UNIVERSITÀ DEGLI STUDI DI CAMERINO

School of Advanced Studies

DOCTORAL COURSE IN

Chemical and Pharmaceutical Sciences, and Biotechnology

***XXXI* cycle**

Deciphering Genetic Traits Toward Alcohol Drinking

Phd Candidate:

Giovanni Deiana

Supervisor:

Prof. Roberto Ciccocioppo

Co-supervisor:

Prof. Valerio Napolioni

Table of Contents

Introduction

Global trends in alcohol use

Being part of many cultural, religious, and social practices, alcohol drinking patterns were shaped by several socioeconomic factors though history (Heinz & Daedelow, 2021). Following economic development levels of total alcohol per capita consumption (APC) are gradually decreasing on a worldwide scale (Holmes & Anderson, 2017). On a regional level Europe shows diverse trends in alcohol consumption, but compared to other macro-regions it is still those showing the highest APC levels (Manthey et al., 2019). Extensive metanalyses including data from 1980 to 2021, shown that while occasional or moderate daily alcohol consumption (from 0 to 44 g per day) is not significantly associated to reduced risk of all-cause mortality, huge consumption (45g or more per day) significantly increased it (Zhao et al., 2023). Differences in alcohol use between genders is narrowing as APC is increasing among female drinkers (Calvo et al., 2020). Cumulative evidence shows that escalation from casual alcohol use to the development of problematic alcohol use seems to be more rapid in women than in men, and that relapse following a stressful event or drug-related cue is also more likely to occur among female drinkers (Becker et al., 2017; Keyes et al., 2019). Interestingly, sexual orientation also influences this trend, as non-heterosexual women are more likely to consume 12 or more drinks on an occasion which is three times the standard binge threshold for women (White, 2020). Furthermore, the recovery from problematic alcohol use for female drinkers is hardened by lower treatment entry and inadequate sample sizes in pharmacological study (Holzhauer et al., 2020). Despite huge variability, cumulative evidence shows that alcohol use incurred substantial social costs, which can be estimated to an average amount of 1306 Int$ per adult or 2.6% of average GDP (Manthey et al., 2021). Efforts to prevent alcohol use in older adults are

not considered cost-effective, interventions aimed at those under 18 years old are therefore the most effective measure to reduce such economic burden (Le et al., 2023).

Alcohol use in communicable and non-communicable diseases

In May 2013, the American Psychiatric Association issued unified the previous distinct definition of alcohol abuse and alcohol dependence into alcohol use disorder (AUD) the fifth edition of the Diagnostic and Statistical Manual of Mental Disorders (DSM–5) (Grant et al., 2015). This division is maintained in the latest of the International Classification of Diseases published by World Health Organization (ICD-11), which distinguishing among episodic harmful use of alcohol, harmful pattern of use of alcohol, and alcohol dependence (Saunders et al., 2019). While the alignment with both current and historical versions of DSM and ICD dependence criteria has been improved in ICD-11, diagnostic consistency with DSM-5 is still weak (Lundin et al., 2021). While ICD-11-related papers outnumbers those of ICD-10 (Chen et al., 2019), the definitory apparatus proposed in DSM–5 stays prominent in alcohol research (Rehm et al., 2019; Tyrer, 2018). The rate of problematic alcohol use developing into alcohol use disorder (AUD) makes it the second most prevalent substance use disorder (estimated at 5.1% in past year), right after nicotine use disorder (estimated at 20% in past year) (Alcohol & Drug Use, 2018). Nicotine and alcohol influence midbrain dopaminergic neurons, leading to mutual reinforcement and cross-tolerance effects, while also heightening susceptibility to risky behaviors (Morel et al., 2019; Poisson et al., 2021). Smoking and alcohol use are indeed associated to impulsive decision making (Ariesen et al., 2023; Gabriel et al., 2019) which increase the risk of experiencing traumatic events, especially among young adults (Bartholomew et al., 2021). Besides other SUD, AUD is frequently associated with a high comorbidity rate with various psychiatric disorders, especially mood and anxiety-related disorders (Castillo-Carniglia et al., 2019; Puddephatt et al., 2022). Stress and maladaptive

response to traumatic events can eventually lead to the development of post-traumatic stress disorder (PTSD) which among anxiety disorders is the one with highest comorbidity rates towards AUD and is itself a predictor for SUD (Smith & Cottler, 2018). Although the interplay of inflammatory processes and cognition has been identified in AUD, mood and anxiety-related disorders, specific molecular mechanisms delineating a clear causal link between them remain elusive (Erickson et al., 2019). The diverse patterns of high comorbidity rates linking AUD to other psychiatric disorders indeed present epistemological challenges that impact the effectiveness of clinical treatment (Carter et al., 2013). Several physical comorbidities are widely recognized as risk factors for alcohol consumption, such as high blood pressure, asthma, hyperlipidemia, and liver enzyme abnormalities (AshaRani et al., 2022). Setting aside other psychiatric disorders, population-attributable fractions of risk due to alcohol consumption are comparatively higher observed for Tuberculosis and Malignant neoplasms among communicable and non-communicable diseases, respectively (Shield et al., 2020). Due to its direct impact on the mucosal immune system, alcohol use is indeed significant burden to infectious diseases epidemiology (Morojele et al., 2021; Szabo & Saha, 2015; Trevejo-Nunez et al., 2015). As carcinogenic metabolite of ethanol, acetaldehyde is one of the biological pathways by which alcohol use leads to cancer development, especially in esophageal tissues (Sasaki et al., 2021). Interestingly, while Asia scores the lowest in APC levels (Manthey et al., 2019), population living in that area score highest for both alcohol-attributable cancer cases (Rumgay et al., 2021) and their demographic landscape was deeply influenced by the spread of tuberculosis infection (Liu et al., 2017).

Genomics in alcohol research

Falling costs of sequencing technology causing a paradigm shift from microarray-based genotyping studies to whole exome and whole genome sequencing allowed the fast rise of

genome-wide association studies (GWAS) (Heather & Chain, 2016). Similar advances in RNA sequencing technology allowed wide analysis of differential gene expression and differential splicing of mRNAs (Stark et al., 2019) and eventually the emerging of transcriptome-wide association studies (TWAS) (Wainberg et al., 2019). Leveraging large sample size and extensive genomic coverage, GWAS test the associations of each sequenced single nucleotide polymorphism (SNPs) to a given phenotype through linear or logistic regression models adjusted by covariates such as age, sex and ancestry (Uffelmann et al., 2021). Across different loci, the nonrandom association of SNPs is quantified by their level of Linkage disequilibrium (LD) that allows to evaluate the subdivision of a population, the heritability of trait and due to signals of natural selection or genetic drift. LD is crucial in post GWAS analysis as, through linear models based on LD (Slatkin, 2008) scores, allows to test correlation heritable traits (Bulik-Sullivan et al., 2015). LD is also employed as a marker to distinguish causal SNPs for a given phenotype and eventually use them as instrumental variables to evaluate causality between genomically correlated traits, such as in Mendelian Randomization analysis (MR) (Davey Smith & Hemani, 2014). Besides the complex and diverse interplay of environmental influences, and history of alcohol exposure that may eventually lead to developing AUD; alcohol use is indeed a hereditable trait (Schuckit, 2014). Consistent evidence of genetic factors contributing to susceptibility towards AUD has been provided by twin, adoption, and family studies (Verhulst et al., 2015). Large-scale GWAS conducted on multiethnic populations have demonstrated that predispositions to alcohol use are heritable irrespective of cultural influences (Saunders et al., 2022). While most variants associated to AUD occur in intergenic or intronic regions and are not directly associated with protein coding few others, like those affecting genes encoding for the enzymes alcohol dehydrogenase (ADH) and aldehyde dehydrogenase (ALDH) (Hendershot et al., 2009; Peng & Yin, 2009), results in actual functional

changes thus having either a protective or detrimental effect (Liu et al., 2019). According to recent metanalysis on functional variants associated to alcohol consumption, other loci attaining genome-wide significant association were *ADH1B, KLB*, *BTF3P13*, *GCKR*, *SLC39A8*, and *DRD2* (Thompson et al., 2020). AUD, being a polygenic phenotype, exhibits a significant degree of shared pleiotropy with various traits, particularly with SUD and mood-related disorders (Icick et al., 2023). While genetic information can be used as an anchor for causal inference (Jennings et al., 2024; Kember et al., 2023), the presence of extensive shared pleiotropy poses an epistemological hurdle for genomic investigations in AUD (McCarthy et al., 2008). Therefore, also in AUD research, integrating ecological, transcriptomic and GWAS derived data is imperative to ensure a dependable interpretation of novel findings driven by genomic insights (Dai et al., 2020).

Aim and relevance of the thesis

Considering the intricate nature of alcohol consumption behaviors and their varied manifestations, this thesis aims to unveil novel insight and perspectives on causal mechanism affecting patterns within alcohol consumption and their interplay with other related traits. In the first chapter of this thesis, we focus our attentions upon the consistent pattern of low alcohol consumptions among of Asian populations (Manthey et al., 2019). Carriers of ADH1B2 (Arg48His, rs1229984) and ALDH22 (Glu504Lys, rs671) genetic variants are recognized for their heightened sensitivity to the adverse effects of alcohol consumption due to their effect on acetaldehyde toxicity. Notably, these two functional genetic variations are mostly prevalent within the Southeast Asia Region (SAER) and are rarely found in non-Asian populations (Polimanti & Gelernter, 2018; Zhang et al., 2021). While the magnitude of selective pressure exerted on population living in SAER is well known; postulated the content of "the aldehyde hypothesis" (Darwin & Stanley, 2022), ours was the first study to systematically evaluate the contribution of infectious diseases and other ecological variables to the selection these two genomic variants in Asian populations (Deiana et al., 2024). Our analyses suggest that *Mycobacteria* may have played a role in the joint selection of *ADH1B*2* and *ALDH2**2. Later in the second chapter, we conducted an extensive review of available data on RNA-seq data for AUD and PTSD to find signals of shared transcriptomic signatures between the two. our analysis evaluated for the first time the explanatory role of shared enriched pathways across dorsolateral prefrontal cortex and basolateral amygdala on bidirectional pattern of comorbidity linking AUD and PTSD (Dell'Aquila & Berle, 2023). Our results underline the role of family A (rhodopsin-like) G-protein coupled receptors (GPCRs), and tyrosine kinases receptors (RTKs), and their clinical affordance in both AUD and PTSD. Lastly, in the third chapter, we investigated the pleiotropic mechanisms linking brain structure and function to alcohol and tobacco use.

Given the complex interplay of alcohol and nicotine with human brain (Alcohol & Drug Use, 2018; He et al., 2022), we applied a wide apparatus of post GWAS analysis including linkage disequilibrium score regression (Bulik-Sullivan et al., 2015), local analysis of [co]variant association (Werme et al., 2022) and genetically inferred causal inference analyses (Hemani et al., 2018; O'Connor & Price, 2018). Our brain-wide investigation highlighted that different pleiotropic mechanisms likely contribute to the relationship of brain structure and function with alcohol drinking and tobacco smoking, suggesting decision-making activities and chemosensory processing as modulators of propensity towards alcohol and tobacco consumption.

References

Abburi, C., Wolfman, S. L., Metz, R. A., Kamber, R., McGehee, D. S., & McDaid, J. (2016). Tolerance to Ethanol or Nicotine Results in Increased Ethanol Self-Administration and Long-Term Depression in the Dorsolateral Striatum. *eNeuro*, *3*(4). https://doi.org/10.1523/ENEURO.0112-15.2016

Agrawal, A., Balci, H., Hanspers, K., Coort, S. L., Martens, M., Slenter, D. N., Ehrhart, F., Digles, D., Waagmeester, A., Wassink, I., Abbassi-Daloii, T., Lopes, E. N., Iyer, A., Acosta, J. M., Willighagen, L. G., Nishida, K., Riutta, A., Basaric, H., Evelo, C. T., . . . Pico, A. R. (2024). WikiPathways 2024: next generation pathway database. *Nucleic Acids Res*, *52*(D1), D679-D689. https://doi.org/10.1093/nar/gkad960

Ajmani, G. S., Suh, H. H., Wroblewski, K. E., & Pinto, J. M. (2017). Smoking and olfactory dysfunction: A systematic literature review and meta-analysis. *Laryngoscope*, *127*(8), 1753-1761. https://doi.org/10.1002/lary.26558

Alcohol, G. B. D., & Drug Use, C. (2018). The global burden of disease attributable to alcohol and drug use in 195 countries and territories, 1990-2016: a systematic analysis for the Global Burden of Disease Study 2016. *Lancet Psychiatry*, *5*(12), 987-1012. https://doi.org/10.1016/S2215-0366(18)30337-7

Alexandra Kredlow, M., Fenster, R. J., Laurent, E. S., Ressler, K. J., & Phelps, E. A. (2022). Prefrontal cortex, amygdala, and threat processing: implications for PTSD. *Neuropsychopharmacology*, *47*(1), 247-259. https://doi.org/10.1038/s41386-021-01155-7

Alfaro-Almagro, F., Jenkinson, M., Bangerter, N. K., Andersson, J. L. R., Griffanti, L., Douaud, G., Sotiropoulos, S. N., Jbabdi, S., Hernandez-Fernandez, M., Vallee, E., Vidaurre, D., Webster, M., McCarthy, P., Rorden, C., Daducci, A., Alexander, D. C., Zhang, H., Dragonu, I., Matthews, P. M., . . . Smith, S. M. (2018). Image processing and Quality Control for the first 10,000 brain imaging datasets from UK Biobank. *Neuroimage*, *166*, 400-424. https://doi.org/10.1016/j.neuroimage.2017.10.034

Apps, M. A., Rushworth, M. F., & Chang, S. W. (2016). The Anterior Cingulate Gyrus and Social Cognition: Tracking the Motivation of Others. *Neuron*, *90*(4), 692-707. https://doi.org/10.1016/j.neuron.2016.04.018

Ariesen, A. D., Neubert, J. H., Gaastra, G. F., Tucha, O., & Koerts, J. (2023). Risky Decision-Making in Adults with Alcohol Use Disorder-A Systematic and Meta-Analytic Review. *J Clin Med*, *12*(8). https://doi.org/10.3390/jcm12082943

AshaRani, P. V., Karuvetil, M. Z., Brian, T. Y. W., Satghare, P., Roystonn, K., Peizhi, W., Cetty, L., Zainuldin, N. A., & Subramaniam, M. (2022). Prevalence and Correlates of Physical Comorbidities in Alcohol Use Disorder (AUD): a Pilot Study in Treatment-Seeking Population. *Int J Ment Health Addict*, 1-18. https://doi.org/10.1007/s11469-021-00734-5

Asiimwe, I. G., Blockman, M., Cohen, K., Cupido, C., Hutchinson, C., Jacobson, B., Lamorde, M., Morgan, J., Mouton, J. P., Nakagaayi, D., Okello, E., Schapkaitz, E., Sekaggya-Wiltshire, C., Semakula, J. R., Waitt, C., Zhang, E. J., Jorgensen, A. L., & Pirmohamed, M. (2022). A genome-wide association study of plasma concentrations of warfarin enantiomers and metabolites in sub-Saharan black-African patients. *Front Pharmacol*, *13*, 967082. https://doi.org/10.3389/fphar.2022.967082

Association, A. P. (2013). *Diagnostic and Statistical Manual of Mental Disorders* (5th ed.). https://doi.org/https://doi.org/10.1176/appi.books.9780890425596

Atli, A., Bulut, M., Bez, Y., Kaplan, I., Ozdemir, P. G., Uysal, C., Selcuk, H., & Sir, A. (2016). Altered lipid peroxidation markers are related to post-traumatic stress disorder (PTSD) and not trauma itself in earthquake survivors. *Eur Arch Psychiatry Clin Neurosci*, *266*(4), 329-336. https://doi.org/10.1007/s00406-015-0638-5

Back, S. E., Brady, K. T., Sonne, S. C., & Verduin, M. L. (2006). Symptom improvement in co-occurring PTSD and alcohol dependence. *J Nerv Ment Dis*, *194*(9), 690-696. https://doi.org/10.1097/01.nmd.0000235794.12794.8a

Bader, G. D., & Hogue, C. W. (2003). An automated method for finding molecular complexes in large protein interaction networks. *BMC Bioinformatics*, *4*, 2. https://doi.org/10.1186/1471-2105-4-2

Bainomugisa, C. K., Sutherland, H. G., Parker, R., McRae, A. F., Haupt, L. M., Griffiths, L. R., Heath, A., Nelson, E. C., Wright, M. J., Hickie, I. B., Martin, N. G., Nyholt, D. R., & Mehta, D. (2021). Using Monozygotic Twins to Dissect Common Genes in Posttraumatic Stress Disorder and Migraine. *Front Neurosci*, *15*, 678350. https://doi.org/10.3389/fnins.2021.678350

Bao, J., Wen, J., Wen, Z., Yang, S., Cui, Y., Yang, Z., Erus, G., Saykin, A. J., Long, Q., Davatzikos, C., & Shen, L. (2023). Brain-wide genome-wide colocalization study for integrating

genetics, transcriptomics and brain morphometry in Alzheimer's disease. *Neuroimage*, *280*, 120346. https://doi.org/10.1016/j.neuroimage.2023.120346

Barrett, T., Wilhite, S. E., Ledoux, P., Evangelista, C., Kim, I. F., Tomashevsky, M., Marshall, K. A., Phillippy, K. H., Sherman, P. M., Holko, M., Yefanov, A., Lee, H., Zhang, N., Robertson, C. L., Serova, N., Davis, S., & Soboleva, A. (2013). NCBI GEO: archive for functional genomics data sets--update. *Nucleic Acids Res*, *41*(Database issue), D991-995. https://doi.org/10.1093/nar/gks1193

Bartel, G., Marko, M., Rameses, I., Lamm, C., & Riecansky, I. (2020). Left Prefrontal Cortex Supports the Recognition of Meaningful Patterns in Ambiguous Stimuli. *Front Neurosci*, *14*, 152. https://doi.org/10.3389/fnins.2020.00152

Bartholomew, R. D., Kerry-Barnard, S., Beckley-Hoelscher, N. D., Phillips, R., Reid, F., Fleming, C., Lesniewska, A., Yoward, F. D., & Oakeshott, P. (2021). Alcohol use, cigarette smoking, vaping and number of sexual partners: A cross-sectional study of sexually active, ethnically diverse, inner city adolescents. *Health Expect*, *24*(3), 1009-1014. https://doi.org/10.1111/hex.13248

Becker, J. B., McClellan, M. L., & Reed, B. G. (2017). Sex differences, gender and addiction. *J Neurosci Res*, *95*(1-2), 136-147. https://doi.org/10.1002/jnr.23963

Benvenuti, F., Colombo, D., Soverchia, L., Cannella, N., Domi, E., & Ciccocioppo, R. (2023). Psilocybin prevents reinstatement of alcohol seeking by disrupting the reconsolidation of alcohol-related memories. *Psychopharmacology (Berl)*, *240*(7), 1521-1530. https://doi.org/10.1007/s00213-023-06384-w

Bountress, K. E., Brick, L. A., Sheerin, C., Grotzinger, A., Bustamante, D., Hawn, S. E., Gillespie, N., Kirkpatrick, R. M., Kranzler, H., Morey, R., Edenberg, H. J., Maihofer, A. X., Disner, S., Ashley-Koch, A., Peterson, R., Lori, A., Stein, D. J., Kimbrel, N., Nievergelt, C., . . . Amstadter, A. B. (2022). Alcohol use and alcohol use disorder differ in their genetic relationships with PTSD: A genomic structural equation modelling approach. *Drug Alcohol Depend*, *234*, 109430. https://doi.org/10.1016/j.drugalcdep.2022.109430

Bountress, K. E., Bustamante, D., Subbie-Saenz de Viteri, S., Chatzinakos, C., Sheerin, C., Daskalakis, N. P., Edenberg, H. J., Psychiatric Genomics Consortium Posttraumatic Stress Disorder Working, G., Peterson, R. E., Webb, B. T., Meyers, J., & Amstadter, A. (2023). Differences in genetic correlations between posttraumatic stress disorder and alcohol-related problems phenotypes compared to alcohol consumption-related

phenotypes. *Psychol Med*, *53*(12), 5767-5777. https://doi.org/10.1017/S0033291722002999

Bountress, K. E., Wendt, F., Bustamante, D., Agrawal, A., Webb, B., Gillespie, N., Edenberg, H., Sheerin, C., Johnson, E., Psychiatric Genomics Consortium Posttraumatic Stress Disorder Working, G., Polimanti, R., & Amstadter, A. (2021). Potential causal effect of posttraumatic stress disorder on alcohol use disorder and alcohol consumption in individuals of European descent: A Mendelian Randomization Study. *Alcohol Clin Exp Res*, *45*(8), 1616-1623. https://doi.org/10.1111/acer.14649

Breen, M. S., Maihofer, A. X., Glatt, S. J., Tylee, D. S., Chandler, S. D., Tsuang, M. T., Risbrough, V. B., Baker, D. G., O'Connor, D. T., Nievergelt, C. M., & Woelk, C. H. (2015). Gene networks specific for innate immunity define post-traumatic stress disorder. *Mol Psychiatry*, *20*(12), 1538-1545. https://doi.org/10.1038/mp.2015.9

Budenz, A., Klein, A., & Prutzman, Y. (2021). The Relationship Between Trauma Exposure and Adult Tobacco Use: Analysis of the National Epidemiologic Survey on Alcohol and Related Conditions (III). *Nicotine Tob Res*, *23*(10), 1716-1726. https://doi.org/10.1093/ntr/ntab057

Bulik-Sullivan, B., Finucane, H. K., Anttila, V., Gusev, A., Day, F. R., Loh, P. R., ReproGen, C., Psychiatric Genomics, C., Genetic Consortium for Anorexia Nervosa of the Wellcome Trust Case Control, C., Duncan, L., Perry, J. R., Patterson, N., Robinson, E. B., Daly, M. J., Price, A. L., & Neale, B. M. (2015). An atlas of genetic correlations across human diseases and traits. *Nat Genet*, *47*(11), 1236-1241. https://doi.org/10.1038/ng.3406

Bycroft, C., Freeman, C., Petkova, D., Band, G., Elliott, L. T., Sharp, K., Motyer, A., Vukcevic, D., Delaneau, O., O'Connell, J., Cortes, A., Welsh, S., Young, A., Effingham, M., McVean, G., Leslie, S., Allen, N., Donnelly, P., & Marchini, J. (2018). The UK Biobank resource with deep phenotyping and genomic data. *Nature*, *562*(7726), 203-209. https://doi.org/10.1038/s41586-018-0579-z

Calder, A. E., & Hasler, G. (2023). Towards an understanding of psychedelic-induced neuroplasticity. *Neuropsychopharmacology*, *48*(1), 104-112. https://doi.org/10.1038/s41386-022-01389-z

Calvo, E., Medina, J. T., Ornstein, K. A., Staudinger, U. M., Fried, L. P., & Keyes, K. M. (2020). Cross-country and historical variation in alcohol consumption among older men and

women: Leveraging recently harmonized survey data in 21 countries. *Drug Alcohol Depend*, *215*, 108219. https://doi.org/10.1016/j.drugalcdep.2020.108219

Carreiras, M., Quinones, I., Chen, H. A., Vazquez-Araujo, L., Small, D., & Frost, R. (2024). Sniffing out meaning: Chemosensory and semantic neural network changes in sommeliers. *Hum Brain Mapp*, *45*(2), e26564. https://doi.org/10.1002/hbm.26564

Carstens, E., & Carstens, M. I. (2022). Sensory Effects of Nicotine and Tobacco. *Nicotine Tob Res*, *24*(3), 306-315. https://doi.org/10.1093/ntr/ntab086

Carter, M., Fisher, C., & Isaac, M. (2013). Recovery From Comorbidity. *SAGE Open*, *3*(4). https://doi.org/10.1177/2158244013512133

Castillo-Carniglia, A., Keyes, K. M., Hasin, D. S., & Cerda, M. (2019). Psychiatric comorbidities in alcohol use disorder. *Lancet Psychiatry*, *6*(12), 1068-1080. https://doi.org/10.1016/S2215-0366(19)30222-6

Chen, D., Zhang, R., Zhao, H., & Feng, J. (2019). A Bibliometric Analysis of the Development of ICD-11 in Medical Informatics. *J Healthc Eng*, *2019*, 1649363. https://doi.org/10.1155/2019/1649363

Cho, H. S., & Yang, Y. (2023). Relationship Between Alcohol Consumption and Risky Sexual Behaviors Among Adolescents and Young Adults: A Meta-Analysis. *Int J Public Health*, *68*, 1605669. https://doi.org/10.3389/ijph.2023.1605669

Collaborators, G. B. D. A. (2018). Alcohol use and burden for 195 countries and territories, 1990-2016: a systematic analysis for the Global Burden of Disease Study 2016. *Lancet*, *392*(10152), 1015-1035. https://doi.org/10.1016/S0140-6736(18)31310-2

Coppens, V., Morrens, M., Destoop, M., & Dom, G. (2019). The Interplay of Inflammatory Processes and Cognition in Alcohol Use Disorders-A Systematic Review. *Front Psychiatry*, *10*, 632. https://doi.org/10.3389/fpsyt.2019.00632

Dai, Y., Hu, R., Pei, G., Zhang, H., Zhao, Z., & Jia, P. (2020). Diverse types of genomic evidence converge on alcohol use disorder risk genes. *J Med Genet*, *57*(11), 733-743. https://doi.org/10.1136/jmedgenet-2019-106490

Darwin, K. H., & Stanley, S. A. (2022). The aldehyde hypothesis: metabolic intermediates as antimicrobial effectors. *Open Biol*, *12*(4), 220010. https://doi.org/10.1098/rsob.220010

Davey Smith, G., & Hemani, G. (2014). Mendelian randomization: genetic anchors for causal inference in epidemiological studies. *Hum Mol Genet*, *23*(R1), R89-98. https://doi.org/10.1093/hmg/ddu328

Decker, A. M., Kapila, Y. L., & Wang, H. L. (2021). The psychobiological links between chronic stress-related diseases, periodontal/peri-implant diseases, and wound healing. *Periodontol 2000*, *87*(1), 94-106. https://doi.org/10.1111/prd.12381

Deiana, G., Sun, R., Huang, J., Napolioni, V., & Ciccocioppo, R. (2024). Contribution of infectious diseases to the selection of ADH1B and ALDH2 gene variants in Asian populations. *Alcohol Clin Exp Res (Hoboken)*. https://doi.org/10.1111/acer.15288

Dell'Aquila, A., & Berle, D. (2023). Predictors of alcohol and substance use among people with post-traumatic stress disorder (PTSD): findings from the NESARC-III study. *Soc Psychiatry Psychiatr Epidemiol*, *58*(10), 1509-1522. https://doi.org/10.1007/s00127-023-02472-6

Demir, S., Bulut, M., Atli, A., Kaplan, I., Kaya, M. C., Bez, Y., Ozdemir, P. G., & Sir, A. (2016). Decreased Prolidase Activity in Patients with Posttraumatic Stress Disorder. *Psychiatry Investig*, *13*(4), 420-426. https://doi.org/10.4306/pi.2016.13.4.420

Dubey, I., Georgescu, A. L., Hommelsen, M., Vogeley, K., Ropar, D., & Hamilton, A. F. C. (2020). Distinct neural correlates of social and object reward seeking motivation. *Eur J Neurosci*, *52*(9), 4214-4229. https://doi.org/10.1111/ejn.14888

Dworkin, E. R., Wanklyn, S., Stasiewicz, P. R., & Coffey, S. F. (2018). PTSD symptom presentation among people with alcohol and drug use disorders: Comparisons by substance of abuse. *Addict Behav*, *76*, 188-194. https://doi.org/10.1016/j.addbeh.2017.08.019

Edmondson, D., & von Kanel, R. (2017). Post-traumatic stress disorder and cardiovascular disease. *Lancet Psychiatry*, *4*(4), 320-329. https://doi.org/10.1016/S2215-0366(16)30377-7

Ercan, A. C., Bahceci, B., Polat, S., Cenker, O. C., Bahceci, I., Koroglu, A., Sahin, K., & Hocaoglu, C. (2017). Oxidative status and prolidase activities in generalized anxiety disorder. *Asian J Psychiatr*, *25*, 118-122. https://doi.org/10.1016/j.ajp.2016.10.008

Erickson, E. K., Grantham, E. K., Warden, A. S., & Harris, R. A. (2019). Neuroimmune signaling in alcohol use disorder. *Pharmacol Biochem Behav*, *177*, 34-60. https://doi.org/10.1016/j.pbb.2018.12.007

Gabriel, D. B. K., Freels, T. G., Setlow, B., & Simon, N. W. (2019). Risky decision-making is associated with impulsive action and sensitivity to first-time nicotine exposure. *Behav Brain Res*, *359*, 579-588. https://doi.org/10.1016/j.bbr.2018.10.008

Gallus, S., Muttarak, R., Franchi, M., Pacifici, R., Colombo, P., Boffetta, P., Leon, M. E., & La Vecchia, C. (2013). Why do smokers quit? *Eur J Cancer Prev*, *22*(1), 96-101. https://doi.org/10.1097/CEJ.0b013e3283552da8

Gargano, M. A., Matentzoglu, N., Coleman, B., Addo-Lartey, E. B., Anagnostopoulos, A. V., Anderton, J., Avillach, P., Bagley, A. M., Bakstein, E., Balhoff, J. P., Baynam, G., Bello, S. M., Berk, M., Bertram, H., Bishop, S., Blau, H., Bodenstein, D. F., Botas, P., Boztug, K., . . . Robinson, P. N. (2024). The Human Phenotype Ontology in 2024: phenotypes around the world. *Nucleic Acids Res*, *52*(D1), D1333-D1346. https://doi.org/10.1093/nar/gkad1005

Genomes Project, C., Auton, A., Brooks, L. D., Durbin, R. M., Garrison, E. P., Kang, H. M., Korbel, J. O., Marchini, J. L., McCarthy, S., McVean, G. A., & Abecasis, G. R. (2015). A global reference for human genetic variation. *Nature*, *526*(7571), 68-74. https://doi.org/10.1038/nature15393

Girgenti, M. J., & Duman, R. S. (2018). Transcriptome Alterations in Posttraumatic Stress Disorder. *Biol Psychiatry*, *83*(10), 840-848. https://doi.org/10.1016/j.biopsych.2017.09.023

Glatt, S. J., Tylee, D. S., Chandler, S. D., Pazol, J., Nievergelt, C. M., Woelk, C. H., Baker, D. G., Lohr, J. B., Kremen, W. S., Litz, B. T., Tsuang, M. T., & Marine Resiliency Study, I. (2013). Blood-based gene-expression predictors of PTSD risk and resilience among deployed marines: a pilot study. *Am J Med Genet B Neuropsychiatr Genet*, *162B*(4), 313-326. https://doi.org/10.1002/ajmg.b.32167

Gordon, E. M., Chauvin, R. J., Van, A. N., Rajesh, A., Nielsen, A., Newbold, D. J., Lynch, C. J., Seider, N. A., Krimmel, S. R., Scheidter, K. M., Monk, J., Miller, R. L., Metoki, A., Montez, D. F., Zheng, A., Elbau, I., Madison, T., Nishino, T., Myers, M. J., . . . Dosenbach, N. U. F. (2023). A somato-cognitive action network alternates with effector regions in motor cortex. *Nature*, *617*(7960), 351-359. https://doi.org/10.1038/s41586-023-05964-2

Gorka, S. M. (2020). Interpersonal trauma exposure and startle reactivity to uncertain threat in individuals with alcohol use disorder. *Drug Alcohol Depend*, *206*, 107727. https://doi.org/10.1016/j.drugalcdep.2019.107727

Govindula, A., Ranadive, N., Nampoothiri, M., Rao, C. M., Arora, D., & Mudgal, J. (2023). Emphasizing the Crosstalk Between Inflammatory and Neural Signaling in Post-

traumatic Stress Disorder (PTSD). *J Neuroimmune Pharmacol*, *18*(3), 248-266. https://doi.org/10.1007/s11481-023-10064-z

Grant, B. F., Goldstein, R. B., Saha, T. D., Chou, S. P., Jung, J., Zhang, H., Pickering, R. P., Ruan, W. J., Smith, S. M., Huang, B., & Hasin, D. S. (2015). Epidemiology of DSM-5 Alcohol Use Disorder: Results From the National Epidemiologic Survey on Alcohol and Related Conditions III. *JAMA Psychiatry*, *72*(8), 757-766. https://doi.org/10.1001/jamapsychiatry.2015.0584

Hamada, K., & Lasek, A. W. (2020). Receptor Tyrosine Kinases as Therapeutic Targets for Alcohol Use Disorder. *Neurotherapeutics*, *17*(1), 4-16. https://doi.org/10.1007/s13311-019-00795-4

He, H., Pan, Z., Wu, J., Hu, C., Bai, L., & Lyu, J. (2022). Health Effects of Tobacco at the Global, Regional, and National Levels: Results From the 2019 Global Burden of Disease Study. *Nicotine Tob Res*, *24*(6), 864-870. https://doi.org/10.1093/ntr/ntab265

He, J., Cabrera-Mendoza, B., De Angelis, F., Pathak, G. A., Koller, D., Curhan, S. G., Curhan, G. C., Mecca, A. P., van Dyck, C. H., & Polimanti, R. (2024). Sex differences in the pleiotropy of hearing difficulty with imaging-derived phenotypes: a brain-wide investigation. *Brain*. https://doi.org/10.1093/brain/awae077

Heather, J. M., & Chain, B. (2016). The sequence of sequencers: The history of sequencing DNA. *Genomics*, *107*(1), 1-8. https://doi.org/10.1016/j.ygeno.2015.11.003

Heinz, A., & Daedelow, L. S. (2021). [Alcohol as cultural asset-a historical, anthropological, and therapeutic perspective on alcohol consumption and its social role in western societies]. *Bundesgesundheitsblatt Gesundheitsforschung Gesundheitsschutz*, *64*(6), 646-651. https://doi.org/10.1007/s00103-021-03327-8 (Alkohol als Kulturgut - eine historisch-anthropologische und therapeutische Perspektive auf Alkoholkonsum und seine soziale Rolle in westlichen Gesellschaften.)

Hemani, G., Zheng, J., Elsworth, B., Wade, K. H., Haberland, V., Baird, D., Laurin, C., Burgess, S., Bowden, J., Langdon, R., Tan, V. Y., Yarmolinsky, J., Shihab, H. A., Timpson, N. J., Evans, D. M., Relton, C., Martin, R. M., Davey Smith, G., Gaunt, T. R., & Haycock, P. C. (2018). The MR-Base platform supports systematic causal inference across the human phenome. *Elife*, *7*. https://doi.org/10.7554/eLife.34408

Hendershot, C. S., Neighbors, C., George, W. H., McCarthy, D. M., Wall, T. L., Liang, T., & Larimer, M. E. (2009). ALDH2, ADH1B and alcohol expectancies: integrating genetic and learning perspectives. *Psychol Addict Behav*, *23*(3), 452-463. https://doi.org/10.1037/a0016629

Hertrich, I., Dietrich, S., Blum, C., & Ackermann, H. (2021). The Role of the Dorsolateral Prefrontal Cortex for Speech and Language Processing. *Front Hum Neurosci*, *15*, 645209. https://doi.org/10.3389/fnhum.2021.645209

Hill, L. J., Di Pietro, V., Hazeldine, J., Davies, D., Toman, E., Logan, A., & Belli, A. (2017). Cystatin D (CST5): An ultra-early inflammatory biomarker of traumatic brain injury. *Sci Rep*, *7*(1), 5002. https://doi.org/10.1038/s41598-017-04722-5

Holmes, A. J., & Anderson, K. (2017). Convergence in National Alcohol Consumption Patterns: New Global Indicators. *Journal of Wine Economics*, *12*(2), 117-148. https://doi.org/10.1017/jwe.2017.15

Holzhauer, C. G., Cucciare, M., & Epstein, E. E. (2020). Sex and Gender Effects in Recovery From Alcohol Use Disorder. *Alcohol Res*, *40*(3), 03. https://doi.org/10.35946/arcr.v40.3.03

Icick, R., Shadrin, A., Holen, B., Karadag, N., Parker, N., O'Connell, K., Frei, O., Bahrami, S., Hoegh, M., Lagerberg, T., Cheng, W., Seibert, T., Djurovic, S., Dale, A., Zhou, H., Edenberg, H., Gelernter, J., Smeland, O., Hindley, G., & Andreassen, O. (2023). Identification of Novel Loci and Cross-Disorder Pleiotropy Through Multi-Ancestry Genome-Wide Analysis of Alcohol Use Disorder in Over One Million Individuals. *Res Sq*. https://doi.org/10.21203/rs.3.rs-3755915/v1

International HapMap, C. (2003). The International HapMap Project. *Nature*, *426*(6968), 789-796. https://doi.org/10.1038/nature02168

Jaffe, A. E., Tao, R., Page, S. C., Maynard, K. R., Pattie, E. A., Nguyen, C. V., Deep-Soboslay, A., Bharadwaj, R., Young, K. A., Friedman, M. J., Williamson, D. E., Traumatic Stress Brain Research, G., Shin, J. H., Hyde, T. M., Martinowich, K., & Kleinman, J. E. (2022). Decoding Shared Versus Divergent Transcriptomic Signatures Across Cortico-Amygdala Circuitry in PTSD and Depressive Disorders. *Am J Psychiatry*, *179*(9), 673-686. https://doi.org/10.1176/appi.ajp.21020162

Janes, A. C., Park, M. T., Farmer, S., & Chakravarty, M. M. (2015). Striatal morphology is associated with tobacco cigarette craving. *Neuropsychopharmacology*, *40*(2), 406-411. https://doi.org/10.1038/npp.2014.185

Jennings, M. V., Martinez-Magana, J. J., Courchesne-Krak, N. S., Cupertino, R. B., Vilar-Ribo, L., Bianchi, S. B., Hatoum, A. S., Atkinson, E. G., Giusti-Rodriguez, P., Montalvo-Ortiz, J. L., Gelernter, J., Artigas, M. S., andMe, I. R. T., Elson, S. L., Edenberg, H. J., Fontanillas, P., Palmer, A. A., & Sanchez-Roige, S. (2024). A phenome-wide association and Mendelian randomisation study of alcohol use variants in a diverse cohort comprising over 3 million individuals. *EBioMedicine*, 105086. https://doi.org/10.1016/j.ebiom.2024.105086

Kanehisa, M., & Goto, S. (2000). KEGG: kyoto encyclopedia of genes and genomes. *Nucleic Acids Res*, *28*(1), 27-30. https://doi.org/10.1093/nar/28.1.27

Kapoor, M., Wang, J. C., Farris, S. P., Liu, Y., McClintick, J., Gupta, I., Meyers, J. L., Bertelsen, S., Chao, M., Nurnberger, J., Tischfield, J., Harari, O., Zeran, L., Hesselbrock, V., Bauer, L., Raj, T., Porjesz, B., Agrawal, A., Foroud, T., . . . Goate, A. (2019). Analysis of whole genome-transcriptomic organization in brain to identify genes associated with alcoholism. *Transl Psychiatry*, *9*(1), 89. https://doi.org/10.1038/s41398-019-0384-y

Kazmi, N., Wallen, G. R., Yang, L., Alkhatib, J., Schwandt, M. L., Feng, D., Gao, B., Diazgranados, N., Ramchandani, V. A., & Barb, J. J. (2022). An exploratory study of pro-inflammatory cytokines in individuals with alcohol use disorder: MCP-1 and IL-8 associated with alcohol consumption, sleep quality, anxiety, depression, and liver biomarkers. *Front Psychiatry*, *13*, 931280. https://doi.org/10.3389/fpsyt.2022.931280

Kember, R. L., Hartwell, E. E., Xu, H., Rotenberg, J., Almasy, L., Zhou, H., Gelernter, J., & Kranzler, H. R. (2023). Phenome-wide Association Analysis of Substance Use Disorders in a Deeply Phenotyped Sample. *Biol Psychiatry*, *93*(6), 536-545. https://doi.org/10.1016/j.biopsych.2022.08.010

Keyes, K. M., Jager, J., Mal-Sarkar, T., Patrick, M. E., Rutherford, C., & Hasin, D. (2019). Is There a Recent Epidemic of Women's Drinking? A Critical Review of National Studies. *Alcohol Clin Exp Res*, *43*(7), 1344-1359. https://doi.org/10.1111/acer.14082

Knox, D., Della Valle, R., Mohammadmirzaei, N., Shultz, B., Biddle, M., Farkash, A., Chamness, M., & Moulton, E. (2021). PI3K-Akt Signaling in the Basolateral Amygdala Facilitates Traumatic Stress Enhancements in Fear Memory. *Int J Neuropsychopharmacol*, *24*(3), 229-238. https://doi.org/10.1093/ijnp/pyaa083

Koller, D., Friligkou, E., Stiltner, B., Pathak, G. A., Lokhammer, S., Levey, D. F., Zhou, H., Hatoum, A. S., Deak, J. D., Kember, R. L., Treur, J. L., Kranzler, H. R., Johnson, E. C., Stein, M. B., Gelernter, J., & Polimanti, R. (2024). Pleiotropy and genetically inferred causality linking

multisite chronic pain to substance use disorders. *Mol Psychiatry*. https://doi.org/10.1038/s41380-024-02446-3

Kutlu, M. G., Cole, R. D., Connor, D. A., Natwora, B., & Gould, T. J. (2018). Tyrosine receptor kinase B receptor activation reverses the impairing effects of acute nicotine on contextual fear extinction. *J Psychopharmacol*, *32*(3), 367-372. https://doi.org/10.1177/0269881118758305

Laing, P. A. F., Steward, T., Davey, C. G., Felmingham, K. L., Fullana, M. A., Vervliet, B., Greaves, M. D., Moffat, B., Glarin, R. K., & Harrison, B. J. (2022). Cortico-Striatal Activity Characterizes Human Safety Learning via Pavlovian Conditioned Inhibition. *J Neurosci*, *42*(25), 5047-5057. https://doi.org/10.1523/JNEUROSCI.2181-21.2022

Lampe, L., Zhang, R., Beyer, F., Huhn, S., Kharabian Masouleh, S., Preusser, S., Bazin, P. L., Schroeter, M. L., Villringer, A., & Witte, A. V. (2019). Visceral obesity relates to deep white matter hyperintensities via inflammation. *Ann Neurol*, *85*(2), 194-203. https://doi.org/10.1002/ana.25396

Lanquetin, A., Leclercq, S., de Timary, P., Segobin, S., Naveau, M., Coulbault, L., Maccioni, P., Lorrai, I., Colombo, G., Vivien, D., Rubio, M., & Pitel, A. L. (2021). Role of inflammation in alcohol-related brain abnormalities: a translational study. *Brain Commun*, *3*(3), fcab154. https://doi.org/10.1093/braincomms/fcab154

Lauten, T. H., Natour, T., & Case, A. J. (2024). Innate and adaptive immune system consequences of post-traumatic stress disorder. *Auton Neurosci*, *252*, 103159. https://doi.org/10.1016/j.autneu.2024.103159

Lavin, C., Melis, C., Mikulan, E., Gelormini, C., Huepe, D., & Ibanez, A. (2013). The anterior cingulate cortex: an integrative hub for human socially-driven interactions. *Front Neurosci*, *7*, 64. https://doi.org/10.3389/fnins.2013.00064

Le, L. K., Faller, J., Chatterton, M. L., Perez, J. K., Chiotelis, O., Tran, H. N. Q., Sultana, M., Hall, N., Lee, Y. Y., Chapman, C., Newton, N., Slade, T., Sunderland, M., Teesson, M., & Mihalopoulos, C. (2023). Interventions to prevent alcohol use: systematic review of economic evaluations. *BJPsych Open*, *9*(4), e117. https://doi.org/10.1192/bjo.2023.81

Lewis, S. J., Arseneault, L., Caspi, A., Fisher, H. L., Matthews, T., Moffitt, T. E., Odgers, C. L., Stahl, D., Teng, J. Y., & Danese, A. (2019). The epidemiology of trauma and post-traumatic stress disorder in a representative cohort of young people in England and

Wales. *Lancet Psychiatry*, *6*(3), 247-256. https://doi.org/10.1016/S2215-0366(19)30031-8

Li, D., Musante, V., Zhou, W., Picciotto, M. R., & Nairn, A. C. (2018). Striatin-1 is a B subunit of protein phosphatase PP2A that regulates dendritic arborization and spine development in striatal neurons. *J Biol Chem*, *293*(28), 11179-11194. https://doi.org/10.1074/jbc.RA117.001519

Li, Q., Wineinger, N. E., Fu, D. J., Libiger, O., Alphs, L., Savitz, A., Gopal, S., Cohen, N., & Schork, N. J. (2017). Genome-wide association study of paliperidone efficacy. *Pharmacogenet Genomics*, *27*(1), 7-18. https://doi.org/10.1097/FPC.0000000000000250

Li, W., Lopez, L., Osher, J., Howard, J. D., Parrish, T. B., & Gottfried, J. A. (2010). Right orbitofrontal cortex mediates conscious olfactory perception. *Psychol Sci*, *21*(10), 1454-1463. https://doi.org/10.1177/0956797610382121

Lim, M., Jassar, H., Kim, D. J., Nascimento, T. D., & DaSilva, A. F. (2021). Differential alteration of fMRI signal variability in the ascending trigeminal somatosensory and pain modulatory pathways in migraine. *J Headache Pain*, *22*(1), 4. https://doi.org/10.1186/s10194-020-01210-6

Lin, W., Zhu, L., & Lu, Y. (2023). Association of smoking with brain gray and white matter volume: a Mendelian randomization study. *Neurol Sci*, *44*(11), 4049-4055. https://doi.org/10.1007/s10072-023-06854-1

Liu, M., Jiang, Y., Wedow, R., Li, Y., Brazel, D. M., Chen, F., Datta, G., Davila-Velderrain, J., McGuire, D., Tian, C., Zhan, X., andMe Research, T., Psychiatry, H. A.-I., Choquet, H., Docherty, A. R., Faul, J. D., Foerster, J. R., Fritsche, L. G., Gabrielsen, M. E., . . . Vrieze, S. (2019). Association studies of up to 1.2 million individuals yield new insights into the genetic etiology of tobacco and alcohol use. *Nat Genet*, *51*(2), 237-244. https://doi.org/10.1038/s41588-018-0307-5

Liu, X., Lu, D., Saw, W. Y., Shaw, P. J., Wangkumhang, P., Ngamphiw, C., Fucharoen, S., Lert-Itthiporn, W., Chin-Inmanu, K., Chau, T. N., Anders, K., Kasturiratne, A., de Silva, H. J., Katsuya, T., Kimura, R., Nabika, T., Ohkubo, T., Tabara, Y., Takeuchi, F., . . . Teo, Y. Y. (2017). Characterising private and shared signatures of positive selection in 37 Asian populations. *Eur J Hum Genet*, *25*(4), 499-508. https://doi.org/10.1038/ejhg.2016.181

Logue, M. W., Zhou, Z., Morrison, F. G., Wolf, E. J., Daskalakis, N. P., Chatzinakos, C., Georgiadis, F., Labadorf, A. T., Girgenti, M. J., Young, K. A., Williamson, D. E., Zhao, X., Grenier, J. G.,

Traumatic Stress Brain Research, G., Huber, B. R., & Miller, M. W. (2021). Gene expression in the dorsolateral and ventromedial prefrontal cortices implicates immune-related gene networks in PTSD. *Neurobiol Stress*, *15*, 100398. https://doi.org/10.1016/j.ynstr.2021.100398

Lotsch, J., Hahner, A., Gossrau, G., Hummel, C., Walter, C., Ultsch, A., & Hummel, T. (2016). Smell of pain: intersection of nociception and olfaction. *Pain*, *157*(10), 2152-2157. https://doi.org/10.1097/j.pain.0000000000000599

Lundin, A., Waern, M., Love, J., Lovestad, S., Hensing, G., & Danielsson, A. K. (2021). Towards ICD-11 for alcohol dependence: Diagnostic agreement with ICD-10, DSM-5, DSM-IV, DSM-III-R and DSM-III diagnoses in a Swedish general population of women. *Drug Alcohol Depend*, *227*, 108925. https://doi.org/10.1016/j.drugalcdep.2021.108925

Mailman, M. D., Feolo, M., Jin, Y., Kimura, M., Tryka, K., Bagoutdinov, R., Hao, L., Kiang, A., Paschall, J., Phan, L., Popova, N., Pretel, S., Ziyabari, L., Lee, M., Shao, Y., Wang, Z. Y., Sirotkin, K., Ward, M., Kholodov, M., . . . Sherry, S. T. (2007). The NCBI dbGaP database of genotypes and phenotypes. *Nat Genet*, *39*(10), 1181-1186. https://doi.org/10.1038/ng1007-1181

Manning, B. D., & Toker, A. (2017). AKT/PKB Signaling: Navigating the Network. *Cell*, *169*(3), 381-405. https://doi.org/10.1016/j.cell.2017.04.001

Manthey, J., Hassan, S. A., Carr, S., Kilian, C., Kuitunen-Paul, S., & Rehm, J. (2021). What are the Economic Costs to Society Attributable to Alcohol Use? A Systematic Review and Modelling Study. *Pharmacoeconomics*, *39*(7), 809-822. https://doi.org/10.1007/s40273-021-01031-8

Manthey, J., Shield, K. D., Rylett, M., Hasan, O. S. M., Probst, C., & Rehm, J. (2019). Global alcohol exposure between 1990 and 2017 and forecasts until 2030: a modelling study. *Lancet*, *393*(10190), 2493-2502. https://doi.org/10.1016/S0140-6736(18)32744-2

Maria-Rios, C. E., & Morrow, J. D. (2020). Mechanisms of Shared Vulnerability to Post-traumatic Stress Disorder and Substance Use Disorders. *Front Behav Neurosci*, *14*, 6. https://doi.org/10.3389/fnbeh.2020.00006

Marshall, G. N., Jaycox, L. H., Engel, C. C., Richardson, A. S., Dutra, S. J., Keane, T. M., Rosen, R. C., & Marx, B. P. (2019). PTSD symptoms are differentially associated with general distress and physiological arousal: Implications for the conceptualization and

measurement of PTSD. *J Anxiety Disord*, *62*, 26-34. https://doi.org/10.1016/j.janxdis.2018.10.003

Mavromatis, L. A., Rosoff, D. B., Cupertino, R. B., Garavan, H., Mackey, S., & Lohoff, F. W. (2022). Association Between Brain Structure and Alcohol Use Behaviors in Adults: A Mendelian Randomization and Multiomics Study. *JAMA Psychiatry*, *79*(9), 869-878. https://doi.org/10.1001/jamapsychiatry.2022.2196

McCarthy, M. I., Abecasis, G. R., Cardon, L. R., Goldstein, D. B., Little, J., Ioannidis, J. P., & Hirschhorn, J. N. (2008). Genome-wide association studies for complex traits: consensus, uncertainty and challenges. *Nat Rev Genet*, *9*(5), 356-369. https://doi.org/10.1038/nrg2344

Mevorach, C., Hodsoll, J., Allen, H., Shalev, L., & Humphreys, G. (2010). Ignoring the elephant in the room: a neural circuit to downregulate salience. *J Neurosci*, *30*(17), 6072-6079. https://doi.org/10.1523/JNEUROSCI.0241-10.2010

Mignot, C., Faria, V., Hummel, T., Frost, M., Michel, C. M., Gossrau, G., & Haehner, A. (2023). Migraine with aura: less control over pain and fragrances? *J Headache Pain*, *24*(1), 55. https://doi.org/10.1186/s10194-023-01592-3

Milacic, M., Beavers, D., Conley, P., Gong, C., Gillespie, M., Griss, J., Haw, R., Jassal, B., Matthews, L., May, B., Petryszak, R., Ragueneau, E., Rothfels, K., Sevilla, C., Shamovsky, V., Stephan, R., Tiwari, K., Varusai, T., Weiser, J., . . . D'Eustachio, P. (2024). The Reactome Pathway Knowledgebase 2024. *Nucleic Acids Res*, *52*(D1), D672-D678. https://doi.org/10.1093/nar/gkad1025

Mondino, M., Luck, D., Grot, S., Januel, D., Suaud-Chagny, M. F., Poulet, E., & Brunelin, J. (2018). Effects of repeated transcranial direct current stimulation on smoking, craving and brain reactivity to smoking cues. *Sci Rep*, *8*(1), 8724. https://doi.org/10.1038/s41598-018-27057-1

Morel, C., Montgomery, S., & Han, M. H. (2019). Nicotine and alcohol: the role of midbrain dopaminergic neurons in drug reinforcement. *Eur J Neurosci*, *50*(3), 2180-2200. https://doi.org/10.1111/ejn.14160

Morojele, N. K., Shenoi, S. V., Shuper, P. A., Braithwaite, R. S., & Rehm, J. (2021). Alcohol Use and the Risk of Communicable Diseases. *Nutrients*, *13*(10). https://doi.org/10.3390/nu13103317

Moura, H. F., Hansen, F., Galland, F., Silvelo, D., Rebelatto, F. P., Ornell, F., Massuda, R., Scherer, J. N., Schuch, F., Kessler, F. H., & von Diemen, L. (2022). Inflammatory cytokines and alcohol use disorder: systematic review and meta-analysis. *Braz J Psychiatry*, *44*(5), 548-556. https://doi.org/10.47626/1516-4446-2021-1893

Muhie, S., Gautam, A., Yang, R., Misganaw, B., Daigle, B. J., Jr., Mellon, S. H., Flory, J. D., Abu-Amara, D., Lee, I., Wang, K., Rampersaud, R., Consortium, P. S. B., Hood, L., Yehuda, R., Marmar, C. R., Wolkowitz, O. M., Ressler, K. J., Doyle, F. J., 3rd, Hammamieh, R., & Jett, M. (2023). Molecular signatures of post-traumatic stress disorder in war-zone-exposed veteran and active-duty soldiers. *Cell Rep Med*, *4*(5), 101045. https://doi.org/10.1016/j.xcrm.2023.101045

Naito, E., Morita, T., & Amemiya, K. (2016). Body representations in the human brain revealed by kinesthetic illusions and their essential contributions to motor control and corporeal awareness. *Neurosci Res*, *104*, 16-30. https://doi.org/10.1016/j.neures.2015.10.013

