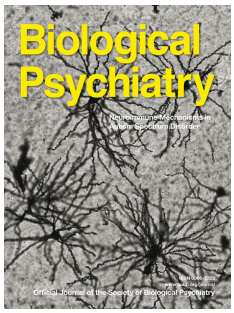
Journal Pre-proof



Novel biological insights into the common heritable liability to substance involvement: a multivariate genome-wide association study

Tabea Schoeler, PhD, Jessie Baldwin, PhD, Andrea Allegrini, PhD, Wikus Barkhuizen, PhD, Andrew McQuillin, PhD, Nicola Pirastu, PhD, Zoltán Kutalik, PhD, Jean-Baptiste Pingault, PhD

PII: S0006-3223(22)01483-4

DOI: https://doi.org/10.1016/j.biopsych.2022.07.027

Reference: BPS 14950

To appear in: *Biological Psychiatry*

Received Date: 4 March 2022

Revised Date: 19 June 2022

Accepted Date: 11 July 2022

Please cite this article as: Schoeler T., Baldwin J., Allegrini A., Barkhuizen W., McQuillin A., Pirastu N., Kutalik Z. & Pingault J.-B., Novel biological insights into the common heritable liability to substance involvement: a multivariate genome-wide association study, *Biological Psychiatry* (2022), doi: https://doi.org/10.1016/j.biopsych.2022.07.027.

Novel biological insights into the common heritable liability to substance involvement: a multivariate genome-wide association study

Tabea Schoeler[1,2*], PhD
Jessie Baldwin[2,3], PhD
Andrea Allegrini[2], PhD
Wikus Barkhuizen[2], PhD
Andrew McQuillin[4], PhD
Nicola Pirastu[5], PhD
Zoltán Kutalik[1,6,7], PhD
Jean-Baptiste Pingault[2,3], PhD

[1] Department of Computational Biology, University of Lausanne, Switzerland
[2] Department of Clinical, Educational and Health Psychology, Division of Psychology and Language Sciences, University College London, London, UK
[3] Social, Genetic and Developmental Psychiatry Centre, Institute of Psychiatry, Psychology and Neuroscience, King's College London, London, UK
[4] Division of Psychiatry, University College London, London, UK
[5] Centre for Global Health Research, Usher Institute, University of Edinburgh, UK
[6] University Center for Primary Care and Public Health, University of Lausanne, Switzerland
[7] Swiss Institute of Bioinformatics, Lausanne, Switzerland

*Correspondence: t.schoeler@ucl.ac.uk

Short title: the shared genetic liability to substance involvement
Key words: Aetiology, substance use, addiction, genome-wide, genetic epidemiology, common liability

Abstract

Background

Consumption of nicotine, alcohol and cannabis commonly co-occurs, which is thought to partly stem from a common heritable liability to substance involvement.

Methods

To elucidate its genetic architecture, we modelled a common liability, inferred from genetic correlations among six measures of dependence and frequency of use of nicotine, alcohol and cannabis.

Results

Forty-two genetic variants were identified in the multivariate genome-wide association study on the common liability to substance involvement, of which 67% were novel and not associated with the six phenotypes. Mapped genes highlighted the role of dopamine (e.g., dopamine D2 gene), and showed enrichment for several components of the central nervous systems (e.g., mesocorticolimbic brain regions) and molecular pathways (dopaminergic, glutamatergic, GABAergic) that are thought to modulate drug reinforcement. Genetic correlations with other traits were most prominent for reward-related behaviours (e.g., risk-taking, cocaine and opioid use) and mood (e.g., depression, insomnia).

Conclusions

These genome-wide results triangulate and expand previous preclinical and human studies focusing on the neurobiological substrates of substance involvement, and help to elucidate the genetic architecture underlying the use of common psychoactive substances.

Background

Substance use constitutes a significant burden for public health(1), and considerable research efforts are made to better understand its aetiology. Substance involvement (i.e., regular use, problematic use/dependence) is typically not restricted to just one class of substance, as dependence of accessible psychoactive substances such as nicotine, alcohol and cannabis often co-occurs(2). This co-occurring pattern of use has been shown to be particularly detrimental to the individual and to society as a whole(3).

Aetiological models posit that the use of multiple substances stems from a common liability to substance involvement(4,5) – a latent continuous trait accounting for the shared risk of developing dependence to different substances. Based on findings from genomic(6–8) and behavioural genetic studies(9,10), it is assumed that this common liability includes a genetic component. Indeed, genetic correlations between use of different classes of psychoactive substances are substantial, as estimated in twin (up to r_g~0.89(11–13)) and genome-wide association (GWA) studies (up to r_g~0.70(7,14–16)). The underlying molecular mechanisms of this common heritable liability to substance involvement are, however, not fully understood. While it has been shown that genotypic variations contribute to the common heritable liability to substance use(17), an investigation into the specific genome-wide effects has yet to be conducted. Furthermore, although increasingly large GWA studies have identified growing numbers of genetic risk variants associated with individual substance use phenotypes(7), it remains unclear as to whether associated risk variants reflect shared (common) or non-shared (substance specific) risk across different phenotypes indexing substance involvement. Some identified genetic variants likely operate through substance-specific pharmacological pathways, as is the case for variants affecting

nicotinic receptors (e.g. genes coding for nicotinic acetylcholine receptors, such as *CHRNA3-CHRNA5*(18)) or alcohol metabolism (e.g. variants in the alcohol dehydrogenases gene family, such as *ADH1B*(7,15), *ADH1C*(14,15)). Other variants may affect common pathways, such as variants associating with two or more classes of psychoactive substances (e.g. *BDNF*(7,19,20), *PDE4B*(7,20,21) or *DRD2*(7,15,18,20,22)) or behavioural phenotypes (e.g., the top variant identified for cannabis use disorder, which also associates with ADHD and risk-taking(23)).

The aforementioned genetic overlap complicates research on causes and consequences of substance use, and distilling shared (common) from non-shared (substance specific) genetic risk is pivotal to the interpretation of genome-wide discoveries. One way of scrutinizing putative pleiotropic variants is to explicitly model the genetic overlap among different phenotypes indexing substance involvement, using multivariate methods such as genomic structural equation modelling (genomic SEM(24)). Applying this method has already been helpful in characterising shared genetic influences across dimensions of psychopathology(24–31) and cognition(32–35). In addition to assessing shared effects of suspected pleiotropic variants, genomic SEM also has the potential to identify novel genetic variants not previously identified in univariate GWA studies on individual phenotypes(29). This is expected, since shared risk is thought to be expressed indirectly via the common liability, resulting in inherently small effects, which hampers detection of pleiotropic variants in univariate GWA analyses. A multivariate GWA can therefore boost discovery of shared variants directly associated with a common heritable liability. While genetically informed methods using polygenic scores have already explored risk factors involved in the common liability to substance involvement(36,37), a multivariate GWA of the shared and non-shared genetic architecture can further deepen our understanding of biological pathways underlying the use of

multiple substances. Indeed, leveraging genetically informed methods would allow us to revisit long-theorized biological pathways underlying dependence, and to triangulate evidence from behavioural genetic(11), brain imaging(38) and preclinical studies(39) focusing on the role of genetics and neural substrates of substance use. Together, such triangulated findings would help researchers and clinicians to better understand biological and developmental pathways involved in risk of initiating and developing dependence to commonly used and abused psychoactive substances.

To unravel the genetic architecture underlying consumption of nicotine, alcohol and cannabis, here we conduct a multivariate GWA analysis on their common heritable liability. To model the common liability, we include phenotypes indexing clinical (diagnosis of dependence) as well as quantitative (frequency of use) measures of use of nicotine, alcohol and cannabis. More specifically, we conduct a multivariate GWA of the common heritable liability, with the aim to

a. identify putative genetic variants associated with the common liability (i.e., shared/pleiotropic variants) and variants specific to the use of different classes of substances (i.e., non-shared)
b. characterize the functional features of genetic variants associated with the common liability
c. assess the genetic correlations between the common liability with other complex traits
d. evaluate the validity of the causal claims imposed by a common liability model of substance involvement

Methods

GWA summary datasets

We screened GWA summary statistics of addiction-related phenotypes for the most commonly used and misused psychoactive substances, namely nicotine, alcohol and cannabis. For each substance class, we included one clinical (diagnosis of dependence) and one quantitative (frequency of use) measure. The following summary statistics were included, derived from samples of individuals of European ancestry: Alcohol use disorder (n=28,757)(14), cannabis use disorder (n=358,534)(23), nicotine dependence (n=244,890)(40), frequency of cigarette (n=245,876)(7), alcohol (n=513,208)(7) and cannabis use (n=24,798(41)). Additional details of each of the included summary statistic files can be found in the Supplement and in sTable 1 (Supplement).

