nature mental health

Article

https://doi.org/10.1038/s44220-023-00034-y

Multivariate genome-wide association meta-analysis of over 1 million subjects identifies loci underlying multiple substance use disorders

Received: 21 September 2022

Accepted: 10 February 2023

Published online: 22 March 2023

Check for updates

Alexander S. Hatoum[1] ✉, Sarah M. C. Colbert[1], Emma C. Johnson[1], Spencer B. Huggett[2], Joseph D. Deak[3,4], Gita A. Pathak[3], Mariela V. Jennings[5], Sarah E. Paul[6], Nicole R. Karcher[1], Isabella Hansen[6], David A. A. Baranger[1], Alexis Edwards[7], Andrew D. Grotzinger[8], Substance Use Disorder Working Group of the Psychiatric Genomics Consortium*, Elliot M. Tucker-Drob[9], Henry R. Kranzler[10,11], Lea K. Davis[12,13,14], Sandra Sanchez-Roige[5,12], Renato Polimanti[3,4], Joel Gelernter[3,15,16,136], Howard J. Edenberg[17,18,136], Ryan Bogdan[6,136] & Arpana Agrawal[1,136]

Genetic liability to substance use disorders can be parsed into loci that confer general or substance-specific addiction risk. We report a multivariate genome-wide association meta-analysis that disaggregates general and substance-specific loci from published summary statistics of problematic alcohol use, problematic tobacco use, cannabis use disorder and opioid use disorder in a sample of 1,025,550 individuals of European descent and 92,630 individuals of African descent. Nineteen independent single-nucleotide polymorphisms were genome-wide significant ($P < 5 \times 10^{-8}$) for the general addiction risk factor (addiction-rf), which showed high polygenicity. Across ancestries, *PDE4B* was significant (among other genes), suggesting dopamine regulation as a cross-substance vulnerability. An addiction-rf polygenic risk score was associated with substance use disorders, psychopathologies, somatic conditions and environments associated with the onset of addictions. Substance-specific loci (9 for alcohol, 32 for tobacco, 5 for cannabis and 1 for opioids) included metabolic and receptor genes. These findings provide insight into genetic risk loci for substance use disorders that could be leveraged as treatment targets.

The lives lost, impacts on individuals and families, and socioeconomic costs attributable to substance use reflect a growing public health crisis[1]. For example, in the United States, 13.5% of deaths among young adults[2] are attributable to alcohol, smoking is the leading risk factor for mortality in males[3], and the odds of dying by opioid overdose are greater than those of dying in a motor vehicle crash[4]. Despite the large impact of substance use and substance use disorders[5], there is limited knowledge of the molecular genetic underpinnings of addiction broadly.

A full list of affiliations appears at the end of the paper. ✉e-mail: ashatoum@wustl.edu

https://doi.org/10.1038/s44220-023-00034-y

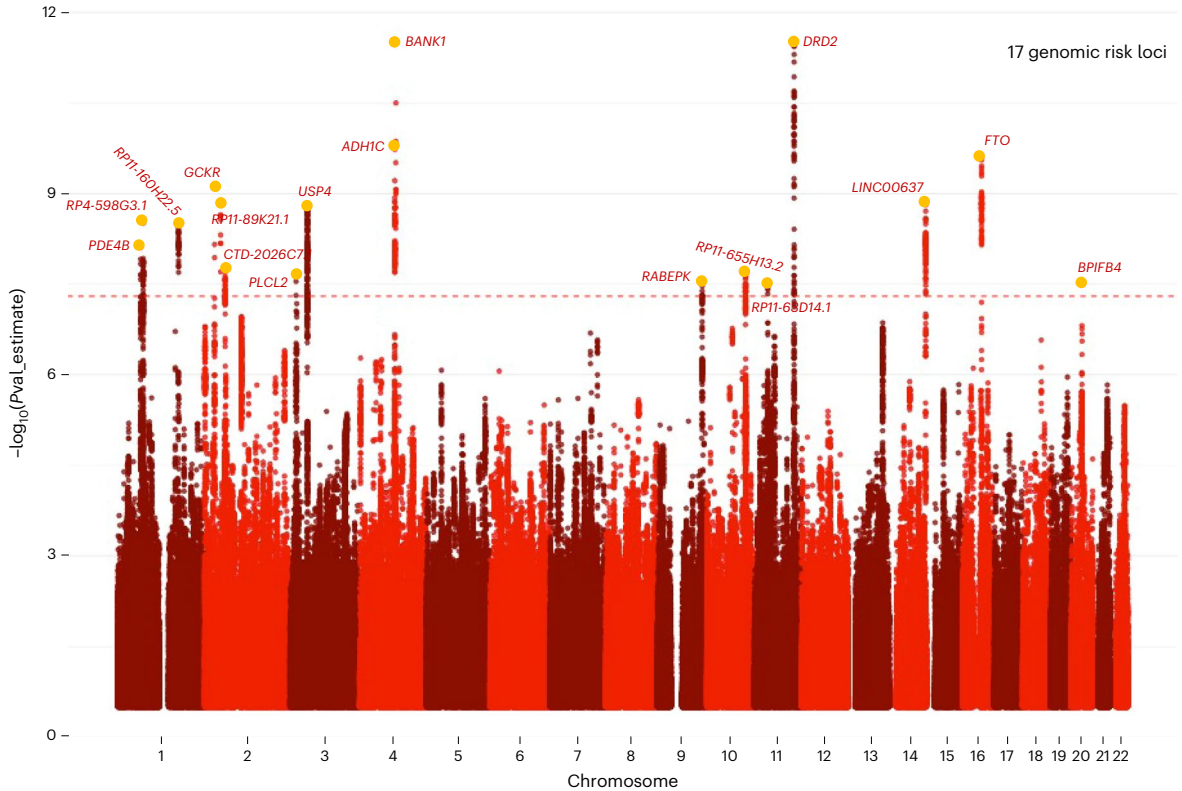

Fig. 1 | Manhattan plot of the addiction-rf GWAS results. The dotted line represents genome-wide significance at 5×10^{-8}. Each SNP peak is annotated with the closest mapped gene from FUMA (Table 1). We have not included all SNPs in the credible set in Table 1, but they are shown in Supplementary Table 4. Significance is set at genome-wide significance Bonferroni correction is a two-sided test ($P < 5 \times 10^{-8}$).

Individual substance use disorders (SUDs) are heritable (h^2, ~50–60%) and highly polygenic[6,7]. Recent large-scale genome-wide association studies (GWASs) have identified loci associated with problematic drinking[8,9], alcohol use disorder (AUD)[10,11], cigarettes smoked per day[12], nicotine dependence[13,14], cannabis use disorder (CUD)[15] and opioid use disorder (OUD)[16]. Echoing evidence from twin and family studies[17], these GWASs show that the genetic architecture of SUDs is characterized by a high degree of commonality[18], that is, a general addiction genetic factor likely conveys vulnerability to multiple SUDs. Even after accounting for genetic correlations with non-problematic substance use and with other psychiatrically relevant traits and disorders, there is considerable variance that is unique to this general risk for addiction, indicating that a liability to addiction reflects more than just the combined genetic liability to substance use and psychopathology[18–21].

We conducted a multivariate GWAS of the largest available discovery GWASs of SUDs, including problematic alcohol use (PAU: $N = 435{,}563$; continuous)[8], problematic tobacco use (PTU: $N = 270{,}120$; continuous)[12,13,18], CUD ($N = 384{,}032$, cases = 14,080)[15] and OUD ($N = 79{,}729$, cases = 10,544 cases)[16]. First, we partitioned single-nucleotide polymorphism (SNP) effects into five sources of variation: **(1)** a general addiction risk factor (referred to as the addiction-rf), and risks specific to **(2)** alcohol, **(3)** nicotine, **(4)** cannabis and **(5)** opioids. Second, we identified biological pathways underlying risk for these five SUD phenotypes using gene, expression quantitative trait locus (eQTL) and pathway enrichment analyses. Third, we examined whether currently available medications could potentially be repurposed to treat SUDs[22]. Fourth, we assessed the association of a polygenic risk score (PRS) derived from the addiction-rf with general SUD phenotypes in an independent case/control sample. Fifth, we examined the extent to which genetic liability to the addiction-rf is shared with other phenotypes (for example, physical and mental health outcomes). Sixth, we tested whether the addiction-rf PRS was associated with medical diagnoses derived from electronic health records (EHRs) and with behavioural phenotypes in largely substance-naive 9–10-year-old children.

Results

Addiction risk factor in European ancestry GWAS

As in our prior study[18], we estimated a single factor model, scaled the variance of the addiction-rf to 1 and allowed loadings to be estimated freely. The single factor model that loaded on OUD ($N_{effective} = 30{,}443$), PAU ($N_{effective} = 300{,}789$), PTU ($N_{effective} = 270{,}120$) and CUD ($N_{effective} = 46{,}351$) fit the data well ($\chi^2(1) = 0.017$, $P = 0.896$, comparative fit index (CFI) = 1, standardized root mean square residual (SRMR) = 0.002). The latent factor loaded significantly on all indicators (standardized loadings on OUD = 0.83, PAU = 0.58, PTU = 0.36, CUD = 0.93; see Supplementary Fig. 1 for full model). The addiction-rf was associated with 19 independent ($r^2 < 0.1$) genome-wide significant (GWS) SNPs that mapped to 17 genomic risk loci (Fig. 1; Table 1; Supplementary Table 1 for lead SNPs and Supplementary Table 2 for genomic risk loci). The most significant SNP (rs6589386, $P = 2.9 \times 10^{-12}$) was intergenic, but closest to *DRD2*, which was GWS in gene-based analyses ($P = 7.9 \times 10^{-12}$; Supplementary Table 3). Further, rs6589386 was an eQTL for *DRD2* in the cerebellum, and Hi-C analyses (in FUMA)[23] revealed that the variant made chromatin contact with the promoter of the gene (Supplementary Fig. 2).

Gene-based analyses identified 42 significantly associated genes (Supplementary Table 3); the most significant signals were *FTO* ($P = 1.86 \times 10^{-13}$), *DRD2* ($P = 7.9 \times 10^{-12}$) and *PDE4B* ($P = 9.63 \times 10^{-11}$). Fine-mapping identified 123 GWS SNPs (of 660 non-independent GWS SNPs) in credible sets as potential causal SNPs based on the posterior probability of inclusion (Supplementary Table 4). Mapping the lead independent SNPs in the credible sets to their nearest gene based on posterior probability of 1, the following SNPs showed the strongest causal potential: rs1937455 (*PDE4B*), rs3739095 (*GTF3C2*), rs6718128

https://doi.org/10.1038/s44220-023-00034-y

Table 1 | Lead GWAS significant variants

rsID	Chr	A1	Position	European ancestry				African ancestry		Cross-ancestry	
				GenomicSEM beta	GenomicSEM *P*	SusieR SNPs in credible set	FUMA GWAS SNPs	ASSET beta	ASSET *P*	METASOFT beta	METASOFT *P*
rs1937455	1	A	66416939	0.013	7.74×10^{-9}	2	43	0.024	3.00×10^{-3}	0.013	0.100
rs1475064	1	A	73882478	0.014	3.06×10^{-9}	3	161	0.09	6.25×10^{-3}	0.011	0.051
rs2860846	1	T	174075924	0.014	3.37×10^{-9}	34	36	NA	NA	NA	NA
rs1260326	2	T	27730940	−0.015	7.60×10^{-10}	17	10	NA	NA	NA	NA
rs570436	2	C	45142673	−0.015	1.31×10^{-9}	15	8	NA	NA	NA	NA
rs2717054	2	G	58046683	−0.014	1.97×10^{-8}	8	44	NA	NA	NA	NA
rs55855024	3	C	16850764	−0.014	2.37×10^{-8}	8	35	NA	NA	NA	NA
rs6795772	3	C	49365269	0.014	1.58×10^{-9}	2	265	NA	NA	NA	NA
rs3114045	4	T	100252560	−0.023	1.85×10^{-10}	3	134	-0.019	0.554	−0.012	2.60×10^{-15}
rs1662031	4	A	100256793	−0.015	1.17×10^{-9}	3	134	NA	NA	NA	NA
rs1813006	4	G	103001649	0.037	3.18×10^{-12}	9	11	NA	NA	NA	NA
rs864882	9	C	127968109	0.015	3.41×10^{-8}	2	26	NA	NA	NA	NA
rs7073987	10	C	110565868	0.015	2.04×10^{-8}	3	61	NA	NA	NA	NA
rs2861190	11	C	38517941	−0.014	3.66×10^{-8}	5	117	NA	NA	NA	NA
rs17602038	11	T	113364691	0.017	6.64×10^{-12}	2	117	NA	NA	NA	NA
rs6589386	11	C	113443753	0.017	2.92×10^{-12}	2	65	NA	NA	NA	NA
rs10083370	14	G	104314182	0.014	1.53×10^{-9}	3	89	NA	NA	NA	NA
rs28567725	16	T	53826028	0.016	2.50×10^{-10}	5	83	0.012	0.457	0.007	6.49×10^{-12}
rs2424952	20	T	31685873	0.030	3.21×10^{-8}	2	4	NA	NA	NA	NA

Lead GWAS variants from European ancestry addiction-rf meta-analysis (in GenomicSEM), African ancestry addiction-rf meta-analysis (based on ASSET) and trans-ancestry meta-analysis (in METASOFT) of the common SNPs underlying PAU, PTU, CUD and OUD. We generated four cross-substance meta-analyses. First, we used a model that leverages genetic overlap across different SUDs via GenomicSEM (Fig. 1). The GWAS of European ancestry individuals, run with GSEM, forms the primary analysis for most downstream analyses (that is, TWAS, genetic correlation, PheWAS and genetic causality). The GSEM results for the addiction-rf are shown first (GenomicSEM beta, GenomicSEM *P*); the number of credible SNPs in each set (SusieR; r^2=0.6) with GWAS lead SNPs (from FUMA) are also shown. Next, results for the cross-substance meta-analysis in African ancestry individuals, using ASSET, is shown (ASSET beta, ASSET *P*). ASSET splits groups of SNPs into pleiotropic versus non-pleiotropic SNPs, which produces a sparser set of GWAS results (as all SNPs must be pleiotropic to estimate a beta and *P*-value), hence, NAs. Finally, to conduct a cross-ancestry meta-analysis, we applied ASSET to the European ancestry sample and then meta-analysed the European and African ancestry summary data using METASOFT (METASOFT beta, METASOFT *P*; see Table 2). Direction of all betas corresponds to the effect allele (A1). rsID, rs number; Chr, chromosome; A1, effect allele; position, genomic position in base pairs. Significance is set at Bonferroni-corrected genome-wide significance in a two-sided test ($P<5\times10^{-8}$).

