Psychological Medicine

www.cambridge.org/psm

Multi-ancestral genome-wide association study of clinically defined nicotine dependence reveals strong genetic correlations with other substance use disorders and health-related traits

Emma C. Johnson[1*], Dongbing Lai[2*], Jared V. Balbona[1], Alex P. Miller[3], Alexander S. Hatoum[1], Joseph D. Deak[4,5], Mariela Jennings[6], David A. A. Baranger[7], Marco Galimberti[8,5], Kittipong Sanichwankul[9], Thorgeir Thorgeirsson[10], Sarah M. C. Colbert[11], Keyrun Adhikari[5,8], Anna R. Docherty[12], Louisa Degenhardt[13], Tobias Edwards[14], Louis Fox[1], Alexandros Giannelis[14], Paul W. Jeffries[1], Tellervo Korhonen[15], Claire L. Morrison[16,17], Yaira Z. Nunez[5,8], Teemu Palviainen[15], Mei-Hsin Su[18], Pamela N. Romero Villela[16,17], Leah Wetherill[2], Emily A. Willoughby[14], Stephanie M. Zellers[15], Laura J. Bierut[1], Jadwiga Buchwald[15], William E. Copeland[19], Robin P. Corley[17], Naomi P. Friedman[16,17], Tatiana M. Foroud[2], Nathan A. Gillespie[18,20], Ian R. Gizer[21], Andrew C. Heath[1], Ian B. Hickie[22], Jaakko Kaprio[15], Matthew C. Keller[16,17], James J. Lee[14], Penelope Lind[20,23,24], Pamela A. Madden[1], Hermine H. M. Maes[18,25,26], Nicholas G. Martin[20], Matt McGue[14], Sarah E. Medland[20,27,28], Elliot C. Nelson[1], John Pearson[20,29], Bernice Porjesz[30], Michael C. Stallings[16,17], Scott Vrieze[14], Kirk C. Wilhelmson[31,32,33], Henry R. Kranzler[34,35], Raymond K. Walters[36,37,38], Renato Polimanti[4,39,40], Robert Malison[4,41], Hang Zhou[4,5], Kari Stefansson[10,42], Sandra Sanchez-Roige[6,43,44], Marc Potenza[4,40,45,46,47], Apiwat Mutirangura[48], Vorasuk Shotelersuk[49,50], Rasmon Kalayasiri[51,52], Howard J. Edenberg[2,53], Joel Gelernter[4,5,8,54] and Arpana Agrawal[1]

[1]Department of Psychiatry, Washington University School of Medicine in St Louis, Saint Louis, MO, USA; [2]Department of Medical and Molecular Genetics, Indiana University School of Medicine, Indianapolis, IN, USA; [3]Department of Psychiatry, Indiana University School of Medicine, Indianapolis, IN, USA; [4]Department of Psychiatry, Yale University School of Medicine, New Haven, CT, USA; [5]Department of Psychiatry, Veterans Affairs Connecticut Healthcare Center, West Haven, CT, USA; [6]Department of Psychiatry, University of California San Diego, La Jolla, CA, USA; [7]Department of Psychological and Brain Sciences, Washington University in St Louis, Saint Louis, MO, USA; [8]Department of Human Genetics, Yale University School of Medicine, New Haven, CT, USA; [9]Suanprung Psychiatric Hospital, Chiang Mai, Thailand; [10]deCODE Genetics/Amgen, Sturlugata 8, IS-101, Reykjavik, Iceland; [11]Department of Genetics and Genomic Sciences, Icahn School of Medicine at Mount Sinai, New York, NY, USA; [12]Department of Psychiatry, University of Utah School of Medicine, Salt Lake City, UT, USA; [13]National Drug and Alcohol Research Centre, University of New South Wales, Sydney, Australia; [14]Department of Psychology, University of Minnesota Twin Cities, Minneapolis, MN, USA; [15]Institute for Molecular Medicine Finland, University of Helsinki, Helsinki, Finland; [16]Department of Psychology and Neuroscience, University of Colorado Boulder, Boulder, CO, USA; [17]Institute for Behavioral Genetics, University of Colorado Boulder, Boulder, CO, USA; [18]Virginia Institute for Psychiatric and Behavioral Genetics, Virginia Commonwealth University, Richmond, VA, USA; [19]Department of Psychiatry, College of Medicine, University of Vermont, Burlington, VT, USA; [20]QIMR Berghofer Medical Research Institute, Brisbane, Queensland, Australia; [21]Department of Psychological Sciences, University of Missouri, Columbia, MO, USA; [22]Brain and Mind Institute, University of Sydney, Sydney, NSW, Australia; [23]School of Biomedical Sciences, Faculty of Medicine, University of Queensland, Brisbane, QLD, Australia; [24]School of Biomedical Sciences, Queensland University of Technology, Brisbane, QLD, Australia; [25]Department of Human and Molecular Genetics, Virginia Commonwealth University, Richmond, VA, USA; [26]Department of Psychiatry, Virginia Commonwealth University, Richmond, VA, USA; [27]School of Psychology and Counselling, Queensland University of Technology, Brisbane, Australia; [28]School of Psychology, University of Queensland, Brisbane, Australia; [29]Faculty of Medicine, The University of Queensland, Brisbane, Australia; [30]Henri Begleiter Neurodynamics Laboratory, Department of Psychiatry, SUNY Downstate Health Science University, Brooklyn, NY, USA; [31]Department of Neurology and Genetics and the Bowles Center of Alcohol Studies, University of North Carolina, Chapel Hill, NC, USA; [32]Rockefeller Neuroscience Institute, West Virginia University, Morgantown, WV, USA; [33]Renaissance Computing Institute, Chapel Hill, NC, USA; [34]Mental Illness Research, Education and Clinical Center, Crescenz VAMC, Philadelphia, PA, USA; [35]Center for Studies of Addiction, University of Pennsylvania Perelman School of Medicine, Philadelphia, PA, USA; [36]Analytic and Translational Genetics Unit, Massachusetts General Hospital, Boston, MA, USA; [37]Harvard Medical School, Boston, MA, USA; [38]Stanley Center for Psychiatric Research, Broad Institute of Harvard and MIT, Cambridge,

Original Article

Cite this article: Johnson, E. C., Lai, D., Balbona, J. V., Miller, A. P., Hatoum, A. S., Deak, J. D., Jennings, M., Baranger, D. A. A., Galimberti, M., Sanichwankul, K., Thorgeirsson, T., Colbert, S. M. C., Adhikari, K., Docherty, A. R., Degenhardt, L., Edwards, T., Fox, L., Giannelis, A., Jeffries, P. W., Korhonen, T., Morrison, C. L., Nunez, Y. Z., Palviainen, T., Su, M.-H., Romero Villela, P. N., Wetherill, L., Willoughby, E. A., Zellers, S. M., Bierut, L. J., Buchwald, J., Copeland, W. E., Corley, R. P., Friedman, N. P., Foroud, T. M., Gillespie, N. A., Gizer, I. R., Heath, A. C., Hickie, I. B., Kaprio, J., Keller, M. C., Lee, J. J., Lind, P., Madden, P. A., Maes, H. H. M., Martin, N. G., McGue, M., Medland, S. E., Nelson, E. C., Pearson, J., Porjesz, B., Stallings, M. C., Vrieze, S., Wilhelmson, K. C., Kranzler, H. R., Walters, R. K., Polimanti, R., Malison, R., Zhou, H., Stefansson, K., Sanchez-Roige, S., Potenza, M., Mutirangura, A., Shotelersuk, V., Kalayasiri, R., Edenberg, H. J., Gelernter, J., & Agrawal, A. (2025). Multi-ancestral genome-wide association study of clinically defined nicotine dependence reveals strong genetic correlations with other substance use disorders and health-related traits. *Psychological Medicine*, **55**, e234, 1–10
https://doi.org/10.1017/S0033291725100883

Received: 05 February 2025
Revised: 28 May 2025
Accepted: 02 June 2025

Keywords:
addiction; genome-wide association study; nicotine dependence; polygenic risk; psychiatric disorders

Corresponding author:
Emma C. Johnson;
Email: emma.c.johnson@wustl.edu

*These authors made equal contributions.

MA, USA; [39]Cooperative Studies Program Clinical Epidemiology Research Center (CSP-CERC), VA Connecticut Healthcare System, West Haven, CT, USA; [40]Wu Tsai Institute, Yale University, New Haven, CT, USA; [41]The Clinical Neuroscience Research Unit, Connecticut Mental Health Center, New Haven, CT, USA; [42]Faculty of Medicine, University of Iceland, Reykjavik, Iceland; [43]Institute for Genomic Medicine, University of California San Diego, La Jolla, CA, USA; [44]Division of Genetic Medicine, Vanderbilt University Medical Center, Nashville, TN, USA; [45]Child Study Center, Yale School of Medicine, New Haven, CT, USA; [46]Connecticut Council on Problem Gambling, Wethersfield, CT, USA; [47]Connecticut Mental Health Center, New Haven, CT, USA; [48]Department of Anatomy, Faculty of Medicine, Chulalongkorn University, King Chulalongkorn Memorial Hospital, Bangkok, Thailand; [49]Center of Excellence for Medical Genomics, Department of Pediatrics, Faculty of Medicine, Chulalongkorn University, Bangkok, Thailand; [50]Excellence Center for Genomics and Precision Medicine, King Chulalongkorn Memorial Hospital, the Thai Red Cross Society, Bangkok, Thailand; [51]Department of Psychiatry, Epidemiology of Psychiatric Disorders and Mental Health Research Unit, Faculty of Medicine, Chulalongkorn University, Bangkok, Thailand; [52]Department of Psychiatry, King Chulalongkorn Memorial Hospital, Bangkok, Thailand; [53]Department of Biochemistry and Molecular Biology, Indiana University School of Medicine, Indianapolis, IN, USA and [54]Department of Neuroscience, Yale University School of Medicine, New Haven, CT, USA

Abstract

Background. Genetic research on nicotine dependence has utilized multiple assessments that are in weak agreement.

