medRxiv preprint doi: https://doi.org/10.1101/2023.01.24.23284960; this version posted January 30, 2023. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-NC-ND 4.0 International license .

Multi-ancestry study of the genetics of problematic alcohol use in >1 million individuals

Hang Zhou[1,2,46], Rachel L. Kember[3,4,46], Joseph D. Deak[1,2], Heng Xu[4], Sylvanus Toikumo[3,4], Kai Yuan[5,6], Penelope A. Lind[7,8,9], Leila Farajzadeh[10,11,12], Lu Wang[1,2], Alexander S. Hatoum[13], Jessica Johnson[14,15], Hyunjoon Lee[16], Travis T. Mallard[5,16,17], Jiayi Xu[1], Keira J.A. Johnston[1], Emma C. Johnson[18], Marco Galimberti[1,2], Cecilia Dao[2,19], Daniel F. Levey[1,2], Cassie Overstreet[1,2], Enda M. Byrne[20], Nathan A. Gillespie[21], Scott Gordon[22], Ian B. Hickie[23], John B. Whitfield[22], Ke Xu[1,2], Hongyu Zhao[24,25], Laura M. Huckins[1], Lea K. Davis[26,27,28], Sandra Sanchez-Roige[27,29], Pamela A. F. Madden[18], Andrew C. Heath[18], Sarah E. Medland[7,9,30], Nicholas G. Martin[22], Tian Ge[5,31,32], Jordan W. Smoller[5,31,33], David M. Hougaard[11,34,], Anders D. Børglum[10,11,12], Ditte Demontis[10,11,12], John H. Krystal[1,2,35,36], J. Michael Gaziano[37,38,39], Howard J. Edenberg[40,41,] Arpana Agrawal[18], Million Veteran Program[*], Amy C. Justice[2,42,43], Murray B. Stein[29,44,45,] Henry R. Kranzler[3,4,47], and Joel Gelernter[1,2,24,35,47]

[1]Department of Psychiatry, Yale School of Medicine, New Haven, CT, USA
[2]Veterans Affairs Connecticut Healthcare System, West Haven, CT, USA
[3]Crescenz Veterans Affairs Medical Center, Philadelphia, PA, USA
[4]Department of Psychiatry, University of Pennsylvania Perelman School of Medicine, Philadelphia, PA, USA
[5]Stanley Center for Psychiatric Research, The Broad Institute of MIT and Harvard, Cambridge, MA, USA
[6]Analytic and Translational Genetics Unit, Department of Medicine, Massachusetts General Hospital, Boston, MA, USA
[7]Psychiatric Genetics, QIMR Berghofer Medical Research Institute, Brisbane, QLD, Australia
[8]School of Biomedical Sciences, Queensland University of Technology, Brisbane, QLD, Australia
[9]Faculty of Medicine, University of Queensland, Brisbane, QLD, Australia
[10]Department of Biomedicine - Human Genetics, Aarhus University, Aarhus, Denmark
[11]The Lundbeck Foundation Initiative for Integrative Psychiatric Research, iPSYCH, Denmark
[12]Center for Genomics and Personalized Medicine, Aarhus, Denmark
[13]Department of Psychological and Brain Sciences, Washington University in St. Louis, Saint Louis, MO, USA
[14]Pamela Sklar Division of Psychiatric Genomics, Department of Psychiatry, Icahn School of Medicine at Mount Sinai, New York, NY, USA
[15]Department of Genetics and Genomic Sciences, Icahn School of Medicine at Mount Sinai, New York, NY, USA
[16]Psychiatric and Neurodevelopmental Genetics Unit, Center for Genomic Medicine, Massachusetts General Hospital, Boston, MA, USA
[17]Department of Psychiatry, Massachusetts General Hospital, Harvard Medical School, Boston, MA, USA
[18]Department of Psychiatry, Washington University School of Medicine, Saint Louis, MO, USA
[19]Department of Chronic Disease Epidemiology, Yale University School of Public Health, New Haven, CT, USA
[20]Child Health Research Centre, The University of Queensland, Brisbane, QLD, Australia
[21]Institute for Psychiatric and Behavioral Genetics, Department of Psychiatry, Virginia Commonwealth University, Richmond, VA, USA
[22]Genetic Epidemiology, QIMR Berghofer Medical Research Institute, Brisbane, QLD, Australia
[23]Brain and Mind Centre, University of Sydney, Camperdown, NSW, Australia
[24]Department of Biostatistics, Yale School of Public Health, New Haven, CT, USA

NOTE: This preprint reports new research that has not been certified by peer review and should not be used to guide clinical practice.

medRxiv preprint doi: https://doi.org/10.1101/2023.01.24.23284960; this version posted January 30, 2023. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-NC-ND 4.0 International license .

[25]Department of Genetics, Yale School of Medicine, New Haven, CT, USA
[26]Vanderbilt Genetics Institute, Vanderbilt University Medical Center, Nashville, TN, USA
[27]Department of Medicine, Division of Medical Genetics, Vanderbilt University Medical Center, Nashville, TN, USA
[28]Department of Psychiatry and Behavioral Sciences, Vanderbilt University Medical Center, Nashville, TN, USA
[29]Department of Psychiatry, University of California San Diego, La Jolla, CA, USA
[30]School of Psychology, University of Queensland, Brisbane, QLD, Australia
[31]Psychiatric and Neurodevelopmental Genetics Unit, Center for Genomic Medicine, Massachusetts General Hospital, Boston, MA, USA
[32]Center for Precision Psychiatry, Massachusetts General Hospital, Boston, MA, USA
[33]Department of Psychiatry, Massachusetts General Hospital, Harvard Medical School, Boston, MA, USA
[34]Center for Neonatal Screening, Department for Congenital Disorders, Statens Serum Institut, Copenhagen, Denmark
[35]Department of Neuroscience, Yale University School of Medicine, New Haven, CT, USA
[36]National Center for PTSD, US Department of Veterans Affairs, West Haven, CT, USA
[37]Massachusetts Veterans Epidemiology and Research Information Center (MAVERIC), Boston Veterans Affairs Healthcare System, Boston, MA, USA
[38]Department of Medicine, Divisions of Aging and Preventative Medicine, Brigham and Women's Hospital, Boston, MA, USA
[39]Department of Medicine, Harvard Medical School, Boston, MA, USA
[40]Department of Biochemistry and Molecular Biology, Indiana University School of Medicine, Indianapolis, IN, USA
[41]Department of Medical and Molecular Genetics, Indiana University School of Medicine, Indianapolis, IN, USA
[42]Department of Internal Medicine, Yale School of Medicine, New Haven, CT, USA
[43]Center for Interdisciplinary Research on AIDS, Yale School of Public Health, New Haven, CT, USA
[44]VA San Diego Healthcare System, Psychiatry Service, San Diego, CA, USA
[45]Herbert Wertheim School of Public Health and Human Longevity Science, University of California San Diego, La Jolla, CA, USA
[46]These authors contributed equally
[47]These authors jointly supervised this work
[*]A list of members and their affiliations appears in the Supplementary Information

Corresponding Authors:

Joel Gelernter, Department of Psychiatry, Yale School of Medicine, Veterans Affairs Connecticut Healthcare System, 116A2, 950 Campbell Ave, West Haven, CT 06516, USA. Email: joel.gelernter@yale.edu.

Hang Zhou, Department of Psychiatry, Yale School of Medicine; Veterans Affairs Connecticut Healthcare System, 43A35, 950 Campbell Ave, West Haven, CT 06516, USA. Email: hang.zhou@yale.edu

medRxiv preprint doi: https://doi.org/10.1101/2023.01.24.23284960; this version posted January 30, 2023. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-NC-ND 4.0 International license .

Abstract

Problematic alcohol use (PAU) is a leading cause of death and disability worldwide. To improve our understanding of the genetics of PAU, we conducted a large cross-ancestry meta-analysis of PAU in 1,079,947 individuals. We observed a high degree of cross-ancestral similarity in the genetic architecture of PAU and identified 110 independent risk variants in within- and cross-ancestry analyses. Cross-ancestry fine-mapping improved the identification of likely causal variants. Prioritizing genes through gene expression and/or chromatin interaction in brain tissues identified multiple genes associated with PAU. We identified existing medications for potential pharmacological studies by drug repurposing analysis. Cross-ancestry polygenic risk scores (PRS) showed better performance in independent sample than single-ancestry PRS. Genetic correlations between PAU and other traits were observed in multiple ancestries, with other substance use traits having the highest correlations. The analysis of diverse ancestries contributed significantly to the findings, and fills an important gap in the literature.

medRxiv preprint doi: https://doi.org/10.1101/2023.01.24.23284960; this version posted January 30, 2023. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-NC-ND 4.0 International license .

Introduction

Alcohol use disorder (AUD) is a chronic relapsing disease associated with a host of adverse medical, psychiatric, and social consequences[1]. Given observed heritability (h^2~50%[2]), there has been substantial progress made in genome-wide association studies (GWAS) of AUD and problematic drinking[3-9], and also for measures of alcohol consumption[10,11]. A prior GWAS of problematic alcohol use (PAU, a phenotype based on a meta-analysis of highly genetically correlated (genetic correlations >0.7) AUD[7], alcohol dependence [AD][5], and AUDIT-P [Alcohol Use Disorders Identification Test-Problem score, a measure of problematic drinking][4,12], N=435,563) identified 29 independent risk variants, predominantly in European (EUR) ancestry subjects[6]. Consistent with genetic studies of other complex traits, and the high polygenicity of PAU, larger and more ancestrally-representative samples need to be examined to outline the genetic architecture of these alcohol use traits.

A key finding from recent studies is that both AUD and AUDIT-P differ phenotypically and genetically from typical alcohol consumption[4,7]. AUD and AUDIT-P index aspects of disordered alcohol intake and correlate with genetic liability to negative psychiatric and psychosocial factors (e.g., higher major depressive disorder [MDD], lower educational attainment). An item-level study of the AUDIT questionnaire confirmed a two-factor structure at the genetic level, underscoring unique genetic influences on alcohol consumption and alcohol-related problems[13] and noted that the genetics of drinking frequency were confounded by socio-economic status. A similar pattern – genetic distinctions between substance use disorder (SUD) vs. non-dependent use – has also been observed for cannabis use disorder and cannabis use[14]. Furthermore, aggregating across multiple substance use disorders suggests that problematic and disordered substance use has a unique genetic architecture that, while shared across SUDs, does not overlap fully with non-dependent substance use *per se*[15].

medRxiv preprint doi: https://doi.org/10.1101/2023.01.24.23284960; this version posted January 30, 2023. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-NC-ND 4.0 International license .

Notwithstanding prior discovery of multiple genome-wide significant (GWS) loci for PAU, there are major gaps in our understanding of its genetic underpinnings. First, the estimated single-nucleotide polymorphism (SNP)-based heritability (h^2) of AUD and PAU ranges from 5.6% to 10.0%[4-7], reflecting substantial "missing heritability" compared to estimates based on genetic epidemiology, which show ~50% heritability[2]. Second, most of the available samples used in human genetic studies – including for AUD – are of EUR ancestry; lack of ancestral diversity is a major problem both for understanding the genetics of these traits, and for potential applications of these genetic discoveries to global populations[16]. Our previous study in the Million Veteran Program (MVP) analyzed AUD in multiple ancestral groups[7]. However, non-EUR samples (N=72,387) were far smaller than EUR samples (N=202,004), resulting in inadequate statistical power and unbalanced gene discovery across ancestral backgrounds, which limits our understanding of the genetic architecture underlying the trait across populations.

To improve our understanding of the biology of PAU in multiple populations, we conducted substantially larger ancestry-specific GWAS of PAU followed by a cross-ancestry meta-analysis in 1,079,947 individuals from multiple cohorts. We identified 85 independent risk variants in EUR participants (almost tripling the number identified in previous studies) and 110 in the within-ancestry and cross-ancestry meta-analyses. We investigated the shared genetic architectures of PAU across different ancestries, performed fine-mapping for causal variants by combining information from multiple ancestries, and tested cross-ancestry polygenic risk score (PRS) associations with AUDIT-P in the UK Biobank (UKB) samples[17]. We combined genes identified by gene-based association analysis, transcriptome-wide association analysis (TWAS) and brain-chromatin interaction analysis, found dozens of genes linking to brain with convergent evidence. Drug repurposing analysis identified potential medications for further pharmacological studies, bringing forward the hope of novel biologically-directed medications strategies with the

medRxiv preprint doi: https://doi.org/10.1101/2023.01.24.23284960; this version posted January 30, 2023. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-NC-ND 4.0 International license .

further potential of personalization. We conducted phenome-wide PRS analyses in biobanks from the PsycheMERGE Network[18] in AFR and EUR-ancestry samples. We tested the genetic correlation between PAU and other traits, especially novel with respect to AFR samples where such analyses could not be conducted previously. These findings substantially augment the number of loci that contribute to risk of PAU, increasing power to investigate the causal relationships of PAU with other diseases, and identify novel druggable targets whose therapeutic potential requires empirical evaluation.