Genomic model of the common heritable liability to substance involvement

We first estimated the genetic correlations (r_g) among the individual phenotypes using genomic structural equation modelling (genomic SEM(24)) version 0.0.3. The method uses an extension of LD-score regression(42) and accounts for sample overlap across studies through the LD-score intercept. In confirmatory factor analysis, we fitted a series of structural equation models. Given our aim to estimate SNP effects that underly the involvement with multiple common psychoactive substances, we started with a single latent factor model onto which the six substance use indicators loaded (Figure 1b). This model has been shown to be a powerful model for SNP discovery(24) and has been proposed by previous studies modelling genetic correlations between phenotypes indexing liability to addiction(17). We tested a number of competing single factor

models and estimated fit indices (i.e., common liability model with and without factor loading constraints; with and without correlated residuals). In addition, we assessed if the model can be improved by the use of a bi-factor model, a hierarchical model and a two-factor model (shown in sFigure 1). The Diagonally Weighted Least Squares (DWLS) estimator was used and model fit was assessed based on the Comparative Fit Index (CFI), the standardized root mean square residual (SRMR) and the Akaike Information Criterion (AIC), an index that balances fit with parsimony.

Multivariate genome-wide association analysis

For the multivariate GWA on the common heritable liability to substance involvement, the summary statistics for the individual substance use phenotypes were restricted to single nucleotide polymorphisms (SNPs) contained in the 1000 genomes phase 3 reference sample (with a minor allele frequency (MAF) > 1%) and SNPs that were present in all GWA summary datasets included in the analysis. Genomic control was applied to all summary statistics showing evidence of uncontrolled confounding (LD score intercept > 1), by multiplying standard errors by the LD score intercept. To identify lead SNPs after conducting the GWA on the common liability, we selected LD-independent SNPs (r^2<0.1 within 250 kb) based on genome-wide significance (p<5x10^{-8}).

To determine whether the effects of the identified lead SNPs are likely to act through the common liability, we applied the heterogeneity test as implemented in Genomic SEM. The resulting Q-statistic (Q_{SNP}) is a χ^2distributed test statistic, with significant Q_{SNP} estimates (p<5x10^{-8}) indicating that the SNP effect does not act entirely through the common liability. Based on Q_{SNP}, we selected only SNPs that did not show evidence of heterogeneity (Q_{SNP} $p\geq5\times10^{-8}$) before conducting

functional follow-up analyses of SNPs associated with the common liability. Positional mapping and expression quantitative trait loci (eQTL) mapping were used to map lead SNPs to genes. PhenoScanner(43) was used to explore previously identified associations of lead SNPs (cf. Supplement for further details).

Pathway enrichment analysis of genetic variants associated with the common liability

To identify the most likely biological pathways underlying the common heritable liability to substance involvement, we used Data-driven Expression-Prioritized Integration for Complex Traits (DEPICT(44)) and Pathway SCoring ALgorithm (PASCAL)(45). DEPCIT was used to test for tissue/cell type enrichment of a set of LD-independent SNPs (r^2<0.05 within 500 kb) outside genome-wide significance ($p<5\times10^{-5}$). PASCAL was used to test for enrichment of all SNPs, using three gene sets (BIOCARTA, KEGG, REACTOME) curated by the Molecular Signatures Database (MSigDB(46)) and gene sets defined by DEPICT. Prior to running the analyses, the GWA on the common liability was filtered according to the Q_{SNP} statistic, retaining only SNPs operating through the common liability (Q_{SNP} $p\geq5\times10^{-8}$). Results were corrected for multiple testing using false discovery rate (FDR) correction (controlled at 5%). Further details regarding the application of the two methods can be found in the Supplement (sMethods).

Genetic correlations between the common liability and other complex traits

Bivariate LD score regression analyses were performed in Genomic SEM, to estimate the genetic correlations between SNPs operating through the common liability (i.e., SNPs with Q_{SNP} $p\geq5\times10^{-8}$) and 35 other traits related to physical features (e.g., height, body mass index), personality (e.g.,

risk-taking, neuroticism), social variables (e.g., socioeconomic status, education) and mental health (e.g., schizophrenia, depression). A complete list of the included GWA summary statistics can be found in **sTable 1** (Supplement). FDR correction (controlled at 5%) was used to adjust for multiple testing.

Evaluation of the causal claims implied by the common liability theory

Mendelian Randomization (MR) analysis was used to evaluate key causal claims imposed by the common liability theory, including that the common liability has direct effects on all its indicators (cf. Supplement for further details on model assumptions). Inverse variance weighted (IVW) MR implemented in TwoSampleMR(47) package was applied to all analyses. The genetic markers instrumenting the common liability were selected based on genome-wide significance ($p<5\times10^{-8}$) and Q_{snp}, retaining only SNPs that operated through the common liability (Q_{snp} $p\geq5\times10^{-8}$). To facilitate comparability of the MR estimates, the beta estimates for the included SNPs were standardized by dividing the z-scores by the square root of the sample size before conducting MR (cf. Supplement).

Results

Genomic model of the common heritable liability to substance involvement

The correlations among the individual cigarette, alcohol and cannabis use phenotypes are presented in the heatmap in **Figure 1a** (cf. Supplement, **sTable 2** for estimates). Genetic correlations varied widely between the individual substance phenotypes, ranging from r_g=-0.01 to r_g=0.74 (mean r_g=0.40; SD=0.21). From the eight assessed structural models (cf. **sFigure 1**, Supplement), we proceeded with the single factor model as depicted in **Figure 1b**, which showed evidence of a good model fit (CFI=0.97, SRMR=0.07, AIC=46.8, cf. **sTable 3**). In this model, equality constrains were imposed on paths belonging to the same pattern of substance use, i.e., equal weights across measures of dependence, and equal weights across measures of frequency of use. Correlated residuals were included to allow for within-substance class associations. Both an unconstrained common liability model with correlated residuals (CFI=0.97, SRMR=0.05, AIC=47.7, **sTable 3**) and the bifactor model showed similar fit to the data (CFI=0.97, SRMR=0.05, AIC=47.7, **sTable 3**). However, we decided to retain the constrained model, as (i) it more parsimoniously reflected the data among the tested models, evident by the lowest AIC and (ii) constraining loadings across substances ensured that the common factor reflected a common liability rather than predominantly one substance above others.

For the selected common factor model, all standardized factor loadings are presented in **Figure 1b**, showing that the constrained loadings were estimated to be 0.39 for frequency measures of substance use and 0.79 for substance dependence measures. On average, the common factor accounted for 38.81% (range 15.21%-62.41%) of the genetic variance in the six substance use phenotypes.

Genetic variants associated with the common heritable liability

6,500,152 SNPs were included in the GWA of the common liability. Since our SNP estimates were derived from overlapping samples, we used the formula developed by Mallard et al.(48) to derive the effective samples size (cf. sMethods, Supplement), which was estimated to be N=187,062. The Manhattan and Q-Q (quantile-quantile) plot of the common liability GWA are shown in **Figure 1c** and **sFigure 2** (Supplement), respectively. The main results for all GWA analyses, including the multivariate analysis on the common liability and the six univariate analyses on the individual substance use phenotypes are summarised in **sTable 5-7** and **sFigure 2** (Supplement). In brief, the GWA on the common liability identified 3,509 genome-wide ($p<5\times10^{-8}$) SNPs, tagging 55 LD-independent SNPs. After removing SNPs showing significant heterogeneity (Q_{SNP} $p<5\times10^{-8}$), 42 SNPs operating through the common liability remained (cf. SNPs highlighted in blue in the Manhattan plot, **Figure 1c**). Of the 42 SNPs, 28 (66.67%) were novel, i.e., have not been associated with any of the individual substance use phenotypes. Positional mapping showed that the top five SNPs (rs10750025, rs4953149, rs281287, rs202665, rs35023999) operating through the common liability lay mostly outside coding regions, located close to *DRD2*, *LINC01833, SEMA6D, SCUBE1* and *ANKK1*, respectively. Further inspection through eQTL mapping indicated that the aforementioned SNPs acted as eQTLs for positionally mapped genes, highlighting their putative role in the common liability via gene expression (cf. **sTable 8**, Supplement). A search in the PhenoScanner database(43) indicated that the five lead SNPs operating through the common liability have previously been linked to a number of behavioural phenotypes, such as neuroticism, irritability, smoking status or time spent in front of the computer (cf. **sTable 9**, Supplement). Of note, 13 of the 55 SNPs associated with the common liability still showed heterogeneous effects

across the individual substance use phenotypes (Q_{SNP} $p<5\times10^{-8}$, highlighted in red in the Manhattan plot, **Figure 1c**). Those SNPs can be considered as false discoveries, which may result from a single or a subset of SNPs with large effects on the individual substance use phenotypes(35). Among all 6,500,152 SNPs included in the common liability GWA, 2,356 (0.04%) showed heterogeneous effects.

For comparison, we also evaluated the GWA results of the individual substance use phenotypes, focusing on significant variants ($p<5\times10^{-8}$) from the original GWA studies that showed heterogenous effects (Q_{SNP} $p<5\times10^{-8}$). The results are summarized in **Figure 2** and further discussed in the Supplement. In brief, a number of variants appeared to be specific with respect to the class of substance, as found for alcohol [e.g., rs1229984, a variant on the alcohol dehydrogenase 1B gene (*ADH1B*)] and cigarette use (e.g., rs76474922 and rs58379124, variants located on the nicotinic receptor genes *CHRNA5 and CHRNB3,* respectively). Only two SNPs were associated with cannabis use phenotypes, of which one variant (rs7783012, FOXP2) appeared to operate via the common liability (Q_{SNP} $p\geq5\times10^{-8}$).

