction risk factor and risk-taking significantly decreased after taking substance use into account (Hatoum et al., 2021). In contrast, genetic susceptibility to substance use had a small effect on the association between this addiction risk factor and neuroticism. Our findings are compatible with these results suggesting that, in contrast to the

internalizing pathway, the effect of the externalizing pathway on SUDs may be partially mediated by substance use.

For SUDs and SUD-proxies, both tobacco and alcohol showed comparable patterns of genetic overlap with externalizing and internalizing traits. However, differences emerged between the tobacco (EverSmk) and alcohol (DrnkWk) use behaviors. EverSmk displayed a higher genetic correlation of shared variants with externalizing and internalizing traits compared to DrnkWk. Moreover, the genetic correlation of shared variants of internalizing traits with DrnkWk was close to 0, suggesting that the internalizing pathway may have a minimal effect on alcohol use. In contrast to tobacco use, DrnkWk is a highly heterogenous trait, combining the frequency of use and quantity per drinking occasion (Mallard et al., 2022; Sanchez-Roige et al., 2021). These dimensions show opposite genetic correlations with traits related to socioeconomic status, psychiatric disorders, and psychological traits (Marees et al., 2019). As a consequence of this heterogeneity, DrnkWk is expected to be less influenced by the internalizing and externalizing pathways than use of tobacco or other substances.

The present study has some limitations. First, we cannot conclude that there is a causal effect of externalizing and internalizing problems on SUBs and SUDs on the basis of MiXeR results alone. We can only assess whether there is a shared genetic susceptibility between both traits and its effect distribution on both traits. Second, some of the MiXeR models were not distinguishable from a complete genetic overlap or the minimum genetic overlap allowed for a given genetic correlation. Therefore, larger GWAS based on a more precise phenotype definition is required to achieve more accurate estimates of the number of shared and unique genetic variants. Third, a lack of well-powered GWAS for other substances only allowed the analysis of substance use for two legal substances: alcohol and tobacco. Thus, our conclusions may not generalize to other substances. Fourth, phenotype definition remains a major topic in psychiatric genetics. A continuous trait for maladaptive alcohol use would assess more accurately the level of substance dependence of an individual. Adding data regarding the course of SUD (age at onset, continuous versus discontinuous use, duration of SUD, speed, and age of transition between use and SUD) may be relevant in refining the genetic component of SUD. This limitation of the dichotomous approach is extensive to other traits used in this study. Furthermore, differences in the procedures used to ascertain cases may be associated with measurement bias. In general, there is a need for more comprehensive assessments of GWAS samples regarding traits of relevance in psychiatry. Unfortunately, most of the phenotypic refinements susceptible to improved transfer to biology or the clinics occur at the expense of sample size, thus reducing statistical power. Fifth, there is considerable comorbidity between externalizing and internalizing behavior, which makes it difficult to distinguish the separate effects of each analyzed pathway. Moreover, many GWAS rely on samples like UK Biobank, which may not accurately represent the general population in terms of gender, socio-economic status, or ancestry, potentially biasing our results (Fry et al., 2017). In addition, since MiXeR assumes that the effect size of causal variants is normally distributed, the cluster of nAChR genes at chromosome 15 was excluded from our Cigsperday GWAS. MiXeR also assumes an additive model of genetic effects, while gene–environment interactions, epistasis, and dominance effects may play an important role, particularly in personality traits such as neuroticism (Boomsma et al., 2018; Keller, Coventry, Heath, & Martin, 2005). Violation of the MiXeR's assumption of additivity may therefore contribute to the weak genetic correlations with neuroticism. Furthermore, we did not provide statistical significance for our results, as formal significance tests based on MiXeR results are not feasible. Finally, although we used several sexual behaviors as representative traits of genetic susceptibility for externalizing behavior, these complex phenotypes are affected by gene–environment correlations attributable to social, economic, and political processes (Abdellaoui, Dolan, Verweij, & Nivard, 2022). Therefore, a cautionary remainder is needed to avoid misinterpretation of genomic findings that may lead to stigmatization.