Nakamura, M., Nestor, P. G., & Shenton, M. E. (2020). Orbitofrontal Sulcogyral Pattern as a Transdiagnostic Trait Marker of Early Neurodevelopment in the Social Brain. *Clin EEG Neurosci*, *51*(4), 275-284. https://doi.org/10.1177/1550059420904180

Nejati, V., Salehinejad, M. A., & Nitsche, M. A. (2018). Interaction of the Left Dorsolateral Prefrontal Cortex (l-DLPFC) and Right Orbitofrontal Cortex (OFC) in Hot and Cold Executive Functions: Evidence from Transcranial Direct Current Stimulation (tDCS). *Neuroscience*, *369*, 109-123. https://doi.org/10.1016/j.neuroscience.2017.10.042

Nunez-Rios, D. L., Martinez-Magana, J. J., Nagamatsu, S. T., Krystal, J. H., Martinez-Gonzalez, K. G., Giusti-Rodriguez, P., & Montalvo-Ortiz, J. L. (2023). Cross-Species Convergence of Brain Transcriptomic and Epigenomic Findings in Posttraumatic Stress Disorder: A Systematic Review. *Complex Psychiatry*, *9*(1-4), 100-118. https://doi.org/10.1159/000529536

O'Connor, L. J., & Price, A. L. (2018). Distinguishing genetic correlation from causation across 52 diseases and complex traits. *Nat Genet*, *50*(12), 1728-1734. https://doi.org/10.1038/s41588-018-0255-0

Obi, I. E., McPherson, K. C., & Pollock, J. S. (2019). Childhood adversity and mechanistic links to hypertension risk in adulthood. *Br J Pharmacol*, *176*(12), 1932-1950. https://doi.org/10.1111/bph.14576

Organization, W. H. (2019). 6B40 Post traumatic stress disorder. In *nternational statistical classification of diseases and related health problems (11th ed.)*. https://icd.who.int/browse/2024-01/mms/en#2070699808

Oughtred, R., Rust, J., Chang, C., Breitkreutz, B. J., Stark, C., Willems, A., Boucher, L., Leung, G., Kolas, N., Zhang, F., Dolma, S., Coulombe-Huntington, J., Chatr-Aryamontri, A., Dolinski, K., & Tyers, M. (2021). The BioGRID database: A comprehensive biomedical resource of curated protein, genetic, and chemical interactions. *Protein Sci*, *30*(1), 187-200. https://doi.org/10.1002/pro.3978

Pacella, M. L., Hruska, B., & Delahanty, D. L. (2013). The physical health consequences of PTSD and PTSD symptoms: a meta-analytic review. *J Anxiety Disord*, *27*(1), 33-46. https://doi.org/10.1016/j.janxdis.2012.08.004

Passos, I. C., Vasconcelos-Moreno, M. P., Costa, L. G., Kunz, M., Brietzke, E., Quevedo, J., Salum, G., Magalhaes, P. V., Kapczinski, F., & Kauer-Sant'Anna, M. (2015). Inflammatory markers in post-traumatic stress disorder: a systematic review, meta-analysis, and meta-regression. *Lancet Psychiatry*, *2*(11), 1002-1012. https://doi.org/10.1016/S2215-0366(15)00309-0

Pathak, G. A., Singh, K., Choi, K. W., Fang, Y., Kouakou, M. R., Lee, Y. H., Zhou, X., Fritsche, L. G., Wendt, F. R., Davis, L. K., & Polimanti, R. (2024). Genetic Liability to Posttraumatic Stress Disorder Symptoms and Its Association With Cardiometabolic and Respiratory Outcomes. *JAMA Psychiatry*, *81*(1), 34-44. https://doi.org/10.1001/jamapsychiatry.2023.4127

Peng, G. S., & Yin, S. J. (2009). Effect of the allelic variants of aldehyde dehydrogenase ALDH2*2 and alcohol dehydrogenase ADH1B*2 on blood acetaldehyde concentrations. *Hum Genomics*, *3*(2), 121-127. https://doi.org/10.1186/1479-7364-3-2-121

Poisson, C. L., Engel, L., & Saunders, B. T. (2021). Dopamine Circuit Mechanisms of Addiction-Like Behaviors. *Front Neural Circuits*, *15*, 752420. https://doi.org/10.3389/fncir.2021.752420

Polimanti, R., Amstadter, A. B., Stein, M. B., Almli, L. M., Baker, D. G., Bierut, L. J., Bradley, B., Farrer, L. A., Johnson, E. O., King, A., Kranzler, H. R., Maihofer, A. X., Rice, J. P., Roberts, A. L., Saccone, N. L., Zhao, H., Liberzon, I., Ressler, K. J., Nievergelt, C. M., . . . Psychiatric Genomics Consortium Posttraumatic Stress Disorder, W. (2017). A putative causal

relationship between genetically determined female body shape and posttraumatic stress disorder. *Genome Med*, *9*(1), 99. https://doi.org/10.1186/s13073-017-0491-4

Polimanti, R., & Gelernter, J. (2018). ADH1B: From alcoholism, natural selection, and cancer to the human phenome. *Am J Med Genet B Neuropsychiatr Genet*, *177*(2), 113-125. https://doi.org/10.1002/ajmg.b.32523

Power, M. C., Deal, J. A., Sharrett, A. R., Jack, C. R., Jr., Knopman, D., Mosley, T. H., & Gottesman, R. F. (2015). Smoking and white matter hyperintensity progression: the ARIC-MRI Study. *Neurology*, *84*(8), 841-848. https://doi.org/10.1212/WNL.0000000000001283

Pradeep, P., Kang, H., & Lee, B. (2023). Glycosylation and behavioral symptoms in neurological disorders. *Transl Psychiatry*, *13*(1), 154. https://doi.org/10.1038/s41398-023-02446-x

Puddephatt, J. A., Irizar, P., Jones, A., Gage, S. H., & Goodwin, L. (2022). Associations of common mental disorder with alcohol use in the adult general population: a systematic review and meta-analysis. *Addiction*, *117*(6), 1543-1572. https://doi.org/10.1111/add.15735

Rabinowitz, J. A., Campos, A. I., Ong, J. S., Garcia-Marin, L. M., Alcauter, S., Mitchell, B. L., Grasby, K. L., Cuellar-Partida, G., Gillespie, N. A., Huhn, A. S., Martin, N. G., Thompson, P. M., Medland, S. E., Maher, B. S., & Renteria, M. E. (2022). Shared Genetic Etiology between Cortical Brain Morphology and Tobacco, Alcohol, and Cannabis Use. *Cereb Cortex*, *32*(4), 796-807. https://doi.org/10.1093/cercor/bhab243

Rahman, S., Zammit, S., Dalman, C., & Hollander, A. C. (2022). Epidemiology of posttraumatic stress disorder: A prospective cohort study based on multiple nationwide Swedish registers of 4.6 million people. *Eur Psychiatry*, *65*(1), e60. https://doi.org/10.1192/j.eurpsy.2022.2311

Rehm, J., Heilig, M., & Gual, A. (2019). ICD-11 for Alcohol Use Disorders: Not a Convincing Answer to the Challenges. *Alcohol Clin Exp Res*, *43*(11), 2296-2300. https://doi.org/10.1111/acer.14182

Rumgay, H., Shield, K., Charvat, H., Ferrari, P., Sornpaisarn, B., Obot, I., Islami, F., Lemmens, V., Rehm, J., & Soerjomataram, I. (2021). Global burden of cancer in 2020 attributable to alcohol consumption: a population-based study. *Lancet Oncol*, *22*(8), 1071-1080. https://doi.org/10.1016/S1470-2045(21)00279-5

Sagar, V., Shanahan, L. K., Zelano, C. M., Gottfried, J. A., & Kahnt, T. (2023). High-precision mapping reveals the structure of odor coding in the human brain. *Nat Neurosci*, *26*(9), 1595-1602. https://doi.org/10.1038/s41593-023-01414-4

Sasaki, F., Kanmura, S., Oda, K., Maeda, H., Kabayama, M., Iwaya, H., Komaki, Y., Arima, S., Tanoue, S., Hashimoto, S., Fujita, H., & Ido, A. (2021). Acetaldehyde breath test as a cancer risk marker in patients with esophageal and hypopharyngeal squamous cell carcinoma. *PLoS One*, *16*(5), e0251457. https://doi.org/10.1371/journal.pone.0251457

Saunders, G. R. B., Wang, X., Chen, F., Jang, S. K., Liu, M., Wang, C., Gao, S., Jiang, Y., Khunsriraksakul, C., Otto, J. M., Addison, C., Akiyama, M., Albert, C. M., Aliev, F., Alonso, A., Arnett, D. K., Ashley-Koch, A. E., Ashrani, A. A., Barnes, K. C., . . . Vrieze, S. (2022). Genetic diversity fuels gene discovery for tobacco and alcohol use. *Nature*, *612*(7941), 720-724. https://doi.org/10.1038/s41586-022-05477-4

Saunders, J. B., Degenhardt, L., Reed, G. M., & Poznyak, V. (2019). Alcohol Use Disorders in ICD-11: Past, Present, and Future. *Alcohol Clin Exp Res*, *43*(8), 1617-1631. https://doi.org/10.1111/acer.14128

Schein, J., Houle, C., Urganus, A., Cloutier, M., Patterson-Lomba, O., Wang, Y., King, S., Levinson, W., Guerin, A., Lefebvre, P., & Davis, L. L. (2021). Prevalence of post-traumatic stress disorder in the United States: a systematic literature review. *Curr Med Res Opin*, *37*(12), 2151-2161. https://doi.org/10.1080/03007995.2021.1978417

Schimmenti, A., Billieux, J., Santoro, G., Casale, S., & Starcevic, V. (2022). A trauma model of substance use: Elaboration and preliminary validation. *Addict Behav*, *134*, 107431. https://doi.org/10.1016/j.addbeh.2022.107431

Schroder, M. S., Culhane, A. C., Quackenbush, J., & Haibe-Kains, B. (2011). survcomp: an R/Bioconductor package for performance assessment and comparison of survival models. *Bioinformatics*, *27*(22), 3206-3208. https://doi.org/10.1093/bioinformatics/btr511

Schuckit, M. A. (2014). A brief history of research on the genetics of alcohol and other drug use disorders. *Journal of Studies on Alcohol and Drugs, Supplement*(s17), 59-67.

Segura, B., Baggio, H. C., Solana, E., Palacios, E. M., Vendrell, P., Bargallo, N., & Junque, C. (2013). Neuroanatomical correlates of olfactory loss in normal aged subjects. *Behav Brain Res*, *246*, 148-153. https://doi.org/10.1016/j.bbr.2013.02.025

Shao, L. X., Liao, C., Gregg, I., Davoudian, P. A., Savalia, N. K., Delagarza, K., & Kwan, A. C. (2021). Psilocybin induces rapid and persistent growth of dendritic spines in frontal cortex in vivo. *Neuron*, *109*(16), 2535-2544 e2534. https://doi.org/10.1016/j.neuron.2021.06.008

Sheerin, C. M., Bountress, K. E., Meyers, J. L., Saenz de Viteri, S. S., Shen, H., Maihofer, A. X., Duncan, L. E., & Amstadter, A. B. (2020). Shared molecular genetic risk of alcohol dependence and posttraumatic stress disorder (PTSD). *Psychol Addict Behav*, *34*(5), 613-619. https://doi.org/10.1037/adb0000568

Shield, K., Manthey, J., Rylett, M., Probst, C., Wettlaufer, A., Parry, C. D. H., & Rehm, J. (2020). National, regional, and global burdens of disease from 2000 to 2016 attributable to alcohol use: a comparative risk assessment study. *Lancet Public Health*, *5*(1), e51-e61. https://doi.org/10.1016/S2468-2667(19)30231-2

Slatkin, M. (2008). Linkage disequilibrium--understanding the evolutionary past and mapping the medical future. *Nat Rev Genet*, *9*(6), 477-485. https://doi.org/10.1038/nrg2361

Smith, N. D., & Cottler, L. B. (2018). The epidemiology of post-traumatic stress disorder and alcohol use disorder. *Alcohol research: current reviews*, *39*(2), 113.

Smith, S. M., Douaud, G., Chen, W., Hanayik, T., Alfaro-Almagro, F., Sharp, K., & Elliott, L. T. (2021). An expanded set of genome-wide association studies of brain imaging phenotypes in UK Biobank. *Nat Neurosci*, *24*(5), 737-745. https://doi.org/10.1038/s41593-021-00826-4

Socodato, R., Henriques, J. F., Portugal, C. C., Almeida, T. O., Tedim-Moreira, J., Alves, R. L., Canedo, T., Silva, C., Magalhaes, A., Summavielle, T., & Relvas, J. B. (2020). Daily alcohol intake triggers aberrant synaptic pruning leading to synapse loss and anxiety-like behavior. *Sci Signal*, *13*(650). https://doi.org/10.1126/scisignal.aba5754

Sollis, E., Mosaku, A., Abid, A., Buniello, A., Cerezo, M., Gil, L., Groza, T., Gunes, O., Hall, P., Hayhurst, J., Ibrahim, A., Ji, Y., John, S., Lewis, E., MacArthur, J. A. L., McMahon, A., Osumi-Sutherland, D., Panoutsopoulou, K., Pendlington, Z., . . . Harris, L. W. (2023). The NHGRI-EBI GWAS Catalog: knowledgebase and deposition resource. *Nucleic Acids Res*, *51*(D1), D977-D985. https://doi.org/10.1093/nar/gkac1010

Stark, R., Grzelak, M., & Hadfield, J. (2019). RNA sequencing: the teenage years. *Nat Rev Genet*, *20*(11), 631-656. https://doi.org/10.1038/s41576-019-0150-2

Sutherland, G. T., Sheedy, D., Stevens, J., McCrossin, T., Smith, C. C., van Roijen, M., & Kril, J. J. (2016). The NSW brain tissue resource centre: Banking for alcohol and major neuropsychiatric disorders research. *Alcohol*, *52*, 33-39. https://doi.org/10.1016/j.alcohol.2016.02.005

Szabo, G., & Saha, B. (2015). Alcohol's effect on host defense. *Alcohol research: current reviews*, *37*(2), 159.

Szklarczyk, D., Gable, A. L., Lyon, D., Junge, A., Wyder, S., Huerta-Cepas, J., Simonovic, M., Doncheva, N. T., Morris, J. H., Bork, P., Jensen, L. J., & Mering, C. V. (2019). STRING v11: protein-protein association networks with increased coverage, supporting functional discovery in genome-wide experimental datasets. *Nucleic Acids Res*, *47*(D1), D607-D613. https://doi.org/10.1093/nar/gky1131

Thompson, A., Cook, J., Choquet, H., Jorgenson, E., Yin, J., Kinnunen, T., Barclay, J., Morris, A. P., & Pirmohamed, M. (2020). Functional validity, role, and implications of heavy alcohol consumption genetic loci. *Sci Adv*, *6*(3), eaay5034. https://doi.org/10.1126/sciadv.aay5034

Tiwari, P., Fanibunda, S. E., Kapri, D., Vasaya, S., Pati, S., & Vaidya, V. A. (2021). GPCR signaling: role in mediating the effects of early adversity in psychiatric disorders. *FEBS J*, *288*(8), 2602-2621. https://doi.org/10.1111/febs.15738

Tomoda, A., Polcari, A., Anderson, C. M., & Teicher, M. H. (2012). Reduced visual cortex gray matter volume and thickness in young adults who witnessed domestic violence during childhood. *PLoS One*, *7*(12), e52528. https://doi.org/10.1371/journal.pone.0052528

Trevejo-Nunez, G., Kolls, J. K., & De Wit, M. (2015). Alcohol use as a risk factor in infections and healing: A clinician's perspective. *Alcohol research: current reviews*, *37*(2), 177.

Tripathi, A., Whitehead, C., Surrao, K., Pillai, A., Madeshiya, A., Li, Y., Khodadadi, H., Ahmed, A. O., Turecki, G., Baban, B., & Pillai, A. (2021). Type 1 interferon mediates chronic stress-induced neuroinflammation and behavioral deficits via complement component 3-dependent pathway. *Mol Psychiatry*, *26*(7), 3043-3059. https://doi.org/10.1038/s41380-021-01065-6

Tyrer, P. (2018). A comparison of DSM and ICD classifications of mental disorder. *Advances in Psychiatric Treatment*, *20*(4), 280-285. https://doi.org/10.1192/apt.bp.113.011296

Uffelmann, E., Huang, Q. Q., Munung, N. S., de Vries, J., Okada, Y., Martin, A. R., Martin, H. C., Lappalainen, T., & Posthuma, D. (2021). Genome-wide association studies. *Nature Reviews Methods Primers*, *1*(1). https://doi.org/10.1038/s43586-021-00056-9

Van Booven, D., Mengying, L., Sunil Rao, J., Blokhin, I. O., Dayne Mayfield, R., Barbier, E., Heilig, M., & Wahlestedt, C. (2021). Alcohol use disorder causes global changes in splicing in the human brain. *Transl Psychiatry*, *11*(1), 2. https://doi.org/10.1038/s41398-020-01163-z

van der Meer, D., Kaufmann, T., Shadrin, A. A., Makowski, C., Frei, O., Roelfs, D., Monereo-Sanchez, J., Linden, D. E. J., Rokicki, J., Alnaes, D., de Leeuw, C., Thompson, W. K., Loughnan, R., Fan, C. C., Westlye, L. T., Andreassen, O. A., & Dale, A. M. (2021). The genetic architecture of human cortical folding. *Sci Adv*, *7*(51), eabj9446. https://doi.org/10.1126/sciadv.abj9446

Veldheer, S., Yingst, J., Zhu, J., & Foulds, J. (2015). Ten-year weight gain in smokers who quit, smokers who continued smoking and never smokers in the United States, NHANES 2003-2012. *Int J Obes (Lond)*, *39*(12), 1727-1732. https://doi.org/10.1038/ijo.2015.127

Verhulst, B., Neale, M. C., & Kendler, K. S. (2015). The heritability of alcohol use disorders: a meta-analysis of twin and adoption studies. *Psychol Med*, *45*(5), 1061-1072. https://doi.org/10.1017/S0033291714002165

Wainberg, M., Forde, N. J., Mansour, S., Kerrebijn, I., Medland, S. E., Hawco, C., & Tripathy, S. J. (2024). Genetic architecture of the structural connectome. *Nat Commun*, *15*(1), 1962. https://doi.org/10.1038/s41467-024-46023-2

Wainberg, M., Sinnott-Armstrong, N., Mancuso, N., Barbeira, A. N., Knowles, D. A., Golan, D., Ermel, R., Ruusalepp, A., Quertermous, T., Hao, K., Bjorkegren, J. L. M., Im, H. K., Pasaniuc, B., Rivas, M. A., & Kundaje, A. (2019). Opportunities and challenges for transcriptome-wide association studies. *Nat Genet*, *51*(4), 592-599. https://doi.org/10.1038/s41588-019-0385-z

Wang, A. C., Jensen, E. H., Rexach, J. E., Vinters, H. V., & Hsieh-Wilson, L. C. (2016). Loss of O-GlcNAc glycosylation in forebrain excitatory neurons induces neurodegeneration. *Proc Natl Acad Sci U S A*, *113*(52), 15120-15125. https://doi.org/10.1073/pnas.1606899113

Weera, M. M., Agim, Z. S., Cannon, J. R., & Chester, J. A. (2019). Genetic correlations between nicotine reinforcement-related behaviors and propensity toward high or low alcohol

preference in two replicate mouse lines. *Genes Brain Behav*, *18*(3), e12515. https://doi.org/10.1111/gbb.12515

Weinberger, A. H., & Seng, E. K. (2023). The Relationship of Tobacco Use and Migraine: A Narrative Review. *Curr Pain Headache Rep*, *27*(4), 39-47. https://doi.org/10.1007/s11916-023-01103-8

Werme, J., van der Sluis, S., Posthuma, D., & de Leeuw, C. A. (2022). An integrated framework for local genetic correlation analysis. *Nat Genet*, *54*(3), 274-282. https://doi.org/10.1038/s41588-022-01017-y

White, A. M. (2020). Gender Differences in the Epidemiology of Alcohol Use and Related Harms in the United States. *Alcohol Res*, *40*(2), 01. https://doi.org/10.35946/arcr.v40.2.01

Wittchen, H. U., Jacobi, F., Rehm, J., Gustavsson, A., Svensson, M., Jonsson, B., Olesen, J., Allgulander, C., Alonso, J., Faravelli, C., Fratiglioni, L., Jennum, P., Lieb, R., Maercker, A., van Os, J., Preisig, M., Salvador-Carulla, L., Simon, R., & Steinhausen, H. C. (2011). The size and burden of mental disorders and other disorders of the brain in Europe 2010. *Eur Neuropsychopharmacol*, *21*(9), 655-679. https://doi.org/10.1016/j.euroneuro.2011.07.018

Xue, Q., Ji, S., Xu, H., & Yu, S. (2024). O-GlcNAcylation: a pro-survival response to acute stress in the cardiovascular and central nervous systems. *Eur J Med Res*, *29*(1), 174. https://doi.org/10.1186/s40001-024-01773-z

Yang, Z., Zhang, Y., Cheng, J., & Zheng, R. (2020). Meta-analysis of brain gray matter changes in chronic smokers. *Eur J Radiol*, *132*, 109300. https://doi.org/10.1016/j.ejrad.2020.109300

Yetter, M. A., Fitzgerald, T. R., Philippi, C. L., & Bruce, S. E. (2024). Pro-inflammatory markers are related to cortical network connectivity in women exposed to interpersonal trauma with PTSD. *Behav Brain Res*, 114942. https://doi.org/10.1016/j.bbr.2024.114942

Zanoaga, M. D., Friligkou, E., He, J., Pathak, G. A., Koller, D., Cabrera-Mendoza, B., Stein, M. B., & Polimanti, R. (2023). Brainwide Mendelian Randomization Study of Anxiety Disorders and Symptoms. *Biol Psychiatry*. https://doi.org/10.1016/j.biopsych.2023.11.006

Zhang, X., Sun, A., & Ge, J. (2021). Origin and Spread of the ALDH2 Glu504Lys Allele. *Phenomics*, *1*(5), 222-228. https://doi.org/10.1007/s43657-021-00017-y

Zhao, J., Stockwell, T., Naimi, T., Churchill, S., Clay, J., & Sherk, A. (2023). Association Between Daily Alcohol Intake and Risk of All-Cause Mortality: A Systematic Review and Meta-

analyses. *JAMA Netw Open*, *6*(3), e236185. https://doi.org/10.1001/jamanetworkopen.2023.6185

Zhou, Y., Zhou, B., Pache, L., Chang, M., Khodabakhshi, A. H., Tanaseichuk, O., Benner, C., & Chanda, S. K. (2019). Metascape provides a biologist-oriented resource for the analysis of systems-level datasets. *Nat Commun*, *10*(1), 1523. https://doi.org/10.1038/s41467-019-09234-6

Chapter 1

Contribution of infectious diseases to the selection of ADH1B and ALDH2 gene variants in Asian populations

Abstract

The gene variants *ADH1B*2* (Arg48His, rs1229984) and *ALDH2*2* (Glu504Lys, rs671) are common in East Asian populations but rare in other populations. We propose that selective pressures from pathogen exposure and dietary changes during the neolithic transition favored these variants. Thus, their current association with differences in alcohol sensitivity likely results from phenotypic plasticity rather than direct natural selection. Samples sourced from the Allele Frequency Database (ALFRED) were utilized to compute the average frequency of *ADH1B*2* and *ALDH2*2* across 88 and 61 countries, respectively. Following computation of the average national allele frequencies, we tested the significance of their correlations with ecological variables. Subsequently, we subjected them to Principal Component Analysis (PCA) and Elastic Net regularization. For comprehensive evaluation, we collected individual-level phenotypic associations, compiling a Phenome-Wide Association Study (PheWAS) spanning multiple ethnicities. Following multiple testing correction, *ADH1B*2* displayed significant correlations with Neolithic transition timing ($r = 0.405$, p.adj = 2.013e-03, $n = 57$) and historical trypanosome burden ($r = -0.418$, p.adj = 0.013, $n = 57$). The first two components of PCA explained 47.7% of the total variability across countries, with the top three contributors being the historical indices of population density and trypanosome and leprosy burdens. Historical burdens of *Mycobacteria tuberculosis* and *leprosy* were the sole predictive variables with positive coefficients that survived Elastic Net regularization. Our analyses suggest that *Mycobacteria* may have played a role in the joint selection of *ADH1B*2* and *ALDH2*2*, expanding the "toxic aldehyde hypothesis" to include *Mycobacterium leprae*. Additionally, our hypothesis, linked to dietary shifts from rice domestication, emphasizes nutritional deficiencies as a key element in the selective pressure exerted by *Mycobacteria*. This offers a plausible explanation for the high frequency of *ADH1B*2* and *ALDH2*2* in Asian populations.

Introduction

According to the latest global status report on alcohol and health by the World Health Organization (WHO), the South-East Asian Region (SEAR) has the second lowest Alcohol-Per capita-Consumption (APC) in the world (Organization, 2019b). Cultural and social factors are known to affect drinking behaviors of Asian subjects which show the lowest alcohol consumption levels compared to other ethnicities (Lui & Zamboanga, 2019). APC levels have been increasing in the SEAR, but the percentage of current drinkers did not significantly change, being constant at around 33% from 2000 to 2016 (Organization, 2019b). This trend stays stable also when looking at differences in APC amongst countries within SEAR (Organization, 2019b). APC levels in SEAR countries match the worldwide distribution of two of the main genetic variants associated with reduced alcohol intake, namely *ADH1B**2 (Arg48His, rs1229984) and *ALDH2**2 (Glu504Lys, rs671) (Zhou et al, 2022). Indeed, these two variants exhibit a robust association with decreased alcohol intake in the East Asian population (Chang et al., 2023; Hashimoto et al., 2016), reaching their peak frequency within this demographic (Polimanti & Gelernter, 2018; Zhang et al., 2021). In contrast, they are found to be rare and less effective in other populations (Schaschl et al., 2022; Whitfield, 2002). Encoding the beta subunit of class I of the alcohol dehydrogenases (ADH), *ADH1B* is responsible for the interconversion between alcohols and aldehydes or ketones with the reduction of NAD+ to NADH (Zakhari, 2006). *ALDH2* encodes the metabolizing enzyme for many reactive and oxidative stress-generated aldehydes, such as acetaldehyde generated by ADH that is converted to acetate (Chen et al., 2014). Due to its high reactivity and genotoxicity, acetaldehyde has been classified by the International Agency for Research on Cancer (IARC) of WHO as carcinogenic to humans in Group 1 tissues (Baan et al., 2007). Compared to the enzyme encoded by the ancestral allele, *ADH1B*2* converts alcohol to acetaldehyde at a fast rate, causing a huge accumulation of blood

acetaldehyde (Whitfield, 2002). The increased amount of acetaldehyde cannot be oxidized into acetate by *ALDH2*2*, which encodes an inactive enzyme (Crabb et al., 1989). The combination of *ADH1B*2* and *ALDH2*2* increases alcohol toxicity, which results in skin rash and coughing resembling an allergic reaction (Whitfield, 2002). Cumulating evidence shows that rice domestication exerted selective pressures on alcohol and acetaldehyde dehydrogenases among those populations that adopted rice farming during the Neolithic transition in SEAR (Landini et al., 2021; Zhu et al., 2021). As rice constitutes the primary staple food for most populations in the region (Muthayya et al., 2014), the diffusion of rice domestication across East Asia primarily followed a demic pattern (Cobo et al., 2019), influencing and molding local cultures and societies (Bray, 1986; Thomson et al., 2018). Rice domestication was paralleled by the spread of infectious diseases that changed the demographic landscape through SEAR as paddy rice forced low relational mobility (Cobo et al., 2019). Thus, it is conceivable that paddy rice may have imposed new burdens of infection, changing the demographic landscape of the area (Cobo et al., 2019; Liu et al., 2017). Indeed, it is remarkable that the emergence of agriculture is associated with an increased allocation of energy to immune functions (Wells & Stock, 2020). Additional phenotypes correlated with alcohol intake, such as risky sexual behaviors (Hutton et al., 2008) and disgust sensitivity (Stafford et al., 2020), as well as broader associations with all risky-type behaviors, have been linked to the historical burden of pathogens (Schaller, 2011). The shared phenotypical association with a reduction in alcohol intake of *ADH1B*2* and *ALDH2*2* (Serajuddin, 2018; Zhou et al., 2022) may therefore be the current result of the effect of the selective pressure exerted by pathogens and changes in dietary habits following the neolithic transition in SEAR. To corroborate this hypothesis, we took advantage of previous works and publicly available data to investigate the correlation between human geography, agriculture, and farming, the exposure to infectious disease, and the

worldwide distribution of the two alleles employing a country-based approach. Furthermore, we attempted to provide a causal explanation linking our findings to changes in aldehyde metabolism caused by *ADH1B*2* and *ALDH2*2*.

Materials and Methods

Allele frequency and ecological data search, extraction, and harmonization

A total of 510 sample populations from the Allele Frequency Database (ALFRED) (Rajeevan et al., 2012), containing 115,303 subjects self-identifying in 284 different ethnicities, was included to calculate the average frequency of *ADH1B*2* across 88 nation-states. For *ALDH2*2*, 212 sample populations composed of 35,540 subjects, self-identifying in 142 ethnicities, were used to calculate its average frequency across 61 nation states. More details on the sample demographics are provided in Tables S1 and S2. Ecological data describing food accessibility, human geography, and historical exposure to infectious diseases were collected from a dataset previously assembled by Fedderke and colleagues (Fedderke et al., 2017) containing 207 nations as described in Table 1 and Table S4. Samples from ALFRED (Rajeevan et al., 2012) were divided by country and reported ethnicity to compute a mean weighted on their relative sample size. National allelic frequencies were computed through a second weighted mean combining previous results accordingly to the relative share of each reported ethnicity in the demographic composition of their corresponding sampling country, as described in The World Factbook (CIA, 2021). The resulting average national allele frequencies were mapped using the rworldmap R package (South, 2011). For Brazil, the United States, and Australia, only native population samples were considered. To standardize our dataset, missing values in the distribution of ALDH2*2 frequency underwent imputation through Multivariate Imputation by Chained Equations (MICE) using random forests. This imputation method was implemented

using the miceRanger R package (Wilson, 2020). The final imputed dataset was the result of five iterations using these variables as predictors in the order: genetic distance to Ethiopia—1500 match (fstdist1500_eth), genetic distance to the United Kingdom—1500 match (fstdist1500_uk) and ethnic fractionalization index (efrac) Table S4. Other parameters were set as default.

Correlation analyses

Given the nonparametric nature of our dataset, we conducted a Kendall hypothesis test on complete observations to explore the potential correlation between the variables of interest, followed by an exploratory Principal Component Analysis (PCA). The Kendal hypothesis test was carried out as implemented by the corr.test function in the psych R package on the following variables: national average frequency of *ADH1B*2* (adh1b_2), national average frequency of *ALDH2*2* (aldh2_2), natural logarithm of distribution of population in 1500 CE (ln_pop1500), natural logarithm of distribution of population density in 1500 CE (ln_pd1500), natural logarithm of distribution of neolithic transition timing until the year 1500 CE (ln_yst1500), natural logarithm of percentage of arable land (ln_arable), natural logarithm of temperature (ln_temp), natural logarithm of precipitation (ln_precip), index of historical prevalence of leishmaniasis (his_leishmanias), index of historical prevalence of schistosomes (his_schistosomes), index of historical prevalence of filariae (his_filariae), index of historical prevalence leprosy (his_leprosy), index of historical prevalence typhus (his_ typhus), index of historical prevalence of malaria (his_malaria), index of historical prevalence of trypanosomes (his_trypanosomes), index of historical prevalence of dengue (his_dengue), and index of historical prevalence tuberculosis (his_tuberculosis). The test considered only complete

observation by selecting the use = “complete” option; results were adjusted by Bonferroni correction.

Principal component analysis

After excluding rows with missing data from the imputed dataset, we conducted PCA using the PCA function within the FactoMineR R package. Subsequently, we analyzed the PCA results using functions from the factoextra R package. The contribution of each variable to the first two dimensions was determined using the fviz_contrib function. To visualize variability across nation states, we utilized the fviz_pca_var function to plot the results within the first two dimensions. The contribution of each variable was depicted by the length and orientation of their eigenvectors. Additionally, the quality of the representation of variability within each nation state was assessed by incorporating the measure of squared cosines of their eigenvectors. Considering the significant correlations among ecological variables post-Bonferroni correction and the orientation of the square cosine of their eigenvectors (cos2), we excluded those surpassing a variance inflation factor (VIF) cutoff of 3.00 as predictors in the subsequent analysis. This step was undertaken to alleviate potential collinearity effects on average allelic distributions. Squared eigenvector cosines measure the similarity between eigenvectors of a matrix. It involves calculating the cosine of the angle between two eigenvectors, squaring it to focus on similarity, with high values indicating strong alignment.

Elastic net regularization analysis

The effectiveness of retained variables in explaining variance was evaluated using Elastic Net regularization, implemented in the glmnet R package through the cv.glmnet function (Friedman et al., 2023). This approach enabled us to alleviate and ultimately eliminate factors that were negatively impacting the fitting to the distributions of average national allelic frequencies, both

for single alleles and combined alleles. The first group of predictors included the following ecological variables: natural logarithm of distribution population in 1500 CE (ln_pop1500), natural logarithm of distribution population density in 1500 CE (ln_pd1500), natural logarithm of distribution neolithic transition timing until the year 1500 CE (ln_yst1500), natural logarithm of distribution percentage of arable land (ln_arable), natural logarithm of distribution precipitation (ln_precip), index of historical prevalence of leishmaniasis (his_leishmanias), index of historical prevalence of schistosomes (his_schistosomes), index of historical prevalence leprosy (his_leprosy), index of historical prevalence typhus (his_ typhus), index of historical prevalence of trypanosomes (his_ trypanosomes), and index of historical prevalence tuberculosis (his_ tuberculosis). This set of predictors was applied to the distribution of national average frequencies of *ADH1B*2* (adh1b_2) and *ALDH2*2* (aldh2_2) as outcome variables. In the second group of predictors, adh1b_2 was included, and this set was combined with aldh2_2 as the sole outcome variable. All regularizations were computed 10 times with gamma values set as default (0, 0.25, 0.5, 0.75, 1).

Phenome-wide association study (PheWAS)

To enhance the interpretation of results, phenotypic associations linked to each allele were extracted from the meta-analysis conducted on the Global Biobank (McInnes et al., 2019), which contains data compiled from summary statistics from the Million Veteran Program (MVP) (Gaziano et al., 2016), the Japanese Biobank (Nagai et al., 2017), and the UK Biobank (Sudlow et al., 2015). These associations were then compiled to create a PheWAS that takes into consideration exposure to infectious diseases and dietary changes. Phenotypic associations categorized as *“Early life factors”, “Disease outcome”*, *”Anthropometrics”*, and *“Biomarkers”*

were filtered for FDR≤ 0.005 calculated according to the Benjamini–Hochberg correction method.

Results

The average allelic frequencies of *ADH1B*2* and *ALDH2*2* were calculated for 88 and 61 nation states (Figures 1C and 2C), respectively. Most countries showed low frequency levels for both alleles, with their distributions reaching their density peak on the lower tail (Figures 1A and 2A). This is especially true for *ALDH2*2*, whose variability (median = 0.072, λ = 7.114, var = 0.004) was lower than that of *ADH1B*2* (median = 0, λ = Inf ,var = 0.047). After imputation, *ALDH2*2* variability was further decreased (var = 0.0004) Table S3. Geographically, the two alleles were similarly distributed, reaching their highest frequencies in SEAR, with Japan being at the extreme upper tail of their distribution (*ADH1B*2* = 76.4%, *ALDH2*2* = 29%) (Figures 1C and 2C). While the two alleles were significantly correlated (r = 0.267, p = 6.411e-03, n = 57), after applying a Bonferroni correction for multiple testing only *ADH1B*2* showed significant correlations, with the Neolithic transition timing until the year 1500 CE (ln_yst1500; r = 0.405, p = 1.480e-05, p.adj = 2.013e-03, n = 57) and the historical burden of trypanosomes infection (his_trypanosomes; r = −0.418, p = 1.010e-04, p.adj = 0.013, n = 57), which were also negatively correlated with each other (r = −0.522, p = 1.537e-06, p.adj = 1.796e-03, n = 57) (Figure 3). After Bonferroni correction, the number of significant correlations between ecological variables was reduced, except for some describing the historical burden of many infectious diseases, which remained positively correlated (Figure 3). Variability across ecological variables was partitioned into 10 main components by PCA; the first two principal components described as 27.8% and 19.9%, respectively, of the total variance (Figure 4A), ensuring within countries a representation of more than 50% of the total variance according to square eigenvectors cosines (cos2). The most relevant contributors to the first component were the historical burden of infectious

diseases led by trypanosomes and population indexes (population density in 1500 CE and population in 1500 CE); the percentage of arable land was the main contributor to the second principal component (Figure 4B). Southern European, Asian, and Middle Eastern countries appear to lay close together due to population density in 1500 CE, population in 1500 CE, and percentage of arable land, while most Oceanian, South American, and North African countries were differently oriented due to the historical burden of infectious diseases (Figure 4C). Due to the squared eigenvectors exhibiting cosine similarities among ecological variables, forming two nearly opposite directional sets and implying a high degree of collinearity (Figure 4C), and with both allele frequencies demonstrating significant inflation at the extremes of their distributions (Figures 1B and 2B), temperature (VIF = 3.20), historical prevalence of filariae (VIF = 4.38), historical prevalence of malaria (VIF = 3.51), and historical prevalence of dengue (VIF = 3.90) were excluded based on their respective VIF values. Elastic net regularization revealed that, at the lowest values of λ, allowing for the best fit, population in 1500 CE, historical prevalence of leprosy, and historical prevalence of tuberculosis were the sole variables maintaining positive coefficients across all nets. For larger λ values, resulting in fit alterations up to one standard error, no variable exhibited a nonzero coefficient except for ln_leprosy. Notably, this pattern held true for nets considering *ALDH2*2* as the single dependent variable, only when *ADH1B*2* was included among predictors (Figure 5). Furthermore, when both allelic distributions were considered, variability in the fit due to different γ values was diminished (Figure 5). In the Global Biobank (McInnes et al., 2019), meta-analyzed phenotypic associations categorized as *"Early life factors," "Disease outcome," "Anthropometrics,"* and *"Biomarkers"* were collected for both alleles, reaching a total amount of 105 for *ADH1B*2* and 58 for *ALDH2*2*. Of those associated with *ADH1B*2*, 33 traits passed a minimum threshold of FDR≤ 0.005, with no one directly linked to an infectious disease; the

most significant was the reduction of insulin-like growth factor 1 (IGF-1) levels (n = 423,324, beta = −0.090, FDR = 2.89E-36) (Table 2). For *ALDH2*2*, six phenotypic associations passed a minimum threshold of FDR≤ 0.005: in particular, the increased propensity to develop nasal polyps was the most significant one (n = 6546, beta = 7.108, FDR = 0.001) (Table 3).

Discussion

According to the results from Elastic Net regularization, the historical burden of tuberculosis (his_tuberculosis) and leprosy (his_leprosy) emerged as the sole positive predictors for the distribution of national average frequencies of *ADH1B*2* and *ALDH2*2*. This association appears to be primarily driven by *ADH1B*2*, as in models considering *ALDH2*2* alone as the dependent variable, his_leprosy was retained only when *ADH1B*2* was included as a predictor. The PCA results suggest that the correlation observed between *ADH1B*2* and the distance to the Neolithic transition (ln_yst1500) may potentially be attributed to other variables exhibiting a positive correlation with it. The collective findings of our analyses indicate that exposure to mycobacterial infections may have played a role in the concurrent selection of these two alleles among Neolithic rice farmers (N. Wang et al., 2018; Z. Wang et al., 2018). Consistent with our findings, a recent exome-wide association study among Han Chinese linked *ALDH2*2* with heightened susceptibility to *Mycobacterium leprae* (p = 2.00 × 10–20, OR = 1.35) (N. Wang et al., 2018). This association was subsequently replicated in another Han Chinese cohort but not observed in African populations, where the maximum frequency of *ALDH2*2* in the sample was below 1% (Gilchrist et al., 2022). Furthermore, clinical research conducted among Korean patients linked *ALDH2*2* with a diminished burden of infection from *Mycobacterium tuberculosis*, particularly among subjects also afflicted by alcoholism, along with *ADH1B*2* (Park et al., 2014). The dissemination of *ADH1B*2* in Southeast Asia occurred sixty thousand years prior to that of *ALDH2*2*, as demonstrated by Li and colleagues (Li et al., 2011), and Luo

and colleagues (Luo et al., 2009). Mutations within this extensive timeframe, taking place in ecological niches, substantiate alterations in their patterns of natural selection. Consequently, these two alleles are presently associated solely with five common phenotypes in our Phenome-Wide Association Studies (PheWAS), all falling within the *"Anthropometrics"* category, such as *"Body fat percentage"*, *"BMI"*, or *"Trunk fat percentage"* (see Tables 2 and 3). Rice-based diets have been linked to deficiencies in lipid absorption, resulting in reduced food intake and BMI measurements (Lim et al., 2016; Yang et al., 2012). *ADH1B*2*, associated with higher levels of LDL-C in Asian populations (Luo et al., 2023), might have conferred survival advantages to carriers due to the decreased energy intake resulting from their dietary habits. Recent evidence of positive selection on variants related to ADH, including *ADH1B*2*, has been identified in populations in Ethiopia following the advent of agriculture (McQuillan et al., 2022). However, while rice domestication was a key catalyst for the Neolithic transition in Southeast Asia and the Pacific, its dissemination appears to have been greatly influenced by demographic factors (Bray, 1986; Cobo et al., 2019), which are closely linked to mycobacterial infections (Benjak et al., 2018; Luo et al., 2015; Okazaki et al., 2019). We propose that alterations in dietary practices in Southeast Asia due to the spread of rice farming resulted in chronic retinol deficiency (Tang et al., 2009) and disrupted lipid metabolism (Lim et al., 2016; Yang et al., 2012), thereby heightening selective pressures from infectious diseases in the region. Indeed, retinoid acid has been shown to enhance antimicrobial responses in epithelial and mucous tissues (Huang et al., 2018), playing a protective role in the response to mycobacterial infections (Fairley et al., 2024; Kim et al., 2019; Tenorio de Menezes et al., 2023), alongside lipid metabolism, which mediates host inflammatory responses (Fairley et al., 2024; Kim & Shin, 2023). Our hypotheses align with the idea that decreased detoxification of aldehydes may confer beneficial effects against pathogens by promoting immune cell activation through the

cytokine interferon-gamma (Darwin & Stanley, 2022). The advent of rice domestication conferred a significant advantage to farming populations in Southeast Asia and the Pacific, facilitating its spread throughout the region during the Neolithic transition. Subsequent shifts in the demographic landscape were profoundly influenced by mycobacterial infections. Elevated levels of circulating aldehydes may have provided an advantage to ancient rice farmers, enhancing their fitness against the high prevalence of mycobacterial infections. This enhancement of immunity could serve as compensation for their retinol deficiency (Tang et al., 2009) and disrupted lipid metabolism, both of which were consequences of a rice-based diet. In summary, our study provides an initial understanding of the natural selection of *ADH1B* and *ALDH2* genes in Asian populations, with a focus on dietary changes following the Neolithic transition within the context of the behavioral immune system theory. Our analysis expands upon the "aldehydes hypothesis" proposed by Darwin and Stanley (Darwin & Stanley, 2022), originally centered on *Mycobacterium tuberculosis*, by advocating for the inclusion of *Mycobacterium leprae* as a target for future research into the beneficial effects of aldehydes. Given the speculative nature of our research and the limitations posed by the available data, several important considerations must be addressed. First, national frequencies of *ADH1B*2* and *ALDH2*2* were calculated using ALFRED (Rajeevan et al., 2012), which provides the geographical origin of each sample along with self-reported ethnicity. However, this dataset lacks more detailed information on individual ancestry, which could impact the accuracy of our analyses. Natural selection is a complex and dynamic process, yet the scarcity of time-framed data on *ADH1B*2* and *ALDH2*2* (Gao et al., 2020), as well as ecological variables, necessitated a country-based approach. We relied on data from Fedderke and colleagues (Fedderke et al., 2017), which were the most reliable sources available to us. It is important to note that while there is some convergent evidence in Ethiopian populations linking the adoption of agriculture,

higher levels of aldehydes (McQuillan et al., 2022), and retinol deficiency to mycobacteria (Keflie et al., 2018), research on *ADH1B*2* and *ALDH2*2* remains biased toward Asian populations and their drinking patterns. However, there is evidence for other loci associated with fatty acid metabolism, cholesterol/triglyceride biosynthesis, and retinoic acid production being affected by the emergence and expansion of rice agriculture (Landini et al., 2021; L.-X. Wang et al., 2016). Notably, among these, the expression of *ALDH1B1* has also been associated with immune response to pathogens, modulating the production of interferon beta in mice (Sun et al., 2024). These factors underscore the complexity and interconnectedness of genetic and environmental influences on human health and disease susceptibility. Indeed, while further research is imperative, our findings, which integrate diverse datasets, provide momentum for the exploration of complex phenotypes, such as alcohol drinking patterns, within a broader framework that encompasses considerations of nutrition and immune response. This holistic approach acknowledges the interconnectedness of genetic, dietary, and immunological factors in shaping human health and disease susceptibility. By embracing this multidimensional perspective, future investigations can yield deeper insights into the intricate interplay between genetics, behavior, and environmental influences on human biology.

References

Baan, R., Straif, K., Grosse, Y., Secretan, B., El Ghissassi, F., Bouvard, V., Altieri, A., Cogliano, V., & Group, W. H. O. I. A. f. R. o. C. M. W. (2007). Carcinogenicity of alcoholic beverages. *Lancet Oncol*, *8*(4), 292-293. https://doi.org/10.1016/s1470-2045(07)70099-2

Benjak, A., Avanzi, C., Singh, P., Loiseau, C., Girma, S., Busso, P., Fontes, A. N. B., Miyamoto, Y., Namisato, M., Bobosha, K., Salgado, C. G., da Silva, M. B., Bouth, R. C., Frade, M. A. C., Filho, F. B., Barreto, J. G., Nery, J. A. C., Buhrer-Sekula, S., Lupien, A., . . . Cole, S. T. (2018). Phylogenomics and antimicrobial resistance of the leprosy bacillus Mycobacterium leprae. *Nat Commun*, *9*(1), 352. https://doi.org/10.1038/s41467-017-02576-z

Bray, F. (1986). *The Rice Economies: Technology and Development in Asian Societies* (1 ed.). University of California Press. https://doi.org/10.2307/jj.2711588

Chang, T. G., Yen, T. T., Wei, C. Y., Hsiao, T. H., & Chen, I. C. (2023). Impacts of ADH1B rs1229984 and ALDH2 rs671 polymorphisms on risks of alcohol-related disorder and cancer. *Cancer Med*, *12*(1), 747-759. https://doi.org/10.1002/cam4.4920

Chen, C. H., Ferreira, J. C., Gross, E. R., & Mochly-Rosen, D. (2014). Targeting aldehyde dehydrogenase 2: new therapeutic opportunities. *Physiol Rev*, *94*(1), 1-34. https://doi.org/10.1152/physrev.00017.2013

CIA. (2021). The World Fact Book 2022.

Cobo, J. M., Fort, J., & Isern, N. (2019). The spread of domesticated rice in eastern and southeastern Asia was mainly demic. *Journal of Archaeological Science*, *101*, 123-130. https://doi.org/10.1016/j.jas.2018.12.001

Crabb, D. W., Edenberg, H. J., Bosron, W. F., & Li, T. K. (1989). Genotypes for aldehyde dehydrogenase deficiency and alcohol sensitivity. The inactive ALDH2(2) allele is dominant. *J Clin Invest*, *83*(1), 314-316. https://doi.org/10.1172/JCI113875

Darwin, K. H., & Stanley, S. A. (2022). The aldehyde hypothesis: metabolic intermediates as antimicrobial effectors. *Open Biol*, *12*(4), 220010. https://doi.org/10.1098/rsob.220010

Fairley, J. K., Ferreira, J. A., Fraga, L. A. O., Lyon, S., Valadao Cardoso, T. M., Boson, V. C., Madureira Nunes, A. C., Medeiros Cinha, E. H., de Oliveira, L. B. P., Magueta Silva, E. B., Marcal, P. H. F., Branco, A. C., Grossi, M. A. F., Jones, D. P., Ziegler, T. R., & Collins, J. M. (2024). High-Resolution Plasma Metabolomics Identifies Alterations in Fatty Acid,

Energy, and Micronutrient Metabolism in Adults Across the Leprosy Spectrum. *J Infect Dis*, *229*(4), 1189-1199. https://doi.org/10.1093/infdis/jiad410

Fedderke, J. W., Klitgaard, R. E., & Napolioni, V. (2017). Genetic adaptation to historical pathogen burdens. *Infect Genet Evol*, *54*, 299-307. https://doi.org/10.1016/j.meegid.2017.07.017

Gao, Y., Zhang, C., Yuan, L., Ling, Y., Wang, X., Liu, C., Pan, Y., Zhang, X., Ma, X., Wang, Y., Lu, Y., Yuan, K., Ye, W., Qian, J., Chang, H., Cao, R., Yang, X., Ma, L., Ju, Y., . . . Xu, S. (2020). PGG.Han: the Han Chinese genome database and analysis platform. *Nucleic Acids Res*, *48*(D1), D971-D976. https://doi.org/10.1093/nar/gkz829

Gaziano, J. M., Concato, J., Brophy, M., Fiore, L., Pyarajan, S., Breeling, J., Whitbourne, S., Deen, J., Shannon, C., Humphries, D., Guarino, P., Aslan, M., Anderson, D., LaFleur, R., Hammond, T., Schaa, K., Moser, J., Huang, G., Muralidhar, S., . . . O'Leary, T. J. (2016). Million Veteran Program: A mega-biobank to study genetic influences on health and disease. *J Clin Epidemiol*, *70*, 214-223. https://doi.org/10.1016/j.jclinepi.2015.09.016

Gilchrist, J. J., Auckland, K., Parks, T., Mentzer, A. J., Goldblatt, L., Naranbhai, V., Band, G., Rockett, K. A., Toure, O. B., Konate, S., Sissoko, S., Djimde, A. A., Thera, M. A., Doumbo, O. K., Sow, S., Floyd, S., Ponnighaus, J. M., Warndorff, D. K., Crampin, A. C., . . . Hill, A. V. S. (2022). Genome-wide association study of leprosy in Malawi and Mali. *PLoS Pathog*, *18*(9), e1010312. https://doi.org/10.1371/journal.ppat.1010312

Hashimoto, M., Watanabe, M., Uematsu, Y., Hattori, S., Miyai, N., Utsumi, M., Oka, M., Hayashida, M., Kinoshita, K., Arita, M., & Takeshita, T. (2016). Relationships of alcohol dehydrogenase 1B (ADH1B) and aldehyde dehydrogenase 2 (ALDH2) genotypes with alcohol sensitivity, drinking behavior and problem drinking in Japanese older men. *Environ Health Prev Med*, *21*(3), 138-148. https://doi.org/10.1007/s12199-016-0507-5

Huang, Z., Liu, Y., Qi, G., Brand, D., & Zheng, S. G. (2018). Role of Vitamin A in the Immune System. *J Clin Med*, *7*(9). https://doi.org/10.3390/jcm7090258

Hutton, H. E., McCaul, M. E., Santora, P. B., & Erbelding, E. J. (2008). The relationship between recent alcohol use and sexual behaviors: gender differences among sexually transmitted disease clinic patients. *Alcohol Clin Exp Res*, *32*(11), 2008-2015. https://doi.org/10.1111/j.1530-0277.2008.00788.x

Keflie, T. S., Samuel, A., Woldegiorgis, A. Z., Mihret, A., Abebe, M., & Biesalski, H. K. (2018). Vitamin A and zinc deficiencies among tuberculosis patients in Ethiopia. *J Clin Tuberc Other Mycobact Dis*, *12*, 27-33. https://doi.org/10.1016/j.jctube.2018.05.002

Kim, E. W., De Leon, A., Jiang, Z., Radu, R. A., Martineau, A. R., Chan, E. D., Bai, X., Su, W. L., Montoya, D. J., Modlin, R. L., & Liu, P. T. (2019). Vitamin A Metabolism by Dendritic Cells Triggers an Antimicrobial Response against Mycobacterium tuberculosis. *mSphere*, *4*(3). https://doi.org/10.1128/mSphere.00327-19

Kim, H., & Shin, S. J. (2023). Revolutionizing control strategies against Mycobacterium tuberculosis infection through selected targeting of lipid metabolism. *Cell Mol Life Sci*, *80*(10), 291. https://doi.org/10.1007/s00018-023-04914-5

Landini, A., Yu, S., Gnecchi-Ruscone, G. A., Abondio, P., Ojeda-Granados, C., Sarno, S., De Fanti, S., Gentilini, D., Di Blasio, A. M., Jin, H., Nguyen, T. T., Romeo, G., Prata, C., Bortolini, E., Luiselli, D., Pettener, D., & Sazzini, M. (2021). Genomic adaptations to cereal-based diets contribute to mitigate metabolic risk in some human populations of East Asian ancestry. *Evol Appl*, *14*(2), 297-313. https://doi.org/10.1111/eva.13090

Li, H., Gu, S., Han, Y., Xu, Z., Pakstis, A. J., Jin, L., Kidd, J. R., & Kidd, K. K. (2011). Diversification of the ADH1B gene during expansion of modern humans. *Ann Hum Genet*, *75*(4), 497-507. https://doi.org/10.1111/j.1469-1809.2011.00651.x

Lim, S. M., Goh, Y. M., Mohtarrudin, N., & Loh, S. P. (2016). Germinated brown rice ameliorates obesity in high-fat diet induced obese rats. *BMC Complement Altern Med*, *16*, 140. https://doi.org/10.1186/s12906-016-1116-y

Liu, X., Lu, D., Saw, W. Y., Shaw, P. J., Wangkumhang, P., Ngamphiw, C., Fucharoen, S., Lert-Itthiporn, W., Chin-Inmanu, K., Chau, T. N., Anders, K., Kasturiratne, A., de Silva, H. J., Katsuya, T., Kimura, R., Nabika, T., Ohkubo, T., Tabara, Y., Takeuchi, F., . . . Teo, Y. Y. (2017). Characterising private and shared signatures of positive selection in 37 Asian populations. *Eur J Hum Genet*, *25*(4), 499-508. https://doi.org/10.1038/ejhg.2016.181

Lui, P. P., & Zamboanga, B. L. (2019). Ethnocultural diversity in alcohol use and misuse. *Am J Orthopsychiatry*, *89*(5), 543-548. https://doi.org/10.1037/ort0000422

Luo, H. R., Wu, G. S., Pakstis, A. J., Tong, L., Oota, H., Kidd, K. K., & Zhang, Y. P. (2009). Origin and dispersal of atypical aldehyde dehydrogenase ALDH2487Lys. *Gene*, *435*(1-2), 96-103. https://doi.org/10.1016/j.gene.2008.12.021

Luo, T., Comas, I., Luo, D., Lu, B., Wu, J., Wei, L., Yang, C., Liu, Q., Gan, M., Sun, G., Shen, X., Liu, F., Gagneux, S., Mei, J., Lan, R., Wan, K., & Gao, Q. (2015). Southern East Asian origin and coexpansion of Mycobacterium tuberculosis Beijing family with Han Chinese. *Proc Natl Acad Sci U S A*, *112*(26), 8136-8141. https://doi.org/10.1073/pnas.1424063112

Luo, Z., Cheng, J., & Wang, Y. (2023). Effects of the genetic variants of alcohol-metabolizing enzymes on lipid levels in Asian populations: a systematic review and meta-analysis. *Nutr Rev*, *81*(8), 921-938. https://doi.org/10.1093/nutrit/nuac100

McInnes, G., Tanigawa, Y., DeBoever, C., Lavertu, A., Olivieri, J. E., Aguirre, M., & Rivas, M. A. (2019). Global Biobank Engine: enabling genotype-phenotype browsing for biobank summary statistics. *Bioinformatics*, *35*(14), 2495-2497. https://doi.org/10.1093/bioinformatics/bty999

McQuillan, M. A., Ranciaro, A., Hansen, M. E. B., Fan, S., Beggs, W., Belay, G., Woldemeskel, D., & Tishkoff, S. A. (2022). Signatures of Convergent Evolution and Natural Selection at the Alcohol Dehydrogenase Gene Region are Correlated with Agriculture in Ethnically Diverse Africans. *Mol Biol Evol*, *39*(10). https://doi.org/10.1093/molbev/msac183

Muthayya, S., Sugimoto, J. D., Montgomery, S., & Maberly, G. F. (2014). An overview of global rice production, supply, trade, and consumption. *Ann N Y Acad Sci*, *1324*, 7-14. https://doi.org/10.1111/nyas.12540

Nagai, A., Hirata, M., Kamatani, Y., Muto, K., Matsuda, K., Kiyohara, Y., Ninomiya, T., Tamakoshi, A., Yamagata, Z., Mushiroda, T., Murakami, Y., Yuji, K., Furukawa, Y., Zembutsu, H., Tanaka, T., Ohnishi, Y., Nakamura, Y., BioBank Japan Cooperative Hospital, G., & Kubo, M. (2017). Overview of the BioBank Japan Project: Study design and profile. *J Epidemiol*, *27*(3S), S2-S8. https://doi.org/10.1016/j.je.2016.12.005

Okazaki, K., Takamuku, H., Yonemoto, S., Itahashi, Y., Gakuhari, T., Yoneda, M., & Chen, J. (2019). A paleopathological approach to early human adaptation for wet-rice agriculture: The first case of Neolithic spinal tuberculosis at the Yangtze River Delta of China. *Int J Paleopathol*, *24*, 236-244. https://doi.org/10.1016/j.ijpp.2019.01.002

Organization, W. H. (2019). *Global status report on alcohol and health 2018*. World Health Organization.