Methods. We conducted a genome-wide association study (GWAS) of nicotine dependence defined using the Diagnostic and Statistical Manual of Mental Disorders (DSM-NicDep) in 61,861 individuals (47,884 of European ancestry [EUR], 10,231 of African ancestry, and 3,746 of East Asian ancestry) and compared the results to other nicotine-related phenotypes.

Results. We replicated the well-known association at the *CHRNA5* locus (lead single-nucleotide polymorphism [SNP]: rs147144681, p = 1.27E−11 in EUR; lead SNP = rs2036527, p = 6.49e−13 in cross-ancestry analysis). DSM-NicDep showed strong positive genetic correlations with cannabis use disorder, opioid use disorder, problematic alcohol use, lung cancer, material deprivation, and several psychiatric disorders, and negative correlations with respiratory function and educational attainment. A polygenic score of DSM-NicDep predicted DSM-5 tobacco use disorder criterion count and all 11 individual diagnostic criteria in the independent National Epidemiologic Survey on Alcohol and Related Conditions-III sample. In genomic structural equation models, DSM-NicDep loaded more strongly on a previously identified factor of general addiction liability than a "problematic tobacco use" factor (a combination of cigarettes per day and nicotine dependence defined by the Fagerström Test for Nicotine Dependence). Finally, DSM-NicDep showed a strong genetic correlation with a GWAS of tobacco use disorder as defined in electronic health records (EHRs).

Conclusions. Our results suggest that combining the wide availability of diagnostic EHR data with nuanced criterion-level analyses of DSM tobacco use disorder may produce new insights into the genetics of this disorder.

Introduction

Nicotine use and nicotine dependence (or tobacco use disorder) are influenced by both genetic liability and environmental factors (Le Foll et al., 2022; MAES et al., 2004). Several different smoking-related phenotypes have been studied, including smoking initiation (SmkInit) (Saunders et al., 2022), cigarettes per day (CPD) (Saunders et al., 2022), the Fagerström Test for Nicotine Dependence (FTND) (Quach et al., 2020), problematic tobacco use (PTU) (Hatoum et al., 2022), tobacco use disorder defined by codes from the International Classification of Diseases (ICD-based tobacco use disorder [ICD-TUD]) (Toikumo et al., 2024, and nicotine dependence (Loukola et al., 2014) defined by the Diagnostic and Statistical Manual of Mental Disorders (DSM-NicDep) (Table 1).

Genome-wide association studies (GWASs) of nicotine use behaviors have largely focused on phenotypes that are easily ascertained through self-report questionnaires (e.g. *When did you start smoking?*, *Are you a current smoker?*, and *How many cigarettes do you smoke in a day?*). The largest GWASs of tobacco smoking behaviors to date identified 140 loci associated with CPD (N = 784,353) and 1,346 loci related to the initiation of regular smoking (SmkInit; N ~ 3.38 million) (Saunders et al., 2022).

GWASs of nicotine dependence phenotypes have generally been smaller. Some GWASs of nicotine dependence-related phenotypes have focused on data collected from short questionnaires, such as the FTND (Quach et al., 2020). The FTND is a six-item questionnaire that includes cigarettes smoked per day as an ordinal indicator. The largest GWAS of FTND (defined as 0–3 for mild, 4–6 for moderate, and 7–10 for severe dependence) identified five loci (Quach et al., 2020). Given the strong genome-wide genetic correlation between CPD and FTND (r_g = 0.95) (Quach et al., 2020), these GWASs were combined into a single phenotype, PTU, in another study (Hatoum et al., 2022) using multi-trait analysis of GWAS. A recent GWAS meta-analysis of ICD-TUD from electronic health records (see Table 1) reached nearly 900,000 samples and identified 88 loci (Toikumo et al., 2024), all of which had been implicated in prior smoking-related GWASs. ICD-TUD showed moderate genetic correlations with CPD (r_g = 0.44) (Toikumo et al., 2024) and FTND (r_g = 0.63).

Our study was motivated by epidemiological and clinical data supporting nosological distinctions between FTND-defined and ICD- or DSM-based diagnoses, including some studies that suggest qualitative and quantitative differences in the associations between DSM- and FTND-defined nicotine dependence and some psychopathology (Breslau & Johnson, 2000; Brook, Koppel, & Pahl, 2009). The FTND is brief and, therefore, easily and frequently collected. It has been especially prioritized in clinical trials of tobacco cessation (Ramon, Morchon, Baena, & Masuet-Aumatell, 2014; Tashkin et al., 2011), likely because FTND scores correlate well with relapse and treatment response, and the scale places a great deal of emphasis on physiological aspects of dependence (e.g. items related to tolerance and withdrawal [Baker, Breslau, Covey, & Shiffman, 2012]). On the other hand, both ICD- and DSM-based nicotine dependence include criteria related to physical and psychological (and social, in DSM-5) impairment due to nicotine/tobacco use, as well as behaviors directed at seeking and using nicotine to the exclusion of other activities. Neither FTND nor the ICD-TUD diagnostic classification maps perfectly to DSM-NicDep (Moolchan et al., 2002; Mwenifumbo & Tyndale, 2011).

Many large GWAS meta-analyses of substance use disorders have relied on cases defined using clinical criteria recommended by the DSM or ICD classification. However, the few GWASs of DSM-NicDep to date have been relatively small (Loukola et al., 2014). We conducted a large meta-analysis of DSM-NicDep, combining data

https://doi.org/10.1017/S0033291725100883 Published online by Cambridge University Press

Table 1. Definitions of different smoking-related phenotypes discussed in this article, the acronyms used to represent them, and the relevant GWASs used in analyses

Smoking phenotype	Shorthand name used in current manuscript	Phenotype definition	Relevant GWAS PubMed ID
Smoking initiation	SmkInit	Whether someone has ever been a regular smoker.	36477530
Cigarettes per day	CPD	Number of cigarettes smoked per day.	36477530
Fagerström Test for Nicotine Dependence	FTND	A six-item questionnaire assessing quantity of cigarettes smoked and compulsion to use, resulting in an ordinal measure of nicotine dependence.	33144568
Tobacco use disorder	ICD-TUD	Binary diagnosis defined using International Classification of Disease (ICD) codes (ICD 10 code F17.200; F17.200).	38632388
Nicotine dependence	DSM-NicDep	Seven criteria designed to assess clinical features of dependence. Including distress and problems caused by using cigarettes, at least 3 of the 7 criteria must be endorsed to reach a diagnosis of nicotine dependence.	NA
Problematic tobacco use	PTU	A combination of FTND and CPD GWAS, produced using multi-trait analysis of GWAS.	34750568

across 16 cohorts and multiple genetic ancestries. The largest analyses of psychiatric traits have focused on individuals whose genetics most closely resemble the European ancestry (EUR) subset of the 1,000 Genomes Project. Therefore, we assessed the genetic correlations between DSM-NicDep and other substance use phenotypes, psychiatric disorders, and related phenotypes in that subset. We also compared the correlations of the various nicotine dependence-related measures (DSM-NicDep, ICD-TUD, and PTU) and other substance use disorders (cannabis use disorder [CanUD], problematic alcohol use [PAU], and opioid use disorder [OUD]) by fitting a previously published common factor model (the Addiction-Risk-Factor) (Hatoum et al., 2023) to the data using genomic structural equation modeling (SEM). Finally, we tested whether a polygenic score (PGS) for DSM-NicDep was associated with DSM-5 tobacco use disorder and its 11 diagnostic criteria and 4 of the 6 FTND criteria in a large, independent sample – the National Epidemiologic Survey on Alcohol and Related Conditions-III (NESARC-III) cohort.