(*ZNF512*), rs4143308 (*RP11-89K21.1*), rs4953152 (*SIX3*), rs41335055 (*CTD-2026C7.1*), rs2678900 (*VRK2*), rs7620024 (*TCTA*), rs283412 (*ADH1C*), rs901406 (*BANK1*), rs359590 (*RABEPK*), rs10083370 (*LINC00637*), rs1477196 (*FTO*) and rs291699 (*CDK5RAP1*) (Supplementary Table 4 and Fig. 1). Pathway analysis of gene-based results revealed several significant gene ontology (GO) terms including double-stranded DNA binding ($P_{Bonferroni}$ = 0.005), sequence-specific double-stranded DNA binding ($P_{Bonferroni}$ = 0.01), regulation of nervous system development (two terms: $P_{Bonferroni}$ = 0.011–0.037), and positive regulation of transcription by RNA polymerase ($P_{Bonferroni}$ = 0.038) (Supplementary Table 6).

Substance-specific risk in European ancestry GWAS

To identify loci associated with only a single substance (that is, not pleiotropic), we used ASSET (Association Analysis Based on Subsets[24]; one-sided $P < 5\times10^{-8}$). SNPs that were associated at GWS with only an individual substance (PAU, PTU, CUD or OUD) were considered substance-specific (for example, *CHRNA5* SNPs were only associated with PTU; Supplementary Fig. 3b–e).

Problematic alcohol use. ASSET analyses revealed nine independent SNPs in six loci associated specifically with PAU (Supplementary Fig. 3b; Supplementary Tables 7 and 8). As expected[8], the top signal was rs1229984 in *ADH1B* ($P = 4.11\times10^{-68}$). Gene-based enrichment analyses also implicated the alcohol dehydrogenase activity zinc-dependent pathway ($P_{Bonferroni}$ = 0.035; Supplementary Table 9).

Problematic tobacco use. PTU was specifically associated with 32 independent SNPs in 12 loci (Supplementary Fig. 3c; Supplementary Tables 10 and 11). The top SNP was rs10519203 ($P = 5.12\times10^{-267}$) in *HYKK* which is also a robust eQTL for *CHRNA5*; the signal is likely driven by the *CHRNA5* missense variant, rs16969968 ($P = 2.79\times10^{-175}$), which has previously been linked to tobacco use (r^2 = 0.87)[12]. Several other SNPs were closest to genes encoding nicotinic acetylcholine receptors, including *CHRNA4, CHRNB4, CHRNB3* and *CHRNB2* (Supplementary Table 10). Gene-based enrichment implicated multiple pathways and gene sets related to nicotinic acetylcholine receptors (Supplementary Table 12). Specific dopamine-related associations were also noted (for example, *PDE1C:* rs215600; $P = 2.35\times10^{-18}$; *DBH:* rs1108581; $P = 1.00\times10^{-14}$).

Cannabis use disorder. ASSET identified five substance-specific loci for CUD (Supplementary Tables 13 and 14), with lead signals at rs11913634 (*FAM19A5;* $P = 1.20\times10^{-15}$), rs8104317 (*CACNA1A;* $P = 1.17\times10^{-13}$), rs72818514 (*ATP10B*; $P = 1.57\times10^{-9}$), rs11715758 (*GNAI2/HYAL3*; $P = 4.84\times10^{-8}$; Supplementary Fig. 3d) and rs11778040 ($P = 1.77\times10^{-9}$; annotated to the *GULOP* pseudogene). rs11778040 also mapped to the previously discovered signal for CUD near *CHRNA2* and *EPHX2*[15] and is an eQTL for *CHRNA2*, *EPHX2* and *CCDC25*. CUD-specific signals showed no significant gene-based enrichment.

Opioid use disorder. The only significant substance-specific signal for OUD was the well-characterized[16] mu opioid receptor (*OPRM1*) SNP,

Table 2 | Top results from the cross-ancestry meta-analysis in METASOFT

Chr	rsID	Gene	A1	Pheno Eur	Pheno AA	Cross beta	Cross *P*	AA beta	AA *P*	Eur beta	Eur *P*
19	rs507766	*FUT2*	T	CUD,PAU	CUD,AUD	−0.009	3.47×10^{-19}	0.007	0.914	−0.004	2.66×10^{-4}
16	rs9928094	*FTO*	A	PTU,OUD,PAU	OUD,AUD	0.006	6.50×10^{-32}	0.011	0.321	0.006	1.63×10^{-11}
2	rs472140	NA	C	PTU,PAU	CUD,PTU	−0.007	3.79×10^{-43}	−0.02	0.722	−0.007	7.66×10^{-13}
4	rs6846266	NA	A	PTU,PAU	OUD,AUD	−0.005	6.05×10^{-23}	−0.011	0.317	−0.005	3.61×10^{-6}
3	rs62250713	*CADM2*	G	CUD,PAU	CUD,AUD	0.009	1.00×10^{-18}	0.018	0.268	0.009	6.97×10^{-11}
1	rs784601	NA	G	CUD,PTU,PAU	CUD,PTU	−0.004	3.57×10^{-15}	−0.014	0.519	−0.004	2.47×10^{-5}
7	rs4727799	*FOXP2*	C	CUD,PTU	PTU,OUD	−0.004	3.90×10^{-15}	−0.024	0.752	−0.004	2.74×10^{-6}
22	rs6002381	NA	T	CUD,PTU,PAU	PTU,AUD	−0.007	5.23×10^{-12}	−0.022	0.551	−0.007	1.13×10^{-6}
12	rs1497253	*IRAG2*	A	CUD,PTU,OUD,PAU	CUD,PTU,AUD	−0.004	4.01×10^{-15}	0.007	0.914	−0.004	2.66×10^{-4}
14	rs7147171	*PP1R13B*	G	PTU,PAU	PTU,OUD,AUD	−0.007	3.71×10^{-12}	−0.01	0.37	−0.007	3.09×10^{-10}
15	rs35175834	*SEMA6D*	G	CUD,PTU,PAU	CUD,PTU,OUD,AUD	−0.007	5.84×10^{-12}	0.007	0.959	−0.007	3.51×10^{-7}
20	rs293553	*C20orf112*	A	PTU,OUD	PTU,OUD	−0.009	2.03×10^{-9}	−0.018	0.24	−0.009	1.43×10^{-7}
9	rs7033815	*SCAI*	G	PTU,OUD	CUD,OUD,AUD	0.006	2.71×10^{-9}	0.018	0.171	0.006	2.83×10^{-6}
17	rs587880	*FBXL20*	G	OUD,PAU	OUD,AUD	−0.009	1.07×10^{-9}	−0.014	0.119	−0.009	1.36×10^{-5}
18	rs4996482	*RP11-397A1**6.1***	C	CUD,PTU,OUD	CUD,OUT	0.006	3.54×10^{-9}	0.028	0.006	0.261	4.18×10^{-7}
5	rs7708715	*TMEM161B-AS1,CTC-498M16.2*	C	CUD,PTU	CUD,OUD,AUD	0.003	3.59×10^{-9}	0.013	0.34	0.003	6.16×10^{-5}
8	rs2321459	*LEKR1*	T	PTU,PAU	PTU,OUD	−0.006	3.68×10^{-9}	−0.009	0.974	−0.006	4.51×10^{-5}
6	rs62394558	NA	G	CUD,PTU	CUD,OUD	−0.003	3.89×10^{-9}	−0.02	0.962	−0.003	8.00×10^{-3}
1*	rs1937439	*PDE4B*	A	CUD,PTU,PAU	CUD,AUD	0.007	5.18×10^{-12}	0.025	0.173	0.007	4.54×10^{-10}
4*	rs10031172	*BANK1*	T	CUD,OUD,PAU	PTU,AUD	−0.006	1.93×10^{-9}	−0.014	0.090	−0.006	8.03×10^{-5}

Chr, chromosome; rsID, rs number; Gene, closest gene; A1, effect allele. Pheno AA lists the phenotypes associated with the SNP in the African ancestry sample GWAS. Pheno Eur lists the phenotypes associated with the SNP in the European sample. Cross beta is the cross-ancestry beta from the METASOFT random-effects meta-analysis. Cross *P* is the *P*-value of the cross-ancestry meta-analysis. AA beta is the transformation of the odds ratio from ASSET when run in the African American sample. AA *P* is the *P*-value from ASSET. Eur beta is the transformation of the odds ratio (to a beta) from ASSET when run in the European sample. Eur *P* is the *P*-value from ASSET. Significance is set at Bonferroni-corrected GWS in a two-sided test ($P<5\times10^{-8}$). CUD, cannabis use disorde; PAU, problematic alcohol use; PTU, problematic tobacco use; OUD, opioid use disorder.

rs1799971 ($P=1.63\times10^{-8}$; Fig. 2e). Gene-based analyses produced no significant findings.

Cross-substance risk in African ancestry GWAS

The ASSET-based meta-analysis of GWAS data for AUD ($N=82{,}705$)[11], tobacco dependence (TD; based on the Fagerström Test for Nicotine Dependence, $N=9{,}925$)[13], CUD ($N=9{,}745$)[15] and OUD ($N=32{,}088$)[16] in individuals of African ancestry yielded only one GWS pleiotropic SNP, rs77193269 ($P=4.92\times10^{-8}$); this SNP was GWS for AUD and TD when considering ASSET loci pleiotropic for two substances (Supplementary Fig. 4b). For substance-specific signals, only one SNP was GWS significant: rs2066702, an *ADH1B* variant that was alcohol-specific (Supplementary Fig. 4a).

Cross-substance risk in cross-ancestry GWAS

We found 68 GWS SNPs (Supplementary Fig. 5), which are challenging to map to nearby regions or candidate genes due to ancestral differences in LD structure. Table 2 lists the SNP with the lowest GWAS *P* value on each chromosome. The most significant association was noted near the *FUT2* gene (rs507766, $P=3.47\times10^{-19}$). Many GWS signals were consistent with genes found in the European GWAS, including *FTO* (rs9928094, $p=6.50\times10^{-32}$) and *PDE4B* (rs1937439, $P=8.56\times10^{-12}$). We also identified two SNPs in genes that have previously been implicated in SUDs including *CADM2* (rs62250713, $P=1.00\times10^{-18}$) and *FOXP2* (rs4727799, $P=3.90\times10^{-15}$), both of which were within $r^2=0.6$ of lead signals from the European GWAS.

Polygenic architecture and power

We used a likelihood estimation-based approach to calculate the probability distribution of effect sizes for the addiction-rf and each of the constituent input GWASs (that is, PAU, PTU, CUD and OUD) to examine relative differences in polygenicity (Methods). The addiction-rf showed a narrow distribution of small effect sizes with almost all values falling close to 0. Contrastingly, the original substance-specific GWASs were characterized by larger average effects (see Supplementary Fig. 6 for shape of probability density distribution). For example, only 26% of genes associated with PTU showed effect sizes as close to the mean threshold of the probability distribution as the addiction-rf did. These findings suggest that the addiction-rf is characterized by greater polygenicity than specific substances.

Transcriptome-wide association and drug repurposing

A transcriptome-wide association study (TWAS)[25] of the addiction-rf using multiple tissues simultaneously from GTEx in MetaXcan (Methods) identified 35 genes in 13 brain regions (Fig. 2; Supplementary Table 15). Gene-set analysis using FUMA[23] revealed that these genes were enriched for gene sets and pathways related to neural cells and T-cell processes (Supplementary Fig. 7; Supplementary Table 16). TWASs with PsychENCODE data found 29 significantly associated genes and 11 genes that overlapped with those identified in the GTEx analysis(*AMT*, *DALRD3*, *GPX1*, *KLHDC8B*, *NCKIPSD*, *NICN1*, *P4HTM*, *PPP6C*, *RHOA*, *SNX17*, *WDR6*; Fig. 2). Linking transcriptome-wide patterns from our GTEx MetaXcan analysis to perturbagens that cross the blood–brain barrier from the Library of Integrated Network-Based