Park, S. K., Park, C. S., Lee, H. S., Park, K. S., Park, B. L., Cheong, H. S., & Shin, H. D. (2014). Functional polymorphism in aldehyde dehydrogenase-2 gene associated with risk of tuberculosis. *BMC Med Genet*, *15*, 40. https://doi.org/10.1186/1471-2350-15-40

Polimanti, R., & Gelernter, J. (2018). ADH1B: From alcoholism, natural selection, and cancer to the human phenome. *Am J Med Genet B Neuropsychiatr Genet*, *177*(2), 113-125. https://doi.org/10.1002/ajmg.b.32523

Rajeevan, H., Soundararajan, U., Kidd, J. R., Pakstis, A. J., & Kidd, K. K. (2012). ALFRED: an allele frequency resource for research and teaching. *Nucleic Acids Res*, *40*(Database issue), D1010-1015. https://doi.org/10.1093/nar/gkr924

Schaller, M. (2011). The behavioural immune system and the psychology of human sociality. *Philos Trans R Soc Lond B Biol Sci*, *366*(1583), 3418-3426. https://doi.org/10.1098/rstb.2011.0029

Schaschl, H., Gollner, T., & Morris, D. L. (2022). Positive selection acts on regulatory genetic variants in populations of European ancestry that affect ALDH2 gene expression. *Sci Rep*, *12*(1), 4563. https://doi.org/10.1038/s41598-022-08588-0

Serajuddin, U. M., H. (2018). *World development indicators*.

South, A. (2011). rworldmap: A new R package for mapping global data. *R Journal*, *3*(1).

Stafford, L. D., Sekulla, A., Morrison, E., Fleischman, D. S., & Harvey, A. J. (2020). Alcohol and disgust: An intimate relationship. *Drug Alcohol Depend*, *206*, 107780. https://doi.org/10.1016/j.drugalcdep.2019.107780

Sudlow, C., Gallacher, J., Allen, N., Beral, V., Burton, P., Danesh, J., Downey, P., Elliott, P., Green, J., Landray, M., Liu, B., Matthews, P., Ong, G., Pell, J., Silman, A., Young, A., Sprosen, T., Peakman, T., & Collins, R. (2015). UK biobank: an open access resource for identifying the causes of a wide range of complex diseases of middle and old age. *PLoS Med*, *12*(3), e1001779. https://doi.org/10.1371/journal.pmed.1001779

Sun, N., Cai, Q., Zhang, Y., Zhang, R. R., Jiang, J., Yang, H., Qin, C. F., & Cheng, G. (2024). The aldehyde dehydrogenase ALDH1B1 exerts antiviral effects through the aggregation of the adaptor MAVS. *Sci Signal*, *17*(818), eadf8016. https://doi.org/10.1126/scisignal.adf8016

Tang, G., Qin, J., Dolnikowski, G. G., Russell, R. M., & Grusak, M. A. (2009). Golden Rice is an effective source of vitamin A. *Am J Clin Nutr*, *89*(6), 1776-1783. https://doi.org/10.3945/ajcn.2008.27119

Tenorio de Menezes, Y. K., Eto, C., de Oliveira, J., Larson, E. C., Mendes, D., Dias, G. B. M., Delgobo, M., Gubernat, A. K., Gleim, J. L., Munari, E. L., Starick, M., Ferreira, F., Mansur, D. S., Costa, D. L., Scanga, C. A., & Bafica, A. (2023). The Endogenous Retinoic Acid

Receptor Pathway Is Exploited by Mycobacterium tuberculosis during Infection, Both In Vitro and In Vivo. *J Immunol*, *211*(4), 601-611. https://doi.org/10.4049/jimmunol.2200555

Thomson, R., Yuki, M., Talhelm, T., Schug, J., Kito, M., Ayanian, A. H., Becker, J. C., Becker, M., Chiu, C. Y., Choi, H. S., Ferreira, C. M., Fulop, M., Gul, P., Houghton-Illera, A. M., Joasoo, M., Jong, J., Kavanagh, C. M., Khutkyy, D., Manzi, C., . . . Visserman, M. L. (2018). Relational mobility predicts social behaviors in 39 countries and is tied to historical farming and threat. *Proc Natl Acad Sci U S A*, *115*(29), 7521-7526. https://doi.org/10.1073/pnas.1713191115

Wang, L.-X., Wen, S., Wang, C.-C., Zhou, B., & Li, H. (2016). Molecular adaption of alcohol metabolism to agriculture in East Asia. *Quaternary International*, *426*, 187-194. https://doi.org/10.1016/j.quaint.2016.03.008

Wang, N., Wang, Z., Wang, C., Fu, X., Yu, G., Yue, Z., Liu, T., Zhang, H., Li, L., Chen, M., Wang, H., Niu, G., Liu, D., Zhang, M., Xu, Y., Zhang, Y., Li, J., Li, Z., You, J., . . . Zhang, F. (2018). Prediction of leprosy in the Chinese population based on a weighted genetic risk score. *PLoS Negl Trop Dis*, *12*(9), e0006789. https://doi.org/10.1371/journal.pntd.0006789

Wang, Z., Mi, Z., Wang, H., Sun, L., Yu, G., Fu, X., Wang, C., Bao, F., Yue, Z., Zhao, Q., Wang, N., Cheng, X., Liu, H., & Zhang, F. (2018). Discovery of 4 exonic and 1 intergenic novel susceptibility loci for leprosy. *Clin Genet*, *94*(2), 259-263. https://doi.org/10.1111/cge.13376

Wells, J. C. K., & Stock, J. T. (2020). Life History Transitions at the Origins of Agriculture: A Model for Understanding How Niche Construction Impacts Human Growth, Demography and Health. *Front Endocrinol (Lausanne)*, *11*, 325. https://doi.org/10.3389/fendo.2020.00325

Whitfield, J. B. (2002). Alcohol dehydrogenase and alcohol dependence: variation in genotype-associated risk between populations. *The American Journal of Human Genetics*, *71*(5), 1247-1250.

Wilson, S. (2020). miceRanger: multiple imputation by chained equations with random forests. *R package version*, *1*(5).

Yang, L., Chen, J. H., Lv, J., Wu, Q., Xu, T., Zhang, H., Liu, Q. H., & Yang, H. K. (2012). Rice protein improves adiposity, body weight and reduces lipids level in rats through modification of

triglyceride metabolism. *Lipids Health Dis*, *11*, 24. https://doi.org/10.1186/1476-511X-11-24

Zakhari, S. (2006). Overview: how is alcohol metabolized by the body? *Alcohol research & health*, *29*(4), 245.

Zhang, X., Sun, A., & Ge, J. (2021). Origin and Spread of the ALDH2 Glu504Lys Allele. *Phenomics*, *1*(5), 222-228. https://doi.org/10.1007/s43657-021-00017-y

Zhou, H., Kalayasiri, R., Sun, Y., Nunez, Y. Z., Deng, H. W., Chen, X. D., Justice, A. C., Kranzler, H. R., Chang, S., Lu, L., Shi, J., Sanichwankul, K., Mutirangura, A., Malison, R. T., & Gelernter, J. (2022). Genome-wide meta-analysis of alcohol use disorder in East Asians. *Neuropsychopharmacology*, *47*(10), 1791-1797. https://doi.org/10.1038/s41386-022-01265-w

Zhu, C., Talhelm, T., Li, Y., Chen, G., Zhu, J., & Wang, J. (2021). Relationship between rice farming and polygenic scores potentially linked to agriculture in China. *R Soc Open Sci*, *8*(8), 210382. https://doi.org/10.1098/rsos.210382

Figures

Figure 1 World-wide *ADH1B*2* frequency. (A) Density plot of frequency distribution of *ADH1B*2*, (B) QQ-plot of frequency distribution of *ADH1B*2* compared to normal distribution, and (C) National distribution of *ADH1B*2* frequency.

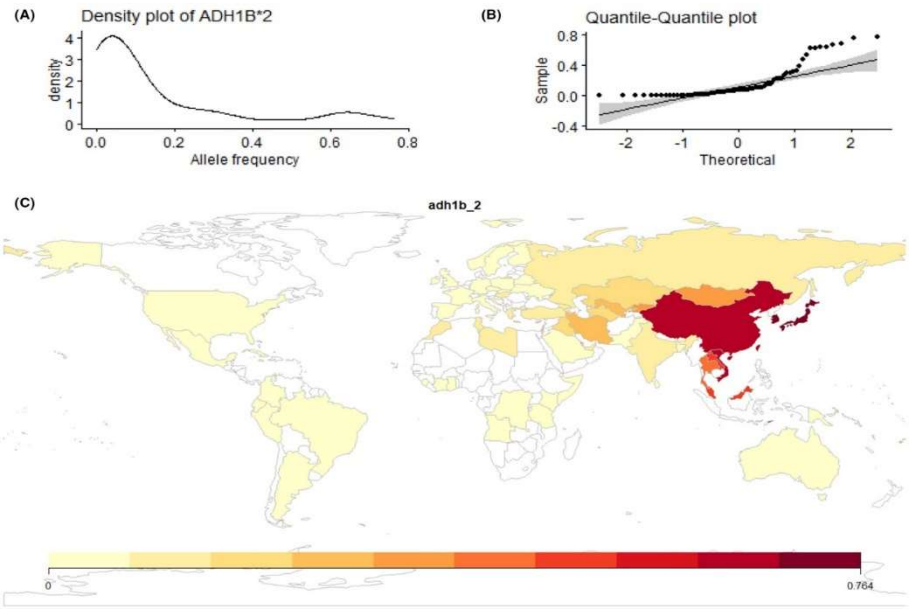


Figure 2 World-wide *ALDH2*2* frequency. (A) Density plot of frequency distribution of *ALDH2*2*, (B) QQ-plot of frequency distribution of *ALDH2*2* compared to normal distribution, and (C) National distribution of *ALDH2*2* frequency.

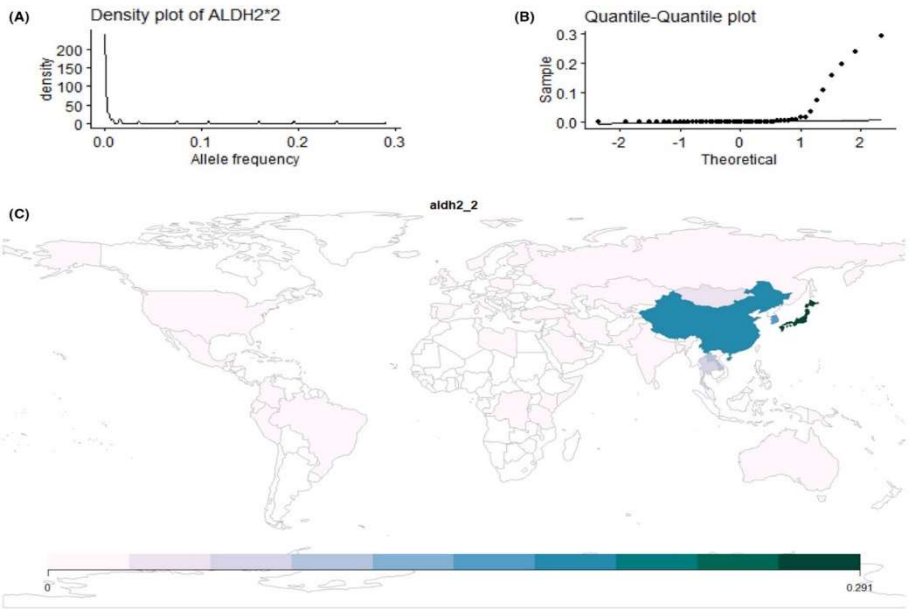


Figure 3 Correlation matrix for Kendall test (upper matrix reports Bonferroni adjusted p values). Asterisks meaning: ***p < 0.001, **p < 0.01, *p < 0.05.

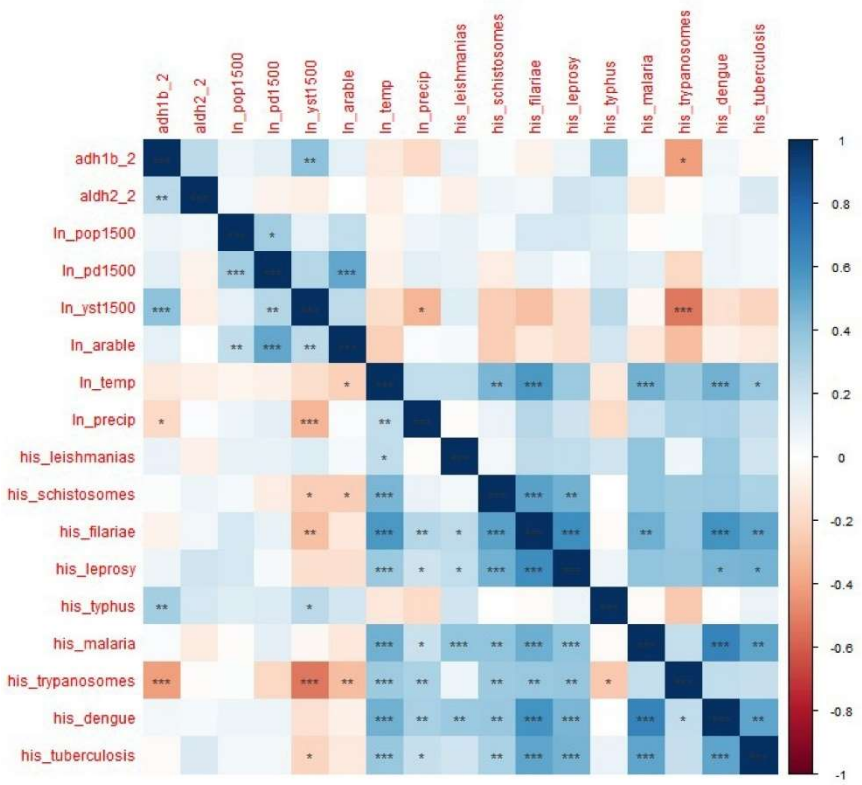


Figure 4 Principal component analysis on significant values. (A) Percentage of variance explained by each principal component, (B) contribution to the first pair of principal components of each variable, (C) plotting of vectors describing the variability explained by the first pair of principal components of each variable.

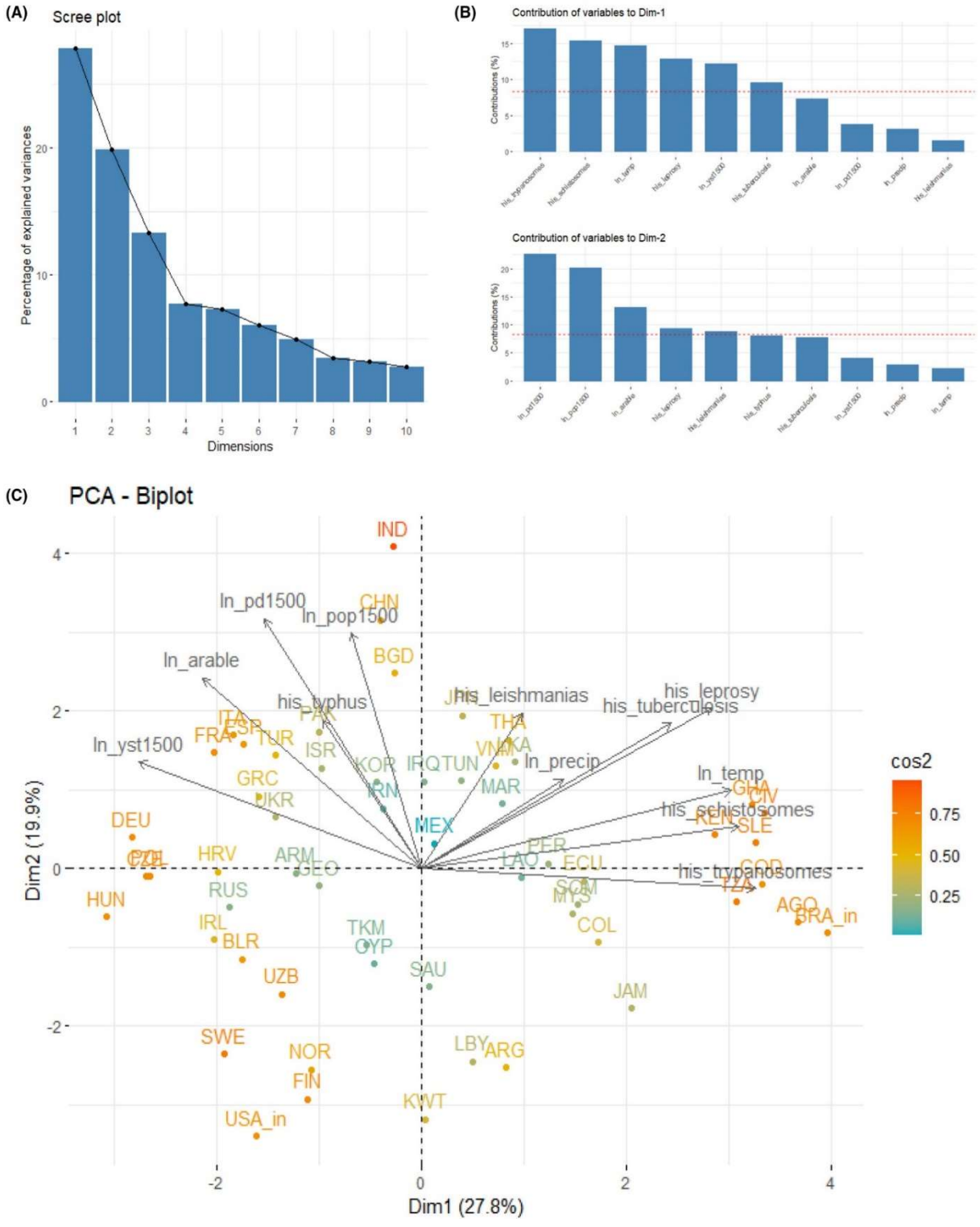
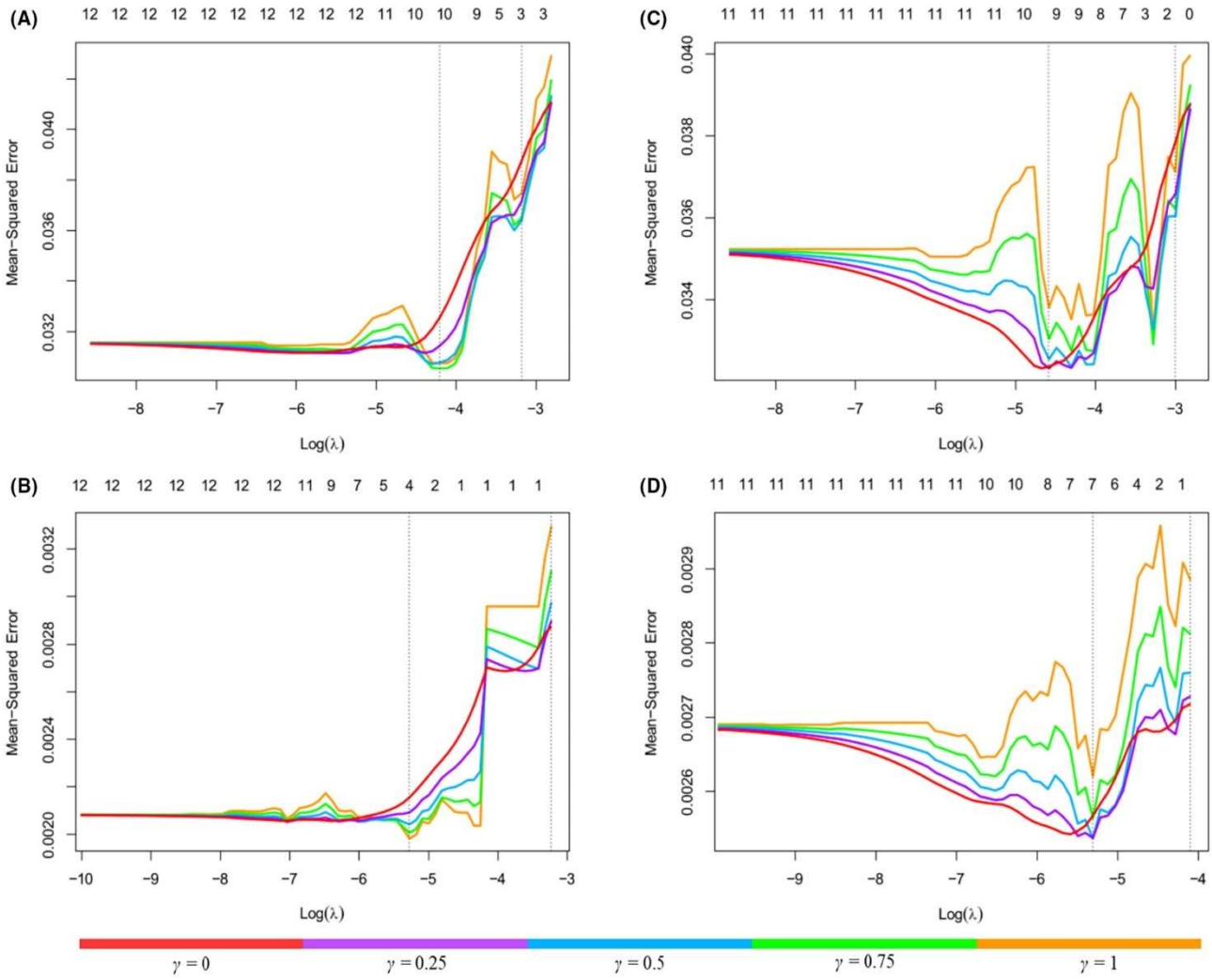


Figure 5 Relaxed Lasso regression model. (A) Elastic net for *ADH1B*2* + *ALDH2*2* = y, (B) Elastic net for *ADH1B*2* = xi; *ALDH2*2* = y, (C) Elastic net for *ADH1B*2* = y, (D) Elastic net for *ALDH2*2* = y. The numbers at the top of each panel refer to the number of predictors with a nonzero coefficient.



Tables

Table 1 Set of variables from Fedderke et al. (2017).

Name	Description
fstdist1500_eth	Genetic distance to Ethiopia (1500 match)
fstdist1500_uk	Genetic distance to the U.K. (1500 match)
efrac	Ethnic fractionalization index
ln_pop1500	Population in 1,500 CE LN
ln_pd1500	Population density in 1,500 CE LN
ln yst	Neolithic transition timing until the year 1,500 CE LN
ln_arable	Percentage of arable land LN
ln_temp	Temperature LN
ln_precip	Precipitation LN
his_leishmanias	Index of historical prevalence of leishmaniasis
his_schistosomes	Index of historical prevalence of schistosomes
his_filariae	Index of historical prevalence of filariae
his_leprosy	Index of historical prevalence leprosy
his_typhus	Index of historical prevalence typhus
his_malaria	Index of historical prevalence malaria
his_trypanosomes	Index of historical prevalence of trypanosomes
his_dengue	Index of historical prevalence of dengue
his_tuberculosis	Index of historical prevalence tuberculosis

Table 2 Phenotypic traits associated to *ADH1B*2*.

Class	Code	Description	N	info	effect	SE	P	FDR
biom	INI10030770	IGF-1	423,324	beta	-0.0895	0.0067	2.78E-38	2.89E-36
biom	INI10030610	Alkaline phosphatase	425,779	beta	0.0754	0.0066	6.07E-27	6.31E-25
biom	INI20030780	LDL cholesterol	424,960	beta	0.0073	0.0066	3.53E-25	3.67E-23
dis_out	HC328	Gout	9,506	or	0.66113	0.0391	3.13E-23	3.26E-21
dis_out	HC1215	TTE gout	15,378	or	0.71527	0.0322	2.55E-22	2.65E-20
biom	INI10030670	Urea	425,469	beta	-0.0666	0.0066	1.02E-20	1.06E-18
dis_out	HC719	TTE mental and behavioural disorders due to use of alcohol	8,788	or	2.21	0.0728	1.40E-20	1.46E-18
biom	INI10030860	Total protein	389,120	beta	0.0649	0.0069	8.53E-18	8.87E-16
biom	INI20030690	Cholesterol	425,743	beta	0.0598	0.0066	1.95E-16	2.02E-14
biom	INI10030880	Urate	425,235	beta	-0.0584	0.0066	1.39E-15	1.45E-13
biom	INI20030640	Apolipoprotein B	423,632	beta	0.0539	0.0066	5.42E-13	5.63E-11
dis_out	HC215	Hypertension	153,912	or	1.1	0.0144	1.81E-11	1.88E-09
biom	INI10030890	Vitamin D	406,303	beta	-0.0484	0.0068	9.89E-10	1.03E-07
dis_out	HC273	Essential hypertension	102,191	or	1.08177	0.1614	1.70E-08	1.77E-06
antro	INI23112	Leg fat mass R	442,807	beta	0.034	0.0051	3.20E-08	3.33E-06
antro	INI23099	Gamma glutamyltransferase	442,599	beta	0.0317	0.0049	7.37E-08	7.66E-06
biom	INI10030730	Body fat percentuage	425,539	beta	0.0431	0.0066	7.68E-08	7.99E-06
antro	INI23116	Whole body fat index	442,781	beta	0.0327	0.0051	1.10E-07	1.15E-05
antro	INI23100	Leg fat mass L	442,115	beta	0.0403	0.0063	1.14E-07	1.19E-05
antro	INI23111	Leg fat percentuage R	442,814	beta	0.0253	0.004	2.80E-07	2.91E-05
antro	INI23115	Non-albumin protein	442,786	beta	0.0247	0.004	4.62E-07	4.80E-05
biom	INI30030860	Leg fat percentuage L	389,120	beta	0.0432	0.0069	4.66E-07	4.84E-05
dis_out	HC932	TTE essential (primary hypertension)	112,342	or	1.08	0.16	5.11E-07	5.31E-05
antro	INI23127	Trunk fat percentuage	442,582	beta	0.0352	0.0059	1.86E-06	0.0001932
biom	INI10030600	Albumin	389,120	beta	0.0414	0.0069	2.34E-06	0.0002436
antro	INI23128	Trunk fat	442,558	beta	0.0383	0.0064	2.62E-06	0.0002729
antro	INI23123	BMI	442,686	beta	0.0287	0.0049	3.62E-06	0.000376
antro	INI23104	Arm fat percentuage L	442,840	beta	0.0378	0.0064	3.66E-06	0.0003802
antro	INI23124	Sodium in urine	442,647	beta	0.0367	0.0063	5.30E-06	0.0005508
biom	INI10030530	Arm fat mass L	412,901	beta	-0.0393	0.0067	5.46E-06	0.0005682
antro	INI23120	Arm fat mass R	442,716	beta	0.0363	0.0063	7.08E-06	0.0007364
antro	INI23119	BMI	442,746	beta	0.028	0.0049	9.14E-06	0.0009506
antro	INI21001	Arm fat percentuage R	448,906	beta	0.0365	0.0064	9.33E-06	0.0009705

Table 3 Phenotypic traits associated to *ALDH2*2*.

Class	Code	Description	N	info	effect	SE	P	FDR
dis_out	HC156	Nasal polyps	6,546	or	7.10785	0.456	1.70E-05	0.000968
antro	INI48	Waist circumference	449,704	beta	-0.137	0.0319	1.78E-05	0.001013
antro	INI23127	Trunk fat percentage	442,582	beta	-0.144	0.0349	3.75E-05	0.002138
antro	INI23099	Body fat percentage	442,599	beta	-0.1183	0.0288	3.97E-05	0.002265
antro	INI23124	Arm fat mass L	442,647	beta	-0.1478	0.0361	4.32E-05	0.002462
antro	INI23104	BMI	442,840	beta	-0.1442	0.0361	6.53E-05	0.003723

Supplemental tables

Supplemental tables (S1-S4) are made available at the following link:

https://onlinelibrary.wiley.com/action/downloadSupplement?doi=10.1111%2Facer.15288&file=acer15288-sup-0001-TablesS1-S4.xlsx

Chapter 2

Shared transcriptomic signatures of Post-traumatic stress disorder and Alcohol use disorder in human brain

Abstract

Post-traumatic stress disorder (PTSD) is a psychiatric condition caused by maladaptive response to traumatic events. PTSD shows high comorbidity rates towards substance use disorders (SUD), with alcohol use disorder (AUD) being the most prominent one. Data obtained from genome wide association studies (GWAS) causally associated PTSD to AUD, the precise nature of molecular mechanisms beneath this association stay elusive. Taking advantage of PubMed and Google Scholar databases; Gene Expression Omnibus repositories we collected transcriptomic data for PTSD and AUD obtained from postmortem brain tissues through bulk RNA-seq technology. Following data harmonizing, differently expressed genes (DEGs) in dorsolateral prefrontal cortex (dlPFC) and basolateral amygdala (BLA) were analyzed using Metscape, a platform to individuate shared enriched pathways and protein-protein integrated networks made publicly available on-line. Across several of shared enrichment patterns, our results underline the role of family A (rhodopsin-like) G-protein coupled receptors (GPCRs), and tyrosine kinases receptors (RTKs) and their clinical affordance in both AUD and PTSD.

Introduction

Post traumatic stress disorder (PTSD) is a psychiatric disorder that may arise as a pathological response to the exposure to a traumatic event (Association, 2013; Organization, 2019a). Due to the heterogenies in its symptomatology and prognosis, PTSD can include both cognitive and physical impairments (Marshall et al., 2019; Pacella et al., 2013). Among civilian population PTSD prevalence ranges between 0.56% to 6.67% across different European states (Wittchen et al., 2011) and 8.0% to 56.7% in the United States (Schein et al., 2021). Sexual assault is consistently reported as the main type of traumatic exposure by males and female patients in the United States (Schein et al., 2021). PTSD shows high comorbidity rates towards several conditions, such as major depression and other anxiety and substance disorders (SUD) (Lewis et al., 2019; Rahman et al., 2022), with alcohol being the most prominent consumed substance (Smith & Cottler, 2018). In US across different ethnic groups, the repeated exposition to trauma increased the odd ratio toward PTSD (OR = 1.61; $p < 0.001$) and alcohol use disorder (AUD) (OR = 1.81; $p < 0.001$), with interpersonal trauma being the best predictor of PTSD and subsequent AUD either after a single (OR = 2.54; $p < 0.001$) or multiple exposure (OR = 1.87 ; $p < 0.001$) (Dell'Aquila & Berle, 2023). Compared to other SUD, the co-occurrence of PTSD and AUD alone is mostly associated to emotional numbing while the increase of hyperarousal symptoms is mediated by other substances (Dworkin et al., 2018). In patients experiencing the co-occurrence of AUD and PTSD, reductions in hyperarousal levels precede positive changes in patterns of alcohol use and the reduction of impairments across all disrupted domains (Back et al., 2006). Evidence on alcohol assumption effecting all the clusters of PTSD symptoms reinforces hypothesis describing AUD as a self-medication response to trauma (Smith & Cottler, 2018). It is important to note that AUD is also associated with the increased propensity towards risk-taking behaviors in decision making process (Ariesen et al., 2023). Therefore,

problematic alcohol use may itself be the trigger for those traumatic events responsible for the development of PTSD (Cho & Yang, 2023; Polimanti et al., 2017). Other possible explanations for the co-occurrence of PTSD and AUD are based environmental factors, such as childhood neglect, or on molecular interactions underlying different styles of emotional learning (Maria-Rios & Morrow, 2020; Schimmenti et al., 2022). A moderate yet significant positive genomic correlation was observed between PTSD and AUD (Sheerin et al., 2020). Data from genome wide association studies (GWAS) confirmed the specific causal effect of PTSD on AUD but not on the number of alcoholic drinks consumed per week (Bountress et al., 2021). The nature of molecular mechanisms outlined by genomic variants shared between PTSD and AUD stay elusive due to their emerging yet diverse pattern of association (Bountress et al., 2022; Bountress et al., 2023). Transcriptomic profiling of brain tissues associate to stress responsivity and reward seeking, highlighted the modulatory role of inflammatory response and dopaminergic, adrenocorticotropic, GABAergic, and glutamatergic pathways in the pathophysiology of PTSD and AUD (Govindula et al., 2023; Maria-Rios & Morrow, 2020). To further explore shared transcriptomic signatures across PTSD to AUD, the present study employs data obtained through bulk RNA sequencing (RNAseq) technology for profiling gene expression. To our knowledge the present study represents the first attempt to directly compare the genome wide transcriptomic profile of subjects affected by either PTSD or AUD through a systematic query across multiple data repositories.

Materials and Methods

Data collection

Between July 2022 and July 2023, we conducted a systematic search for data and literature, using PubMed and Google Scholar databases. Gene Expression Omnibus repositories (Barrett et al., 2013) were also interrogated. To ensure an extensive data retrieval while accurately trawling through all three databases, two specific Boolean queries for PTSD and AUD were formulated as it respectively follows: ((expression OR RNA) AND Brain AND PTSD), ((expression OR RNA) AND (Brain OR (Alcohol OR AUD). Results were filtered following the following selection criteria: 1) Being a case-control study 2) Employing RNA sequencing in post-mortem human brain tissue 3) Match at least one brain region from the other phenotype. Following the acquisition of the complete lists of differentially expressed genes (DEGs) either from open repositories, full-text articles or through direct communication with investigators; we excluded data whose original sample demographics exhibited overlaps to avoid any potential bias in further analysis.

Meta Analysis

DEGs lists obtained relative to same brain areas and phenotype were metanalyzed combining their unadjusted p values through the Fisher Z weighted test using sample sizes as weights using the *combine.test* function including the *z.transform* option from the *survcomp* package (Schroder et al., 2011). If needed, unadjusted p values were converted to 1-tailed p values according to Log2 fold change value (log2fc) to suit Fisher's test assumptions. Fold change values were combined through a weighted mean using sample size with the "*weighted.mean*" R function.

Metascape analysis

Functional enrichment analysis was performed using Metascape (Zhou et al., 2019), an online platform that provides a rigorous pipeline for multi-platform OMICs data analysis and interpretation that was employed as described in the following paragraphs. Only DEGs passing nominal significant threshold ($p \leq 0.05$) were included in Metascape analysis. Gene annotation was performed for dbGap (NCBI) (Mailman et al., 2007), GWAS (NHGRI-EBI) (Sollis et al., 2023), Human Phenotype Ontology (HPO) (Gargano et al., 2024) libraries, merging one-to-many relations to absorb most redundancies into representative clusters. Enrichment analysis was performed on Canonical Pathways, Reactome Gene Sets (Milacic et al., 2024), KEGG Pathway (Kanehisa & Goto, 2000) and WikiPathways (Agrawal et al., 2024) setting thresholds on nominal significance ($p \leq 0.01$) applying a cutoff on minimum enrichment levels (1.5) and overlaps (3). Data from STRING (Szklarczyk et al., 2019) and BioGrid (Oughtred et al., 2021) databases was leveraged to extract relevant protein complexes embedded within networks of enriched genes. Through mature complex identification algorithm (MCODE) (Bader & Hogue, 2003) included in Metscape's Protein-protein Interaction Enrichment Analysis, subsets of proteins forming physical interactions with at least one other member were used to build protein networks containing between 3 and 500 proteins. Pathway and process enrichment analysis has been then independently applied to each MCODE component selecting the three best-scoring terms by p-value as their corresponding functional description.

Results

Data collection and harmonization

Transcriptomic data for AUD were obtained from specimens coming from the same cohort of European ancestry built by the New South Wales Brain Tissue Resource Centre (NSWBTRC)

(Sutherland et al., 2016). Due to sample overlaps reported by the NSWBTRC only data from Kapoor and colleagues (Kapoor et al., 2019) and Van Booven and colleagues (Van Booven et al., 2021) were employed in our analysis. Subjects from 3 studies identified for PTSD showed primarily white non-Hispanic or African American ancestry. Approximately 50% of the sample analyzed in Logue and colleagues (Logue et al., 2021) overlaps with that reported by Girgenti and colleagues (Girgenti & Duman, 2018). Considered the sample size of Logue and colleagues (Logue et al., 2021) was lower than Girgenti and colleagues (Girgenti & Duman, 2018), and that it was not possible to obtain full data access it was excluded from our analysis. Data provided by Jaffe and colleagues (Jaffe et al., 2022) accounting for the co-occurrence of major depressive disorder and PTSD was also excluded from our analysis. Brain areas whose transcriptomic profile was not examined across both phenotypes were removed leaving 2 DEGs lists for basolateral amygdala (BLA) and 3 for dorsolateral prefrontal cortex (dlPFC). Data obtained from dlPFC by Girgenti and colleagues (Girgenti & Duman, 2018) and Jaffe and colleagues (Jaffe et al., 2022) were combined in a single list of 14414 DEGs though Fisher Z weighted test. Full details on sample demographic are reported in Table 1. After filtering non-nominally significant DEGs, they were combined according to Metscape guidelines into a single operative dataset of 2523 gene entries. A total of 238 genes were dysregulated in AUD while 52 in PTSD, shared DEGs across dlPFC were 56 while 139 for BLA; carbonic anhydrase 12 (*CA12*) was the only DEG present in all lists (Figure 1A).

Metascape analysis

Following annotation and enrichment analysis, 101 terms were hierarchically clustered into 20 networks (Figure 1B). Most enriched pathways members within the 4 fully converging networks were: Signaling by GPCR (GO= R-HSA-372790, rank= 3, gene percentage= 9.59%, pvalue=

6.606e-60, q= 8.912e-57), Extracellular matrix organization (GO= R-HSA-1474244, rank= 4, gene percentage= 6.12%, pvalue= 4.677e-58, q= 4.677e-55), Signaling by Receptor Tyrosine Kinases (GO= R-HSA-9006934, rank= 6, gene percentage= 7.44%, pvalue = 6.760e-49, q= 3.890e-4), Hemostasis (GO= R-HSA-109582, rank = 8, gene percentage= 7.61%, pvalue= 3.388e-42, q= 1.513e-39) (Figure 2). With the exception of Signaling by GPCR, other fully converging networks clustered together with 6 others whose most enriched pathways were: Pleural mesothelioma (GO= WP5087, rank= 5, gene percentage=7.28%, pvalue= 5.88844e-57, q= 4.7863e-54), Proteoglycans in cancer (GO= hsa05205, rank= 10, gene percentage= 3.97%, pvalue=2.95121e-36, q= 6.60693e-34), NABA CORE MATRISOME (GO= M5884, rank= 11, gene percentage= 4.47%, pvalue=2.5704E-34, q= 5.49541E-32), MAPK signaling pathway (GO= hsa04010, rank= 14, gene percentage= 4.63%, pvalue= 5.01187E-34, q= 9.12011E-32), Intracellular signaling by second messengers (GO= R-HSA-9006925, rank= 15, gene percentage= 4.69%, pvalue= 8.51138E-34, q= 3.63078E-31), Calcium signaling pathway (GO= hsa04020, rank= 17, gene percentage= 4.02%, pvalue= 8.31764E-31, q= 3,01995E-28) (Figure 3). A total of 35 related protein-protein interaction networks embedded within genes enriched in Cytokine Signaling in Immune system (GO= R-HSA-1280215, pvalue= 5.01187E-21), Pathways in cancer (GO= hsa05200, pvalue= 7.94328E-20), and Extracellular matrix organization (GO= R-HSA-1474244, pvalue= 6.30957E-19) were divided into 13 MCODE components (Figure 4B-C). Most of protein-protein interactions were clustered within the first four MCODE components whose best scoring terms respectively were: Interferon alpha/beta signaling (MCODE1, GO= R-HSA-909733, pvalue= 6.30957E-20), Phosphatidyl inositol phosphate pathway (MCODE2, GO= WP5411, pvalue=3.98107E-14), Defective B3GALTL causes PpS (MCODE3, GO= R-HSA-5083635, pvalue= 1.99526E-09) and MAPK signaling pathway (MCODE4, GO= hsa04010, pvalue= 7.94328E-10) (Figure 4A).

Discussion

To our knowledge the present study represents the first attempt to leverage data coming from genome wide transcriptomic profiling of human brain tissues to analyze common alteration in gene expression due to PTSD and AUD. Our analysis focused on data from basolateral amygdala (BLA) and dorsolateral prefrontal cortex (dlPFC) which play a modulatory role in threat and fear processing and are both involved in AUD and PTSD (Alexandra Kredlow et al., 2022; Gorka, 2020; Laing et al., 2022). Although carbonic anhydrase 12 (*CA12*) was the only gene present across all DEGs lists, it was embedded in an integrated network related to Transport of small molecules (R-HSA-382551) whose pathways was enriched both in dlPFC and BLA for AUD, while for PTSD only in dlPFC. In previous GWAS *CA12* variants have been associated to decreased response to paliperidone (Li et al., 2017) and warfarin (Asiimwe et al., 2022), together with increased inflammation levels (Hill et al., 2017). While the use of antipsychotics drugs such as paliperidone are still under evaluation as potential treatments for PTSD (https://clinicaltrials.gov/study/NCT00766064), the disruption of vascular system and (Edmondson & von Kanel, 2017; Pathak et al., 2024) and the high concentration of pro-inflammatory biomarkers have been consistently reported as a determining risk factors for PTSD development in brain tissues of traumatized individuals (Lauten et al., 2024; Passos et al., 2015; Yetter et al., 2024); similarly to those affected by AUD (Coppens et al., 2019; Lanquetin et al., 2021; Moura et al., 2022). Across all analyzed tissues, the significance of molecular mechanisms related to stress response is outlined by integrated networks of enriched pathways targeting family A (rhodopsin-like) G-protein coupled receptors (GPCRs), and tyrosine kinases receptors (RTKs). Early stress has been associated to abnormal expression of 5-HTA1/2 GPCRs (Tiwari et al., 2021), which interestingly have been targeted as a potential novel treatment for AUD due to their potential modulatory role towards resumption of

alcohol seeking behaviors through the induction of neuroplasticity (Benvenuti et al., 2023; Calder & Hasler, 2023; Shao et al., 2021). Murine models shown that RTKs are involved both in fear and addiction (Kutlu et al., 2018), they modulate the use of ethanol intake as a coping strategy towards anxiety (Hamada & Lasek, 2020), triggering aberrant synaptic pruning that might actually worsen anxiety symptoms (Socodato et al., 2020). Extracellular matrix organization was the only integrated network of pathways enriched across analyzed tissues that contributed to protein-protein interaction enrichment. Alterations of intracellular enzymes necessary for cell growth, and extracellular matrix organization due to collagen turnover resulting in higher oxidative levels are one of the biomarkers of chronic stress-related diseases (Decker et al., 2021; Demir et al., 2016; Ercan et al., 2017; Obi et al., 2019). The application of MCODE algorithm to enriched protein-protein interaction shown that the 4 MCODE components embedded the largest interaction networks, which were either associated to Interferon alpha/beta signaling (MCODE01), phosphatidylinositol (MCODE02), O-glycosylation (MCODE 3), mitogen-activated protein kinase (MCODE4). While post-deployment US Marine experiencing PTSD shown an increase enrichment of networks associated to interferon signaling (Breen et al., 2015; Glatt et al., 2013), follow up studies dismissed these associations as the result of matrix degradation products leading to chronic inflammatory response (Muhie et al., 2023). Murine models for chronic restraint stress displayed similar results, showing that chronic stress increases Interferon alpha/beta receptor 1 (IFNAR-1) signaling causing the infiltration of macrophages into prefrontal cortex and behavioral abnormalities (Tripathi et al., 2021). The activation of phosphatidylinositol 3-kinase (PI3K)/protein kinase B (Akt) mitigates damages related to disruption of inflammatory response altering the functioning of cardiovascular system (Manning & Toker, 2017), which is indeed characterized by increased lipid peroxidation (Atli et al., 2016). Murine models also show that the consolidation of fear

memories causes an increased activation of PI3K/Akt pathways specific to BLA (Knox et al., 2021). O-glycosylation is a common posttranslational modification of protein implied in several neurodegenerative and neurodevelopment conditions (Pradeep et al., 2023). Among O-glicans, O-GlcNAc modification exhibits rapid and transient upregulation causing stress-induced hyperglycemia and upregulation of immune-response genes while providing energy for high-energy organs, such as the brain and heart (Xue et al., 2024). Interestingly, forebrain specific O-linked GlcNAc transferase conditional knockout mouse models displayed impaired amygdala-dependent cued freezing behavior in the open field, fear conditioning in behavior tests, excessive grooming, impaired nest building, reduced exploration, and increased anxiety (A. C. Wang et al., 2016). The enrichment of mitogen-activated protein kinase (MAPK) pathways has been reported in studies on PTSD employing monozygotic twins (Bainomugisa et al., 2021). Studies on metabolic dysregulation caused by elevated inflammation and oxidative stress in soldiers experiencing PTSD shown that, together with impaired glycolytic processes, the interplay between of MAPK and PI3K-AKT pathways plays crucial role in the dysregulation of insulin signaling which exacerbate oxidative stress derived impairments contributing to persistent PTSD and related medical comorbidities (Muhie et al., 2023). Being highly polygenic the causal relationship linking PTSD to AUD stays elusive (Bountress et al., 2022; Bountress et al., 2023; Bountress et al., 2021; Sheerin et al., 2020). Pro-inflammatory biomarkers have been previously associated to increased alcohol consumption and anxiety (Kazmi et al., 2022), our results point at molecular pathways involved in extracellular matrix organization in response to oxidative stress as the main transcriptomic signature shared between PTSD and AUD. While the role of increased oxidative stress has been consistently reported for both analyzed traits, some limitations should be addressed. The use of bulk RNA-seq, the certified overlaps across samples provided by NSWBTRC, and the huge diversity in tissues whose transcriptomic profiles

were available for both traits, limited the range of our analysis. Concerns over possible bias due to PTSD cohort including multiple ancestries should be considered at the light converging results from similar analysis on cross-species of PTSD transcriptomic profiles limited to European descendants (Nunez-Rios et al., 2023). Our results also reinforce the clinical affordance of serotonin GPCRs as molecular targets for AUD, due to their role converging enrichment in and across PTSD transcriptome; further research on their interplay with tyrosine kinases receptors within the context of neuroplasticity in both disorders is advised.