Pathway enrichment analyses of genes associated with the common heritable liability

Testing for tissue and cell type enrichment in DEPICT revealed 22 pathways associated (with FDR controlled at 5%) with the common liability (**Figure 3A**), which were all part of central nervous system tissues. In PASCAL (**Figure 3B**), 481 pathways were significantly (FDR controlled at 5%) enriched for the common liability, of which the top pathways related to broader categories of neurotransmitter functioning (e.g., neural system, transmission across chemical synapses). Overall, the pattern of regional enrichment and neuronal signalling pathways highlighted the role

of a widespread network of brain areas involved in the common liability, in line with theories suggesting that genetic risk to addiction is not solely the manifestation of altered limbic reward processes(49). Since the highlighted pathways were most prominently enriched for the common liability, and less so for individual substance use phenotypes, the aforementioned brain circuits may tap into somewhat distinct features characterizing the common liability. A more detailed discussion on the pattern of enrichment is provided in the Supplement and estimates obtained from DEPCIT and PASCAL are included in **sTable 11-12** (Supplement).

Genetic correlations between the common liability and other complex traits

Using the input from the Q_{SNP}-filtered GWA of the common liability and the GWA summary statistics for 41 traits (cf. **sTable 1** in Supplement for details), we found significant correlations with 36 complex traits after correction for multiple testing (**Figure 4**). As expected, the largest positive correlations were present between the common liability and its "constituents", i.e., the cigarette, alcohol and cannabis use phenotypes used to derive the common liability (mean r_g=0.68). Among the other traits, the largest genetic correlations were present for cocaine and opioid dependence (both r_g=0.60), number of sexual partners (r_g=0.49), ADHD (r_g=0.48) and risk tolerance (r_g=0.41). Moderate genetic correlations were also present for a number of traits relating to mood, including insomnia (r_g=0.35) and depression (r_g=0.34). No significant (FDR controlled at 5%) associations were found with birth weight, openness, obsessive-compulsive disorder, anorexia and cortical surface area.

Evaluation of the causal relationships implied by a common liability model

Figure 5 displays the results from Mendelian Randomization (MR) analyses, assessing paths running from the common liability to the individual substance use phenotypes. Using 42 Q_{SNP}-filtered LD-independent SNPs from the common liability GWA, the MR findings provide support for a causal interpretation of the initial descriptive common liability model (cf. Figure 1b) – that is, the common liability increases the risk of use and dependence of nicotine, alcohol and cannabis. More specifically, the loadings obtained from the genomic factor model of the common heritable were recovered using genetic markers instrumenting the common liability. As shown, the standardized causal effects obtained in MR were comparable to the factor loadings of the indicators (highlighted in red in Figure 5), as evident for measures of dependence [mean MR estimate: 0.79 (0.10 SD)] and measures of frequency of substance use [mean MR estimate: 0.33 (0.12 SD)]. The hypothesized structural model also asserts absence of causal effects between indicators belonging to a different class of substance (e.g., cannabis dependence → alcohol dependence), which was in line with our MR results. Finally, reverse causation (effects of the specific substance use indicators on the common liability to substance involvement) was indicated for three of the indicators. Further discussion on the interpretation of reverse causation in this context is included in the Supplement, together with the full set of MR results (cf. sTable 13).

Conclusions

To dissect shared from non-shared genetic liability to use and dependence of nicotine, alcohol and cannabis, we conducted a multivariate genome-wide association (GWA) study of a common heritable liability to substance involvement. The modelled liability constituted primarily a common liability to problematic substance use, as measures of dependence dominated in this model. The top genetic variant operating through the common liability (rs10750025, located on the dopamine receptor D2 [*DRD2*] gene) provides support for the role dopamine in risk of addiction. Functional follow-up of common liability-associated genes further highlighted the role of widespread neuronal signalling pathways and neurotransmitter functioning beyond dopamine, such as GABAergic and glutamatergic pathways. Brain areas implicated in the common liability to substance involvement spanned limbic and cortical areas involved in reward, motivation, memory and cognitive control. The genetic overlap between the common liability and other complex traits was most prominent for other measures of addiction (e.g., cocaine and opioid use), as well as impulsive behaviours (e.g., risk-taking, ADHD) and mood (e.g., depression, insomnia). For cigarette and alcohol use, risk genes not operating via the common liability translated into specific pharmacogenomic pathways, such as nicotinic acetylcholine receptor functioning. Distinct pathways for cannabis use were, however, not identified.

Shared and non-shared genetic risks involved in risk of nicotine, alcohol and cannabis involvement

In line with existing evidence, we found substantial genetic correlations between measures of cigarette, alcohol and cannabis use. This allowed us to model the common heritable liability to substance involvement, which explained substantial variance (average=39%) in genetic liabilities

to individual substance use phenotypes. Since more variance was explained for measures indexing dependence (62%) than measures of frequency of use (15%), the common liability captures mostly a problematic substance use pattern (cf. Supplement for further discussion on this finding).

DRD2 was identified as the lead gene operating via the common liability – a pathway believed to be a common mechanism by which addictive substances exert their acute pleasurable effects. *DRD2* in particular is a frequently studied gene implicated in addictive behaviours, given its central role in modulating the dopamine reward system that mediates the reinforcing effects of addictive substances. Indeed, *DRD2* has been identified in numerous genome-wide studies on cigarette(7,8), alcohol(7,8,20), cannabis use(8) and problematic substance use(50). Other notable genes linked to the common liability are further discussed in the Supplement.

Our results also highlight the role of neural signalling pathways involved in the common heritable liability to substance involvement, particularly synaptic functioning and a range of neurotransmitter systems beyond dopamine (GABA, glutamate, serotonin). Indeed, while dopaminergic mechanisms have been the traditional focus in addiction research, a growing body of research is now assessing the role of wider-ranging and interconnected neurotransmission systems in addiction vulnerability, involving GABAergic, glutamatergic and serotoninergic projections that contribute to modulating reward reinforcement and drug-seeking behaviour(51–53). In line with this, enrichment analysis implicated the central nervous system and a network of brain areas in the common liability, including circuits involved in the processing of information related to reward (limbic structures), motivation (basal ganglia), memory (hippocampus) and cognitive control (frontal lobe areas). Since the discussed pathways were most prominently related to the common liability, rather than the individual substance use liabilities, the identified

pathways may reflect common neural substrates characterizing addiction vulnerability (cf. Supplement for a more detailed discussion on substance-specific risk).

Finally, it is assumed that the brain reward pathways partly link to addictive behaviours via some intermediate complex behaviours, such as risk-taking, sensation seeking or impulsivity. While this remains to be formally tested, this idea corroborates with our findings of genetic correlations between the common liability and maladaptive behaviours, including ADHD, risk-taking and cocaine and opioid dependence. We also found large genetic correlations between the common liability with internalising symptoms, including depression and insomnia. This is in line with recent evidence of a link between a polygenic index for addiction and a number of traits indexing impulsivity and externalizing behaviours in drug-naïve children(50).

Implications for the aetiology of substance involvement

Bi-directional MR was used to evaluate key causal claims imposed by the common liability, namely (1) the common liability has direct effects on all its indicators (i.e., the individual substance use phenotypes) and (2) there is no reverse causation (a third claim typically made by strict latent factor is further discussed in the Supplement).

Overall, MR findings provided support for assumption (1), as causality ran from the common liability to all of the individual substance use phenotypes. While effects in the reverse direction were also present for three indicators – at odds with assumption (2) – this may reflect unaccounted pleiotropy (cf. Supplement for further discussion on this point). Together, since the common liability was identified as a shared cause of all substance use phenotypes, the results suggest that a common liability to substance involvement may usefully capture most relationships

between substances. Embracing such conceptualization would have important implications for intervention. First, targeting modifiable features of the common liability should reduce risk of addiction to nicotine, alcohol and cannabis. For example, pharmacological treatments targeting dopamine, glutamate and GABA function may reduce craving and the euphoric/rewarding responses to cigarettes, alcohol and cannabis(54–59). Second, interventional targeting of only one specific class of substance (e.g., nicotine) unlikely leads to reductions in use of another class (e.g., alcohol). This conclusion is somewhat inconsistent with previous evidence in rodents showing reductions in alcohol use following the administration nicotinic treatments (e.g. varenicline(60,61)), although evidence from RCTs in humans is mixed(62,63) and efficacy may not translate into the long-term(64). Finally, while our MR analyses included overlapping sample, risk of bias is limited when using partly overlapping samples as shown in recent work(65).

Future directions

To maximize power for genome-wide discovery, the common liability to substance involvement modelled in this work included only commonly used substances (i.e., nicotine, alcohol and cannabis) that are typically available as quantitative and clinical measures in large-scale genotyped samples. As such, our model including one common liability to substance involvement may not generalize to broader liability models of addiction. An important goal for future work would thus be to extend this multivariate analysis to a broader array of addiction phenotypes (e.g., initiation, tolerance, craving, withdrawal, relapse) characterizing different classes of psychoactive substances(e.g., cocaine, opioid use), once well-powered GWA data becomes available. Such efforts will allow exploration of more fine-grained structural models. Furthermore, multivariate

approaches as employed here are not just important in terms of GWA discovery, but also essential to reducing biases; in our study, a substantial proportion of GWA-significant SNPs associated with the individual substance use phenotypes appeared to be mediated by the common heritable liability. As such, future GWA studies powerful enough to detect small genetic effects will likely tag an increasing number of SNPs with horizontal pleiotropic effects when examining addiction phenotypes (i.e., direct effects on several phenotypes). As such, modelling heritable latent factors as done in this study, and/or accounting for its contribution as recently proposed(66,67) is therefore paramount when using genetically informed causal inference methods that are sensitive to the presence of heritable confounding.