Results

Ancestrally diverse data collection. We collected newly genotyped subjects (most from MVP) and previously published data from multiple cohorts (MVP[19], FinnGen[20], UKB[17], Psychiatric Genomics Consortium (PGC)[5], iPSYCH[21,22], Queensland Berghofer Medical Research Institute (QIMR) cohorts[23-25], Yale-Penn 3[26], and East Asian cohorts[27]) resulting in a total of 1,079,947 subjects (Table 1). Five ancestral groups were analyzed (Figure 1a): EUR (N=903,147), AFR (N=122,571), Latin American (LA, N=38,962), East Asian (EAS, N=13,551, all published in ref. 27), and South Asian (SAS, N=1,716). As in our previous study[6], we utilized data on International Classification of Diseases (ICD)-diagnosed AUD (N_{case}=136,182; $N_{control}$=692,594), DSM-IV AD (N_{case}=29,770; $N_{control}$=70,282) and AUDIT-P (N=151,119), together defined as problematic alcohol use (PAU; based on high genetic correlations across these measures). The total number of AUD and AD cases was 165,952, almost double the 85,391 cases in the previously largest study[28].

Figure 1. Genetic architecture of problematic alcohol use (PAU). a, Sample sizes in different ancestral groups. **b**, Relationship between sample size and number of independent

medRxiv preprint doi: https://doi.org/10.1101/2023.01.24.23284960; this version posted January 30, 2023. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-NC-ND 4.0 International license .

variants identified. Kranzler et al., 2019: cross-ancestry meta-analysis for AUD; Zhou et al., 2020: PAU in EUR. **c**, Lookup for cross-ancestry replication in AFR for the 85 independent variants in EUR meta-analysis. Of the 85 variants, 76 could be analyzed in AFR (see Methods). Sign test was performed for the number of variants with same direction of effect (64/76, p=1.0×10^{-9}). 23 variants were nominally significant in AFR and 6 were significant after multiple correction (p<0.05/76). **d**, Observed-scale and liability-scale SNP-based heritability (h^2) in multiple ancestries. **e**, Cross-ancestry genetic-effect correlation (ρ_{ge}) and genetic-impact correlation (ρ_{gi}) between EUR, AFR and LA ancestries. Error bar is the 95% confidence interval. **f**, Genome-wide association results for PAU in the cross-ancestry meta-analysis. Red line is significance threshold of 5×10^{-8}. EUR, European; AFR, African; LA, Latin American; EAS, East Asian; SAS, South Asian; GWS, genome-wide significant.

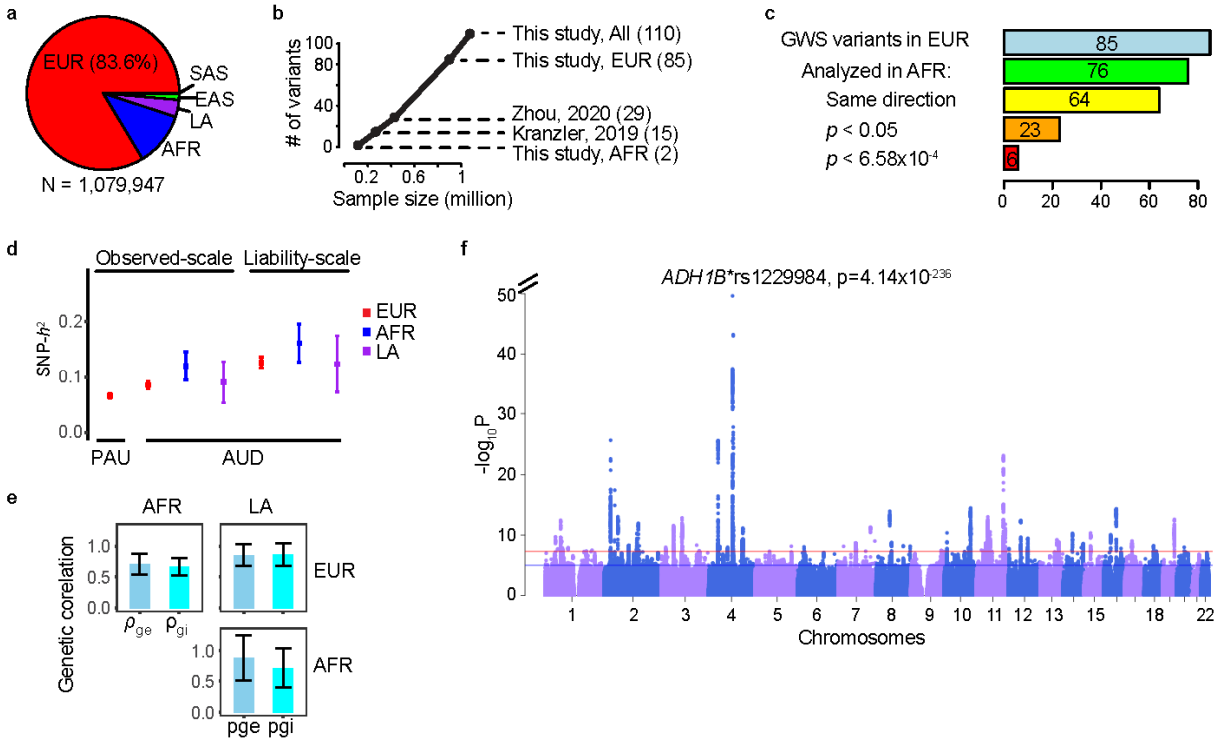


Table 1. Demographics for cohorts in the meta-analysis of PAU. Cohorts are described in

medRxiv preprint doi: https://doi.org/10.1101/2023.01.24.23284960; this version posted January 30, 2023. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-NC-ND 4.0 International license .

the Methods. UKB-EUR1: genetically defined White-British by UK Biobank; UKB-EUR2: genetically defined European non-White-British participants (see Methods); AGDS, the Australian Genetics of Depression Study; TWINS, the Australian twin-family study of alcohol use disorder; GBP, the Australian Genetics of Bipolar Disorder Study; iPSYCH1, phase 1 of iPSYCH; iPSYCH2, phase 2 of iPSYCH; YP3, Yale-Penn 3; $N_{effective}$, effective sample size; Thai, study of the genetics of methamphetamine dependence in Thailand; GSA, Illumina Global Screening Array; MEGA, Illumina Multi-Ethnic Global Array; Cyto, Illumina Cyto12 array.

Cohorts	**Traits**	**N_{case}**	**$N_{control}$**	**N_{total}**	**N_{female} (%)**	**$N_{effective}$**	**Ref[a]**
			European ancestry (EUR)				
MVP	AUD	80,028	368,113	448,141	33,345 (7.4)	262,947	[6] and new
FinnGen	AUD	8,866	209,926	218,792	123,579 (56.5)	34,027	New[b]
UKB-EUR1	AUDIT-P	-	-	132,001	74,113 (56.1)	132,001	[6] and new
UKB-EUR2	AUDIT-P	-	-	17,898	10,529 (58.5)	17,898	New
PGC	AD	9,938	30,992	40,930	20,933 (51.1)	23,075	[5][d]
QIMR_AGDS	AD	6,726	4,467	11,193	8,605 (76.9)	10,737	New
QIMR_TWINS	AD	2,772	5,630	8,402	4,922 (58.6)	7,430	[5] and new
QIMR_GBP	AD	1,287	751	2,038	1,435 (70.4)	1,897	New
iPSYCH1	AD	2,117	13,238	15,355	8,077 (52.6)	7,301	New
iPSYCH2	AD	1,024	5,732	6,756	3,607 (53.4)	3,475	New
YP3	AD	567	1,074	1,641	854 (52.0)	1,484	New
Sub_total	**PAU**	**113,325**	**639,923**	**903,147**	**289,999 (32.1)**	**502,272**	
			African ancestry (AFR)				
MVP	AUD	36,330	79,100	115,430	16,084 (13.9)	99,583	[7] and new
PGC	AD	3,335	2,945	6,280	3,124 (49.7)	4,991	[5]
YP3	AD	451	410	861	430 (50.0)	959	New
Sub_total	**AUD**	**40,116**	**82,455**	**122,571**	**19,638 (16.0)**	**105,433**	
			Latin American ancestry (LA)				
MVP	**AUD**	**10,150**	**28,812**	**38,962**	**3,731 (9.6)**	**30,023**	[7] and new
			East Asian[a] ancestry (EAS)				
MVP	AUD	701	6,254	6,955	747 (10.7)	2,521	[27]
Han Chinese–GSA	AD	533	2,848	3,381	1,012 (29.9)	1,796	
Thai METH–MEGA	AD	794	1,576	2,370	1,008 (42.5)	2,112	
Thai METH–GSA	AD	127	405	532	263 (49.4)	387	
Han Chinese–Cyto	AD	99	214	313	0 (0)	271	
Sub_total	**AUD**	**2,254**	**11,297**	**13,551**	**3,030 (22.4)**	**7,087**	
			South Asian ancestry (SAS)				
MVP	AUD	107	389	496	67 (13.5)	336	[7] and new
UKB-SAS	AUDIT-P	-	-	1,220	535 (43.9)	1,220	New
Sub_total	**PAU**	**107**	**389**	**1,716**	**602 (35.1)**	**1,556**	
Total	**PAU**	**165,952**	**762,876**	**1,079,947**	**317,000 (29.4)**	**646,371**	

medRxiv preprint doi: https://doi.org/10.1101/2023.01.24.23284960; this version posted January 30, 2023. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-NC-ND 4.0 International license .

Note: [a]data either published in previous alcohol GWAS or newly included for this project. [b]FinnGen summary statistics were downloaded from FinnGen data freeze v5 (https://r5.finngen.fi/). [c]Included related subjects from UKB. [d]Reran the PGC AD GWAS in EUR excluding two Australian cohorts.

Genome-wide association results for PAU. We performed GWAS and within-ancestry meta-analyses for PAU in five ancestral groups and then completed a cross-ancestry meta-analysis. In the EUR meta-analysis, 113,325 cases of AUD/AD, 639,923 controls and 149,899 participants with AUDIT-P scores were analyzed (Supplementary Figure 1a). After conditional analysis, 85 independent variants at 75 loci reached GWS (Methods, Supplementary Table 1) (see also Figure 1b). Of these variants, 41 are in protein-coding genes; of these, 5 are missense variants (*GCKR**rs1260326; *ADH1B**rs75967634; *ADH1B**rs1229984; *SCL39A8**rs13107325; *BDNF**rs6265).

Due to the smaller sample numbers, the non-EUR GWAS yielded fewer variants associated with PAU than did the EUR GWAS (Supplementary Table 1). The AFR meta-analysis found two independent *ADH1B* missense variants (rs1229984 and rs2066702) associated with AUD (Figure 1b, Supplementary Figure 1b); these were reported previously[7,26]. In the LA samples from MVP, only *ADH1B**rs1229984 (lead SNP) was identified (Supplementary Figure 1c). Two independent risk variants, *ADH1B**rs1229984 and *BRAP**rs3782886, were reported in EAS previously[27,29]. In the small SAS meta-analysis, one intergenic variant (rs12677811) was associated with AUD; however, this SNP was present only in the UKB (Supplementary Figure 1d).

Of the 85 lead variants identified in the EUR GWAS, 76 were either directly analyzed or had proxy variants in AFR (Methods, Supplementary Table 2, Figure 1c), 64 of which had the

medRxiv preprint doi: https://doi.org/10.1101/2023.01.24.23284960; this version posted January 30, 2023. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-NC-ND 4.0 International license .

same direction of effect (sign test p=1.00×10^{-9}). Of these, 23 were nominally associated (p<0.05) and 6 were significantly associated with AUD after multiple-testing correction (p<6.58×10^{-4}). In LA, 15 of the EUR GWS variants were nominally significant (p<0.05) and 2 were significantly associated with AUD (rs12048727 and rs1229984). In EAS, 5 variants were nominally significant and two were significantly associated with AUD (rs1229984 and rs10032906). Only two variants were nominally associated with PAU in SAS (rs1229984 was not present in SAS).

We estimated the SNP-based heritability (h^2) for PAU and AUD (excluding AUDIT-P from UKB) in EUR, AFR and LA; significant h^2 estimates (range from 0.066 to 0.127) were observed (Figure 1d, Supplementary Table 3).

High genetic correlations were observed across the EUR, AFR, and LA ancestries (Figure 1e, Supplementary Table 4). The genetic-effect correlation (ρ_{ge}) is 0.71 (SE=0.09, p=6.16×10^{-17}) between EUR and AFR, 0.85 (SE=0.09, p=3.14×10^{-22}) between EUR and LA, and 0.88 (SE=0.18, p=1.58×10^{-6}) between AFR and LA. The genetic-impact correlation (ρ_{gi}) is 0.67 (SE=0.07, p=2.78×10^{-21}) between EUR and AFR, 0.86 (SE=0.09, p=3.52×10^{-20}) between EUR and LA, and 0.72 (SE=0.16, p=9.63×10^{-6}) between AFR and LA. The estimates involving smaller study populations were not robust (Bonferroni p>0.05).