https://doi.org/10.1038/s44220-023-00034-y

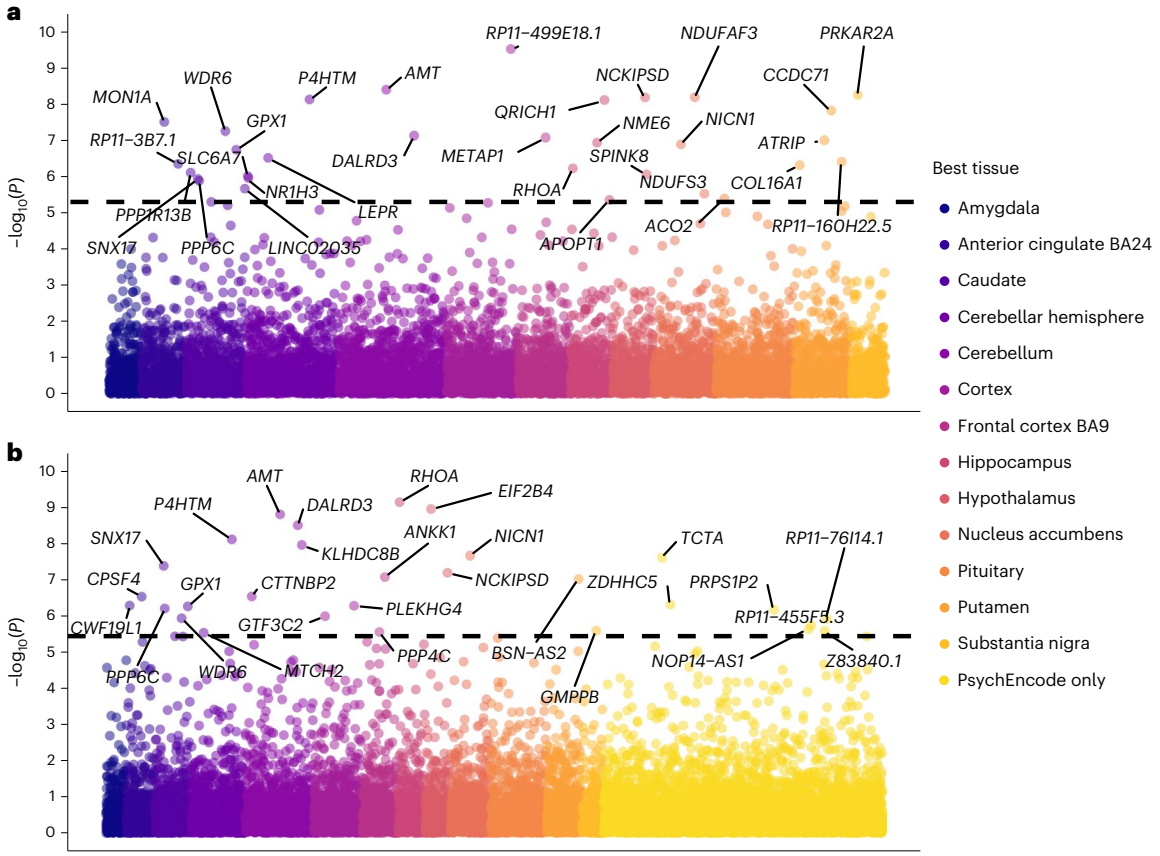

Fig. 2 | Manhattan plot of the transcriptome-wide association study results for addiction-rf. a,**b**, TWAS of the addiction-rf, plotted as a Manhattan plot. The analyses in **a** were conducted in S-MultiXcan with GTeX v8 data. The analysis in **b** was run using S-PrediXcan with weights trained from PsychENCODE. The *y*-axis is presented as $-\log_{10}(P)$, the colour of the data point represents the tissue in which correlation between gene expression and outcome was the highest. The dotted black line represents Bonferroni-corrected TWAS significance of a two-sided test (**a**, 9,944 genes, $P_{Bonferroni} = 5 \times 10^{-6}$ and the line is at 5.3; **b**, 13,850 genes, $P_{Bonferroni} = 3.6 \times 10^{-6}$, line is at 5.4).

Cellular Signatures (LINCS)[26] database identified 104 medications approved by the US Food and Drug Administration (FDA) that reverse the addiction-rf transcriptional profile (Supplementary Table 17). Medications currently used to treat SUDs (for example, varenicline for smoking cessation), other psychiatric conditions (for example, reboxetine for depression) as well as those used for other purposes (for example, mifepristone is currently used for pregnancy termination and is currently under clinical investigation for treating AUD; riluzole is a treatment for amyotrophic lateral sclerosis) were identified.

Linkage disequilibrium score regression and genetic correlations

After Bonferroni correction ($P < 0.05/1{,}547 = 3.20 \times 10^{-5}$), the addiction-rf was genetically correlated with 251 phenotypes (Fig. 3; Supplementary Table 18). Notably, 38 of these (15%) were somatic diseases linked to specific substances (for example, lung cancer with tobacco and pain-related conditions with opioids). As expected, we found significant genetic correlations (rG) between the addiction-rf and serious, transdiagnostic psychopathological behaviours, including suicide attempt (rG = 0.62, $P = 2.89 \times 10^{-33}$) and self-medication (for example, using non-prescribed drugs or alcohol for anxiety, rG = 0.64, $P = 3.18 \times 10^{-6}$). The addiction-rf was correlated with, but remained separable based on 95% confidence intervals (rG = 0.63 ± 0.037, $P = 2.33 \times 10^{-231}$), from an externalizing factor[27] that included similar indices of problematic substance use and behavioural measures.

Latent causal variable analysis

We used MASSIVE to conduct latent causal variable (LCV)[28] analyses on the same 251 phenotypes significant in our genetic correlation analyses (Supplementary Table 19). After multiple corrections ($P = 0.05/250 = 1.98 \times 10^{-4}$), the only significant causal processes were medication codes. Specifically, addiction-rf was estimated as a potential risk factor for "Medication for cholesterol, blood pressure or diabetes: cholesterol lowering medication" (genetic causality proportion = −0.739(0.078), $P = 4.51 \times 10^{-21}$), "treatment/medication code: atorvastatin" (genetic causality proportion = −0.373(0.050), $P = 7.93 \times 10^{-14}$) and "Medication for cholesterol, blood pressure, diabetes, or take exogenous hormones: cholesterol lowering medication" (genetic causality proportion = −0.315(0.071), $P = 8.31 \times 10^{-6}$). The negative genetic causality proportion estimates suggest a causal role of addiction on physical disease (addiction-rf is trait 2 in all instances).

Polygenic risk score analyses

PRS analyses with measures addiction and SUDs. In the independent Yale–Penn 3 sample[16] (European ancestry, $N = 1{,}986$), the addiction-rf PRS was significantly associated with a phenotypic factor loading on several SUDs ($P < 0.001$), polysubstance use disorder (two or more SUDs; $P < 2 \times 10^{-16}$), and each individual SUD (DSM-IV[29]: TD, cocaine use disorder (CoUD), AUD, CUD and OUD (all $P < 7.71 \times 10^{-6}$; Fig. 4; Supplementary Table 20). Nagelkerke's R^2 values ranged from 2.4% for CUD to 5.9% for TD, and 6.6% for a phenotype similar to the addiction-rf that represents phenotypic commonality across AUD, CUD, OUD, TD and CoUD. Odds ratios varied from 1.41 for CUD to 1.73 for OUD.

https://doi.org/10.1038/s44220-023-00034-y

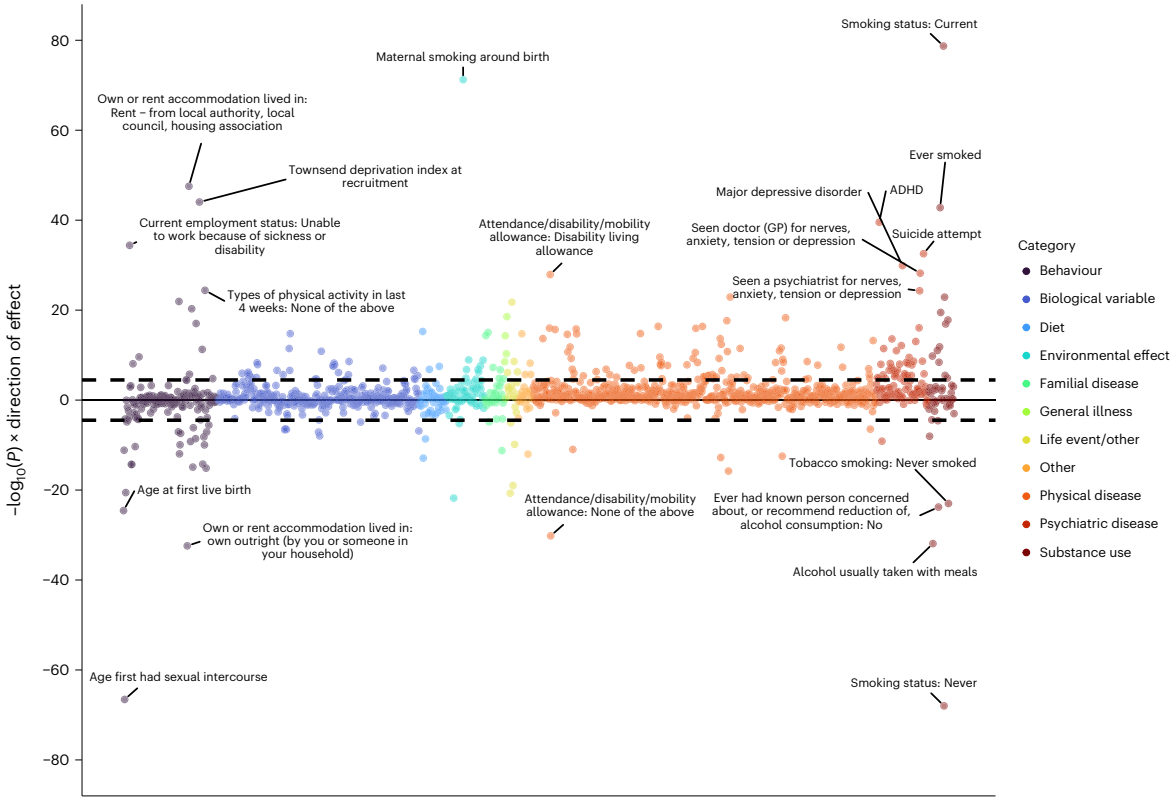

Fig. 3 | PheWAS of genetic correlations using MASSIVE. Genetic correlations between 1,547 traits and the addiction-rf, calculated in MASSIVE, mapped by their statistical significance ($-\log_{10}(P)$ on the *y*-axis), and broad category. **The top 20 correlations are annotated**; all results can be found in the Supplementary Results. The black dashed line represents Bonferroni significance for association of a two-sided test ($P_{Bonferroni} = 0.05/1{,}574 = 3.232 \times 10^{-5}$).

Phenome-wide association studies in electronic health records data. In the BioVU sample (European ancestry, $N = 66{,}914$)[30], the addiction-rf *PRS* was associated with SUDs ($P = 3.31 \times 10^{-29}$; Supplementary Fig. 8), various types of substance involvement (for example, tobacco use disorder $P = 9.79 \times 10^{-24}$, alcoholism (so named in EHR, we note the term 'alcohol use disorder' is more appropriate), $P = 1.12 \times 10^{-21}$), chronic airway obstruction ($P = 4.99 \times 10^{-10}$) and several psychiatric disorders, with the strongest being bipolar disorder ($P = 2.44 \times 10^{-11}$). Controlling for any SUD diagnosis to account for causal effects found similar associations with 'alcoholism', mood disorders, respiratory disease and heart disease (Supplementary Fig. 9a). Controlling for tobacco use disorder diagnosis did not significantly modify associations (Supplementary Fig. 9b).

Behavioural phenotypes in substance-naive children. Among 4,491 substance-naive children aged 9–10 years who completed the baseline session of the Adolescent Brain and Cognitive Development (ABCD) Study[31], the addiction-rf PRS was positively correlated (after Bonferroni correction) with Behavior Activation System Scale (BAS) fun-seeking (an aspect of externalizing behaviour; $P = 2.09 \times 10^{-5}$), family history of drug addiction ($P = 7.04 \times 10^{-7}$), family history of hospitalization due to mental health concerns (including suicidal behaviour; $P = 4.64 \times 10^{-6}$), childhood externalizing behaviours (for example, antisocial; $P = 1.62 \times 10^{-5}$), childhood thought problems ($P = 3.51 \times 10^{-6}$), sleep duration ($P = 1.52 \times 10^{-7}$), parental externalizing and substance use behaviours (for example, prenatal tobacco exposure; $P = 2.87 \times 10^{-11}$), maternal pregnancy characteristics (for example, urinary tract infection during pregnancy, $P = 2.70 \times 10^{-7}$), socioeconomic disadvantage (for example, child's neighbourhood deprivation; $P = 9.84 \times 10^{-7}$) and child's likeliness to play sports ($P = 2.80 \times 10^{-6}$) (Supplementary Fig. 10; Supplementary Table 21 for results from all phenotypes and Supplementary Table 23 for measure inclusion criteria).

Discussion

We found 17 genomic loci significantly associated with addiction-rf, and 47 substance-specific loci. Post-hoc fine-mapping, annotation, and exploratory drug repurposing analyses highlight the potential therapeutic relevance of the discovered loci. The addiction-rf PRS was associated with many medical conditions characterized by high morbidity and mortality rates, including psychiatric illnesses, self-harming behaviours, and somatic diseases that could be consequences of chronic substance use (for example, chronic airway obstruction) or precursors to heavy substance use (for example, chronic pain). Finally, in a sample of drug-naive children, the addiction-rf PRS was correlated with parental substance use problems and externalizing behaviour.

Our analyses suggest that the regulation or modulation of dopaminergic genes, rather than variation in dopaminergic genes themselves, is central to general addiction liability. *DRD2* was the top gene signal, which was mapped via chromatin refolding, suggesting a regulatory mechanism. The role of striatal dopamine in positive drug reinforcement is well established[32]. *DRD2* plays a role in reward sensitivity and may also be central to executive functioning[33]—the interplay of reward and cognition is likely relevant throughout the course of addiction. These complementary observations reinforce the role of dopamine signalling in addiction[32].