Taken together, our results confirm that a common heritable liability partially explains the high co-occurrence of use and abuse of nicotine, alcohol, and cannabis. Functions of the implicated genes converged on broad central nervous system pathways beyond the dopaminergic pathways long-hypothesised in risk of addiction.

Data access

Summary statistics of the common liability GWA analysis are accessible on GWAS catalog (https://www.ebi.ac.uk/gwas/) at the accession number XXXX. References to all publicly available summary statistic files included in this work are listed in sTable 1.

Code availability

The code used to conduct the analyses presented in this work is available on GitHub (https://github.com/TabeaSchoeler/TS2021_CommonLiabAddiction)

URLs

MsigDB data (https://www.gsea-msigdb.org/gsea/index.jsp)

PASCAL (https://www2.unil.ch/cbg/index.php?title=Pascal)

G:Profiler (https://biit.cs.ut.ee/gprofiler/page/r/)

PhenoScanner (https://github.com/phenoscanner/phenoscanner)

TwoSampleMR (https://mrcieu.github.io/TwoSampleMR/)

GenomicSEM (https://github.com/MichelNivard/GenomicSEM)

DEPICT (https://data.broadinstitute.org/mpg/depict/)

Neale Lab UKBB summary statistics (http://www.nealelab.is/uk-biobank/)

GWAS ATLAS (https://atlas.ctglab.nl/)

Acknowledgements

This study would not have possible without the use of publicly available genome-wide summary data and software tools. The authors gratefully acknowledge these resources, and thank the research participants, the research teams and institutions that have contributed to this research. We are grateful to Dr. Judit Cabana-Domínguez and Prof. Bru Cormand for providing the results from the cocaine dependence genome-wide association study.

A previous version of this article was published as a preprint on medRxiv:

https://doi.org/10.1101/2021.11.19.21266548

Funding

T.S. is funded by a Wellcome Trust Sir Henry Wellcome fellowship (grant 218641/Z/19/Z). J.R.B is funded by a Wellcome Trust Sir Henry Wellcome fellowship (grant 215917/Z/19/Z). A.M. is supported by the UCLH NIHR BRC. Z.K. is funded by the Swiss National Science Foundation (# 310030-189147). JB.P. has received funding from the European Research Council (ERC) under the European Union's Horizon 2020 research and innovation programme (grant agreement No. 863981) and is supported by the Medical Research Foundation 2018 Emerging Leaders 1st Prize in Adolescent Mental Health (MRF-160-0002-ELP-PINGA). The funding organisations listed were not involved in the design and conduct of the study; collection, management, analysis, and interpretation of the data; preparation, review, or approval of the manuscript; or decision to submit the manuscript for publication.

Competing interests

All authors report no biomedical financial interests or potential conflicts of interest.

References

1. Degenhardt L, Charlson F, Ferrari A, Santomauro D, Erskine H, Mantilla-Herrara A, *et al.* (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: 987–1012.

2. Palmer RHC, Button TM, Rhee SH, Corley RP, Young SE, Stallings MC, *et al.* (2012): Genetic etiology of the common liability to drug dependence: Evidence of common and specific mechanisms for DSM-IV dependence symptoms. *Drug Alcohol Depend* 123: S24–S32.

3. Connor JP, Gullo MJ, White A, Kelly AB (2014): Polysubstance use - diagnostic challenges, patterns of use and health. *Curr Opin Psychiatry* 27: 269–275.

4. Vanyukov MM, Tarter RE, Kirillova GP, Kirisci L, Reynolds MD, Kreek MJ, *et al.* (2012): Common liability to addiction and "gateway hypothesis": Theoretical, empirical and evolutionary perspective. *Drug Alcohol Depend* 123: 3–17.

5. Vanyukov MM, Tarter RE, Kirisci L, Kirillova GP, Maher BS, Clark DB (2003): Liability to substance use disorders: 1. Common mechanisms and manifestations. *Neurosci Biobehav Rev* 27: 507–515.

6. Nivard M, Verweij K, Minică C, Treur J, Cannabis Consortium I, Vink J, Boomsma D (2016): Connecting the dots, genome-wide association studies in substance use. *Mol Psychiatry* 21: 733–735.

7. Liu M, Jiang Y, Wedow R, Li Y, Brazel DM, Chen F, *et al.* (2019): Association studies of up to 1.2 million individuals yield new insights into the genetic etiology of tobacco and alcohol use. *Nat Genet* 51: 237–244.

8. Jang S-K, Saunders G, Liu M, Jiang Y, Liu DJ, Vrieze S (2020): Genetic correlation, pleiotropy, and causal associations between substance use and psychiatric disorder. *Psychol Med* 1–11.

9. Kendler KS, Schmitt E, Aggen SH, Prescott CA (2008): Genetic and environmental influences on alcohol, caffeine, cannabis, and nicotine use from early adolescence to middle adulthood. *Arch Gen Psychiatry* 65: 674–682.

10. Ystrom E, Reichborn-Kjennerud T, Neale MC, Kendler KS (2014): Genetic and environmental risk factors for illicit substance use and use disorders: joint analysis of self and co-twin ratings. *Behav Genet* 44: 1–13.

11. Swan GE, Carmelli D, Cardon LR (1996): The consumption of tobacco, alcohol, and coffee in caucasian male twins: A multivariate genetic analysis. *J Subst Abuse* 8: 19–31.

12. Agrawal A, Grant JD, Lynskey MT, Madden PAF, Heath AC, Bucholz KK, Sartor CE (2016): The genetic relationship between cannabis and tobacco cigarette use in European- and African-American female twins and siblings. *Drug Alcohol Depend* 163: 165–171.

13. True WR, Xian H, Scherrer JF, Madden PAF, Bucholz KK, Heath AC, *et al.* (1999): Common genetic vulnerability for nicotine and alcohol dependence in men. *Arch Gen Psychiatry* 56: 655.

14. Walters RK, Polimanti R, Johnson EC, McClintick JN, Adams MJ, Adkins AE, *et al.* (2018): Transancestral GWAS of alcohol dependence reveals common genetic underpinnings with psychiatric disorders. *Nat Neurosci* 21: 1656–1669.

15. Zhou H, Sealock JM, Sanchez-Roige S, Clarke T-K, Levey DF, Cheng Z, *et al.* (2020): 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.

16. Quach BC, Bray MJ, Gaddis NC, Liu M, Palviainen T, Minica CC, *et al.* (2020): Expanding the genetic architecture of nicotine dependence and its shared genetics with multiple traits. *Nat Commun* 11: 5562.

17. Hatoum AS, Johnson EC, Colbert SMC, Polimanti R, Zhou H, Walters RK, *et al.* (2021): The addiction risk factor: A unitary genetic vulnerability characterizes substance use disorders and their associations with common correlates. *Neuropsychopharmacology* 2021.01.26.21250498.

18. Quach BC, Bray MJ, Gaddis NC, Liu M, Palviainen T, Minica CC, *et al.* (2020): Expanding the

Genetic Architecture of Nicotine Dependence and its Shared Genetics with Multiple Traits: Findings from the Nicotine Dependence GenOmics (iNDiGO) Consortium. *bioRxiv* 2020.01.15.898858.

19. Warnault V, Darcq E, Morisot N, Phamluong K, Wilbrecht L, Massa SM, *et al.* (2016): The BDNF Valine 68 to Methionine Polymorphism increases compulsive alcohol drinking in mice that Is reversed by tropomyosin receptor kinase B activation. *Biol Psychiatry* 79: 463–473.
20. Evangelou E, Gao H, Chu C, Ntritsos G, Blakeley P, Butts AR, *et al.* (2019): New alcohol-related genes suggest shared genetic mechanisms with neuropsychiatric disorders. *Nat Hum Behav*. https://doi.org/10.1038/s41562-019-0653-z
21. Brazel DM, Jiang Y, Hughey JM, Turcot V, Zhan X, Gong J, *et al.* (2019): Exome Chip meta-analysis fine maps causal variants and elucidates the genetic architecture of rare coding variants in smoking and alcohol use. *Biol Psychiatry* 85: 946–955.
22. Henderson JA, Buchwald DS, Howard B V, Henderson PN, Li Y, Tyndale RF, *et al.* (2020): Genetics of smoking behavior in American Indians. *Cancer Epidemiol Biomarkers Prev* cebp.0026.2020.
23. Johnson EC, Demontis D, Thorgeirsson TE, Walters RK, Polimanti R, Hatoum AS, *et al.* (2020): A large-scale genome-wide association study meta-analysis of cannabis use disorder. *The Lancet Psychiatry* 7: 1032–1045.
24. Grotzinger AD, Rhemtulla M, de Vlaming R, Ritchie SJ, Mallard TT, Hill WD, *et al.* (2019): Genomic structural equation modelling provides insights into the multivariate genetic architecture of complex traits. *Nat Hum Behav* 3: 513–525.
25. Luningham JM, Poore HE, Yang J, Waldman ID (2018): Testing structural models of

psychopathology at the genomic level. *bioRxiv* 502039.