In the cross-ancestry meta-analysis of all available datasets, we identified 100 independent variants at 90 loci (Figure 1f, Supplementary Table 1); 80 are novel findings for PAU. Of these, 53 variants were located in protein-coding genes, of which 9 are missense variants: *GCKR**rs1260326; *ADH1B**rs75967634, rs1229984, and rs2066702; *SCL39A8**rs13107325; *OPRM1**rs1799971; *SLC25A37**rs2942194; *BDNF**rs6265; and

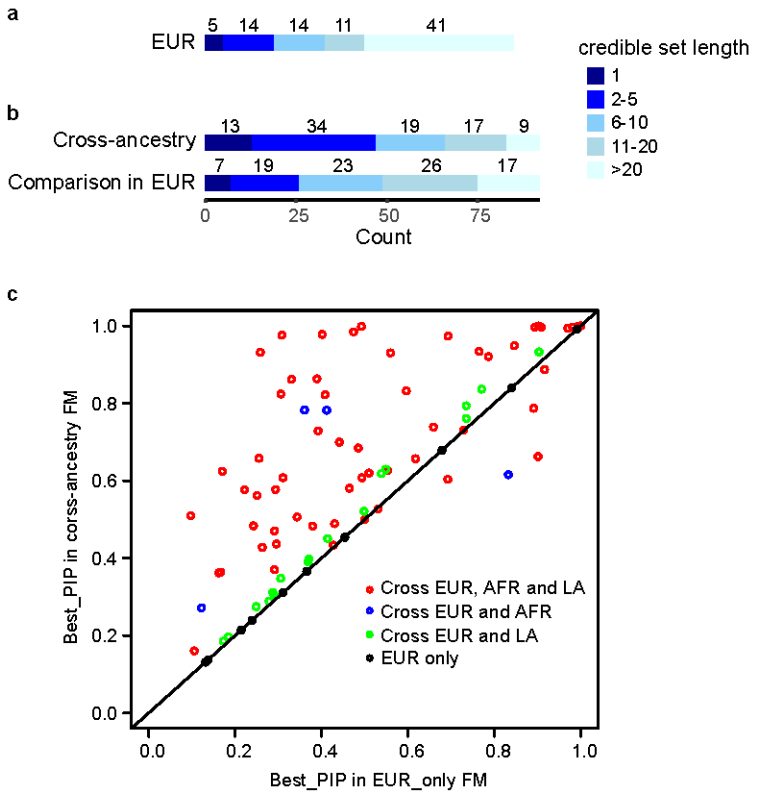
medRxiv preprint doi: https://doi.org/10.1101/2023.01.24.23284960; this version posted January 30, 2023. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-NC-ND 4.0 International license .

*BRAP**rs3782886. The cross-ancestry meta-analysis identified 24 more risk variants than the EUR meta-analysis, but 9 EUR variants fell below GWS (p-values ranging from 5.26×10^{-6} to 9.84×10^{-8}). In total, 110 unique variants were associated with PAU in either the within-ancestry or cross-ancestry analyses (Figure 1b, Supplementary Table 1).

Within- and cross-ancestry causal variant fine-mapping. We performed within-ancestry fine-mapping for the 85 clumped regions with independent lead variants in EUR (Supplementary Tables 5 and 6). A median number of 115 SNPs were included in each region to estimate the credible sets with 99% posterior inclusion probability (PIP) of causal variants. After fine-mapping, the median number of SNPs constituting the credible sets was reduced to 20. Among the 85 regions, there were 5 credible sets that include only a single variant with PIP ≥99% (presumably indicating successful identification of specific causal variants): rs1260326 in *GCKR*, rs472140 and rs1229984 in *ADH1B*, rs2699453 (intergenic), and rs2098112 (intergenic). Another 19 credible sets contained ≤5 variants (Figure 2a).

Figure 2. Fine-mapping for PAU. a, Fine-mapping of causal variants in 85 regions in EUR. **b**, 92 regions in cross-ancestry analysis were fine-mapped and a direct comparison was done for these regions in EUR. **c**, Comparison for the highest PIPs from cross-ancestry and EUR-only fine-mapping in the 92 regions. Red dots are the regions fine-mapped across EUR, AFR, and LA; blue dots are the regions fine-mapped across EUR and AFR; green dots are the regions fine-mapped across EUR and LA; black dots are the regions only fine-mapped in EUR. FM, fine-mapping.

medRxiv preprint doi: https://doi.org/10.1101/2023.01.24.23284960; this version posted January 30, 2023. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-NC-ND 4.0 International license .



We performed cross-ancestry fine-mapping to identify credible sets with 99% PIP for causal variants proximate to 92 independent lead variants in the cross-ancestry meta-analysis (Supplementary Tables 7 and 8). The median number of SNPs in the credible sets was 9. 13 credible sets contain only a single variant with PIP ≥99%; 47 credible sets contain ≤5 variants (Figure 2b). For example, fine-mapping the region proximate to lead SNP rs12354219 (which maps to *DYPD* on chromosome 1) identified rs7531138 as the most likely potential causal variant (PIP=48%), although this variant and rs12354219 (PIP=11%) are in high linkage disequilibrium (LD) in different populations (r^2 ranges from 0.76 to 0.99). In a cross-ancestry meta-analysis rs7531138 showed significant association with schizophrenia (p=1.04×10^{-8}), but rs12354219 (p=6.18×10^{-8}) did not[30] (although the two p-values were very similar). rs7531138 is also a lead SNP associated with educational attainment (p=1.74×10^{-11}), unlike rs12354219 (p>5×10^{-8})[31].

medRxiv preprint doi: https://doi.org/10.1101/2023.01.24.23284960; this version posted January 30, 2023. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-NC-ND 4.0 International license .

To compare within- and cross-ancestry fine-mapping, we performed fine-mapping for the above 92 regions using the same SNP sets and EUR-only LD information (Figure 2b & 2c). The median number of SNPs in the credible sets is 13, with 7 credible sets containing a single variant and 26 containing ≤5 variants, indicating that cross-ancestry fine-mapping improved causal variant identification, consistent with other studies reporting improved fine-mapping by including other ancestries[11].

Gene-based association analysis. We used MAGMA[32,33] to perform gene-based association analyses. 130 genes in EUR, 9 in AFR and 6 in LA (for AFR and LA populations, all mapped to the ADH gene cluster), and 7 in EAS (mapped to either the ADH gene cluster or the *ALDH2* region[27]) were associated with PAU or AUD (Supplementary Table 9). There were no significant findings in SAS.

Transcriptome-wide association analyses (TWAS). We used S-PrediXcan[34] to identify predicted gene expression associations with PAU in 13 brain tissues[35]. 426 significant gene-tissue associations were identified, representing 89 unique genes (Supplementary Table 10). Five genes showed associations with PAU in all available brain tissues, including *AMT* (Aminomethyltransferase), *YPEL3* (Yippee Like 3), *EVI2A* (Ecotropic Viral Integration Site 2A), *EVI2B* (Ecotropic Viral Integration Site 2B), and *CTA-223H9.9* (lncRNA). We also observed associations between PAU and the expression of alcohol dehydrogenase genes (*ADH1B* in the putamen (basal ganglia), *ADH1C* in 10 brain tissues, and *ADH5* in cerebellar hemisphere and cerebellum). Among the brain tissues, caudate (basal ganglia) had the most genes whose expression was associated with PAU (42 genes), followed by the putamen (basal ganglia) (39

medRxiv preprint doi: https://doi.org/10.1101/2023.01.24.23284960; this version posted January 30, 2023. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-NC-ND 4.0 International license .

genes). TWAS that integrated evidence across 13 brain tissues using S-MultiXcan[36] to test joint effects of gene expression variation identified 121 genes (81 shared with S-PrediXcan) whose expression was associated with PAU (Supplementary Table 11).

Linking risk genes to brain chromatin interaction. We used H-MAGMA[37] to implicate risk genes associated with PAU by incorporating brain chromatin interaction profiles. 1,030 gene-chromatin associations were identified in 6 brain Hi-C annotations, representing 401 unique genes (Supplementary Table 12). 58 genes showed association with chromatin interaction in all 6 annotations, including *ADH1B*, *ADH1C*, *DRD2*, *EVI2A* and others that also showed evidence by TWAS in brain tissues.

Convergent evidence linking association to brain. We examined overlapped genes by both gene-based association analysis and TWAS in brain tissues and/or H-MAGMA analysis using Hi-C brain annotations. Among the 130 genes associated with PAU in EUR, 60 were also implicated by TWAS findings either by single brain tissue (S-PrediXcan) or across brain tissues (S-MultiXcan), 82 have evidence of brain chromatin interaction, and 38 have evidence from both TWAS and Hi-C annotations including *ADH1B*, *DRD2*, *KLB* and others (Supplementary Table 9).

Probabilistic fine-mapping of TWAS. We performed fine-mapping for TWAS using FOCUS[38], a method that estimates credible gene sets predicted to include the causal gene that can be prioritized for functional assays. We detected 53 credible sets at a nominal confidence level (set at 90% PIP). These contained 145 gene-tissue associations with an average PIP of 32% (Supplementary Table 13). For the 19 gene-tissue associations having PIP >90%, 9 are from brain tissues (e.g., *ZNF184* expression in hypothalamus (PIP=0.94%), *MTCH2* expression in

medRxiv preprint doi: https://doi.org/10.1101/2023.01.24.23284960; this version posted January 30, 2023. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-NC-ND 4.0 International license .

nucleus accumbens (basal ganglia) (PIP=99%), *SLC4A8* expression in dorsolateral prefrontal cortex (PIP=98%), *YPEL3* expression in cerebellum (PIP=100%), and *CHD9* expression in dorsolateral prefrontal cortex (PIP=100%).

Drug repurposing. Independent genetic signals from the cross-ancestry meta-analysis were searched in OpenTargets.org[39] for druggability and medication target status based on nearest genes. Among them, *OPRM1* implicated naltrexone and *GABRA4* implicated acamprosate, both current treatments for AUD. Additionally, the genes *DRD2*, *CACNA1C*, *DPYD*, *PDE4B*, *KLB*, *BRD3*, *NCAM1*, *FTOP*, and *MAPT*, were identified as druggable genes.

From the drug repurposing analysis using S-PrediXcan results, 287 compounds were significantly correlated with the transcriptional pattern associated with risk for PAU (Supplementary Table 14). Of these 287, 141 medications were anti-correlated with the transcriptional pattern. Of those, trichostatin-a (p=3.29×10^{-35}), melperone (p=6.88×10^{-11}), triflupromazine (p=7.37×10^{-10}), spironolactone (p=2.45×10^{-9}), amlodipine (p=1.42×10^{-6}) and clomethiazole (p=1.30×10^{-5}) reversed the transcriptional profile associated with increased PAU risk, targeted a gene near an independent significant locus in the cross-ancestry GWAS.

Cross-ancestry polygenic risk score association. We tested the cross-ancestry PRS association with AUDIT-P in UKB using AUD summary data from EUR (leaving out the UKB AUDIT-P data), AFR, and LA. PRS-CSx was applied to calculate the posterior effect sizes for each SNP by leveraging LD diversity across discovery samples[40]. We validated the PRS associations with AUDIT-P in UKB-EUR2 and tested them in UKB-EUR1 (see Table 1). In the UKB-EUR1 samples, EUR-based AUD PRS is significantly associated with AUDIT-P (Z-

medRxiv preprint doi: https://doi.org/10.1101/2023.01.24.23284960; this version posted January 30, 2023. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-NC-ND 4.0 International license .

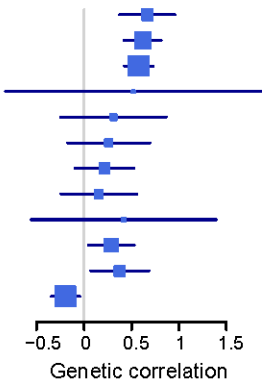
score=11.6, p=3.14×10^{-31}, ΔR^2=0.11%). By incorporating GWAS data from multiple ancestries, the AUD PRS is more significantly associated with AUDIT-P and explains more variance (Z-score=13.6, p=2.44×10^{-42}, ΔR^2=0.15%) than the single ancestry AUD PRS.

Genetic correlations. We confirmed significant positive genetic correlations (r_g) in EUR between PAU and substance use and psychiatric traits[6] (Supplementary Table 15). AD[5] showed the highest correlation with PAU (r_g=0.85, SE=0.07, p=4.49×10^{-34}), followed by maximum habitual alcohol intake[9] (r_g=0.79, SE=0.03, p=1.24×10^{-191}), opioid use disorder[41] (r_g=0.78, SE=0.04, p=1.20×10^{-111}), drinks per week[11] (r_g=0.76, SE=0.02, p<1×10^{-200}), smoking trajectory[42] (r_g=0.63, SE=0.02, p=2.47×10^{-176}), and cannabis use disorder[14] (r_g=0.61, SE=0.04, p=4.85×10^{-63}). We next tested r_g between AUD and 13 published traits with large GWAS in AFR (Figure 3, Supplementary Table 16). As in EUR, the traits with the strongest correlations were substance use traits. Maximum habitual alcohol intake[9] (r_g=0.67, SE=0.15, p=8.13×10^{-6}) showed the highest correlation with AUD, followed by opioid use disorder[41] (r_g=0.62, SE=0.10, p=6.70×10^{-10}), and smoking trajectory[42] (r_g=0.57, SE=0.08, p=3.64×10^{-4}). Major depressive disorder[43] and smoking initiation[11] showed nominally significant (p<0.05) positive correlation with AUD and type 2 diabetes[44] showed a nominally significant negative correlation.