In summary, we found a different genetic relationship between SUBs and SUDs with externalizing and internalizing traits, suggesting a larger involvement of the externalizing and internalizing pathways in severe substance use. In addition, the externalizing pathway seems to play a more prominent role throughout the progression of substance use than the internalizing pathway. These findings improve our understanding of the origins of substance use and SUDs, which may inform attempts to advance prevention and treatment strategies.

References

Abdellaoui, A., Dolan, C. V., Verweij, K. J. H., & Nivard, M. G. (2022). Gene–environment correlations across geographic regions affect genome-wide association studies. *Nature Genetics*, **54**(9), 1345–1354. https://doi.org/10.1038/s41588-022-01158-0

Abdellaoui, A., Smit, D. J. A., van den Brink, W., Denys, D., & Verweij, K. J. H. (2021). Genomic relationships across psychiatric disorders including substance use disorders. *Drug and Alcohol Dependence*, **220**, 108535. https://doi.org/10.1016/j.drugalcdep.2021.108535

Al-Soufi, L., & Costas, J. (2023). Genetic susceptibility for schizophrenia after adjustment by genetic susceptibility for smoking: Implications in identification of risk genes and genetic correlation with related traits. *Psychological Medicine*, 1–11. https://doi.org/10.1017/S0033291723000326

Bartels, M., Cacioppo, J. T., Van Beijsterveldt, T. C. E. M., & Boomsma, D. I. (2013). Exploring the association between well-being and psychopathology in adolescents. *Behavior Genetics*, **43**(3), 177–190. https://doi.org/10.1007/s10519-013-9589-7

Baselmans, B., Hammerschlag, A. R., Noordijk, S., Ip, H., van der Zee, M., de Geus, E., … van 't Ent, D. (2021). The genetic and neural substrates of externalizing behavior. *Biological Psychiatry Global Open Science*, 1–11. https://doi.org/10.1016/j.bpsgos.2021.09.007

Boomsma, D. I., Helmer, Q., Nieuwboer, H. A., Hottenga, J. J., de Moor, M. H., van den Berg, S. M., … de Geus, E. J. (2018). An extended twin-pedigree study of neuroticism in the Netherlands Twin Register. *Behavior Genetics*, **48**(1), 1–11. https://doi.org/10.1007/s10519-017-9872-0

Bulik-Sullivan, B., Finucane, H. K., Anttila, V., Gusev, A., Day, F. R., Loh, P.-R., … Neale, B. M. (2015). An atlas of genetic correlations across human diseases and traits. *Nature Genetics*, **47**(11), 1236–1241. https://doi.org/10.1038/ng.3406

Chang, L.-H., Whitfield, J. B., Liu, M., Medland, S. E., Hickie, I. B., Martin, N. G., … Gillespie, N. A. (2019). Associations between polygenic risk for tobacco and alcohol use and liability to tobacco and alcohol use, and psychiatric disorders in an independent sample of 13,999 Australian adults. *Drug and Alcohol Dependence*, **205**, 107704. https://doi.org/10.1016/j.drugalcdep.2019.107704

Chen, L. G., Tubbs, J. D., Liu, Z., Thach, T. Q., & Sham, P. C. (2024, June 1). Mendelian randomization: Causal inference leveraging genetic data. *Psychological Medicine*, **54**, 1461–1474. https://doi.org/10.1017/S0033291724000321

Day, F. R., Ong, K. K., & Perry, J. R. B. (2018). Elucidating the genetic basis of social interaction and isolation. *Nature Communications*, **9**(1), 1–6. https://doi.org/10.1038/s41467-018-04930-1

Deak, J. D., & Johnson, E. C. (2021). Genetics of substance use disorders: A review. *Psychological Medicine*, **51**, 2189–2200. https://doi.org/10.1017/S0033291722000629

Degenhardt, L., Charlson, F., Ferrari, A., Santomauro, D., Erskine, H., Mantilla-Herrara, A., … Vos, T. (2018). The global burden of disease attributable to alcohol and drug use in 195 countries and territories, 1990–2016: A systematic analysis for the Global Burden of Disease Study 2016. *The Lancet Psychiatry*, **5**(12), 987–1012. https://doi.org/10.1016/S2215-0366(18)30337-7