Other regulatory effects on dopaminergic pathways were supported by the signal at *PDE4B*, which has been implicated in prior GWASs of disinhibition traits[27]. The phosphodiesterase (PDE) system has been proposed as a dopaminergic regulation mechanism[34]. Furthermore, animal studies suggest that the PDE system is associated with downregulation of drug-seeking behaviours across opioids, alcohol and psychostimulants[35]. Notably, The *PDE4B* antagonist, ibudilast, has been shown to reduce heavy drinking among patients with AUD[36,37] and also shown to reduce inflammation in methamphetamine use disorder[38], and was significant in our drug repurposing analysis.

https://doi.org/10.1038/s44220-023-00034-y

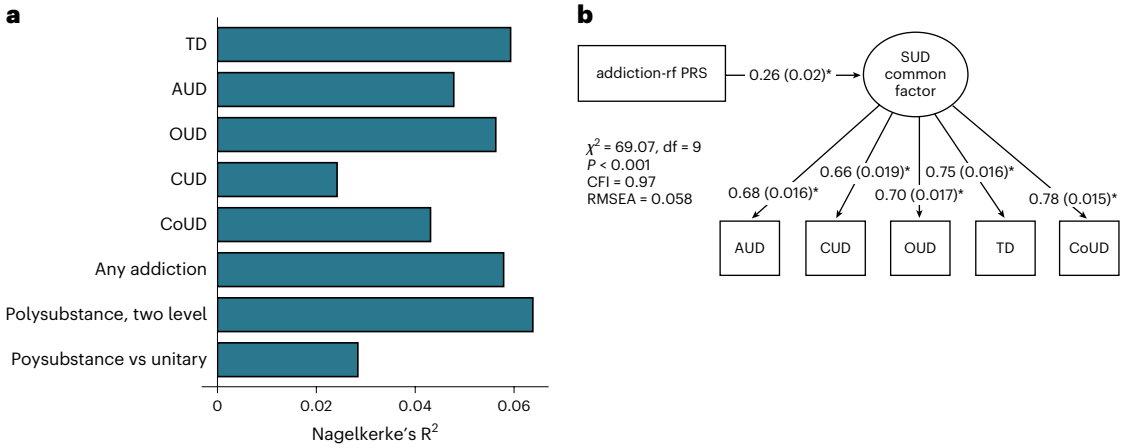

Fig. 4 | Polygenic risk score prediction in Yale–Penn 3. a, PRS of the addiction-rf predicts lifetime AUD, CUD, OUD, TD and CoUD, and variables representing more than one lifetime SUD diagnosis versus no SUDs diagnosis (polysubstance use disorder, two level), more than one lifetime diagnosis versus one lifetime diagnosis (polysubstance versus unitary), as well as any SUD diagnosis (any addiction) in an independent sample (Yale–Penn 3; $N = 1{,}986$ individuals of European genetic ancestry). **b**, The addiction-rf PRS was associated with a comparable phenotypic SUD common factor in the Yale–Penn 3 sample. Analyses control for age, sex and 10 genetic principal components of ancestry; all path estimates are fully standardized. *, Estimates were significant at $P < 0.001$ of a two-sided test (LAVAAN does not report P-values lower than 0.001). CFI, comparative fit index; RMSEA, root mean square error of approximation.

The addiction-rf PRS was associated with general and specific SUD liabilities in an independent sample. The addiction-rf PRS predicted ~6% of OUD variance, which is nearly half the total SNP-heritability of OUD[16]. The addiction-rf PRS also predicted variance in cocaine use disorder (CoUD); as CoUD was not included in the development of the addiction-rf (due to a lack of a well-powered CoUD GWASs), these findings highlight the generalizability of the addiction-rf beyond alcohol, tobacco, cannabis and opioids.

Substance-specific genetic signals fell primarily into three broad categories: drug-specific metabolism (for example, *ADH1B* for PAU), drug receptors (for example, *CHRNA5* for PTU, *OPRM1* for OUD) and general neurotransmitter mechanisms (for example, *CACNA1A* for CUD). Surprisingly, even after accounting for the addiction-rf, dopaminergic genes (*DBH* and *PDE1C* in particular) were implicated in substance-specific effects for tobacco (PTU). In contrast, CUD-specific genes did not include well-studied receptor targets (for example, *CNR1*) or metabolic mechanisms (for example, cytochrome P450 genes).

The current addiction-rf is distinct from recent genetic factors[21,27,39] that were based upon analyses of SUDs with other substance use, psychiatric and behavioural traits. We focus on SUDs rather than measures of substance use or other externalizing traits, which prior data indicate have differing aetiologies and relationships with psychiatric health[9,40,41]. Our study also parses substance-general (that is, addiction-rf) and substance-specific loci. This approach distinguishes the addiction-rf from other genetic factors that include substance use measures. For example, despite genetic overlap between the addiction-rf and a recent index of externalizing behaviours (rG = 0.63)[27], a significant portion of the variance in the addiction-rf was distinct.

Our analyses highlight the robust genetic association of the addiction-rf with serious mental and somatic illness. The addiction-rf PRS was more strongly associated with using drugs to cope with internalizing disorder symptoms (anxiety, depression; rG = 0.60–0.62) than with the individual psychiatric traits and disorders themselves (rG = 0.3), suggesting that genetic correlations between SUDs and mood disorders may partially be attributable to a predisposition to use substances to alleviate negative mood states ('self-medication')[42].

The phenome-wide association study (PheWAS) provided insight into potentially complex mechanisms of genetic liability to environmental pathways of risk. In addition to indices of socioeconomic status (SES), the addiction-rf was correlated with maternal tobacco smoking during pregnancy and with attention deficit hyperactivity disorder, in line with evidence that effects ascribed to the prenatal environment may also be mediated by the inheritance of risk loci[43,44]. The addiction-rf PRS was associated with a family history of serious mental illness, which likely represents an amalgam of genetic and environmental vulnerability[45]. Finally, disability and SES were also associated with polygenic risk, further supporting the association between environmental risk factors and common genetic effects on SUD liability[9,41,46].

This study has limitations. First, our GWAS in individuals of African ancestry had few discoveries, underscoring the need for systematic data collection on SUDs in globally representative populations. Still, we chose to analyse and present these data as their exclusion only furthers disparities in genetic discoveries. Second, although we discovered many loci, they accounted for only a small proportion of the total variance. More samples, particularly from diverse populations, and the integration of rarer variants are needed to discover the biological pathways that fall below genome-wide significance or are missed in GWAS. Finally, despite interesting associations between our PRS and SUDs, our findings do not apply to prognostication of future disease risk.

Conclusion

A common and highly polygenic genetic architecture underlies multiple SUDs, a finding that merits integration into medical knowledge on addictions.

Methods

Summary statistics from each SUD-related GWAS

Summary statistics from the largest available discovery GWAS were used to represent genetic risk for each construct. These include four measures of problematic substance use or SUD: (1) PAU[8], (2) PTU[12,13,18], (3) CUD[15], (4) OUD[16]. All GWAS summary statistics were filtered to retain variants with minor allele frequencies >0.01 and INFO score >0.90 for GSCAN[12] and PGC[15] and INFO score >0.70 for the MVP[8,16].

For the current cross-trait GWAS, we maintained the same quality control (QC) metrics and only analysed SNPs that were present in all four input GWASs, that is, variants that passed QC thresholds at all levels, resulting in 3,513,381 SNPs in samples of European ancestry and 5,303,643 SNPs in samples of African American ancestry. The linkage disequilibrium (LD) scores used for the genomic structural equation modelling (GenomicSEM)[47] were estimated in the European ancestry

samples only using the 1000 Genomes European data[48]. We restricted analyses to HapMap3 SNPs[49] as these tend to be well imputed and produce accurate estimates of heritability. We used the effective *N*, that was estimated for each GWAS[50]. For traits with a binary distribution, the effective sample size for an equivalently powered case-control study under a 50–50 case control balance was estimated using the equation: $N_{effective} = 4/((1/N_{case}) + (1/N_{control}))$[51],where *N* represents the sample size. Continuous and quasi-continuous traits used the given *N* or if from MTAG, the equation $N_{effective} = ((Z/\beta)^2)/(2 \times MAF \times (1 - MAF))$, where MAF is the minor allele frequency, *Z* is the *z*-score of the effect size and β is the beta of the effect size[8], to approximate an equivalently powered GWAS of a single trait. Effective *N* values ranged from 46,351 (CUD) to 300,789 (PAU) and are described for each substance-specific GWAS in the Results. Individual GWAS details can be found in the Supplementary Methods.

Genome-wide analyses in European ancestry

We conducted a GWAS of a unidimensional addiction risk factor (addiction-rf) underlying the genetic covariance among PAU, PTU, CUD and OUD by applying GenomicSEM[47] to these European ancestry summary statistics. GenomicSEM conducts genome-wide association analyses in two stages. First, a multivariate version of LD score regression is used to estimate the genetic covariance matrix among all GWAS phenotypes, which is then combined with each individual SNP to calculate SNP-specific genetic covariance matrices. Second, these matrices are then used to estimate the SEM using the lavaan package in R[52]. Variable and unknown extents of sample overlap across contributing GWASs are automatically accounted for in the estimation procedure. The unifactor model fit the data well[53] ($\chi^2(1) = 0.017$, $P = 0.896$, CFI = 1, SRMR = 0.002; residual $r = 0.51$, $P = 0.016$; Supplementary Fig. 1; see also our prior work[18] and Methods).

As the sample size of summary data derived from African American samples (*N* range = 9,835–56,648) was not sufficient for LD score[54] analyses, we used ASSET[24] to conduct the addiction-rf GWAS, as opposed to GenomicSEM, as described in the subsequent **ASSET** section below.

ASSET trans-ancestry analyses

ASSET[24] was used to identify pleiotropic (that is, SNPs that show associations with more than one SUD) and substance-specific (that is, SNPs only associated with a single SUD) SNPs within the European and African American ancestry samples (in addition to GenomicSEM in Europeans). ASSET was used in our African American ancestry addiction-rf GWAS because the sample size was not sufficient for the genomic structural equation modelling (SEM) approach used in the European addiction-rf GWAS. As a result, there are important differences in the primary addiction-rf GWAS and GWAS run in ASSET. First, the ASSET-based addiction-rf GWAS contains SNPs that may influence two, three, or all four individual SUDs, while the GenomicSEM-based addiction-rf GWAS in European ancestry samples includes SNPs associated with a common factor across all included SUDs. We used ASSET to identify pleiotropic SNPs in the European ancestry sample to facilitate method-consistent cross-ancestry meta-analysis GWAS (see subsequent 'Cross-ancestry meta-analysis' section below) and cross validate primary GenomicSEM results.

ASSET does not leverage the genetic correlation to identify variants of interest (as GenomicSEM does); instead, subset searches scaffold effects into pleiotropic and non-pleiotropic variants based on effect size and standard error derivations that estimate the degree to which the SNP–trait association is due to pooled effects across the phenotypes, versus a single phenotype driving variant association. Loci were designated as substance specific when they were significantly associated with only one SUD. Because ASSET does not automatically account for sample overlap; we used the linkage disequilibrium score regression intercept (LDSC) to adjust for overlap within the European ancestry ASSET covariance term.

Cross-ancestry meta-analysis

We conducted a cross-ancestry meta-analysis of ASSET-derived (to maintain analytic consistency) European and African ancestry addiction-rf summary statistics. First, SNPs with evidence of SUD pleiotropy (that is, effects on two, three, or all four SUDs, including different sets of SUDs in each ancestry) in both ancestral groups were extracted. SNPs with evidence of cross-ancestral heterogeneity (that is, Cochran's *Q* statistic $<5 \times 10^{-8}$) were removed, leaving 317,447 SNPs. A meta-analysis in METASOFT[55] using a random-effects meta-analysis with ancestry group as a random effect was used to identify cross-ancestral effects. We report the random-effects beta and *P*-value as cross-ancestry effects.

Substance specific genetics in European ancestry individuals

To validate substance-specific SNPs, we used ASSET for discovery of these variants and, in the European ancestry GWAS, also examined Q-SNP results derived from GenomicSEM. Q-SNP[14] indexes violation of the null hypothesis that a SNP acts on a trait entirely through a common factor (for example, the addiction-rf). For example, if a SNP has a particular effect on one SUD trait (such as SNPs in *CHRNA5* influencing PTU), then it should have significant Q-SNP statistics because it violates the assumption that its effect on PTU is via the addiction-rf. We identified Q-SNPs by estimating the association between each SNP and the addiction-rf. Then, we fit a model where the SNP predicted the indicators underlying the addiction-rf, that is, PAU, PTU, CUD, OUD. We compared the χ^2 difference statistic between the two models; those with significant decrement of fit (χ^2 for Δd.f. = 4) in the model where the SNP predicted the addiction-rf alone relative to the SNP predicting the indicators themselves was considered a significant Q-SNP above GWS (that is, Q $P < 5 \times 10^{-8}$). SNPs with significant Q-SNP statistics were removed from the addiction-rf summary statistics for all post-hoc analyses, including fine-mapping, gene-based tests, transcriptome-wide association analyses, LD score genetic correlations and PRS analyses.

Q-SNP analysis also identified several SNPs that appeared to be specific to a single substance. However, as Q-SNP cannot be used for precise identification of substance-specific (trait-specific) SNPs, we relied on ASSET analyses (with a one-sided *P*-value), to identify the subset of SNPs with effects (at GWS, $P < 5 \times 10^{-8}$) limited to only one SUD-related trait (for example, PAU-specific vs. PAU common with OUD). It is worth noting that the ASSET analysis determines both common addiction and substance specific SNPs. Here we would like to note that the common addiction SNPs from ASSET results were used for our cross-ancestry analysis, while specific SNPs in our results are described seperately for each population.

Post-hoc analyses of European ancestry GWAS results

Estimation of expected SNP effect sizes. We estimated the distribution of genetic effect-sizes of the addiction-rf (GenomicSEM) and the four input GWASs (PAU, PTU, CUD, OUD) using genetic effect-size distribution inference from summary-level data (GENESIS). GENESIS is a likelihood-based approach[56]. In this approach, GWAS summary statistics and an external panel of LD (in our case, the 1000 Genomes Phase 3 reference panel) are used to estimate a projected distribution of SNP effect sizes. A flexible normal mixture model based on the number of tagged SNPs and LD scores is estimated. A three-component model is fit, where SNP effect sizes are estimated to belong to one of three components based on bins of effect sizes (large, medium and small). If the distribution of SNPs is multivariate normal, the estimation of the SNPs with large and medium effect sizes can be done via their independent effect sizes. The third component represents SNPs with null and small effect sizes, and these should follow a similar distribution. Therefore, this model reweights SNPs and generates a projected distribution of effect sizes, and from this projection, we can draw conclusions about the distribution of effect sizes[54].