26. Lee PH, Anttila V, Won H, Feng Y-CA, Rosenthal J, Zhu Z, *et al.* (2019): Genomic Relationships, Novel Loci, and Pleiotropic Mechanisms across Eight Psychiatric Disorders. *Cell* 179: 1469-1482.e11.

27. Mallard TT, Linnér RK, Grotzinger AD, Sanchez-Roige S, Seidlitz J, Okbay A, *et al.* (2020): Multivariate GWAS of psychiatric disorders and their cardinal symptoms reveal two dimensions of cross-cutting genetic liabilities. *bioRxiv* 603134.

28. Thorp JG, Campos AI, Grotzinger AD, Gerring Z, An J, Ong J-S, *et al.* (2020): Symptom-level genetic modelling identifies novel risk loci and unravels the shared genetic architecture of anxiety and depression. *medRxiv* 2020.04.08.20057653.

29. Grotzinger AD, Mallard TT, Akingbuwa WA, Ip HF, Adams MJ, Lewis CM, *et al.* (2020): Genetic Architecture of 11 Major Psychiatric Disorders at Biobehavioral, Functional Genomic, and Molecular Genetic Levels of Analysis. *medRxiv* 2020.09.22.20196089.

30. Karlsson Linner R, Mallard TT, Barr PB, Sanchez-Roige S, Madole JW, Driver MN, *et al.* (2020): Multivariate genomic analysis of 1.5 million people identifies genes related to addiction, antisocial behavior, and health. *bioRxiv* 2020.10.16.342501.

31. Peyre H, Schoeler T, Liu C, Williams CM, Hoertel N, Havdahl A, Pingault J (2021): Combining multivariate genomic approaches to elucidate the comorbidity between autism spectrum disorder and attention deficit hyperactivity disorder. *J Child Psychol Psychiatry* jcpp.13479.

32. Davies G, De la Fuente, Javier, Grotzinger AD, Tucker-Drob EM, Deary IJ (n.d.): Genetic "General Intelligence," Objectively Determined and Measured. *bioRxiv*.

33. Demange PA, Malanchini M, Mallard TT, Biroli P, Cox SR, Grotzinger AD, *et al.* (2020):

Investigating the Genetic Architecture of Non-Cognitive Skills Using GWAS-by-Subtraction. *bioRxiv* 2020.01.14.905794.

34. Hatoum AS, Morrison CL, Mitchell EC, Lam M, Reineberg AE, Palmer RHC, *et al.* (2020): GWAS of Over 427,000 Individuals Establishes GABAergic and Synaptic Molecular Pathways as Key for Cognitive Executive Functions. *bioRxiv* 674515.
35. de la Fuente J, Davies G, Grotzinger AD, Tucker-Drob EM, Deary IJ (2020): A general dimension of genetic sharing across diverse cognitive traits inferred from molecular data. *Nat Hum Behav*. https://doi.org/10.1038/s41562-020-00936-2
36. Iob E, Schoeler T, Cecil CM, Walton E, McQuillin A, Pingault J-B (2020): Identifying risk factors involved in the common versus specific liabilities to substance use: A genetically informed approach. *Addict Biol* e12944.
37. Carey CE, Agrawal A, Bucholz KK, Hartz SM, Lynskey MT, Nelson EC, *et al.* (2016): Associations between Polygenic Risk for Psychiatric Disorders and Substance Involvement. *Front Genet* 7. https://doi.org/10.3389/fgene.2016.00149
38. Fowler J, Volkow N, Kassed C, Chang L (2007): Imaging the addicted human brain. *Sci Pract Perspect* 3: 4–16.
39. Mayfield J, Arends MA, Harris RA, Blednov YA (2016): Genes and alcohol consumption. pp 293–355.
40. Canela-Xandri O, Rawlik K, Tenesa A (2018): An atlas of genetic associations in UK Biobank. *Nat Genet* 50: 1593–1599.
41. Neale Lab (2017): Rapid GWAS of thousands of phenotypes for 337,000 samples in the UK. Retrieved from http://www.nealelab.is/blog/2017/7/19/rapid-gwas-of-thousands-of-

phenotypes-for-337000-samples-in-the-uk-biobank

42. Bulik-Sullivan BK, Loh P-R, Finucane HK, Ripke S, Yang J, Patterson N, *et al.* (2015): LD Score regression distinguishes confounding from polygenicity in genome-wide association studies. *Nat Genet* 47: 291–295.

43. Kamat MA, Blackshaw JA, Young R, Surendran P, Burgess S, Danesh J, *et al.* (2019): PhenoScanner V2: an expanded tool for searching human genotype–phenotype associations ((J. Kelso, editor)). *Bioinformatics* 35: 4851–4853.

44. Pers TH, Karjalainen JM, Chan Y, Westra H-J, Wood AR, Yang J, *et al.* (2015): Biological interpretation of genome-wide association studies using predicted gene functions. *Nat Commun* 6: 5890.

45. Lamparter D, Marbach D, Rueedi R, Kutalik Z, Bergmann S (2016): Fast and Rigorous Computation of Gene and Pathway Scores from SNP-Based Summary Statistics ((J. Listgarten, editor)). *PLOS Comput Biol* 12: e1004714.

46. Subramanian A, Tamayo P, Mootha VK, Mukherjee S, Ebert BL, Gillette MA, *et al.* (2005): Gene set enrichment analysis: A knowledge-based approach for interpreting genome-wide expression profiles. *Proc Natl Acad Sci* 102: 15545–15550.

47. Hemani G, Zheng J, Elsworth B, Wade KH, Haberland V, Baird D, *et al.* (2018): The MR-Base platform supports systematic causal inference across the human phenome. *Elife* 7. https://doi.org/10.7554/eLife.34408

48. Mallard TT, Linnér RK, Okbay A, Grotzinger AD, de Vlaming R, Meddens SFW, *et al.* (2019): Not just one p: Multivariate GWAS of psychiatric disorders and their cardinal symptoms reveal two dimensions of cross-cutting genetic liabilities. *bioRxiv* 603134.

49. Goldstein RZ, Volkow ND (2002): Drug addiction and its underlying neurobiological basis: neuroimaging evidence for the involvement of the frontal cortex. *Am J Psychiatry* 159: 1642–1652.

50. Hatoum AS, Colbert SMC, Johnson EC, Huggett SB, Deak JD, Pathak G, *et al.* (2022): Multivariate genome-wide association meta-analysis of over 1 million subjects identifies loci underlying multiple substance use disorders. *medRxiv* 2022.01.06.22268753.

51. Al-Hasani R, Gowrishankar R, Schmitz GP, Pedersen CE, Marcus DJ, Shirley SE, *et al.* (2021): Ventral tegmental area GABAergic inhibition of cholinergic interneurons in the ventral nucleus accumbens shell promotes reward reinforcement. *Nat Neurosci*. https://doi.org/10.1038/s41593-021-00898-2

52. Fischer AG, Ullsperger M (2017): An update on the role of serotonin and its interplay with ropamine for Reward. *Front Hum Neurosci* 11. https://doi.org/10.3389/fnhum.2017.00484

53. Kalivas PW (2009): The glutamate homeostasis hypothesis of addiction. *Nat Rev Neurosci* 10: 561–572.

54. Liu Y, Sun H, Bao Y, Li S, Beveridge TJR, Di X, *et al.* (2009): Subjective, cognitive/psychomotor, and physiological effects of aripiprazole in Chinese light and heavy smokers. *Drug Alcohol Depend* 101: 42–52.

55. Kranzler HR, Covault J, Pierucci-Lagha A, Chan G, Douglas K, Arias AJ, Oncken C (2008): Effects of aripiprazole on subjective and physiological responses to alcohol. *Alcohol Clin Exp Res* 32: 573–579.

56. Rassnick S, Pulvirenti L, Koob GF (1992): Oral ethanol self-administration in rats is reduced by the administration of dopamine and glutamate receptor antagonists into the nucleus

accumbens. *Psychopharmacology (Berl)* 109: 92–98.

57. Kenny PJ, Chartoff E, Roberto M, Carlezon WA, Markou A (2009): NMDA receptors regulate nicotine-enhanced brain reward function and intravenous nicotine self-administration: Role of the Ventral Tegmental Area and Central Nucleus of the Amygdala. *Neuropsychopharmacology* 34: 266–281.

58. Gray KM, Watson NL, Carpenter MJ, LaRowe SD (2010): N-Acetylcysteine (NAC) in young marijuana users: An Open-Label Pilot Study. *Am J Addict* 19: 187–189.

59. Haney M, Hart CL, Vosburg SK, Comer SD, Reed SC, Cooper ZD, Foltin RW (2010): Effects of baclofen and mirtazapine on a laboratory model of marijuana withdrawal and relapse. *Psychopharmacology (Berl)* 211: 233–244.

60. Feduccia AA, Simms JA, Mill D, Yi HY, Bartlett SE (2014): Varenicline decreases ethanol intake and increases dopamine release via neuronal nicotinic acetylcholine receptors in the nucleus accumbens. *Br J Pharmacol* 171: 3420–3431.