Figure 3. Genetic correlations between AUD and traits in AFR. totalPCL, total index of recent symptom severity by PTSD checklist for DSM-IV. Traits labeled in bold font are genetically correlated with AUD after Bonferroni correction (p<3.83×10^{-3}). Error bar is 95% confidence interval.

medRxiv preprint doi: https://doi.org/10.1101/2023.01.24.23284960; this version posted January 30, 2023. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-NC-ND 4.0 International license .

Traits	Study	N	p
Maximum habitual alcohol intake	Deak, 2022	29,132	**8.13×10^{-6}**
Opioid use disorder	Kember, 2022	88,497	**6.70×10^{-10}**
Smoking trajectory	Xu, 2020	54,867	**3.64×10^{-13}**
Post-traumatic stress disorder	Stein, 2021	51,036	0.45
avoidance		25,414	0.27
totalPCL		25,318	0.24
reexperiencing		25,414	0.18
hyperarousal		25,521	0.45
Cigarettes per day	Saunders, 2022	20,157	0.40
Smoking initiation	Saunders, 2022	119,589	0.02
Major depressive disorder	Levey, 2021	85,083	0.02
Type II diabetes	Vujkovic, 2020	76,750	0.01



PRS for phenome-wide associations. We calculated PRS for PAU (based on the meta-analysis of PAU in EUR) in 131,500 individuals of European ancestry, PRS for AUD (based on the meta-analysis of AUD in AFR) in 27,494 individuals of African ancestry in 4 biobanks (Vanderbilt University Medical Center BioVU, Penn Medicine BioBank, Mount Sinai Icahn School of Medicine Bio*Me* ™, and Mass General Brigham Biobank) from the PsycheMERGE Network, and conducted phenome-wide association studies (PheWAS). After Bonferroni correction, 58 of the 1,493 tested phenotypes were significantly associated with the PAU PRS in EUR, including 26 mental disorders, 8 respiratory traits, 5 neurological conditions, 4 infectious diseases, and 4 neoplasms (Supplementary Table 17, Supplementary Figure 2). For the 793 phenotypes tested in AFR, alcoholism (OR=1.25, SE=0.04, $p=2.62\times10^{-7}$), alcohol-related disorders (OR=1.21, SE=0.04, $p=4.11\times10^{-7}$), and tobacco use disorder (OR=1.09, SE=0.02, $p=6.98\times10^{-6}$) showed significant association with AUD PRS (Supplementary Table 18, Supplementary Figure 3).

We also conducted PheWAS in Yale-Penn, a deeply phenotyped cohort with comprehensive psychiatric assessments (substance use disorders [SUDs] and psychiatric disorders) and assessments for physical and psychosocial traits[26,45]. In EUR, the PRS of PAU

medRxiv preprint doi: https://doi.org/10.1101/2023.01.24.23284960; this version posted January 30, 2023. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-NC-ND 4.0 International license .

was associated with 123 traits, including 26 in alcohol, 39 in opioid, 24 in cocaine and 17 in tobacco categories (Supplementary Table 19, Supplementary Figure 4), indicating high comorbidity and shared genetic components across SUDs. In AFR, the AUD PRS was associated with the DSM-5 AUD criterion count, alcohol-induced blackouts, frequency of alcohol use, and 3 individual AUD criteria: unsuccessful effort to decrease use, used more than intended, and continued use despite social/interpersonal problems (Supplementary Table 20, Supplementary Figure 5).

Discussion

We report here the largest multi-ancestry GWAS for PAU to date, comprising over 1 million individuals and including 165,952 AUD/AD cases, more than double the largest previous study[28]. Considering the results from this study and previous GWAS, in all ancestral populations, we observed a nearly linear relationship between sample size and the number of risk variants discovered.

Convergent evidence supports substantial shared genetic architecture for PAU across multiple ancestries. First, of the 76 independent risk variants detected in EUR and represented in other populations, the majority have the same direction of effect in AFR (84.2%) and LA (81.6%). Twenty-three variants (30.3%) in AFR and 15 (19.7%) in LA were nominally replicated (p<=0.05), which is considerable given the appreciably lower sample size of these ancestral groups. Second, there are high cross-ancestry genetic correlations among EUR, AFR, and LA, ranging from 0.71 (between EUR and AFR) to 0.88 (between AFR and LA). Third, cross-ancestry meta-analysis substantially improved the power for gene discovery and resulted in the identification of 24 additional variants beyond the EUR-only results.

medRxiv preprint doi: https://doi.org/10.1101/2023.01.24.23284960; this version posted January 30, 2023. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-NC-ND 4.0 International license .

A total of 110 variants were associated with PAU in either within-ancestry or cross-ancestry analyses; of these, 9 are missense variants. These include rs1799971 in *OPRM1* which encodes the *μ* opioid receptor, which plays roles in regulating pain, reward, and addictive behaviors. This variant was also associated with opioid use disorder (OUD) in multiple large GWAS[41,46,47]. Previously, there were inconsistent candidate gene association results for *OPRM1**rs1799971 and AUD (reviewed in ref. 48). This is the first GWAS to confirm the association of rs1799971 in PAU; the risk allele is the same as for OUD. In contrast to an apparent EUR-specific effect of rs1799971 on OUD, the *OPRM1* association with PAU (p=6.16×10^{-9}) was detected in the cross-ancestry meta-analysis. Further investigation in larger non-EUR samples is needed to assess the association of this SNP with SUDs in different population groups. Rs6265 in *BDNF* (brain-derived neurotrophic factor) encodes a member of the nerve growth factor family of proteins and has been investigated intensively in the past decades[49]; studies showed that this variant is associated with smoking traits[10] and externalizing behavior[50]. Rs13107325 in *SLC39A8* (Solute Carrier Family 39 Member 8) has been associated with schizophrenia[51], substance uses[6,7,10] and many glycemic traits, and is critical for glycosylation pathways[52,53].

Previous studies have shown that PAU is a brain-related trait with evidence of functional and heritability enrichment in multiple brain regions[6,7]. We performed gene-based association, TWAS in brain tissues, and H-MAGMA analysis in brain annotations. We identified 38 genes that were supported across multiple levels of analysis. For example, *ADH1B* expression in putamen was associated with PAU by TWAS, and with chromatin interaction in all 6 brain annotations by H-MAGMA, indicating additional potential biological mechanisms for the association of *ADH1B* with PAU risk through gene expression and/or chromatin interactions in brain, potentially independent of the well-known hepatic effect on alcohol metabolism. *DRD2*

medRxiv preprint doi: https://doi.org/10.1101/2023.01.24.23284960; this version posted January 30, 2023. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-NC-ND 4.0 International license .

expression in cerebellar hemisphere and chromatin interaction in all brain annotations were also associated with PAU risk. Alcohol metabolism, as is well-reported, has affects that modulate alcohol's aversive and reinforcing effects[54], but also contributes to brain histone acetylation, gene expression and alcohol-related associative learning in mice[55]. The detailed molecular pathways and mechanisms involving changes in human brain need to be elucidated.

Independent genetic signals supported the two main AUD pharmacological treatments acamprosate and naltrexone: *GABRA4* is a target of acamprosate while *OPRM1* is a target for naltrexone. We identified genes known to be druggable; our multivariate analysis also provided evidence for several repurposable drugs. Trichostatin-a, a histone deacetylase inhibitor, showed effects on H3 and H4 acetylation and neuropeptide Y expression in the amygdala and prevented the development of alcohol withdrawal-related anxiety in rats[56]. Clinical trials showed that melperone, a dopamine and serotonin receptor antagonist, has inconsistent effects on alcoholic craving[57,58]. Spironolactone, a mineralocorticoid receptor antagonist, reduced alcohol use in both rats and humans in a recent study[59]. Clomethiazole, a GABA receptor antagonist, also showed effect of treatment for alcohol withdrawal syndrome[60]. Future clinical trials may use the evidence from this drug-repurposing analysis to prioritize drugs for further study.

PAU was positively genetically correlated with many psychiatric and substance use disorders and negatively with cognitive performance. Most of our genetic correlations with PAU, and all those in previous studies, were restricted to EUR populations, presumably because of insufficient statistical power in other populations. The PheWAS PRS also identified associations with medical phenotypes in EUR. With increasing number of AFR GWAS now published, mainly from MVP, we were able to estimate genetic correlations between AUD and a limited set of

medRxiv preprint doi: https://doi.org/10.1101/2023.01.24.23284960; this version posted January 30, 2023. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-NC-ND 4.0 International license .

traits in AFR. As in EUR, AUD in AFR was genetically correlated with substance use traits including OUD, smoking trajectory (which identifies groups of individuals that follow a similar progression of smoking behavior), and maximum habitual alcohol intake. PheWAS of PRS in AFR from PsycheMERGE and Yale-Penn confirmed that AUD is genetically correlated with substance use traits. The lack of a wider set of phenotypes for comparison by ancestry is a continuing limitation.

Additional limitations include that the differences in ascertainment and phenotypic heterogeneity across cohorts might bias the results. Despite the high genetic correlation between AUD and AUDIT-P, they are not identical traits. Also, differences in ascertainment amongst the cohorts may have introduced additional biases; for example, considering the QIMR AGDS and GBP cohorts, the former have high major depression comorbidity, and the latter have high bipolar disorder comorbidity. (This heterogeneity would, however, have been more likely to limit discovery than to create false-positives.) Additionally, while we set out to include all available samples for problematic drinking in multiple ancestries, the sample sizes in the non-EUR ancestries were still small for gene discoveries and downstream analyses. The collection of substantial numbers of non-European subjects is a critical next step in this field.

In summary, we report here a large multi-ancestry GWAS and meta-analysis for PAU, in which we focused our analyses in three main directions. First, we demonstrated that there is substantial shared genetic architecture of PAU across multiple populations. Second, we analyzed gene prioritization for PAU using multiple approaches, including cross-ancestry fine-mapping, gene-based association, brain-tissue TWAS and fine-mapping, and H-MAGMA for chromatin interaction. We identified many genes associated with PAU with biological support,

medRxiv preprint doi: https://doi.org/10.1101/2023.01.24.23284960; this version posted January 30, 2023. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-NC-ND 4.0 International license .

extending our understanding of the brain biology that substantially modifies PAU risk and expands opportunities for investigation using *in vitro* methods and animal models. These genes are potential actionable targets for downstream functional studies and possible targets of pharmacological intervention based on the drug repurposing results. Third, we investigated the genetic relationship between PAU and many traits, which for the first time was possible in AFR populations. Future increases in sample size will doubtless yield additional gains; this is particularly needed in non-EUR populations both for primary GWAS analyses and the analysis of other traits for comparison and to estimate pleiotropy.

Methods

Study design. In the previous PAU study[6], the r_g between MVP AUD and PGC alcohol dependence (AD) was 0.98, which justified the meta-analysis of AUD (includes AUD and AD) across the two datasets; and the r_g between AUD and UKB AUDIT-P was 0.71, which justified the proxy-phenotype meta-analysis of PAU (including AUD, AD and AUDIT-P) across all datasets. In this study, we use the same definitions, defining AUD by meta-analyzing AUD and AD across all datasets, and defining PAU by meta-analyzing AUD, AD and AUDIT-P (Table 1).

MVP dataset. MVP enrollment and genotyping have been described previously[19,61]. MVP is a biobank supported by the US Department of Veterans Affairs (VA) with rich phenotypic data collected using questionnaires and the VA electronic health record system (EHR). The Central VA Institutional Review Board (IRB) and site-specific IRBs approved the MVP study. All relevant ethical regulations for work with human subjects were followed in the conduct of the study and informed consent was obtained from all participants.

medRxiv preprint doi: https://doi.org/10.1101/2023.01.24.23284960; this version posted January 30, 2023. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-NC-ND 4.0 International license .

MVP genotype data were processed by the MVP Release 4 (R4) Data Team. 729,324 samples were genotyped using Affymetrix Axiom Biobank Array. Rigorous sample-level quality control (QC) served to remove samples with duplicates, call rates <98.5%, sex mismatches, >7 relatives, or excess heterozygosity. After QC, MVP R4 data contains 658,582 participants and 667,995 variants (pre-imputation). Pre-imputation QC removed variants with high missingness (>1.5%), that were monomorphic, or with Hardy-Weinberg Equilibrium (HWE) p-value $\leq 1\times10^{-6}$, leaving 590,511 variants for imputation. As in our previous work[7], we ran principal component analysis (PCA)[62,63] for the R4 data and 1000 Genome phase3 reference panels[64]. The Euclidean distances between each MVP participant and the centers of the five reference ancestral groups were calculated using the first 10 PCs, with each participant assigned to the nearest reference ancestry. A second round PCA within each assigned ancestral group was performed and outliers with PC scores >6 standard deviations from the mean of any of the 10 PCs were removed. This two-stage approach resulted in the assignment of 468,869 European ancestry (EUR), 122,024 African ancestry (AFR), 41,662 Latin American (LA), 7,364 East Asian (EAS) and 536 South Asian (SAS) individuals for analysis.