References

Agrawal, A., Balci, H., Hanspers, K., Coort, S. L., Martens, M., Slenter, D. N., Ehrhart, F., Digles, D., Waagmeester, A., Wassink, I., Abbassi-Daloii, T., Lopes, E. N., Iyer, A., Acosta, J. M., Willighagen, L. G., Nishida, K., Riutta, A., Basaric, H., Evelo, C. T., . . . Pico, A. R. (2024). WikiPathways 2024: next generation pathway database. *Nucleic Acids Res*, *52*(D1), D679-D689. https://doi.org/10.1093/nar/gkad960

Alexandra Kredlow, M., Fenster, R. J., Laurent, E. S., Ressler, K. J., & Phelps, E. A. (2022). Prefrontal cortex, amygdala, and threat processing: implications for PTSD. *Neuropsychopharmacology*, *47*(1), 247-259. https://doi.org/10.1038/s41386-021-01155-7

Ariesen, A. D., Neubert, J. H., Gaastra, G. F., Tucha, O., & Koerts, J. (2023). Risky Decision-Making in Adults with Alcohol Use Disorder-A Systematic and Meta-Analytic Review. *J Clin Med*, *12*(8). https://doi.org/10.3390/jcm12082943

Asiimwe, I. G., Blockman, M., Cohen, K., Cupido, C., Hutchinson, C., Jacobson, B., Lamorde, M., Morgan, J., Mouton, J. P., Nakagaayi, D., Okello, E., Schapkaitz, E., Sekaggya-Wiltshire, C., Semakula, J. R., Waitt, C., Zhang, E. J., Jorgensen, A. L., & Pirmohamed, M. (2022). A genome-wide association study of plasma concentrations of warfarin enantiomers and metabolites in sub-Saharan black-African patients. *Front Pharmacol*, *13*, 967082. https://doi.org/10.3389/fphar.2022.967082

Association, A. P. (2013). *Diagnostic and Statistical Manual of Mental Disorders* (5th ed.). https://doi.org/https://doi.org/10.1176/appi.books.9780890425596

Atli, A., Bulut, M., Bez, Y., Kaplan, I., Ozdemir, P. G., Uysal, C., Selcuk, H., & Sir, A. (2016). Altered lipid peroxidation markers are related to post-traumatic stress disorder (PTSD) and not trauma itself in earthquake survivors. *Eur Arch Psychiatry Clin Neurosci*, *266*(4), 329-336. https://doi.org/10.1007/s00406-015-0638-5

Back, S. E., Brady, K. T., Sonne, S. C., & Verduin, M. L. (2006). Symptom improvement in co-occurring PTSD and alcohol dependence. *J Nerv Ment Dis*, *194*(9), 690-696. https://doi.org/10.1097/01.nmd.0000235794.12794.8a

Bader, G. D., & Hogue, C. W. (2003). An automated method for finding molecular complexes in large protein interaction networks. *BMC Bioinformatics*, *4*, 2. https://doi.org/10.1186/1471-2105-4-2

Bainomugisa, C. K., Sutherland, H. G., Parker, R., McRae, A. F., Haupt, L. M., Griffiths, L. R., Heath, A., Nelson, E. C., Wright, M. J., Hickie, I. B., Martin, N. G., Nyholt, D. R., & Mehta, D. (2021). Using Monozygotic Twins to Dissect Common Genes in Posttraumatic Stress Disorder and Migraine. *Front Neurosci*, *15*, 678350. https://doi.org/10.3389/fnins.2021.678350

Barrett, T., Wilhite, S. E., Ledoux, P., Evangelista, C., Kim, I. F., Tomashevsky, M., Marshall, K. A., Phillippy, K. H., Sherman, P. M., Holko, M., Yefanov, A., Lee, H., Zhang, N., Robertson, C. L., Serova, N., Davis, S., & Soboleva, A. (2013). NCBI GEO: archive for functional genomics data sets--update. *Nucleic Acids Res*, *41*(Database issue), D991-995. https://doi.org/10.1093/nar/gks1193

Benvenuti, F., Colombo, D., Soverchia, L., Cannella, N., Domi, E., & Ciccocioppo, R. (2023). Psilocybin prevents reinstatement of alcohol seeking by disrupting the reconsolidation of alcohol-related memories. *Psychopharmacology (Berl)*, *240*(7), 1521-1530. https://doi.org/10.1007/s00213-023-06384-w

Bountress, K. E., Brick, L. A., Sheerin, C., Grotzinger, A., Bustamante, D., Hawn, S. E., Gillespie, N., Kirkpatrick, R. M., Kranzler, H., Morey, R., Edenberg, H. J., Maihofer, A. X., Disner, S., Ashley-Koch, A., Peterson, R., Lori, A., Stein, D. J., Kimbrel, N., Nievergelt, C., . . . Amstadter, A. B. (2022). Alcohol use and alcohol use disorder differ in their genetic relationships with PTSD: A genomic structural equation modelling approach. *Drug Alcohol Depend*, *234*, 109430. https://doi.org/10.1016/j.drugalcdep.2022.109430

Bountress, K. E., Bustamante, D., Subbie-Saenz de Viteri, S., Chatzinakos, C., Sheerin, C., Daskalakis, N. P., Edenberg, H. J., Psychiatric Genomics Consortium Posttraumatic Stress Disorder Working, G., Peterson, R. E., Webb, B. T., Meyers, J., & Amstadter, A. (2023). Differences in genetic correlations between posttraumatic stress disorder and alcohol-related problems phenotypes compared to alcohol consumption-related phenotypes. *Psychol Med*, *53*(12), 5767-5777. https://doi.org/10.1017/S0033291722002999

Bountress, K. E., Wendt, F., Bustamante, D., Agrawal, A., Webb, B., Gillespie, N., Edenberg, H., Sheerin, C., Johnson, E., Psychiatric Genomics Consortium Posttraumatic Stress Disorder Working, G., Polimanti, R., & Amstadter, A. (2021). Potential causal effect of posttraumatic stress disorder on alcohol use disorder and alcohol consumption in

individuals of European descent: A Mendelian Randomization Study. *Alcohol Clin Exp Res*, *45*(8), 1616-1623. https://doi.org/10.1111/acer.14649

Breen, M. S., Maihofer, A. X., Glatt, S. J., Tylee, D. S., Chandler, S. D., Tsuang, M. T., Risbrough, V. B., Baker, D. G., O'Connor, D. T., Nievergelt, C. M., & Woelk, C. H. (2015). Gene networks specific for innate immunity define post-traumatic stress disorder. *Mol Psychiatry*, *20*(12), 1538-1545. https://doi.org/10.1038/mp.2015.9

Calder, A. E., & Hasler, G. (2023). Towards an understanding of psychedelic-induced neuroplasticity. *Neuropsychopharmacology*, *48*(1), 104-112. https://doi.org/10.1038/s41386-022-01389-z

Cho, H. S., & Yang, Y. (2023). Relationship Between Alcohol Consumption and Risky Sexual Behaviors Among Adolescents and Young Adults: A Meta-Analysis. *Int J Public Health*, *68*, 1605669. https://doi.org/10.3389/ijph.2023.1605669

Coppens, V., Morrens, M., Destoop, M., & Dom, G. (2019). The Interplay of Inflammatory Processes and Cognition in Alcohol Use Disorders-A Systematic Review. *Front Psychiatry*, *10*, 632. https://doi.org/10.3389/fpsyt.2019.00632

Decker, A. M., Kapila, Y. L., & Wang, H. L. (2021). The psychobiological links between chronic stress-related diseases, periodontal/peri-implant diseases, and wound healing. *Periodontol 2000*, *87*(1), 94-106. https://doi.org/10.1111/prd.12381

Dell'Aquila, A., & Berle, D. (2023). Predictors of alcohol and substance use among people with post-traumatic stress disorder (PTSD): findings from the NESARC-III study. *Soc Psychiatry Psychiatr Epidemiol*, *58*(10), 1509-1522. https://doi.org/10.1007/s00127-023-02472-6

Demir, S., Bulut, M., Atli, A., Kaplan, I., Kaya, M. C., Bez, Y., Ozdemir, P. G., & Sir, A. (2016). Decreased Prolidase Activity in Patients with Posttraumatic Stress Disorder. *Psychiatry Investig*, *13*(4), 420-426. https://doi.org/10.4306/pi.2016.13.4.420

Dworkin, E. R., Wanklyn, S., Stasiewicz, P. R., & Coffey, S. F. (2018). PTSD symptom presentation among people with alcohol and drug use disorders: Comparisons by substance of abuse. *Addict Behav*, *76*, 188-194. https://doi.org/10.1016/j.addbeh.2017.08.019

Edmondson, D., & von Kanel, R. (2017). Post-traumatic stress disorder and cardiovascular disease. *Lancet Psychiatry*, *4*(4), 320-329. https://doi.org/10.1016/S2215-0366(16)30377-7

Ercan, A. C., Bahceci, B., Polat, S., Cenker, O. C., Bahceci, I., Koroglu, A., Sahin, K., & Hocaoglu, C. (2017). Oxidative status and prolidase activities in generalized anxiety disorder. *Asian J Psychiatr*, *25*, 118-122. https://doi.org/10.1016/j.ajp.2016.10.008

Gargano, M. A., Matentzoglu, N., Coleman, B., Addo-Lartey, E. B., Anagnostopoulos, A. V., Anderton, J., Avillach, P., Bagley, A. M., Bakstein, E., Balhoff, J. P., Baynam, G., Bello, S. M., Berk, M., Bertram, H., Bishop, S., Blau, H., Bodenstein, D. F., Botas, P., Boztug, K., . . . Robinson, P. N. (2024). The Human Phenotype Ontology in 2024: phenotypes around the world. *Nucleic Acids Res*, *52*(D1), D1333-D1346. https://doi.org/10.1093/nar/gkad1005

Girgenti, M. J., & Duman, R. S. (2018). Transcriptome Alterations in Posttraumatic Stress Disorder. *Biol Psychiatry*, *83*(10), 840-848. https://doi.org/10.1016/j.biopsych.2017.09.023

Glatt, S. J., Tylee, D. S., Chandler, S. D., Pazol, J., Nievergelt, C. M., Woelk, C. H., Baker, D. G., Lohr, J. B., Kremen, W. S., Litz, B. T., Tsuang, M. T., & Marine Resiliency Study, I. (2013). Blood-based gene-expression predictors of PTSD risk and resilience among deployed marines: a pilot study. *Am J Med Genet B Neuropsychiatr Genet*, *162B*(4), 313-326. https://doi.org/10.1002/ajmg.b.32167

Gorka, S. M. (2020). Interpersonal trauma exposure and startle reactivity to uncertain threat in individuals with alcohol use disorder. *Drug Alcohol Depend*, *206*, 107727. https://doi.org/10.1016/j.drugalcdep.2019.107727

Govindula, A., Ranadive, N., Nampoothiri, M., Rao, C. M., Arora, D., & Mudgal, J. (2023). Emphasizing the Crosstalk Between Inflammatory and Neural Signaling in Post-traumatic Stress Disorder (PTSD). *J Neuroimmune Pharmacol*, *18*(3), 248-266. https://doi.org/10.1007/s11481-023-10064-z

Hamada, K., & Lasek, A. W. (2020). Receptor Tyrosine Kinases as Therapeutic Targets for Alcohol Use Disorder. *Neurotherapeutics*, *17*(1), 4-16. https://doi.org/10.1007/s13311-019-00795-4

Hill, L. J., Di Pietro, V., Hazeldine, J., Davies, D., Toman, E., Logan, A., & Belli, A. (2017). Cystatin D (CST5): An ultra-early inflammatory biomarker of traumatic brain injury. *Sci Rep*, *7*(1), 5002. https://doi.org/10.1038/s41598-017-04722-5

Jaffe, A. E., Tao, R., Page, S. C., Maynard, K. R., Pattie, E. A., Nguyen, C. V., Deep-Soboslay, A., Bharadwaj, R., Young, K. A., Friedman, M. J., Williamson, D. E., Traumatic Stress Brain

Research, G., Shin, J. H., Hyde, T. M., Martinowich, K., & Kleinman, J. E. (2022). Decoding Shared Versus Divergent Transcriptomic Signatures Across Cortico-Amygdala Circuitry in PTSD and Depressive Disorders. *Am J Psychiatry*, *179*(9), 673-686. https://doi.org/10.1176/appi.ajp.21020162

Kanehisa, M., & Goto, S. (2000). KEGG: kyoto encyclopedia of genes and genomes. *Nucleic Acids Res*, *28*(1), 27-30. https://doi.org/10.1093/nar/28.1.27

Kapoor, M., Wang, J. C., Farris, S. P., Liu, Y., McClintick, J., Gupta, I., Meyers, J. L., Bertelsen, S., Chao, M., Nurnberger, J., Tischfield, J., Harari, O., Zeran, L., Hesselbrock, V., Bauer, L., Raj, T., Porjesz, B., Agrawal, A., Foroud, T., . . . Goate, A. (2019). Analysis of whole genome-transcriptomic organization in brain to identify genes associated with alcoholism. *Transl Psychiatry*, *9*(1), 89. https://doi.org/10.1038/s41398-019-0384-y

Kazmi, N., Wallen, G. R., Yang, L., Alkhatib, J., Schwandt, M. L., Feng, D., Gao, B., Diazgranados, N., Ramchandani, V. A., & Barb, J. J. (2022). An exploratory study of pro-inflammatory cytokines in individuals with alcohol use disorder: MCP-1 and IL-8 associated with alcohol consumption, sleep quality, anxiety, depression, and liver biomarkers. *Front Psychiatry*, *13*, 931280. https://doi.org/10.3389/fpsyt.2022.931280

Knox, D., Della Valle, R., Mohammadmirzaei, N., Shultz, B., Biddle, M., Farkash, A., Chamness, M., & Moulton, E. (2021). PI3K-Akt Signaling in the Basolateral Amygdala Facilitates Traumatic Stress Enhancements in Fear Memory. *Int J Neuropsychopharmacol*, *24*(3), 229-238. https://doi.org/10.1093/ijnp/pyaa083

Kutlu, M. G., Cole, R. D., Connor, D. A., Natwora, B., & Gould, T. J. (2018). Tyrosine receptor kinase B receptor activation reverses the impairing effects of acute nicotine on contextual fear extinction. *J Psychopharmacol*, *32*(3), 367-372. https://doi.org/10.1177/0269881118758305

Laing, P. A. F., Steward, T., Davey, C. G., Felmingham, K. L., Fullana, M. A., Vervliet, B., Greaves, M. D., Moffat, B., Glarin, R. K., & Harrison, B. J. (2022). Cortico-Striatal Activity Characterizes Human Safety Learning via Pavlovian Conditioned Inhibition. *J Neurosci*, *42*(25), 5047-5057. https://doi.org/10.1523/JNEUROSCI.2181-21.2022

Lanquetin, A., Leclercq, S., de Timary, P., Segobin, S., Naveau, M., Coulbault, L., Maccioni, P., Lorrai, I., Colombo, G., Vivien, D., Rubio, M., & Pitel, A. L. (2021). Role of inflammation in alcohol-related brain abnormalities: a translational study. *Brain Commun*, *3*(3), fcab154. https://doi.org/10.1093/braincomms/fcab154

Lauten, T. H., Natour, T., & Case, A. J. (2024). Innate and adaptive immune system consequences of post-traumatic stress disorder. *Auton Neurosci*, *252*, 103159. https://doi.org/10.1016/j.autneu.2024.103159

Lewis, S. J., Arseneault, L., Caspi, A., Fisher, H. L., Matthews, T., Moffitt, T. E., Odgers, C. L., Stahl, D., Teng, J. Y., & Danese, A. (2019). The epidemiology of trauma and post-traumatic stress disorder in a representative cohort of young people in England and Wales. *Lancet Psychiatry*, *6*(3), 247-256. https://doi.org/10.1016/S2215-0366(19)30031-8

Li, Q., Wineinger, N. E., Fu, D. J., Libiger, O., Alphs, L., Savitz, A., Gopal, S., Cohen, N., & Schork, N. J. (2017). Genome-wide association study of paliperidone efficacy. *Pharmacogenet Genomics*, *27*(1), 7-18. https://doi.org/10.1097/FPC.0000000000000250

Logue, M. W., Zhou, Z., Morrison, F. G., Wolf, E. J., Daskalakis, N. P., Chatzinakos, C., Georgiadis, F., Labadorf, A. T., Girgenti, M. J., Young, K. A., Williamson, D. E., Zhao, X., Grenier, J. G., Traumatic Stress Brain Research, G., Huber, B. R., & Miller, M. W. (2021). Gene expression in the dorsolateral and ventromedial prefrontal cortices implicates immune-related gene networks in PTSD. *Neurobiol Stress*, *15*, 100398. https://doi.org/10.1016/j.ynstr.2021.100398

Mailman, M. D., Feolo, M., Jin, Y., Kimura, M., Tryka, K., Bagoutdinov, R., Hao, L., Kiang, A., Paschall, J., Phan, L., Popova, N., Pretel, S., Ziyabari, L., Lee, M., Shao, Y., Wang, Z. Y., Sirotkin, K., Ward, M., Kholodov, M., . . . Sherry, S. T. (2007). The NCBI dbGaP database of genotypes and phenotypes. *Nat Genet*, *39*(10), 1181-1186. https://doi.org/10.1038/ng1007-1181

Manning, B. D., & Toker, A. (2017). AKT/PKB Signaling: Navigating the Network. *Cell*, *169*(3), 381-405. https://doi.org/10.1016/j.cell.2017.04.001

Maria-Rios, C. E., & Morrow, J. D. (2020). Mechanisms of Shared Vulnerability to Post-traumatic Stress Disorder and Substance Use Disorders. *Front Behav Neurosci*, *14*, 6. https://doi.org/10.3389/fnbeh.2020.00006

Marshall, G. N., Jaycox, L. H., Engel, C. C., Richardson, A. S., Dutra, S. J., Keane, T. M., Rosen, R. C., & Marx, B. P. (2019). PTSD symptoms are differentially associated with general distress and physiological arousal: Implications for the conceptualization and measurement of PTSD. *J Anxiety Disord*, *62*, 26-34. https://doi.org/10.1016/j.janxdis.2018.10.003

Milacic, M., Beavers, D., Conley, P., Gong, C., Gillespie, M., Griss, J., Haw, R., Jassal, B., Matthews, L., May, B., Petryszak, R., Ragueneau, E., Rothfels, K., Sevilla, C., Shamovsky, V., Stephan, R., Tiwari, K., Varusai, T., Weiser, J., . . . D'Eustachio, P. (2024). The Reactome Pathway Knowledgebase 2024. *Nucleic Acids Res*, *52*(D1), D672-D678. https://doi.org/10.1093/nar/gkad1025

Moura, H. F., Hansen, F., Galland, F., Silvelo, D., Rebelatto, F. P., Ornell, F., Massuda, R., Scherer, J. N., Schuch, F., Kessler, F. H., & von Diemen, L. (2022). Inflammatory cytokines and alcohol use disorder: systematic review and meta-analysis. *Braz J Psychiatry*, *44*(5), 548-556. https://doi.org/10.47626/1516-4446-2021-1893

Muhie, S., Gautam, A., Yang, R., Misganaw, B., Daigle, B. J., Jr., Mellon, S. H., Flory, J. D., Abu-Amara, D., Lee, I., Wang, K., Rampersaud, R., Consortium, P. S. B., Hood, L., Yehuda, R., Marmar, C. R., Wolkowitz, O. M., Ressler, K. J., Doyle, F. J., 3rd, Hammamieh, R., & Jett, M. (2023). Molecular signatures of post-traumatic stress disorder in war-zone-exposed veteran and active-duty soldiers. *Cell Rep Med*, *4*(5), 101045. https://doi.org/10.1016/j.xcrm.2023.101045

Nunez-Rios, D. L., Martinez-Magana, J. J., Nagamatsu, S. T., Krystal, J. H., Martinez-Gonzalez, K. G., Giusti-Rodriguez, P., & Montalvo-Ortiz, J. L. (2023). Cross-Species Convergence of Brain Transcriptomic and Epigenomic Findings in Posttraumatic Stress Disorder: A Systematic Review. *Complex Psychiatry*, *9*(1-4), 100-118. https://doi.org/10.1159/000529536

Obi, I. E., McPherson, K. C., & Pollock, J. S. (2019). Childhood adversity and mechanistic links to hypertension risk in adulthood. *Br J Pharmacol*, *176*(12), 1932-1950. https://doi.org/10.1111/bph.14576

Organization, W. H. (2019). 6B40 Post traumatic stress disorder. In *nternational statistical classification of diseases and related health problems (11th ed.)*. https://icd.who.int/browse/2024-01/mms/en#2070699808

Oughtred, R., Rust, J., Chang, C., Breitkreutz, B. J., Stark, C., Willems, A., Boucher, L., Leung, G., Kolas, N., Zhang, F., Dolma, S., Coulombe-Huntington, J., Chatr-Aryamontri, A., Dolinski, K., & Tyers, M. (2021). The BioGRID database: A comprehensive biomedical resource of curated protein, genetic, and chemical interactions. *Protein Sci*, *30*(1), 187-200. https://doi.org/10.1002/pro.3978

Pacella, M. L., Hruska, B., & Delahanty, D. L. (2013). The physical health consequences of PTSD and PTSD symptoms: a meta-analytic review. *J Anxiety Disord*, *27*(1), 33-46. https://doi.org/10.1016/j.janxdis.2012.08.004

Passos, I. C., Vasconcelos-Moreno, M. P., Costa, L. G., Kunz, M., Brietzke, E., Quevedo, J., Salum, G., Magalhaes, P. V., Kapczinski, F., & Kauer-Sant'Anna, M. (2015). Inflammatory markers in post-traumatic stress disorder: a systematic review, meta-analysis, and meta-regression. *Lancet Psychiatry*, *2*(11), 1002-1012. https://doi.org/10.1016/S2215-0366(15)00309-0

Pathak, G. A., Singh, K., Choi, K. W., Fang, Y., Kouakou, M. R., Lee, Y. H., Zhou, X., Fritsche, L. G., Wendt, F. R., Davis, L. K., & Polimanti, R. (2024). Genetic Liability to Posttraumatic Stress Disorder Symptoms and Its Association With Cardiometabolic and Respiratory Outcomes. *JAMA Psychiatry*, *81*(1), 34-44. https://doi.org/10.1001/jamapsychiatry.2023.4127

Polimanti, R., Amstadter, A. B., Stein, M. B., Almli, L. M., Baker, D. G., Bierut, L. J., Bradley, B., Farrer, L. A., Johnson, E. O., King, A., Kranzler, H. R., Maihofer, A. X., Rice, J. P., Roberts, A. L., Saccone, N. L., Zhao, H., Liberzon, I., Ressler, K. J., Nievergelt, C. M., . . . Psychiatric Genomics Consortium Posttraumatic Stress Disorder, W. (2017). A putative causal relationship between genetically determined female body shape and posttraumatic stress disorder. *Genome Med*, *9*(1), 99. https://doi.org/10.1186/s13073-017-0491-4

Pradeep, P., Kang, H., & Lee, B. (2023). Glycosylation and behavioral symptoms in neurological disorders. *Transl Psychiatry*, *13*(1), 154. https://doi.org/10.1038/s41398-023-02446-x

Rahman, S., Zammit, S., Dalman, C., & Hollander, A. C. (2022). Epidemiology of posttraumatic stress disorder: A prospective cohort study based on multiple nationwide Swedish registers of 4.6 million people. *Eur Psychiatry*, *65*(1), e60. https://doi.org/10.1192/j.eurpsy.2022.2311

Schein, J., Houle, C., Urganus, A., Cloutier, M., Patterson-Lomba, O., Wang, Y., King, S., Levinson, W., Guerin, A., Lefebvre, P., & Davis, L. L. (2021). Prevalence of post-traumatic stress disorder in the United States: a systematic literature review. *Curr Med Res Opin*, *37*(12), 2151-2161. https://doi.org/10.1080/03007995.2021.1978417

Schimmenti, A., Billieux, J., Santoro, G., Casale, S., & Starcevic, V. (2022). A trauma model of substance use: Elaboration and preliminary validation. *Addict Behav*, *134*, 107431. https://doi.org/10.1016/j.addbeh.2022.107431

Schroder, M. S., Culhane, A. C., Quackenbush, J., & Haibe-Kains, B. (2011). survcomp: an R/Bioconductor package for performance assessment and comparison of survival models. *Bioinformatics*, *27*(22), 3206-3208. https://doi.org/10.1093/bioinformatics/btr511

Shao, L. X., Liao, C., Gregg, I., Davoudian, P. A., Savalia, N. K., Delagarza, K., & Kwan, A. C. (2021). Psilocybin induces rapid and persistent growth of dendritic spines in frontal cortex in vivo. *Neuron*, *109*(16), 2535-2544 e2534. https://doi.org/10.1016/j.neuron.2021.06.008

Sheerin, C. M., Bountress, K. E., Meyers, J. L., Saenz de Viteri, S. S., Shen, H., Maihofer, A. X., Duncan, L. E., & Amstadter, A. B. (2020). Shared molecular genetic risk of alcohol dependence and posttraumatic stress disorder (PTSD). *Psychol Addict Behav*, *34*(5), 613-619. https://doi.org/10.1037/adb0000568

Smith, N. D., & Cottler, L. B. (2018). The epidemiology of post-traumatic stress disorder and alcohol use disorder. *Alcohol research: current reviews*, *39*(2), 113.

Socodato, R., Henriques, J. F., Portugal, C. C., Almeida, T. O., Tedim-Moreira, J., Alves, R. L., Canedo, T., Silva, C., Magalhaes, A., Summavielle, T., & Relvas, J. B. (2020). Daily alcohol intake triggers aberrant synaptic pruning leading to synapse loss and anxiety-like behavior. *Sci Signal*, *13*(650). https://doi.org/10.1126/scisignal.aba5754

Sollis, E., Mosaku, A., Abid, A., Buniello, A., Cerezo, M., Gil, L., Groza, T., Gunes, O., Hall, P., Hayhurst, J., Ibrahim, A., Ji, Y., John, S., Lewis, E., MacArthur, J. A. L., McMahon, A., Osumi-Sutherland, D., Panoutsopoulou, K., Pendlington, Z., . . . Harris, L. W. (2023). The NHGRI-EBI GWAS Catalog: knowledgebase and deposition resource. *Nucleic Acids Res*, *51*(D1), D977-D985. https://doi.org/10.1093/nar/gkac1010

Sutherland, G. T., Sheedy, D., Stevens, J., McCrossin, T., Smith, C. C., van Roijen, M., & Kril, J. J. (2016). The NSW brain tissue resource centre: Banking for alcohol and major neuropsychiatric disorders research. *Alcohol*, *52*, 33-39. https://doi.org/10.1016/j.alcohol.2016.02.005

Szklarczyk, D., Gable, A. L., Lyon, D., Junge, A., Wyder, S., Huerta-Cepas, J., Simonovic, M., Doncheva, N. T., Morris, J. H., Bork, P., Jensen, L. J., & Mering, C. V. (2019). STRING v11: protein-protein association networks with increased coverage, supporting functional discovery in genome-wide experimental datasets. *Nucleic Acids Res*, *47*(D1), D607-D613. https://doi.org/10.1093/nar/gky1131

Tiwari, P., Fanibunda, S. E., Kapri, D., Vasaya, S., Pati, S., & Vaidya, V. A. (2021). GPCR signaling: role in mediating the effects of early adversity in psychiatric disorders. *FEBS J*, *288*(8), 2602-2621. https://doi.org/10.1111/febs.15738

Tripathi, A., Whitehead, C., Surrao, K., Pillai, A., Madeshiya, A., Li, Y., Khodadadi, H., Ahmed, A. O., Turecki, G., Baban, B., & Pillai, A. (2021). Type 1 interferon mediates chronic stress-induced neuroinflammation and behavioral deficits via complement component 3-dependent pathway. *Mol Psychiatry*, *26*(7), 3043-3059. https://doi.org/10.1038/s41380-021-01065-6

Van Booven, D., Mengying, L., Sunil Rao, J., Blokhin, I. O., Dayne Mayfield, R., Barbier, E., Heilig, M., & Wahlestedt, C. (2021). Alcohol use disorder causes global changes in splicing in the human brain. *Transl Psychiatry*, *11*(1), 2. https://doi.org/10.1038/s41398-020-01163-z

Wang, A. C., Jensen, E. H., Rexach, J. E., Vinters, H. V., & Hsieh-Wilson, L. C. (2016). Loss of O-GlcNAc glycosylation in forebrain excitatory neurons induces neurodegeneration. *Proc Natl Acad Sci U S A*, *113*(52), 15120-15125. https://doi.org/10.1073/pnas.1606899113

Wittchen, H. U., Jacobi, F., Rehm, J., Gustavsson, A., Svensson, M., Jonsson, B., Olesen, J., Allgulander, C., Alonso, J., Faravelli, C., Fratiglioni, L., Jennum, P., Lieb, R., Maercker, A., van Os, J., Preisig, M., Salvador-Carulla, L., Simon, R., & Steinhausen, H. C. (2011). The size and burden of mental disorders and other disorders of the brain in Europe 2010. *Eur Neuropsychopharmacol*, *21*(9), 655-679. https://doi.org/10.1016/j.euroneuro.2011.07.018

Xue, Q., Ji, S., Xu, H., & Yu, S. (2024). O-GlcNAcylation: a pro-survival response to acute stress in the cardiovascular and central nervous systems. *Eur J Med Res*, *29*(1), 174. https://doi.org/10.1186/s40001-024-01773-z

Yetter, M. A., Fitzgerald, T. R., Philippi, C. L., & Bruce, S. E. (2024). Pro-inflammatory markers are related to cortical network connectivity in women exposed to interpersonal trauma with PTSD. *Behav Brain Res*, 114942. https://doi.org/10.1016/j.bbr.2024.114942

Zhou, Y., Zhou, B., Pache, L., Chang, M., Khodabakhshi, A. H., Tanaseichuk, O., Benner, C., & Chanda, S. K. (2019). Metascape provides a biologist-oriented resource for the analysis of systems-level datasets. *Nat Commun*, *10*(1), 1523. https://doi.org/10.1038/s41467-019-09234-6

Figures

Figure 1 Overlaps across analyzed lists of differently expressed genes (DEGs): (A) Gene level overlaps where each purple curve links identical DEGs; (B) Enrichment term level overlaps, where each blue curve links DEGs that belong to the same enriched ontology term. The inner circle represents DEGs lists, where hits are arranged along the arch. DEGs that hit multiple lists are colored in dark orange, and genes unique to a list are shown in light orange.

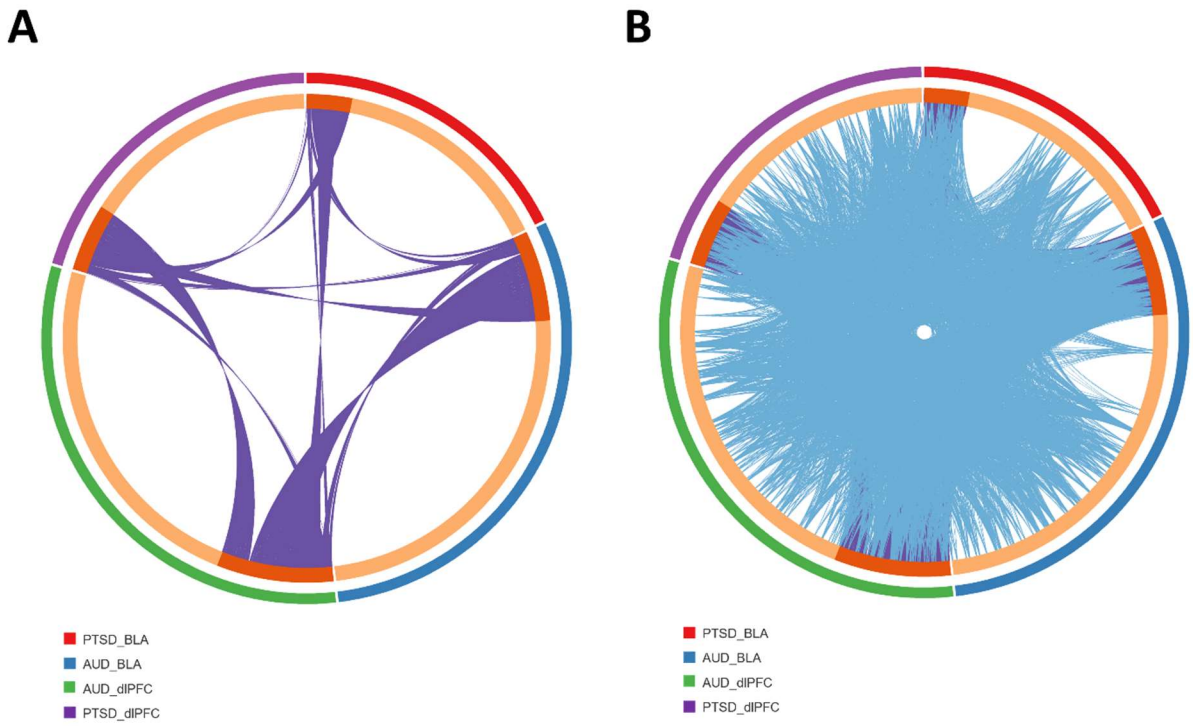


Figure 2 Heatmap of top 20 clusters of enriched terms across analyzed lists of differently expressed genes (DEGs) colored by p-values. Ration between enriched DEGs and total amount of genes included in each pathway is reported in red.

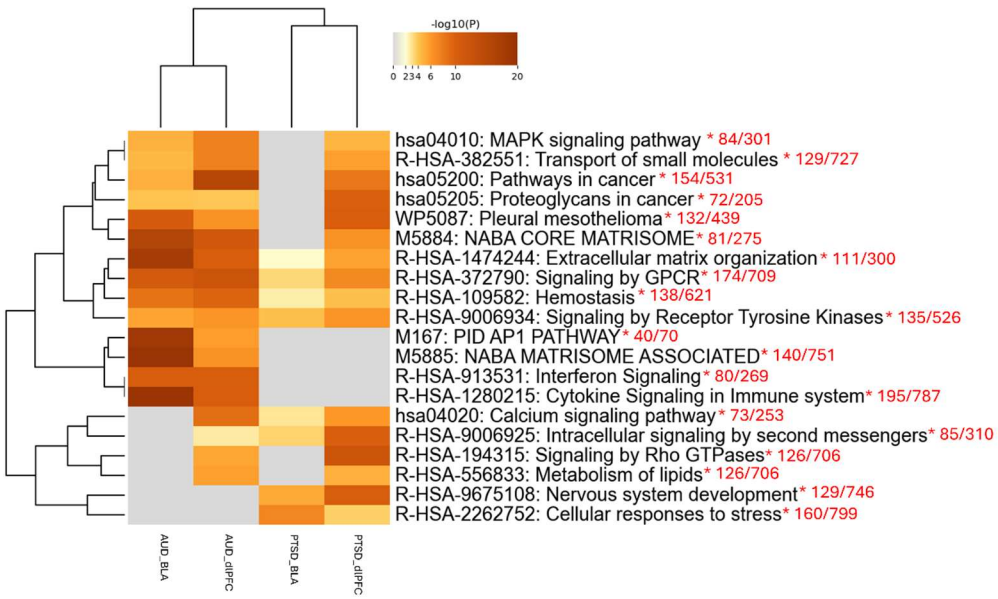


Figure 3 Network of enriched terms across analyzed lists of differently expressed gene (DEGs) (A) Networks are colored by cluster ID based on the best-scoring term by enrichment, where nodes that share the same cluster ID are typically close to each other; (B) Nodes are colored pies based on the percentage of identity to each DEG list; (C) Nodes are colored by p-value, where terms containing more genes tend to have a more significant p-value.

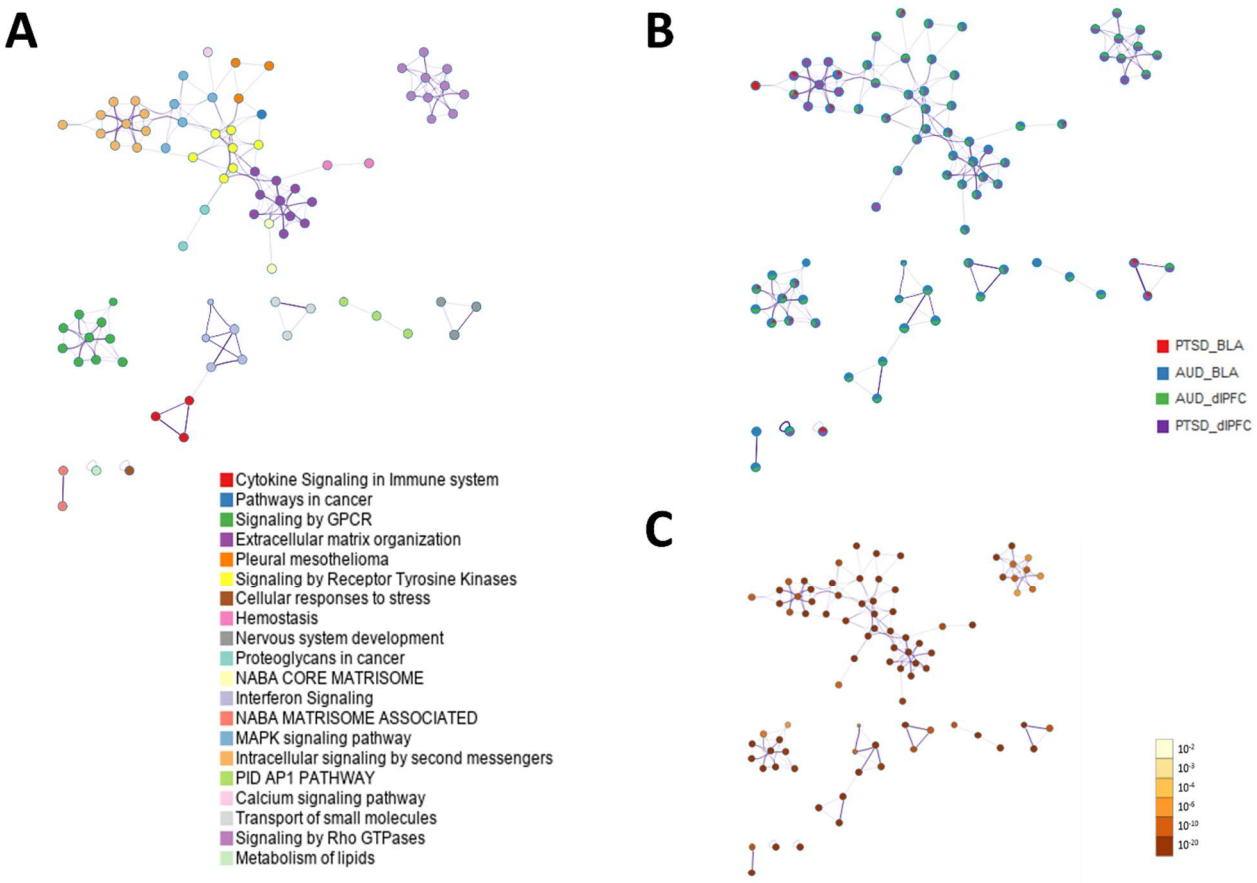
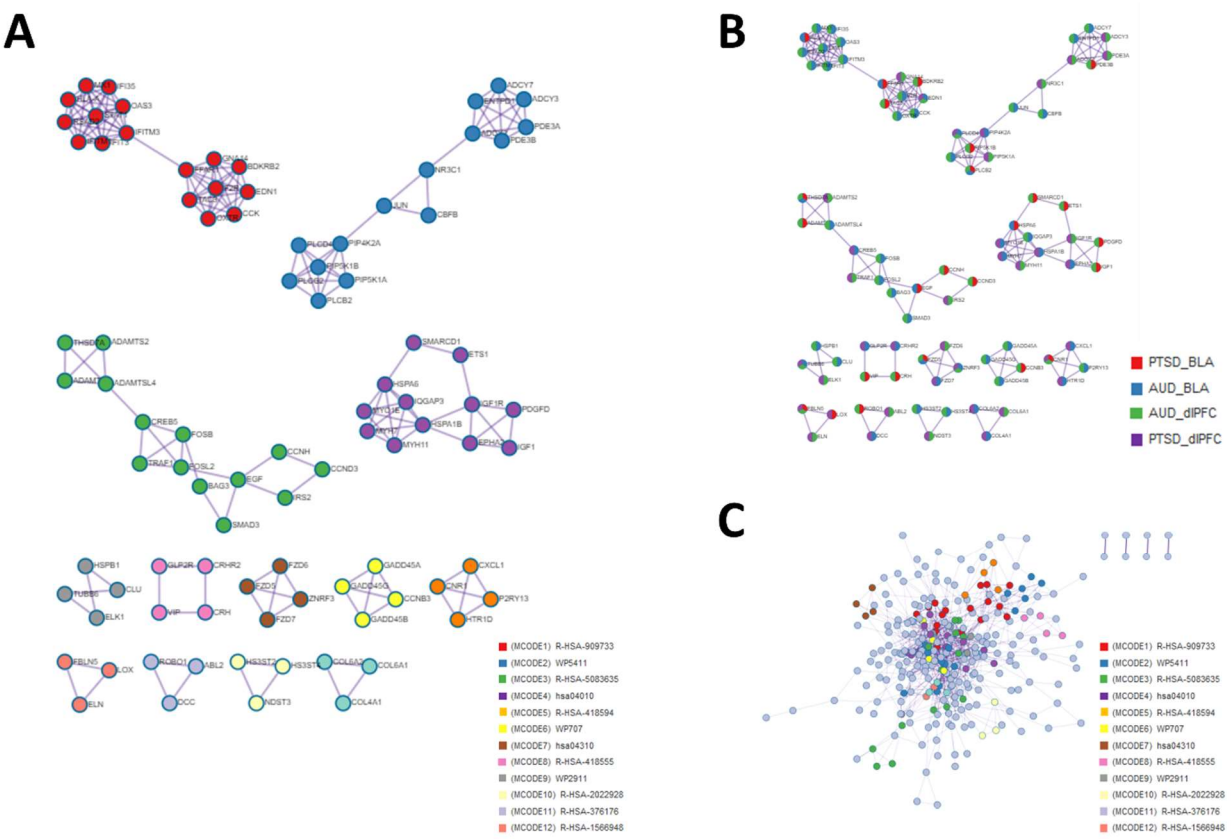


Figure 4 Protein-protein interaction network and MCODE components identified lists of differently expressed gene (DEGs). (A) MCODE colored by the independently enriched three best-scoring terms by p-value; (B) MCODE components where nodes are colored pies based on the percentage of identity to each DEG list; (C) Protein-protein interaction network where nodes are colored by the independently enriched three best-scoring terms by p-value



Tables

Table 1 Tissues and sample demographics. Grey colored columns stand as headers.

pheno	PMID	inclusion_status	mean_age (SD)	N	N_male	dlPFC	dACC	vmPFC	sgPFC	OFC	SFC	amy_ML	amy_BLA	amy_CN	NAC	HIP	VS	CN	PUT
AUD	30419947	dismissed	NA	120	NA	X													
PTSD	34646915	dismissed	40.98(10.4)	73	34	X		X											
PTSD	35791611	included	45.2(13,2)	206	140	X	X					X	X						
PTSD	33332381	included	45.6(11.9)	98	47	X	X		X	X									
AUD	31477794	dismissed	57.5 (1.6)	60	46								X	X	X				
AUD	33414398	included	57.5 (1.6)	60	46						X		X	X	X				
AUD	34791764	dismissed	57.94(8.9)	50	40	X													
AUD	32331824	dismissed	NA	20	NA														X
AUD	33332381	dismissed	55.97(9.4)	36	36						X				X				
AUD	35523767	dismissed	56,29(10,6)	99	68												X	X	X
AUD	30765688	included	55.3	138	101	X													
AUD	25450227	dismissed	NA	31	NA											X			
AUD	21464311	dismissed	37.3(7.6)	16	16											X			

Chapter 3

Brain-wide pleiotropy investigation of alcohol drinking and tobacco smoking behaviors

Abstract

To investigate the pleiotropic mechanisms linking brain structure and function to alcohol drinking and tobacco smoking, we integrated genome-wide data generated by the GWAS and Sequencing Consortium of Alcohol and Nicotine use (GSCAN; up to 805,431 participants) with information related to 3,935 brain imaging-derived phenotypes (IDPs) available from UK Biobank (N=33,224). We observed global genetic correlation of smoking behaviors with white matter hyperintensities, the morphology of the superior longitudinal fasciculus, and the mean thickness of pole-occipital. With respect to the latter brain IDP, we identified a local genetic correlation with age at which the individual began smoking regularly (chr2:36,120,744-36,867,389: rho=1, $p=1.01\times10^{-5}$). This pleiotropic effect appears to be related to *STRN* (*Striatin*) gene, previously associated with structural connectivity and possibly linked to tobacco cigarette craving. Our genetically informed causal inference analysis using latent causal variable approach and Mendelian randomization linked the activity of prefrontal and premotor cortex and that of superior and inferior precentral sulci, and cingulate sulci to alcoholic drinks per week (genetic causality proportion, gcp=0.38, $p=8.9\times10^{-4}$, rho=-0.18±0.07; inverse variance weighting, IVW beta=-0.04, 95%CI=-0.07 – -0.01). This relationship could be related to the role of these brain regions in the modulation of reward-seeking motivation and the processing of social cues. Overall, our brain-wide investigation highlighted that different pleiotropic mechanisms likely contribute to the relationship of brain structure and function with alcohol drinking and tobacco smoking, suggesting decision-making activities and chemosensory processing as modulators of propensity towards alcohol and tobacco consumption.

Introduction

Alcohol drinking and tobacco smoking are among the leading causes of death worldwide due to their high prevalence (Collaborators, 2018; He et al., 2022). This widespread use is partially attributed to the complex interplay of alcohol and nicotine with human brain, as shown by previous studies (Abburi et al., 2016; Morel et al., 2019; Weera et al., 2019). Additionally, several analyses observed consistent pleiotropy linking alcohol drinking and tobacco smoking to brain structure and function. For instance, early tobacco smoking initiation was genetically correlated with an increased precuneus surface area and decreased cortical thickness and surface area of the inferior temporal gyrus (Rabinowitz et al., 2022). Similarly, alcohol drinking showed genetic correlation with an increased total cortical surface area and decreased average cortical thickness (Rabinowitz et al., 2022). Because genetic information can be used as an anchor for causal inference (Davey Smith & Hemani, 2014), investigators also explored possible direct effects between drinking/smoking behaviors and brain morphology. For example, smoking initiation and alcohol drinking appear to have a possible causal association with decreased gray matter volume and the multivariable analysis pointed to alcohol drinking as the potential primary driver of this relationship (Lin et al., 2023). In another study, genetically predicted global cortical thickness showed an effect on alcohol drinking behaviors that was independent of neuropsychiatric phenotypes, substance use, trauma, and neurodegeneration (Mavromatis et al., 2022). Focusing on specific hypotheses, these previous investigations advanced our understanding of the brain mechanisms contributing to alcohol drinking and tobacco smoking behaviors. However, large-scale datasets allow investigators to expand further the depth of the analyses. Indeed, recent brain-wide pleiotropy analyses provided new insights into the role of brain structure and function on neuropsychiatric and behavioral traits (Bao et al., 2023; He et al., 2024; Koller et al., 2024; Zanoaga et al., 2023). In the present study,

we systematically investigated the pleiotropic mechanisms linking alcohol drinking and tobacco smoking to brain structure and function. Specifically, we integrated genome-wide data generated by the genome-wide association studies (GWAS) and Sequencing Consortium of Alcohol and Nicotine use (GSCAN; up to 805,431 participants) (Saunders et al., 2022) with information related to 3,935 brain imaging-derived phenotypes (IDPs; Supplemental Table 1) obtained from six magnetic resonance imaging (MRI) modalities (Smith et al., 2021), exploring the contribution of different pleiotropic mechanisms in the interplay between drinking/smoking behaviors and human brain.

Materials and Methods

Study Design

In the present study, we investigated the pleiotropic mechanisms linking alcohol drinking and tobacco smoking behaviors to brain structure and function by applying multiple analytic approaches to large-scale genome-wide datasets (Figure 1). A global genetic correlation analysis was performed to assess the overall genetic overlap between alcohol drinking and tobacco smoking and brain IDPs. Considering global genetic correlations surviving multiple testing correction, we leveraged the local analysis of [co]variant association (LAVA) approach (Werme et al., 2022) to identify chromosomal regions with strong statistical evidence of shared genetic mechanisms. To assess the presence of causal relationships underlying the global genetic overlap observed, we applied the latent causal variable (LCV) approach (O'Connor & Price, 2018) to pairwise combinations of brain IDP and alcohol/tobacco-related behaviors reaching nominally significant genetic correlation. Mendelian Randomization (MR) analysis was also conducted to follow up on the false discovery rate (FDR) significant results obtained from the LCV analysis.

Data sources

Genome-wide association statistics regarding IDPs were derived from the UK Biobank (UKB). The UKB is a large population-based prospective cohort containing in-depth genetic and health information from over 500,000 participants (Bycroft et al., 2018). In UKB, brain imaging was conducted using six MRI modalities: T1-weighted structural image, T2-weighted fluid-attenuated inversion recovery (T2_FLAIR) structural image, diffusion MRI (dMRI), resting-state functional MRI (rfMRI), task functional MRI (tfMRI), and susceptibility-weighted imaging (SWI). A total of 3,935 brain IDPs (Supplemental Table 1) were defined from MRI scans (Alfaro-Almagro et al., 2018). GWAS of brain IDPs in up to 33,224 UKB participants of European descent was previously described (Smith et al., 2021). Genome-wide association statistics regarding behaviors related to alcohol drinking and tobacco smoking were derived from GSCAN. In its latest GWAS (Saunders et al., 2022), this large collaborative effort meta-analyzed genome-wide information regarding smoking initiation (SmkInit) and the age at which the individual began smoking regularly (AgeSmk), cigarettes smoked per day (CigDay), smoking cessation (SmkCes), and alcoholic drinks per week (DrnkWk). Since brain IDP GWAS data were available only for individuals of European descent, we used publicly available GSCAN GWAS data for the same ancestry group (SmkInit N=805,431; AgeSmk N=323,386; CigDay N=326,497; SmkCes N= 388,313; DrnkWk N= 666,978). The sample overlap due to UKB inclusion in both IDP and GSCAN GWAS does not affect genetic correlation, LAVA, and LCV analyses. However, to avoid potential sample overlap bias in Mendelian randomization (MR) analysis, we also analyzed GSCAN GWAS data excluding the UKB cohort (SmkInit N= 357,235; AgeSmk N= 175,835; CigDay N= 183,196; SmkCes N= 188,701; DrnkWk N= 304,322).

Linkage disequilibrium score regression

Single nucleotide polymorphism (SNP)-based heritability and global genetic correlation were estimated using the linkage disequilibrium score regression (LDSC) method (Bulik-Sullivan et al., 2015). These analyses were performed using the HapMap 3 reference panel (International HapMap, 2003) and LD scores derived from European reference populations available from the 1000 Genomes Project (Genomes Project et al., 2015). Statistically significant genetic correlations were determined considering FDR $q<0.05$ to account for the number of brain IDPs tested with respect to each GSCAN phenotype.

Local analysis of [co]variant association

Local genetic correlation was assessed across 2,495 semi-independent chromosomal regions (~1Mb window) using LAVA (Werme et al., 2022). The LAVA univariate analysis was performed to estimate local SNP-based heritability for each pair of GSCAN phenotype and brain IDP. Considering chromosomal regions with at least nominally significant local SNP-based heritability ($p<0.05$), we estimated local genetic correlation between GSCAN phenotypes and brain IDPs using LAVA bivariate analysis. FDR multiple testing correction (FDR $q<0.05$) accounting for the number of chromosomal regions tested was applied to define statistically significant local genetic correlations. To further characterize the genomic regions identified using the LAVA approach, we leveraged information available from the GWAS catalog including genetic regions and reports from previous associations (Sollis et al., 2023).

Genetically Inferred Causal Inference

We performed a LCV analysis (O'Connor & Price, 2018) to estimate whether the global genetic correlation observed between GSCAN phenotypes and brain IDPs was due to possible cause-effect relationships. Considering pair combinations that reached at least nominally significance in the LDSC genetic correlation analysis ($p<0.05$), we estimated the genetic

causality proportions (gcp) between two traits. The gcp statistics can range from -1 to 1, where gcp=0 indicates no genetic causality, gcp=1 indicates a full genetic causality of trait #1 on trait #2, and gcp=-1 indicates full genetic causality of trait #2 on trait #1. In the present study, LCV analyses were performed considering brain IDPs phenotypes as trait #1 and GSCAN phenotypes as trait #2. Accordingly, positive gcp estimates indicate a causal effect of brain IDPs on GSCAN phenotypes, while a negative gcp estimate indicates a causal effect in the reverse direction. The sign of the genetically inferred causal effect is defined by the sign of the LCV rho statistics (i.e., rho>0 corresponds to positive causal effects, while rho<0 corresponds to negative causal effects). FDR correction accounting for the number of tests performed (FDR $q<0.05$) was applied to define statistically significant LCV results. To complement LCV results, we performed a MR analysis. Although both methods evaluate causal effect relationships, the LCV and MR results are based on different assumptions. Accordingly, an effect consistent between these two approaches can be considered more reliable. The MR analysis was performed using the TwoSampleMR package (Hemani et al., 2018) to estimate inverse variance weighting (IVW) estimates. Independent genetic instruments for TwoSampleMR analyses were identified considering SNPs with an exposure GWAS P-value threshold of 1×10^{-5}, that were LD-independent ($r^2=0.001$ within a 10,000-kb window). Because MR analyses can be biased by sample overlap, this analysis was conducted using UKB brain-IDP GWAS and GSCAN GWAS data excluding UKB participants.

Results

After FDR multiple testing correction (FDR $q<0.05$), we identified three genetic correlations linking behaviors related to tobacco smoking with brain structure and function (Supplemental Table 2). AgeSmk showed a negative genetic correlation with the mean thickness of Pole-occipital in the left hemisphere generated by Destrieux (a2009s) parcellation of the white

surface (aparc-a2009s lh thickness Pole-occipital, IDP 1219; rg=-0.23, $p=7.74\times10^{-6}$). A positive genetic correlation was observed between SmCes and the total volume of white matter hyperintensities from T1 and T2_FLAIR images (T2 FLAIR BIANCA WMH volume, IDP 1437; rg=0.16, $p=1.03\times10^{-5}$). CigDay also showed a positive correlation with the mean second level (L2) of right superior longitudinal fasciculus on fractional anisotropy skeleton from dMRI data (dMRI TBSS L2 Superior longitudinal fasciculus R, IDP 1736; rg=0.14, $p=1.24\times10^{-5}$). To further investigate the dynamics underlying these relationships, we conducted a local genetic correlation analysis and identified one region (hg38 chr2:35,895,678-36,640,246) showing statistically significant local genetic correlation between AgeSmk and aparc-a2009s lh thickness Pole-occipital (rho=1, $p=1.01\times10^{-5}$). In this region, 178 genome-wide significant associations ($p<5\times10^{-8}$) were reported in the GWAS catalog (Sollis et al., 2023) (Supplemental Table 3). Among them, several were related to brain related phenotypes including smoking initiation (rs62134085 $p=1\times10^{-18}$), educational attainment (rs305191 $p=2\times10^{-14}$), chronotype (rs848552 $p=5\times10^{-14}$), self-reported math ability (rs6708545 $p=1\times10^{-10}$), cognitive performance (rs6728742 $p=1\times10^{-8}$), and cortical thickness (rs1017154 $p=3\times10^{-8}$). As mentioned above, positive gcp estimates in the LCV analysis indicate a causal effect of brain IDPs on GSCAN phenotypes, while the sign of the genetically inferred causal effect is defined by the sign of the LCV rho statistics (i.e., rho>0 corresponds to positive causal effects, while rho<0 corresponds to negative causal effects). Considering nominally significant genetic correlations between brain IDPs and GSCAN phenotypes (Supplemental Table 2), we found 19 relationships linking brain structure and function to behaviors related to alcohol drinking and tobacco smoking (gcp>0, FDR q<0.05; Figure 2, Supplemental Table 4). Among them, 12 were related to brain connectivity analysis derived from rfMRI: three were related to DrnkWk (e.g, partial correlation of edge 363 in rfMRI dimensionality 100, ICA100 edge 363, IDP 2791; gcp=0.77, $P=1.12\times10^{-16}$,

rho=0.18±0.07), four to AgeSmk (e.g., ICA100 edge 838, IDP 3266, Supplemental Figure 1; gcp=0.79, P=3.72×10^{-23}, rho=0.20±0.08), three to CigDay (e.g., ICA25 edge 184, IDP 2402, Supplemental Figure 2; gcp=0.47, P=2.13×10^{-20}, rho=-0.20±0.08), and two to SmkCes (e.g., ICA25 edge 190, IDP 2408; gcp=0.86, P=9.43×10^{-22}, rho=0.21±0.07). We observed significant LCV results not related to brain connectivity only with respect to AgeSmk. These included brain IDPs related to cortical thickness (i.e., mean thickness of V1 in the right hemisphere generated by parcellation of the white surface using BA_exvivo parcellation, IDP 1111; gcp=0.67, P=3.33×10^{-9}, rho=-0.24±0.07), regional brain volumes (i.e., volume of rostral anterior cingulate in the right hemisphere generated by parcellation of the white surface using DKT parcellation, IDP 492; gcp=0.6, P=1.45×10^{-7}, rho=0.18±0.06), and cortical areas (e.g., area of orbital-inferior frontal gyrus in the right hemisphere generated by parcellation of the white surface using a2009s parcellation, IDP 958; gcp=-0.67, P=7.07×10^{-10}, rho=0.22±0.08). Among LCV effects surviving FDR-significance, we observed consistent effects in the MR analyses with respect to rfMRI connectivity ICA100 edge 772 (IDP 3200, Figure 3) on DrnWk (LCV gcp=0.38, p=8.9×10^{-4}, rho=-0.18±0.07; IVW beta=-0.04, 95%CI=-0.07 – -0.01). Direction consistency was observed for other 12 of the LCV significant results (Supplemental Table 5).