Biological characterization

FUMA[23] was used for post-hoc bioinformatic analyses of our five GWASs (that is, the addiction-rf (from GenomicSEM), PAU-specific, PTU-specific, CUD-specific, OUD-specific (from ASSET) loci) in European ancestry samples and to determine lead and independent variants. Within FUMA, gene-based tests and gene-set enrichment were conducted via MAGMA[57]; gene annotation, and identification of SNP-to-gene associations via eQTLs and/or chromatin interactions (via Hi-C data) in PsychENCODE[58] and Roadmap Epigenomics tissues for prefrontal cortex, hippocampus, ventricles and neural progenitor cells[59,60]. For each specific SUD, the distribution of *P*-values included all non-pleiotropic SNPs identified by ASSET (that is, SNPs only associated with a single SUD, *n* SNP CUD-specific = 312,661, *n* SNP PTU-specific = 560,983, *n* SNP PAU-specific = 193,647, *n* SNP OUD-specific = 425,665).

Fine-mapping with SusieR. We fine-mapped the association statistics of the four phenotypes (the addiction-rf, PAU-specific, PTU-specific, CUD-specific; OUD-specific only had one significant locus, and that locus has a known mechanism of effect) that had more than one GWS SNP in a 1 Mb region around the lead SNP to determine the 95% credible set using susieR[61] with at most 10 causal variants (this analysis reduces the total number of SNPs at a lead genome-wide signal to those that can credibly be considered as causal SNPs). The credible set reports include the likelihood of being a causal variant; the marginal posterior inclusion probability (PIP) ranges from 0 to 1, with values closer to 1 being most likely causal.

Transcriptome-wide association analysis. We conducted two transcriptome-wide analyses. First, we used MetaXcan/S-MultiXcan[38] to conduct a cross-tissue analysis of all brain tissues in the GTEx v8 data[37]. S-MultiXcan returns a broad *z*-score across all tissues in the model, along with the top and lowest scores at each tissue. S-MultiXcan combines information across individual tissues, which improves the power for discovery by reducing the multiple correction burden. It also produces *z*-score and *P*-values for top-associated tissues. Second, we also used S-PrediXcan[62] to predict transcription using the weights trained on psychiatric cases versus controls transcriptional differences from the frontal and temporal cortex using the PsychENCODE[63] dataset. As these data were very densely sampled for psychiatrically relevant traits, it serves to complement the relatively healthy GTEx sample.

Drug repurposing

Our technique for drug signature matching used data from the LINCS L1000 database[64]. The LINCS L1000 database catalogues in vitro gene expression profiles (signatures) from thousands of compounds in over 80 human cell lines (level 5 data from phase I: GSE92742 and phase II: GSE70138)[26]. We selected compounds that were currently FDA approved or in clinical trials (via https://clue.io/repurposing#download-data; updated 24 March 2020). Our analyses included signatures of 829 chemical compounds (590 FDA approved, 239 in clinical trials) in five neuronal cell-lines (NEU, NPC, MNEU.E, NPC.CAS9 and NPC.TAK), a total of 3,897 signatures were present as not all compounds were tested in all cell lines in the LINCS dataset.

In vitro medication signatures were matched with the addiction-rf signatures from the transcriptome-wide association analyses (conducted using S-MultiXcan)[25,62] via multi-level meta-regression. We computed weighted (by its proportion of heritability explained ($h_{\text{MULTI-XCAN}}^2$)) Pearson correlations between transcriptome-wide brain associations and in vitro L1000 compound signatures using the metafor package in R[65]. We treated each L1000 compound as a fixed effect incorporating the effect size (r_{weighted}) and sampling variability ($\text{se}_{\text{r\_weighted}}^2$) from all signatures of a compound (for example, across all time points, cell lines and doses). Analyses included time since perturbagen exposure as a random effect. Only genes that were Bonferroni significant in the S-PrediXcan analysis (transcriptome-wide correction = 0.05/14,389 = 3.48×10^{-6}) were entered into the model. We only report those perturbagens that were associated after Bonferroni correction (perturbagen correction = 0.05/3,897 = 1.28×10^{-5}).

PRS analyses in Yale–Penn

Yale–Penn 3. The Yale–Penn[16,66] sample includes 11,332 genotyped and phenotyped individuals recruited across three phases (that is, Yale–Penn 1, Yale–Penn 2 and Yale–Penn 3) based on the time of recruitment and genotyping array used. All cohorts were ascertained via recruitment at substance use treatment centres or targeted advertisements for genetic studies of cocaine, opioid and alcohol dependence, resulting in a sample highly enriched for problematic substance use, as well as control subjects and relatives. All participants were assessed using the Semi-Structured Assessment for Drug Dependence and Alcoholism (SSADDA)[67]. Analyses based on Yale–Penn 1 and 2 have been published previously[66], and were used in the discovery sample of the present study. Here, we used data from Yale–Penn 3[16] for replication analyses and as a target sample for PRS analyses; the Yale–Penn 3 sample is independent from our discovery GWASs. Yale–Penn 3 comprises 3,026 genotyped and phenotyped Americans of European (EUR; *N* = 1,986) and African (AFR; *N* = 1,040) ancestry passing standard QC. Genotyping was performed at the Gelernter lab at Yale University using the Illumina Multi-ethnic Global Array containing 1,779,819 markers, followed by genotype imputation using Minimac3[68] and the Haplotype Reference Consortium reference panel[69] as implemented on the Michigan imputation server (https://imputationserver.sph.umich.edu).

For the present analysis, only Yale–Penn 3 EUR subjects (*N* = 1,986) were included. DSM-IV[29] substance abuse and dependence diagnoses (combined as abuse or dependence to represent use disorder) based on SSADDA assessments were used to determine case and control status for AUD, CUD, CoUD, TD and OUD. Of the 1,986 EUR subjects, 42.4% met criteria for AUD (*N* = 843), 25.9% met criteria for CUD (*N* = 515), 25.3% met criteria for CoUD (*N* = 503), 31% met criteria for TD (*N* = 615) and 22.6% met criteria for OUD (*N* = 448). The mean age of Yale–Penn 3 EUR subjects is 41.5 years (s.e. = 15.1) and 51.5% are female (*N* = 1,023).

We calculated the addiction-rf PRS using the PRS-CS auto approach[70]. This method assumes a general distribution of effect sizes across the genome, and then reweights SNPs based on this assumption, their effect size in the original GWAS, and their LD; weights for every SNP were then summed to create a final score. PRS were associated with phenotypes (OUD, TD, CUD, AUD, CoUD) in Yale–Penn 3 via a logistic regression controlling for the first 10 ancestral principal components, age, sex and age by sex. PRS were scaled to unit variance. These logistic regression analyses were also examined for the following contrasts: (1) those with any SUD (*n* = 985) versus those with no SUD (*n* = 1,001), to represent 'any SUD'; (2) those with at least two SUDs (*n* = 729) versus those with less than two (including zero) SUDs (*n* = 1,257) to represent 'polysubstance use disorder'; and (3) those with at least two SUDs (*n* = 729) versus those with one SUD (*n* = 256) to represent polysubstance use disorder within those with SUD. The association between the addiction-rf PRS and the SUD common factor was estimated with lavaan[52] where the common factor loaded on the five SUDs.

Genetic correlations and latent causal variable modelling

To examine phenotypes that were genetically correlated with the addiction-rf, we calculated genetic correlations using LD score regression[54,71] through the MASSIVE pipeline[72], which conducts LD score regression[13,46] and Latent Causal Variable Analysis[28] on 1,547 summary statistics for various phenotypic traits, including a mixture of ICD codes and self-reported traits from the UK Biobank and publicly available meta-analyses from GWAS consortia.

Phenome-wide association studies

PheWASs in adult samples. As MASSIVE includes a fairly sparse set of diagnoses (not all ICD codes are available) for genetic correlation

analyses, we conducted additional and theoretically relevant PheWASs using the addiction-rf PRS. We used EHR data for 66,914 genotyped individuals of European ancestry from the Vanderbilt University Medical Center biobank (BioVU)[30]. BioVU is a repository of leftover blood samples (~240,000 samples) from clinical testing, that are sequenced, de-identified and linked to clinical and demographic data. Genotyping and QC of this sample have been described elsewhere[30]. The addiction-rf PRS was used to predict 1,335 diseases in a logistic regression model, controlling for median age on record, reported gender and first 10 genetic ancestral principal components. For an individual to be considered a case, they were required to have two separate ICD codes for the index phenotype, and each phenotype needed at least 100 cases to be included in the analysis. A Bonferroni-corrected phenome-wide significance threshold of $0.05/1{,}335 = 3.7 \times 10^{-5}$ was used[73].

ABCD PheWAS of phenotypes collected in childhood. To identify phenotypes that were associated with the addiction-rf before the onset of regular substance use, we used data from the ABCD Study (release 2.0 for genomic data and 3.0 for phenotypes) to conduct a phenome-wide association analysis of behavioural, social and environmental phenotypes in adolescence. The ABCD Study is an ongoing multi-site longitudinal study of child health and development (Methods)[31,74]. Children ($N = 11{,}875$; including twins and siblings) ages 8.9–11 years were recruited from 22 sites across the United States to complete the ABCD Study baseline assessment. We restricted our sample to participants of genomically confirmed European ancestry (based on principal components) who were not missing any covariate measures ($N = 4{,}490$).

PRS were generated using the PRS-CS software package[70] consistent with our other (that is, Yale–Penn 3, BioVU) PRS analyses described above. Associations between the addiction-rf PRS and phenotypes were estimated using mixed-effects models in the lme4[75] package in R. PRS were scaled to unit variance. Family ID and site were included as random effects to account for non-independence of measurement associated with relatedness and scanner/site. We controlled for the first 10 ancestral principal components, age, sex and age by sex. We used a Bonferroni-corrected phenome-wide significance threshold of $0.05/1{,}480 = 3.38 \times 10^{-5}$; all results are presented in the Supplementary Table 21.

Reporting summary

Further information on research design is available in the Nature Portfolio Reporting Summary linked to this article.

Data availability

The MVP summary statistics were obtained via an approved dbGaP application (phs001672.v4.p1). For details on the MVP, see https://www.research.va.gov/mvp/ and ref. [76]. This research is based on data from the MVP, Office of Research and Development, Veterans Health Administration, and was supported by the Veterans Administration Cooperative Studies Program award G002.

Publicly available data were also downloaded from the psychiatric genomics consortium (https://www.med.unc.edu/pgc/) and the GSCAN consortium (https://conservancy.umn.edu/handle/11299/201564).

The datasets used for the BioVU analyses described were obtained from Vanderbilt University Medical Center's biorepository, which is supported by numerous sources: institutional funding, private agencies and federal grants. These include the National Institutes of Health-funded Shared Instrumentation grant S10RR025141; and Clinical and Translational Science Awards (CTSA) grants UL1TR002243, UL1TR000445 and UL1RR024975. Genomic data are also supported by investigator-led projects that include U01HG004798, R01NS032830, RC2GM092618, P50GM115305, U01HG006378, U19HL065962 and R01HD074711; and additional funding sources listed at https://victr.vumc.org/biovu-funding/.

Data from Yale–Penn 1 are available through dbGAP accession no phs000425.v1.p1 including 1,889 African American subjects and 1,020 European-American subjects. Yale–Penn 1 data are also available through dbGAP accession no phs000952.v1.p1 including 1,531 African American subjects and 1,339 self-reported European-American subjects. Summary statistics for all Yale–Penn data are available on request to J.G. (joel.gelernter@yale.edu).

References

1. Degenhardt, L. et al. The impact of cohort substance use upon likelihood of transitioning through stages of alcohol and cannabis use and use disorder: findings from the Australian National Survey on Mental Health and Wellbeing. *Drug Alcohol Rev.* **37**, 546–556 (2018).
2. Peacock, A. et al. Global statistics on alcohol, tobacco and illicit drug use: 2017 status report. *Addiction* **113**, 1905–1926 (2018).
3. Reitsma, M. B. et al. Spatial, temporal, and demographic patterns in prevalence of smoking tobacco use and initiation among young people in 204 countries and territories, 1990–2019. *Lancet Public Health* **6**, e472–e481 (2021).
4. Odds of dying. *Injury Facts* https://injuryfacts.nsc.org/all-injuries/preventable-death-overview/odds-of-dying/ (accessed 3 December 2021).
5. Vanyukov, M. M. An eternal epidemic: 1. Why substance use problems persist. Preprint at *PsyArXiv* https://psyarxiv.com/tkm5u/ (2021).
6. Deak, J. D. & Johnson, E. C. Genetics of substance use disorders: a review. *Psychol. Med.* **51**, 2189–2200 (2021).
7. Gelernter, J. & Polimanti, R. Genetics of substance use disorders in the era of big data. *Nat. Rev. Genet.* **22**, 712–729 (2021).
8. Zhou, H. et al. Genome-wide meta-analysis of problematic alcohol use in 435,563 individuals yields insights into biology and relationships with other traits. *Nat. Neurosci.* **23**, 809–818 (2020).
9. Mallard, T. T. et al. Item-level genome-wide association study of the alcohol use disorders identification test in three population-based cohorts. *Am. J. Psychiatry* **179**, 58–70 (2021).
10. Walters, R. K. et al. Transancestral GWAS of alcohol dependence reveals common genetic underpinnings with psychiatric disorders. *Nat. Neurosci.* **21**, 1656–1669 (2018).
11. Kranzler, H. R. et al. Genome-wide association study of alcohol consumption and use disorder in 274,424 individuals from multiple populations. *Nat. Commun.* **10**, 1499 (2019).
12. Liu, M. et al. Association studies of up to 1.2 million individuals yield new insights into the genetic etiology of tobacco and alcohol use. *Nat. Genet.* **51**, 237–244 (2019).
13. Hancock, D. B. et al. Genome-wide association study across European and African American ancestries identifies a SNP in DNMT3B contributing to nicotine dependence. *Mol. Psychiatry* **23**, 1911–1919 (2018).
14. Quach, B. C. et al. Expanding the genetic architecture of nicotine dependence and its shared genetics with multiple traits. *Nat. Commun.* **11**, 5562 (2020).
15. Johnson, E. C. et al. A large-scale genome-wide association study meta-analysis of cannabis use disorder. *Lancet Psychiatry* **7**, 1032–1045 (2020).
16. Zhou, H. et al. Association of OPRM1 functional coding variant with opioid use disorder: a genome-wide association study. *JAMA Psychiatry* **77**, 1072–1080 (2020).
17. Kendler, K. S. et al. Recent advances in the genetic epidemiology and molecular genetics of substance use disorders. *Nat. Neurosci.* **15**, 181–189 (2012).
18. Hatoum, A. S. et al. The addiction risk factor: a unitary genetic vulnerability characterizes substance use disorders and their associations with common correlates. *Neuropsychopharmacology* **47**, 1739–1745 (2022).