Imputation was done by the MVP R4 Data Team. The entire cohort was pre-phased using SHAPEIT4 (v4.1.3)[65], then imputed using Minimac4[66] with African Genome Resources reference panel by Sanger Institute and 1000 Genomes Project phase3 as reference. Single nucleotide variants with imputation score <0.8, or HWE p-value $\leq 1\times10^{-6}$, or minor allele frequency (MAF) lower than the threshold set in each ancestral group based upon their sample size (EA, 0.0005; AA, 0.001; LA, 0.005; EAA, 0.01; SAA, 0.01) were removed before association analysis.

medRxiv preprint doi: https://doi.org/10.1101/2023.01.24.23284960; this version posted January 30, 2023. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-NC-ND 4.0 International license .

Participants with at least one inpatient or two outpatient International Classification of Diseases (ICD)-9/10 codes for AUD were assigned as AUD cases, while participants with zero ICD codes for AUD were controls. Those with one outpatient diagnosis were excluded from the analysis. In total, 80,028; 36,330; 10,150; 701; and 107 cases were included in EUR, AFR, LA, EAS, and SAS, respectively; 368,113; 79,100; 28,812; 6,254; and 389 controls were included in EUR, AFR, LA, EAS, and SAS, respectively. BOLT-LMM[67] was used to correct for relatedness, with age, sex, and the first 10 PCs as covariates.

UK Biobank (UKB). UKB released genotype and imputed data for ~500,000 individuals from across the United Kingdom[17] which were accessed through application 41910. UKB defined White-British (WB) participants genetically. For the non-WB individuals, we used PCA to classify them into different genetic groups as for MVP. Subjects with available AUDIT-P score were included in this study. The final sample included 132,001 WB (hereafter called UKB-EUR1) and 17,898 non-WB Europeans (hereafter called UKB-EUR2), and 1,220 SAS. SNPs with genotype call rate >0.95, HWE p-value $>1\times10^{-6}$, imputation score ≥0.8 and MAF ≥0.001 in EUR1 and EUR2 and ≥0.01 in SAS were kept for GWAS, BOLT-LMM was used for association correcting for relatedness, age, sex, and the first 10 PCs.

FinnGen. Summary statistics for AUD from FinnGen data freeze 5 were downloaded from the FinnGen website (http://r5.finngen.fi/). Details of the genotyping, imputation and quality control for FinnGen data were described previously[20]. There were 8,866 AUD cases defined by ICD-8/9/10 codes and 209,926 controls. Association analysis was performed using SAIGE[68] mixed-model with age, sex and 10 PCs as covariates. Positions of the variants were lifted over to build 37 (GRCh37/hg19) for meta-analysis.

medRxiv preprint doi: https://doi.org/10.1101/2023.01.24.23284960; this version posted January 30, 2023. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-NC-ND 4.0 International license .

iPSYCH: The iPSYCH[21,22] samples were selected from a baseline birth cohort comprising all singletons born in Denmark between May 1, 1981, and December 31, 2008. The iPSYCH study was approved by the Scientific Ethics Committee in the Central Denmark Region (Case No 1-10-72-287-12) and the Danish Data Protection Agency.

AUD was diagnosed according to the ICD-10 criteria (F10.1 – F10.9 diagnosis codes). The iPSYCH cohort was established to investigate genetic risk for major psychiatric disorders (i.e., attention-deficit/hyperactivity disorder, schizophrenia, bipolar disorder, major depressive disorder, autism spectrum disorder) but not AUD (or PAU), so comorbidity of psychiatric disorders among these AUD cases is higher than expected for cases selected randomly from the population. Therefore, we generated a control group around five times as large as the case groups, and to correct for the bias introduced by high comorbidity of psychiatric disorders among cases, we included within the control group individuals with the above listed psychiatric disorders (without comorbid AUD) at a proportion equal to what was observed among the cases.

The samples were genotyped in two genotyping rounds referred to as iPSYCH1 and iPSYCH2. iPSYCH1 samples were genotyped using Illumina's PsychChip array and iPSYCH2 samples using Illumina´s GSA v.2 (Illumina, San Diego, CA, USA). Quality control and GWAS were performed using the Ricopili pipeline[69]. More details can be found in ref. 70. GWAS were performed separately for iPSYCH1 (2,117 cases and 13,238 controls) and iPSYCH2 (1,024 cases and 5,732 controls) using dosages for imputed genotypes and additive logistic regression with the first 5 PCs (from the final PCAs) as covariates using PLINK v1.9[71]. Only variants with a MAF >0.01 and imputation score >0.8 were included in the final summary statistics.

medRxiv preprint doi: https://doi.org/10.1101/2023.01.24.23284960; this version posted January 30, 2023. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-NC-ND 4.0 International license .

Queensland Berghofer Medical Research Institute (QIMR) cohorts. The Australian Genetics of Depression Study (AGDS) recruited >20,000 participants with major depression between 2017 and 2020. Recruitment and subject characteristics have been reported[24]: Participants completed an online self-report questionnaire. Lifetime AUD was assessed on DSM-5 criteria using the Composite International Diagnostic Interview (CIDI). A total of 6,726 subjects with and 4,467 without AUD were included in the present study.

The Australian twin-family study of alcohol use disorder (TWINS, including Australian Alcohol and Nicotine Studies) participants were recruited from adult twins and their relatives who had participated in questionnaire- and interview-based studies on alcohol and nicotine use and alcohol-related events or symptoms (as described in Heath et al.[72]). They were predominantly of EUR ancestry. Young adult twins and their non-twin siblings were participants in the Nineteen and Up study (19Up)[25]. 2,772 cases and 5,630 controls were defined using DSM-III-R and DSM-IV criteria. Most alcohol-dependent cases were mild, with 70% of those meeting alcohol dependence criteria reporting only three or four dependence symptoms and ≤5% reporting seven dependence symptoms.

The Australian Genetics of Bipolar Disorder Study (GBP) recruited >5,000 participants living with bipolar disorder between 2018 and 2021. The sample's recruitment and characteristics have been reported[23]: Participants completed an online self-report questionnaire. Lifetime DSM-5 AUD was assessed using the CIDI.

medRxiv preprint doi: https://doi.org/10.1101/2023.01.24.23284960; this version posted January 30, 2023. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-NC-ND 4.0 International license .

Genotyping of QIMR cohorts was performed using Illumina Global Screening Array v2. Pre-imputation QC removed variants with GenTrain score <0.6, MAF <0.01, SNP call rate <95%, and Hardy-Weinberg equilibrium deviation ($p<1x10^{-6}$). Variants were then imputed using the Michigan Imputation Server with the Haplotype Reference Consortium reference panel[66]. Association analysis was performed using SAIGE and the LOCO=TRUE flag with age, sex, 10 PCs and two imputation variables as covariates. Participants of non-EUR ancestry (defined as >6 standard deviations from the PC1 and PC2 centroids) were excluded. Association analyses were limited to variants with a MAF≥0.0001, MAC≥5, and an R^2≥0.1.

Psychiatric Genomics Consortium (PGC). Lifetime DSM-IV diagnosis of AD in both EUR and AFR ancestries were analyzed by PGC, with details reported previously[5]. This included 5,638 individuals from Australia. To avoid overlap with the new QIMR cohorts, we re-analyzed the PGC data without two Australian cohorts: Australian Alcohol and Nicotine Studies and Brisbane Longitudinal Twin Study. This yielded 9,938 cases and 30,992 controls of EUR ancestry and 3,335 cases and 2,945 controls of AFR ancestry.

Yale-Penn 3. There are 3 phases of the Yale-Penn study defined by genotyping epoch; the first two were incorporated in the PGC study, thus they are included in the meta-analyses. Here, we included Yale-Penn 3 subjects as a separate sample. Lifetime AD was diagnosed based on DSM-IV criteria. Genotyping was performed in the Gelernter laboratory at Yale using the Illumina Multi-Ethnic Global Array, then imputed using Michigan Imputation Server with Haplotype Reference Consortium reference. We did PCA analyses to classify EAs (567 cases and 1,074 controls) and AAs (451 cases and 410 controls). Variants with MAF >0.01, HWE p-value $>1\times10^{-6}$ and imputation INFO score ≥0.8 were retained for association analyses using

medRxiv preprint doi: https://doi.org/10.1101/2023.01.24.23284960; this version posted January 30, 2023. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-NC-ND 4.0 International license .

linear mixed models implemented in GEMMA[73] and corrected for age, sex and 10 PCs.

East Asian cohorts. Summary statistics for AUD/AD GWAS from 5 EAS cohorts (MVP EAS, Han Chinese–GSA, Thai METH–MEGA, Thai METH–GSA and Han Chinese–Cyto) were included in the cross-ancestry meta-analysis. Analyses of these five cohorts were previously published and the detailed QC can be found in ref. 27.

Meta-analyses. Meta-analyses were performed using METAL[74] with effective sample size weighting. For all the case-control samples, we calculated effective sample size as:

$$n_{effective} = \frac{4}{\frac{1}{n_{case}} + \frac{1}{n_{control}}}$$

For AUDIT-P in UKB, a continuous trait, we used actual sample sizes for meta-analysis. For all meta-analyses within or across ancestries, variants with a heterogeneity test p-value $<5\times10^{-8}$ and variants with effective sample size <15% of the total effective sample size were removed. For the cross-ancestry and EUR within-ancestry meta-analyses, we required that variants were present in at least two cohorts. For the AFR and SAS within-ancestry meta-analyses, which are small samples, this was not required.

Independent variants and conditional analyses. We identified the lead variants using PLINK with parameters of clumping region 500 kb and LD r^2 0.1. We then ran conditional analyses using GCTA-COJO[75] to define conditionally independent variants among the lead variants using the 1000 Genomes Project phase3 as the LD reference panel. Any two independent variants <1 Mb apart whose clumped regions overlapped were merged into one locus.

medRxiv preprint doi: https://doi.org/10.1101/2023.01.24.23284960; this version posted January 30, 2023. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-NC-ND 4.0 International license .

Cross-ancestry lookup. For the 85 independent variants associated in EUR, we looked up the associations in non-EUR groups. If the variants were not observed in another ancestry, we substituted proxy SNPs defined as associated with PAU ($p<5\times10^{-8}$) and in high LD with the EUR lead SNP ($r^2\geq0.8$).

SNP-based heritability (h^2). SNP-based h^2 for common SNPs mapped to HapMap3[76] was estimated in EUR, AFR and LA ancestries using LDSC[77]; corresponding populations in the 1000 Genomes Project phase3 were used as LD reference panels. For PAU in EUR, we only estimated the observed-scale h^2. For AUD, both observed-scale h^2 and liability-scale h^2 were estimated, using population lifetime prevalence estimates of 0.326, 0.220, and 0.229 in EUR, AFR, and LA, respectively[1]. These prevalence estimates were for lifetime DSM-5 AUD in the United States, which could introduce bias given the different definitions and prevalence in different cohorts. By default, LDSC removes SNPs with sample size <90th percentile N/2. Here we skipped this filtering and kept all SNPs for analyses because we did basic filtering based on the number of cohorts and sample size. The final number of SNPs in the analyses range from 527,994 to 1.17 M.

Cross-ancestry genetic correlation. We estimated the genetic correlations between different ancestries using Popcorn[78], which can estimate both the genetic-effect correlation (ρ_{ge}) as correlation coefficient of the per-allele SNP effect sizes and the genetic-impact correlation (ρ_{gi}) as correlation coefficient of the ancestry-specific allele-variance-normalized SNP effect sizes. Populations in 1000 Genomes were used as reference for their corresponding population. A large sample size and number of SNPs are required for accurate estimation[78], which explains

medRxiv preprint doi: https://doi.org/10.1101/2023.01.24.23284960; this version posted January 30, 2023. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-NC-ND 4.0 International license .

the non-robust estimates for EAS and SAS samples.