Discussion

The present study uncovered new information regarding the contribution of pleiotropic mechanisms to the complex interplay of alcohol drinking and tobacco smoking with brain structure and function. Building on previous studies that reported genetically informed relationships of drinking and smoking behaviors with brain cortical morphology and grey matter volume (Lin et al., 2023; Mavromatis et al., 2022; Rabinowitz et al., 2022), our brain-wide analyses identified genetic overlaps linking alcohol drinking and tobacco smoking with

previously unexplored brain-related phenotypes, including white matter hyperintensities, specific brain substructures, and brain connectivity. The global genetic correlation analysis identified three smoking-related results surviving multiple testing correction. Specifically, SmCes phenotype (current versus former smoker) was genetically correlated with increased total volume of white matter hyperintensities (T2 FLAIR BIANCA WMH volume, IDP 1437). Tobacco smoking has been previously linked to the progression of white matter hyperintensity in a dose-response relationship (Power et al., 2015). However, although this previous study did not observe an association between years since quitting tobacco smoking (Power et al., 2015), our finding highlights a possible genetic relationship between SmCes and white matter hyperintensities. This may be due to the fact that former smokers are more likely to have quit because of health conditions and heavier tobacco use in the past (Gallus et al., 2013). Additionally, SmCes is known to be associated with long-term weight gain (Veldheer et al., 2015), which can lead to white matter hyperintensities via inflammation (Lampe et al., 2019). We also observed a positive genetic correlation between CigDay and the morphology of the mean second level of the right superior longitudinal fasciculus (dMRI TBSS L2 Superior longitudinal fasciculus R, IDP 1736). This brain region is part of a brain network involved in spatial awareness and proprioception (Naito et al., 2016). Interestingly, the right superior longitudinal fasciculus has been associated with olfactory performance (Carreiras et al., 2024; Segura et al., 2013). In this context, the relationship of this brain region with CigDay may be related to chemosensory processing. While there is a well-established relationship between tobacco smoking and olfactory dysfunction (Ajmani et al., 2017), our result suggests possible shared genetic mechanisms predisposing individuals with certain chemosensory abilities to tobacco smoking quantity. There was also a negative genetic correlation between AgeSmk and the mean thickness of pole-occipital (parc-a2009s lh thickness Pole-occipital, IDP 1219). This

region is involved in visual attention (Mevorach et al., 2010) and appears to play a role in the reactivity toward smoking cues and tobacco craving (Mondino et al., 2018; Yang et al., 2020). In young adults, a reduced mean thickness of the left occipital pole has been associated with exposure to domestic violence (Tomoda et al., 2012), which is also a risk factor for tobacco smoking (Budenz et al., 2021). With respect to the relationship between AgeSmk and IDP 1219, we also observed a significant local genetic correlation in hg38 chr2:35,895,678-36,640,246 region. In this region, the GWAS catalog (Sollis et al., 2023) reports a number of genome-wide significant associations, including several related to smoking initiation, chronotype, educational attainment, cognitive performance, and cortical thickness. The majority of the associations reported were related to variants mapping to *CRIM1* and *FEZ2*. *CRIM1* gene has been linked to regulatory mechanisms over axon projection targeting (Sahni, Itoh, et al., 2021; Sahni, Shnider, et al., 2021). *FEZ2* is a member of a hub protein family involved in neuronal development, neurological disorders, viral infection, and autophagy (Teixeira et al., 2019). While these genes were not previously linked to substance use behaviors, their functions neurodevelopment suggests mechanisms pointing to the effect of brain development on tobacco-smoking behaviors later in life. To understand potential cause-effect relationships linking alcohol drinking and tobacco smoking to brain structure and function, we conducted a genetically informed causal inference analysis, observing convergent results between the LCV and MR methods that support an inverse effect of rfMRI connectivity ICA100 edge 772 (IDP 3200) on DrnWk phenotype. Because LCV and MR approaches are based on different assumptions, findings supported by both can be considered highly reliable. IDP 3200 reflects the activity of prefrontal and premotor cortex in the left hemisphere and that of superior frontal sulci, superior and inferior precentral sulci, and cingulate sulci (Figure 3). The left premotor cortex is involved in visual attention and in the integration of visual data at a semantic level

(Bartel et al., 2020; Hertrich et al., 2021). Precentral and cingulate regions are associated with the modulation of reward-seeking motivation (Dubey et al., 2020) and are also involved in the processing of context-related social cues (Apps et al., 2016; Gordon et al., 2023; Lavin et al., 2013). Conversely, the right orbitofrontal region is associated with social comprehension (Nakamura et al., 2020). In this context, the effect of IDP 3200 on DrnWk could reflect the impact of social cognition and decision-making process on the propensity towards alcohol consumption. While the other LCV results did not show statistical significance in the MR analysis, we observed directional consistency for 12 of them (Supplemental Table 5). Among them, we observed that CigDay was inversely affected by fluctuations in rfMRI connectivity in edge 184 for dimensionality 25 (IDP 2402) and edge 505 for dimensionality 100 (IDP 2933). IDP 2402 reflects increased activity across left hemisphere in dorsolateral prefrontal cortex and frontal gyri, decreased activity around right orbitofrontal areas, along with increased activation of parietal lobe (Supplemental Figure 2). Being part of the superior longitudinal fasciculus, these brain areas are associated with spatial awareness and proprioception (Naito et al., 2016) and olfactory performance (Carreiras et al., 2024; Segura et al., 2013). While olfactory perception is especially mediated by the orbitofrontal cortex (Sagar et al., 2023), brain lesions localized within the right orbitofrontal cortex have been recorded to specifically hinder the formation of conscious olfactory percepts (Li et al., 2010). Interestingly, stimulating the left dorsolateral prefrontal cortex and inhibiting the right orbitofrontal cortex resulted in a risk-averse response in human subjects enrolled in a transcranial direct current stimulation study (Nejati et al., 2018). IDP 2933 reflects the association between lower activation levels in the ventral posterolateral nucleus to higher ones in the right hemisphere along the posterior cingulate cortex, dorsolateral prefrontal cortex, and frontal gyri (Supplemental Figure 3). Intriguingly, opposite fluctuation patterns were associated with increased pain in patients

experiencing migraine attacks, possibly due to a strengthened network of nociceptive information processing (Lim et al., 2021). Nicotine may have anti-nociceptive effects (Carstens & Carstens, 2022) and tobacco smoking has been linked to coping mechanisms related to migraine and chronic head pain (Weinberger & Seng, 2023). Additionally, the interplay between nociception and olfaction has been proposed at both molecular and functional levels (Lotsch et al., 2016; Mignot et al., 2023). In this context, our results reinforce the hypothesis of shared genetic mechanisms predisposing individuals with certain chemosensory abilities to tobacco smoking behaviors. In conclusion, our brain-wide analyses highlighted that different pleiotropic mechanisms likely contribute to the relationship of brain structure and function with alcohol drinking and tobacco smoking, opening new directions in understanding the processes underlying these complex behaviors. However, we also acknowledge two main limitations. While we leveraged large-scale genome-wide datasets, these were generated including only participants of European descent, because of the lack of large genetic and imaging studies in other human populations. Accordingly, our findings may not be generalizable to other population groups. Another important limitation is related to genetically informed analysis. While we used multiple methods relying on different assumptions, our results may still be affected by unaccounted confounders. Thus, our findings will need to be confirmed by evidence generated by complementary study designs (e.g., prospective studies). Finally, while no genetically inferred effect of alcohol drinking and tobacco smoking was observed, future studies with larger datasets may be able to characterize this effect direction.

References

Abburi, C., Wolfman, S. L., Metz, R. A., Kamber, R., McGehee, D. S., & McDaid, J. (2016). Tolerance to Ethanol or Nicotine Results in Increased Ethanol Self-Administration and Long-Term Depression in the Dorsolateral Striatum. *eNeuro*, *3*(4). https://doi.org/10.1523/ENEURO.0112-15.2016

Ajmani, G. S., Suh, H. H., Wroblewski, K. E., & Pinto, J. M. (2017). Smoking and olfactory dysfunction: A systematic literature review and meta-analysis. *Laryngoscope*, *127*(8), 1753-1761. https://doi.org/10.1002/lary.26558

Alfaro-Almagro, F., Jenkinson, M., Bangerter, N. K., Andersson, J. L. R., Griffanti, L., Douaud, G., Sotiropoulos, S. N., Jbabdi, S., Hernandez-Fernandez, M., Vallee, E., Vidaurre, D., Webster, M., McCarthy, P., Rorden, C., Daducci, A., Alexander, D. C., Zhang, H., Dragonu, I., Matthews, P. M., . . . Smith, S. M. (2018). Image processing and Quality Control for the first 10,000 brain imaging datasets from UK Biobank. *Neuroimage*, *166*, 400-424. https://doi.org/10.1016/j.neuroimage.2017.10.034

Apps, M. A., Rushworth, M. F., & Chang, S. W. (2016). The Anterior Cingulate Gyrus and Social Cognition: Tracking the Motivation of Others. *Neuron*, *90*(4), 692-707. https://doi.org/10.1016/j.neuron.2016.04.018

Bao, J., Wen, J., Wen, Z., Yang, S., Cui, Y., Yang, Z., Erus, G., Saykin, A. J., Long, Q., Davatzikos, C., & Shen, L. (2023). Brain-wide genome-wide colocalization study for integrating genetics, transcriptomics and brain morphometry in Alzheimer's disease. *Neuroimage*, *280*, 120346. https://doi.org/10.1016/j.neuroimage.2023.120346

Bartel, G., Marko, M., Rameses, I., Lamm, C., & Riecansky, I. (2020). Left Prefrontal Cortex Supports the Recognition of Meaningful Patterns in Ambiguous Stimuli. *Front Neurosci*, *14*, 152. https://doi.org/10.3389/fnins.2020.00152

Budenz, A., Klein, A., & Prutzman, Y. (2021). The Relationship Between Trauma Exposure and Adult Tobacco Use: Analysis of the National Epidemiologic Survey on Alcohol and Related Conditions (III). *Nicotine Tob Res*, *23*(10), 1716-1726. https://doi.org/10.1093/ntr/ntab057

Bulik-Sullivan, B., Finucane, H. K., Anttila, V., Gusev, A., Day, F. R., Loh, P. R., ReproGen, C., Psychiatric Genomics, C., Genetic Consortium for Anorexia Nervosa of the Wellcome Trust Case Control, C., Duncan, L., Perry, J. R., Patterson, N., Robinson, E. B., Daly, M. J.,

Price, A. L., & Neale, B. M. (2015). An atlas of genetic correlations across human diseases and traits. *Nat Genet*, *47*(11), 1236-1241. https://doi.org/10.1038/ng.3406

Bycroft, C., Freeman, C., Petkova, D., Band, G., Elliott, L. T., Sharp, K., Motyer, A., Vukcevic, D., Delaneau, O., O'Connell, J., Cortes, A., Welsh, S., Young, A., Effingham, M., McVean, G., Leslie, S., Allen, N., Donnelly, P., & Marchini, J. (2018). The UK Biobank resource with deep phenotyping and genomic data. *Nature*, *562*(7726), 203-209. https://doi.org/10.1038/s41586-018-0579-z

Carreiras, M., Quinones, I., Chen, H. A., Vazquez-Araujo, L., Small, D., & Frost, R. (2024). Sniffing out meaning: Chemosensory and semantic neural network changes in sommeliers. *Hum Brain Mapp*, *45*(2), e26564. https://doi.org/10.1002/hbm.26564

Carstens, E., & Carstens, M. I. (2022). Sensory Effects of Nicotine and Tobacco. *Nicotine Tob Res*, *24*(3), 306-315. https://doi.org/10.1093/ntr/ntab086

Collaborators, G. B. D. A. (2018). Alcohol use and burden for 195 countries and territories, 1990-2016: a systematic analysis for the Global Burden of Disease Study 2016. *Lancet*, *392*(10152), 1015-1035. https://doi.org/10.1016/S0140-6736(18)31310-2

Davey Smith, G., & Hemani, G. (2014). Mendelian randomization: genetic anchors for causal inference in epidemiological studies. *Hum Mol Genet*, *23*(R1), R89-98. https://doi.org/10.1093/hmg/ddu328

Dubey, I., Georgescu, A. L., Hommelsen, M., Vogeley, K., Ropar, D., & Hamilton, A. F. C. (2020). Distinct neural correlates of social and object reward seeking motivation. *Eur J Neurosci*, *52*(9), 4214-4229. https://doi.org/10.1111/ejn.14888

Gallus, S., Muttarak, R., Franchi, M., Pacifici, R., Colombo, P., Boffetta, P., Leon, M. E., & La Vecchia, C. (2013). Why do smokers quit? *Eur J Cancer Prev*, *22*(1), 96-101. https://doi.org/10.1097/CEJ.0b013e3283552da8

Genomes Project, C., Auton, A., Brooks, L. D., Durbin, R. M., Garrison, E. P., Kang, H. M., Korbel, J. O., Marchini, J. L., McCarthy, S., McVean, G. A., & Abecasis, G. R. (2015). A global reference for human genetic variation. *Nature*, *526*(7571), 68-74. https://doi.org/10.1038/nature15393

Gordon, E. M., Chauvin, R. J., Van, A. N., Rajesh, A., Nielsen, A., Newbold, D. J., Lynch, C. J., Seider, N. A., Krimmel, S. R., Scheidter, K. M., Monk, J., Miller, R. L., Metoki, A., Montez, D. F., Zheng, A., Elbau, I., Madison, T., Nishino, T., Myers, M. J., . . . Dosenbach, N. U. F.

(2023). A somato-cognitive action network alternates with effector regions in motor cortex. *Nature*, *617*(7960), 351-359. https://doi.org/10.1038/s41586-023-05964-2

He, H., Pan, Z., Wu, J., Hu, C., Bai, L., & Lyu, J. (2022). Health Effects of Tobacco at the Global, Regional, and National Levels: Results From the 2019 Global Burden of Disease Study. *Nicotine Tob Res*, *24*(6), 864-870. https://doi.org/10.1093/ntr/ntab265

He, J., Cabrera-Mendoza, B., De Angelis, F., Pathak, G. A., Koller, D., Curhan, S. G., Curhan, G. C., Mecca, A. P., van Dyck, C. H., & Polimanti, R. (2024). Sex differences in the pleiotropy of hearing difficulty with imaging-derived phenotypes: a brain-wide investigation. *Brain*. https://doi.org/10.1093/brain/awae077

Hemani, G., Zheng, J., Elsworth, B., Wade, K. H., Haberland, V., Baird, D., Laurin, C., Burgess, S., Bowden, J., Langdon, R., Tan, V. Y., Yarmolinsky, J., Shihab, H. A., Timpson, N. J., Evans, D. M., Relton, C., Martin, R. M., Davey Smith, G., Gaunt, T. R., & Haycock, P. C. (2018). The MR-Base platform supports systematic causal inference across the human phenome. *Elife*, *7*. https://doi.org/10.7554/eLife.34408

Hertrich, I., Dietrich, S., Blum, C., & Ackermann, H. (2021). The Role of the Dorsolateral Prefrontal Cortex for Speech and Language Processing. *Front Hum Neurosci*, *15*, 645209. https://doi.org/10.3389/fnhum.2021.645209

International HapMap, C. (2003). The International HapMap Project. *Nature*, *426*(6968), 789-796. https://doi.org/10.1038/nature02168

Koller, D., Friligkou, E., Stiltner, B., Pathak, G. A., Lokhammer, S., Levey, D. F., Zhou, H., Hatoum, A. S., Deak, J. D., Kember, R. L., Treur, J. L., Kranzler, H. R., Johnson, E. C., Stein, M. B., Gelernter, J., & Polimanti, R. (2024). Pleiotropy and genetically inferred causality linking multisite chronic pain to substance use disorders. *Mol Psychiatry*. https://doi.org/10.1038/s41380-024-02446-3

Lampe, L., Zhang, R., Beyer, F., Huhn, S., Kharabian Masouleh, S., Preusser, S., Bazin, P. L., Schroeter, M. L., Villringer, A., & Witte, A. V. (2019). Visceral obesity relates to deep white matter hyperintensities via inflammation. *Ann Neurol*, *85*(2), 194-203. https://doi.org/10.1002/ana.25396

Lavin, C., Melis, C., Mikulan, E., Gelormini, C., Huepe, D., & Ibanez, A. (2013). The anterior cingulate cortex: an integrative hub for human socially-driven interactions. *Front Neurosci*, *7*, 64. https://doi.org/10.3389/fnins.2013.00064

Li, W., Lopez, L., Osher, J., Howard, J. D., Parrish, T. B., & Gottfried, J. A. (2010). Right orbitofrontal cortex mediates conscious olfactory perception. *Psychol Sci*, *21*(10), 1454-1463. https://doi.org/10.1177/0956797610382121

Lim, M., Jassar, H., Kim, D. J., Nascimento, T. D., & DaSilva, A. F. (2021). Differential alteration of fMRI signal variability in the ascending trigeminal somatosensory and pain modulatory pathways in migraine. *J Headache Pain*, *22*(1), 4. https://doi.org/10.1186/s10194-020-01210-6

Lin, W., Zhu, L., & Lu, Y. (2023). Association of smoking with brain gray and white matter volume: a Mendelian randomization study. *Neurol Sci*, *44*(11), 4049-4055. https://doi.org/10.1007/s10072-023-06854-1

Lotsch, J., Hahner, A., Gossrau, G., Hummel, C., Walter, C., Ultsch, A., & Hummel, T. (2016). Smell of pain: intersection of nociception and olfaction. *Pain*, *157*(10), 2152-2157. https://doi.org/10.1097/j.pain.0000000000000599

Mavromatis, L. A., Rosoff, D. B., Cupertino, R. B., Garavan, H., Mackey, S., & Lohoff, F. W. (2022). Association Between Brain Structure and Alcohol Use Behaviors in Adults: A Mendelian Randomization and Multiomics Study. *JAMA Psychiatry*, *79*(9), 869-878. https://doi.org/10.1001/jamapsychiatry.2022.2196

Mevorach, C., Hodsoll, J., Allen, H., Shalev, L., & Humphreys, G. (2010). Ignoring the elephant in the room: a neural circuit to downregulate salience. *J Neurosci*, *30*(17), 6072-6079. https://doi.org/10.1523/JNEUROSCI.0241-10.2010

Mignot, C., Faria, V., Hummel, T., Frost, M., Michel, C. M., Gossrau, G., & Haehner, A. (2023). Migraine with aura: less control over pain and fragrances? *J Headache Pain*, *24*(1), 55. https://doi.org/10.1186/s10194-023-01592-3

Mondino, M., Luck, D., Grot, S., Januel, D., Suaud-Chagny, M. F., Poulet, E., & Brunelin, J. (2018). Effects of repeated transcranial direct current stimulation on smoking, craving and brain reactivity to smoking cues. *Sci Rep*, *8*(1), 8724. https://doi.org/10.1038/s41598-018-27057-1

Morel, C., Montgomery, S., & Han, M. H. (2019). Nicotine and alcohol: the role of midbrain dopaminergic neurons in drug reinforcement. *Eur J Neurosci*, *50*(3), 2180-2200. https://doi.org/10.1111/ejn.14160

Naito, E., Morita, T., & Amemiya, K. (2016). Body representations in the human brain revealed by kinesthetic illusions and their essential contributions to motor control and corporeal awareness. *Neurosci Res*, *104*, 16-30. https://doi.org/10.1016/j.neures.2015.10.013

Nakamura, M., Nestor, P. G., & Shenton, M. E. (2020). Orbitofrontal Sulcogyral Pattern as a Transdiagnostic Trait Marker of Early Neurodevelopment in the Social Brain. *Clin EEG Neurosci*, *51*(4), 275-284. https://doi.org/10.1177/1550059420904180

Nejati, V., Salehinejad, M. A., & Nitsche, M. A. (2018). Interaction of the Left Dorsolateral Prefrontal Cortex (l-DLPFC) and Right Orbitofrontal Cortex (OFC) in Hot and Cold Executive Functions: Evidence from Transcranial Direct Current Stimulation (tDCS). *Neuroscience*, *369*, 109-123. https://doi.org/10.1016/j.neuroscience.2017.10.042

O'Connor, L. J., & Price, A. L. (2018). Distinguishing genetic correlation from causation across 52 diseases and complex traits. *Nat Genet*, *50*(12), 1728-1734. https://doi.org/10.1038/s41588-018-0255-0

Power, M. C., Deal, J. A., Sharrett, A. R., Jack, C. R., Jr., Knopman, D., Mosley, T. H., & Gottesman, R. F. (2015). Smoking and white matter hyperintensity progression: the ARIC-MRI Study. *Neurology*, *84*(8), 841-848. https://doi.org/10.1212/WNL.0000000000001283

Rabinowitz, J. A., Campos, A. I., Ong, J. S., Garcia-Marin, L. M., Alcauter, S., Mitchell, B. L., Grasby, K. L., Cuellar-Partida, G., Gillespie, N. A., Huhn, A. S., Martin, N. G., Thompson, P. M., Medland, S. E., Maher, B. S., & Renteria, M. E. (2022). Shared Genetic Etiology between Cortical Brain Morphology and Tobacco, Alcohol, and Cannabis Use. *Cereb Cortex*, *32*(4), 796-807. https://doi.org/10.1093/cercor/bhab243

Sagar, V., Shanahan, L. K., Zelano, C. M., Gottfried, J. A., & Kahnt, T. (2023). High-precision mapping reveals the structure of odor coding in the human brain. *Nat Neurosci*, *26*(9), 1595-1602. https://doi.org/10.1038/s41593-023-01414-4

Sahni, V., Itoh, Y., Shnider, S. J., & Macklis, J. D. (2021). Crim1 and Kelch-like 14 exert complementary dual-directional developmental control over segmentally specific corticospinal axon projection targeting. *Cell Rep*, *37*(3), 109842. https://doi.org/10.1016/j.celrep.2021.109842

Sahni, V., Shnider, S. J., Jabaudon, D., Song, J. H. T., Itoh, Y., Greig, L. C., & Macklis, J. D. (2021). Corticospinal neuron subpopulation-specific developmental genes prospectively indicate mature segmentally specific axon projection targeting. *Cell Rep*, *37*(3), 109843. https://doi.org/10.1016/j.celrep.2021.109843

Saunders, G. R. B., Wang, X., Chen, F., Jang, S. K., Liu, M., Wang, C., Gao, S., Jiang, Y., Khunsriraksakul, C., Otto, J. M., Addison, C., Akiyama, M., Albert, C. M., Aliev, F., Alonso, A., Arnett, D. K., Ashley-Koch, A. E., Ashrani, A. A., Barnes, K. C., . . . Vrieze, S. (2022). Genetic diversity fuels gene discovery for tobacco and alcohol use. *Nature*, *612*(7941), 720-724. https://doi.org/10.1038/s41586-022-05477-4

Segura, B., Baggio, H. C., Solana, E., Palacios, E. M., Vendrell, P., Bargallo, N., & Junque, C. (2013). Neuroanatomical correlates of olfactory loss in normal aged subjects. *Behav Brain Res*, *246*, 148-153. https://doi.org/10.1016/j.bbr.2013.02.025

Smith, S. M., Douaud, G., Chen, W., Hanayik, T., Alfaro-Almagro, F., Sharp, K., & Elliott, L. T. (2021). An expanded set of genome-wide association studies of brain imaging phenotypes in UK Biobank. *Nat Neurosci*, *24*(5), 737-745. https://doi.org/10.1038/s41593-021-00826-4

Sollis, E., Mosaku, A., Abid, A., Buniello, A., Cerezo, M., Gil, L., Groza, T., Gunes, O., Hall, P., Hayhurst, J., Ibrahim, A., Ji, Y., John, S., Lewis, E., MacArthur, J. A. L., McMahon, A., Osumi-Sutherland, D., Panoutsopoulou, K., Pendlington, Z., . . . Harris, L. W. (2023). The NHGRI-EBI GWAS Catalog: knowledgebase and deposition resource. *Nucleic Acids Res*, *51*(D1), D977-D985. https://doi.org/10.1093/nar/gkac1010

Teixeira, M. B., Alborghetti, M. R., & Kobarg, J. (2019). Fasciculation and elongation zeta proteins 1 and 2: From structural flexibility to functional diversity. *World J Biol Chem*, *10*(2), 28-43. https://doi.org/10.4331/wjbc.v10.i2.28

Tomoda, A., Polcari, A., Anderson, C. M., & Teicher, M. H. (2012). Reduced visual cortex gray matter volume and thickness in young adults who witnessed domestic violence during childhood. *PLoS One*, *7*(12), e52528. https://doi.org/10.1371/journal.pone.0052528

Veldheer, S., Yingst, J., Zhu, J., & Foulds, J. (2015). Ten-year weight gain in smokers who quit, smokers who continued smoking and never smokers in the United States, NHANES 2003-2012. *Int J Obes (Lond)*, *39*(12), 1727-1732. https://doi.org/10.1038/ijo.2015.127

Weera, M. M., Agim, Z. S., Cannon, J. R., & Chester, J. A. (2019). Genetic correlations between nicotine reinforcement-related behaviors and propensity toward high or low alcohol preference in two replicate mouse lines. *Genes Brain Behav*, *18*(3), e12515. https://doi.org/10.1111/gbb.12515

Weinberger, A. H., & Seng, E. K. (2023). The Relationship of Tobacco Use and Migraine: A Narrative Review. *Curr Pain Headache Rep*, *27*(4), 39-47. https://doi.org/10.1007/s11916-023-01103-8

Werme, J., van der Sluis, S., Posthuma, D., & de Leeuw, C. A. (2022). An integrated framework for local genetic correlation analysis. *Nat Genet*, *54*(3), 274-282. https://doi.org/10.1038/s41588-022-01017-y

Yang, Z., Zhang, Y., Cheng, J., & Zheng, R. (2020). Meta-analysis of brain gray matter changes in chronic smokers. *Eur J Radiol*, *132*, 109300. https://doi.org/10.1016/j.ejrad.2020.109300

Zanoaga, M. D., Friligkou, E., He, J., Pathak, G. A., Koller, D., Cabrera-Mendoza, B., Stein, M. B., & Polimanti, R. (2023). Brainwide Mendelian Randomization Study of Anxiety Disorders and Symptoms. *Biol Psychiatry*. https://doi.org/10.1016/j.biopsych.2023.11.006

Figures

Figure 1 Workflow of the analyses performed.

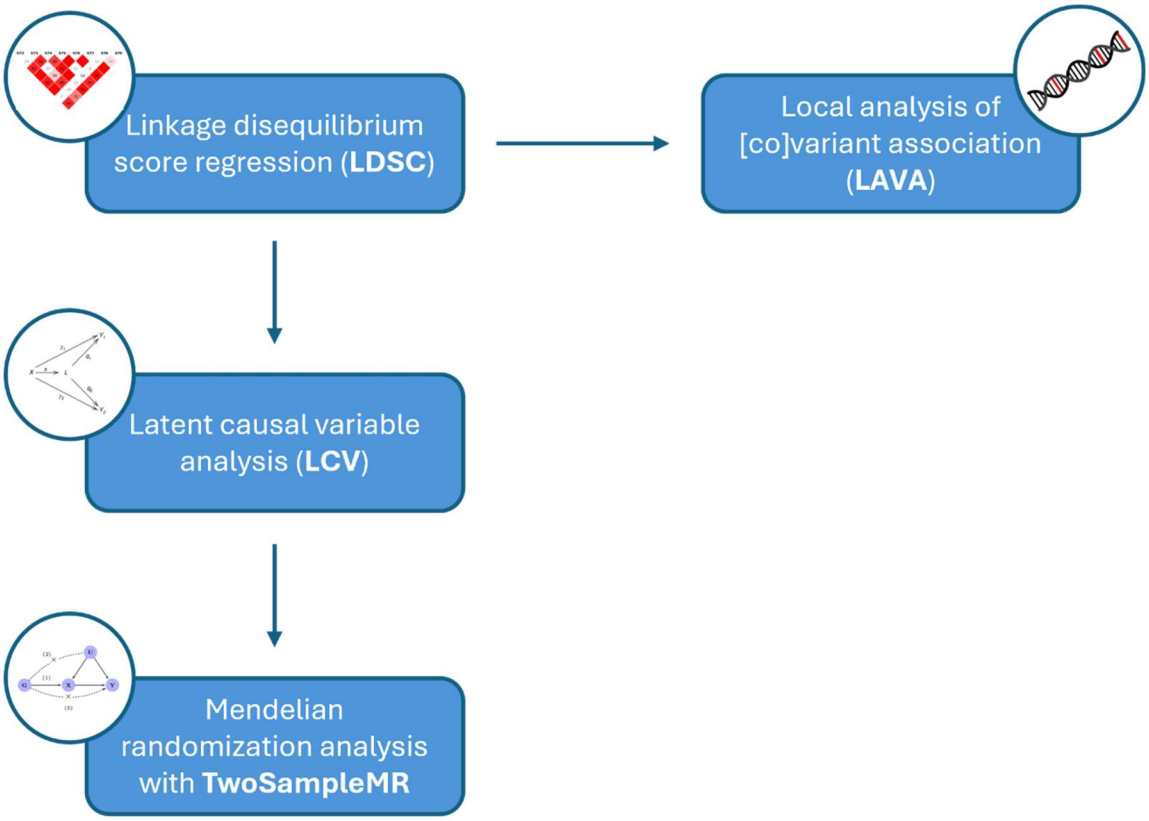


Figure 2 Genetic causal proportion (false discovery rate, FDR q<0.05) linking brain imaging-derived phenotypes (IDP) to alcohol drinking and tobacco smoking behaviors. Abbreviations aparc-a2009s rh area G-front-inf-Orbital (IDP 0958); aparc-DKTatlas rh area rostralanteriorcingulate (IDP 0864); aparc-a2009s rh area S-interm-prim-Jensen (IDP 1000); aparc-Desikan rh area rostralanteriorcingulate (IDP 0707); aparc-a2009s rh area G-subcallosal (IDP 0977); BA-exvivo rh thickness V1 (IDP 1111); aparc-DKTatlas rh volume rostralanteriorcingulate (IDP 0492); ICA100 edge 838 (IDP 3266); ICA25 edge 190 (IDP 2408); ICA25 edge 184 (IDP 2402); ICA100 edge 363 (IDP 2791); ICA100 edge 649 (IDP 3077); ICA100 edge 534 (IDP 2962); ICA100 edge 1438 (IDP 3866); ICA100 edge 628 (IDP 3056); ICA100 edge 974 (IDP 3402); ICA100 edge 772 (IDP 3200); ICA100 edge 162 (IDP 2590); ICA100 edge 505 (IDP 2933)

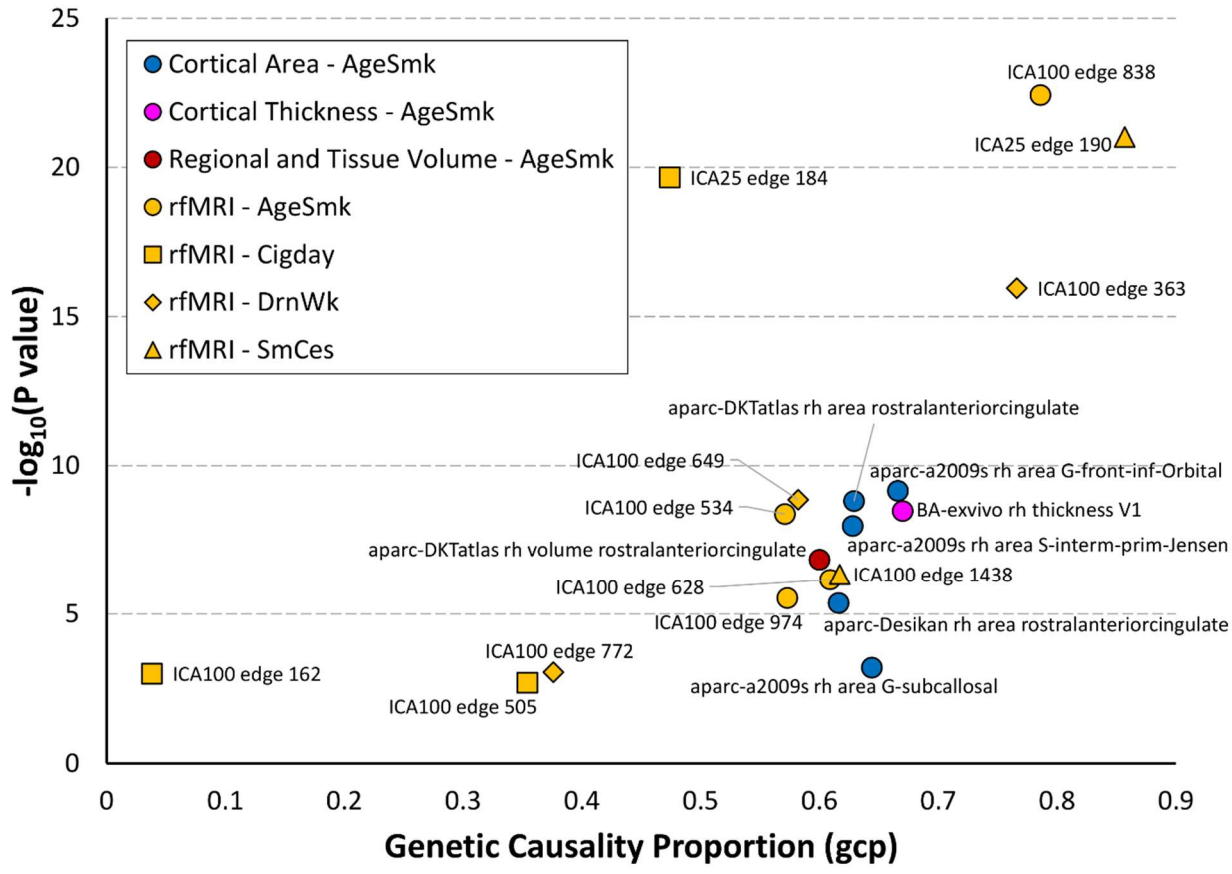
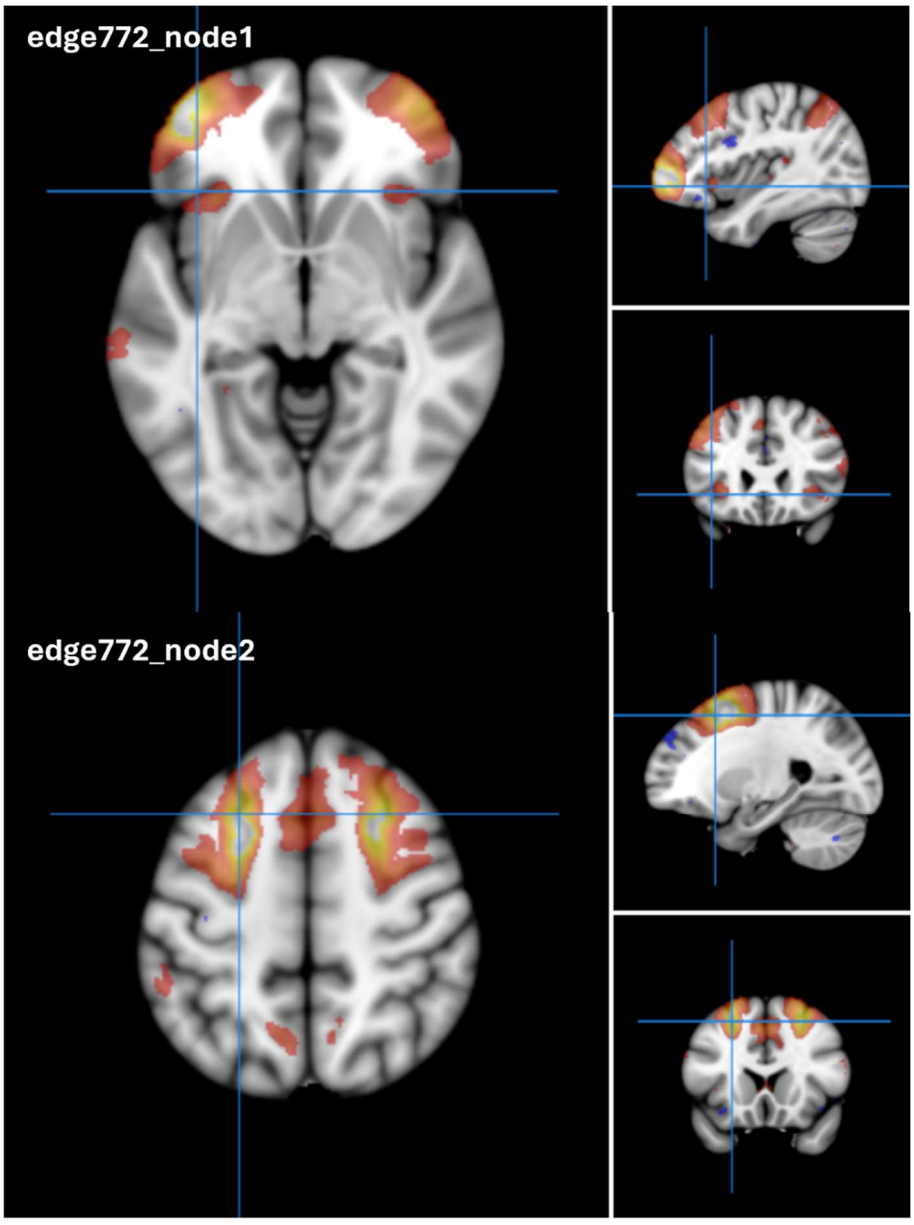
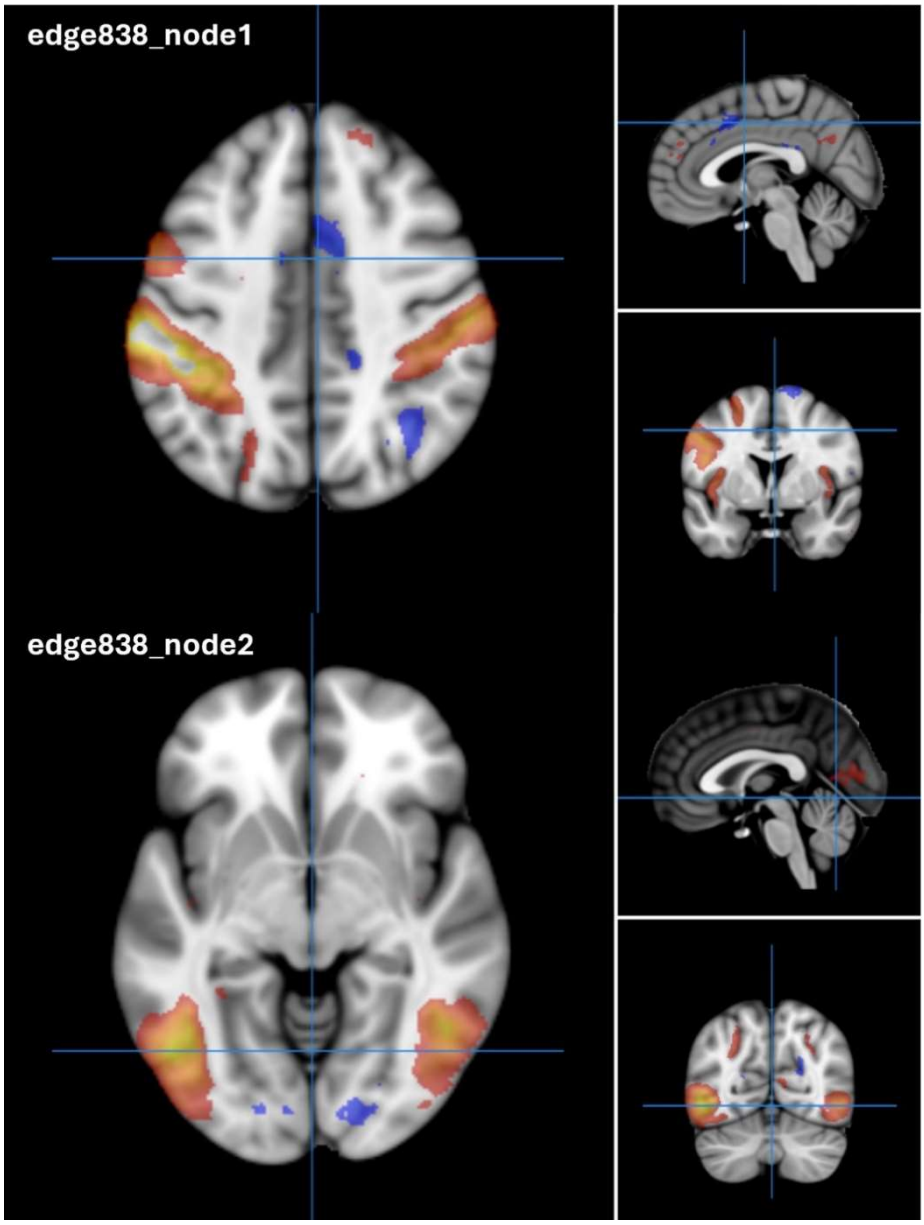


Figure 3 IDP 3200 reflecting edge 772 of dimensionality 100 separated by spatial ICA in resting-state functional magnetic resonance imaging.

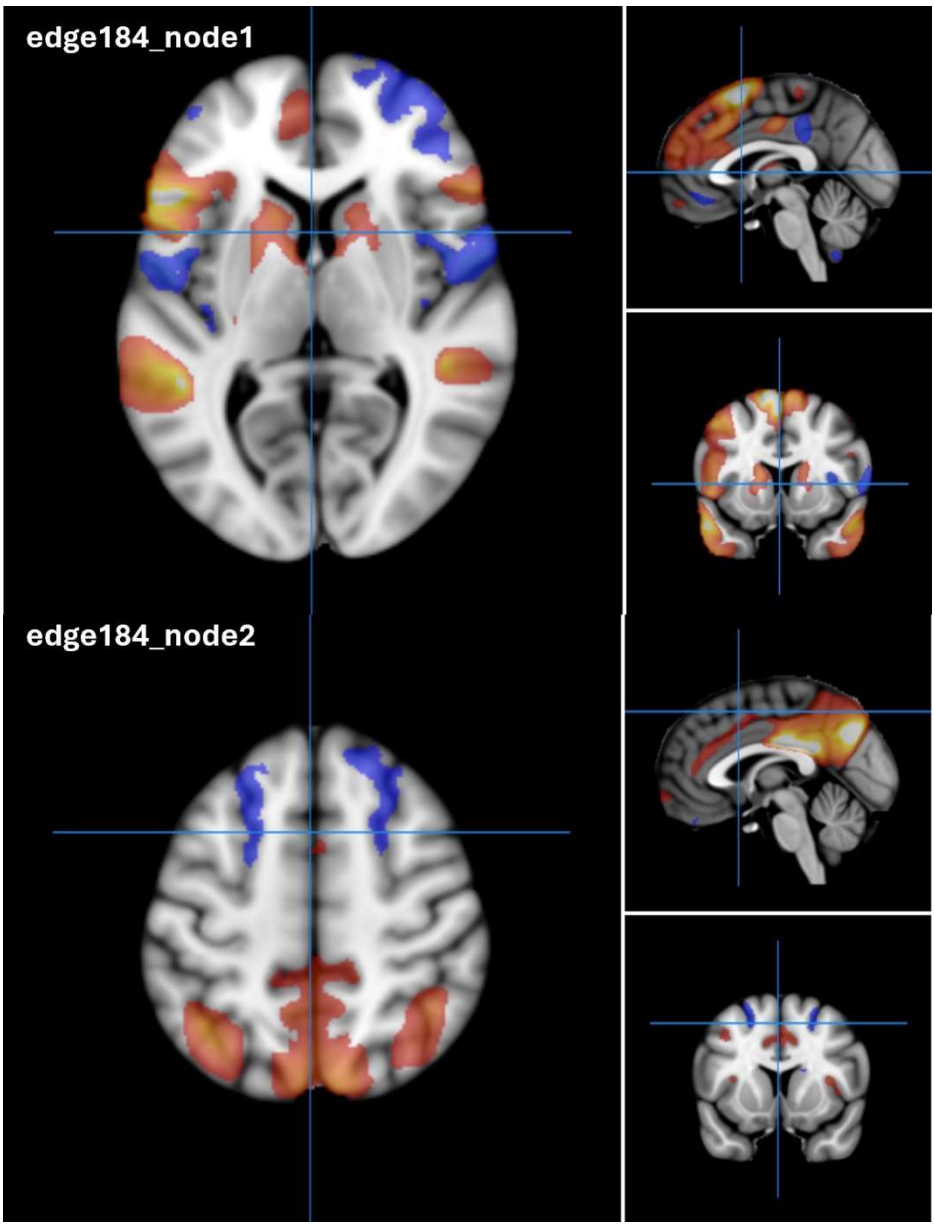


Supplemental Figures

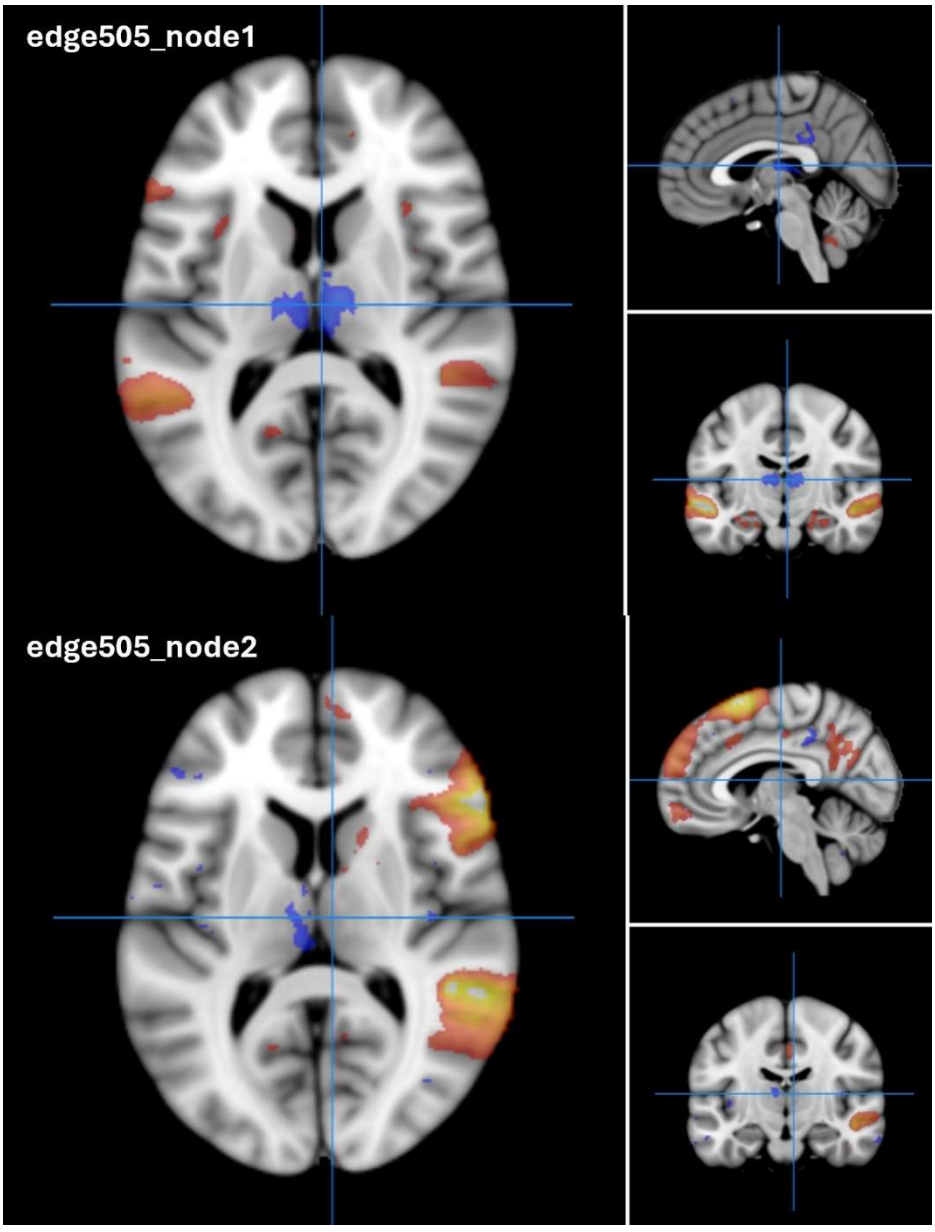
Supplemental Figure 1 IDP 3266 reflecting edge 838 of dimensionality 100 separated by spatial ICA in resting-state functional magnetic resonance imaging.



Supplemental Figure 2 IDP 2402 reflecting edge 184 of dimensionality 25 separated by spatial ICA in resting-state functional magnetic resonance imaging.



Supplemental Figure 3 IDP 2933 reflecting edge 505 of dimensionality 100 separated by spatial ICA in resting-state functional magnetic resonance imaging.



Supplemental tables

Supplemental Table 1 Information for brain imaging-derived phenotypes (IDPs) can be found in Table S1 of the preprint made available at the following link: https://www.medrxiv.org/content/medrxiv/early/2023/08/28/2023.08.25.23294639/DC1/embed/media-1.xlsx?download=true

Supplemental Table 2 Global genetic correlation between GSCAN phenotypes (smoking initiation SmkInit; Age at which the individual began smoking regularly, AgeSmk; cigarettes smoked per day, CigDay; smoking cessation, SmkCes; alcoholic drinks per week, DrnkWk) and brain imaging-derived phenotypes (IDPs; abbreviations described in Supplemental Table 1). Genetic correlation was calculated for 3,723 IDPs with heritability > 0. Here only results surviving false discovery rate (FDR q<0.05) are show.

GSCAN phenotype	Brain IDP	rg	se	p
AgeSmk	1219	-0.234	0.052	7.74E-06
SmkCes	1437	0.164	0.037	1.04E-05
CigDay	1736	0.143	0.033	1.24E-05

Supplemental Table 3 Genome-wide significant associations reported in the GWAS catalog for chr2:36,120,744-36,867,389 locus.