Methods

GWASs of DSM-NicDep

We performed meta-analyses of GWASs of DSM-defined nicotine dependence (hereafter referred to as "DSM-NicDep") across a total of 18 cohorts, 3 of which included samples of multiple ancestries, using a sample size-weighted meta-analysis implemented in METAL (Willer, Li, & Abecasis, 2010). We used nicotine-exposed controls where possible (see Supplemental Materials for more details on each cohort). The deCODE and Finnish Twin Cohort samples were entirely of EUR. For other cohorts, genetic ancestry similarity was inferred by comparing an individual's genome to the genomes from global reference populations using statistical methods such as principal components analysis. Specifically, all Psychiatric Genomics Consortium cohorts included in this meta-analysis used the 1,000 Genomes Phase 3 as a reference panel for defining ancestries based on principal components, as did the Thai sample. There were 16 cohorts with samples that were most genetically similar to EUR global reference populations, 4 with samples that were most genetically similar to African ancestry (AFR) global reference populations, and 1 cohort whose participants were most genetically similar to East Asian ancestry (EAS) global reference populations (hereafter referred to as European, African, or East Asian ancestries). All GWASs controlled for age, sex, and 10 principal components, except for the Minnesota Center for Twin and Family Research (MCTFR) cohort, which covaried for 20 principal components, and the deCODE cohort, which controlled for population stratification by adjusting for a genome-wide inflation factor; more details, including cohort-specific covariates, are provided in Supplemental Table 1. Twelve cohorts provided summary statistics, while five cohorts provided individual genotype and phenotype data for analysis (Supplemental Table 1).

We used FUMA v1.5.2 (Watanabe, Taskesen, van Bochoven, & Posthuma, 2017) to identify independent, genome-wide significant risk loci, annotate variants, and perform gene-wise analyses via MAGMA (de Leeuw, Mooij, Heskes, & Posthuma, 2015). We used the default FUMA parameters to define "independent significant single-nucleotide polymorphisms (SNPs)" as those that reached genome-wide significance ($p < 5e{-}8$) and were independent of each other at $r^2 < 0.6$, and "lead SNPs" as those SNPs that were independent of each other at $r^2 < 0.1$. "Genomic risk loci" were defined by merging linkage disequilibrium (LD) blocks of independent significant SNPs within a 250-kb distance. We performed gene mapping in FUMA using positional mapping (based on ANNOVAR annotations), expression quantitative trait locus mapping (using GTEx V8 [Aguet et al., 2017], CommonMind [Huckins et al., 2019], and Braineac [Trabzuni et al., 2011] data), and chromatin interaction mapping. We also performed the MAGMA gene, gene-set, and gene expression analyses (using GTEx V8 data).

We performed most follow-up analyses described below (genomic SEM and PGSs) only in the EUR data because the AFR and EAS subsets were underpowered.

Genetic correlations with other substance use disorders, psychiatric disorders, and other relevant phenotypes

We used LD score regression (LDSC) (Bulik-Sullivan et al., 2015; Bulik-Sullivan et al., 2015) to estimate the SNP-heritability of DSM-NicDep and the genetic correlations between DSM-NicDep and other substance use phenotypes, using published GWASs of PAU (Zhou et al., 2023), FTND (Quach et al., 2020), ICD-TUD (Toikumo et al., 2024), CPD (Saunders et al., 2022), cannabis ever-use (Pasman et al., 2018), CanUD (Levey et al., 2023), OUD

https://doi.org/10.1017/S0033291725100883 Published online by Cambridge University Press

(Deak et al., 2022), SmkInit (Saunders et al., 2022), and smoking cessation (Saunders et al., 2022). We also estimated genetic correlations between DSM-NicDep and other phenotypes, including psychiatric disorders, behavioral traits, respiratory health, and socioeconomic status-related phenotypes. Details on the individual GWAS used in genetic correlation analyses are provided in the Supplemental Methods. We further tested whether genetic correlations for DSM-NicDep and FTND were different from each other using a block-jackknife method (Bulik-Sullivan, Finucane, et al., 2015; Coleman et al., 2020).

We also estimated genetic correlations between the AFR DSM-NicDep GWAS and eight traits for which GWAS data were available from a reasonably large AFR sample. With a total sample size <5,000, it was not feasible to estimate meaningful genetic correlations using LDSC for the DSM-NicDep EAS samples.

Genomic SEM

We applied confirmatory factor analysis to the covariance matrix generated by LDSC using genomic SEM (Grotzinger et al., 2019) with weighted least squares estimation. As presented in Hatoum et al. (2022)), the indicators were allowed to load freely on a single latent factor (*Addiction-Risk-Factor*), but we updated the OUD (Deak et al., 2022), PAU (Zhou et al., 2023), and CanUD (Levey et al., 2023) GWASs. We compared *Addiction-Risk-Factor* models with DSM-NicDep, PTU (a combination of FTND and CPD [Hatoum, Colbert, et al., 2023), or ICD-TUD as the tobacco-related indicator. The variance of the common latent factor was scaled to 1.0.

PGS analyses

We created PGSs for DSM-NicDep, FTND, ICD-TUD, and CPD in the EUR subset of the NESARC-III sample. NESARC-III was genotyped using the Affymetrix Axiom® Exome Array (Zhang et al., 2022), which limited our ability to impute SNPs due to a lack of appreciable nonexonic coverage and resulted in some regions with low SNP densities. Details of quality control and imputation are available in the Supplemental Materials. We used PRS-CS (Ge, Chen, Ni, Feng, & Smoller, 2019), a Bayesian method that uses continuous shrinkage before weight SNP effect sizes, and used the "auto" function, which allows the global scaling parameter to be automatically learned from the data. We then used the "score" function in Plink 1.9 (Chang et al., 2015) to create PGS for DSM-NicDep, FTND, ICD-TUD, and CPD in NESARC-III. We used logistic regression models to estimate associations between the PGS and endorsement of the 11 individual diagnostic criteria for DSM-5 TUD and 4 of the 6 FTND criteria (excluding CPD and smoking when ill) in NESARC-III (N = 12,482; DSM-TUD N_{cases} = 4,205). Linear regression models were used to estimate associations between PGSs and the total DSM-5 TUD criterion count and total FTND criterion count. All regression models covaried for age, sex, and 10 within-ancestry principal components. We first ran a model with only the DSM-NicDep PGS as a predictor, then a model with only the FTND PGS as a predictor, and finally a model where all four PGSs (DSM-NicDep, FTND, ICD-TUD, and CPD) were included jointly as predictors.

We also created PGSs for DSM-NicDep, FTND, and CPD in the BioVU biobank (we did not create a PGS for ICD-TUD, because BioVU was included in that published GWAS meta-analysis). We used the methods described above to create the PGS and tested whether the DSM-NicDep, FTND, and CPD PGS were associated with ICD-TUD (N = 66,914; N_{cases} = 9,666) in BioVU in a logistic regression model that controlled for age, sex, and 10 ancestry principal components.

Results

GWAS of DSM-NicDep

There were a total of 15 cohorts included in the EUR meta-analysis (N_{cases} = 20,923, $N_{controls}$ = 26,961), 4 cohorts included in the AFR meta-analysis (N_{cases} = 5,293, $N_{controls}$ = 4,938), and 1 cohort included in the EAS GWAS (N_{cases} = 2,007, $N_{controls}$ = 1,739), for a total cross-ancestry sample size of N = 61,861 (28,223 cases; Supplemental Table 1). We identified one genome-wide significant locus, near the cholinergic receptor nicotinic alpha 5 subunit (*CHRNA5*) gene on chromosome 15 (lead SNP: rs147144681, p =1.27E−11 in EUR; lead SNP = rs2036527, p = 6.49e−13 in cross-ancestry analysis; Supplemental Tables 2 and 3). Using a lifetime prevalence of 24% (Breslau, Johnson, Hiripi, & Kessler, 2001), we estimated the liability scale SNP-heritability of DSM-NicDep to be 0.07 (SE = 0.01) in the EUR samples. The estimated liability scale SNP-heritability was similar in the AFR samples, but with a much larger standard error (SNP-h^2 = 0.08, SE = 0.09).

In the MAGMA gene-based analysis of the EUR data, three genes were significant after multiple testing corrections: *CHRNA5*, *CHRNA3*, and *IREB2* (Supplemental Table 4). The following three gene sets were significantly associated: "GOBP_TRANSCRIPTION_BY_RNA_POLYMERASE_III," "GOMF_PRE_MRNA_5_SPLICE_SITE_BINDING," and "REACTOME_HIGHLY_CALCIUM_PERMEABLE_NICOTINIC_ACETYLCHOLINE_RECEPTORS." None of the 30 tissue types from GTEx v8 were significant.

Genetic correlations with substance use and other phenotypes

In the EUR data, DSM-NicDep was strongly correlated with the other measures of nicotine dependence (r_gs with FTND and ICD-TUD ranged from 0.81 to 1.01). The genetic correlations between DSM-NicDep and published GWASs of substance use disorders were of moderate-to-high magnitude: CanUD, PAU, and OUD (r_g = 0.64–0.84). Overall, DSM-NicDep was significantly correlated with 23 of the 26 phenotypes tested (Figure 1 and Supplemental Table 5). Compared to the other tobacco-related phenotypes, DSM-NicDep showed the strongest correlations with many traits, albeit with wider confidence intervals (CIs) due to smaller sample sizes. When correcting for 24 comparisons, the genetic correlations between DSM-NicDep and SmkInit, cannabis ever-use, CanUD, PAU, and the Townsend Deprivation Index were significantly larger (p < 0.002) than the corresponding genetic correlations between FTND and these phenotypes.