Within- and cross-ancestry fine-mapping. We did fine-mapping using MsCAVIAR[79], which can leverage LD information from multiple ancestries to improve fine-mapping of causal variants. To reduce bias introduced by populations with small sample size, here we performed fine-mapping using summary statistics from the EUR, AFR and LA populations. Three sets of analyses were conducted. The first is within-ancestry fine-mapping for the 85 regions with independent variants in EUR using EUR summary data and 1000 Genomes Project phase3 EUR LD reference data. For each region, we selected SNPs which clumped (within 500 kb and LD r^2>0.1) with the lead SNP and with p<0.05 for fine-mapping. We then calculated the pair-wise LD among the selected SNPs. If two SNPs were in perfect LD (r^2=1, indicating that they are likely to be inherited together), we randomly removed one from the analysis. The second is cross-ancestry fine-mapping for the 100 regions with independent variants identified in cross-ancestry meta-analyses. For each region, we performed clumping (within 500 kb and LD r^2>0.1) in EUR, AFR, and LA summary data for the lead SNP separately, to select 3 sets of SNPs (p<0.05) for fine-mapping, corresponding LD reference panels from 1000 Genomes Project were used. For each set of SNPs, we calculated the pair-wise LD and randomly removed one SNP if r^2=1. If the lead SNP was not presented in the EUR SNP set, we did not perform fine-mapping for this region. Loci with limited numbers of variants cannot have convergent results, so they are not included in the results. After that, this cross-ancestry analysis included 92 regions. For the 10 regions in which the lead SNPs are missing in both AFR and LA populations, we did within-ancestry fine-mapping in EUR instead to keep the lead SNP (cross-ancestry fine-mapping will only analyze the SNPs common in analyzed ancestries). Third, because the credible set length identified is related to the number of variants in the input, to provide a more direct comparison between the cross-ancestry fine-mapping and the fine-mapping using

medRxiv preprint doi: https://doi.org/10.1101/2023.01.24.23284960; this version posted January 30, 2023. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-NC-ND 4.0 International license .

information only from EUR, we used the same lists of SNPs from the above 92 regions in the cross-ancestry fine-mapping as for the EUR-only fine-mapping. "Credible set" was defined as plausible causal variants with accumulated posterior inclusion probability (PIP) > 99%. For each credible set, we report the variant with the highest PIP. We assumed that each locus contains only one causal variant by default, and increased to three at maximum if the analysis unable to converge.

Gene-based association analyses. We performed gene-based association analysis for PAU or AUD in multiple ancestries using MAGMA implemented in FUMA[32,33]. Default settings were applied. Bonferroni corrections for the number of genes tested (range from 18,390 to 19,002 in different ancestries) were used to determine genome-wide significant genes.

Transcriptome-wide association study (TWAS). For PAU in EUR, we performed TWAS using S-PrediXcan to integrate transcriptomic data from GTEx. With prior knowledge that PAU is a brain-related disorder (evidenced by significant enrichment of gene expression in several brain tissues)[6], 13 brain tissues were analyzed. The transcriptome prediction model database and the covariance matrices of the SNPs within each gene model were downloaded from the PredictDB repository (http://predictdb.org/). Significance of the gene-tissue association was determined following Bonferroni correction for the total number of gene-tissue pairs (p<0.05/166,064=3.01×10^{-7}). We also used S-MultiXcan[36] to integrate evidence across the 13 brain tissues using multivariate regression to improve association detection. In total, 18,383 genes were tested in S-MultiXcan, leading to a significance p-value threshold of 2.72×10^{-6}.

Association with chromatin interactions in brain. We used H-MAGMA[37], a computational

medRxiv preprint doi: https://doi.org/10.1101/2023.01.24.23284960; this version posted January 30, 2023. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-NC-ND 4.0 International license .

tool that incorporates brain chromatin interaction profiles from Hi-C, to identify risk genes associated with PAU based on EUR inputs. Six brain annotations were used: fetal brain, adult brain, adult midbrain dopaminergic, iPSC-derived astrocyte, iPSC-derived neuron and cortical neuron. In total, 319,903 gene-chromatin associations were analyzed across the six brain annotations. Significant genes were those with a p-value below the Bonferroni corrected value for the total number of tests (p<0.05/319,903=1.56×10^{-7}).

Probabilistic fine-mapping of TWAS. We did fine-mapping for TWAS in EUR using FOCUS[38], a method that models correlation among TWAS signals to assign a PIP for every gene in the risk region to explain the observed association signal. The estimated credible set containing the causal gene can be prioritized for functional assays. FOCUS used 1000 Genomes Project EUR samples as the LD reference and multiple eQTL reference panel weights that include GTEx_v7[80], The Metabolic Syndrome in Men[81], Netherlands Twin Register[82], Young Finns Study[83], and CommonMind Consortium[84]. Under the model of PAU as substantially a brain disorder, we did fine-mapping while prioritizing predictive models using a brain tissue-prioritized approach.

Drug repurposing. To match inferred transcriptional patterns of PAU with transcriptional patterns induced by perturbagens, we related our S-PrediXcan results to signatures from the Library of Integrated Network-based Cellular Signatures (LINCs) L1000 database[85]. This database catalogues *in vitro* gene expression profiles (signatures) from thousands of compounds >80 human cell lines (level 5 data from phase I: GSE92742 and phase II: GSE70138). Our analyses included signatures of 829 chemical compounds in five neuronal cell-lines (NEU, NPC, MNEU.E, NPC.CAS9 and NPC.TAK). To test significance of the association

medRxiv preprint doi: https://doi.org/10.1101/2023.01.24.23284960; this version posted January 30, 2023. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-NC-ND 4.0 International license .

between PAU signatures and LINCs perturbagen signatures we followed the procedure from So et al[86]. Briefly, we computed weighted (by proportion of heritability explained) Pearson correlations between transcriptome-wide brain associations and *in vitro* L1000 compound signatures using the *metafor* package[87] in R. We treated each L1000 compound as a fixed effect incorporating the effect size ($r_{weighted}$) and sampling variability (se^2) from all signatures of a compound (e.g., across all time points and doses). We only report those perturbagens that were associated after Bonferroni correction ($p<0.05/829=6.03\times10^{-5}$).

Cross-ancestry polygenic risk score. We used PRS-CSx[40], a method that couples genetic effects and LD across ancestries via a shared continuous shrinkage prior, to calculate the posterior effect sizes for SNPs mapped to HapMap3. Three sets of AUD GWAS summary data were use as input and corresponding posterior effect sizes in each ancestry were generated: EUR (without AUDIT-P from UKB, $N_{effective}$=352,373), AFR ($N_{effective}$=105,433), and LA ($N_{effective}$=30,023). Three sets of AUD PRS based on the posterior effect sizes were calculated for UKB-EUR1 and UKB-EUR2 individuals using PLINK, following standardization (zero mean and unit variance) for each PRS. For each related pair (≥3rd-degree, kinship coefficient ≥0.0442 as calculated by UKB), we removed the subject with the lower AUDIT-P score, or randomly if they had the same score, leaving 123,565 individuals in UKB-EUR1 and 17,401 in UKB-EUR2. Then we ran linear regression for AUDIT-P in UKB-EUR2 as a validation dataset using PRS_{EUR}, PRS_{AFR} and PRS_{LA} as independent variables. The corresponding regression coefficients were used as weights in the test dataset (UKB-EUR1) to calculate the final PRS: $PRS_{final} = \omega_{EUR}*PRS_{EUR} + \omega_{AFR}*PRS_{AFR} + \omega_{LA}*PRS_{LA}$. We used linear regression to test the association between AUDIT-P and PRS_{final} after standardization, correcting for age, sex, and the first 10 PCs. We also ran a null model of association between AUDIT-P and covariates only, to calculate the variance explained (R^2) by PRS_{final}. For comparison, we also calculated PRS in

medRxiv preprint doi: https://doi.org/10.1101/2023.01.24.23284960; this version posted January 30, 2023. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-NC-ND 4.0 International license .

UKB-EUR1 using only the AUD summary data in EUR using PRS-CS[88], then calculated the variance explained by PRS_{single}. The improved PRS association was measured as the difference of the variance explained (ΔR^2).

Genetic correlation. Genetic correlations (r_g) between PAU or AUD and traits of interest were estimated using LDSC[89]. For EUR, we tested r_g between PAU and 49 traits using published summary data and the EUR LD reference from the 1000 Genomes Project. r_gs with p-value $<1.02\times10^{-3}$ were considered significant. For AFR, we tested r_g between AUD and 13 published traits in AFR using MVP in-sample LD (most of the analyzed AFR were from MVP) built from 1000 randomly-selected AFR subjects by cov-LDSC[90]. r_gs with p-value $<3.85\times10^{-3}$ (0.05/13) in AFR were considered as significant. For comparison, we also tested r_gs using 1000 Genomes AFR as LD reference, which showed similar estimates.

PAU PRS for phenome-wide associations.

We calculated PRS using PRS-continuous shrinkage (PRS-CS) for PAU (in EUR) and AUD (in AFR) in four independent datasets [Vanderbilt University Medical Center's Biobank (BioVU), Mount Sinai (Bio*Me* ™), Mass General Brigham Biobank (MGBB)[91] and Penn Medicine Biobank (PMBB)[92]] from the PsycheMERGE Network, followed by phenome-wide association studies. Details for each dataset are described below.

BioVU: Genotyping of individuals was performed using the Illumina MEGEX array. Genotypes were filtered for SNP and individual call rates, sex discrepancies, and excessive heterozygosity using PLINK. Imputation was conducted using the Michigan Imputation Server based on the

medRxiv preprint doi: https://doi.org/10.1101/2023.01.24.23284960; this version posted January 30, 2023. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-NC-ND 4.0 International license .

Haplotype Reference Consortium reference panel. PCA using FlashPCA2[93] combined with CEU, YRI and CHB reference sets from the 1000 Genomes Project Phase 3 was conducted to determine participants of AFR and EUR ancestry. One individual from each pair of related individuals was removed (pi-hat>0.2). This resulted in 12,384 AFR and 66,903 EUR individuals for analysis.

BioMe™: The Bio*Me™* Biobank: The Illumina Global Screening Array was used to genotype the Bio*Me™* samples. The SNP-level quality control (QC) removed SNPs with (1) MAF <0.0001 (2) HWE p-value $\leq 1\times10^{-6}$ and (3) call rate <98%. The individual-level QC removed participants with (1) sample call rate <98% and (2) heterozygosity F coefficient ≥3 standard deviations. In addition, one individual from each pair of related samples with a genomic relatedness (proportion IBD) >0.125 was removed (--rel-cutoff=0.125 in PLINK). Imputation was performed using 1000 Genomes Phase 3 data. Each ancestry was confirmed by the genetic PC plot. A final sample size of 4,727 AFR and 9,544 EUR individuals were included for this study.

MGBB: Individuals in the Mass General Brigham Biobank (MGBB) were genotyped using the Illumina Multi-Ethnic Global array with hg19 coordinates. Variant-level quality control filters removed variants with a call rate <98% and those that were duplicated across batches, monomorphic, not confidently mapped to a genomic location, or associated with genotyping batch. Sample-level quality control filters removed individuals with a call rate less than 98%, excessive autosomal heterozygosity (±3 standard deviations from the mean), or discrepant self-reported and genetically inferred sex. PCs of ancestry were calculated in the 1000 Genomes Phase 3 reference panel and subsequently projected onto the MGBB dataset, where a Random Forest classifier was used to assign ancestral group membership for individuals with a

medRxiv preprint doi: https://doi.org/10.1101/2023.01.24.23284960; this version posted January 30, 2023. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-NC-ND 4.0 International license .

prediction probability >90%. The Michigan Imputation Server was then used to impute missing genotypes with the Haplotype Reference Consortium dataset serving as the reference panel. Imputed genotype dosages were converted to hard-call format and subjected to further quality control, where SNPs were removed if they exhibited poor imputation quality (INFO<0.8), low minor allele frequency (<1%), deviations from Hardy-Weinberg equilibrium ($p<1\times10^{-10}$), or missingness (variant call rate <98%). Only unrelated individuals (pi-hat<0.2) of EUR ancestry were included in the present study. These procedures yielded a final analytic sample of 25,698 individuals in the MGBB.

PMBB: Genotyping of individuals was performed using the Illumina Global Screening Array. Quality control removed SNPs with marker call rate <95% and sample call rate <90%, and individuals with sex discrepancies. Imputation was performed using Eagle2[94] and Minimac4 on the TOPMed Imputation Server. One individual from each pair of related individuals (pi-hat threshold of 0.25) were removed from analysis. PCA was conducted using smartpca[62,63] and the Hapmap3 dataset to determine genetic ancestry. This resulted in 10,383 AFR and 29,355 EUR individuals for analysis.

PheWAS: The AFR AUD PRS and EUR PAU PRS scores in each dataset were standardized for the PheWAS analyses. International Classification of Diseases (ICD)-9 and -10 codes were extracted from the electronic health record and mapped to phecodes. Individuals were considered cases if they had two instances of the phecode. We conducted PheWAS by fitting a logistic regression for each phecode within each biobank. Covariates included sex, age and the top 10 PCs. PheWAS results were meta-analyzed within each ancestral group across biobanks (AFR=27,494, EUR=131,500) using the PheWAS package[95] in R.

medRxiv preprint doi: https://doi.org/10.1101/2023.01.24.23284960; this version posted January 30, 2023. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-NC-ND 4.0 International license .

Yale-Penn: Quality control and creation of the PheWAS dataset have been described previously[45]. We calculated PRS for PAU in EUR and AUD in AFR (using summary statistics that leave out the Yale-Penn 3 and PGC sample which includes Yale-Penn 1 and 2). We conducted PheWAS by fitting logistic regression models for binary traits and linear regression models for continuous traits. We used sex, age at recruitment, and the top 10 genetic PCs as covariates. We applied a Bonferroni correction to control for multiple comparisons.

Data Availability: The full summary-level association data from the meta-analysis are available upon request to the corresponding authors and through dbGaP (accession number phs001672).