Phenotype	STRONGEST SNP	P-VALUE	MAPPED_GENE
Height	rs711245	1E-300	CRIM1, FEZ2
Height	rs9295	2E-254	FEZ2, CRIM1
Height	rs861051	2E-178	CRIM1
Height	rs6544015	4E-74	FEZ2
Height	rs17018786	1E-65	CRIM1
Height	rs711255	6E-56	CRIM1
Height	rs3770817	1E-51	CRIM1, FEZ2
Height	rs848602	5E-42	FEZ2
Height	rs711245	9E-39	CRIM1, FEZ2
Height	rs17018786	3E-36	CRIM1
Height	rs3770879	3E-32	CRIM1
Height	rs12997487	9E-32	CRIM1
Lung function (FEV1/FVC)	rs7571231	1E-29	CRIM1
Height	rs3770962	3E-29	CRIM1
Lung function (FEV1/FVC)	rs1179500	5E-28	CRIM1
Height	rs3755206	7E-27	CRIM1

QRS duration	rs3770901	2E-26	CRIM1
QRS duration	rs7562790	7E-23	CRIM1
Height	rs3755206	4E-21	CRIM1
Weight	rs9295	1E-20	FEZ2, CRIM1
Body surface area	rs9295	4E-20	FEZ2, CRIM1
Height	rs9295	9E-20	FEZ2, CRIM1
Height	rs17018786	4E-19	CRIM1
Smoking initiation	rs62134085	1E-18	MRPL50P1 - RPL21P36
Blood pressure (pleiotropy model 1 DBP adjusted for estimated causal effects x SBP)	rs861051	4E-18	CRIM1
Height	rs10172196	5E-18	FEZ2
Height	rs3755206	8E-17	CRIM1
Smoking initiation	rs13023731	1E-16	MRPL50P1 - RPL21P36
Core binding factor acute myeloid leukemia	rs13405170; rs11124500; rs1186353; rs1183799; rs17018460; rs13430583; rs1167437; rs1183798	8E-16	MRPL50P1 - RPL21P36; MRPL50P1 - RPL21P36; MRPL50P1 - RPL21P36; MRPL50P1 - RPL21P36; MRPL50P1 - RPL21P36; MRPL50P1 - RPL21P36; MRPL50P1 - RPL21P36; MRPL50P1 - RPL21P36
Body fat percentage	rs13410783	8E-15	FEZ2
Core binding factor acute myeloid leukemia	rs13405170; rs11124500; rs1186353; rs1183799; rs17018460; rs13430583; rs1167437; rs1183798	1E-14	MRPL50P1 - RPL21P36; MRPL50P1 - RPL21P36; MRPL50P1 - RPL21P36; MRPL50P1 - RPL21P36; MRPL50P1 - RPL21P36; MRPL50P1 - RPL21P36; MRPL50P1 - RPL21P36; MRPL50P1 - RPL21P36
Whole body fat mass (UKB data field 23100)	rs9295	1E-14	FEZ2, CRIM1
QRS duration	rs7562790	2E-14	CRIM1
Educational attainment	rs305191	2E-14	MRPL50P1 - RPL21P36
Height	rs848602	2E-14	FEZ2
Height	rs711245	4E-14	CRIM1, FEZ2
Chronic obstructive pulmonary disease liability (machine learning-based score)	rs4670562	5E-14	CRIM1
Chronotype	rs848552	5E-14	CRIM1
Morning person	rs848552	5E-14	CRIM1
ECG latent space	rs3821160	7E-14	CRIM1
Height	rs11900940	2E-13	MRPL50P1 - RPL21P36
Whole body fat mass (UKB data field 23100)	rs10197570	3E-13	FEZ2

Diverticular disease	rs10172196	4E-13	FEZ2
Lung function (forced vital capacity)	rs888078	9E-13	CRIM1
Height	rs3755206	1E-12	CRIM1
Blood pressure (pleiotropy model 2 SBP adjusted for estimated causal effects x DBP)	rs861051	1E-12	CRIM1
Hip circumference adjusted for BMI	rs3821150	2E-12	FEZ2
Height	rs11124513	2E-12	CRIM1
Height	rs17018663	7E-12	CRIM1
Metabolic biomarkers (multivariate analysis)	rs13410783	2E-11	FEZ2
Multi-trait sex score	rs11691767	2E-11	FEZ2
Electrocardiogram morphology (amplitude at temporal datapoints)	rs56036545	3E-11	CRIM1
Lung function (FEV1)	rs1477048	4E-11	CRIM1
Educational attainment	rs6721505	4E-11	RPL21P36 - CRIM1-DT
Height	rs145029810	9E-11	CRIM1
QRS duration	rs1523787	1E-10	CRIM1
Body mass index	rs3770799	1E-10	FEZ2
Height	rs3770820	1E-10	CRIM1, FEZ2
Height (standard GWA)	rs12712501	1E-10	CRIM1
Self-reported math ability (MTAG)	rs6708545	1E-10	RPL21P36 - CRIM1-DT
Hip circumference adjusted for BMI	rs7571231	1E-10	CRIM1
Corneal resistance factor (MTAG)	rs10172196	2E-10	FEZ2
Highest math class taken (MTAG)	rs6708545	2E-10	RPL21P36 - CRIM1-DT
Lung function (FEV1/FVC)	rs4670546	2E-10	CRIM1
Total PHF-tau (SNP x SNP interaction)	rs12613323 x rs1018491	2E-10	MRPL50P1 - RPL21P36 x SIGLEC1 - HSPA12B
Total PHF-tau (SNP x SNP interaction)	rs1953119 x rs4286311	2E-10	LINC00609 x LRRTM4 - RPL38P2
QRS duration	rs7562790	3E-10	CRIM1
Height	rs3821169	3E-10	CRIM1
Body mass index (MTAG)	rs3770799	5E-10	FEZ2
Central corneal thickness	rs848546	5E-10	CRIM1
Hip circumference adjusted for BMI	rs884215	5E-10	CRIM1
Waist circumference adjusted for body mass index	rs2276671	7E-10	CRIM1, FEZ2
Serum total protein levels	rs1045543	7E-10	CRIM1, FEZ2

Total PHF-tau (SNP x SNP interaction)	rs10196561 x rs10114094	7E-10	HDAC4-AS1 - LINC02991 x HMGB3P24 - RNF38
Offspring birth weight	rs1179494	9E-10	FEZ2
Body mass index	rs6544006	1.00E-09	RPL21P36 - CRIM1-DT
Haemorrhoidal disease	rs4670149	1.00E-09	FEZ2
Protein quantitative trait loci (liver)	rs141708688	1.00E-09	CRIM1
Smoking initiation (ever regular vs never regular) (MTAG)	rs11124501	1.00E-09	MRPL50P1 - RPL21P36
Morningness	rs848552	1.00E-09	CRIM1
Aspartate aminotransferase to alanine aminotransferase ratio	rs848608	2.00E-09	FEZ2
Body mass index	rs3770799	2.00E-09	FEZ2
Body fat percentage	rs13410783	2.00E-09	FEZ2
Waist circumference adjusted for body mass index	rs17488036	2.00E-09	FEZ2
Multi-trait sex score	rs10189344	2.00E-09	FEZ2
Diffuse plaques (SNP x SNP interaction)	rs10951234 x rs4670134	2.00E-09	PRR15-DT x MRPL50P1 - RPL21P36
Mean reticulocyte volume	rs11891181	3.00E-09	CRIM1
Central corneal thickness (MTAG)	rs10172196	3.00E-09	FEZ2
Intraocular pressure	rs10172196	3.00E-09	FEZ2
Multi-trait sex score	rs13010413	3.00E-09	FEZ2
Body mass index	rs3770799	4.00E-09	FEZ2
Protein quantitative trait loci (liver)	rs141708688	4.00E-09	CRIM1
Lung function (FVC)	rs2216099	4.00E-09	CRIM1
Seropositivity for shigella phage sp18 peptide (twist_65448)	rs1017153	5.00E-09	CRIM1
Self-reported math ability	rs181698387	5.00E-09	CRIM1-DT - CRIM1
Neuroblastoma	rs17018667	6.00E-09	CRIM1
Serum uric acid levels	rs2287084	6.00E-09	CRIM1
Educational attainment (MTAG)	rs6752813	7.00E-09	RPL21P36 - CRIM1-DT
Educational attainment	rs17480064	7.00E-09	CRIM1
Chronic elevation of alanine aminotransferase (cALT) levels	rs848559	8.00E-09	CRIM1
Body mass index	rs3770890	8.00E-09	CRIM1
QRS duration	rs1523787	1.00E-08	CRIM1
Protein quantitative trait loci (liver)	rs73922873	1.00E-08	RPL21P36 - CRIM1-DT
Body mass index	rs3770890	1.00E-08	CRIM1

Body mass index	rs3770890	1.00E-08	CRIM1
Cognitive performance (MTAG)	rs6728742	1.00E-08	CRIM1
Whole body fat mass (UKB data field 23100)	rs12615281	1.00E-08	CRIM1
Serum urate levels	rs2287084	1.00E-08	CRIM1
Hip circumference adjusted for BMI	rs115545989	1.00E-08	FEZ2
Chronic elevation of alanine aminotransferase (cALT) levels	rs848559	2.00E-08	CRIM1
Height	rs17019115	2.00E-08	FEZ2
Axial length	rs60806750	2.00E-08	RPL21P36 - CRIM1-DT
Hip circumference adjusted for BMI	rs756203	2.00E-08	CRIM1
Hip circumference adjusted for BMI	rs711245	2.00E-08	CRIM1, FEZ2
Neurofibrillary tangles (SNP x SNP interaction)	rs17332034 x rs10518756	2.00E-08	MRPL50P1 - RPL21P36 x TMEM87A
Neurofibrillary tangles (SNP x SNP interaction)	rs7401172 x rs13432611	2.00E-08	BRMS1L - ILF2P2 x CPS1 - RPS27P10
Electrocardiographic traits (multivariate)	rs4670556	3.00E-08	CRIM1
Cortical thickness	rs1017154	3.00E-08	CRIM1
Waist circumference adjusted for body mass index	rs111693785	3.00E-08	FEZ2
Body mass index	rs62132389	3.00E-08	MRPL50P1 - RPL21P36
Whole body fat mass (UKB data field 23100)	rs7604086	3.00E-08	CRIM1
Waist circumference adjusted for body mass index	rs848606	3.00E-08	FEZ2, CRIM1
Multi-trait sex score	rs848517	3.00E-08	CRIM1, FEZ2
Height	rs3770820	4.00E-08	CRIM1, FEZ2
Alistipes shahii abundance in stool	rs12473111	4.00E-08	MRPL50P1 - RPL21P36
Body mass index	rs2699189	4.00E-08	MRPL50P1 - RPL21P36
Body mass index	rs14291	4.00E-08	FEZ2
Body mass index (MTAG)	rs3770890	4.00E-08	CRIM1
Morning person	rs848552	4.00E-08	CRIM1

Supplemental Table 4 Latent causal variable (LCV) analysis between GSCAN phenotypes (smoking initiation SmkInit; Age at which the individual began smoking regularly, AgeSmk; cigarettes smoked per day, CigDay; smoking cessation, SmkCes; alcoholic drinks per week, DrnkWk) and brain imaging-derived phenotypes (IDPs; abbreviations described in Supplemental Table 1). LCV analysis was performed for GSCAN-IDP pairs with nominally significant genetic correlations. Here only results surviving false discovery rate (FDR q<0.05) are show.

GSCAN phenotype	Brain IDP	gcp	gcp pvalue	rho	rhoSE
AgeSmk	3266	0.786	3.73E-23	0.204	0.080
SmCes	2408	0.857	9.44E-22	0.215	0.078
CigDay	2402	0.474	2.14E-20	-0.198	0.080
DrnWk	2791	0.766	1.12E-16	0.184	0.073
AgeSmk	958	0.666	7.08E-10	0.222	0.079
DrnWk	3077	0.582	1.39E-09	-0.178	0.087
AgeSmk	864	0.629	1.56E-09	0.161	0.057
AgeSmk	1111	0.670	3.34E-09	-0.239	0.069
AgeSmk	2962	0.571	4.24E-09	0.179	0.087
AgeSmk	1000	0.628	1.07E-08	0.252	0.085
AgeSmk	492	0.600	1.45E-07	0.184	0.057
SmCes	3866	0.617	4.62E-07	-0.188	0.082
AgeSmk	3056	0.609	6.97E-07	0.175	0.083
AgeSmk	3402	0.573	2.93E-06	0.210	0.062
AgeSmk	707	0.616	4.32E-06	0.142	0.060
AgeSmk	977	0.644	6.27E-04	0.277	0.085
DrnWk	3200	0.376	8.95E-04	-0.175	0.072
CigDay	2590	0.038	0.001	-0.198	0.068
CigDay	2933	0.354	0.002	-0.170	0.065

Supplemental Table 5 Mendelian randomization (MR) analysis between GSCAN phenotypes (smoking initiation SmkInit; Age at which the individual began smoking regularly, AgeSmk; cigarettes smoked per day, CigDay; smoking cessation, SmkCes; alcoholic drinks per week, DrnkWk) and brain imaging-derived phenotypes (IDPs; abbreviations described in Supplemental Table 1). MR analysis was performed for GSCAN-IDP pairs surviving false discovery rate (FDR q<0.05) using TwoSampleMR package (Hemani et al., 2018). Converging LCV and MR statistically significant results are highlighted in red. Direction consistency between LCV and MR analyses are indicated in light blue.

GSCAN phenotype	Brain IDP	gcp	gcp pvalue	rho	rhoSE	IVW b	se	p	n_IVS	LCV rho - IVW beta concordance
DrnWk	3200	0.38	8.95E-04	-0.175	0.072	-0.037	0.016	0.021	18	Y
AgeSmk	3402	0.57	2.93E-06	0.210	0.062	-0.029	0.020	0.158	23	N
AgeSmk	977	0.64	6.27E-04	0.277	0.085	-0.026	0.020	0.195	27	N
CigDay	2933	0.35	0.002	-0.170	0.065	-0.025	0.021	0.227	23	Y
SmCes	2408	0.86	9.44E-22	0.215	0.078	-0.018	0.019	0.355	19	N
AgeSmk	1000	0.63	1.07181E-08	0.252	0.085	0.017	0.019	0.363	18	Y
AgeSmk	2962	0.57	4.24E-09	0.179	0.087	-0.018	0.020	0.370	23	N
AgeSmk	3266	0.79	3.73E-23	0.204	0.080	0.018	0.026	0.484	14	Y
AgeSmk	492	0.60	1.45E-07	0.184	0.057	-0.013	0.020	0.509	29	N
AgeSmk	1111	0.67	3.34E-09	-0.239	0.069	-0.011	0.018	0.551	34	Y
CigDay	2590	0.04	0.001	-0.198	0.068	0.013	0.025	0.596	21	N
SmCes	3866	0.62	4.62E-07	-0.188	0.082	-0.010	0.025	0.703	21	Y
DrnWk	3077	0.58	1.39E-09	-0.178	0.087	-0.006	0.016	0.718	24	Y
AgeSmk	3056	0.61	6.97E-07	0.175	0.083	0.006	0.019	0.749	24	Y
DrnWk	2791	0.77	1.12E-16	0.184	0.073	0.003	0.014	0.808	26	Y
CigDay	2402	0.47	2.14E-20	-0.198	0.080	-0.004	0.018	0.809	27	Y
AgeSmk	707	0.62	4.32E-06	0.142	0.060	0.003	0.019	0.891	22	Y
AgeSmk	958	0.67	7.08E-10	0.222	0.079	0.001	0.018	0.961	33	Y
AgeSmk	864	0.63	1.56E-09	0.161	0.057	0.000	0.017	0.999	32	Y

Appendix

Published results

Received: 8 July 2023 | Accepted: 19 February 2024

DOI: 10.1111/acer.15288

ALCOHOL CLINICAL & EXPERIMENTAL RESEARCH

RESEARCH ARTICLE

Contribution of infectious diseases to the selection of *ADH1B* and *ALDH2* gene variants in Asian populations

Giovanni Deiana[1] | Ruinan Sun[2] | Jie Huang[3,4] | Valerio Napolioni[5] | Roberto Ciccocioppo[1]

[1]Center for Neuroscience, Pharmacology Unit, School of Pharmacy, University of Camerino, Camerino, Italy

[2]Department of Public and Ecosystem Health, Cornell University College of Veterinary Medicine, Ithaca, New York, USA

[3]School of Public Health and Emergency Management, Southern University of Science and Technology, Shenzhen, China

[4]Institute for Global Health and Development, Peking University, Beijing, China

[5]School of Biosciences and Veterinary Medicine, University of Camerino, Camerino, Italy

Correspondence
Valerio Napolioni, School of Biosciences and Veterinary Medicine, University of Camerino, Via Gentile da Varano III, 62032 Camerino, Italy.
Email: valerio.napolioni@unicam.it

Funding information
Ministero dell'Universitàe della Ricerca, Grant/Award Number: PRIN 2017SXEXT5; The NIH/NIAAA stands for National Institute of Health/National Institute on Alcohol Abuse and Alcoholism, Grant/Award Number: AA014351 and AA017447

Abstract

Background: The gene variants *ADH1B**2 (Arg48His, rs1229984) and *ALDH2**2 (Glu504Lys, rs671) are common in East Asian populations but rare in other populations. We propose that selective pressures from pathogen exposure and dietary changes during the neolithic transition favored these variants. Thus, their current association with differences in alcohol sensitivity likely results from phenotypic plasticity rather than direct natural selection.

Methods: Samples sourced from the Allele Frequency Database (ALFRED) were utilized to compute the average frequency of *ADH1B**2 and *ALDH2**2 across 88 and 61 countries, respectively. Following computation of the average national allele frequencies, we tested the significance of their correlations with ecological variables. Subsequently, we subjected them to Principal Component Analysis (PCA) and Elastic Net regularization. For comprehensive evaluation, we collected individual-level phenotypic associations, compiling a Phenome-Wide Association Study (PheWAS) spanning multiple ethnicities.

Results: Following multiple testing correction, *ADH1B**2 displayed significant correlations with Neolithic transition timing ($r = 0.405$, $p.adj = 2.013e\text{-}03$, $n = 57$) and historical trypanosome burden ($r = -0.418$, $p.adj = 0.013$, $n = 57$). The first two components of PCA explained 47.7% of the total variability across countries, with the top three contributors being the historical indices of population density and trypanosome and leprosy burdens. Historical burdens of the *Mycobacteria* tuberculosis and leprosy were the sole predictive variables with positive coefficients that survived Elastic Net regularization.

Conclusions: Our analyses suggest that *Mycobacteria* may have played a role in the joint selection of *ADH1B**2 and *ALDH2**2, expanding the "toxic aldehyde hypothesis" to include *Mycobacterium leprae*. Additionally, our hypothesis, linked to dietary shifts from rice domestication, emphasizes nutritional deficiencies as a key element in the selective pressure exerted by *Mycobacteria*. This offers a plausible explanation for the high frequency of *ADH1B**2 and *ALDH2**2 in Asian populations.

KEYWORDS
alcohol use disorder, East-Asia, environment, evolution, natural selection

INTRODUCTION

According to the latest global status report on alcohol and health by the World Health Organization (WHO), the South-East Asian Region (SEAR) has the second lowest Alcohol-Per capita-Consumption (APC) in the world (World Health Organization, 2018). Cultural and social factors are known to affect the drinking behaviors of Asian subjects, who show the lowest alcohol consumption levels compared to other ethnicities (Lui & Zamboanga, 2019). APC levels have been increasing in the SEAR, but the percentage of current drinkers did not significantly change, being constant at around 33% from 2000 to 2016 (World Health Organization, 2018). This trend stays stable also when looking at differences in APC among countries within SEAR (World Health Organization, 2018). APC levels in SEAR countries match the worldwide distribution of two of the main genetic variants associated with reduced alcohol intake, namely, *ADH1B**2 (Arg48His, rs1229984) and *ALDH2**2 (Glu504Lys, rs671) (Zhou et al., 2022). Indeed, these two variants exhibit a robust association with decreased alcohol intake in the East Asian population (Chang et al., 2023; Hashimoto et al., 2016), reaching their peak frequency within this demographic (Polimanti & Gelernter, 2018; Zhang et al., 2021). In contrast, they are found to be rare and less effective in other populations (Schaschl et al., 2022; Whitfield, 2002).

Encoding the beta subunit of class I of the alcohol dehydrogenases (ADH), *ADH1B* is responsible for the interconversion between alcohols and aldehydes or ketones with the reduction of NAD^+ to NADH (Zakhari, 2006). *ALDH2* encodes the metabolizing enzyme for many reactive and oxidative stress-generated aldehydes, such as acetaldehyde generated by ADH that is converted to acetate (Chen et al., 2014). Due to its high reactivity and genotoxicity, acetaldehyde has been classified by the International Agency for Research on Cancer (IARC) of WHO as carcinogenic to humans in Group 1 tissues (Baan et al., 2007). Compared to the enzyme encoded by the ancestral allele, *ADH1B*2* converts alcohol to acetaldehyde at a fast rate, causing a huge accumulation of blood acetaldehyde (Whitefield, 2002). The increased amount of acetaldehyde cannot be oxidized into acetate by *ALDH2*2*, which encodes an inactive enzyme (Crabb et al., 1989). The combination of *ADH1B*2* and *ALDH2*2* increases alcohol toxicity, which results in skin rash and coughing resembling an allergic reaction (Whitfield, 2002).

Cumulating evidence shows that rice domestication exerted selective pressures on alcohol and acetaldehyde dehydrogenases among those populations that adopted rice farming during the Neolithic transition in SEAR (Landini et al., 2021; Zhu et al., 2021). As rice constitutes the primary staple food for most populations in the region (Muthayya et al., 2014), the diffusion of rice domestication across East Asia primarily followed a demic pattern (Cobo et al., 2019), influencing and molding local cultures and societies (Bray, 1994; Thomson et al., 2018). Rice domestication was paralleled by the spread of infectious diseases that changed the demographic landscape through SEAR as paddy rice forced low relational mobility (Cobo et al., 2019). Thus, it is conceivable that paddy rice may have imposed new burdens of infection, changing the demographic landscape of the area (Cobo et al., 2019; Liu et al., 2017). Indeed, it is remarkable that the emergence of agriculture is associated with an increased allocation of energy to immune functions (Wells & Stock, 2020).

Additional phenotypes correlated with alcohol intake, such as risky sexual behaviors (Hutton et al., 2008) and disgust sensitivity (Stafford et al., 2020), as well as broader associations with all risky-type behaviors, have been linked to the historical burden of pathogens (Schaller, 2011). The shared phenotypical association with a reduction in alcohol intake of *ADH1B*2* and *ALDH2*2* (Serajuddin & Maeda, 2018; Zhou et al., 2022) may therefore be the current result of the effect of the selective pressure exerted by pathogens and changes in dietary habits following the neolithic transition in SEAR.

To corroborate this hypothesis, we took advantage of previous works and publicly available data to investigate the correlation between human geography, agriculture, and farming, the exposure to infectious disease, and the worldwide distribution of the two alleles employing a country-based approach. Furthermore, we attempted to provide a causal explanation linking our findings to changes in aldehyde metabolism caused by *ADH1B**2 and *ALDH2**2.

METHODS

Allele frequency and ecological data search, extraction, and harmonization

A total of 510 sample populations from the Allele Frequency Database (ALFRED) (Rajeevan et al., 2012), containing 115,303 subjects self-identifying in 284 different ethnicities, was included to calculate the average frequency of *ADH1B*2* across 88 nation-states. For *ALDH2*2*, 212 sample populations composed of 35,540 subjects, self-identifying in 142 ethnicities, were used to calculate its average frequency across 61 nation states. More details on the sample demographics are provided in Tables S1 and S2. Ecological data describing food accessibility, human geography, and historical exposure to infectious diseases were collected from a dataset previously assembled by Fedderke et al. (2017) containing 207 nations as described in Table 1 and Table S4.

Samples from ALFRED (Rajeevan et al., 2012) were divided by country and reported ethnicity to compute a mean weighted on their relative sample size. National allelic frequencies were computed through a second weighted mean combining previous results accordingly to the relative share of each reported ethnicity in the demographic composition of their corresponding sampling country, as described in The World Factbook (The World Factbook, 2021). The resulting average national allele frequencies were mapped using the rworldmap R package (South, 2011). For Brazil, the United States, and Australia, only native population samples were considered.

To standardize our dataset, missing values in the distribution of *ALDH2**2 frequency underwent imputation through Multivariate Imputation by Chained Equations (MICE) using random forests. This imputation method was implemented using the miceRanger R package

ALCOHOL CLINICAL & EXPERIMENTAL RESEARCH

TABLE 1 Set of variables from Fedderke et al. (2017).

Name	Description
fstdist1500_eth	Genetic distance to Ethiopia (1500 match)
fstdist1500_uk	Genetic distance to the U.K. (1500 match)
efrac	Ethnic fractionalization index
ln_pop1500	Population in 1500 CE LN
ln_pd1500	Population density in 1500 CE LN
ln yst	Neolithic transition timing until the year 1500 CE LN
ln_arable	Percentage of arable land LN
ln_temp	Temperature LN
ln_precip	Precipitation LN
his_leishmanias	Index of historical prevalence of leishmaniasis
his_schistosomes	Index of historical prevalence of schistosomes
his_filariae	Index of historical prevalence of filariae
his_leprosy	Index of historical prevalence leprosy
his_typhus	Index of historical prevalence typhus
his_malaria	Index of historical prevalence malaria
his_trypanosomes	Index of historical prevalence of trypanosomes
his_dengue	Index of historical prevalence of dengue
his_tuberculosis	Index of historical prevalence tuberculosis

Note: List of variables employed in our analysis. When "LN" is reported values are distributed in a natural logarithm form.

Abbreviation: LN, Natural logarithm.

(Wilson, 2021). The final imputed dataset was the result of five iterations using these variables as predictors in the order: genetic distance to Ethiopia—1500 match (fstdist1500_eth), genetic distance to the United Kingdom—1500 match (fstdist1500_uk) and ethnic fractionalization index (efrac) Table S4. Other parameters were set as default.

Correlation analyses

Given the nonparametric nature of our dataset, we conducted a Kendall hypothesis test on complete observations to explore the potential correlation between the variables of interest, followed by an exploratory Principal Component Analysis (PCA).

The Kendal hypothesis test was carried out as implemented by the corr.test function in the psych R package on the following variables: national average frequency of *ADH1B*2* (adh1b_2), national average frequency of *ALDH2*2* (aldh2_2), natural logarithm of distribution of population in 1500 CE (ln_pop1500), natural logarithm of distribution of population density in 1500 CE (ln_pd1500), natural logarithm of distribution of neolithic transition timing until the year 1500 CE (ln_yst1500), natural logarithm of percentage of arable land (ln_arable), natural logarithm of temperature (ln_temp), natural logarithm of precipitation (ln_precip), index of historical prevalence of leishmaniasis (his_leishmanias), index of historical prevalence of schistosomes (his_schistosomes), index of historical prevalence of filariae (his_filariae), index of historical prevalence leprosy (his_leprosy), index of historical prevalence typhus (his_ typhus), index of historical prevalence of malaria (his_malaria), index of historical prevalence of trypanosomes (his_trypanosomes), index of historical prevalence of dengue (his_dengue), and index of historical prevalence tuberculosis (his_tuberculosis). The test considered only complete observation by selecting the use = "complete" option; results were adjusted by Bonferroni correction.

Principal component analysis

After excluding rows with missing data from the imputed dataset, we conducted PCA using the PCA function within the FactoMineR R package. Subsequently, we analyzed the PCA results using functions from the factoextra R package. The contribution of each variable to the first two dimensions was determined using the fviz_contrib function. To visualize variability across nation states, we utilized the fviz_pca_var function to plot the results within the first two dimensions. The contribution of each variable was depicted by the length and orientation of their eigenvectors. Additionally, the quality of the representation of variability within each nation state was assessed by incorporating the measure of squared cosines of their eigenvectors.

Considering the significant correlations among ecological variables post-Bonferroni correction and the orientation of the square cosine of their eigenvectors (cos2), we excluded those surpassing a variance inflation factor (VIF) cutoff of 3.00 as predictors in the subsequent analysis. This step was undertaken to alleviate potential collinearity effects on average allelic distributions. Squared eigenvector cosines measure the similarity between eigenvectors of a matrix. It involves calculating the cosine of the angle between two eigenvectors, squaring it to focus on similarity, with high values indicating strong alignment.

Elastic net regularization analysis

The effectiveness of retained variables in explaining variance was evaluated using Elastic Net regularization, implemented in the glmnet R package through the cv.glmnet function (Friedman et al., 2023). This approach enabled us to alleviate and ultimately eliminate factors that were negatively impacting the fitting to the distributions of average national allelic frequencies, both for single alleles and combined alleles. The first group of predictors included the following ecological variables: natural logarithm of distribution population in 1500 CE (ln_pop1500), natural logarithm of distribution population density in 1500 CE (ln_pd1500), natural logarithm of distribution neolithic transition timing until the year 1500 CE (ln_yst1500), natural logarithm of distribution percentage of arable land (ln_arable), natural logarithm of distribution precipitation (ln_precip), index of historical prevalence of leishmaniasis (his_leishmanias), index of historical prevalence of schistosomes (his_schistosomes),

index of historical prevalence leprosy (his_leprosy), index of historical prevalence typhus (his_ typhus), index of historical prevalence of trypanosomes (his_ trypanosomes), and index of historical prevalence tuberculosis (his_ tuberculosis). This set of predictors was applied to the distribution of national average frequencies of *ADH1B*2* (adh1b_2) and *ALDH2*2* (aldh2_2) as outcome variables. In the second group of predictors, adh1b_2 was included, and this set was combined with aldh2_2 as the sole outcome variable. All regularizations were computed 10 times with gamma values set as default (0, 0.25, 0.5, 0.75, 1).

Phenome-wide association study (PheWAS)

To enhance the interpretation of results, phenotypic associations linked to each allele were extracted from the meta-analysis conducted on the Global Biobank (McInnes et al., 2019), which contains data compiled from summary statistics from the Million Veteran Program (MVP) (Gaziano et al., 2016), the Japanese Biobank (Nagai et al., 2017), and the UK Biobank (Sudlow et al., 2015). These associations were then compiled to create a PheWAS that takes into consideration exposure to infectious diseases and dietary changes. Phenotypic associations categorized as "Early life factors," "Disease outcome," "Anthropometrics," and "Biomarkers" were filtered for FDR ≤0.005 calculated according to the Benjamini–Hochberg correction method.

RESULTS

The average allelic frequencies of *ADH1B*2* and *ALDH2*2* were calculated for 88 and 61 nation states (Figures 1C and 2C), respectively. Most countries showed low frequency levels for both alleles, with their distributions reaching their density peak on the lower tail (Figures 1A and 2A). This is especially true for *ALDH2*2*, whose variability (var = 0.004) was lower than that of *ADH1B*2* (var = 0.047). After imputation, *ALDH2*2* variability was further decreased (var = 0.0004) Table S3.

Geographically, the two alleles were similarly distributed, reaching their highest frequencies in SEAR, with Japan being at

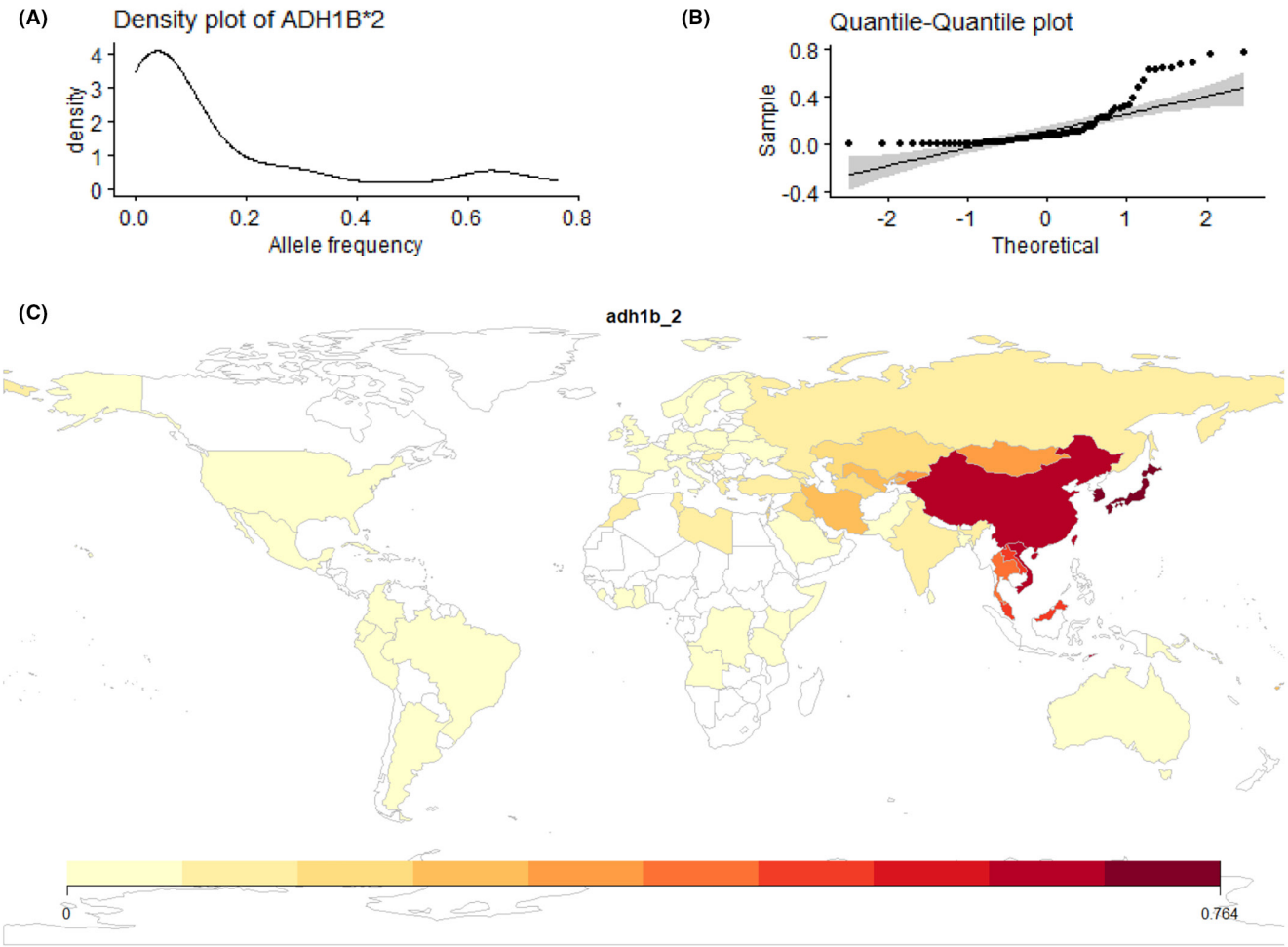


FIGURE 1 World-wide *ADH1B*2* frequency. (A) Density plot of frequency distribution of *ADH1B*2*, (B) QQ-plot of frequency distribution of *ADH1B*2* compared to normal distribution, and (C) National distribution of *ADH1B*2* frequency.

ALCOHOL CLINICAL & EXPERIMENTAL RESEARCH

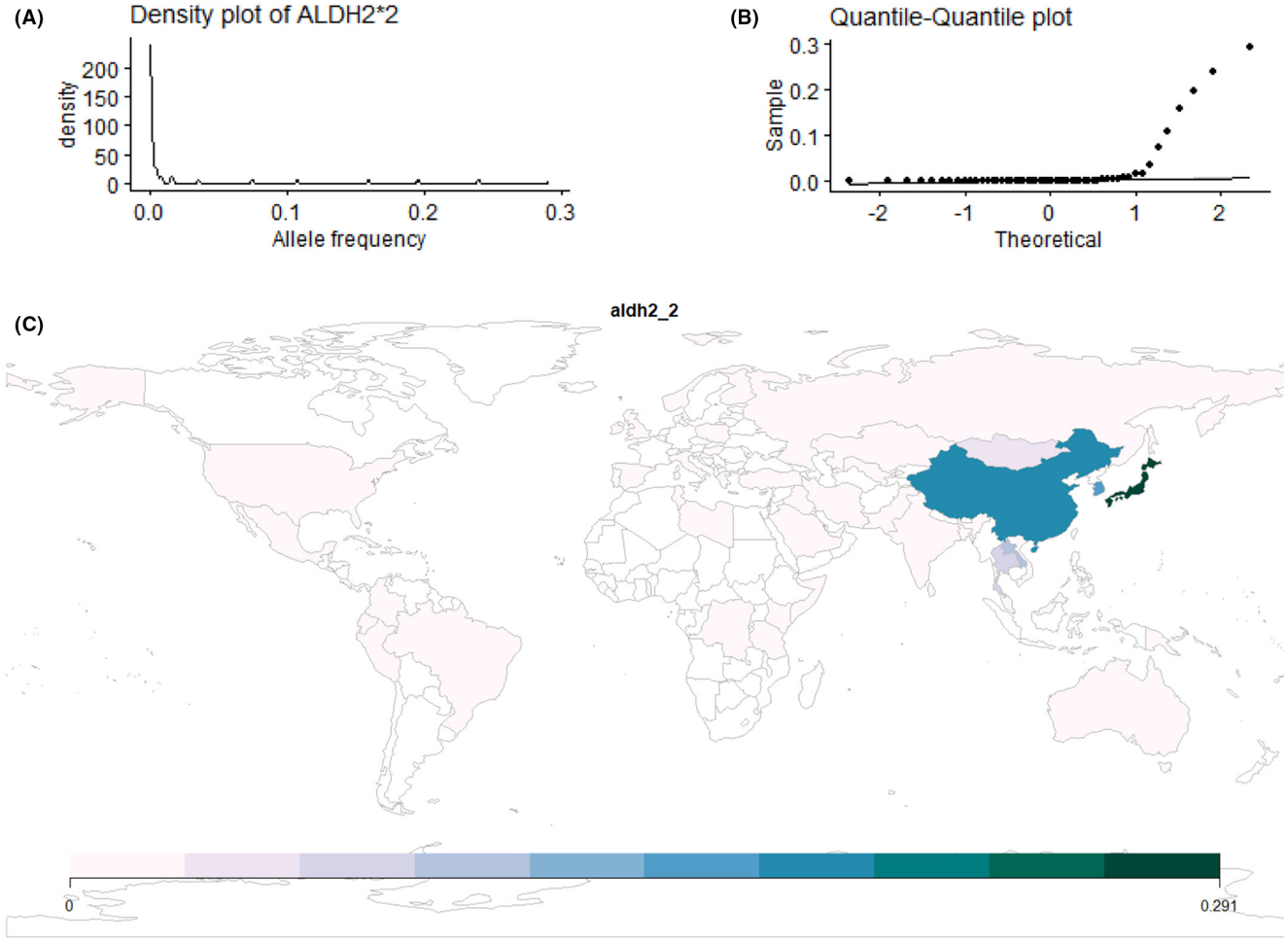


FIGURE 2 World-wide *ALDH2*2* frequency. (A) Density plot of frequency distribution of *ALDH2*2*, (B) QQ-plot of frequency distribution of *ALDH2*2* compared to normal distribution, and (C) National distribution of *A LDH2*2* frequency.

the extreme upper tail of their distribution (*ADH1B*2* = 76.4%, *ALDH2*2* = 29%) (Figures 1C and 2C).

While the two alleles were significantly correlated ($r = 0.267$, $p = 6.411e\text{-}03$, $n = 57$), after applying a Bonferroni correction for multiple testing only *ADH1B*2* showed significant correlations, with the Neolithic transition timing until the year 1500 CE (ln_yst1500; $r = 0.405$, $p = 1.480e\text{-}05$, $p.adj = 2.013e\text{-}03$, $n = 57$) and the historical burden of trypanosomes infection (his_trypanosomes; $r = -0.418$, $p = 1.010e\text{-}04$, $p.adj = 0.013$, $n = 57$), which were also negatively correlated with each other ($r = -0.522$, $p = 1.537e\text{-}06$, $p.adj = 1.796e\text{-}03$, $n = 57$) (Figure 3). After Bonferroni correction, the number of significant correlations between ecological variables was reduced, except for some describing the historical burden of many infectious diseases, which remained positively correlated (Figure 3).

Variability across ecological variables was partitioned into 10 main components by PCA; the first two principal components described as 27.8% and 19.9%, respectively, of the total variance (Figure 4A), ensuring within countries a representation of more than 50% of the total variance according to square eigenvectors cosines (cos^2). The most relevant contributors to the first component were the historical burden of infectious diseases led by trypanosomes and population indexes (population density in 1500 CE and population in 1500 CE); the percentage of arable land was the main contributor to the second principal component (Figure 4B). Southern European, Asian, and Middle Eastern countries appear to lay close together due to population density in 1500 CE, population in 1500 CE, and percentage of arable land, while most Oceanian, South American, and North African countries were differently oriented due to the historical burden of infectious diseases (Figure 4C).

Due to the squared eigenvectors exhibiting cosine similarities among ecological variables, forming two nearly opposite directional sets and implying a high degree of collinearity (Figure 4C), and with both allele frequencies demonstrating significant inflation at the extremes of their distributions (Figures 1B and 2B), temperature (VIF = 3.20), historical prevalence of filariae (VIF = 4.38), historical prevalence of malaria (VIF = 3.51), and historical prevalence of dengue (VIF = 3.90) were excluded based on their respective VIF values.

Elastic net regularization revealed that, at the lowest values of λ, allowing for the best fit, population in 1500 CE, historical prevalence of leprosy, and historical prevalence of tuberculosis were the sole variables maintaining positive coefficients across all nets. For

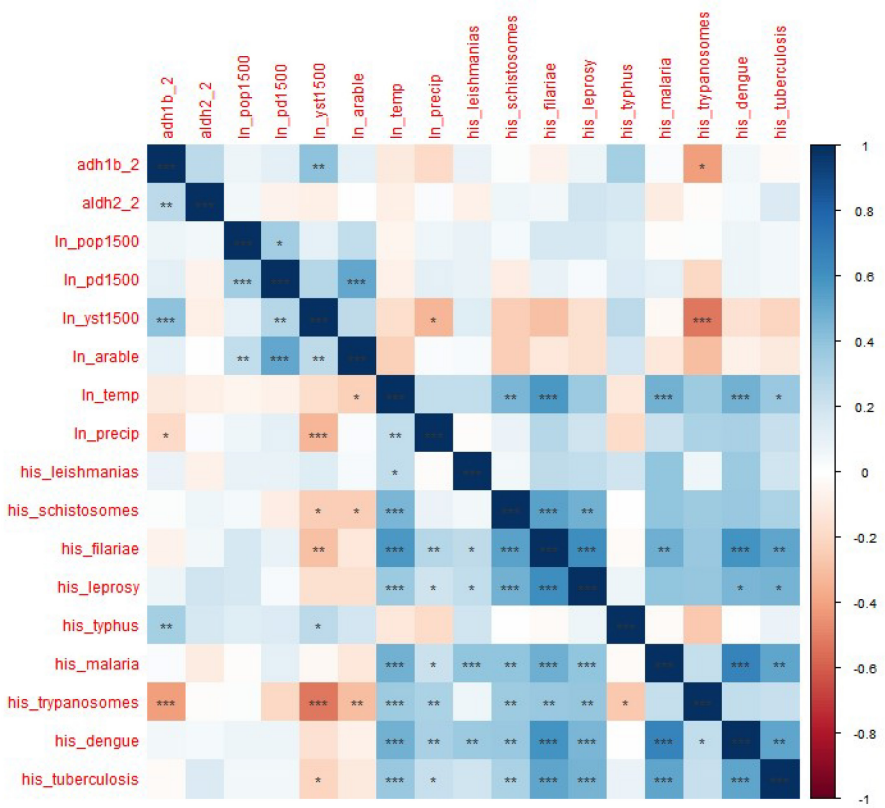


FIGURE 3 Correlation matrix for Kendall test (upper matrix reports Bonferroni adjusted *p* values). Asterisks meaning: ***$p < 0.001$, **$p < 0.01$, *$p < 0.05$.

larger λ values, resulting in fit alterations up to one standard error, no variable exhibited a nonzero coefficient except for ln_leprosy. Notably, this pattern held true for nets considering *ALDH2*2* as the single dependent variable, only when *ADH1B*2* was included among predictors (Figure 5). Furthermore, when both allelic distributions were considered, variability in the fit due to different γ values was diminished (Figure 5).

In the Global Biobank (McInnes et al., 2019), meta-analyzed phenotypic associations categorized as "Early life factors," "Disease outcome," "Anthropometrics," and "Biomarkers" were collected for both alleles, reaching a total amount of 105 for *ADH1B*2* and 58 for *ALDH2*2*. Of those associated with *ADH1B*2*, 33 traits passed a minimum threshold of FDR ≤0.005, with no one directly linked to an infectious disease; the most significant was the reduction of insulin-like growth factor 1 (IGF-1) levels ($n = 423{,}324$, beta = −0.090, FDR = 2.89E-36) (Table 2). For *ALDH2*2*, six phenotypic associations passed a minimum threshold of FDR ≤0.005: in particular, the increased propensity to develop nasal polyps was the most significant one ($n = 6546$, beta = 7.108, FDR = 0.001) (Table 3).

DISCUSSION

According to the results from Elastic Net regularization, the historical burden of tuberculosis (his_tuberculosis) and leprosy (his_leprosy) emerged as the sole positive predictors for the distribution of national average frequencies of *ADH1B*2* and *ALDH2*2*. This association appears to be primarily driven by *ADH1B*2*, as in models considering *ALDH2*2* alone as the dependent variable, his_leprosy was retained only when *ADH1B*2* was included as a predictor. The PCA results suggest that the correlation observed between *ADH1B*2* and the distance to the Neolithic transition (ln_yst1500) may potentially be attributed to other variables exhibiting a positive correlation with it. The collective findings of our analyses indicate that exposure to mycobacterial infections may have played a role in the concurrent selection of these two alleles among Neolithic rice farmers (Wang, Mi, et al., 2018; Wang, Wang, et al., 2018).

Consistent with our findings, a recent exome-wide association study among Han Chinese linked *ALDH2*2* with heightened susceptibility to Mycobacterium leprae ($p = 2.00 \times 10{-}20$, OR = 1.35) (Wang, Mi, et al., 2018). This association was subsequently replicated in another Han Chinese cohort but not observed in African populations, where the maximum frequency of *ALDH2*2* in the sample was below 1% (Gilchrist et al., 2022). Furthermore, clinical research conducted among Korean patients linked *ALDH2*2* with a diminished burden of infection from Mycobacterium tuberculosis, particularly among subjects also afflicted by alcoholism, along with *ADH1B*2* (Park et al., 2014).

The dissemination of *ADH1B*2* in Southeast Asia occurred sixty thousand years prior to that of *ALDH2*2*, as demonstrated by Li et al. (2011) and Luo et al. (2009). Mutations within this extensive

ALCOHOL CLINICAL & EXPERIMENTAL RESEARCH

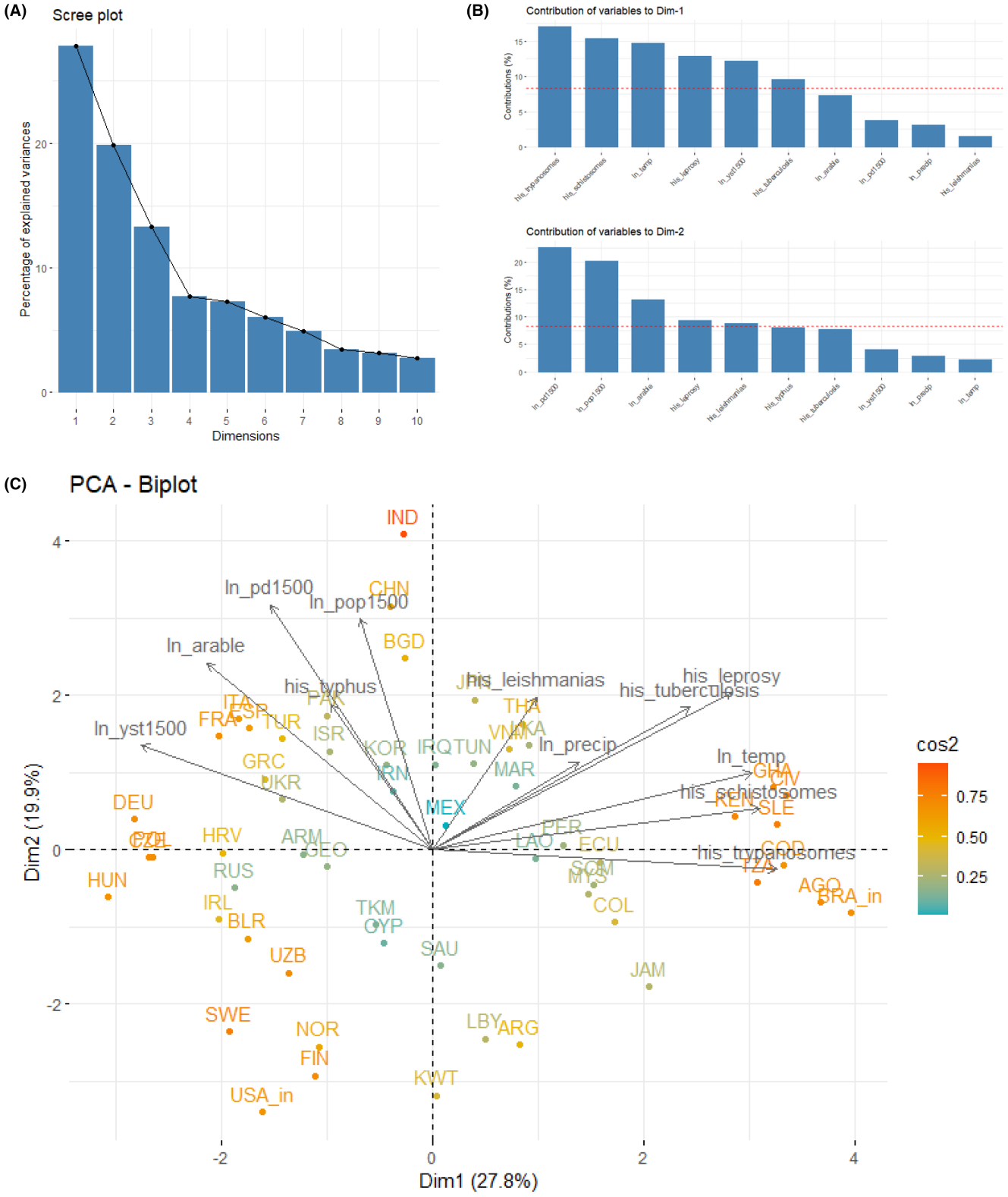


FIGURE 4 Principal component analysis on significant values. (A) Percentage of variance explained by each principal component, (B) contribution to the first pair of principal components of each variable, (C) plotting of vectors describing the variability explained by the first pair of principal components of each variable.

timeframe, taking place in ecological niches, substantiate alterations in their patterns of natural selection. Consequently, these two alleles are presently associated solely with five common phenotypes in our Phenome-Wide Association Studies (PheWAS), all falling within the "Anthropometrics" category, such as "Body fat percentage," "BMI," or "Trunk fat percentage" (see Tables 2 and 3).

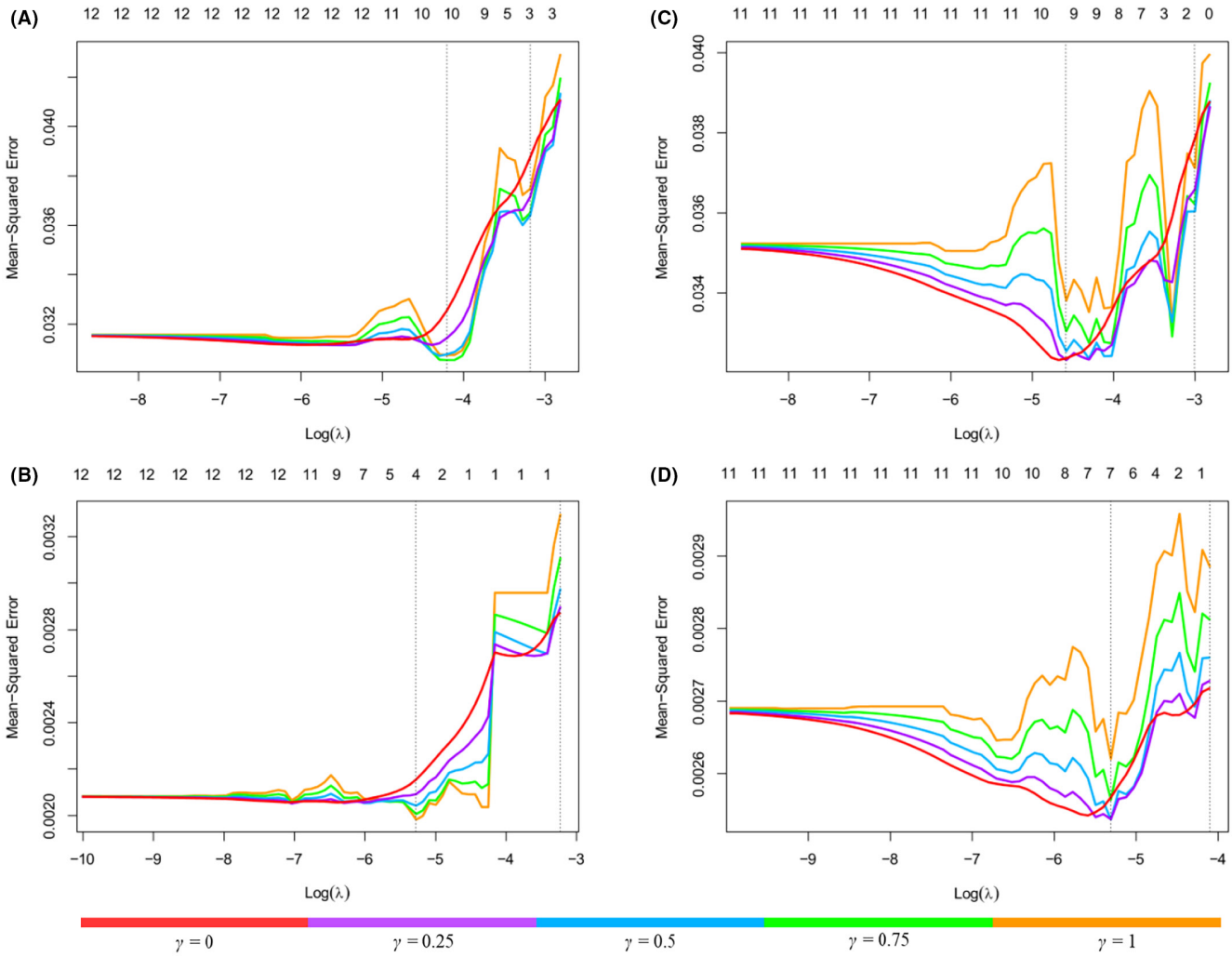


FIGURE 5 Relaxed Lasso regression model. (A) Elastic net for *ADH1B*2* + *ALDH2*2* = *y*, (B) Elastic net for *ADH1B*2* = x_i; *ALDH2*2* = *y*, (C) Elastic net for *ADH1B*2* = *y*, (D) Elastic net for *ALDH2*2* = *y*. The numbers at the top of each panel refer to the number of predictors with a nonzero coefficient.

Rice-based diets have been linked to deficiencies in lipid absorption, resulting in reduced food intake and BMI measurements (Lim et al., 2016; Yang et al., 2012). *ADH1B**2, associated with higher levels of LDL-C in Asian populations (Luo et al., 2023), might have conferred survival advantages to carriers due to the decreased energy intake resulting from their dietary habits. Recent evidence of positive selection on variants related to ADH, including *ADH1B**2, has been identified in populations in Ethiopia following the advent of agriculture (McQuillan et al., 2022). However, while rice domestication was a key catalyst for the Neolithic transition in Southeast Asia and the Pacific, its dissemination appears to have been greatly influenced by demographic factors (Bray, 1994), which are closely linked to mycobacterial infections (Benjak et al., 2018; Luo et al., 2015; Okazaki et al., 2019). We propose that alterations in dietary practices in Southeast Asia due to the spread of rice farming resulted in chronic retinol deficiency (Tang et al., 2009) and disrupted lipid metabolism (Lim et al., 2016; Yang et al., 2012), thereby heightening selective pressures from infectious diseases in the region. Indeed, retinoid acid has been shown to enhance antimicrobial responses in epithelial and mucous tissues (Huang et al., 2018), playing a protective role in the response to mycobacterial infections (Fairley et al., 2023; Kim et al., 2019; Tenório De Menezes et al., 2023), alongside lipid metabolism, which mediates host inflammatory responses (Fairley et al., 2023; Kim & Shin, 2023).

Our hypotheses align with the idea that decreased detoxification of aldehydes may confer beneficial effects against pathogens by promoting immune cell activation through the cytokine interferon-gamma (Darwin & Stanley, 2022). The advent of rice domestication conferred a significant advantage to farming populations in Southeast Asia and the Pacific, facilitating its spread throughout the region during the Neolithic transition. Subsequent shifts in the demographic landscape were profoundly influenced by mycobacterial infections. Elevated levels of circulating aldehydes may have provided an advantage to ancient rice farmers, enhancing their fitness against the high prevalence of mycobacterial infections. This enhancement of immunity could serve as

ALCOHOL CLINICAL & EXPERIMENTAL RESEARCH

TABLE 2 Phenotypic traits associated to *ADH1B*2*.