Presumably due to the smaller sample size of the AFR data (reflected in the imprecise estimate of SNP-heritability, SNP-h^2 = 0.08, SE = 0.09), the standard errors of the genetic correlations were quite large (Supplemental Table 6). However, the genetic correlation between DSM-NicDep and ICD-TUD in the AFR ancestry was significant at p < 0.05 (r_g = 0.58, p = 0.024) (Toikumo et al., 2024).

https://doi.org/10.1017/S0033291725100883 Published online by Cambridge University Press

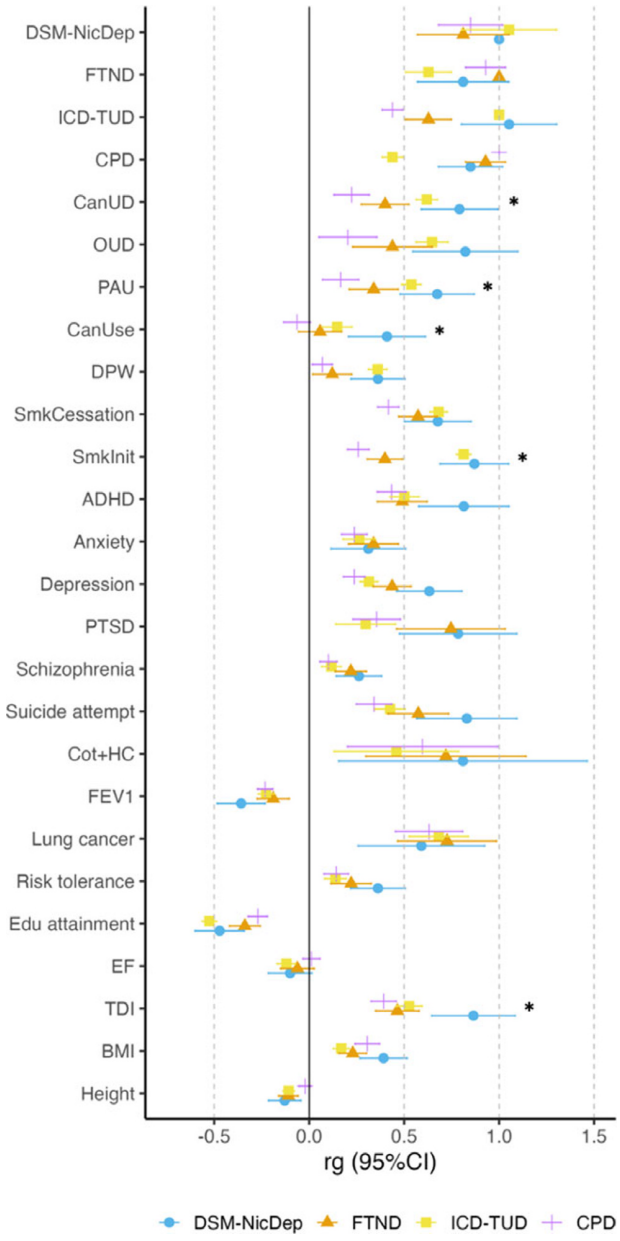

Figure 1. Comparing genetic correlations (r_g) for DSM-NicDep, FTND, ICD-TUD, and CPD with other traits in European ancestry data. Traits include other substance use disorders (CanUD, cannabis use disorder [Levey et al., 2023]; OUD, opioid use disorder [Deak et al., 2022]; PAU, problematic alcohol use [Zhou et al., 2023]; ICD-TUD, ICD-based tobacco use disorder [Toikumo et al., 2024]), substance use behaviors (CanUse, cannabis ever-use [Pasman et al., 2018]; DPW, drinks per week [Saunders et al., 2022]; SmkInit, smoking initiation [Saunders et al., 2022]; SmkCessation, smoking cessation [Saunders et al., 2022]), psychiatric disorders and other mental health phenotypes (ADHD, attention deficit hyperactivity disorder [Demontis et al., 2023]; PTSD, post-traumatic stress disorder [Nievergelt et al., 2019]), biomarkers (Cot + HC, cotinine +3-hydroxycotinine [Buchwald et al., 2021]), lung health-related traits (FEV1, forced expiratory volume in 1 s), risk tolerance (Linnér et al., 2019), socioeconomic status-related traits (Edu attainment, educational attainment [Lee et al., 2018]; TDI, Townsend Deprivation Index]), executive function (EF [Hatoum, Morrison, et al., 2023]), and anthropometric measures (BMI, body mass index [Yengo et al., 2018]; height [Yengo et al., 2022]). * indicates r_gs that significantly differ between DSM-NicDep and FTND at $\alpha = 0.002$ (Bonferroni correction for 24 comparisons).

Genomic structural equation models of broad addiction liability

In the EUR data, we found that a common genetic factor model with DSM-NicDep, PAU, OUD, and CanUD as indicators, similar to the

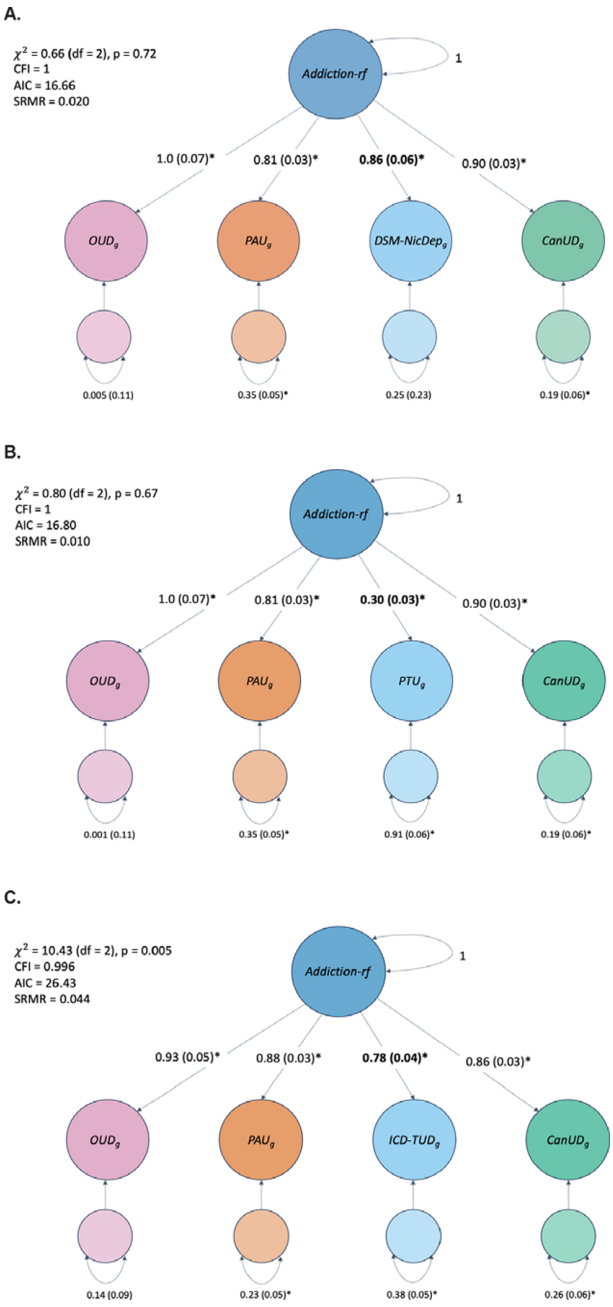

Figure 2. A modified Addiction-Risk-Factor model. This model is patterned upon the common factor model in Figure 1A of Hatoum et al., 2022, but updated with new, larger versions of the OUD (Deak et al., 2022), PAU (Zhou et al., 2023), and CanUD GWAS (Levey et al., 2023) and using three different phenotypes for tobacco GWAS. (a) DSM-NicDep. (b) PTU (Hatoum et al., 2022) GWAS. (c) ICD-TUD (Toikumo et al., 2024). *Significant loadings at $p < 0.05$. Addiction-rf, the Addiction-Risk-Factor; CanUD, cannabis use disorder; DSM-NicDep, nicotine dependence; ICD-TUD, ICD-based tobacco use disorder; OUD, opioid use disorder; PAU, problematic alcohol use.