Acknowledgements

This research used data from the Million Veteran Program and was supported by funding from the Department of Veterans Affairs Office of Research and Development, Million Veteran Program Grant #I01CX001849, #I01BX004820, #I01BX003341, and the VA Cooperative Studies Program (CSP) study #575B, MVP004, and MVP025. This publication does not represent the views of the Department of Veterans Affairs or the United States Government. Supported also by NIH (NIAAA) P50 AA12870 (J.H.K.) and K01 AA028292 (R.L.K.), a NARSAD Young Investigator Grant 27835 from the Brain & Behavior Research Foundation (H.Z.), NCI R21 CA252916 (H.Z.), R01 AA026364 (J.G.), NIAAA T32 AA028259 (J.D.D.), NIMH R01 MH124839 (L.M.H.), NIAAA K01 AA030083 (A.S.H.), NIDA K01 DA051759 (E.C.J.), TRDRP (T29KT0526, T32IR5226, S.S.R.) and NIDA DP1 DA054394 (S.S.R.). This research used data from UK Biobank (project ID: 41910), a population-based sample of participants whose contributions we gratefully acknowledge. The data access is supported by Yale SCORE pilot

medRxiv preprint doi: https://doi.org/10.1101/2023.01.24.23284960; this version posted January 30, 2023. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-NC-ND 4.0 International license .

grant (U54AA027989). We want to acknowledge the participants and investigators of the FinnGen study. D.D. was supported by the Novo Nordisk Foundation (NNF20OC0065561), the Lundbeck Foundation (R344-2020-1060). The iPSYCH team was supported by grants from the Lundbeck Foundation (R102-A9118, R155-2014-1724, and R248-2017-2003), NIH/NIMH (1U01MH109514-01 and 1R01MH124851-01 to A.D.B.) and the Universities and University Hospitals of Aarhus and Copenhagen. The Danish National Biobank resource was supported by the Novo Nordisk Foundation. High-performance computer capacity for handling and statistical analysis of iPSYCH data on the GenomeDK HPC facility was provided by the Center for Genomics and Personalized Medicine and the Centre for Integrative Sequencing, iSEQ, Aarhus University, Denmark (grant to A.D.B.). The Australian Genetics of Depression Study (AGDS) was primarily funded by the National Health and Medical Research Council (NHMRC) of Australia Grant No. 1086683 to N.G.M. N.G.M is supported by a NHMRC Investigator Grant (No. APP 1172990). We are indebted to all of the participants for giving their time to contribute to this study. We wish to thank all the people who helped in the conception, implementation, media campaign and data cleaning. We thank Richard Parker, Simone Cross, and Lenore Sullivan for their valuable work coordinating all the administrative and operational aspects of the AGDS project. We would also like to thank the research participants for making this work possible. The Australian Genetics of Bipolar Disorder Study (GBP) data collection was funded and data analysis was supported by the Australian NHMRC (No. APP1138514) to S.E.M. S.E.M. is supported by a NHMRC Investigator Grant (No. APP1172917). We thank the participants for giving their time and support for this project. We acknowledge and thank M. Steffens for her generous donations and fundraising support. The NHMRC (APP10499110) and the NIH (K99R00, R00DA023549) funded the 19Up study. Genotyping was funded by the NHMRC (389891). We thank the twins and their families for their willing participation in our studies. Funding for the Australian adult twin studies in which information on alcohol use and smoking status was obtained came from the US NIH (AA07535, AA07728, AA11998, AA13320,

medRxiv preprint doi: https://doi.org/10.1101/2023.01.24.23284960; this version posted January 30, 2023. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-NC-ND 4.0 International license .

AA13321, AA14041, AA17688, DA012854 and DA019951); the Australian NHMRC (241944, 339462, 389927, 389875, 389891, 389892, 389938, 442915, 442981, 496739, 552485 and 552498); and the Australian Research Council (A7960034, A79906588, A79801419, DP0770096, DP0212016 and DP0343921). We acknowledge the work over many years of staff of the Genetic Epidemiology group at QIMR Berghofer Medical Research Institute (formerly the Queensland Institute of Medical Research) in managing the studies which generated the data used in this analysis. We also acknowledge and appreciate the willingness of study participants to complete multiple, and sometimes lengthy, questionnaires and interviews. Many of the participants were contacted originally through the Australian Twin Registry. This research also used summary data from the Psychiatric Genomics Consortium (PGC) Substance Use Disorders (SUD) working group. The PGC-SUD is supported by NIH grant R01DA054869. PGC-SUD gratefully acknowledges its contributing studies and the participants in those studies, without whom this effort would not be possible. We acknowledge the Penn Medicine BioBank (PMBB) for providing data and thank the patient-participants of Penn Medicine who consented to participate in this research program. We would also like to thank the PMBB team and Regeneron Genetics Center for providing genetic variant data for analysis. The PMBB is approved under IRB protocol# 813913 and supported by Perelman School of Medicine at University of Pennsylvania, a gift from the Smilow family, and the National Center for Advancing Translational Sciences of the NIH under CTSA award number UL1TR001878. This study was supported in part through the resources and staff expertise provided by the Charles Bronfman Institute for Personalized Medicine and The Bio*Me*™ Biobank Program at the Icahn School of Medicine at Mount Sinai. Research reported in this paper was supported by the Office of Research Infrastructure of the NIH under award numbers S10OD018522 and S10OD026880. The content is solely the responsibility of the authors and does not necessarily represent the official views of the NIH.

medRxiv preprint doi: https://doi.org/10.1101/2023.01.24.23284960; this version posted January 30, 2023. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-NC-ND 4.0 International license .

Disclosure: H.R.K. is a member of advisory boards for Dicerna Pharmaceuticals, Sophrosyne Pharmaceuticals, and Enthion Pharmaceuticals; a consultant to Sobrera Pharmaceuticals; the recipient of research funding and medication supplies for an investigator-initiated study from Alkermes; and a member of the American Society of Clinical Psychopharmacology's Alcohol Clinical Trials Initiative, which was supported in the past 3 years by Alkermes, Dicerna, Ethypharm, Lundbeck, Mitsubishi, Otsuka, and Pear Therapeutics. M.B.S. has in the past 3 years been a consultant for Actelion, Acadia Pharmaceuticals, Aptinyx, Bionomics, BioXcel Therapeutics, Clexio, EmpowerPharm, Epivario, GW Pharmaceuticals, Janssen, Jazz Pharmaceuticals, Roche/Genentech and Oxeia Biopharmaceuticals. M.B.S. has stock options in Oxeia Biopharmaceuticals and Epivario. He also receives payment from the following entities for editorial work: *Biological Psychiatry* (published by Elsevier), *Depression and Anxiety* (published by Wiley) and *UpToDate*. J.G. and H.R.K. hold US patent 10,900,082 titled: "Genotype-guided dosing of opioid agonists," issued January 26, 2021. J.G. is paid for his editorial work on the journal *Complex Psychiatry*. J.H.K. has consulting agreements (less than US$10,000 per year) with the following: AstraZeneca Pharmaceuticals, Biogen, Idec, MA, Biomedisyn Corporation, Bionomics, Limited (Australia), Boehringer Ingelheim International, COMPASS Pathways, Limited, United Kingdom, Concert Pharmaceuticals, Inc., Epiodyne, Inc., EpiVario, Inc., Heptares Therapeutics, Limited (UK), Janssen Research & Development, Otsuka America, Pharmaceutical, Inc., Perception Neuroscience Holdings, Inc., Spring Care, Inc., Sunovion Pharmaceuticals, Inc., Takeda Industries and Taisho Pharmaceutical Co., Ltd. J.H.K. serves on the scientific advisory boards of Bioasis Technologies, Inc., Biohaven Pharmaceuticals, BioXcel Therapeutics, Inc. (Clinical Advisory Board), BlackThorn Therapeutics, Inc., Cadent Therapeutics (Clinical Advisory Board), Cerevel Therapeutics, LLC., EpiVario, Inc., Lohocla Research Corporation, PsychoGenics, Inc.; is on the board of directors of Inheris Biopharma,

medRxiv preprint doi: https://doi.org/10.1101/2023.01.24.23284960; this version posted January 30, 2023. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-NC-ND 4.0 International license .

Inc.; has stock options with Biohaven Pharmaceuticals Medical Sciences, BlackThorn Therapeutics, Inc., EpiVario, Inc. and Terran Life Sciences; and is editor of *Biological Psychiatry* with income greater than $10,000. I.B.H. is the Co-Director of Health and Policy at the Brain and Mind Centre (BMC) University of Sydney. The BMC operates an early-intervention youth services at Camperdown under contract to Headspace. He is the Chief Scientific Advisor to, and a 3.2% equity shareholder in, InnoWell Pty Ltd. InnoWell was formed by the University of Sydney (45% equity) and PwC (Australia; 45% equity) to deliver the $30 M Australian Government-funded Project Synergy (2017-20; a three-year program for the transformation of mental health services) and to lead transformation of mental health services internationally through the use of innovative technologies. J.W.S. is a member of the Leon Levy Foundation Neuroscience Advisory Board, the Scientific Advisory Board of Sensorium Therapeutics (with equity), and has received grant support from Biogen, Inc. He is PI of a collaborative study of the genetics of depression and bipolar disorder sponsored by 23andMe for which 23andMe provides analysis time as in-kind support but no payments. All other authors report no biomedical financial interests or potential conflicts of interest.

medRxiv preprint doi: https://doi.org/10.1101/2023.01.24.23284960; this version posted January 30, 2023. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-NC-ND 4.0 International license .

References

1. Grant, B.F., *et al.* Epidemiology of DSM-5 Alcohol Use Disorder: Results From the National Epidemiologic Survey on Alcohol and Related Conditions III. *JAMA Psychiatry* **72**, 757-766 (2015).
2. Verhulst, B., Neale, M.C. & Kendler, K.S. The heritability of alcohol use disorders: a meta-analysis of twin and adoption studies. *Psychol Med* **45**, 1061-1072 (2015).
3. Gelernter, J. & Polimanti, R. Genetics of substance use disorders in the era of big data. *Nat Rev Genet* **22**, 712-729 (2021).
4. Sanchez-Roige, S., *et al.* Genome-Wide Association Study Meta-Analysis of the Alcohol Use Disorders Identification Test (AUDIT) in Two Population-Based Cohorts. *Am J Psychiatry* **176**, 107-118 (2019).
5. Walters, R.K., *et al.* Transancestral GWAS of alcohol dependence reveals common genetic underpinnings with psychiatric disorders. *Nat Neurosci* **21**, 1656-1669 (2018).
6. 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).
7. 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).
8. Gelernter, J., *et al.* Genome-wide Association Study of Maximum Habitual Alcohol Intake in >140,000 U.S. European and African American Veterans Yields Novel Risk Loci. *Biol Psychiatry* **86**, 365-376 (2019).
9. Deak, J.D., *et al.* Genome-Wide Investigation of Maximum Habitual Alcohol Intake in US Veterans in Relation to Alcohol Consumption Traits and Alcohol Use Disorder. *JAMA Netw Open* **5**, e2238880 (2022).
10. 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).
11. Saunders, G.R.B., *et al.* Genetic diversity fuels gene discovery for tobacco and alcohol use. *Nature* (2022).
12. Saunders, J.B., Aasland, O.G., Babor, T.F., de la Fuente, J.R. & Grant, M. Development of the Alcohol Use Disorders Identification Test (AUDIT): WHO Collaborative Project on Early Detection of Persons with Harmful Alcohol Consumption--II. *Addiction* **88**, 791-804 (1993).
13. 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 (2022).
14. Johnson, E.C., *et al.* A large-scale genome-wide association study meta-analysis of cannabis use disorder. *Lancet Psychiatry* **7**, 1032-1045 (2020).
15. 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).
16. Peterson, R.E., *et al.* Genome-wide Association Studies in Ancestrally Diverse Populations: Opportunities, Methods, Pitfalls, and Recommendations. *Cell* **179**, 589-603 (2019).
17. Bycroft, C., *et al.* The UK Biobank resource with deep phenotyping and genomic data. *Nature* **562**, 203-209 (2018).
18. Zheutlin, A.B., *et al.* Penetrance and Pleiotropy of Polygenic Risk Scores for Schizophrenia in 106,160 Patients Across Four Health Care Systems. *Am J Psychiatry* **176**, 846-855 (2019).
19. 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).

medRxiv preprint doi: https://doi.org/10.1101/2023.01.24.23284960; this version posted January 30, 2023. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-NC-ND 4.0 International license .