19. Abdellaoui, A., Smit, D. J. A., van den Brink, W., Denys, D. & Verweij, K. J. H. Genomic relationships across psychiatric disorders including substance use disorders. *Drug Alcohol Depend.* **220**, 108535 (2021).
20. Waldman, I. D., Poore, H. E., Luningham, J. M. & Yang, J. Testing structural models of psychopathology at the genomic level. *World Psychiatry* **19**, 350–359 (2020).
21. Grotzinger, A. D. et al. Genetic architecture of 11 major psychiatric disorders at biobehavioral, functional genomic and molecular genetic levels of analysis. *Nat. Genet.* **54**, 548–559 (2022).
22. Duan, Q. et al. LINCS Canvas Browser: interactive web app to query, browse and interrogate LINCS L1000 gene expression signatures. *Nucleic Acids Res.* **42**, W449–W460 (2014).
23. Watanabe, K., Taskesen, E., van Bochoven, A. & Posthuma, D. Functional mapping and annotation of genetic associations with FUMA. *Nat. Commun.* **8**, 1826 (2017).
24. Bhattacharjee, S. et al. A subset-based approach improves power and interpretation for the combined analysis of genetic association studies of heterogeneous traits. *Am. J. Hum. Genet.* **90**, 821–835 (2012).
25. Barbeira, A. N. et al. Integrating predicted transcriptome from multiple tissues improves association detection. *PLoS Genet.* **15**, e1007889 (2019).
26. Subramanian, A. et al. A next generation connectivity map: L1000 platform and the first 1,000,000 profiles. *Cell* **171**, 1437–1452 e17 (2017).
27. Karlsson Linner, R. et al. Multivariate analysis of 1.5 million people identifies genetic associations with traits related to self-regulation and addiction. *Nat. Neurosci.* **24**, 1367–1376 (2021).
28. O'Connor, L. J. & Price, A. L. Distinguishing genetic correlation from causation across 52 diseases and complex traits. *Nat. Genet.* **50**, 1728–1734 (2018).
29. Association, A. P. *Diagnostic and Statistical Manual (DSM-IV)* (American Psychiatric Press, 1994).
30. Roden, D. M. et al. Development of a large-scale de-identified DNA biobank to enable personalized medicine. *Clin. Pharmacol. Ther.* **84**, 362–369 (2008).
31. Volkow, N. D. et al. The conception of the ABCD study: from substance use to a broad NIH collaboration. *Dev. Cogn. Neurosci.* **32**, 4–7 (2018).
32. Wise, R. A. & Robble, M. A. Dopamine and addiction. *Annu. Rev. Psychol.* **71**, 79–106 (2020).
33. Hatoum, A.S. et al. Genome-wide association study shows that executive functioning is influenced by gabaergic processes and is a neurocognitive genetic correlate of psychiatric disorders. *Biol. Psychiatry* **93**, 59–70 (2022).
34. Snyder, G. L. & Vanover, K. E. PDE inhibitors for the treatment of schizophrenia. *Adv Neurobiol* **17**, 385–409 (2017).
35. Olsen, C. M. & Liu, Q. S. Phosphodiesterase 4 inhibitors and drugs of abuse: current knowledge and therapeutic opportunities. *Front. Biol.* **11**, 376–386 (2016).
36. Burnette, E. M., Ray, L. A., Irwin, M. R. & Grodin, E. N. Ibudilast attenuates alcohol cue-elicited frontostriatal functional connectivity in alcohol use disorder. *Alcohol. Clin. Exp. Res.* **45**, 2017–2028 (2021).
37. Grodin, E. N. et al. Ibudilast, a neuroimmune modulator, reduces heavy drinking and alcohol cue-elicited neural activation: a randomized trial. *Transl. Psychiatry* **11**, 355 (2021).
38. Li, M. J., Briones, M. S., Heinzerling, K. G., Kalmin, M. M. & Shoptaw, S. J. Ibudilast attenuates peripheral inflammatory effects of methamphetamine in patients with methamphetamine use disorder. *Drug Alcohol Depend.* **206**, 107776 (2020).
39. Schoeler, T. et al. Novel biological insights into the common heritable liability to substance involvement: a multivariate genome-wide association study. *Biol. Psychiatry* **93**, 524–535 (2022).
40. Sanchez-Roige, S., Palmer, A. A. & Clarke, T. K. Recent efforts to dissect the genetic basis of alcohol use and abuse. *Biol. Psychiatry* **87**, 609–618 (2020).
41. Marees, A. T. et al. Genetic correlates of socio-economic status influence the pattern of shared heritability across mental health traits. *Nat. Hum. Behav.* **5**, 1065–1073 (2021).
42. Khantzian, E. J. Addiction as a self-regulation disorder and the role of self-medication. *Addiction* **108**, 668–669 (2013).
43. Thapar, A. & Rice, F. Family-based designs that disentangle inherited factors from pre- and postnatal environmental exposures: in vitro fertilization, discordant sibling pairs, maternal versus paternal comparisons, and adoption designs. *Cold Spring Harb. Perspect. Med.* **11**, a038877 (2021).
44. Haan, E. et al. Prenatal smoking, alcohol and caffeine exposure and maternal reported ADHD symptoms in childhood: triangulation of evidence using negative control and polygenic risk score analyses. *Addiction* **117**, 1458–1471 (2021).
45. Cornelis, M. C., Zaitlen, N., Hu, F. B., Kraft, P. & Price, A. L. Genetic and environmental components of family history in type 2 diabetes. *Hum. Genet.* **134**, 259–267 (2015).
46. Wendt, F. R. et al. Multivariate genome-wide analysis of education, socioeconomic status and brain phenome. *Nat. Hum. Behav.* **5**, 482–496 (2021).
47. Grotzinger, A. D. et al. Genomic structural equation modelling provides insights into the multivariate genetic architecture of complex traits. *Nat. Hum. Behav.* **3**, 513–525 (2019).
48. Genomes Project, C. et al. A global reference for human genetic variation. *Nature* **526**, 68–74 (2015).
49. International HapMap, C. et al. Integrating common and rare genetic variation in diverse human populations. *Nature* **467**, 52–58 (2010).
50. Grotzinger, A. D., de la Fuente, J., Privé, F., Nivard, M. G. & Tucker-Drob, E. M. Pervasive downward bias in estimates of liability-scale heritability in genome-wide association study meta-analysis: a simple solution. *Biol. Psychiatry*, **93**, 29–36 (2023).
51. Turley, P. et al. Multi-trait analysis of genome-wide association summary statistics using MTAG. *Nat. Genet.* **50**, 229–237 (2018).
52. Rosseel, Y. Lavaan: an R package for structural equation modeling. *J. Stat. Software* **48**, 1–36 (2012).
53. Hu, L. & Bentler, P. M. Cutoff criteria for fit indexes in covariance strcuture analysis: conventional criteria versus new alternatives. *Struct. Equation Modell.* **6**, 1 (1999).
54. Bulik-Sullivan, B. K. et al. LD score regression distinguishes confounding from polygenicity in genome-wide association studies. *Nat. Genet.* **47**, 291–295 (2015).
55. Han, B. & Eskin, E. Random-effects model aimed at discovering associations in meta-analysis of genome-wide association studies. *Am. J. Hum. Genet.* **88**, 586–598 (2011).
56. Zhang, Y., Qi, G., Park, J. H. & Chatterjee, N. Estimation of complex effect-size distributions using summary-level statistics from genome-wide association studies across 32 complex traits. *Nat. Genet.* **50**, 1318–1326 (2018).
57. de Leeuw, C. A., Mooij, J. M., Heskes, T. & Posthuma, D. MAGMA: generalized gene-set analysis of GWAS data. *PLoS Comput. Biol.* **11**, e1004219 (2015).
58. Psych, E. C. et al. The PsychENCODE project. *Nat. Neurosci.* **18**, 1707–1712 (2015).
59. Roadmap Epigenomics, C. et al. Integrative analysis of 111 reference human epigenomes. *Nature* **518**, 317–330 (2015).
60. Ernst, J. & Kellis, M. ChromHMM: automating chromatin-state discovery and characterization. *Nat. Methods* **9**, 215–216 (2012).

61. Wang, G., Sarkar, A., Carbonetto, P. & Stephens, M. A simple new approach to variable selection in regression, with application to genetic fine mapping. *R. Stat. Soc. B* **82**, 1273–1300 (2020).
62. Barbeira, A. N. et al. Exploring the phenotypic consequences of tissue specific gene expression variation inferred from GWAS summary statistics. *Nat. Commun.* **9**, 1825 (2018).
63. Gandal, M. J. et al. Transcriptome-wide isoform-level dysregulation in ASD, schizophrenia, and bipolar disorder. *Science* **362**, eaat8127 (2018).
64. Liu, C. et al. Compound signature detection on LINCS L1000 big data. *Mol. Biosyst.* **11**, 714–722 (2015).
65. Viechtbauer, W. Accounting for heterogeneity via random-effects models and moderator analyses in meta-analysis. *Z. Psychol.* **215**, 104–121 (2007).
66. Sherva, R. et al. Genome-wide association study of cannabis dependence severity, novel risk variants, and shared genetic risks. *JAMA Psychiatry* **73**, 472–480 (2016).
67. Pierucci-Lagha, A. et al. Diagnostic reliability of the semi-structured assessment for drug dependence and alcoholism (SSADDA). *Drug Alcohol Depend.* **80**, 303–312 (2005).
68. Das, S. et al. Next-generation genotype imputation service and methods. *Nat. Genet.* **48**, 1284–1287 (2016).
69. McCarthy, S. et al. A reference panel of 64,976 haplotypes for genotype imputation. *Nat. Genet.* **48**, 1279–1283 (2016).
70. Ge, T., Chen, C. Y., Ni, Y., Feng, Y. A. & Smoller, J. W. Polygenic prediction via Bayesian regression and continuous shrinkage priors. *Nat. Commun.* **10**, 1776 (2019).
71. Bulik-Sullivan, B. et al. An atlas of genetic correlations across human diseases and traits. *Nat. Genet.* **47**, 1236–1241 (2015).
72. Cuellar-Partida, G. et al. Complex-Traits Genetics Virtual Lab: A community-driven web platform for post-GWAS analyses. Preprint at *bioRxiv* https://doi.org/10.1101/518027 (2021).
73. Carroll, R. J., Bastarache, L. & Denny, J. C. R PheWAS: data analysis and plotting tools for phenome-wide association studies in the R environment. *Bioinformatics* **30**, 2375–2376 (2014).
74. Lisdahl, K. M. et al. Substance use patterns in 9-10 year olds: baseline findings from the adolescent brain cognitive development (ABCD) study. *Drug Alcohol Depend.* **227**, 108946 (2021).
75. Bates, D., Machler, M., Bolker, B. M. & Walker, S. C. Fitting linear mixed-effects models using lme4. *J. Stat. Software* **67**, 1–48 (2015).
76. Gaziano, J. M. et al. Million Veteran Program: a mega-biobank to study genetic influences on health and disease. *J. Clin. Epidemiol.* **70**, 214–223 (2016).

Acknowledgements

The authors would like to thank Cold Harbor Labroatory for posting a preprint of this work on MedRxiv (https://www.medrxiv.org/). The authors thank Million Veteran Program (MVP) staff, researchers and volunteers, who have contributed to MVP, and especially participants who previously served their country in the military and now generously agreed to enroll in the study. **Funding:** K01 AA030083 (A.S.H.), T32 DA007261 (A.S.H.), DA54869 (A.A., J.G., H.E.), R01 DA54750 (A.A., R.B.), K02 DA32573 (A.A.), R21 AA027827 (R.B.), U01 DA055367 (R.B.), K01 DA51759 (E.C.J.), K23 MH121792 (N.R.K.), DP1 DA54394 (S.S.-R.), T32 MH014276 (G.A.P.), R01 AA027522 (A.E.), F31 AA029934 (S.E.P.), R01 MH120219 (E.M.T.-D., A.D.G.), RF1 AG073593 (E.M.T.-D., A.D.G.), P30 AG066614 (E.M.T.-D.), P2CHD042849 (E.M.T.-D.), R33 DA047527 (R.P., G.A.P.) and T32 AA028259 (J.D.D.)

Author contributions

A.S.H. designed the study and conducted analyses. S.M.C.C., E.C.J., S.B.H., J.D.D., G.A.P., M.V.J., S.S.-R., S.E.P., N.R.K., I.H. and D.A.A.B. conducted various analyses. A.A., R.B., H.J.E. and J.G. supervised the study. A.D.G. and E.M.T.-D. provided statistical guidance. A.E., H.R.K., R.P., L.K.D. and S.S.-R. guided interpretation of key findings. A.S.H., H.J.E., J.G., R.B. and A.A. drafted the manuscript. A.S.H., S.M.C.C., J.D.D., M.V.J., S.E.P., N.R.K. and I.H. organized the data. The consortium members provided insight into various aspects of analyses and interpretation and data for some of the discovery GWAS that were inputs to these analyses. All named authors reviewed, edited and approved the submission.