Addiction-Risk-Factor presented in Hatoum et al., (2022), fit the data well (Figure 2a and Supplemental Table 7). There were two differences between the Hatoum et al. *Addiction-Risk-Factor* and our modified model: (1) We used larger, more recent versions of the OUD (Deak et al., 2022), PAU (Zhou et al., 2023), and CanUD (Levey et al., 2023) GWASs; and (2) we removed the residual correlation between PAU and OUD, as this path was no longer significant. We compared the model with DSM-NicDep as the tobacco-related indicator (Figure 2a) to one where DSM-NicDep was substituted by the

https://doi.org/10.1017/S0033291725100883 Published online by Cambridge University Press

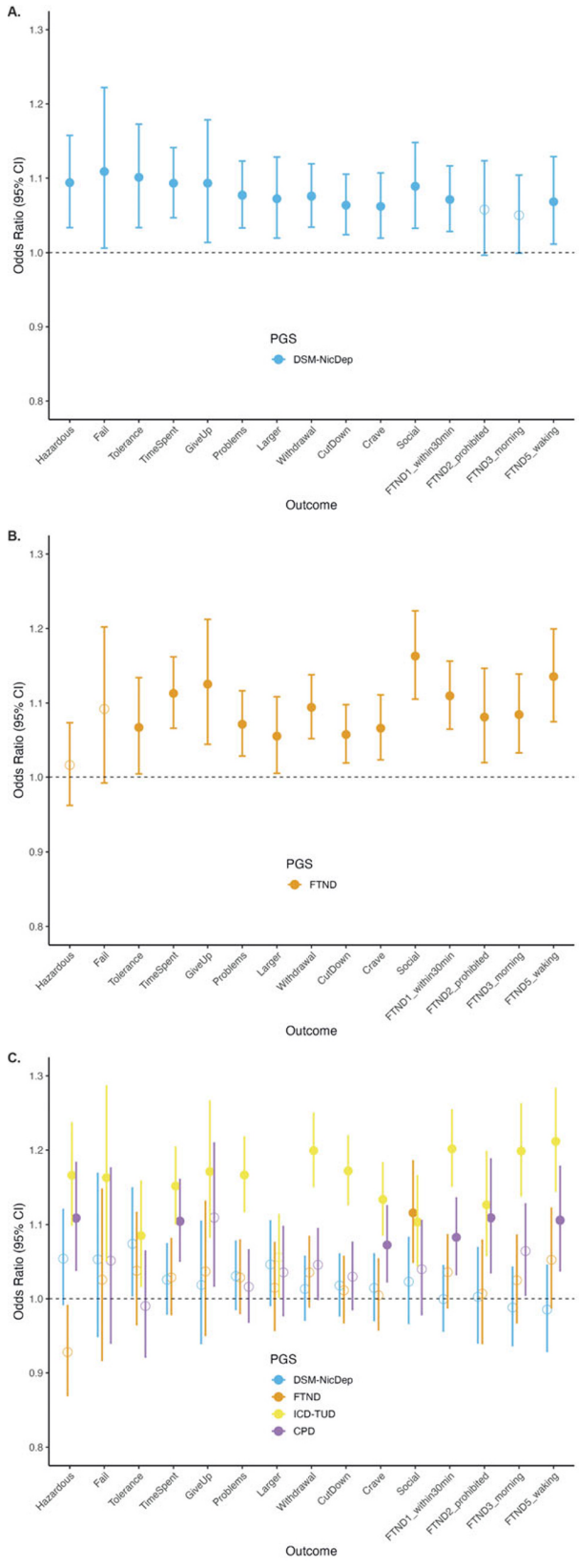

Figure 3. Polygenic scores (PGSs) for DSM-NicDep (a), FTND (b), and DSM-NicDep, FTND, ICD-TUD, and CPD (c) predict individual DSM-5 nicotine use disorder and FTND criteria and total criterion or item counts, respectively, in the European ancestry subset of the NESARC-III sample. Filled circles represent estimates that were significant after FDR correction, while open circles represent estimates that were not significant after FDR correction. Hazardous = Recurrent use in physically hazardous situations; Fail = Recurrent use resulting in failure to fulfill major role obligations at work, school, or home; Tolerance = Marked need for increased amount to get the same effect or diminished effect of the same amount; TimeSpent = Great deal of time spent in activities necessary to obtain or use; GiveUp = Important recreational, social, or occupational activities given up or reduced; Problems = Use despite knowledge of persistent/recurrent physical/psychological problems; Larger = Taken over larger amounts/longer periods than intended; Withdrawal = Withdrawal syndrome or use to relieve/avoid syndrome; Cutdown = Persistent desire or unsuccessful attempts to cut down or control use; Crave = Craving or strong urge or desire to use; Social = Persistent use despite recurring social/interpersonal problems caused or exacerbated by use; FTND1_within30min = How soon after you wake up do you smoke your first cigarette?; FTND2_prohibited = Do you find it difficult to refrain from smoking in places where it is forbidden?; FTND3_morning = Which cigarette would you hate most to give up?; FTND5_waking = Do you smoke more frequently during the first hours after waking than during the rest of the day?

original PTU GWAS meta-analysis (FTND + CPD) from Hatoum et al.[4] (Figure 2b and Supplemental Table 8), and one where NicDep was substituted by ICD-TUD (Toikumo et al., 2024) (Figure 2c and Supplemental Table 9). Each of the modified models fit the data well, but the loading for DSM-NicDep was the largest of the three tobacco-related indicators, almost threefold larger than the loading for PTU in the equivalent model (0.86 vs. 0.30).

PGSs for DSM-NicDep, FTND, CPD, and ICD-TUD

Results from the item-level logistic regression models in the EUR subset of the NESARC-III sample are shown in Figure 3. As the sole PGS predictor (Figure 3a and Supplemental Table 10), the DSM-NicDep PGS was significantly associated with all 11 DSM criteria ($p_{FDR} < 0.05$) and 2 of the 4 FTND items (how soon after waking is the first cigarette smoked, and whether one smokes more frequently during the early hours of the day). The DSM-NicDep PGS was also associated with the total number of endorsed DSM-5 TUD criteria ($\beta = 0.116$, SE = 0.025, $p_{FDR} = 7.1e{-}5$, variance explained = 0.17%) and total number of endorsed FTND items ($\beta = 0.033$, SE = 0.010, $p_{FDR} = 0.003$, variance explained = 0.08%). When the FTND PGS was the sole predictor, it was associated with all 4 FTND items and 9 of 11 DSM criteria (Figure 3b), as well as the total number of endorsed DSM-5 TUD criteria ($\beta = 0.118$, SE = 0.025, $p_{FDR} = 7.2e{-}6$, variance explained = 0.2%) and total number of endorsed FTND items ($\beta = 0.052$, SE = 0.010, $p_{FDR} = 1.3e{-}6$, variance explained = 0.2%; Supplemental Table 11). When all four PGS were included as predictors (Figure 3c and Supplemental Table 12), the DSM-NicDep PGS was no longer associated with any individual DSM or FTND items nor the total criterion counts. Overall, the ICD-TUD PGS was the strongest predictor of endorsement of individual DSM and FTND items, but the variance explained was still quite small (maximum variance explained = 0.716%).

In a model that included DSM-NicDep, FTND, and CPD PGSs as joint predictors, the DSM-NicDep PGS was significantly associated with TUD in the EUR subset of the BioVU biobank (odds ratio [OR] = 1.013, 95% CI = [1.01, 1.016], $p = 5.7e{-}18$), as was the PGS for CPD (OR = 1.017, 95% CI = [1.014, 1.02], $p = 1.3e{-}31$). The PGS for FTND was also significantly associated, but with an even smaller effect size (OR = 1.006, 95% CI = [1.003, 1.009], $p = 1.6e{-}5$). Altogether, the PGS explained 0.58% of the variance in ICD-TUD, while the DSM-NicDep PGS alone explained 0.12% of the variance in ICD-TUD after accounting for the variance explained by the other PGS (Supplemental Table 13).

Discussion

In this first cross-ancestry GWAS of DSM-NicDep ($N = 61,861$), we replicated the well-known *CHRNA5* risk locus. DSM-NicDep was

https://doi.org/10.1017/S0033291725100883 Published online by Cambridge University Press

genetically correlated with other substance use disorders, smoking-related phenotypes, and psychiatric disorders. In genomic SEM analyses, we found that DSM-NicDep was more closely related to the genetic underpinnings of other DSM- or ICD-defined substance use disorders than PTU (a phenotype that combined FTND and CPD measures).

ICD and DSM criteria focus on physical and psychological (and social, in DSM-5) impairment due to substance use and are mostly identical across substances. FTND, on the other hand, places greater emphasis on physiological aspects of dependence (e.g. items related to tolerance and withdrawal [Baker et al., 2012]). Because most GWASs of substance use disorders utilize ICD and DSM criteria, we hypothesized that this discrepancy might underlie the lower genetic correlation observed between nicotine dependence and other substance use disorders in prior twin and genome-wide correlation studies that used FTND. Indeed, the loading for DSM-NicDep in the Addiction-Risk-Factor genomic SEM was nearly three times that of the loading for PTU (0.83 vs. 0.30) (Figure 1). Notably, the loadings for OUD and PAU also increased in our model (0.83–1.0 and 0.58–0.81, respectively), likely due to the larger sample sizes of these more recent GWAS. We also noted no significant residual correlation between OUD and PAU in our analysis.

In the large, independent BioVU biobank, the DSM-NicDep and CPD PGS showed significant associations with ICD-TUD of similar magnitude (ORs = 1.013–1.017), while the association with the FTND PGS was smaller (OR = 1.006). A very small proportion of variance was explained by the three PGSs (0.58%), reflecting underpowered GWASs and PGSs that are not currently suitable for individual-level prediction. PGS analyses of individual criteria in NESARC-III revealed that the DSM-NicDep PGS was associated with all individual DSM-5 TUD criteria and two of the FTND-specific items when it was the only predictor in the model (Figure 3a). However, when all four tobacco-related PGSs (ICD-TUD, FTND, CPD, and DSM-NicDep) were included as predictors, DSM-NicDep was no longer associated with any individual items (Figure 3b), nor the total criterion counts. In this multi-PGS model, the FTND PGS was associated with only 1 item (social, one of the three DSM-IV abuse criteria in the DSM-5 TUD diagnosis) while the ICD-TUD PGS was associated with 10 of the 11 DSM-5 criteria and all 4 FTND items, consistent with the ICD-TUD GWAS being much more statistically powerful than FTND or DSM-NicDep. Future item-level GWASs using novel structural equation modeling methods that bring together DSM and FTND items might be of high value in parsing whether genetic and clinical heterogeneity align.