Demontis, D., Walters, G. B., Athanasiadis, G., Walters, R., Therrien, K., Nielsen, T. T., … Børglum, A. D. (2023). Genome-wide analyses of ADHD identify 27 risk loci, refine the genetic architecture and implicate several cognitive domains. *Nature Genetics*, **55**(2), 198–208. https://doi.org/10.1038/s41588-022-01285-8

Edwards, A. C., Gardner, C. O., Hickman, M., & Kendler, K. S. (2016). A prospective longitudinal model predicting early adult alcohol problems: Evidence for a robust externalizing pathway. *Psychological Medicine*, **46**(5), 957–968. https://doi.org/10.1017/S0033291715002457

Facal, F., Flórez, G., Blanco, V., Rodríguez, J., Pereiro, C., Fernández, J. M., … the GenPol Study Group. (2021). Genetic predisposition to alcohol dependence: The combined role of polygenic risk to general psychopathology and to high alcohol consumption. *Drug and Alcohol Dependence*, **221**, 108556. https://doi.org/10.1016/j.drugalcdep.2021.108556

Frei, O., Holland, D., Smeland, O. B., Shadrin, A. A., Fan, C. C., Maeland, S., … Dale, A. M. (2019). Bivariate causal mixture model quantifies polygenic overlap between complex traits beyond genetic correlation. *Nature Communications*, **10**(1), 1–11. https://doi.org/10.1038/s41467-019-10310-0

Fry, A., Littlejohns, T. J., Sudlow, C., Doherty, N., Adamska, L., Sprosen, T., … Allen, N. E. (2017). Comparison of Sociodemographic and Health-Related Characteristics of UK Biobank Participants with Those of the General Population. *American Journal of Epidemiology*, **186**(9), 1026–1034. https://doi.org/10.1093/aje/kwx246

Gelernter, J., & Polimanti, R. (2021). Genetics of substance use disorders in the era of big data. *Nature Reviews Genetics*, **22**(11), 712–729. https://doi.org/10.1038/s41576-021-00377-1

Griffith, J. W., Zinbarg, R. E., Craske, M. G., Mineka, S., Rose, R. D., Waters, A. M., & Sutton, J. M. (2010). Neuroticism as a common dimension in the internalizing disorders. *Psychological Medicine*, **40**(7), 1125–1136. https://doi.org/10.1017/S0033291709991449

Hatoum, A. S., Colbert, S. M. C., Johnson, E. C., Huggett, S. B., Deak, J. D., Pathak, G. A., … Agrawal, A. (2023). Multivariate genome-wide association meta-analysis of over 1 million subjects identifies loci underlying multiple substance use disorders. *Nature Mental Health*, **1**(3), 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., … Agrawal, A. (2021). The addiction risk factor: A unitary genetic vulnerability characterizes substance use disorders and their associations with common correlates. *Neuropsychopharmacology*, (September), 1–7. https://doi.org/10.1038/s41386-021-01209-w

Heitzeg, M. M., Hardee, J. E., & Beltz, A. M. (2018). Sex differences in the developmental neuroscience of adolescent substance use risk. *Current Opinion in Behavioral Sciences*, **23**, 21–26. https://doi.org/10.1016/j.cobeha.2018.01.020

Hindley, G., Frei, O., Shadrin, A. A., Cheng, W., O'Connell, K. S., Icick, R., … Andreassen, O. A. (2022). Charting the landscape of genetic overlap between mental disorders and related traits beyond genetic correlation. *American Journal of Psychiatry*, **179**(11), 833–843. https://doi.org/10.1176/appi.ajp.21101051

Holland, D., Frei, O., Desikan, R., Fan, C. C., Shadrin, A. A., Smeland, O. B., … Dale, A. M. (2020). Beyond SNP heritability: Polygenicity and discoverability of phenotypes estimated with a univariate Gaussian mixture model. *PLoS Genetics*, **16**(5), 1–30. https://doi.org/10.1371/journal.pgen.1008612