Competing interests

H.R.K. is a member of advisory boards for Dicerna Pharmaceuticals, Sophrosyne Pharmaceuticals and Enthion Pharmaceuticals; a consultant to Sobrera Pharmaceuticals; and a member of the American Society of Clinical Psychopharmacology's Alcohol Clinical Trials Initiative, which was supported in the last three years by Alkermes, Dicerna, Ethypharm, Lundbeck, Mitsubishi and Otsuka. H.R.K. and J.G. hold US Patent 10900,082: 'Genotype-guided dosing of opioid agonists' issued on 26 January 2021. The remaining authors declare no competing interests.

Additional information

Supplementary information The online version contains supplementary material available at https://doi.org/10.1038/s44220-023-00034-y.

Correspondence and requests for materials should be addressed to Alexander S. Hatoum.

Peer review information *Nature Mental Health* thanks Ditte Demontis and Eske Derks for their contribution to the peer review of this work.

[1]Department of Psychiatry, Washington University School of Medicine, Saint Louis, MO, USA. [2]Department of Psychology, Emory University, Atlanta, GA, USA. [3]Department of Psychiatry, Division of Human Genetics, Yale School of Medicine, New Haven, CT, USA. [4]Veterans Affairs Connecticut Healthcare System, West Haven, CT, USA. [5]Department of Psychiatry, UC San Diego School of Medicine, San Diego, CA, USA. [6]Department of Psychological and Brain Sciences, Washington University in Saint Louis, Saint Louis, MO, USA. [7]Virginia Institute of Psychiatric and Behavioral Genetics, Virginia Commonwealth University, Richmond, VA, USA. [8]Institute for Behavioral Genetics, University of Colorado-Boulder, Boulder, CO, USA. [9]Department of Psychology and Population Research Center, University of Texas at Austin, Austin, TX, USA. [10]Center for Studies of Addiction, Department of Psychiatry, University of Pennsylvania Perelman School of Medicine, Philadelphia, PA, USA. [11]VISN 4 MIRECC, Crescenz VAMC, Philadelphia, PA, USA. [12]Department of Medicine, Division of Genetic Medicine, Vanderbilt University, Nashville, TN, USA. [13]Department of Psychiatry and Behavioral Sciences, Vanderbilt University Medical Center, Nashville, TN, USA. [14]Department of Biomedical Informatics, Vanderbilt University Medical Center, Nashville, TN, USA. [15]Department of

Genetics, Yale School of Medicine, New Haven, CT, USA. [16]Department of Neuroscience, Yale School of Medicine, New Haven, CT, USA. [17]Department of Medical and Molecular Genetics, Indiana University, School of Medicine, Indianapolis, IN, USA. [18]Department of Biochemistry and Molecular Biology, Indiana University School of Medicine, Indianapolis, IN, USA. [136]These authors jointly supervised this work: Joel Gelernter, Howard J. Edenberg, Ryan Bogdan, Arpana Agrawal. *A list of authors and their affiliations appears at the end of the paper. ✉e-mail: ashatoum@wustl.edu

Substance Use Disorder Working Group of the Psychiatric Genomics Consortium

Daniel E. Adkins[19,20], Amy E. Adkins[21], Mervi Alanne-Kinnunen[22], Jeffry C. Alexander[23,24,25], Fazil Aliev[26], Silviu-Alin Bacanu[23,24,25], Anthony Batzler[27], Joanna M. Biernacka[28], Laura J. Bierut[1], Tim B. Bigdeli[29], Anna Blagonravova[30], Jason D. Boardman[31], Joseph M. Boden[32], Dorret I. Boomsma[33], Sandra A. Brown[34], Kathleen K. Bucholz[1], Danfeng Chen[35], Li-Shiun Chen[1], Doo-Sup Choi[36], S. Patricia Chou[37], Sven Cichon[38], William E. Copeland[39], Robin P. Corley[40], Franziska Degenhardt[41,42], Marta Di Forti[43], Nancy Diazgranados[44], Danielle M. Dick[45], Benjamin W. Domingue[46], Johan G. Eriksson[47,48], Lindsay A. Farrer[49,50], Jerome C. Foo[51], Tatiana M. Foroud[17], Louis Fox[1], Josef Frank[51], Mark A. Frye[52], Wolfgang Gaebel[53], Raul R. Gainetdinov[54], Ina Giegling[55], Nathan A. Gillespie[56], Alison M. Goate[57], David Goldman[58,59], Scott Gordon[60], Laura M. Hack[61], Dana B. Hancock[62], Kathleen Mullan Harris[63], Annette M. Hartmann[64], Andrew C. Heath[1], Stefanie Heilmann-Heimbach[65], Stefan Herms[38,41,42], Victor Hesselbrock[66], John K. Hewitt[67], Ian Hickie[68], Colin Hodgkinson[69], Per Hoffmann[38,41,42], Christian Hopfer[70], John Horwood[32], Jouke Jan Hottenga[33], Daniel Patrick Howrigan[35,71], William G. Iacono[72], Marcus Ising[73], Eric O. Johnson[74], Jaakko Kaprio[22], Victor M. Karpyak[52], Kenneth S. Kendler[23,24,25], Martin A. Kennedy[75], Margaret Keyes[76], Alexander Kibitov[77], Falk Kiefer[78], Bettina Konte[64], John Kramer[79], Kenneth Krauter[80], Evgeny M. Krupitsky[81], Samuel Kuperman[82], Jari Lahti[83], Marius Lahti-Pulkkinen[83], Dongbing Lai[17], Anastasia Levchenko[54], Lannie Ligthart[33], Penelope A. Lind[84], Susanne Lucae[73], Michael T. Lynskey[85], Pamela A. F. Madden[1], Hermine H. Maes[86], Patrik K. E. Magnusson[87], Brion S. Maher[88], Karl Mann[78], Satu Männistö[89], Nicholas G. Martin[60], Hamdi Mbarek[33], Matt McGue[76], Matthew B. McQueen[90], Sarah E. Medland[84], Jacquelyn L. Meyers[91], Grant W. Montgomery[92], Bertram Müller-Myhsok[93], Benjamin M. Neale[35,94], Elliot C. Nelson[1], Markus M. Nöthen[65], John I. Nurnberger[95], Aarno Palotie[22,35,94,96,97], Teemu Palviainen[22], John F. Pearson[98], Nancy L. Pedersen[87], Brenda W. J. H. Penninx[99], Roseann E. Peterson[56], Bernice Porjesz[100], Ulrich W. Preuss[101,102], Diego Quattrone[43], Katri Räikkönen[103], Maureen D. Reynolds[104], John P. Rice[1], Monika Ridinger[105], Marcella Rietschel[51], Brien P. Riley[23,24,25], Samuli Ripatti[22,35,94,106], Richard J. Rose[107], Dan Rujescu[64], Ksenia V. Rybakova[108], Euijung Ryu[109], Nancy L. Saccone[110], Jessica E. Salvatore[111], Norbert Scherbaum[112], Marc A. Schuckit[113], Melanie Schwandt[114], Pei-Hong Shen[69], Richard Sherva[115], Judy Silberg[116], Michael C. Stallings[117], Dan J. Stein[118], Fabian Streit[51], Jana Strohmaier[51], Ralph E. Tarter[119], Nathaniel Thomas[120], Michael M. Vanyukov[121], Scott Vrieze[76], Tamara L. Wall[122], Raymond K. Walters[35,94], Bradley T. Webb[23,24,25], Robbee Wedow[123,124,125,126,127], Frank Wendt[128], Leah Wetherill[129], John B. Whitfield[60], Stephanie Witt[51], Norbert Wodarz[130], Margaret J. Wright[131], Sarah M. Hartz[1], Stephanie Zellers[76], Haitao Zhang[132], Hongyu Zhao[133], Hang Zhou[134], Peter Zill[135] & Lea Zillich[51]

[19]Department of Psychiatry, University of Utah, Salt Lake City, UT, USA. [20]Department of Sociology, University of Utah, Salt Lake City, UT, USA. [21]Department of Psychology and College Behavioral and Emotional Health Institute, Virginia Commonwealth University, Richmond, VA, USA. [22]Institute for Molecular Medicine Finland, University of Helsinki, Helsinki, Finland. [23]Virginia Commonwealth University Alcohol Research Center, Richmond, VA, USA. [24]Department of Psychiatry, Virginia Commonwealth University, Richmond, VA, USA. [25]Virginia Institute for Psychiatric and Behavioral Genetics, Richmond, VA, USA. [26]Department of Psychiatry Rutgers University, Robert Wood Johnson Medical School, New Brunswick, NJ, USA. [27]Department of Quantitative Health Sciences, Mayo Clinic, Rochester, MN, USA. [28]Department of Psychiatry and Psychology, Mayo Clinic, Rochester, MN, USA. [29]Department of Psychiatry and Behavioral Sciences, State University of New York Downstate Medical Center, Brooklyn, NY, USA. [30]Clinical Laboratory Investigations, Privolzhsky Research Medical University, Nizhny Novgorod, Russia. [31]Department of Sociology, Institute of Behavioral Science, University of Colorado, Boulder, CO, USA. [32]Department of Psychological Medicine, University of Otago, Christchurch, New Zealand. [33]Department of Biological Psychology, Amsterdam Public Health Research Institute, Vrije Universiteit Amsterdam, Amsterdam, The Netherlands. [34]Department of Psychiatry and Psychology, University of California San Diego, La Jolla, CA, USA. [35]Stanley Center for Psychiatric Research, Broad Institute of MIT and Harvard, Cambridge, MA, USA. [36]Department of Molecular Pharmacology and Experimental Therapeutics, Mayo Clinic, Rochester, MN, USA. [37]NIH/NIAAA, Epidemiology and Biometry Branch, Bethesda, MD, USA. [38]Human Genomics Research Group, Department of Biomedicine, University of Basel Institute of Medical Genetics and Pathology, University Hospital Basel, Basel, Switzerland. [39]Department of Psychiatry, University of Vermont, Burlington, VT, USA. [40]Institute for Behavioral Genetics, University of Colorado Boulder, Boulder, CO, USA. [41]Institute of Human Genetics, University of Bonn, Bonn, Germany. [42]Department of Genomics, Life and Brain Center, University of Bonn, Bonn, Germany. [43]Social, Genetic and Developmental Psychiatry Centre Institute of Psychiatry, Psychology and Neuroscience, King's College London, London, UK. [44]NIH/NIAAA, Office of the Clinical Director NIAAA Intramural Research Program, Bethesda, MD, USA. [45]Department of Psychiatry, Rutgers Addiction Research Center at the Brain Health Institute, Rutgers University, Robert Wood Johnson Medical School, Piscataway, NJ, USA. [46]Graduate School of Education, Stanford University, Stanford, CA, USA. [47]Department of General Practice and Primary Health Care, University of Helsinki, Helsinki, Finland. [48]National Institute for Health and Welfare, Finland University of Helsinki, Helsinki, Finland. [49]Department of Medicine (Biomedical Genetics), Boston University School of Medicine, Boston, MA, USA. [50]Departments of Neurology, Ophthalmology, Epidemiology, and Biostatistics, Boston University Schools of Medicine and Public Health, Boston, MA, USA. [51]Department of Genetic Epidemiology in Psychiatry, Central Institute of Mental Health, Medical Faculty Mannheim, Heidelberg University, Mannheim, Germany. [52]Department of Psychiatry and Psychology, Mayo Clinic, Rochester, MN, USA. [53]LVR-Klinikum Düsseldorf, Heinrich-Heine-University Düsseldorf, Department of Psychiatry and Psychotherapy, University of Düsseldorf, Duesseldorf, Germany. [54]Institute of Translational Biomedicine, Saint Petersburg State University, Saint Petersburg, Russia. [55]Comprehensive Center of Clinical Neuroscience and Mental Health, Medical University of Vienna, Austria, Vienna, Austria. [56]Department of Psychiatry, Virginia Institute for Psychiatric and Behavioral Genetics, Virginia Commonwealth University, Richmond, VA, USA. [57]Department of Genetics and Genomic Sciences, Icahn School of Medicine at Mount Sinai, New York, NY, USA. [58]NIH/NIAAA, Laboratory of Neurogenetics, NIAAA, NIH, Rockville, MD, USA. [59]NIH/NIAAA, Office of the Clinical Director, Rockville, MD, USA. [60]Genetic Epidemiology, QIMR Berghofer Medical Research Institute, Brisbane, Queensland, Australia. [61]Department of Psychiatry and Behavioral Sciences, Stanford University School of Medicine, Stanford, CA, USA. [62]GenOmics, Bioinformatics, and Translational Research Center, RTI International Research, Triangle Park, NC, USA.