We expected that DSM-NicDep would be more genetically correlated with psychopathology than FTND, as has been found in clinical and epidemiological studies (Breslau & Johnson, 2000). The point estimates of our genetic correlation analysis suggested that PTU (CPD + FTND) was the least associated with most indices of psychopathology, while both DSM-NicDep and ICD-TUD were more strongly associated with psychosocial indices. DSM-NicDep appeared to additionally index material deprivation more strongly than the other traits, while all four tobacco GWASs (DSM-NicDep, ICD-TUD, FTND, and CPD) were equivalently related to respiratory (e.g. FEV1 and lung cancer) and metabolic markers of tobacco exposure.

The largest GWAS of CPD (Saunders et al., 2022) identified 140 genomic risk loci, compared with 5 in the largest GWAS of FTND (Quach et al., 2020) and 1 in the current study of DSM-NicDep. This likely reflects the much larger sample size in the CPD study. Despite a similar sample size, the FTND GWAS identified more loci than the current DSM-NicDep GWAS, possibly because that study used a linear model to analyze a categorical three-level (mild–moderate–severe) measure while our analyses relied on binary diagnoses. A recent GWAS of ICD-TUD that had a much larger sample size than ours (N = 898,680) identified 88 loci (Toikumo et al., 2024. In the current study, DSM-NicDep was highly correlated with ICD-TUD. Given their high genetic correlation, future efforts may consider combining these diagnostic modalities as sources of information on tobacco use disorder.

We did not find any significant loci in the much smaller AFR or EAS analyses, nor could we precisely estimate SNP-heritability or genetic correlations due to the limited power. The lack of diversity in genetic data is unfortunate, as nicotine dependence and tobacco use disorder are leading contributors to mortality in worldwide populations (Le Foll et al., 2022; Reitsma et al., 2021). A related limitation was our inability to conduct sex- and birth cohort-stratified analyses; rates of tobacco use vary markedly according to both sex and birth cohort, which may in turn modify the extent to which genetic liability influences risk for DSM-NicDep. Amassing more genetic datasets with DSM-derived nicotine dependence would enable such valuable analyses.

In summary, our analyses highlight the importance of considering diagnostic assessment in genetic studies of substance use disorders. We found that DSM-NicDep was more closely related to a general addiction liability factor compared to a "PTU" phenotype that combined FTND and CPD. The DSM-NicDep PGS was significantly associated with ICD-TUD in an independent biobank and was associated with individual DSM-5 TUD criteria in another independent sample. Compared to FTND, DSM-NicDep was more strongly genetically correlated with ICD-TUD, PAU, SmkInit, cannabis ever-use, and material deprivation. Given the strong genetic correlation between DSM-NicDep and ICD-TUD, future analyses may consider combining data from DSM-NicDep and ICD-based studies of TUD to maximize sample size for gene discovery.

Supplementary material. The supplementary material for this article can be found at http://doi.org/10.1017/S0033291725100883.

Acknowledgments. The authors gratefully acknowledge our contributing studies and the participants in those studies, without whom this effort would not be possible.

AGDS: The authors would like to thank all the participants for giving their time to contribute to this study. The authors wish to thank all the people who helped in the conception, implementation, beta testing, media campaign, and data cleaning. The authors would specifically like to acknowledge Ken Kendler, Patrick Sullivan, Andrew McIntosh, and Cathryn Lewis for input on the questionnaire; Lorelle Nunn, Mary Ferguson, Lucy Winkler, and Natalie Garden for data and sample collection; Natalia Zmicerevska, Alissa Nichles, and Candace Brennan for participant recruitment support; Jonathan Davies, Luke Lowrey, and Valeriano Antonini for support with IT aspects; and Vera Morgan and Ken Kirkby for help with the media campaign.

GBP: The authors would like to thank the participants for giving their time and support for this project. The authors acknowledge and thank M. Steffens for her generous donations in loving memory of J. Banks.

Funding support for the **Comorbidity and Trauma Study (CATS)** (dbGAP accession number: phs000277.v1.p1) was provided by the National Institute on Drug Abuse (R01 DA17305) and GWAS genotyping services at the CIDR at The Johns Hopkins University were supported by the National Institutes of Health (contract N01-HG-65403). The **Christchurch Health and Development Study (CHDS**: dbGAP in progress) has been supported by funding from the Health Research Council of New Zealand, the National Child Health Research Foundation (Cure Kids), the Canterbury Medical Research Foundation, the New Zealand Lottery Grants Board, the University of Otago, the Carney Centre for Pharmacogenomics, the James Hume Bequest Fund, US NIH grant MH077874,

and NIDA grant "A developmental model of gene–environment interplay in SUDs" (R01DA024413) 2007–2012. The **Collaborative Study on the Genetics of Alcoholism (COGA)** – Principal Investigators B. Porjesz, V. Hesselbrock, A. Agrawal; Scientific Director, and A. Agrawal and Translational Director D. Dick – includes 10 different centers: University of Connecticut (V. Hesselbrock); Indiana University (H.J. Edenberg, T. Foroud, Y. Liu, and M.H. Plawecki); University of Iowa Carver College of Medicine (S. Kuperman and J. Kramer); SUNY Downstate Health Sciences University (B. Porjesz, J. Meyers, C. Kamarajan and A. Pandey); Washington University in St. Louis (L. Bierut, J. Rice, K. Bucholz, A. Agrawal, and S. Hartz); University of California at San Diego (M. Schuckit); Rutgers University (J. Tischfield, D. Dick, R. Hart, and J. Salvatore); The Children's Hospital of Philadelphia, University of Pennsylvania (L. Almasy); Icahn School of Medicine at Mount Sinai (A. Goate and P. Slesinger); and Howard University (D. Scott). Other COGA collaborators include: L. Bauer (University of Connecticut); J. Nurnberger Jr., L. Wetherill, X. Xuei, D. Lai, and S. O'Connor (Indiana University); G. Chan (University of Iowa and University of Connecticut); D.B. Chorlian, J. Zhang, P. Barr, S. Kinreich, G. Pandey, and Z. Neale (SUNY Downstate); N. Mullins (Icahn School of Medicine at Mount Sinai); A. Anokhin, S. Hartz, E. Johnson, V. McCutcheon, and S. Saccone (Washington University); J. Moore, F. Aliev, Z. Pang, and S. Kuo (Rutgers University); A. Merikangas (The Children's Hospital of Philadelphia and University of Pennsylvania); and H. Chin and A. Parsian are the NIAAA Staff Collaborators. The authors expressed that they continue to be inspired by their memories of Henri Begleiter and Theodore Reich, the founding PI and Co-PI of COGA, and also owe a debt of gratitude to other past organizers of COGA, including Ting-Kai Li, P. Michael Conneally, Raymond Crowe, and Wendy Reich, for their critical contributions. This national collaborative study is supported by NIH Grant U10AA008401 from the National Institute on Alcohol Abuse and Alcoholism (NIAAA) and the National Institute on Drug Abuse (NIDA). Support for the **Study of Addiction: Genetics and Environment (SAGE)** was provided through the NIH Genes, Environment and Health Initiative (GEI; U01 HG004422; dbGaP study accession phs000092.v1.p1). SAGE is one of the genome-wide association studies funded as part of the Gene Environment Association Studies (GENEVA) under GEI. Assistance with phenotype harmonization and genotype cleaning, as well as with general study coordination, was provided by the GENEVA Coordinating Center (U01 HG004446). Assistance with data cleaning was provided by the National Center for Biotechnology Information. Support for the collection of datasets and samples was provided by the Collaborative Study on the Genetics of Alcoholism (COGA; U10 AA008401), the Collaborative Genetic Study of Nicotine Dependence (COGEND; P01 CA089392; see also phs000404.v1.p1), and the Family Study of Cocaine Dependence (FSCD; R01 DA013423 and R01 DA019963). Funding support for genotyping, which was performed at the Johns Hopkins University Center for Inherited Disease Research (CIDR), was provided by the NIH GEI (U01HG004438), the National Institute on Alcohol Abuse and Alcoholism, the National Institute on Drug Abuse, and the NIH contract "High throughput genotyping for studying the genetic contributions to human disease" (HHSN268200782096C). The **GSMS** project (phs000852.v1.p1) was supported by the National Institute on Drug Abuse (U01DA024413 and R01DA11301), the National Institute of Mental Health (R01MH063970, R01MH063671, R01MH048085, K01MH093731, and K23MH080230), NARSAD, and the William T. Grant Foundation. The authors are grateful to all the GSMS and CCC study participants who contributed to this work. The following grants supported data collection and analysis of **CADD** (dbGAP in progress): DA011015, DA012845, DA021913, DA021905, DA032555, and DA035804. **ADAA** was funded by NIH grant R01 AA017444. **Brisbane Longitudinal Twin Study (BLTS)** was supported by the United States National Institute on Drug Abuse (R00DA023549) and by the Australian Research Council (DP0343921, DP0664638, 464914, 619667, and FT110100548). **Yale-Penn** (phs000425.v1.p1 and phs000952.v1.p1) was supported by National Institutes of Health Grants RC2 DA028909, R01 DA12690, R01 DA12849, R01 DA18432, R01 AA11330, and R01 AA017535 and the Veterans Affairs Connecticut and Philadelphia Veterans Affairs Mental Illness Research, Educational, and Clinical Centers. **Australian Alcohol and Nicotine studies (OZ-ALC-NAG**; phs000181.v1.p1) were supported by National Institutes of Health Grants AA07535, AA07728, AA13320, AA13321, AA14041, AA11998, AA17688, DA12854, and DA019951; by Grants from the Australian National Health and Medical Research Council (241944, 339462, 389927,389875, 389891, 389892, 389938, 442915, 442981, 496739, 552485, and 552498); by Grants from the Australian Research Council (A7960034, A79906588, A79801419, DP0770096, DP0212016, and DP0343921); and by the 5th Framework Programme (FP-5) GenomEUtwin Project (QLG2-CT-2002-01254). Genome-wide association study genotyping at the Center for Inherited Disease Research was supported by a grant to the late Richard Todd, MD, PhD, former Principal Investigator of Grant AA13320. The **Finnish Twin Cohort/Nicotine Addiction Genetics-Finland** study was supported by Academy of Finland (grants # 213506 and 129680 to JK), NIH DA12854 (PAFM), Global Research Award for Nicotine Dependence/Pfizer Inc. (JK), Wellcome Trust Sanger Institute, UK, and the European Community's Seventh Framework Programme ENGAGE Consortium (HEALTH-F4-2007- 201413). In **Finntwin12**, support for data collection and genotyping has come from the National Institute of Alcohol Abuse and Alcoholism (grants AA-12502, AA-00145, and AA-09203 to RJR and AA15416 and K02AA018755 to DMD), the Academy of Finland (grants 100499, 205585, 118555, 141054, and 264146 to JK), and the Wellcome Trust Sanger Institute, UK.