Class	Code	Description	*N*	Info	Effect	SE	*p*	FDR
biom	INI10030770	IGF-1	423,324	beta	−0.0895	0.0067	2.78E-38	2.89E-36
biom	INI10030610	Alkaline phosphatase	425,779	beta	0.0754	0.0066	6.07E-27	6.31E-25
biom	INI20030780	LDL cholesterol	424,960	beta	0.0073	0.0066	3.53E-25	3.67E-23
dis_out	HC328	Gout	9506	or	0.66113	0.0391	3.13E-23	3.26E-21
dis_out	HC1215	TTE gout	15,378	or	0.71527	0.0322	2.55E-22	2.65E-20
biom	INI10030670	Urea	425,469	beta	−0.0666	0.0066	1.02E-20	1.06E-18
dis_out	HC719	TTE mental and behavioral disorders due to use of alcohol	8788	or	2.21	0.0728	1.40E-20	1.46E-18
biom	INI10030860	Total protein	389,120	beta	0.0649	0.0069	8.53E-18	8.87E-16
biom	INI20030690	Cholesterol	425,743	beta	0.0598	0.0066	1.95E-16	2.02E-14
biom	INI10030880	Urate	425,235	beta	−0.0584	0.0066	1.39E-15	1.45E-13
biom	INI20030640	Apolipoprotein B	423,632	beta	0.0539	0.0066	5.42E-13	5.63E-11
dis_out	HC215	Hypertension	153,912	or	1.1	0.0144	1.81E-11	1.88E-09
biom	INI10030890	Vitamin D	406,303	beta	−0.0484	0.0068	9.89E-10	1.03E-07
dis_out	HC273	Essential hypertension	102,191	or	1.08177	0.1614	1.70E-08	1.77E-06
antro	INI23112	Leg fat mass R	442,807	beta	0.034	0.0051	3.20E-08	3.33E-06
antro	INI23099	Gamma glutamyltransferase	442,599	beta	0.0317	0.0049	7.37E-08	7.66E-06
biom	INI10030730	Body fat percentage	425,539	beta	0.0431	0.0066	7.68E-08	7.99E-06
antro	INI23116	Whole body fat index	442,781	beta	0.0327	0.0051	1.10E-07	1.15E-05
antro	INI23100	Leg fat mass L	442,115	beta	0.0403	0.0063	1.14E-07	1.19E-05
antro	INI23111	Leg fat percentage R	442,814	beta	0.0253	0.004	2.80E-07	2.91E-05
antro	INI23115	Non-albumin protein	442,786	beta	0.0247	0.004	4.62E-07	4.80E-05
biom	INI30030860	Leg fat percentage L	389,120	beta	0.0432	0.0069	4.66E-07	4.84E-05
dis_out	HC932	TTE essential (primary hypertension)	112,342	or	1.08	0.16	5.11E-07	5.31E-05
antro	INI23127	Trunk fat percentage	442,582	beta	0.0352	0.0059	1.86E-06	0.0001932
biom	INI10030600	Albumin	389,120	beta	0.0414	0.0069	2.34E-06	0.0002436
antro	INI23128	Trunk fat	442,558	beta	0.0383	0.0064	2.62E-06	0.0002729
antro	INI23123	BMI	442,686	beta	0.0287	0.0049	3.62E-06	0.000376
antro	INI23104	Arm fat percentage L	442,840	beta	0.0378	0.0064	3.66E-06	0.0003802
antro	INI23124	Sodium in urine	442,647	beta	0.0367	0.0063	5.30E-06	0.0005508
biom	INI10030530	Arm fat mass L	412,901	beta	−0.0393	0.0067	5.46E-06	0.0005682
antro	INI23120	Arm fat mass R	442,716	beta	0.0363	0.0063	7.08E-06	0.0007364
antro	INI23119	BMI	442,746	beta	0.028	0.0049	9.14E-06	0.0009506
antro	INI21001	Arm fat percentage R	448,906	beta	0.0365	0.0064	9.33E-06	0.0009705

Note: Full list of the phenotypic traits associated to *ADH1B*2* retrieved from the Global Biobank (McInnes et al., 2019) for FDR ≤0.005, divided in class by Biometrics(biom), Disease outcome (dis_out) and Antropometrics (antro). The column "info" reports whether the effect value is an odds ratio (or) or its Beta (beta).

compensation for their retinol deficiency (Tang et al., 2009) and disrupted lipid metabolism, both of which were consequences of a rice-based diet.

In summary, our study provides an initial understanding of the natural selection of *ADH1B* and *ALDH2* genes in Asian populations, with a focus on dietary changes following the Neolithic transition within the context of the behavioral immune system theory. Our analysis expands upon the "aldehydes hypothesis" proposed by Darwin and Stanley (2022), originally centered on Mycobacterium tuberculosis, by advocating for the inclusion of Mycobacterium leprae as a target for future research into the beneficial effects of aldehydes.

TABLE 3 Phenotypic traits associated to *ALDH2*2*.

Class	Code	Description	*N*	Info	Effect	SE	*p*	FDR
dis_out	HC156	Nasal polyps	6546	or	7.10785	0.456	1.70E-05	0.000968
antro	INI48	Waist circumference	449,704	beta	−0.137	0.0319	1.78E-05	0.001013
antro	INI23127	Trunk fat percentage	442,582	beta	−0.144	0.0349	3.75E-05	0.002138
antro	INI23099	Body fat percentage	442,599	beta	−0.1183	0.0288	3.97E-05	0.002265
antro	INI23124	Arm fat mass L	442,647	beta	−0.1478	0.0361	4.32E-05	0.002462
antro	INI23104	BMI	442,840	beta	−0.1442	0.0361	6.53E-05	0.003723

Note: Full list of the phenotypic traits associated to *ALDH2*2* retrieved from the Global Biobank (McInnes et al., 2019) for FDR ≤0.005, divided in class by Biometrics (biom), Disease outcome (dis_out) and Antropometrics (antro). The column "info" reports whether the effect value is an odds ratio (or) or its Beta (beta).

Given the speculative nature of our research and the limitations posed by the available data, several important considerations must be addressed. First, national frequencies of *ADH1B**2 and *ALDH2**2 were calculated using ALFRED (Rajeevan et al., 2012), which provides the geographical origin of each sample along with self-reported ethnicity. However, this dataset lacks more detailed information on individual ancestry, which could impact the accuracy of our analyses. Natural selection is a complex and dynamic process, yet the scarcity of time-framed data on *ADH1B**2 and *ALDH2**2 (Gao et al., 2020), as well as ecological variables, necessitated a country-based approach. We relied on data from Fedderke et al. (2017), which were the most reliable sources available to us. It is important to note that while there is some convergent evidence in Ethiopian populations linking the adoption of agriculture, higher levels of aldehydes (McQuillan et al., 2022), and retinol deficiency to mycobacteria (Keflie et al., 2018), research on *ADH1B**2 and *ALDH2**2 remains biased toward Asian populations and their drinking patterns. However, there is evidence for other loci associated with fatty acid metabolism, cholesterol/triglyceride biosynthesis, and retinoic acid production being affected by the emergence and expansion of rice agriculture (Landini et al., 2021; Wang et al., 2016). Notably, among these, the expression of *ALDH1B1* has also been associated with immune response to pathogens, modulating the production of interferon beta in mice (Sun et al., 2024). These factors underscore the complexity and interconnectedness of genetic and environmental influences on human health and disease susceptibility.

Indeed, while further research is imperative, our findings, which integrate diverse datasets, provide momentum for the exploration of complex phenotypes, such as alcohol drinking patterns, within a broader framework that encompasses considerations of nutrition and immune response. This holistic approach acknowledges the interconnectedness of genetic, dietary, and immunological factors in shaping human health and disease susceptibility. By embracing this multidimensional perspective, future investigations can yield deeper insights into the intricate interplay between genetics, behavior, and environmental influences on human biology.

FUNDING INFORMATION

This research was funded by NIH/NIAAA grant AA017447 (to RC), grant AA014351 (to RC) and PRIN 2017SXEXT5 (to RC).

CONFLICT OF INTEREST STATEMENT

The authors declare no conflict of interest.

DATA AVAILABILITY STATEMENT

Data sharing is not applicable to this article as no new data were created or analyzed in this study.

ORCID

Giovanni Deiana https://orcid.org/0009-0000-6148-3972
Valerio Napolioni https://orcid.org/0000-0002-4378-6838

REFERENCES

Baan, R., Straif, K., Grosse, Y., Secretan, B., el Ghissassi, F., Bouvard, V. et al. (2007) Carcinogenicity of alcoholic beverages. *The Lancet Oncology*, 8(4), 292–293. Available from: https://doi.org/10.1016/S1470-2045(07)70099-2

Benjak, A., Avanzi, C., Singh, P., Loiseau, C., Girma, S., Busso, P. et al. (2018) Phylogenomics and antimicrobial resistance of the leprosy bacillus mycobacterium leprae. *Nature Communications*, 9(1), 352. Available from: https://doi.org/10.1038/s41467-017-02576-z

Bray, F. (1994) *The Rice economies: technology and development in Asian societies*. Berkeley, CA: University of California Press.

Chang, T., Yen, T., Wei, C., Hsiao, T.H. & Chen, I.C. (2023) Impacts of *ADH1B* rs1229984 and *ALDH2* rs671 polymorphisms on risks of alcohol-related disorder and cancer. *Cancer Medicine*, 12(1), 747–759. Available from: https://doi.org/10.1002/cam4.4920

Chen, C.-H., Ferreira, J.C.B., Gross, E.R. & Mochly-Rosen, D. (2014) Targeting aldehyde dehydrogenase 2: new therapeutic opportunities. *Physiological Reviews*, 94(1), 1–34. Available from: https://doi.org/10.1152/physrev.00017.2013

Cobo, J.M., Fort, J. & Isern, N. (2019) The spread of domesticated rice in eastern and southeastern Asia was mainly demic. *Journal of Archaeological Science*, 101, 123–130. Available from: https://doi.org/10.1016/j.jas.2018.12.001

Crabb, D.W., Edenberg, H.J., Bosron, W.F. & Li, T.K. (1989) Genotypes for aldehyde dehydrogenase deficiency and alcohol sensitivity. The inactive ALDH2(2) allele is dominant. *Journal of Clinical Investigation*, 83(1), 314–316. Available from: https://doi.org/10.1172/JCI113875

Darwin, K.H. & Stanley, S.A. (2022) The aldehyde hypothesis: metabolic intermediates as antimicrobial effectors. *Open Biology*, 12(4), 220010. Available from: https://doi.org/10.1098/rsob.220010

Fairley, J.K., Ferreira, J.A., Fraga, L.A.O., Lyon, S., Valadão Cardoso, T.M., Campos Boson, V. et al. (2023) High-resolution plasma metabolomics identifies alterations in fatty acid, energy, and micronutrient

metabolism in adults across the leprosy spectrum. *The Journal of Infectious Diseases*, jiad410. Available from: https://doi.org/10.1093/infdis/jiad410

Fedderke, J.W., Klitgaard, R.E. & Napolioni, V. (2017) Genetic adaptation to historical pathogen burdens. *Infection, Genetics and Evolution*, 54, 299–307. Available from: https://doi.org/10.1016/j.meegid.2017.07.017

Friedman, J., Hastie, T., Tibshirani, R., Narasimhan, B., Tay, K., Simon, N. et al. (2023) *glmnet: Lasso and elastic-net regularized generalized linear models*. Available from: https://cran.r-project.org/web/packages/glmnet/ [Accessed 21st September 2023].

Gao, Y., Zhang, C., Yuan, L., Ling, Y.C., Wang, X., Liu, C. et al. (2020) PGG.Han: the Han Chinese genome database and analysis platform. *Nucleic Acids Research*, 48(D1), D971–D976. Available from: https://doi.org/10.1093/nar/gkz829

Gaziano, J.M., Concato, J., Brophy, M., Fiore, L., Pyarajan, S., Breeling, J. et al. (2016) Million veteran program: a mega-biobank to study genetic influences on health and disease. *Journal of Clinical Epidemiology*, 70, 214–223. Available from: https://doi.org/10.1016/j.jclinepi.2015.09.016

Gilchrist, J.J., Auckland, K., Parks, T., Mentzer, A.J., Goldblatt, L., Naranbhai, V. et al. (2022) Genome-wide association study of leprosy in Malawi and Mali. *PLoS Pathogens*, 18(9), e1010312. Available from: https://doi.org/10.1371/journal.ppat.1010312

Hashimoto, M., Watanabe, M., Uematsu, Y., Hattori, S., Miyai, N., Utsumi, M. et al. (2016) Relationships of alcohol dehydrogenase 1B (ADH1B) and aldehyde dehydrogenase 2 (ALDH2) genotypes with alcohol sensitivity, drinking behavior and problem drinking in Japanese older men. *Environmental Health and Preventive Medicine*, 21(3), 138–148. Available from: https://doi.org/10.1007/s12199-016-0507-5

Huang, Z., Liu, Y., Qi, G., Brand, D. & Zheng, S. (2018) Role of vitamin a in the immune system. *Journal of Clinical Medicine*, 7(9), 258. Available from: https://doi.org/10.3390/jcm7090258

Hutton, H.E., McCaul, M.E., Santora, P.B. & Erbelding, E.J. (2008) The relationship between recent alcohol use and sexual behaviors: gender differences among sexually transmitted disease clinic patients. *Alcoholism: Clinical and Experimental Research*, 32, 2008–2015. Available from: https://doi.org/10.1111/j.1530-0277.2008.00788.x

Keflie, T.S., Samuel, A., Woldegiorgis, A.Z., Mihret, A., Abebe, M. & Biesalski, H.K. (2018) Vitamin a and zinc deficiencies among tuberculosis patients in Ethiopia. *Journal of Clinical Tuberculosis and Other Mycobacterial Diseases*, 12, 27–33. Available from: https://doi.org/10.1016/j.jctube.2018.05.002

Kim, E.W., De Leon, A., Jiang, Z., Radu, R.A., Martineau, A.R., Chan, E.D. et al. (2019) Vitamin a metabolism by dendritic cells triggers an antimicrobial response against mycobacterium tuberculosis. *mSphere*, 4(3), e00327-19. Available from: https://doi.org/10.1128/mSphere.00327-19

Kim, H. & Shin, S.J. (2023) Revolutionizing control strategies against mycobacterium tuberculosis infection through selected targeting of lipid metabolism. *Cellular and Molecular Life Sciences*, 80(10), 291. Available from: https://doi.org/10.1007/s00018-023-04914-5

Landini, A., Yu, S., Gnecchi-Ruscone, G.A., Abondio, P., Ojeda-Granados, C., Sarno, S. et al. (2021) Genomic adaptations to cereal-based diets contribute to mitigate metabolic risk in some human populations of east Asian ancestry. *Evolutionary Applications*, 14(2), 297–313. Available from: https://doi.org/10.1111/eva.13090

Li, H., Gu, S., Han, Y., Xu, Z., Pakstis, A.J., Jin, L. et al. (2011) Diversification of the ADH1B gene during expansion of modern humans: ADH1B gene diversification. *Annals of Human Genetics*, 75(4), 497–507. Available from: https://doi.org/10.1111/j.1469-1809.2011.00651.x

Lim, S.M., Goh, Y.M., Mohtarrudin, N. & Loh, S.P. (2016) Germinated brown rice ameliorates obesity in high-fat diet induced obese rats. *BMC Complementary and Alternative Medicine*, 16(1), 140. Available from: https://doi.org/10.1186/s12906-016-1116-y

Liu, X., Lu, D., Saw, W.-Y., Shaw, P.J., Wangkumhang, P., Ngamphiw, C. et al. (2017) Characterising private and shared signatures of positive selection in 37 Asian populations. *European Journal of Human Genetics*, 25(4), 499–508. Available from: https://doi.org/10.1038/ejhg.2016.181

Lui, P.P. & Zamboanga, B.L. (2019) Ethnocultural diversity in alcohol use and misuse. *American Journal of Orthopsychiatry*, 89(5), 543–548. Available from: https://doi.org/10.1037/ort0000422

Luo, H.-R., Wu, G.-S., Pakstis, A.J., Tong, L., Oota, H., Kidd, K.K. et al. (2009) Origin and dispersal of atypical aldehyde dehydrogenase ALDH2*487Lys. *Gene*, 435(1–2), 96–103. Available from: https://doi.org/10.1016/j.gene.2008.12.021

Luo, T., Comas, I., Luo, D., Lu, B., Wu, J., Wei, L. et al. (2015) Southern East Asian origin and coexpansion of *Mycobacterium tuberculosis* Beijing family with Han Chinese. *Proceedings of the National Academy of Sciences of the United States of America*, 112(26), 8136–8141. Available from: https://doi.org/10.1073/pnas.1424063112

Luo, Z., Cheng, J. & Wang, Y. (2023) Effects of the genetic variants of alcohol-metabolizing enzymes on lipid levels in Asian populations: a systematic review and meta-analysis. *Nutrition Reviews*, 81(8), 921–938.

McInnes, G., Tanigawa, Y., DeBoever, C., Lavertu, A., Olivieri, J.E., Aguirre, M. et al. (2019) Global biobank engine: enabling genotype-phenotype browsing for biobank summary statistics. *Bioinformatics*, 35(14), 2495–2497. Available from: https://doi.org/10.1093/bioinformatics/bty999

McQuillan, M.A., Ranciaro, A., Hansen, M.E.B., Fan, S., Beggs, W., Belay, G. et al. (2022) Signatures of convergent evolution and natural selection at the alcohol dehydrogenase gene region are correlated with agriculture in ethnically diverse Africans. *Molecular Biology and Evolution*, 39(10), msac183. Available from: https://doi.org/10.1093/molbev/msac183

Muthayya, S., Sugimoto, J.D., Montgomery, S. & Maberly, G.F. (2014) An overview of global rice production, supply, trade, and consumption: global rice production, consumption, and trade. *Annals of the New York Academy of Sciences*, 1324(1), 7–14. Available from: https://doi.org/10.1111/nyas.12540

Nagai, A., Hirata, M., Kamatani, Y., Muto, K., Matsuda, K., Kiyohara, Y. et al. (2017) Overview of the BioBank Japan project: study design and profile. *Journal of Epidemiology*, 27(3), S2–S8. Available from: https://doi.org/10.1016/j.je.2016.12.005

Okazaki, K., Takamuku, H., Yonemoto, S., Itahashi, Y., Gakuhari, T., Yoneda, M. et al. (2019) A paleopathological approach to early human adaptation for wet-rice agriculture: the first case of Neolithic spinal tuberculosis at the Yangtze River Delta of China. *International Journal of Paleopathology*, 24, 236–244.

Park, S.K., Park, C.-S., Lee, H.-S., Park, K.S., Park, B.L., Cheong, H.S. et al. (2014) Functional polymorphism in aldehyde dehydrogenase-2 gene associated with risk of tuberculosis. *BMC Medical Genetics*, 15(1), 40. Available from: https://doi.org/10.1186/1471-2350-15-40

Polimanti, R. & Gelernter, J. (2018) *ADH1B*: from alcoholism, natural selection, and cancer to the human phenome. *American Journal of Medical Genetics Part B: Neuropsychiatric Genetics*, 177(2), 113–125. Available from: https://doi.org/10.1002/ajmg.b.32523

Rajeevan, H., Soundararajan, U., Kidd, J.R., Pakstis, A.J. & Kidd, K.K. (2012) ALFRED: an allele frequency resource for research and teaching. *Nucleic Acids Research*, 40(D1), D1010–D1015. Available from: https://doi.org/10.1093/nar/gkr924

Schaller, M. (2011) The behavioural immune system and the psychology of human sociality. *Philosophical Transactions of the Royal Society, B: Biological Sciences*, 366(1583), 3418–3426. Available from: https://doi.org/10.1098/rstb.2011.0029

Schaschl, H., Göllner, T. & Morris, D.L. (2022) Positive selection acts on regulatory genetic variants in populations of European ancestry

that affect ALDH2 gene expression. *Scientific Reports*, 12(1), 4563. Available from: https://doi.org/10.1038/s41598-022-08588-0

Serajuddin, U. & Maeda, H. (2018) *World development indicators*. Washington, DC: The World Bank.

South, A. (2011) Rworldmap: a new R package for mapping global data. *The R Journal*, 3(1), 35–43.

Stafford, L.D., Sekulla, A., Morrison, E., Fleischman, D.S. & Harvey, A.J. (2020) Alcohol and disgust: an intimate relationship. *Drug and Alcohol Dependence*, 206, 107780. Available from: https://doi.org/10.1016/j.drugalcdep.2019.107780

Sudlow, C., Gallacher, J., Allen, N., Beral, V., Burton, P., Danesh, J. et al. (2015) UK biobank: an open access resource for identifying the causes of a wide range of complex diseases of middle and old age. *PLoS Medicine*, 12(3), e1001779. Available from: https://doi.org/10.1371/journal.pmed.1001779

Sun, N., Cai, Q., Zhang, Y., Zhang, R.R., Jiang, J., Yang, H. et al. (2024) The aldehyde dehydrogenase ALDH1B1 exerts antiviral effects through the aggregation of the adaptor MAVS. *Science Signaling*, 17(818), eadf8016. Available from: https://doi.org/10.1126/scisignal.adf8016

Tang, G., Qin, J., Dolnikowski, G.G., Russell, R.M. & Grusak, M.A. (2009) Golden Rice is an effective source of vitamin a. *The American Journal of Clinical Nutrition*, 89(6), 1776–1783. Available from: https://doi.org/10.3945/ajcn.2008.27119

Tenório De Menezes, Y.K., Eto, C., De Oliveira, J., Larson, E.C., Mendes, D.A.G.B., Dias, G.B.M. et al. (2023) The endogenous retinoic acid receptor pathway is exploited by mycobacterium tuberculosis during infection, both in vitro and in vivo. *The Journal of Immunology*, 211(4), 601–611. Available from: https://doi.org/10.4049/jimmunol.2200555

The World Factbook. (2021) *The world factbook*. Available from: https://www.cia.gov/the-world-factbook/ [Accessed 4th July 2023].

Thomson, R., Yuki, M., Talhelm, T., Schug, J., Kito, M., Ayanian, A.H. et al. (2018) Relational mobility predicts social behaviors in 39 countries and is tied to historical farming and threat. *Proceedings of the National Academy of Sciences of the United States of America*, 115(29), 7521–7526. Available from: https://doi.org/10.1073/pnas.1713191115

Wang, L.X., Wen, S., Wang, C.C., Zhou, B. & Li, H. (2016) Molecular adaptation of alcohol metabolism to agriculture in East Asia. *Quaternary International*, 426, 187–194. Available from: https://doi.org/10.1016/j.quaint.2016.03.008

Wang, N., Wang, Z., Wang, C., Fu, X', Yu, G., Yue, Z. et al. (2018) Prediction of leprosy in the Chinese population based on a weighted genetic risk score. *PLoS Neglected Tropical Diseases*, 12(9), e0006789. Available from: https://doi.org/10.1371/journal.pntd.0006789

Wang, Z., Mi, Z., Wang, H., Sun, L., Yu, G., Fu, X. et al. (2018) Discovery of 4 exonic and 1 intergenic novel susceptibility loci for leprosy. *Clinical Genetics*, 94(2), 259–263. Available from: https://doi.org/10.1111/cge.13376

Wells, J.C.K. & Stock, J.T. (2020) Life history transitions at the origins of agriculture: a model for understanding how niche construction impacts human growth, demography and health. *Frontiers in Endocrinology*, 11, 325. Available from: https://doi.org/10.3389/fendo.2020.00325

Whitfield, J.B. (2002) Alcohol dehydrogenase and alcohol dependence: variation in genotype-associated risk between populations. *American Journal of Human Genetics*, 71(5), 1247–50.

Wilson, S. (2021) *miceRanger: Multiple imputation by chained equations with random forests*. Available from: https://cran.r-project.org/web/packages/miceRanger/ [Accessed 21st September 2023].

World Health Organization. (2018) *Global status report on alcohol and health 2018*. Geneva: World Health Organization. Available from: https://apps.who.int/iris/handle/10665/274603 [Accessed 3rd July 2023].

Yang, L., Chen, J.-H., Lv, J., Wu, Q., Xu, T., Zhang, H. et al. (2012) Rice protein improves adiposity, body weight and reduces lipids level in rats through modification of triglyceride metabolism. *Lipids in Health and Disease*, 11(1), 24. Available from: https://doi.org/10.1186/1476-511X-11-24

Zakhari, S. (2006) Overview: how is alcohol metabolized by the body? *Alcohol Research & Health: The Journal of the National Institute on Alcohol Abuse and Alcoholism*, 29(4), 245–254.

Zhang, X., Sun, A. & Ge, J. (2021) Origin and spread of the ALDH2 Glu504Lys allele. *Phenomics*, 1(5), 222–228. Available from: https://doi.org/10.1007/s43657-021-00017-y

Zhou, H., Kalayasiri, R., Sun, Y., Nuñez, Y.Z., Deng, H.W., Chen, X.D. et al. (2022) Genome-wide meta-analysis of alcohol use disorder in east Asians. *Neuropsychopharmacology*, 47(10), 1791–1797. Available from: https://doi.org/10.1038/s41386-022-01265-w

Zhu, C., Talhelm, T., Li, Y., Chen, G., Zhu, J. & Wang, J. (2021) Relationship between rice farming and polygenic scores potentially linked to agriculture in China. *Royal Society Open Science*, 8(8), 210382. Available from: https://doi.org/10.1098/rsos.210382

SUPPORTING INFORMATION

Additional supporting information can be found online in the Supporting Information section at the end of this article.

How to cite this article: Deiana, G., Sun, R., Huang, J., Napolioni, V. & Ciccocioppo, R. (2024) Contribution of infectious diseases to the selection of *ADH1B* and *ALDH2* gene variants in Asian populations. *Alcohol: Clinical and Experimental Research*, 00, 1–12. Available from: https://doi.org/10.1111/acer.15288

Gene Discovery and Biological Insights into Anxiety Disorders from a Multi-Ancestry Genome-wide Association Study of >1.2 Million Participants

Eleni Friligkou[1,2], Solveig Løkhammer[1,3,4], Brenda Cabrera-Mendoza[1,2], Jie Shen[1,5], Jun He[1,2], Giovanni Deiana[1,6], Mihaela Diana Zanoaga[1,7], Zeynep Asgel[1,8], Abigail Pilcher[1], Luciana Di Lascio[1,9], Ana Makharashvili[1,2], Dora Koller[1,10], Daniel S. Tylee[1], Gita A. Pathak[1,2], Renato Polimanti[1,2,11,12]

[1]Department of Psychiatry, Yale School of Medicine, New Haven, CT, USA

[2]VA CT Healthcare Center, West Haven, CT, USA

[3]Department of Clinical Science, University of Bergen, Bergen, Norway

[4]Dr. Einar Martens Research Group for Biological Psychiatry, Center for Medical Genetics and Molecular Medicine, Haukeland University Hospital, Bergen, Norway

[5]Department of Cardiology, Children's Hospital of Soochow University, Suzhou, China.

[6] Center for Neuroscience, Pharmacology Unit, School of Pharmacy, University of Camerino

[7] Department of Computational Biology, University of Lausanne, Lausanne, Switzerland

[8]Department of Child and Adolescent Psychiatry, NYU Langone Health, New York Metropolitan Area, USA

[9]IRCCS Istituto Clinico Humanitas, Rozzano, Milan, Italy; Humanitas University, Pieve Emanuele, Milan, Italy.

[10]Department of Genetics, Microbiology and Statistics, Faculty of Biology, University of Barcelona, Catalonia, Spain

[11]Wu Tsai Institute, Yale University, New Haven, CT, USA

[12]Department of Chronic Disease Epidemiology, Yale School of Public Health, New Haven, CT, USA

Correspondence: Renato Polimanti, PhD. Department of Psychiatry, Yale University School of Medicine. 60 Temple, Suite 7A, New Haven, CT 06511, USA. Phone: +1 (203) 737-6391. E-mail: renato.polimanti@yale.edu

ABSTRACT

We leveraged information from more than 1.2 million participants to investigate the genetics of anxiety disorders across five continental ancestral groups. Ancestry-specific and cross-ancestry genome-wide association studies identified 51 anxiety-associated loci, 39 of which are novel. Additionally, polygenic risk scores derived from individuals of European descent were associated with anxiety in African, Admixed-American, and East Asian groups. The heritability of anxiety was enriched for genes expressed in the limbic system, the cerebral cortex, the cerebellum, the metencephalon, the entorhinal cortex, and the brain stem. Transcriptome- and proteome-wide analyses highlighted 115 genes associated with anxiety through brain-specific and cross-tissue regulation. We also observed global and local genetic correlations with depression, schizophrenia, and bipolar disorder and putative causal relationships with several physical health conditions. Overall, this study expands the knowledge regarding the genetic risk and pathogenesis of anxiety disorders, highlighting the importance of investigating diverse populations and integrating multi-omics information.

MAIN

Anxiety disorders affect many people worldwide, with a lifetime prevalence of approximately 34%, impacting overall health and all-cause mortality[1,2]. Although, according to the DSM-5[3], this diagnostic category contains multiple anxiety disorders with distinct sets of diagnostic criteria, there is a significant phenotypic overlap among them; 48-68% of individuals with an anxiety diagnosis fulfill the criteria for at least a second one[4]. Likewise, twin and genome-wide association studies (GWAS) demonstrated a shared genetic basis for the whole spectrum of anxiety diagnoses[5]. Specifically, anxiety disorders have an estimated twin-based heritability of 20–60% across subtypes and are highly polygenic[6,7]. GWAS contributed to understanding the genetic architecture of anxiety disorders by identifying up to ten risk loci[8-10]. These were primarily conducted in individuals of European descent, with limited samples available from African American participants enrolled in the Million Veteran Program (MVP)[10]. While the lack of diversity is a known issue in human genetic research[11], genetic research on anxiety appears to progress slower and be less diverse compared to current studies of other internalizing disorders, such as major depressive disorder (MDD) [12] and posttraumatic stress disorder (PTSD)[13]. To fill this gap, we conducted a multi-ancestry GWAS meta-analysis combining newly generated cohorts with previously reported data, reaching a total sample size of 1,266,780 participants (97,383 anxiety cases; Table 1). We included individuals of European (EUR, N=1,096,458), African (AFR, N=118,071), Admixed-American (AMR, N=36,634), South Asian (SAS, N=10,534), and East Asian (EAS, N=5,083) ancestries available from the All of Us Research Program (AoU)[14], the Lundbeck Foundation Initiative for Integrative Psychiatric Research (iPSYCH)[15], the FinnGen Project [16], MVP[10], the Psychiatric Genomics Consortium (PGC)[17], and the UK Biobank (UKB)[18]. Studying these cohorts, we identified multiple loci associated with anxiety and gained insights into the role of polygenic risk, pleiotropy, tissue-specific regulation, and transcriptomic and proteomic variations in the pathogenesis of anxiety disorders.

RESULTS

Genetic Discovery for Common Anxiety Factor, SNP-Heritability, and Genetic Correlation.

We combined GWAS of anxiety from six cohorts consisting of five ancestries (Table 1). In the EUR cohort, statistically significant single nucleotide polymorphism (SNP)-based heritability was observed in all cohorts, with Z-scores ranging from 2.64 (PGC) to 15.13 (FinnGen), and values proportional to the effective sample size of each cohort. The genetic

correlation (rg) ranged from 0.46 ($p=2.1\times10^{-3}$), between PGC and iPSYCH, to 1.02 ($p=2.0\times10^{-5}$), between PGC and AoU, with a median estimate of 0.72 (Figure 1; Supplemental Table 1). To quantify the genetically inferred differences across anxiety phenotypes across EUR datasets (Supplemental Figure 1), we applied PheMED (Phenotypic Measurement of Effective Dilution)[14]. FinnGen was considered the reference sample because it had the highest SNP-based heritability z score among the datasets investigated. MVP exhibited the largest phenotypic dilution ($\varphi=2.92$, $p=5\times10^{-12}$), followed by AoU ($\varphi=2.35$, $p=6.0\times10^{-13}$), PGC ($\varphi=1.81$, $p=2.0\text{-}\times10^{-13}$) and UKB ($\varphi=1.73$, $p<1x10^{-300}$), while the iPSYCH anxiety phenotype was not significantly diluted when compared to FinnGen ($\varphi=1.09$, $p=0.09$). Leveraging data from cohorts assessed with different approaches, we observed that FinnGen and iPSYCH show more consistent genetic effects than the other samples investigated, with MVP showing the highest phenotypic dilution factor. We did not observe statistically significant SNP-based heritability among non-EUR groups due to their limited sample size (Supplemental Table 2). Therefore, we could not explore the genetic correlation and phenotypic dilution among anxiety datasets in other populations.

We combined the genome-wide information of the 1,096,458 EUR individuals with genomic structural equation modeling (gSEM)[15] to account for the variability in the genetic correlation and the phenotypic heterogeneity among cohorts. The common anxiety factor (ANX; $\chi^2(9)=13.4$, p=0.15, comparative fit index=1, standardized root mean square residual=0.08) loaded significantly on all indicators (standardized loadings on AoU=0.85±0.05, FinnGen=0.82±0.03, MVP=0.80±0.04, iPSYCH=0.80±0.05, PGC=0.80±0.8, UKB=0.95±0.05; Figure 2). ANX showed a SNP-based heritability of 0.05±0.002 z-score=25.1) with no evidence of systematic bias due to population stratification or other confounders (linkage disequilibrium score regression[16] LDSC intercept=1.01±0.01, ratio=0.01±0.02). We identified 35 linkage disequilibrium (LD)-independent ($r^2<0.1$) variants with genome-wide significant association ($p<5\times10^{-8}$) with ANX (Figure 3; Supplemental Table 3). A conditional analysis[17] yielded five additional genome-wide significant (GWS) variants, leading to a total of 40 genome-wide significant independent SNPs (Supplemental Table 3). Of these, 29 were LD-independent (LD $r^2<0.1$) from variants reported by previous anxiety GWAS studies.[8-10] Among the novel associations, the most significant one was rs6689226 (beta=0.01, $p=4.8\times10^{-12}$), located in an intergenic region downstream of *LINC01360* gene on chromosome 1.

Due to the limited sample size of other population groups, we could not apply the gSEM approach. These ancestry-specific meta-analyses were conducted using a sample-

size weighted approach[18] to combine genetic effects derived from different assessments (Table 1). In the AFR GWAS meta-analysis of UKB, MVP, and AoU, we observed a GWS association for rs575403075 (beta=-0.11, $p=2.8\times10^{-8}$), located within a candidate cis-regulatory element with enhancer-like signature (ENCODE accession number: EH38E2597848) on chromosome 7[19]. This AFR-specific locus (non-AFR minor allele frequency<0.01) was previously associated with anxiety in the AFR GWAS in MVP.[10] No GWS associations were observed for the other populations. Because of the limited sample size of the other population groups, we observed cross-ancestry replication (p<0.01) only for two EUR GWS associations (Supplemental Table 3): rs12457101 (EUR beta=0.01, $p=2.1\times10^{-9}$; SAS beta=0.12, p=0.008) and rs10078721 (EUR beta=0.01, $p=4.9\times10^{-8;}$ AFR beta=0.02, p=0.009). EUR-derived polygenic risk scores (PRS) were significantly associated with anxiety phenotypes (fifth quintile vs. first quintile of PRS distribution) in AFR-AoU (OR=1.23, 95%CI=1.12-1.36), AMR-AoU (OR=1.55, 95%CI=1.39-1.74), EAS-AoU (OR=1.69, 95%CI=1.16-2.49), and EUR-AoU (OR=1.79, 95%CI=1.69-1.89) (Figure 4; Supplemental Table 4). The strength of the cross-ancestry PRS associations was proportional to the genetic distance between EUR and the other populations investigated[20]. Likely due to the limited sample size, null EUR-derived PRS association was observed with respect to SAS-AoU (Figure 4; Supplemental Table 4).

A cross-ancestry meta-analysis revealed 41 GWS LD-independent loci, of which ten were novel (Supplemental Table 5), considering previously reported variants[8-10] and index variants identified in the present ancestry-specific GWAS. The most significant among the novel cross-ancestry GWS associations was rs2510682 (cross-ancestry z=5.92, $p=3.3\times10^{-9}$; EUR beta=-0.12, $p=6.2\times10^{-8}$; AFR beta=-0.02, p=0.05; AMR beta=-0.04, p=0.18; EAS beta=-0.04, p=0.69; SAS beta=-0.04, p=0.44) located in an intronic region of *CNTN5* gene. For this locus and most of the other cross-ancestry associations, the signal was almost exclusively driven by the EUR sample. However, rs11681562 showed a cross-ancestry GWS association (z=-5.67, $p=1.4\times10^{-8}$) driven by both EUR and AFR samples (EUR beta=-0.01, $p=4.3\times10^{-7}$; AFR beta=-0.02, p=0.005; EAS beta=-0.05, p=0.64; SAS beta=-0.07, p=0.12).

Transcriptomic and Proteomic Analyses.

Partitioning the EUR-ANX heritability to 205 annotations based on tissue- and cell-specific transcriptomic profiles (Supplemental Table 6), we identified that ANX genetic liability was enriched for genes expressed in multiple brain regions, including the limbic system ($p=2.3\times10^{-5}$), the cerebral cortex ($p=4.2\times10^{-5}$), the cerebellum ($p=7.4\times10^{-5}$), the

metencephalon ($p=1.1\times10^{-4}$), the entorhinal cortex ($p=1.5\times10^{-4}$), and the brain stem ($p=2.0\times10^{-4}$).

We performed a transcriptome-wide association study (TWAS) of the EUR-ANX GWAS using S-PrediXcan[21]. Initially, we conducted a tissue-specific TWAS considering 13 brain tissues available from GTEx v8[22]. We observed 152 transcriptome-wide significant associations related to 39 genes accounting for the number of genes and brain tissues tested (N=165,710, $p<3.0\times10^{-7}$) with a Bonferroni correction (Supplemental Table 7). The strongest association was related to the genetically regulated transcriptomic variation of the *DRD2* locus in the cerebellar hemisphere ($z=-6.8$, $p=1.1\times10^{-11}$). Interestingly, no associations survived Bonferroni correction for this gene among the other brain tissues tested. Conversely, four genes showed Bonferroni significant associations with 10 or more brain tissues: *CD40* (from hippocampus $z=-6.0$, $p=1.5\times10^{-9}$ to nucleus accumbens $z=-5.1$, $p=2.7\times10^{-7}$), *GNL3* (from spinal cord $z=5.5$, $p=4.1\times10^{-8}$ to cerebellum $z=5.3$, $p=1.0\times10^{-7}$), *NEK4* (from cerebellar hemisphere $z=5.6$, $p=5.6\times10^{-8}$ to spinal cord $z=5.2$, $p=2.5\times10^{-7}$), and *SMIM4* (from anterior cingulate cortex $z=-5.9$, $p=3.4\times10^{-9}$ to spinal cord $z=-5.8$, $p=6.3\times10^{-9}$). To investigate cross-tissue transcriptomic regulation further, we combined information from 49 tissues available from GTEx V8[22] in a multi-tissue TWAS using S-MultiXcan[23]. After Bonferroni correction accounting for the number of genes tested (N=22,045; $p<2.3\times10^{-6}$), we identified 94 loci with genetically regulated transcriptomic associations with ANX (Figure 3, Supplemental Table 8). Fifty-nine of the 94 loci were not observed in the brain-specific TWAS, including the most significant transcriptome-wide association related to *MED19* (S-MultiXcan $p=5.5\times10^{-15}$). The second strongest multi-tissue association was *LINC01360* (S-MultiXcan $p=1.5\times10^{-11}$) near rs6689226, the most significant novel variant in the present EUR-ANX GWAS.

To understand the potential role of brain proteomic regulation in the pathogenesis of anxiety, we tested the EUR-ANX GWAS against protein quantitative trait loci (pQTL) in the dorsolateral prefrontal cortex (dlPFC) with FUSION[24]. We identified 24 loci with evidence of genetically regulated proteomic association with ANX after Bonferroni correction accounting for the number of genes tested (N=1,629; $p<3.1\times10^{-5}$) (Figure 3; Supplemental Table 9). We observed six genes identified by the proteome-wide association study (PWAS) to also be Bonferroni significant in the multi-tissue TWAS: *CTNND1* (PWAS $p=3.5\times10^{-16}$; multi-tissue TWAS $p=2.1\times10^{-8}$), *CNNM2* (PWAS $p=8.3\times10^{-14}$; multi-tissue TWAS $p=2.3\times10^{-7}$), *RAB27B* (PWAS $p=9.4\times10^{-14}$; multi-tissue TWAS $p=5.9\times10^{-8}$), *KHK* (PWAS $p=4.2\times10^{-9}$; multi-tissue TWAS $p=8.1\times10^{-7}$), NEK4 (PWAS $p=9.4\times10^{-8}$; multi-tissue TWAS $p=7.2\times10^{-8}$), and *BTN2A1* (PWAS $p=9.6\times10^{-9}$; multi-tissue TWAS $p=1.2\times10^{-7}$). Of those, *CTNND1* and *NEK4*

also demonstrated Bonferroni-significant expression in areas of the cortex, the basal ganglia, and the cerebellum in the brain TWAS. *CGREF1* was a PWAS significant hit (PWAS $p=3.3\times10^{-7}$), with a significant brain TWAS expression in the frontal cortex (TWAS $p=1.13\times10^{-7}$) and the putamen (TWAS $p=2.73\times10^{-7}$). The Bonferroni-significant PWAS loci were enriched for multiple synaptic locations and biological processes, including synapse (gene ontology, GO:0045202, $p=4.6\times10^{-4}$), synaptic vesicle membrane (GO:0030672, $p=5.9\times10^{-4}$), presynapse, synaptic vesicle (GO:0008021, p=0.001), process in the synapse (SYNGO:synprocess, p=0.003), and process in the presynapse (SYNGO:presynprocess, p=0.008).

Pleiotropy with Human Traits and Diseases.

We used MixeR[25] to investigate the polygenic architecture of anxiety in the context of other psychiatric disorders. A total of 12,622±834 influential variants were estimated for EUR-ANX. A similar estimate was observed for MDD (N=11,428±453, $p_{ANX\text{-}difference}=0.21$). Conversely, for anorexia nervosa (AN, N=7,869±387, $p_{ANX\text{-}difference}=1.9\times10^{-6}$), attention-deficit hyperactivity disorder (ADHD, N=7,978±387, $p_{ANX\text{-}difference}=4.4\times10^{-7}$), bipolar disorder (BIP, N=8,772±387, $p_{ANX\text{-}difference}=1.4\times10^{-5}$), PTSD (N=7,585±533, $p_{ANX\text{-}difference}=3.6\times10^{-7}$), and schizophrenia (SCZ, N=9,636±262, $p_{ANX\text{-}difference}=6.4\times10^{-4}$), the number of influential variants was significantly lower. No reliable estimate was obtained for autism spectrum disorder (ASD), obsessive-compulsive disorder (OCD), and Tourette syndrome (TS, Supplemental Table 10). Considering statistically meaningful bivariate MixeR models (Akaike information criterion for the best model versus model with minimal possible polygenic overlap>0; Supplemental Table 11), ANX shared 91±4%, 75±9%, and 68±1% of its influential variants with MDD, BIP, and SCZ, respectively (Figure 5A; Supplemental Figure 2).

To estimate the local genetic correlation of ANX with other psychiatric traits, we used Local Analysis of (co-)Variant Association (LAVA)[26]. After Bonferroni correction for the number of tests performed (N=12,368, $p<4.1\times10^{-6}$), we observed 47 local genetic correlations of ANX with five psychiatric disorders (i.e., ADHD, BIP, MDD, PTSD, and SCZ) across 35 genomic regions (Figure 5B; Supplemental Table 12). Consistent with the extensive genetic overlap observed in MixeR, we observed local genetic correlations between ANX and MDD in 39 genomic regions. Twelve regions showed evidence of local genetic correlation between ANX and SCZ. Only two regions were genetically correlated between ANX and ADHD, BIP, and PTSD, respectively. ANX demonstrated local genetic correlations with two or more psychiatric disorders in 10 regions. Among them, ANX showed

local genetic correlation with MDD (rho=0.97, $p=2.3\times10^{-6}$), SCZ (rho=0.84, $p=7.2\times10^{-12}$), and BIP (rho=0.73, $p=8.3\times10^{-8}$) within locus 1719 (chr11:112,755,447-113,889,019; Supplemental Figure 3. Within locus 2281 (chr18:52,512,524-53,762,996), ANX local genetic correlation was found with MDD (rho=0.88, $p=1.7\times10^{-12}$) and SCZ (rho=0.67, $p=1.9\times10^{-7}$). Among the most significant genomic regions, we also observed locus 852 (chr5:91,956,906-93,814,604), where ANX was genetically correlated with ADHD (rho=0.97, $p=4.7\times10^{-7}$) and MDD (rho=1, $p=3.4\times10^{-8}$).

To decompose the pleiotropic mechanisms linking ANX to multiple psychiatric disorders, we applied LAVA multiple regression models[26] to the five genomic regions with the strongest evidence of local genetic correlation with ANX (Supplemental Table 13). ANX was entered in the regression models as an outcome, and the other genetically correlated traits as predictors. We observed that the local genetic correlations of ANX with MDD, SCZ, and BIP in locus 1719 were not independent of each other. Shared pleiotropic mechanisms involving ANX were also observed in locus 62 (chr1:72,513,120-73,992,170) for MDD and SCZ and locus 852 (chr5:91,956,906-93,814,604) for ADHD and MDD. Conversely, for locus 1582 (chr10:106,142,284-107,877,787), we identified independent pleiotropic mechanisms of ANX with MDD (rho=0.91, $p=5.2\times10^{-9}$; gamma=0.63, p=0.04) and SCZ (rho=1, $p=2.8\times10^{-8}$; gamma=0.78, p=0.004). For locus 2281 (chr18:52,512,524-53,762,996), the ANX-MDD local genetic correlation (rho=0.88, $p=1.7\times10^{-12}$; gamma=0.78, p=0.003) was the primary pleiotropic driver that also accounted for ANX-SCZ relationship (rho=0.67, $p=1.9\times10^{-7}$; gamma=0.16, p=0.35).

To investigate the shared genetic mechanisms between ANX and human traits and diseases, we performed a phenome-wide genetic correlation analysis leveraging 11,175 phenotypes available from UKB, MVP, and FinnGen[27-29]. After Bonferroni correction ($p<4.47\times10^{-6}$), 1,929 showed a statistically significant genetic correlation with ANX (Figure 6, Supplemental Table 14). Among these, the median |rg| was 0.42 (25^{th} percentile=0.29; 75^{th} percentile=0.55), with 77 of the top 100 results (all |rg|>0.72) being mental health outcomes (e.g., FinnGen anxiety disorders rg=0.91, $p<1x10^{-300}$; UKB self-reported depression rg=0.83, $p=1.0\times10^{-53}$) and related drug prescriptions (e.g., UKB citalopram medication rg=0.93, $p=7.3\times10^{-8}$; FinnGen depression medications rg=0.82, $p<1x10^{-300}$). Among non-psychiatric outcomes, the top results included myalgia (FinnGen rg=0.85, $p=4.0\times10^{-12}$), nausea/vomiting (UKB rg=0.84, $p=1.5\times10^{-11}$), diseases of pulp and periapical tissues (FinnGen rg=0.84, $p=4.6\times10^{-7}$), dysuria (MVP rg=0.79, $p=1.7\times10^{-6}$), cystitis (FinnGen rg=0.77, $p=8.1\times10^{-9}$), constipation (UKB rg=0.75, $p=1.2\times10^{-20}$), pain related to

temporomandibular disorders (FinnGen rg=0.75, $p=1.4\times10^{-16}$), unspecified dorsalgia (FinnGen rg=0.73, $p=3.1\times10^{-45}$), and pain management (UKB rg=0.73, $p=6.1\times10^{-30}$).

To investigate potential causal relationships underlying those genetic correlations, we employed the latent causal variable approach (LCV)[30]. We identified 24 traits with statistically significant genetic causal proportion (gcp) after a false discovery rate (FDR) multiple testing correction (FDR $q<0.05$, $p<4.1\times10^{-6}$) (Supplemental Table 15). Positive causal effects were observed from other phenotypes to ANX in all cases (LCV rho>0). Only three of these phenotypes were related to mental health: “Number of things worried about during worst period of anxiety” (UKB gcp=0.89, $p=4.1\times10^{-63}$; rho=0.67±0.10), “Ever had period of mania/excitability” (UKB gcp=0.46, $p=1.4\times10^{-28}$; rho=0.51±0.08), and “Number of cigarettes currently smoked daily” (UKB gcp=0.54, $p=2.2\times10^{-7}$; rho=0.23±0.07). Seven of the phenotypes identified were related to physical health concerns, with the strongest relationship implicating Dupuytren's disease (MVP, gcp=0.71, $p=7.1\times10^{-93}$; rho=0.13±0.05). Five were related to the cardiovascular system: “Atherosclerotic cardiovascular disease” (MVP, gcp=0.60, $p=8.9\times10^{-29}$; rho=0.40±0.09), “Other disorders of circulatory system” (UKB, gcp=0.43, $p=5.2\times10^{-26}$; rho=0.46±0.07), “Medication for transient ischemic attack” (MVP, gcp=0.63, $p=6.4\times10^{-17}$; rho=0.24±0.0.8), “Other specified cardiac dysrhythmias” (UKB, gcp=0.93, $p=7.0\times10^{-15}$; rho=0.40±0.08), and “Heart Surgery” (UKB, gcp=0.96, $p=7.0\times10^{-15}$; rho=0.40±0.0.08). We also observed two environmental variables with FDR-significant effects: “Workplace had a lot of diesel exhaust” (UKB, gcp=0.95, $p=2.9\times10^{-24}$; rho=0.45±0.10) and “Number of gap periods in employment” (UKB, gcp=0.60, $p=5.4\times10^{-7}$; rho=0.26±0.08).

DISCUSSION

The present study analyzed genome-wide information of more than 1.2 million participants, giving valuable insights into the genetic liability of anxiety disorders. The ANX GWAS in EUR identified 40 GWS associations, 29 of which were novel. Unfortunately, the limited sample size of other ancestry groups permitted us to only confirm a previously reported locus in AFR[10]. However, the cross-ancestry GWAS added ten novel anxiety risk loci to our gene discovery. The novel findings of the EUR-specific and cross-ancestry analyses quadrupled the gene discoveries reported by previous studies[8-10]. Among the novel loci identified, the strongest ANX association was with rs6689226, located near *LINC01360*. While the regulatory mechanisms related to this long non-coding RNA (lncRNA) are unclear, this variant was previously associated with several mental health outcomes such as smoking initiation[31], externalizing behaviors[32], and cross-disorder psychopathology[33]. Our multi-

tissue transcriptomic analysis identified genetically regulated expression of *LINC01360* in the testis. LncRNAs are more expressed in the testis than other tissues, and some of them play a role in androgen production[34], which may be of pathophysiological relevance in certain animal models of anxiety[35]. In the cross-ancestry GWAS, the strongest ANX association was rs2510682, located in an intron region of the *CNTN5* gene, which encodes a protein mediating cell surface interactions during the development of the nervous system[36]. While it has not been associated with anxiety by previous GWAS[8-10], *CNTN5* variants were linked to neuroticism[37], suicidal behaviors[38] , and ASD[39]. The second most significant cross-ancestry association, rs11681562, was also a novel ANX locus. Variants in high LD with rs11681562 (r^2>0.8) have been previously associated with multiple phenotypes related to human cortical folding (i.e., vertex-wise sulcal depth, vertex-wise cortical thickness, vertex-wise cortical surface area, cortical thickness, and cortical surface area)[40]. Rs11681562 is an eQTL for multiple genes located in a chromosomal region with high gene density. In this region, our transcriptome-wide and proteome-wide analyses converge on *KHK* as potentially linked to anxiety pathogenesis. This gene is responsible for producing an enzyme involved in fructose metabolism[41]. Animal studies highlighted the potential role of early-life high fructose exposure in long-term depression- and anxiety-like behaviors[42]. *KHK* has also been identified as a PTSD risk locus in transcriptomic and proteomic analyses in the brain and blood[43]. However, we cannot exclude that other genes in this region may contribute to the association observed.

The transcriptome- and proteome-wide analyses converged on five additional loci: *CTNND1*, *CNNM2*, *RAB27B*, *BTN2A1*, and *NEK4*. In both approaches, the strongest statistical evidence was observed for the *CTNND1* gene. This encodes a protein regulating the dendritic spine and synapse development through rho-family GTPases and cadherins[44]. The overexpression of *CTNND1* has been linked to improved memory and reduced anxiety in mice[45]. The *CNNM2* gene, encoding a cyclin protein involved in Mg^{2+} transport, has been associated with impaired neuronal development and epilepsy[46]. *RAB27B* has been identified as a PTSD and depression risk locus in transcriptomic and proteomic analyses in the brain and blood[47]. Reduced *BTN2A1* expression in the placenta was linked to immune system processes in the context of maternal anxiety and depression[48]. *NEK4* has been previously associated with ADHD, BIP, migraines, and cross-disorder psychopathology in genetic and multi-omics analyses[49-52].

In line with the cross-disorder overlap observed in the single-locus discoveries described above, our genome-wide analyses indicated pleiotropic mechanisms between ANX and other psychiatric disorders. We observed a similar degree of polygenicity and a

high genetic correlation between ANX and MDD (rg=0.9), consistent with previous studies[53,54] and their shared symptomatology[55,56]. However, while 99% of MDD influential variants appear to be shared with ANX, our MixeR analysis suggests that about 10% of the ANX influential variants do not overlap with those of MDD. This may suggest, from a genetic perspective, that MDD is considered an ANX subtype.

LAVA identified several genomic regions with statistical evidence for the local genetic correlation of ANX with multiple psychiatric disorders. In line with the global pleiotropy analysis, ANX shared the most local genetic correlations with MDD. We also observed multiple regions where ANX showed local genetic correlation with two or more mental illnesses. Among them, locus 1719 (chr11:112,755,447-113,889,019) linked ANX to MDD, SCZ, and BIP. Based on data available from the GWAS catalog[57], over 400 GWS associations involving more than 100 phenotypes have been previously reported in this genomic region (Supplemental Table 16), with the majority related to mental health outcomes, such as anxiety, SCZ, BIP, MDD, PTSD, substance use, and educational attainment. Within this region, we observed the strongest brain-specific TWAS finding, involving the association of ANX with the genetically regulated transcriptomic variation of *DRD2* in the cerebellar hemispheres but not in other brain tissues. This locus is well-known in psychiatric genetics, and considerable research has characterized the function of the encoded dopamine D2 receptor (DR2) protein[58]. In mice, changes in cerebellar D2R levels in Purkinje cells alter sociability and preference for social novelty[59]. The present findings suggest a role for cerebellar D2R in the pathophysiology of human anxiety disorders.

Additionally, we observed a strong genetic overlap between ANX and health-related traits beyond psychiatric disorders. Among the non-psychiatric phenotypes in the 100 highest genetic correlations, we observed several gastrointestinal (43%) and pain-related (17%) phenotypes involving different body areas (e.g., muscle, temporomandibular, and thoracic). Anxiety is highly comorbid with gastrointestinal (GI) disorders. Irritable bowel syndrome has a common genetic basis with anxiety and mood disorders[60,61]. However, the genomic links between anxiety and the broader spectrum of GI disorders are still unclear. A growing body of literature supports the complex interplay between pain and anxiety disorders[62,63]. A recent brain-wide analysis integrating imaging and genetic data highlighted the potential role of pain sensitivity in the pathogenesis of anxiety disorders[64]. The results of our LCV analyses also identified genetic evidence of a causal impact of physical health phenotypes on anxiety disorders, particularly for multiple cardiovascular conditions. This is in line with the known comorbidity between these disorders[65] and the high anxiety prevalence after cardiac events[66].