20. Kurki, M.I., *et al.* FinnGen: Unique genetic insights from combining isolated population and national health register data. *medRxiv* **doi: https://doi.org/10.1101/2022.03.03.22271360**(2022).
21. Pedersen, C.B., *et al.* The iPSYCH2012 case-cohort sample: new directions for unravelling genetic and environmental architectures of severe mental disorders. *Mol Psychiatry* **23**, 6-14 (2018).
22. Bybjerg-Grauholm, J., *et al.* The iPSYCH2015 Case-Cohort sample: updated directions for unravelling genetic and environmental architectures of severe mental disorders. *medRxiv* **doi: https://doi.org/10.1101/2020.11.30.20237768**(2020).
23. Lind, P., *et al.* Preliminary results from the Australian Genetics of Bipolar Disorder Study: A nation-wide cohort. *Psyarxive* **10.31234/osf.io/yd3sx**(2022).
24. Byrne, E.M., *et al.* Cohort profile: the Australian genetics of depression study. *BMJ Open* **10**, e032580 (2020).
25. Couvy-Duchesne, B., *et al.* Nineteen and Up study (19Up): understanding pathways to mental health disorders in young Australian twins. *BMJ Open* **8**, e018959 (2018).
26. Gelernter, J., *et al.* Genome-wide association study of alcohol dependence:significant findings in African- and European-Americans including novel risk loci. *Mol Psychiatry* **19**, 41-49 (2014).
27. Zhou, H., *et al.* Genome-wide meta-analysis of alcohol use disorder in East Asians. *Neuropsychopharmacology* **47**, 1791-1797 (2022).
28. Kember, R.L., *et al.* Genetic underpinnings of the transition from alcohol consumption to alcohol use disorder: shared and unique genetic architectures in a cross-ancestry sample. *medRxiv* **https://doi.org/10.1101/2021.09.08.21263302**(2021).
29. Gelernter, J., *et al.* Genomewide Association Study of Alcohol Dependence and Related Traits in a Thai Population. *Alcohol Clin Exp Res* **42**, 861-868 (2018).
30. Trubetskoy, V., *et al.* Mapping genomic loci implicates genes and synaptic biology in schizophrenia. *Nature* **604**, 502-508 (2022).
31. Okbay, A., *et al.* Polygenic prediction of educational attainment within and between families from genome-wide association analyses in 3 million individuals. *Nat Genet* **54**, 437-449 (2022).
32. 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).
33. Watanabe, K., Taskesen, E., van Bochoven, A. & Posthuma, D. Functional mapping and annotation of genetic associations with FUMA. *Nat Commun* **8**, 1826 (2017).
34. 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).
35. GTEx Consortium. The GTEx Consortium atlas of genetic regulatory effects across human tissues. *Science* **369**, 1318-1330 (2020).
36. Barbeira, A.N., *et al.* Integrating predicted transcriptome from multiple tissues improves association detection. *PLoS genetics* **15**, e1007889 (2019).
37. Sey, N.Y.A., *et al.* A computational tool (H-MAGMA) for improved prediction of brain-disorder risk genes by incorporating brain chromatin interaction profiles. *Nat Neurosci* **23**, 583-593 (2020).
38. Mancuso, N., *et al.* Probabilistic fine-mapping of transcriptome-wide association studies. *Nat Genet* **51**, 675-682 (2019).
39. Ochoa, D., *et al.* The next-generation Open Targets Platform: reimagined, redesigned, rebuilt. *Nucleic Acids Res* **51**, D1353-D1359 (2023).
40. Ruan, Y., *et al.* Improving polygenic prediction in ancestrally diverse populations. *Nat Genet* **54**, 573-580 (2022).

medRxiv preprint doi: https://doi.org/10.1101/2023.01.24.23284960; this version posted January 30, 2023. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-NC-ND 4.0 International license .

41. Kember, R.L., *et al.* Cross-ancestry meta-analysis of opioid use disorder uncovers novel loci with predominant effects in brain regions associated with addiction. *Nat Neurosci* **25**, 1279-1287 (2022).
42. Xu, K., *et al.* Genome-wide association study of smoking trajectory and meta-analysis of smoking status in 842,000 individuals. *Nat Commun* **11**, 5302 (2020).
43. Levey, D.F., *et al.* Bi-ancestral depression GWAS in the Million Veteran Program and meta-analysis in >1.2 million individuals highlight new therapeutic directions. *Nat Neurosci* **24**, 954-963 (2021).
44. Vujkovic, M., *et al.* Discovery of 318 new risk loci for type 2 diabetes and related vascular outcomes among 1.4 million participants in a multi-ancestry meta-analysis. *Nat Genet* **52**, 680-691 (2020).
45. Kember, R.L., *et al.* Phenome-wide Association Analysis of Substance Use Disorders in a Deeply Phenotyped Sample. *Biol Psychiatry*, Aug 18:S0006-3223(0022)01515-01513. doi: 01510.01016/j.biopsych.02022.01508.01010. Epub ahead of print. PMID: 36273948. (2022).
46. 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).
47. Deak, J.D., *et al.* 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. *Mol Psychiatry* **27**, 3970-3979 (2022).
48. Schwantes-An, T.H., *et al.* Association of the OPRM1 Variant rs1799971 (A118G) with Non-Specific Liability to Substance Dependence in a Collaborative de novo Meta-Analysis of European-Ancestry Cohorts. *Behav Genet* **46**, 151-169 (2016).
49. Notaras, M., Hill, R. & van den Buuse, M. The BDNF gene Val66Met polymorphism as a modifier of psychiatric disorder susceptibility: progress and controversy. *Mol Psychiatry* **20**, 916-930 (2015).
50. 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).
51. Schizophrenia Working Group of the Psychiatric Genomics, C. Biological insights from 108 schizophrenia-associated genetic loci. *Nature* **511**, 421-427 (2014).
52. Mealer, R.G., *et al.* The schizophrenia risk locus in SLC39A8 alters brain metal transport and plasma glycosylation. *Sci Rep* **10**, 13162 (2020).
53. Mealer, R.G., *et al.* Glycobiology and schizophrenia: a biological hypothesis emerging from genomic research. *Mol Psychiatry* **25**, 3129-3139 (2020).
54. Amit, Z. & Smith, B.R. A multi-dimensional examination of the positive reinforcing properties of acetaldehyde. *Alcohol* **2**, 367-370 (1985).
55. Mews, P., *et al.* Alcohol metabolism contributes to brain histone acetylation. *Nature* **574**, 717-721 (2019).
56. Pandey, S.C., Ugale, R., Zhang, H., Tang, L. & Prakash, A. Brain chromatin remodeling: a novel mechanism of alcoholism. *J Neurosci* **28**, 3729-3737 (2008).
57. Carlsson, C. & Gullberg, B. A double blind study with melperone and placebo in the treatment of chronic alcoholics. *Int J Clin Pharmacol Biopharm* **16**, 331-332 (1978).
58. Carlsson, C., Gullberg, B., Hostery, U. & Christensson, E. A double-blind study of melperone and placebo in hospitalized chronic alcoholics in postintoxication phase. *Int J Clin Pharmacol Biopharm* **17**, 341-345 (1979).
59. Farokhnia, M., *et al.* Spironolactone as a potential new pharmacotherapy for alcohol use disorder: convergent evidence from rodent and human studies. *Mol Psychiatry* **27**, 4642-4652 (2022).

medRxiv preprint doi: https://doi.org/10.1101/2023.01.24.23284960; this version posted January 30, 2023. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-NC-ND 4.0 International license .

60. Sychla, H., Grunder, G. & Lammertz, S.E. Comparison of Clomethiazole and Diazepam in the Treatment of Alcohol Withdrawal Syndrome in Clinical Practice. *Eur Addict Res* **23**, 211-218 (2017).
61. Hunter-Zinck, H., *et al.* Genotyping Array Design and Data Quality Control in the Million Veteran Program. *Am J Hum Genet* **106**, 535-548 (2020).
62. Price, A.L., *et al.* Principal components analysis corrects for stratification in genome-wide association studies. *Nat Genet* **38**, 904-909 (2006).
63. Galinsky, K.J., *et al.* Fast Principal-Component Analysis Reveals Convergent Evolution of ADH1B in Europe and East Asia. *Am J Hum Genet* **98**, 456-472 (2016).
64. 1000 Genomes Project Consortium. A global reference for human genetic variation. *Nature* **526**, 68-74 (2015).
65. Delaneau, O., Zagury, J.F., Robinson, M.R., Marchini, J.L. & Dermitzakis, E.T. Accurate, scalable and integrative haplotype estimation. *Nat Commun* **10**, 5436 (2019).
66. Das, S., *et al.* Next-generation genotype imputation service and methods. *Nat Genet* **48**, 1284-1287 (2016).
67. Loh, P.R., Kichaev, G., Gazal, S., Schoech, A.P. & Price, A.L. Mixed-model association for biobank-scale datasets. *Nat Genet* **50**, 906-908 (2018).
68. Zhou, W., *et al.* Efficiently controlling for case-control imbalance and sample relatedness in large-scale genetic association studies. *Nat Genet* **50**, 1335-1341 (2018).
69. Lam, M., *et al.* RICOPILI: Rapid Imputation for COnsortias PIpeLIne. *Bioinformatics* **36**, 930-933 (2020).
70. Demontis, D., *et al.* Genome-wide analyses of ADHD identify 27 risk loci, refine the genetic architecture and implicate several cognitive domains. *medRxiv* **https://doi.org/10.1101/2022.02.14.22270780**, 2022.2002.2014.22270780 (2022).
71. Chang, C.C., *et al.* Second-generation PLINK: rising to the challenge of larger and richer datasets. *Gigascience* **4**, 7 (2015).
72. Heath, A.C., *et al.* A quantitative-trait genome-wide association study of alcoholism risk in the community: findings and implications. *Biol Psychiatry* **70**, 513-518 (2011).
73. Zhou, X. & Stephens, M. Efficient multivariate linear mixed model algorithms for genome-wide association studies. *Nat Methods* **11**, 407-409 (2014).
74. Willer, C.J., Li, Y. & Abecasis, G.R. METAL: fast and efficient meta-analysis of genomewide association scans. *Bioinformatics* **26**, 2190-2191 (2010).
75. Yang, J., *et al.* Conditional and joint multiple-SNP analysis of GWAS summary statistics identifies additional variants influencing complex traits. *Nat Genet* **44**, 369-375, S361-363 (2012).
76. International HapMap Consortium, *et al.* Integrating common and rare genetic variation in diverse human populations. *Nature* **467**, 52-58 (2010).
77. Bulik-Sullivan, B.K., *et al.* LD Score regression distinguishes confounding from polygenicity in genome-wide association studies. *Nat Genet* **47**, 291-295 (2015).
78. Brown, B.C., Asian Genetic Epidemiology Network Type 2 Diabetes, C., Ye, C.J., Price, A.L. & Zaitlen, N. Transethnic Genetic-Correlation Estimates from Summary Statistics. *Am J Hum Genet* **99**, 76-88 (2016).
79. LaPierre, N., *et al.* Identifying causal variants by fine mapping across multiple studies. *PLoS genetics* **17**, e1009733 (2021).
80. GTEx Consortium. Genetic effects on gene expression across human tissues. *Nature* **550**, 204-213 (2017).
81. Laakso, M., *et al.* The Metabolic Syndrome in Men study: a resource for studies of metabolic and cardiovascular diseases. *J Lipid Res* **58**, 481-493 (2017).

medRxiv preprint doi: https://doi.org/10.1101/2023.01.24.23284960; this version posted January 30, 2023. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-NC-ND 4.0 International license .

82. Ligthart, L., *et al.* The Netherlands Twin Register: Longitudinal Research Based on Twin and Twin-Family Designs. *Twin Res Hum Genet* **22**, 623-636 (2019).
83. Turpeinen, H., *et al.* A genome-wide expression quantitative trait loci analysis of proprotein convertase subtilisin/kexin enzymes identifies a novel regulatory gene variant for FURIN expression and blood pressure. *Hum Genet* **134**, 627-636 (2015).
84. Hoffman, G.E., *et al.* CommonMind Consortium provides transcriptomic and epigenomic data for Schizophrenia and Bipolar Disorder. *Sci Data* **6**, 180 (2019).
85. Subramanian, A., *et al.* A Next Generation Connectivity Map: L1000 Platform and the First 1,000,000 Profiles. *Cell* **171**, 1437-1452 e1417 (2017).
86. So, H.C., *et al.* Analysis of genome-wide association data highlights candidates for drug repositioning in psychiatry. *Nat Neurosci* **20**, 1342-1349 (2017).
87. Viechtbauer, W. Conducting Meta-Analyses in R with the metafor Package. *Journal of Statistical Software* **36(3), 1–48**(2010).
88. 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).
89. Bulik-Sullivan, B., *et al.* An atlas of genetic correlations across human diseases and traits. *Nat Genet* **47**, 1236-1241 (2015).
90. Luo, Y., *et al.* Estimating heritability and its enrichment in tissue-specific gene sets in admixed populations. *Hum Mol Genet* **30**, 1521-1534 (2021).
91. Boutin, N.T., *et al.* The Evolution of a Large Biobank at Mass General Brigham. *J Pers Med* **12**(2022).
92. Verma, A., *et al.* The Penn Medicine BioBank: Towards a Genomics-Enabled Learning Healthcare System to Accelerate Precision Medicine in a Diverse Population. *J Pers Med* **12**(2022).
93. Abraham, G., Qiu, Y. & Inouye, M. FlashPCA2: principal component analysis of Biobank-scale genotype datasets. *Bioinformatics* **33**, 2776-2778 (2017).
94. Loh, P.R., *et al.* Reference-based phasing using the Haplotype Reference Consortium panel. *Nat Genet* **48**, 1443-1448 (2016).
95. Denny, J.C., *et al.* Systematic comparison of phenome-wide association study of electronic medical record data and genome-wide association study data. *Nature biotechnology* **31**, 1102-1110 (2013).