61. Wouda JA, Riga D, De Vries W, Stegeman M, van Mourik Y, Schetters D, *et al.* (2011): Varenicline attenuates cue-induced relapse to alcohol, but not nicotine seeking, while reducing inhibitory response control. *Psychopharmacology (Berl)* 216: 267–277.

62. Litten RZ, Ryan ML, Fertig JB, Falk DE, Johnson B, Dunn KE, *et al.* (2013): A double-blind, placebo-controlled trial assessing the efficacy of varenicline tartrate for alcohol dependence. *J Addict Med* 7: 277–286.

63. O'Malley SS, Zweben A, Fucito LM, Wu R, Piepmeier ME, Ockert DM, *et al.* (2018): Effect of varenicline combined with medical management on alcohol use disorder with comorbid cigarette smoking. *JAMA Psychiatry* 75: 129.

64. Bold KW, Zweben A, Fucito LM, Piepmeier ME, Muvvala S, Wu R, *et al.* (2019): Longitudinal findings from a Randomized Clinical Trial of Varenicline for alcohol use disorder with comorbid cigarette smoking. *Alcohol Clin Exp Res* 43: 937–944.

65. Mounier N, Kutalik Z (2021): Bias correction for inverse variance weighting Mendelian randomization. *bioRxiv* 2021.03.26.437168.

66. Morrison J, Knoblauch N, Marcus JH, Stephens M, He X (2020): Mendelian randomization accounting for correlated and uncorrelated pleiotropic effects using genome-wide summary statistics. *Nat Genet* 52: 740–747.

67. Darrous L, Mounier N, Kutalik Z (2020): Simultaneous estimation of bi-directional causal effects and heritable confounding from GWAS summary statistics. *medRxiv* 2020.01.27.20018929.

========== FIGURE 1==========

Figure 1. Multivariate genome-wide association study of the shared genetic architecture of cigarette, alcohol and cannabis use

Panel A. Heat map displaying the genetic correlations among the six substance use phenotypes. Shown are the genetic correlations between each of the cigarette, alcohol and cannabis phenotypes, with SNP-heritability estimates displayed down the diagonal. The mean genetic correlation is r_g=0.4 [sd=0.21, median=0.34 and range (-0.01-0.74)]. **Panel B.** Genomic structural equation model fitted on the genetic covariation matrices of the individual cigarette, alcohol and cannabis use phenotypes. Squares represent observed variables (the measured cigarette, alcohol and cannabis use phenotypes). The circle represents the latent variable, i.e., the common heritable liability to substance involvement, derived through factor analysis of the genetic correlations between the individual substance use phenotypes. Single-headed arrows are regression paths constrained to be equal across measures of frequency of use and dependence. **Panel C.** Manhattan plot of the SNP effects obtained from the multivariate genome-wide association analysis on the common liability. Labels are provided for the LD-independent genome-wide significant SNPs (i.e., SNPs above the horizontal line, with $p<5\times10^{-8}$) and gene names obtained through positional mapping. The x-axis refers to chromosomal position, the y-axis refers to the *p*-value on a -log10 scale. Genetic variants coloured in red index variants that showed heterogeneous effects across the individual cigarette, alcohol and cannabis use phenotypes (Q_{SNP} $p<5\times10^{-8}$), indicating that their effects operate not entirely through the common liability. Genetic variants coloured in blue index genetic variants that did not show heterogeneous effects across the individual cigarette, alcohol and cannabis use phenotypes (Q_{SNP} $p\geq5\times10^{-8}$), indicating that their effects are likely to operate through the common liability.

========== FIGURE 2==========

Figure 2. Associations of genetic variants with the common liability (blue) and the individual substance use phenotypes (red)

Plotted are the standardized beta coefficients (β_{std}) and their confidence intervals (cf. Supplement for details and corresponding formula) obtained from the multivariate genome-wide association (GWA) analysis on the common liability (column 1) and the univariate GWA analyses on the individual substance use phenotypes (columns 2-7). Displayed are genetic variants associated ($p<5\times10^{-8}$) with at least one of the individual substance use phenotypes and/or the common liability. Bars coloured in grey index genetic variants that are not significantly associated ($p\geq5\times10^{-8}$) with their respective phenotype. Bars coloured in red index genetic variants that showed heterogeneous effects across the individual cigarette, alcohol and cannabis use phenotypes (Q_{SNP} $p<5\times10^{-8}$), indicating that their effects operate not entirely through the common liability. Bars coloured in blue index genetic variants that did not show heterogeneous effects across the individual cigarette, alcohol and cannabis use phenotypes (Q_{SNP} $p\geq5\times10^{-8}$), indicating that their effects are unlikely to entirely operate through the common liability. The complete set of estimates can be found in **sTable 7**. The asterisks (*) highlight genetic variants that were identified as LD-independent SNPs following clumping

========== FIGURE 3==========

Figure 3. Pathway enrichment analyses of genes associated with the common heritable liability

Shown are the results obtained from pathway enrichment analysis conducted in DEPICT and PASCAL. The common liability GWA results (filtered according to Q_{SNP} $p<5\times10^{-8}$) and the individual substance use GWA summary statistics were used as the input. The violet shading indexes the significance level corresponding to each tested pathway. The asterisk marks pathways that remained significant after correction for multiple testing (False Discovery Rate controlled at 5%). **Panel A** highlights results obtained from the tissue/cell type enrichment analysis done in DEPICT. Displayed

are the -log10(*p*-value) for all pathways that were significant ($p<0.05$) in at least one of the included GWA studies. **Panel B** depicts results obtained from pathway analysis done in PASCAL, using gene-sets curated by the Molecular Signatures Database (n=1077 sets) and DEPICT (n=14462 sets). For the common liability, n=478 pathways were significant after FDR correction for multiple testing. Displayed in the figure are the 15 most significant pathways per GWA study. The full set of results is listed in **sTable 10-11** (Supplement).

========== FIGURE 4==========

Figure 4. Genetic correlations between the common liability and other traits

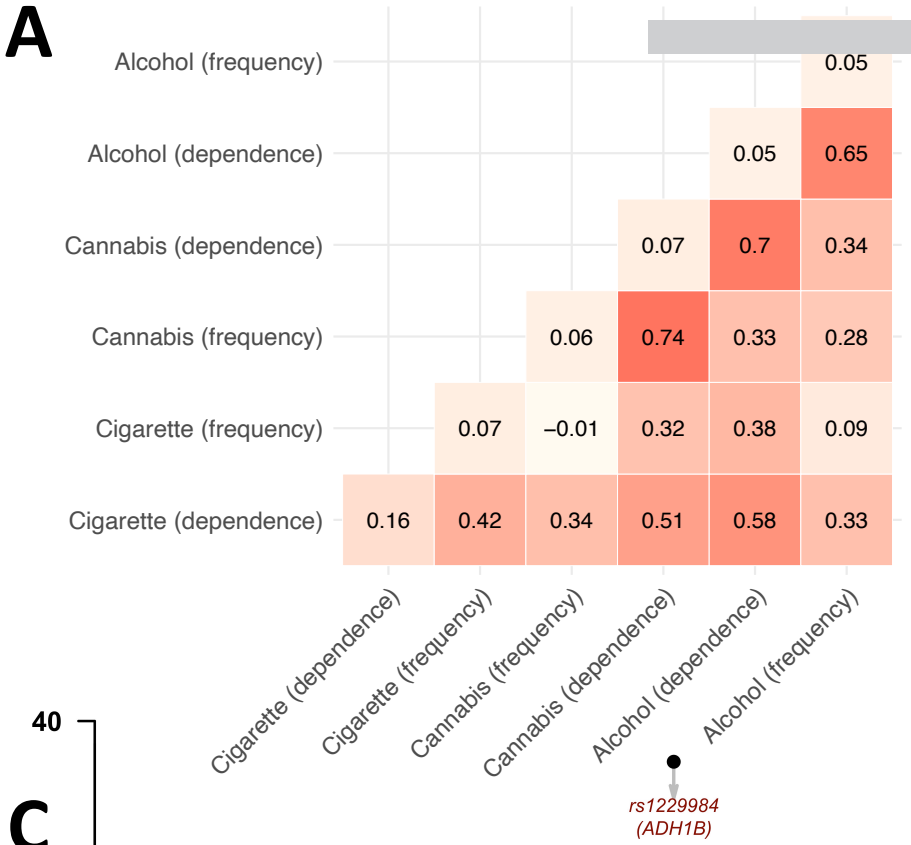
Shown are the genetic correlations (r_g) between the common liability GWA (filtered according to Q_{SNP} $p<5\times10^{-8}$) and 41 other phenotypes, including 35 other traits (highlighted in blue) and the six individual substance use phenotypes used to derive the common liability (highlighted in grey). The asterisk indexes significant genetic correlations after correction for multiple testing (false discovery rate controlled at 5%, corrected for 41 tests). The full set of results is reported in **sTable 12** (Supplement).