Our common ANX factor demonstrated high loadings to all input datasets (standardized loadings≥0.8), suggesting the generalizability of the results across different cohorts. However, FinnGen and iPSYCH showed more consistent genetic effects than the other samples investigated, with MVP showing the highest degree of phenotypic dilution. One potential explanation is that MVP was the only cohort where anxiety was assessed quantitatively. Those results highlight that more homogeneous and detailed assessments may improve the statistical power and interpretability of genetic studies. Unfortunately, the limited sample size and statistical power in non-EUR populations precluded identifying novel loci in these ancestries. However, we demonstrated evidence that the polygenic risk for ANX detected in EUR is translatable to other population groups. Similar to prior observations[67], the strength of the cross-ancestry PRS association was proportional to the genetic distance between the population group of the training and target samples. To improve inclusion and equity in anxiety genetics, future studies should prioritize recruiting and assessing diverse cohorts.

In conclusion, this large-scale genome-wide and multi-omics investigation yielded novel insights into the biology of anxiety disorders. We significantly expanded the number of known anxiety risk loci, highlighting the importance of analyzing data from ancestrally diverse participants. Our multi-omics analyses identified novel pathways putatively involved in the pathogenesis of anxiety and refined plausible mechanisms related to previously reported loci. We also disentangled the pleiotropic mechanisms linking ANX to MDD and other psychiatric disorders, reinforcing the importance of studying anxiety disorders in the context of physical health.

METHODS

Study Populations.

The present study leveraged genome-wide data from six cohorts including AoU[68], FinnGen[69], iPSYCH[70], MVP[10], PGC[71], and UKB[28] (Table 1). This permitted us to investigate a total of 1,266,780 participants with diverse ancestral backgrounds. AoU is a research program aiming to create a representative cohort of the US population to accelerate biomedical research and improve healthcare through precision medicine[68]. Among the samples investigated in the present study, the AoU cohort was the most ancestrally diverse, with information on five genetically inferred population groups (i.e., AFR, AMR, EAS, EUR, and SAS). Details regarding AoU recruitment, assessment, and whole-genome sequencing

have been previously described[68], and the genomic quality control is described at https://support.researchallofus.org/hc/en-us/articles/4617899955092-All-of-Us-Genomic-Quality-Report. In the present study, we analyzed genetic and phenotypic information for 201,334 individuals. The definition of anxiety disorder cases and controls was derived from AoU electronic health records (EHR) using the SNOMED (Systematized Nomenclature of Medicine) classification system[72]. Cases were defined as AoU participants with a lifetime diagnosis of generalized anxiety disorder (SNOMED ID: 434613), panic disorder (SNOMED ID: 4138454), phobic disorder (SNOMED ID: 4304010), chronic stress disorder (SNOMED ID: 436074), acute stress disorder (SNOMED ID: 440083), posttraumatic stress disorder (SNOMED ID: 436676), obsessive-compulsive disorder (SNOMED ID: 440374), or mixed anxiety and depressive disorder (SNOMED ID: 4338031). Any participant without a documented EHR diagnosis falling into the above categories until the maximum follow-up age was considered a control. Ancestry-stratified GWAS was performed using the following quality control criteria: biallelic variants with minor allele frequency >1%, Hardy-Weinberg equilibrium $p<10^{-6}$, call rate >95%, and per-individual genotyping rate >95%. A PLINK 2[73] logistic regression model estimated genetic associations, including covariates for sex, age, and the first ten within-ancestry principal components.

FinnGen is a project focused on developing a large Finnish cohort combining genotype data from biobanks and digital health record data from health registries[27]. In the present study, we used genome-wide association statistics generated from Release 9 (May 11, 2023). A detailed description of FinnGen data is available at https://finngen.gitbook.io/documentation/v/r9/. The FinnGen GWAS was performed using REGENIE[74] including covariates for age, sex, the top-ten within-ancestry PCs, and genotyping batch. In our analysis, we used GWAS data generated from 40,191 cases and 277,526 controls of EUR descent for the phenotype "anxiety disorders" (KRA_PSY_ANXIETY), defined based on International Classification of Diseases (ICD-10 F40-F48; ICD-9 300.0-300.3, 300.6-300.9,3078A, 309; ICD-8 300.0-300.2, 30030, 300.5-300.9, 305, 30680, 30799). Details regarding the KRA_PSY_ANXIETY definition are available at https://risteys.finregistry.fi/endpoints/KRA_PSY_ANXIETY.

iPSYCH is a large Danish population-based cohort aimed at unraveling the genetic and environmental architecture of severe mental disorders[75]. In the present study, we used genome-wide association statistics generated from the analysis of 12,655 cases and 19,225 controls of EUR descent[70]. Cases were defined as individuals with an anxiety and stress-related diagnosis according to the ICD-10 F40.0-F41.9 and F43.0-F43.9. codes. Controls included individuals without any ICD-10 diagnoses of anxiety, stress-related disorders, or

mood disorders. Genetic associations were estimated using logistic regression models with the imputed marker dosages using the first four principal components as ancestry covariates. The analysis was stratified by genotyping batch, and the results were meta-analyzed using an inverse variance–weighted fixed-effect models.

MVP is a biobank funded by the US Department of Veterans Affairs to understand how genes, lifestyle, military experiences, and exposures affect health and wellness[76]. The GWA statistics used in the present study were generated from an analysis of 61,796 and 241,541 participants of AFR and EUR descent, respectively[10]. The anxiety phenotype was the total score of the generalized anxiety disorder 2-item scale, which ranged from 0 to 6. The genetic association analysis was performed using a PLINK 2[73] logistic regression model with covariates for age, sex, and the first ten within-ancestry principal components.

PGC genome-wide association statistics were derived from a meta-analysis conducted by the Anxiety NeuroGenetics STudy (ANGST)[71]. Cases were defined as individuals with diagnoses of generalized anxiety disorder, panic disorder, social phobia, agoraphobia, or specific phobias[71]. A total of 5,761 cases and 11,765 controls of EUR descent were included. The GWAS was performed using a logistic regression model with covariates for sex, age, and the first ten within-ancestry principal components.

UKB is a large population-based study that collected information regarding over 500,000 participants[28]. GWA statistics used in the present study were obtained from the Pan-UKB analysis. Details regarding the quality control, ancestry assignment, and statistical analyses are available at https://pan.ukbb.broadinstitute.org/. Briefly, ancestry-stratified GWASs were performed using the Scalable and Accurate Implementation of Generalized (SAIGE)[77] mixed models, including a kinship matrix as a random effect and the remaining covariates (i.e., top-10 within-ancestry PC, sex, age, age^2, sex$\times$age, and sex$\times$age^2) as fixed effects. We analyzed UKB a total of 10,751 cases and 383,235 controls of AFR, EUR, and SAS descents for the phecode 300 “Anxiety, dissociative and somatoform disorders”.

Phenotype Definition Comparisons.

Because of the different anxiety definitions, we compared their genetic architecture using two methods. LD Score regression[78] was used to estimate their SNP-based heritability and their pairwise genetic correlations. LD scores were calculated using HapMap 3 variants and the 1000 Genomes Project reference populations corresponding to each ancestry group[79]. Phenotype dilution among anxiety definitions was quantified using PheMED[14],

considering the FinnGen cohort as the reference sample because it showed the highest SNP-based heritability z-score. Because of the lack of significant SNP-based heritability in non-EUR samples, genetic correlation, and phenotype dilution analyses were performed only in EUR datasets.

Ancestry-Specific and Cross-Ancestry Meta-Analyses and Conditional Analysis.

Due to moderate genetic correlation (median rg=0.72) and statistically significant phenotype dilution among anxiety phenotypes, we combined the EUR GWASs using the gSEM approach[15]. As recommended[80], we used the effective sample size for each EUR anxiety GWAS. For binary definitions, this was calculated as $4/({}^{1}/_{N_{cases}} + {}^{1}/_{N_{controls}})$. For quantitative phenotypes, we used the total sample size. The effective sample size of the common ANX factor obtained from the gSEM analysis was estimated as $mean(1/(2 \times MAF \times (1 - MAF) \times SE^2)$, where MAF is the minor allele frequency and SE is the standard error. Due to the limited sample sizes, we could not use gSEM to combine genome-wide information available for AFR (AoU, UKB, and MVP) and SAS (AoU and UKB). Thus, we performed ancestry-specific meta-analyses using the sample-size weighted approach available in METAL[18]. To present ancestry-specific effects with the same statistics, AFR and SAS z-scores obtained from the sample-size weighted approach were converted to betas using the following formula: $beta = z/sqrt(2 \times p \times (1 - p) \times (n + z^2)$, where p is the effect allele frequency and n is the sample size. The sample-size weighted approach, using effective sample sizes, was also used to perform the cross-ancestry meta-analysis. Independent associations were identified by clumping GWS associations considering LD r^2<0.1. To identify secondary associations accounting for the primary GWS signals, we applied the stepwise model selection procedure available in the COJO approach implemented in the GCTA package[17,81]. Among the GWS associations identified by the conditional analysis, we considered only those with LD r^2<0.8 with respect to the corresponding index GWS variant.

Cross-Ancestry Polygenic Risk Scoring.

To investigate how anxiety polygenic risk translates across ancestry groups, we derived PRS from the EUR ANX factor and tested them in other populations available from the AoU cohort (i.e., AFR, AMR, EAS, and SAS). To compare within-ancestry to cross-ancestry PRS associations, we derived a PRS from the EUR ANX factor excluding AoU-EUR and used the latter as the EUR target sample. Posterior variant-level effect sizes of the

EUR ANX factor were computed using PRS-CS[82] and the 1000 Genomes European population as LD reference. ANX PRS were computed using PLINK[73]. We tested the quintiles of PRS distribution using the first quintile as reference with logistic regression models available in the *stats* R package in AoU population groups.

Heritability Analyses.

LDSC[78] was used to estimate the SNP-based heritability of anxiety leveraging ancestry-specific genome-wide association statistics using the methods described above. For the EUR ANX factor, we also performed a partitioned heritability analysis[83], testing the enrichment of 205 tissue- and cell-type specific transcriptomic profiles. These annotations were derived from GTEx V8[22] (53 tissues) and the Franke lab[84,85] (152 tissues/cell types).

Transcriptome- and Proteome-Wide Association Studies.

To understand the role of genetically regulated transcriptomic and proteomic variation in anxiety pathogenesis, we performed transcriptome-wide and proteome-wide association analyses. A brain-specific transcriptome-wide investigation was conducted using S-PrediXcan[21] to integrate EUR-ANX genome-wide association statistics with GTEx eQTL data available for 13 brain tissues. A Bonferroni correction accounting for the number of tests performed (N=165,710, $p<3\times10^{-7}$) was applied to define the statistically significant associations. To explore cross-tissue transcriptomic regulation, we performed a multi-tissue transcriptome-wide association analysis using S-MultiXcan[23]. This permitted us to boost the statistical power of the association analysis via a joint multi-tissue analysis accounting for transcriptomic correlation across the 49 GTEx tissues tested. Multi-tissue genetically regulated transcriptomic associations were defined after Bonferroni correction accounting for the number of the genes tested (N=22,045; $p<2.27\times10^{-6}$).

Proteome-wide associations were estimated using the FUSION approach[24] to integrate EUR-ANX GWAS with dlPFC pQTLs available from Religious Orders Study and Memory and Aging Project (ROSMAP)[86] and the Banner Sun Health Research Institute (BSHRI)[87]. Analyses were performed separately in two pQTL datasets and the results were meta-analyzed using an inverse-variance weighted approach available in METAL[18]. Bonferroni correction accounting for the number of genes tested (N=1,629; $p<3.07\times10^{-5}$) was applied to define proteome-wide significant genes. To understand their implications on synaptic localization and function, these loci were analyzed using SynGO[88], an online knowledge base that organizes research on synaptic proteins using GO annotations. SynGO

enrichments were estimated for each subset of cellular components and biological processes by a one-sided Fisher exact test.

Pleiotropy Analyses.

We used multiple methods and data sources to estimate the pleiotropy between EUR ANX and human traits and diseases. Initially, we focused on understanding the shared genetic mechanisms of ANX with other psychiatric disorders previously investigated using GWAS, including ADHD[89], AN[90], ASD[91], BIP[92], MDD[93], OCD[94], PTSD[95], SCZ, and TS[96]. After excluding the MHC region (chr6:26,000,000-34,000,000), we applied mixture models available from MixeR[25] to estimate the number of influential variants ANX shares with other psychiatric disorders. We then applied LAVA[26] to estimate local genetic correlations of ANX with mental illnesses. Because we observed local genetic correlation with multiple psychiatric disorders in the same genomic regions, we used LAVA multivariate models[26] to dissect whether the observed local genetic correlations were common or distinct among the implicated disorders. For MixeR and LAVA analyses, European populations available from 1000 Genomes Project Phase 3 were used as LD reference.

To assess ANX pleiotropic mechanisms across the human phenotypic spectrum, we performed an LDSC genetic correlation analysis[78] for 11,175 phenotypes available from UKB (details available at https://pan.ukbb.broadinstitute.org/), FinnGen (Release 9, details available at https://finngen.gitbook.io/documentation/v/r9/), and MVP[29]. Bonferroni correction accounting for the number of tests performed was applied to define statistically significant genetic correlations ($p<4.47\times10^{-6}$). To investigate the potential causality underlying the observed genetic correlations, we conducted a phenome-wide LCV[30] analysis using FinnGen, UKB, and MVP data. This was performed using LD scores calculated from European populations available from the 1000 Genomes Project Phase 3. LCV permitted us to calculate gcp estimates. Positive and negative gcp values reflect the direction of the putative causal effect (i.e., phenotype #1 → phenotype #2 and phenotype #2 → phenotype #1, respectively). In the present study, positive gcp values correspond to the effect of a phenotype on ANX. Information regarding the sign of the LCV effect is determined by the LCV rho statistics: rho>0 corresponds to a positive effect, while rho<0 corresponds to a negative effect.

ACKNOWLEDGMENTS

This study was supported by grants from the National Institutes of Health (RF1 MH132337, R33 DA047527, and K99 AG078503), One Mind, the Alzheimer's Association (Research Fellowship AARF-22-967171), the American Foundation for Suicide Prevention (PDF-1-022-21), Horizon 2020 (Marie Sklodowska-Curie Individual Fellowship 101028810), University of Bergen (International Training Grant), and the Yale Franke Program in Science and Humanities. We also acknowledge the contribution of the participants and the investigators involved in the UK Biobank, the FinnGen Project, the Million Veteran Program, the All of Us Research Program, the iPSYCH study, and the Psychiatric Genomics Consortium. The All of Us Research Program is supported by the National Institutes of Health, Office of the Director: Regional Medical Centers: 1 OT2 OD026549; 1 OT2 OD026554; 1 OT2 OD026557; 1 OT2 OD026556; 1 OT2 OD026550; 1 OT2 OD 026552; 1 OT2 OD026553; 1 OT2 OD026548; 1 OT2 OD026551; 1 OT2 OD026555; IAA #: AOD 16037; Federally Qualified Health Centers: HHSN 263201600085U; Data and Research Center: 5 U2C OD023196; Biobank: 1 U24 OD023121; The Participant Center: U24 OD023176; Participant Technology Systems Center: 1 U24 OD023163; Communications and Engagement: 3 OT2 OD023205; 3 OT2 OD023206; and Community Partners: 1 OT2 OD025277; 3 OT2 OD025315; 1 OT2 OD025337; 1 OT2 OD025276.

COMPETING INTERESTS

Dr. Polimanti is paid for their editorial work on the journal Complex Psychiatry and reports a research grant from Alkermes. The other authors declare no competing interests.

REFERENCES

1. Szuhany, K.L. & Simon, N.M. Anxiety Disorders: A Review. *JAMA* **328**, 2431-2445 (2022).
2. Mutz, J., Choudhury, U., Zhao, J. & Dregan, A. Frailty in individuals with depression, bipolar disorder and anxiety disorders: longitudinal analyses of all-cause mortality. *BMC Med* **20**, 274 (2022).
3. Association, A.P. *Diagnostic and Statistical Manual of Mental Disorders*.
4. Penninx, B.W.J.H., Pine, D.S., Holmes, E.A. & Reif, A. Anxiety disorders. *The Lancet* **397**, 914-927 (2021).
5. Meier, S.M. & Deckert, J. Genetics of Anxiety Disorders. *Curr Psychiatry Rep* **21**, 16 (2019).
6. Hettema, J.M., Neale, M.C. & Kendler, K.S. A review and meta-analysis of the genetic epidemiology of anxiety disorders. *Am J Psychiatry* **158**, 1568-78 (2001).
7. Purves, K.L. *et al.* A major role for common genetic variation in anxiety disorders. *Mol Psychiatry* **25**, 3292-3303 (2020).
8. Hettema, J.M. *et al.* Genome-wide association study of shared liability to anxiety disorders in Army STARRS. *Am J Med Genet B Neuropsychiatr Genet* **183**, 197-207 (2020).
9. Li, W. *et al.* Genome-wide meta-analysis, functional genomics and integrative analyses implicate new risk genes and therapeutic targets for anxiety disorders. *Nat Hum Behav* (2023).
10. Levey, D.F. *et al.* Reproducible Genetic Risk Loci for Anxiety: Results From ~200,000 Participants in the Million Veteran Program. *Am J Psychiatry* **177**, 223-232 (2020).
11. Sirugo, G., Williams, S.M. & Tishkoff, S.A. The Missing Diversity in Human Genetic Studies. *Cell* **177**, 26-31 (2019).
12. Meng, X. *et al.* Multi-ancestry genome-wide association study of major depression aids locus discovery, fine mapping, gene prioritization and causal inference. *Nat Genet* (2024).
13. Nievergelt, C.M. *et al.* Discovery of 95 PTSD loci provides insight into genetic architecture and neurobiology of trauma and stress-related disorders. *medRxiv* (2023).
14. Burstein, D. *et al.* Detecting and Adjusting for Hidden Biases due to Phenotype Misclassification in Genome-Wide Association Studies. *medRxiv* (2023).
15. Grotzinger, A.D. *et al.* Genomic structural equation modelling provides insights into the multivariate genetic architecture of complex traits. *Nat Hum Behav* **3**, 513-525 (2019).
16. Bulik-Sullivan, B.K. *et al.* LD Score regression distinguishes confounding from polygenicity in genome-wide association studies. *Nature Genetics* **47**, 291-295 (2015).
17. Yang, J. *et al.* Conditional and joint multiple-SNP analysis of GWAS summary statistics identifies additional variants influencing complex traits. *Nat Genet* **44**, 369-75, S1-3 (2012).
18. Willer, C.J., Li, Y. & Abecasis, G.R. METAL: fast and efficient meta-analysis of genomewide association scans. *Bioinformatics* **26**, 2190-1 (2010).
19. Moore, J.E. *et al.* Expanded encyclopaedias of DNA elements in the human and mouse genomes. *Nature* **583**, 699-710 (2020).
20. Peterson, R.E. *et al.* Genome-wide Association Studies in Ancestrally Diverse Populations: Opportunities, Methods, Pitfalls, and Recommendations. *Cell* **179**, 589-603 (2019).
21. 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).

22. Consortium, G. The GTEx Consortium atlas of genetic regulatory effects across human tissues. *Science* **369**, 1318-1330 (2020).
23. Barbeira, A.N. *et al.* Integrating predicted transcriptome from multiple tissues improves association detection. *PLoS Genet* **15**, e1007889 (2019).
24. Gusev, A. *et al.* Integrative approaches for large-scale transcriptome-wide association studies. *Nat Genet* **48**, 245-52 (2016).
25. Frei, O. *et al.* Bivariate causal mixture model quantifies polygenic overlap between complex traits beyond genetic correlation. *Nat Commun* **10**, 2417 (2019).
26. Werme, J., van der Sluis, S., Posthuma, D. & de Leeuw, C.A. An integrated framework for local genetic correlation analysis. *Nat Genet* **54**, 274-282 (2022).
27. Kurki, M.I. *et al.* FinnGen provides genetic insights from a well-phenotyped isolated population. *Nature* **613**, 508-518 (2023).
28. Bycroft, C. *et al.* The UK Biobank resource with deep phenotyping and genomic data. *Nature* **562**, 203-209 (2018).
29. Verma, A. *et al.* Diversity and Scale: Genetic Architecture of 2,068 Traits in the VA Million Veteran Program. *medRxiv* (2023).
30. O'Connor, L.J. & Price, A.L. Distinguishing genetic correlation from causation across 52 diseases and complex traits. *Nat Genet* **50**, 1728-1734 (2018).
31. 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).
32. Karlsson Linnér, 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).
33. plee0@mgh.harvard.edu, C.-D.G.o.t.P.G.C.E.a. & Consortium, C.-D.G.o.t.P.G. Genomic Relationships, Novel Loci, and Pleiotropic Mechanisms across Eight Psychiatric Disorders. *Cell* **179**, 1469-1482.e11 (2019).
34. Otsuka, K. *et al.* A Testis-Specific Long Noncoding RNA,. *Front Endocrinol (Lausanne)* **12**, 665874 (2021).
35. Domonkos, E., Hodosy, J., Ostatníková, D. & Celec, P. On the Role of Testosterone in Anxiety-Like Behavior Across Life in Experimental Rodents. *Front Endocrinol (Lausanne)* **9**, 441 (2018).
36. Zuloaga, D.G., Heck, A.L., De Guzman, R.M. & Handa, R.J. Roles for androgens in mediating the sex differences of neuroendocrine and behavioral stress responses. *Biol Sex Differ* **11**, 44 (2020).
37. Wendt, F.R. *et al.* Sex-Specific Genetic and Transcriptomic Liability to Neuroticism. *Biol Psychiatry* **93**, 243-252 (2023).
38. Wendt, F.R. *et al.* Sex-stratified gene-by-environment genome-wide interaction study of trauma, posttraumatic-stress, and suicidality. *Neurobiol Stress* **14**, 100309 (2021).
39. Narita, A. *et al.* Clustering by phenotype and genome-wide association study in autism. *Transl Psychiatry* **10**, 290 (2020).
40. Shadrin, A.A. *et al.* Vertex-wise multivariate genome-wide association study identifies 780 unique genetic loci associated with cortical morphology. *Neuroimage* **244**, 118603 (2021).
41. Brouwers, M.C.G.J. Fructose 1-phosphate, an evolutionary signaling molecule of abundancy. *Trends Endocrinol Metab* **33**, 680-689 (2022).
42. Hyldgaard Andersen, S., Black, T., Grassi-Oliveira, R. & Wegener, G. Can early-life high fructose exposure induce long-term depression and anxiety-like behaviours? - A preclinical systematic review. *Brain Res* **1814**, 148427 (2023).
43. Pathak, G.A. *et al.* Genetically regulated multi-omics study for symptom clusters of posttraumatic stress disorder highlights pleiotropy with hematologic and cardio-metabolic traits. *Mol Psychiatry* **27**, 1394-1404 (2022).
44. Elia, L.P., Yamamoto, M., Zang, K. & Reichardt, L.F. p120 catenin regulates dendritic spine and synapse development through Rho-family GTPases and cadherins. *Neuron* **51**, 43-56 (2006).

45. Ryu, T. *et al.* Improved memory and reduced anxiety in δ-catenin transgenic mice. *Exp Neurol* **318**, 22-31 (2019).
46. Li, X. *et al.* Case Report: CNNM2 Mutations Cause Damaged Brain Development and Intractable Epilepsy in a Patient Without Hypomagnesemia. *Front Genet* **12**, 705734 (2021).
47. Deng, Y.T. *et al.* Identifying causal genes for depression via integration of the proteome and transcriptome from brain and blood. *Mol Psychiatry* **27**, 2849-2857 (2022).
48. Martinez, C.A. *et al.* Prenatal stress, anxiety and depression alter transcripts, proteins and pathways associated with immune responses at the maternal-fetal interface†. *Biol Reprod* **106**, 449-462 (2022).
49. Al-Mubarak, B.R. *et al.* Whole exome sequencing in ADHD trios from single and multi-incident families implicates new candidate genes and highlights polygenic transmission. *Eur J Hum Genet* **28**, 1098-1110 (2020).
50. Shang, M.Y. *et al.* Genetic associations between bipolar disorder and brain structural phenotypes. *Cereb Cortex* **33**, 6990-7000 (2023).
51. Gedik, H. *et al.* Identifying potential risk genes and pathways for neuropsychiatric and substance use disorders using intermediate molecular mediator information. *Front Genet* **14**, 1191264 (2023).
52. Meyers, T.J. *et al.* Transcriptome-wide association study identifies novel candidate susceptibility genes for migraine. *HGG Adv* **4**, 100211 (2023).
53. Wendt, F.R. *et al.* Characterizing the effect of background selection on the polygenicity of brain-related traits. *Genomics* **113**, 111-119 (2021).
54. Wendt, F.R., Pathak, G.A., Tylee, D.S., Goswami, A. & Polimanti, R. Heterogeneity and Polygenicity in Psychiatric Disorders: A Genome-Wide Perspective. *Chronic Stress (Thousand Oaks)* **4**, 2470547020924844 (2020).
55. Marshall, M. The hidden links between mental disorders. *Nature* **581**, 19-21 (2020).
56. Huang, Y.C. *et al.* Anxiety comorbidities in patients with major depressive disorder: the role of attachment. *Int J Psychiatry Clin Pract* **23**, 286-292 (2019).
57. Sollis, E. *et al.* The NHGRI-EBI GWAS Catalog: knowledgebase and deposition resource. *Nucleic Acids Res* **51**, D977-D985 (2023).
58. Ike, K.G.O. *et al.* The human neuropsychiatric risk gene Drd2 is necessary for social functioning across evolutionary distant species. *Mol Psychiatry* (2023).
59. Cutando, L. *et al.* Cerebellar dopamine D2 receptors regulate social behaviors. *Nat Neurosci* **25**, 900-911 (2022).
60. Avramidou, M. *et al.* Epidemiology of gastrointestinal symptoms in young and middle-aged Swiss adults: prevalences and comorbidities in a longitudinal population cohort over 28 years. *BMC Gastroenterol* **18**, 21 (2018).
61. Eijsbouts, C. *et al.* Genome-wide analysis of 53,400 people with irritable bowel syndrome highlights shared genetic pathways with mood and anxiety disorders. *Nat Genet* **53**, 1543-1552 (2021).
62. Kremer, M., Becker, L.J., Barrot, M. & Yalcin, I. How to study anxiety and depression in rodent models of chronic pain? *Eur J Neurosci* **53**, 236-270 (2021).
63. Gómez Penedo, J.M. *et al.* The Complex Interplay of Pain, Depression, and Anxiety Symptoms in Patients With Chronic Pain: A Network Approach. *Clin J Pain* **36**, 249-259 (2020).
64. Zanoaga, M.D. *et al.* Brainwide Mendelian Randomization Study of Anxiety Disorders and Symptoms. *Biol Psychiatry* (2023).
65. Celano, C.M., Daunis, D.J., Lokko, H.N., Campbell, K.A. & Huffman, J.C. Anxiety Disorders and Cardiovascular Disease. *Curr Psychiatry Rep* **18**, 101 (2016).
66. Murphy, B., Le Grande, M., Alvarenga, M., Worcester, M. & Jackson, A. Anxiety and Depression After a Cardiac Event: Prevalence and Predictors. *Front Psychol* **10**, 3010 (2019).
67. (!!! INVALID CITATION !!! 25).

68. Denny, J.C. *et al.* The "All of Us" Research Program. *N Engl J Med* **381**, 668-676 (2019).
69. Kurki, M.I. *et al.* FinnGen: Unique genetic insights from combining isolated population and national health register data. *medRxiv*, 2022.03.03.22271360 (2022).
70. Meier, S.M. *et al.* Genetic Variants Associated With Anxiety and Stress-Related Disorders: A Genome-Wide Association Study and Mouse-Model Study. *JAMA Psychiatry* **76**, 924-932 (2019).
71. Otowa, T. *et al.* Meta-analysis of genome-wide association studies of anxiety disorders. *Mol Psychiatry* **21**, 1391-9 (2016).
72. Chang, E. & Mostafa, J. The use of SNOMED CT, 2013-2020: a literature review. *J Am Med Inform Assoc* **28**, 2017-2026 (2021).
73. Purcell, S. *et al.* PLINK: a tool set for whole-genome association and population-based linkage analyses. *Am J Hum Genet* **81**, 559-75 (2007).
74. Mbatchou, J. *et al.* Computationally efficient whole-genome regression for quantitative and binary traits. *Nat Genet* **53**, 1097-1103 (2021).
75. 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).
76. Gaziano, J.M. *et al.* Million Veteran Program: A mega-biobank to study genetic influences on health and disease. *J Clin Epidemiol* **70**, 214-23 (2016).
77. 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).
78. Bulik-Sullivan, B. *et al.* An atlas of genetic correlations across human diseases and traits. *Nat Genet* **47**, 1236-41 (2015).
79. Altshuler, D.M. *et al.* Integrating common and rare genetic variation in diverse human populations. *Nature* **467**, 52-8 (2010).
80. Grotzinger, A.D., Fuente, J., Privé, F., Nivard, M.G. & Tucker-Drob, E.M. Pervasive Downward Bias in Estimates of Liability-Scale Heritability in Genome-wide Association Study Meta-analysis: A Simple Solution. *Biol Psychiatry* **93**, 29-36 (2023).
81. Yang, J., Lee, S.H., Goddard, M.E. & Visscher, P.M. GCTA: a tool for genome-wide complex trait analysis. *Am J Hum Genet* **88**, 76-82 (2011).
82. Ge, T., Chen, C.-Y., Ni, Y., Feng, Y.-C.A. & Smoller, J.W. Polygenic prediction via Bayesian regression and continuous shrinkage priors. *Nature Communications* **10**, 1776 (2019).
83. Finucane, H.K. *et al.* Partitioning heritability by functional annotation using genome-wide association summary statistics. *Nat Genet* **47**, 1228-35 (2015).
84. Pers, T.H. *et al.* Biological interpretation of genome-wide association studies using predicted gene functions. *Nat Commun* **6**, 5890 (2015).
85. Fehrmann, R.S. *et al.* Gene expression analysis identifies global gene dosage sensitivity in cancer. *Nat Genet* **47**, 115-25 (2015).
86. Beach, T.G. *et al.* Arizona Study of Aging and Neurodegenerative Disorders and Brain and Body Donation Program. *Neuropathology* **35**, 354-89 (2015).
87. Beach, T.G. *et al.* The Sun Health Research Institute Brain Donation Program: description and experience, 1987-2007. *Cell Tissue Bank* **9**, 229-45 (2008).
88. Koopmans, F. *et al.* SynGO: An Evidence-Based, Expert-Curated Knowledge Base for the Synapse. *Neuron* **103**, 217-234.e4 (2019).
89. Demontis, D. *et al.* Genome-wide analyses of ADHD identify 27 risk loci, refine the genetic architecture and implicate several cognitive domains. *Nat Genet* **55**, 198-208 (2023).
90. Watson, H.J. *et al.* Genome-wide association study identifies eight risk loci and implicates metabo-psychiatric origins for anorexia nervosa. *Nat Genet* **51**, 1207-1214 (2019).

91. Grove, J. *et al.* Identification of common genetic risk variants for autism spectrum disorder. *Nat Genet* **51**, 431-444 (2019).
92. Mullins, N. *et al.* Genome-wide association study of more than 40,000 bipolar disorder cases provides new insights into the underlying biology. *Nat Genet* **53**, 817-829 (2021).
93. 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).
94. Yu, D. *et al.* Interrogating the Genetic Determinants of Tourette's Syndrome and Other Tic Disorders Through Genome-Wide Association Studies. *Am J Psychiatry* **176**, 217-227 (2019).
95. Stein, M.B. *et al.* Genome-wide association analyses of post-traumatic stress disorder and its symptom subdomains in the Million Veteran Program. *Nat Genet* **53**, 174-184 (2021).
96. Auton, A. *et al.* A global reference for human genetic variation. *Nature* **526**, 68-74 (2015).

Table 1. Cohorts and Anxiety Definitions investigated. Population groups included African (AFR), Admixed-American (AMR), East Asian (EAS), European (EUR), and South Asian (SAS) descent. Case and control sample sizes are reported for binary anxiety definitions, while the total sample size is reported for quantitative anxiety phenotypes. Other abbreviations include International Classification of Diseases (ICD) and Systemized Nomenclature of Medicine (SNOMED).

Cohort	Population	Sample Size	Anxiety Definition
All of Us Research Program (AoU)	AFR	4,937 - 45,092	Generalized Anxiety Disorder (SNOMED ID: 434613), Panic Disorder (SNOMED ID: 4138454), Phobic Disorder (SNOMED ID: 4304010), Chronic Stress Disorder (SNOMED ID: 436074), Acute Stress Disorder (SNOMED ID: 440083), Posttraumatic Stress Disorder (SNOMED ID: 436676), Obsessive-compulsive Disorder (SNOMED ID: 440374), or Mixed Anxiety and Depressive Disorder (SNOMED ID: 4338031)
	AMR	4,157 - 32,477	
	EAS	337 - 4,746	
	EUR	18,461 - 88,954	
	SAS	133 - 2,040	
FinnGen Study	EUR	40,191 - 277,526	KRA_PSY_ANXIETY endpoint
Lundbeck Foundation Initiative for Integrative Psychiatric Research (iPSYCH)	EUR	12,655 - 19,225	ICD-10 F40.0-F41.9; F43.0-F43.9
Million Veteran Program (MVP)	AFR	61,796	Generalized Anxiety Disorder 2-item scale
	EUR	241,541	
Psychiatric Genomics Consortium (PGC)	EUR	5,761 - 11,765	Generalized anxiety disorder, panic disorder, social phobia, agoraphobia, or specific phobias
UK Biobank (UKB)	AFR	130 - 6,116	Phecode 300, “Anxiety, dissociative and somatoform disorders”
	EUR	10,449 - 369,930	
	SAS	172 - 8,189	

FIGURES

Figure 1. Genetic correlations among anxiety phenotypes assessed in participants of European descent (EUR). Full statistics are reported in Supplemental Table 1.

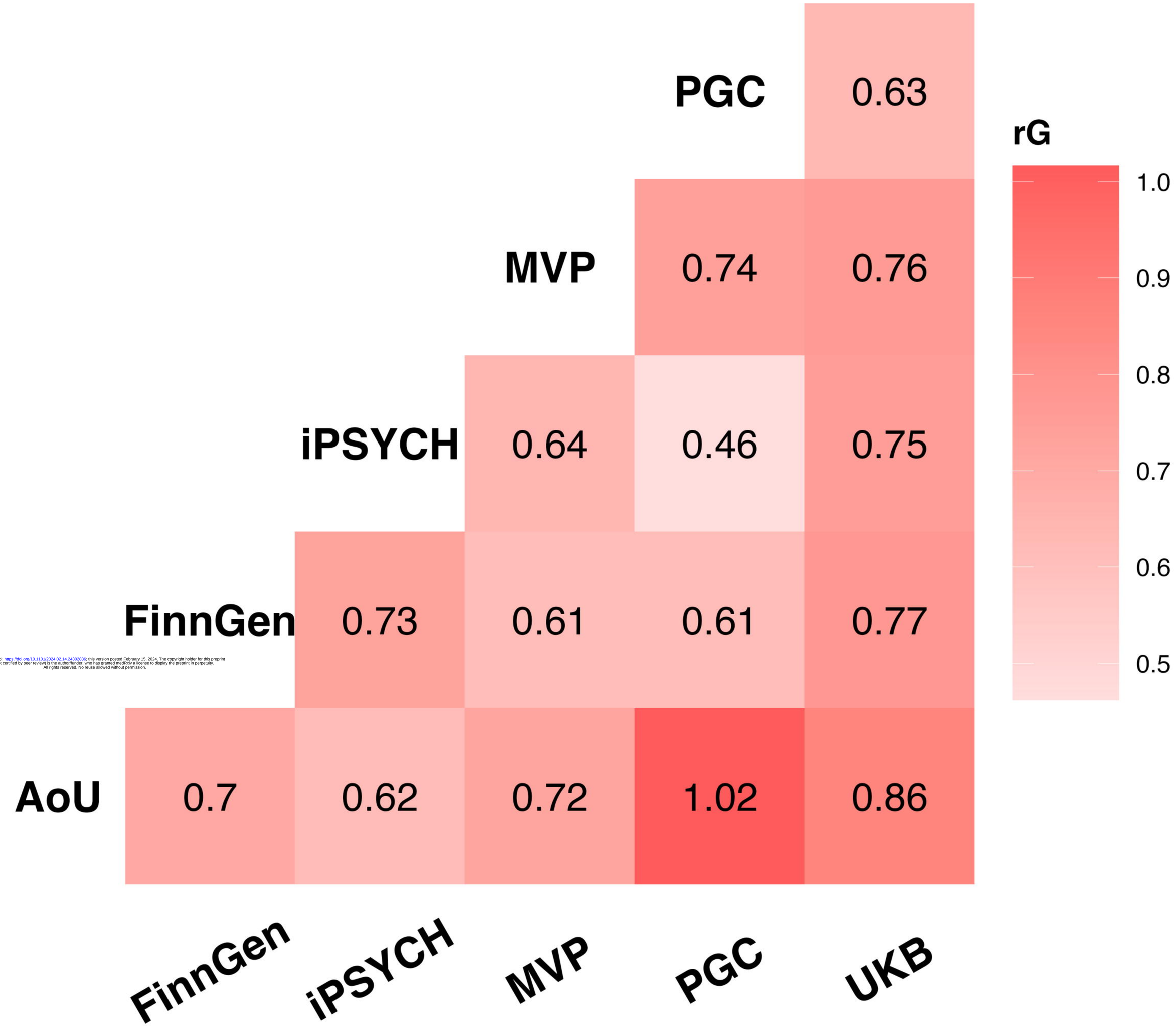
Figure 2. Factor structure of anxiety phenotypes assessed in individuals of European descent. Factor loadings and model fit for the confirmatory factor analysis model of the common anxiety factor (ANX) are reported.

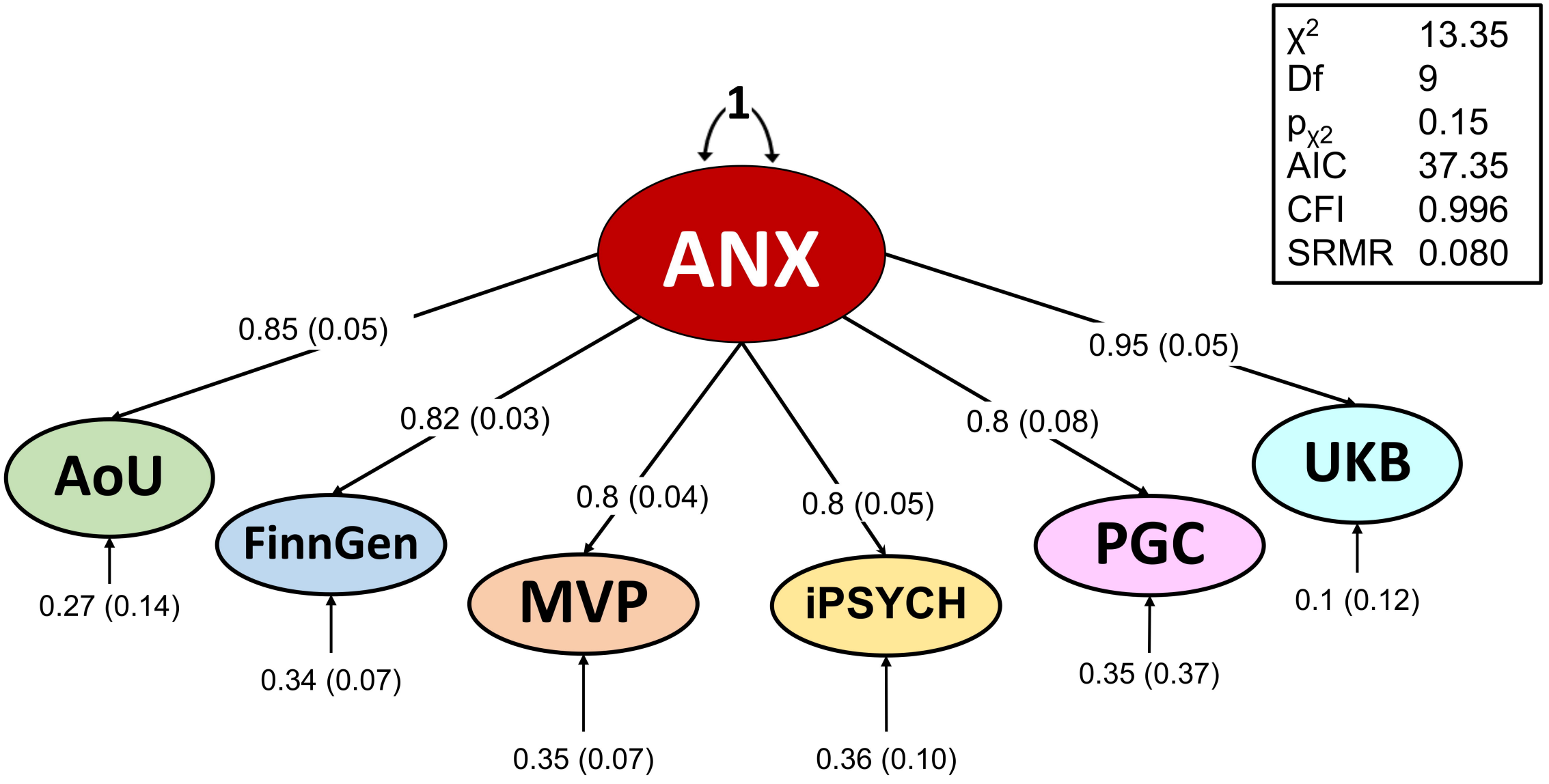
Figure 3. Manhattan plots of genome-wide, transcriptome-wide, and proteome-wide association statistics (bottom, center, and top, respectively) related to the common anxiety factor (ANX). Transcriptome-wide data are those obtained from the multiple-tissue analysis (Supplemental Table 8). Dashed lines represent Bonferroni multiple testing correction applied in each analysis. The labeled genes are those with convergent evidence across analyses.

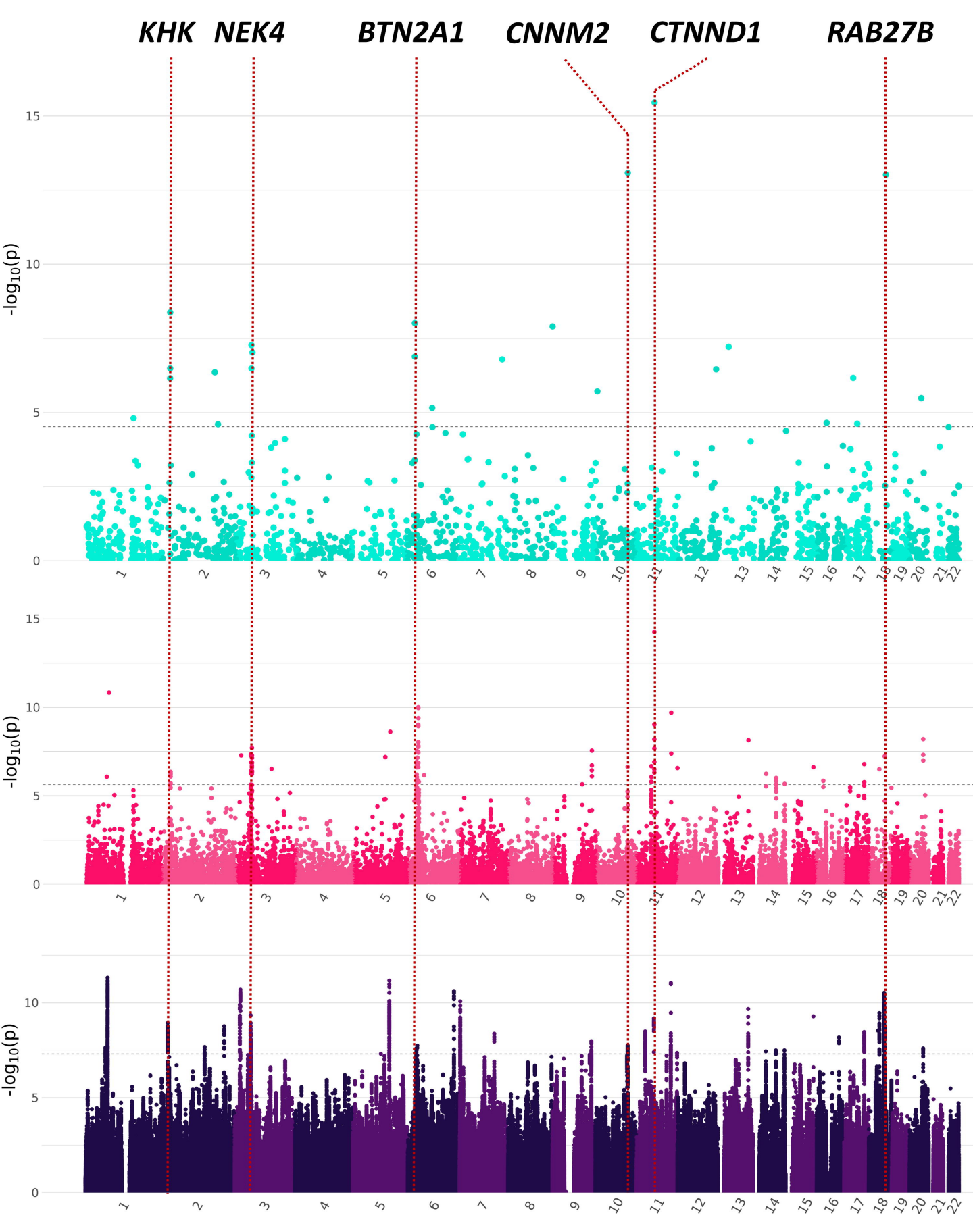
Figure 4. Within-ancestry and cross-ancestry polygenic risk score (PRS) associations of the common anxiety factor (ANX). Full statistics are available in Supplemental Table 4. In the within-ancestry, the EUR-ANX PRS (excluding EUR AoU from the training sample) is tested against the EUR AoU sample. In the cross-ancestry analysis, EUR-ANX PRSs (with and without EUR AoU in the training sample) are tested against AoU samples of African (AoU), Admixed-American (AMR), East Asian (EAS), and South Asian (SAS) descent.

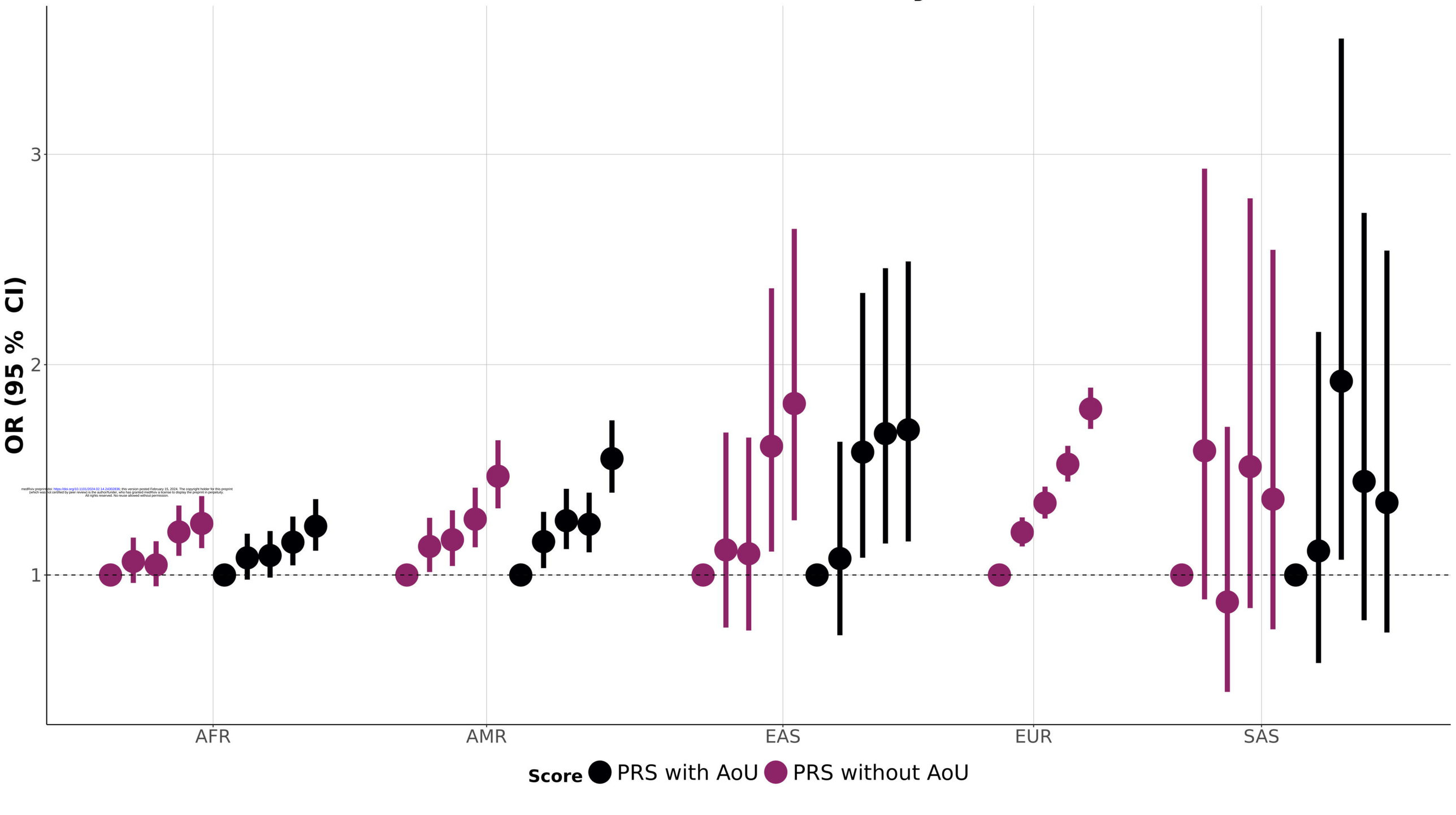
Figure 5 A. Venn diagrams of unique and shared causal variants showing polygenic overlap (gray) of the common anxiety factor (ANX) with major depressive disorder (MDD), bipolar disorder (BIP), and schizophrenia (SCZ). The numbers indicate the estimates of causal variants and their standard errors (in thousands), explaining 90% of the SNP heritability in each phenotype. The size of the circles reflects the degree of polygenicity. **B.** Chord diagram representing ANX local genetic correlations with respect to attention deficit and hyperactivity disorder (ADHD), BIP, MDD, post-traumatic stress disorder (PTSD), and SCZ.

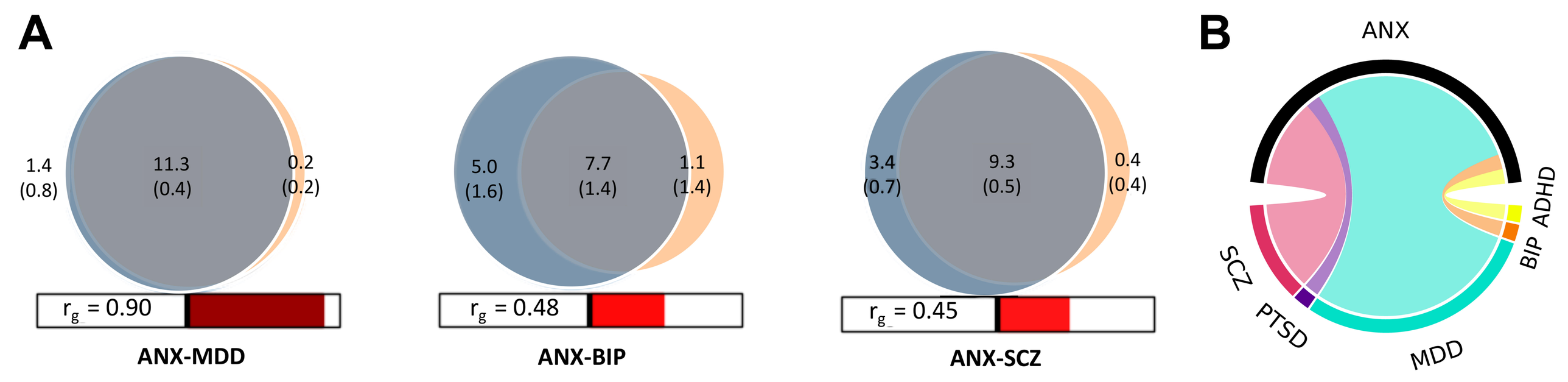
Figure 6. Phenome-wide genetic correlation of the common anxiety factor (ANX). The x-axis reports the genetic correlation of ANX with traits available from UK Biobank, FinnGen, and Million Veteran Program. The y-axis corresponds to two-tailed $-\log_{10}$(p-value). Bonferroni-significant results are color-coded based on the corresponding categories. Full statistics of the Bonferroni-significant results are available in Supplemental Table 14. Non-significant results are reported in grey.




χ^2 13.35
Df 9
$p_{\chi 2}$ 0.15
AIC 37.35
CFI 0.996
SRMR 0.080
1
ANX
0.85 (0.05)
0.82 (0.03)
0.8 (0.04)
0.8 (0.05)
0.8 (0.08)
0.95 (0.05)
AoU
FinnGen
MVP
iPSYCH
PGC
UKB
0.27 (0.14)
0.34 (0.07)
0.35 (0.07)
0.36 (0.10)
0.35 (0.37)
0.1 (0.12)


KHK
NEK4
BTN2A1
CNNM2
CTNND1
RAB27B
-log10(p)
15
10
5
0
1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22




A
1.4
(0.8)
11.3
(0.4)
0.2
(0.2)
r_g = 0.90
ANX-MDD
5.0
(1.6)
7.7
(1.4)
1.1
(1.4)
r_g = 0.48
ANX-BIP
3.4
(0.7)
9.3
(0.5)
0.4
(0.4)
r_g = 0.45
ANX-SCZ
B
ANX
BIP
ADHD
SCZ
PTSD
MDD