Howard, D. M., Adams, M. J., Clarke, T.-K., Hafferty, J. D., Gibson, J., Shirali, M., … McIntosh, A. M. (2019). Genome-wide meta-analysis of depression identifies 102 independent variants and highlights the importance of the prefrontal brain regions. *Nature Neuroscience*, **22**(3), 343–352. https://doi.org/10.1038/s41593-018-0326-7

Hussong, A. M., Jones, D. J., Stein, G. L., Baucom, D. H., & Boeding, S. (2011). An internalizing pathway to alcohol use and disorder. *Psychology of Addictive Behaviors*, **25**(3), 390–404. https://doi.org/10.1037/a0024519

Icick, R., Shadrin, A., Holen, B., Karadag, N., Lin, A., Hindley, G., … Andreassen, O. A. (2022). Genetic overlap between mood instability and alcohol-related phenotypes suggests shared biological underpinnings. *Neuropsychopharmacology*, **47**(11), 1883–1891. https://doi.org/10.1038/s41386-022-01401-6

Icick, R., Shadrin, A., Holen, B., Karadag, N., Parker, N., O'Connell, K. S., … Andreassen, O. A. (2023). *Identification of novel loci and cross-disorder pleiotropy through multi-ancestry genome-wide analysis of alcohol use disorder in over one million individuals. PREPRINT (Version 1) available at Research Square.* https://doi.org/10.21203/rs.3.rs-3755915/v1

Ip, H. F., van der Laan, C. M., Krapohl, E. M. L., Brikell, I., Sánchez-Mora, C., Nolte, I. M., … Boomsma, D. I. (2021). Genetic association study of childhood aggression across raters, instruments, and age. *Translational Psychiatry*, **11**(1). https://doi.org/10.1038/s41398-021-01480-x

Johnson, E. C., Demontis, D., Thorgeirsson, T. E., Walters, R. K., Polimanti, R., Hatoum, A. S., … Børglum, A. D. (2020). A large-scale genome-wide association study meta-analysis of cannabis use disorder. *The Lancet Psychiatry*, **7**(12), 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., … Wagner, G. G. (2019). Genome-wide association analyses of risk tolerance and risky behaviors in over 1 million individuals identify hundreds of loci and shared genetic influences. *Nature Genetics*, **51**(2), 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., … Dick, D. M. (2021). Multivariate analysis of 1.5 million people identifies genetic associations with traits related to self-regulation and addiction. *Nature Neuroscience*, **24**(10), 1367–1376. https://doi.org/10.1038/s41593-021-00908-3

Keller, M. C., Coventry, W. L., Heath, A. C., & Martin, N. G. (2005). Widespread evidence for non-additive genetic variation in Cloninger's and Eysenck's personality dimensions using a twin plus sibling design. *Behavior Genetics*, **35**(6), 707–721. https://doi.org/10.1007/s10519-005-6041-7

Kim, S., Kim, K., Hwang, M. Y., Ko, H., Jung, S. H., Shim, I., … Won, H. H. (2022). Shared genetic architectures of subjective well-being in East Asian and European ancestry populations. *Nature Human Behaviour*, **6**(7), 1014–1026. https://doi.org/10.1038/s41562-022-01343-5

Krueger, R. F., Hicks, B. M., Patrick, C. J., Carlson, S. R., Iacono, W. G., & McGue, M. (2002). Etiologic connections among substance dependence, antisocial behavior, and personality: Modeling the externalizing spectrum.