========== FIGURE 5==========

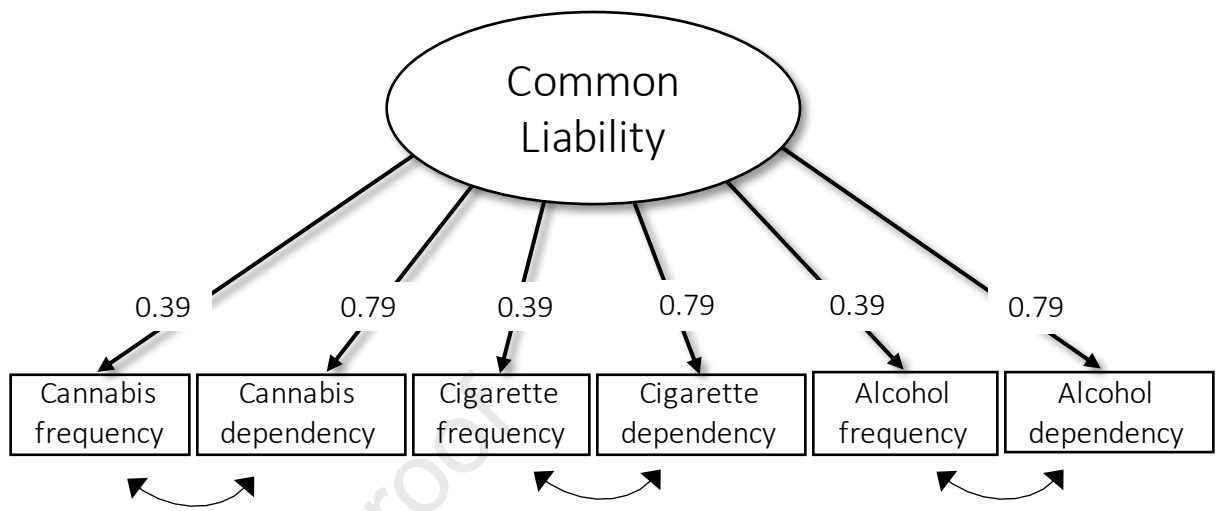
Figure 5. Mendelian Randomization analysis assessing causality between the common liability and the individual substance use phenotypes

Shown are the standardized beta coefficients (β_{std}) obtained from Mendelian Randomization (MR) analysis assessing the effects of the common liability on the six individual substance use phenotypes. Included were 42 genome-wide significant genetic variants ($p<5\times10^{-8}$) operating through the common liability (Q_{SNP} $p>5\times10^{-8}$) as instruments for the exposure. The red dots indicate the standardized loadings per substance use phenotype on the common liability as estimated in the structural model shown in **Figure 1B**. The full set of MR results can be found in **sTable 13**.