Author contribution. ECJ, DL, HJE, JG, and AA conceived the study idea. ECJ, DL, APM, ASH, JDD, MJ, JVB, MG, KS, TT, TE, PJ, TK, CM, YN, TP, MS, PNRV, LW, and IRG contributed to data analysis. ECJ, DL, and AA drafted the manuscript, and all authors edited, reviewed, and approved the final manuscript.

Funding statement. We acknowledge the following sources of support: K01DA051759 (ECJ); T32DA007261 (JVB); K01DA058807 (JDD); K01AA031724 (APM); R01DA054869 (AA, HJE, and JG); DP1DA054394 and T32IR5226 (SSR); R01DA042755, U01DA042217, R01AG077742, R01DA054087, R01DA044283, DA05147, DA13240, DA02441, AA09367, AA11886, and MH066140 (TE, AG, EAW, SV, MM, and JJL); R01DA030976 (KCW and IRG); R01MH100141 (PNRV); R01MH123489, R01MH123619, and R01MH134284 (ARD); AUS NHMRC 464914 (IBH and NGM); R00DA023549 (NAG); and HSRI 67–095 (VS). LD acknowledges support from NHMRC Investigator Grant L3 (2016825), NHMRC Senior Principal Research Fellowship (1135991), the National Drug and Alcohol Research Centre (NDARC), and UNSW funded by the Australian Government Department of Health. The views expressed in this publication do not necessarily represent the position of the Australian Government.

The AGDS was primarily funded by the National Health and Medical Research Council (NHMRC) of Australia (Grant No. APP1086683) to NGM. This work was further supported by NHMRC grants (No. 1145645, 1078901, and 1087889). NGM is supported by an NHMRC Investigator Grant (No. APP 1172990).

GBP: Data collection was funded, and data analysis was supported by the Australian National Health and Medical Research Council (No. APP1138514) to SEM. SEM is supported by a National Health and Medical Research Council Investigator Grant (No. APP1172917).

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

The Psychiatric Genomics Consortium's Substance Use Disorders (PGC-SUD) Working Group receives support from the National Institute on Drug Abuse via R01DA054869. We gratefully acknowledge prior support from the National Institute on Alcohol Abuse and Alcoholism and support from the National Institutes of Mental Health to the overall PGC. Statistical analyses were carried out on the NL Genetic Cluster Computer (http://www.geneticcluster.org) hosted by SURFsara.

Competing interests. The authors declare none.

References

Aguet, F., Brown, A. A., Castel, S. E., Davis, J. R., He, Y., Jo, B., … Biospecimen Collection Source Site—NDRI. (2017). Genetic effects on gene expression across human tissues. *Nature*, **550**(7675), 204–213. https://doi.org/10.1038/nature24277.

Baker, T. B., Breslau, N., Covey, L., & Shiffman, S. (2012). DSM criteria for tobacco use disorder and tobacco withdrawal: A critique and proposed revisions for DSM-5. *Addiction*, **107**(2), 263–275. https://doi.org/10.1111/j.1360-0443.2011.03657.x.

Breslau, N., & Johnson, E. O. (2000). Predicting smoking cessation and major depression in nicotine-dependent smokers. *American Journal of Public Health*, **90**(7), 1122–1127. https://doi.org/10.2105/AJPH.90.7.1122.

Breslau, N., Johnson, E. O., Hiripi, E., & Kessler, R. (2001). Nicotine dependence in the United States: Prevalence, trends, and smoking persistence. *Archives of General Psychiatry*, **58**(9), 810–816. https://doi.org/10.1001/archpsyc.58.9.810.

Brook, J. S., Koppel, J., & Pahl, K. (2009). Predictors of DSM and Fagerström-defined nicotine dependence in African American and Puerto Rican young adults. *Substance Use & Misuse*, **44**(6), 809–822. https://doi.org/10.1080/10826080802483985.

Buchwald, J., Chenoweth, M. J., Palviainen, T., Zhu, G., Benner, C., Gordon, S., … Lehtimäki, T. (2021). Genome-wide association meta-analysis of nicotine metabolism and cigarette consumption measures in smokers of European descent. *Molecular Psychiatry*, **26**(6), 2212–2223.

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

Bulik-Sullivan, B. K., Loh, P.-R., Finucane, H. K., Ripke, S., Yang, J., Patterson, N., … Schizophrenia Working Group of the Psychiatric Genomics Consortium. (2015). LD score regression distinguishes confounding from polygenicity in genome-wide association studies. *Nature Genetics*, **47**(3), 291–295.

Chang, C. C., Chow, C. C., Tellier, L. C., Vattikuti, S., Purcell, S. M., & Lee, J. J. (2015). Second-generation PLINK: Rising to the challenge of larger and richer datasets. *GigaScience*, **4**(1), 7. https://doi.org/10.1186/s13742-015-0047-8.

Coleman, J. R. I., Peyrot, W. J., Purves, K. L., Davis, K. A. S., Rayner, C., Choi, S. W., … on the behalf of Major Depressive Disorder Working Group of the Psychiatric Genomics Consortium. (2020). Genome-wide gene-environment analyses of major depressive disorder and reported lifetime traumatic experiences in UK biobank. *Molecular Psychiatry*, **25**(7), 1430–1446. https://doi.org/10.1038/s41380-019-0546-6.

de Leeuw, C. A., Mooij, J. M., Heskes, T., & Posthuma, D. (2015). MAGMA: Generalized gene-set analysis of GWAS data. *PLoS Computational Biology*, **11**(4). https://doi.org/10.1371/journal.pcbi.1004219.

Deak, J. D., Zhou, H., Galimberti, M., Levey, D. F., Wendt, F. R., Sanchez-Roige, S., … Gelernter, J. (2022). Genome-wide association study in individuals of European and African ancestry and multi-trait analysis of opioid use disorder identifies 19 independent genome-wide significant risk loci. *Molecular Psychiatry*, **27**(10), 3970–3979. https://doi.org/10.1038/s41380-022-01709-1.

Demontis, D., Walters, G. B., Athanasiadis, G., Walters, R., Therrien, K., Nielsen, T. T., … Zeng, B. (2023). Genome-wide analyses of ADHD identify 27 risk loci, refine the genetic architecture and implicate several cognitive domains. *Nature Genetics*, **55**(2), 198–208.

Ge, T., Chen, C.-Y., Ni, Y., Feng, Y.-C. A., & Smoller, J. W. (2019). Polygenic prediction via Bayesian regression and continuous shrinkage priors. *Nature Communications*, **10**(1), 1–10.

Grotzinger, A. D., Rhemtulla, M., de Vlaming, R., Ritchie, S. J., Mallard, T. T., Hill, W. D., … Tucker-Drob, E. M. (2019). Genomic structural equation modelling provides insights into the multivariate genetic architecture of complex traits. *Nature Human Behaviour*, **3**(5), 513–525. https://doi.org/10.1038/s41562-019-0566-x.

Hatoum, A. S., Colbert, S. M. C., Johnson, E. C., Huggett, S. B., Deak, J. D., Pathak, G. A., … Hansen, I. (2023). Multivariate genome-wide association meta-analysis of over 1 million subjects identifies loci underlying multiple substance use disorders. *Nature Mental Health*, **1**(3), 210–223.