[63]Department of Sociology and Carolina Population Center, University of North Carolina at Chapel Hill, Chapel Hill, NC, USA. [64]Department of Psychiatry and Psychotherapy, Medical University of Vienna, Austria, Vienna, Austria. [65]Institute of Human Genetics, University of Bonn School of Medicine and University Hospital Bonn, Bonn, Germany. [66]Department of Psychiatry, University of Connecticut School of Medicine, Farmington, CT, USA. [67]Institute for Behavioral Genetics, University of Colorado Boulder, Boulder, CO, USA. [68]Brain and Mind Centre, University of Sydney, Sydney, Australia. [69]Laboratory of Neurogenetics, DICBR, NIAAA, NIH, USA, NIH/NIAAA, Laboratory of Neurogenetics, NIAAA, NIH, USA, Rockville, MD, USA. [70]Department of Psychiatry, University of Colorado School of Medicine, Aurora, CO, USA. [71]Department of Medicine Analytic and Translational Genetics Unit, Massachusetts General Hospital and Harvard Medical School, Boston, MA, USA. [72]Department of Psychology, University of Minnesota, Minneapolis, MN, USA. [73]Max-Planck-Institute of Psychiatry, Munich, Germany. [74]GenOmics, Bioinformatics, and Translational Research Center Biostatistics and Epidemiology Division, RTI International, Fellows Program Research, Triangle Park, NC, USA. [75]Department of Pathology and Biomedical Science, University of Otago, Christchurch, New Zealand. [76]Department of Psychology, University of Minnesota, Minneapolis, MN, USA. [77]Serbsky National Medical Research Center on Psychiatry and Addictions, Moscow, Russia. [78]Department of Addictive Behavior and Addiction Medicine, Central Institute of Mental Health Medical Faculty Mannheim, Heidelberg University, Mannheim, Germany. [79]Department of Psychiatry, Roy J and Lucille A Carver College of Medicine, University of Iowa, Iowa City, IA, USA. [80]Department of Molecular, Cellular, and Developmental Biology, University of Colorado Boulder, Boulder, CO, USA. [81]Addictions V.M. Bekhterev National Medical Research Center for Psychiatry and Neurology, Saint Petersburg, Russia. [82]Univeristy of Iowa Carver College of Medicine, Department of Psychiatry University of Iowa, Roy J and Lucille A Carver College of Medicine, Iowa City, IA, USA. [83]Department of Psychology and Logopedics Faculty of Medicine, University of Helsinki, Helsinki, Finland. [84]Psychiatric Genetics QIMR Berghofer Medical Research Institute Brisbane, Queensland, Australia. [85]Institute of Psychiatry, Psychology and Neuroscience, King's College, London, UK. [86]Department of Human and Molecular Genetics and Psychiatry, Virginia Institute for Psychiatric and Behavioral Genetics, Virginia Commonwealth University, Richmond, VA, USA. [87]Department of Medical Epidemiology and Biostatistics, Karolinska Instituet, Stockholm, Sweden. [88]Department of Mental Health, Johns Hopkins Bloomberg School of Public Health, Baltimore, MD, USA. [89]Public Health and Welfare, Finnish Institute for Health and Welfare, Helsinki, Finland. [90]Department of Integrative Physiology, University of Colorado Boulder, Boulder, CO, USA. [91]Department of Psychiatry and Behavioral Sciences, Henri Begleiter Neurodynamics Laboratory, SUNY Downstate Medical Center, Brooklyn, NY, USA. [92]The Institute for Molecular Bioscience, The University of Queensland, Brisbane, Queensland, Australia. [93]Department of Statistical Genetics, Max-Planck-Institute of Psychiatry, Munich, Germany. [94]Analytic and Translational Genetics Unit, Department of Medicine, Massachusetts General Hospital and Harvard Medical School, Boston, MA, USA. [95]Stark Neurosciences Research Institute, Departments of Psychiatry and Medical and Molecular Genetics, Indiana University School of Medicine, Indianapolis, IN, USA. [96]Department of Medicine, Department of Neurology and Department of Psychiatry, Massachusetts General Hospital, Boston, MA, USA. [97]University of Helsinki, Helsinki, Finland. [98]Biostatistics and Computational Biology Unit, University of Otago, Christchurch, New Zealand. [99]Department of Psychiatry, Amsterdam Public Health Research Institute, Amsterdam UMC, Vrije Universiteit, Amsterdam, The Netherlands. [100]Henri Begleiter Neurodynamics Laboratory, Department of Psychiatry and Behavioral Sciences, SUNY Downstate Medical Center, Brooklyn, NY, USA. [101]Martin-Luther-University Halle-Wittenberg, Department of Psychiatry, Psychotherapy and Psychosomatics, Halle (Saale), Germany. [102]RKH Hospital Ludwigsburg, Department of Psychiatry, Psychotherapy and Psychosomatics, Ludwigsburg, Germany. [103]Department of Psychology and Logopedics, University of Helsinki, Helsinki, Finland. [104]Department of Pharmaceutical Sciences, School of Pharmacy, University of Pittsburgh, Pittsburgh, PA, USA. [105]Department of Psychiatry and Psychotherapy, University of Regensburg Psychiatric Health Care Aargau, Regensburg, Germany. [106]Department of Public Health, University of Helsinki, Helsinki, Finland. [107]Department of Psychological and Brain Sciences, Indiana University, Bloomington, IN, USA. [108]V.M. Bekhterev National Medical Research Center for Psychiatry, Saint Petersburg, Russia. [109]Department of Quantitative Health Sciences, Divison of Computational Biology, Mayo Clinic, Rochester, MN, USA. [110]Department of Genetics, Washington University School of Medicine, Saint Louis, MO, USA. [111]Department of Psychiatry, Rutgers Biomedical and Health Sciences, Rutgers Robert Wood Johnson Medical School, Piscataway, NJ, USA. [112]LVR-Hospital Essen, Department of Psychiatry and Psychotherapy Medical Faculty, University of Duisburg-Essen, Essen, Germany. [113]Department of Psychiatry, University of California San Diego, La Jolla, CA, USA. [114]Office of the Clinical Director NIH/NIAAA, Bethesda, MD, USA. [115]Department of Medicine (Biomedical Genetics), Boston University School of Medicine, Boston, MA, USA. [116]Department of Human and Molecular Genetics, Virginia Institute for Psychiatric and Behavioral Genetics, Virginia Commonwealth University, Richmond, VA, USA. [117]Institute for Behavioral Genetics, University of Colorado Boulder, Boulder, CO, USA. [118]SAMRC Unit on Risk and Resilience in Mental Disorders, Department of Psychiatry and Neuroscience Institute, University of Cape Town, Cape Town, South Africa. [119]Department of Pharmaceutical Sciences, School of Pharmacy, University of Pittsburgh, Pittsburgh, PA, USA. [120]Department of Psychology and College Behavioral and Emotional Health Institute, Virginia Commonwealth University, Richmond, VA, USA. [121]Departments of Pharmaceutical Sciences, Psychiatry, and Human Genetics, School of Pharmacy, University of Pittsburgh, Pittsburgh, PA, USA. [122]Department of Psychiatry, School of Medicine, University of California San Diego, La Jolla, CA, USA. [123]Sociology, Purdue University, West Lafayette, IN, USA. [124]Statistics, Purdue University, West Lafayette, IN, USA. [125]Medical and Molecular Genetics, Indiana University School of Medicine, West Lafayette, IN, USA. [126]AnalytiXIN, West Lafayette, IN, USA. [127]Computer Science, Purdue University, West Lafayette, IN, USA. [128]Department of Psychiatry, Division of Human Genetics, Yale School of Medicine and VA CT Healthcare Center, West Haven, CT, USA. [129]Department of Medical and Molecular Genetics, Division of Hereditary Genomics, School of Medicine, Indiana University, Indianapolis, IN, USA. [130]Department of Psychiatry and Psychotherapy, University of Regensburg, Regensburg, Germany. [131]Queensland Brain Institute, University of Queensland, Brisbane, Queensland, Australia. [132]Epidemiology and Biometry Branch, National Institute on Alcohol Abuse and Alcoholism, National Institutes of Health, Bethesda, MD, USA. [133]Department of Biostatistics, Yale School of Public Health, Yale University, New Haven, CT, USA. [134]Department of Psychiatry, Yale School of Medicine and VA CT Healthcare Center, West Haven, CT, USA. [135]Department of Psychiatry, Psychiatric Hospital, Ludwig-Maximilians-University, Munich, Germany.

nature portfolio

Corresponding author(s): Alexander S. Hatoum

Last updated by author(s): 01/18/2023

Reporting Summary

Nature Portfolio wishes to improve the reproducibility of the work that we publish. This form provides structure for consistency and transparency in reporting. For further information on Nature Portfolio policies, see our Editorial Policies and the Editorial Policy Checklist.

Statistics

For all statistical analyses, confirm that the following items are present in the figure legend, table legend, main text, or Methods section.

n/a	Confirmed
☐	☒ The exact sample size (*n*) for each experimental group/condition, given as a discrete number and unit of measurement
☒	☐ A statement on whether measurements were taken from distinct samples or whether the same sample was measured repeatedly
☐	☒ The statistical test(s) used AND whether they are one- or two-sided *Only common tests should be described solely by name; describe more complex techniques in the Methods section.*
☐	☒ A description of all covariates tested
☒	☐ A description of any assumptions or corrections, such as tests of normality and adjustment for multiple comparisons
☐	☒ A full description of the statistical parameters including central tendency (e.g. means) or other basic estimates (e.g. regression coefficient) AND variation (e.g. standard deviation) or associated estimates of uncertainty (e.g. confidence intervals)
☐	☒ For null hypothesis testing, the test statistic (e.g. *F*, *t*, *r*) with confidence intervals, effect sizes, degrees of freedom and *P* value noted *Give P values as exact values whenever suitable.*
☒	☐ For Bayesian analysis, information on the choice of priors and Markov chain Monte Carlo settings
☒	☐ For hierarchical and complex designs, identification of the appropriate level for tests and full reporting of outcomes
☐	☒ Estimates of effect sizes (e.g. Cohen's *d*, Pearson's *r*), indicating how they were calculated

Our web collection on statistics for biologists contains articles on many of the points above.

Software and code

Policy information about availability of computer code

Data collection	No software was used in data collection
Data analysis	This paper made use of GenomicSEM v0.0.5(https://github.com/GenomicSEM/GenomicSEM), FUMA (https://fuma.ctglab.nl/), MASSIVE (genoma.io), PRS-CS (https://github.com/getian107/PRSc), LD score (https://github.com/bulik/ldsc) , S-Predixcan (https://github.com/hakyimlab/MetaXcan), SUsieR (https://cran.r-project.org/web/packages/susieR/index.html), and LAVAAN v0.6 (https://lavaan.ugent.be/). R 4.1.0-4.2.0(https://www.r-project.org). lme4 v1.1 (https://cran.r-project.org/web/packages/lme4/index.html). Lmertest v1.3 (https://cran.r-project.org/web/packages/lmerTest/index.html), metafor v3.4.0, GENESIS (https://github.com/yandorazhang/GENESIS), METASOFT v2.0.1, ASSET v 2.4.0., PheWAS V1.0 (https://github.com/PheWAS/PheWAS). Not all packages had versions as some are github pages. Software is cited in the text.

For manuscripts utilizing custom algorithms or software that are central to the research but not yet described in published literature, software must be made available to editors and reviewers. We strongly encourage code deposition in a community repository (e.g. GitHub). See the Nature Portfolio guidelines for submitting code & software for further information.

Data

Policy information about availability of data

All manuscripts must include a data availability statement. This statement should provide the following information, where applicable:

- Accession codes, unique identifiers, or web links for publicly available datasets
- A description of any restrictions on data availability
- For clinical datasets or third party data, please ensure that the statement adheres to our policy

The MVP summary statistics were obtained via an approved dbGaP application (phs001672.v4.p1). The authors thank Million Veteran Program (MVP) staff, researchers, and volunteers, who have contributed to MVP, and especially participants who previously served their country in the military and now generously agreed to enroll in the study. (For details, see https://www.research.va.gov/mvp/ and Gaziano, J.M. et al. Million Veteran Program: A mega-biobank to study genetic influences on health and disease. J Clin Epidemiol 70, 214-23 (2016)). This research is based on data from the Million Veteran Program, Office of Research and Development, Veterans Health Administration, and was supported by the Veterans Administration (VA) Cooperative Studies Program (CSP) award #G002.

Publicly available data were also taken from the psychiatric genomics consortium: https://www.med.unc.edu/pgc/, and the GSCAN consortium: https://conservancy.umn.edu/handle/11299/201564

The electronic health record data that support the findings of this study are available from Vanderbilt University Medical Center, but restrictions apply to the availability of these data, which were used under license for the current study. The data are only available from the institution with appropriate material transfer agreements or data use agreements and permission of Vanderbilt University Medical Center. The data in question must first be reviewed by the Integrated Data Access and Services Core to ensure that the de-identification is complete and no potentially identifying information remains. Please contact the Vanderbilt Institute for Clinical and Translational Research (research.support.services@vumc.org) for more information.

Data from Yale-Penn 1 are available through dbGAP accession nos. phs000425.v1.p1 and phs000952.v1.p1. Summary statistics for all Yale-Penn data are available on request.

ABCD data is made publicly available through the NDA: https://nda.nih.gov/data_dictionary.html?source=ABCD%2BRelease%2B3.0&submission=ALL

Field-specific reporting

Please select the one below that is the best fit for your research. If you are not sure, read the appropriate sections before making your selection.

☐ Life sciences ☒ Behavioural & social sciences ☐ Ecological, evolutionary & environmental sciences

For a reference copy of the document with all sections, see nature.com/documents/nr-reporting-summary-flat.pdf

Behavioural & social sciences study design

All studies must disclose on these points even when the disclosure is negative.

Study description	Genome-wide association study, a quantitative analysis that searches common variants for degree of statistical assocation with an outcome.
Research sample	1,025,550 individuals of European and 92,630 individuals of African descent genome-wide summary statistics for four substance use disorders: Opioid Use Disorder, problematic alcohol use/alcohol use disorder, cannabis use disorder, problematic tobacco use/ tobacco use disorder. Existing samples were used as to pool a large sample of existing dataset towards a common addiction model.
Sampling strategy	Samples were taken from all publicly available GWAS of substance use disorders that show significant genome-wide significant findings. More can be found in the original publications. Samples span clinical and convenience samples.
Data collection	Not relevant as all data were previously published
Timing	Not relevant as all data were previously published
Data exclusions	No African or European American summary statistics were excluded.
Non-participation	No participants were involved in this study. The study is an analysis of already published summary statistics.
Randomization	This is not an experiment.

Reporting for specific materials, systems and methods

We require information from authors about some types of materials, experimental systems and methods used in many studies. Here, indicate whether each material, system or method listed is relevant to your study. If you are not sure if a list item applies to your research, read the appropriate section before selecting a response.

Materials & experimental systems

n/a	Involved in the study
☒	☐ Antibodies
☒	☐ Eukaryotic cell lines
☒	☐ Palaeontology and archaeology
☒	☐ Animals and other organisms
☒	☐ Human research participants
☒	☐ Clinical data
☒	☐ Dual use research of concern

Methods

n/a	Involved in the study
☒	☐ ChIP-seq
☒	☐ Flow cytometry
☒	☐ MRI-based neuroimaging