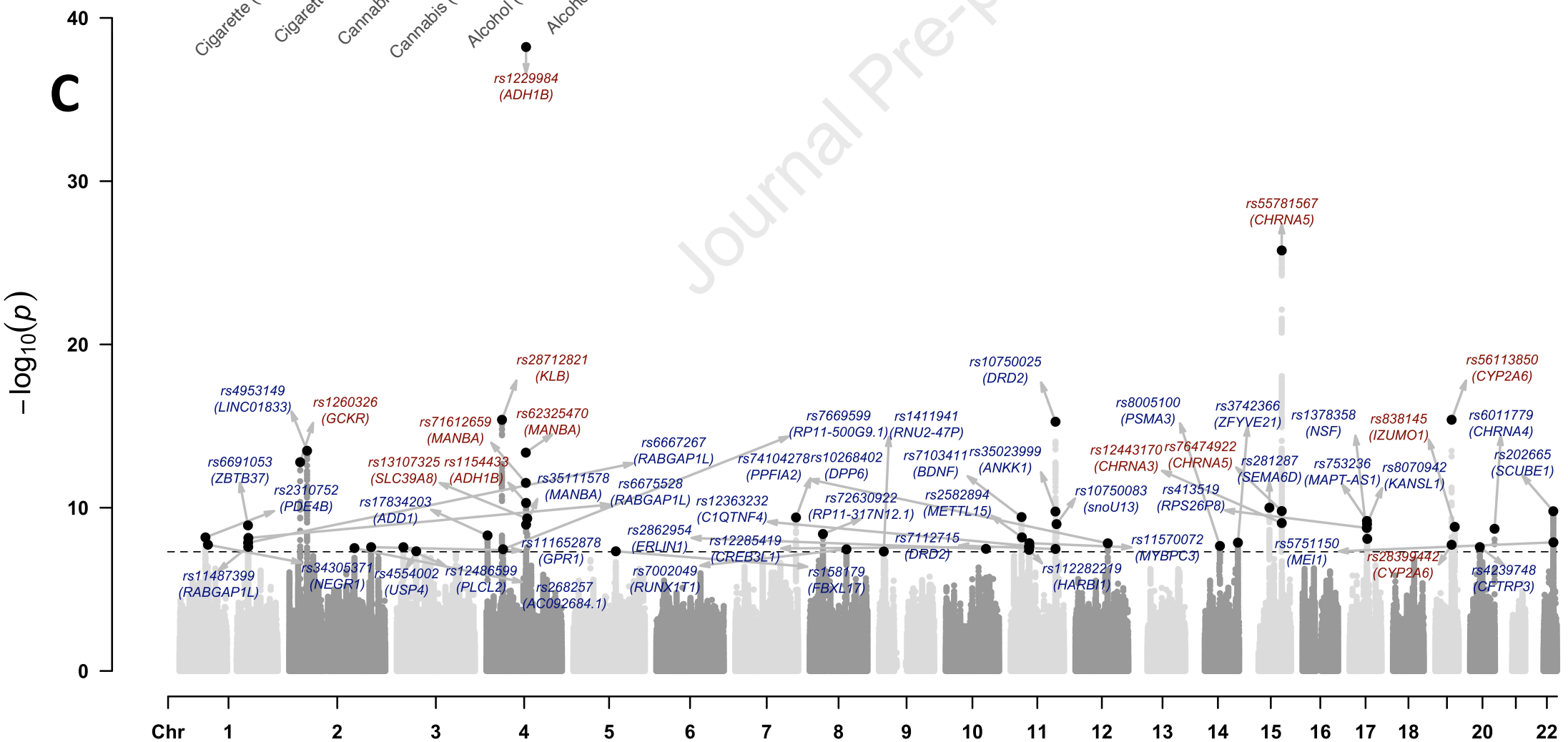
A

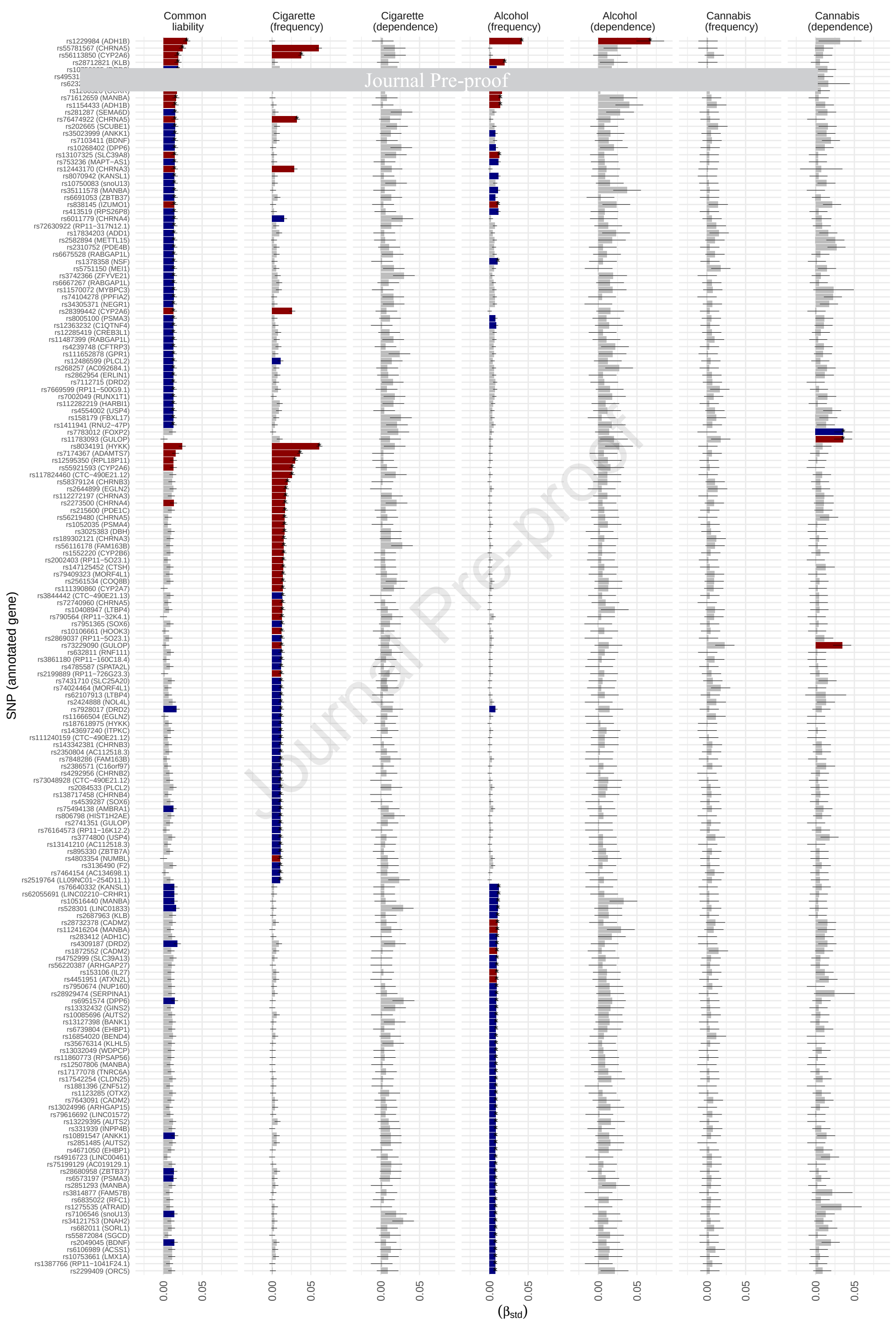


B



C





A

Common liability | Alcohol | Cigarette | Cannabis

Brain
Central Nervous System
Occipital Lobe
Visual Cortex
Prosencephalon
Telencephalon
Cerebrum
Cerebral Cortex
Limbic System
Hippocampus
Mesencephalon
Parahippocampal Gyrus
Entorhinal Cortex
Temporal Lobe
Diencephalon
Frontal Lobe
Hypothalamus
Basal Ganglia
Hypothalamo Hypophyseal System
Hypothalamus Middle
Neurosecretory Systems
Corpus Striatum
Parietal Lobe
Brain Stem
Retina
Rhombencephalon
Metencephalon
Epidermis
Cerebellum
Exocrine Glands
Prostate
Genitalia Male
Neural Stem Cells
Atrial Appendage
Heart Atria
Endocrine Cells
Endocrine Glands
Genitalia
Gonads
Ovary
Adnexa Uteri
Adipocytes
Cervix Uteri
Genitalia Female
Esophagus
Adipose Tissue
Subcutaneous Fat
Adipose Tissue White
Glucagon Secreting Cells
Enteroendocrine Cells
Subcutaneous Fat Abdominal
Abdominal Fat
Fallopian Tubes
Adult Stem Cells
Foreskin
Islets of Langerhans
Penis
Keratinocytes
Synovial Membrane
Joint Capsule
Joints
Fibroblasts
Synovial Fluid
Leukocytes
Cicatrix
Granulation Tissue
Spleen
Keloid
Embryoid Bodies
Arteries
Chorion
Extraembryonic Membranes
Osteoblasts

Common liability | Alcohol (frequency) | Alcohol (dependence) | Cigarette (frequency) | Cigarette (dependence) | Cannabis (frequency) | Cannabis (dependence)

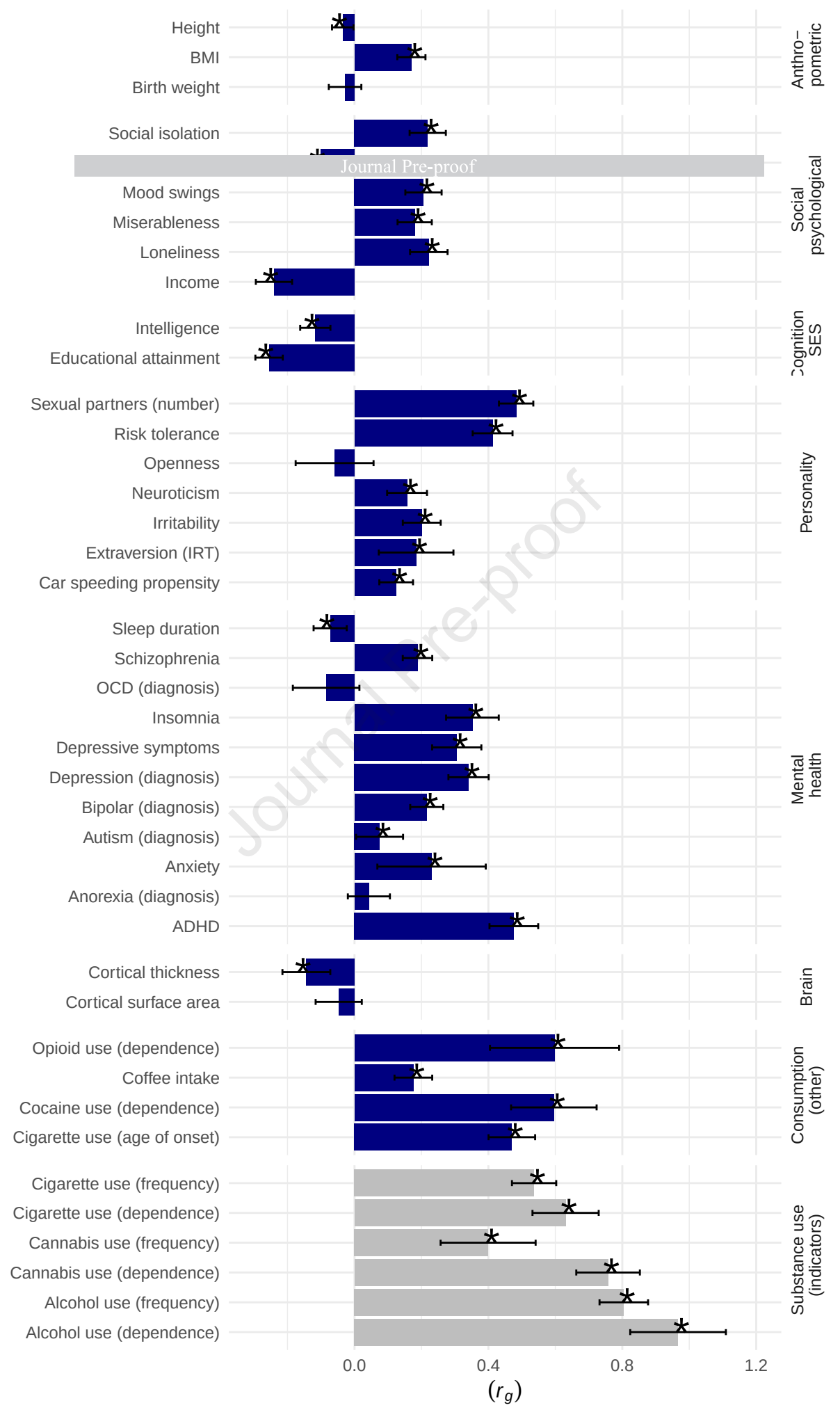
B

Common liability | Alcohol | Cigarette | Cannabis

reactome depolarization of ...
abnormal excitatory postsyn...
reactome transmission acros...
presynaptic membrane
dendrite
abnormal inhibitory postsyn...
synapse
obsolete synapse part
ionotropic glutamate recept...
neuron spine
dendritic spine
ligand−gated channel activity
ligand−gated ion channel ac...
abnormal locomotor behavior
periplasmic space
outer membrane−bounded peri...
absence seizures
CACNG2 (calcium voltage−gat...
postsynaptic membrane
voltage−gated cation channe...
neuron projection terminus
neuron cell−cell adhesion
perikaryon
regulation of neuronal syna...
regulation of synaptic tran...
GABRA1 (gamma−aminobutyric ...
increased synaptic depression
regulation of transmission ...
mitogen−activated protein k...
learning
seizures
axon terminus
regulation of nervous syste...
CLK3 (CDC like kinase 3 [So...
abnormal locomotor activation
ensg00000137345
ensg00000206456
MOG (myelin oligodendrocyte...
LGI1 (leucine rich glioma i...
POU2F3 (POU class 2 homeobo...
impaired righting response
AKAP8 (A−kinase anchoring p...
abnormal brain development
enhanced long term potentia...
LCA5 (lebercilin LCA5 [Sour...
RTN3 (reticulon 3 [Source:H...
ZNF106 (zinc finger protein...
GATA1 (GATA binding protein...
abnormal long term depression
abnormal medulla oblongata ...
NUFIP2 (nuclear FMR1 intera...
abnormal neuron morphology
DDX21 (DExD−box helicase 21...
abnormal midbrain development
SORBS2 (sorbin and SH3 doma...
PATL1 (PAT1 homolog 1, proc...
abnormal motor neuron morph...
increased neuron apoptosis
cAMP binding
reactome highly calcium per...
reactome presynaptic nicoti...
reactome acetylcholine bind...
STK11 (serine/threonine kin...

Common liability | Alcohol (frequency) | Alcohol (dependence) | Cigarette (frequency) | Cigarette (dependence) | Cannabis (frequency) | Cannabis (dependence)

$-\log_{10}(p)$ 1 2 3


Height
BMI
Birth weight
Anthro–pometric
Social isolation
Journal Pre-proof
Mood swings
Miserableness
Loneliness
Income
Social psychological
Intelligence
Educational attainment
Cognition SES
Sexual partners (number)
Risk tolerance
Openness
Neuroticism
Irritability
Extraversion (IRT)
Car speeding propensity
Personality
Sleep duration
Schizophrenia
OCD (diagnosis)
Insomnia
Depressive symptoms
Depression (diagnosis)
Bipolar (diagnosis)
Autism (diagnosis)
Anxiety
Anorexia (diagnosis)
ADHD
Mental health
Cortical thickness
Cortical surface area
Brain
Opioid use (dependence)
Coffee intake
Cocaine use (dependence)
Cigarette use (age of onset)
Consumption (other)
Cigarette use (frequency)
Cigarette use (dependence)
Cannabis use (frequency)
Cannabis use (dependence)
Alcohol use (frequency)
Alcohol use (dependence)
Substance use (indicators)
0.0
0.4
0.8
1.2
(r_g)