Hatoum, A. S., Johnson, E. C., Colbert, S. M. C., Polimanti, R., Zhou, H., Walters, R. K., … Agrawal, A. (2022). The addiction risk factor: A unitary genetic vulnerability characterizes substance use disorders and their associations with common correlates. *Neuropsychopharmacology*, **47**(10), 1739–1745.

Hatoum, A. S., Morrison, C. L., Mitchell, E. C., Lam, M., Benca-Bachman, C. E., Reineberg, A. E., … Friedman, N. P. (2023). Genome-wide association study shows that executive functioning is influenced by GABAergic processes and is a neurocognitive genetic correlate of psychiatric disorders. *Biological Psychiatry*, **93**(1), 59–70.

Huckins, L. M., Dobbyn, A., Ruderfer, D. M., Hoffman, G., Wang, W., Pardiñas, A. F., … CommonMind Consortium, The Schizophrenia Working Group of the Psychiatric Genomics Consortium, iPSYCH-GEMS Schizophrenia Working Group. (2019). Gene expression imputation across multiple brain regions provides insights into schizophrenia risk. *Nature Genetics*, **51**(4), 659–674. https://doi.org/10.1038/s41588-019-0364-4.

Le Foll, B., Piper, M. E., Fowler, C. D., Tonstad, S., Bierut, L., Lu, L., … Hall, W. D. (2022). Tobacco and nicotine use. *Nature Reviews Disease Primers*, **8**(1), 19. https://doi.org/10.1038/s41572-022-00346-w.

Lee, J. J., Wedow, R., Okbay, A., Kong, E., Maghzian, O., Zacher, M., … 23andMe Research Team, COGENT (Cognitive Genomics Consortium), Social Science Genetic Association Consortium. (2018). Gene discovery and polygenic prediction from a genome-wide association study of educational attainment in 1.1 million individuals. *Nature Genetics*, **50**(8), 1112–1121. https://doi.org/10.1038/s41588-018-0147-3.

Levey, D. F., Galimberti, M., Deak, J. D., Wendt, F. R., Bhattacharya, A., Koller, D., … Veterans Affairs Million Veteran Program. (2023). Multi-ancestry genome-wide association study of cannabis use disorder yields insight into disease biology and public health implications. *Nature Genetics*, **55**(12), 2094–2103. https://doi.org/10.1038/s41588-023-01563-z.

Linnér, R. K., Biroli, P., Kong, E., Meddens, S. F. W., Wedow, R., Fontana, M. A., … Hammerschlag, A. R. (2019). Genome-wide association analyses of risk tolerance and risky behaviors in over 1 million individuals identify hundreds of loci and shared genetic influences. *Nature Genetics*, **51**(2), 245.

Loukola, A., Wedenoja, J., Keskitalo-Vuokko, K., Broms, U., Korhonen, T., Ripatti, S., … Kaprio, J. (2014). Genome-wide association study on detailed profiles of smoking behavior and nicotine dependence in a twin sample. *Molecular Psychiatry*, **19**(5), 615–624. https://doi.org/10.1038/mp.2013.72.

Maes, H. H., Sullivan, P. F., Bulik, C. M., Neale, M. C., Prescott, C. A., Eaves, L. J., & Kendler, K. S. (2004). A twin study of genetic and environmental influences on tobacco initiation, regular tobacco use and nicotine dependence. *Psychological Medicine*, **34**(7), 1251–1261. https://doi.org/10.1017/S0033291704002405.

Moolchan, E. T., Radzius, A., Epstein, D. H., Uhl, G., Gorelick, D. A., Cadet, J. L., & Henningfield, J. E. (2002). The Fagerstrom test for nicotine dependence and the diagnostic interview schedule: Do they diagnose the same smokers? *Addictive Behaviors*, **27**(1), 101–113. https://doi.org/10.1016/S0306-4603(00)00171-4.

Mwenifumbo, J. C., & Tyndale, R. F. (2011). DSM-IV, ICD-10 and FTND: Discordant tobacco dependence diagnoses in adult smokers. *Journal of Addiction Research & Therapy*, **2**(1), 105.

Nievergelt, C. M., Maihofer, A. X., Klengel, T., Atkinson, E. G., Chen, C.-Y., Choi, K. W., … Koenen, K. C. (2019). International meta-analysis of PTSD genome-wide association studies identifies sex- and ancestry-specific genetic risk loci. *Nature Communications*, **10**(1), 4558. https://doi.org/10.1038/s41467-019-12576-w.

Pasman, J. A., Verweij, K. J. H., Gerring, Z., Stringer, S., Sanchez-Roige, S., Treur, J. L., … The Substance Use Disorders Working Group of the Psychiatric Genomics Consortium, International Cannabis Consortium (2018). GWAS of lifetime cannabis use reveals new risk loci, genetic overlap with psychiatric traits, and a causal influence of schizophrenia. *Nature Neuroscience*, **21**(9), 1161–1170. https://doi.org/10.1038/s41593-018-0206-1.

Quach, B. C., Bray, M. J., Gaddis, N. C., Liu, M., Palviainen, T., Minica, C. C., … Hancock, D. B. (2020). Expanding the genetic architecture of nicotine dependence and its shared genetics with multiple traits. *Nature Communications*, **11**(1), 5562. https://doi.org/10.1101/2020.01.15.898858.

Ramon, J. M., Morchon, S., Baena, A., & Masuet-Aumatell, C. (2014). Combining varenicline and nicotine patches: A randomized controlled trial study in smoking cessation. *BMC Medicine*, **12**, 1–9.

Reitsma, M. B., Kendrick, P. J., Ababneh, E., Abbafati, C., Abbasi-Kangevari, M., Abdoli, A., … Gakidou, E. (2021). Spatial, temporal, and demographic patterns in prevalence of smoking tobacco use and attributable disease burden in 204 countries and territories, 1990–2013; 2019: A systematic analysis from the global burden of disease study 2019. *The Lancet*, **397**(10292), 2337–2360. https://doi.org/10.1016/S0140-6736(21)01169-7.

Saunders, G. R. B., Wang, X., Chen, F., Jang, S.-K., Liu, M., Wang, C., … 23andMe Research Team; Biobank Japan Project. (2022). Genetic diversity fuels gene discovery for tobacco and alcohol use. *Nature*, **612**(7941), 720–724. https://doi.org/10.1038/s41586-022-05477-4.

Tashkin, D. P., Rennard, S., Hays, J. T., Ma, W., Lawrence, D., & Lee, T. C. (2011). Effects of Varenicline on smoking cessation in patients with mild to

https://doi.org/10.1017/S0033291725100883 Published online by Cambridge University Press

moderate COPD: A randomized controlled trial. *Chest*, **139**(3), 591–599. https://doi.org/10.1378/chest.10-0865.

Toikumo, S., Jennings, M. V., Pham, B. K., Lee, H., Mallard, T. T., Bianchi, S. B., … PsycheMERGE Substance Use Disorder Workgroup. (2024). Multi-ancestry meta-analysis of tobacco use disorder identifies 461 potential risk genes and reveals associations with multiple health outcomes. *Nature Human Behaviour*, **8**(6), 1177–1193. https://doi.org/10.1038/s41562-024-01851-6.

Trabzuni, D., Ryten, M., Walker, R., Smith, C., Imran, S., Ramasamy, A., … Hardy, J. (2011). Quality control parameters on a large dataset of regionally dissected human control brains for whole genome expression studies. *Journal of Neurochemistry*, **119**(2), 275–282.

Watanabe, K., Taskesen, E., van Bochoven, A., & Posthuma, D. (2017). Functional mapping and annotation of genetic associations with FUMA. *Nature Communications*, **8**(1), 1826. https://doi.org/10.1038/s41467-017-01261-5.

Willer, C. J., Li, Y., & Abecasis, G. R. (2010). METAL: Fast and efficient meta-analysis of genomewide association scans. *Bioinformatics*, **26**(17), 2190–2191.

Yengo, L., Sidorenko, J., Kemper, K. E., Zheng, Z., Wood, A. R., Weedon, M. N., … GIANT Consortium. (2018). Meta-analysis of genome-wide association studies for height and body mass index in ∼700000 individuals of European ancestry. *Human Molecular Genetics*, **27**(20), 3641–3649. https://doi.org/10.1093/hmg/ddy271.

Yengo, L., Vedantam, S., Marouli, E., Sidorenko, J., Bartell, E., Sakaue, S., … Raghavan, S. (2022). A saturated map of common genetic variants associated with human height. *Nature*, **610**(7933), 704–712.

Zhang, H., Grant, B. F., Hodgkinson, C. A., Ruan, W. J., Kerridge, B. T., Huang, B., … Chou, S. P. (2022). Strong and weak cross-inheritance of substance use disorders in a nationally representative sample. *Molecular Psychiatry*, **27**(3), 1742–1753. https://doi.org/10.1038/s41380-021-01370-0.

Zhou, H., Kember, R. L., Deak, J. D., Xu, H., Toikumo, S., Yuan, K., … Million Veteran Program. (2023). Multi-ancestry study of the genetics of problematic alcohol use in over 1 million individuals. *Nature Medicine*, **29**(12), 3184–3192. https://doi.org/10.1038/s41591-023-02653-5.

https://doi.org/10.1017/S0033291725100883 Published online by Cambridge University Press