Journal of Abnormal Psychology, **111**(3), 411–424. https://doi.org/10.1037/0021-843X.111.3.411

Leppert, B., Millard, L. A. C., Riglin, L., Davey Smith, G., Thapar, A., Tilling, K., … Stergiakouli, E. (2020). A cross-disorder PRS-pheWAS of 5 major psychiatric disorders in UK Biobank. *PLOS Genetics*, **16**(5), e1008185. https://doi.org/10.1371/journal.pgen.1008185

Levey, D. F., Gelernter, J., Polimanti, R., Zhou, H., Cheng, Z., Aslan, M., … Stein, M. B. (2020). Reproducible Genetic Risk Loci for Anxiety: Results From ~200,000 Participants in the Million Veteran Program. *American Journal of Psychiatry*, **177**(3), 223–232. https://doi.org/10.1176/appi.ajp.2019.19030256

Liu, M., Jiang, Y., Wedow, R., Li, Y., Brazel, D. M., Chen, F., … Vrieze, S. (2019). Association studies of up to 1.2 million individuals yield new insights into the genetic etiology of tobacco and alcohol use. *Nature Genetics*, **51**(2), 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., … Sanchez-Roige, S. (2022). Item-level genome-wide association study of the alcohol use disorders identification test in three population-based cohorts. *The American Journal of Psychiatry*, **179**(1), 58–70. https://doi.org/10.1176/appi.ajp.2020.20091390

Marees, A. T., Smit, D. J. A., Ong, J. S., Macgregor, S., An, J., Denys, D., … Derks, E. M. (2019). Potential influence of socioeconomic status on genetic correlations between alcohol consumption measures and mental health. *Psychological Medicine*. https://doi.org/10.1017/S0033291719000357

Meque, I., Dachew, B. A., Maravilla, J. C., Salom, C., & Alati, R. (2019). Externalizing and internalizing symptoms in childhood and adolescence and the risk of alcohol use disorders in young adulthood: A meta-analysis of longitudinal studies. *Australian and New Zealand Journal of Psychiatry*, **53**(10), 965–975. https://doi.org/10.1177/0004867419844308

Mills, M. C., Tropf, F. C., Brazel, D. M., van Zuydam, N., Vaez, A., Agbessi, M., … Day, F. R. (2021). Identification of 371 genetic variants for age at first sex and birth linked to externalising behaviour. *Nature Human Behaviour*, **5**(12), 1717–1730. https://doi.org/10.1038/s41562-021-01135-3

Nagel, M., Jansen, P. R., Stringer, S., Watanabe, K., De Leeuw, C. A., Bryois, J., … Posthuma, D. (2018). Meta-analysis of genome-wide association studies for neuroticism in 449,484 individuals identifies novel genetic loci and pathways. *Nature Genetics*, **50**(7), 920–927. https://doi.org/10.1038/s41588-018-0151-7

Nagel, M., Watanabe, K., Stringer, S., Posthuma, D., & Van Der Sluis, S. (2018). Item-level analyses reveal genetic heterogeneity in neuroticism. *Nature Communications*, **9**(1). https://doi.org/10.1038/s41467-018-03242-8

Ning, K., Gondek, D., Patalay, P., & Ploubidis, G. B. (2020). The association between early life mental health and alcohol use behaviours in adulthood: A systematic review. *PLoS ONE*, **15**(2), 1–28. https://doi.org/10.1371/journal.pone.0228667

Pasman, J. A., Verweij, K. J. H., Gerring, Z., Stringer, S., Sanchez-Roige, S., Treur, J. L., … 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. *Nature Neuroscience*, **21**(9), 1161–1170. https://doi.org/10.1038/s41593-018-0206-1

Polimanti, R., Peterson, R. E., Ong, J.-S., MacGregor, S., Edwards, A. C., Clarke, T.-K., … Derks, E. M. (2019). Evidence of causal effect of major depression on alcohol dependence: Findings from the psychiatric genomics consortium. *Psychological Medicine*, **49**(07), 1218–1226. https://doi.org/10.1017/S0033291719000667

Quach, B. C., Bray, M. J., Gaddis, N. C., Liu, M., Palviainen, T., Minica, C. C., … Hancock, D. B. (2020). Expanding the genetic architecture of nicotine dependence and its shared genetics with multiple traits. *Nature Communications*, **11**(1). https://doi.org/10.1038/s41467-020-19265-z

Sallis, H. M., Davey Smith, G., & Munafò, M. R. (2019). Cigarette smoking and personality: Interrogating causality using Mendelian randomisation. *Psychological Medicine*, **49**(13), 2197–2205. https://doi.org/10.1017/S0033291718003069

Sanchez-Roige, S., Cox, N. J., Johnson, E. O., Hancock, D. B., & Davis, L. K. (2021). Alcohol and cigarette smoking consumption as genetic proxies for alcohol misuse and nicotine dependence. *Drug and Alcohol Dependence*, **221**, 108612. https://doi.org/10.1016/j.drugalcdep.2021.108612

Sanchez-Roige, S., Palmer, A. A., Fontanillas, P., Elson, S. L., Adams, M. J., Howard, D. M., … Wilson, C. H. (2019). Genome-wide association study meta-analysis of the alcohol use disorders identification test (AUDIT) in two population-based cohorts. *American Journal of Psychiatry*, **176**(2), 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., … Vrieze, S. (2022). Genetic diversity fuels gene discovery for tobacco and alcohol use. *Nature*, **612**(7941), 720–724. https://doi.org/10.1038/s41586-022-05477-4

Soler Artigas, M., Sánchez-Mora, C., Rovira, P., Richarte, V., Garcia-Martínez, I., Pagerols, M., … Ribasés, M. (2020). Attention-deficit/hyperactivity disorder and lifetime cannabis use: Genetic overlap and causality. *Molecular Psychiatry*, **25**(10), 2493–2503. https://doi.org/10.1038/s41380-018-0339-3

Treur, J. L., Demontis, D., Smith, G. D., Sallis, H., Richardson, T. G., Wiers, R. W., … Munafò, M. R. (2019). Investigating causality between liability to ADHD and substance use, and liability to substance use and ADHD risk, using Mendelian randomization. *Addiction Biology*, **26**(1), e12849. https://doi.org/10.1111/adb.12849

Vilar-Ribó, L., Sánchez-Mora, C., Rovira, P., Richarte, V., Corrales, M., Fadeuilhe, C., … Soler Artigas, M. (2020). Genetic overlap and causality between substance use disorder and attention-deficit and hyperactivity disorder. *American Journal of Medical Genetics Part B: Neuropsychiatric Genetics*. https://doi.org/10.1002/ajmg.b.32827

Vink, J. M., Treur, J. L., Pasman, J. A., & Schellekens, A. (2020). Investigating genetic correlation and causality between nicotine dependence and ADHD in a broader psychiatric context. *American Journal of Medical Genetics Part B: Neuropsychiatric Genetics*. https://doi.org/10.1002/ajmg.b.32822

Walters, R. K., Polimanti, R., Johnson, E. C., McClintick, J. N., Adams, M. J., Adkins, A. E., … Agrawal, A. (2018). Transancestral GWAS of alcohol dependence reveals common genetic underpinnings with psychiatric disorders. *Nature Neuroscience*, **21**(12), 1656–1669. https://doi.org/10.1038/s41593-018-0275-1

Wootton, R. E., Richmond, R. C., Stuijfzand, B. G., Lawn, R. B., Sallis, H. M., Taylor, G. M. J., … Munafò, M. R. (2020). Evidence for causal effects of lifetime smoking on risk for depression and schizophrenia: A Mendelian randomisation study. *Psychological Medicine*, **50**(14), 2435–2443. https://doi.org/10.1017/S0033291719002678

Wray, N. R., Ripke, S., Mattheisen, M., Trzaskowski, M., Byrne, E. M., Abdellaoui, A., … Sullivan, P. F. (2018). Genome-wide association analyses identify 44 risk variants and refine the genetic architecture of major depression. *Nature Genetics*, **50**(5), 668–681. https://doi.org/10.1038/s41588-018-0090-3

Yao, Y., Xu, Y., Cai, Z., Liu, Q., Ma, Y., Li, A. N., … Li, M. D. (2021). Determination of shared genetic etiology and possible causal relations between tobacco smoking and depression. *Psychological Medicine*, **51**(11), 1870–1879. https://doi.org/10.1017/S003329172000063X