

## **The impact of assortative mating, participation bias, and socioeconomic status on the polygenic risk of behavioral and psychiatric traits**

Brenda Cabrera-Mendoza<sup>1,2</sup>, Frank R Wendt<sup>1,2,3,4</sup>, Gita A Pathak<sup>1,2</sup>, Loic Yengo<sup>5</sup>, Renato Polimanti<sup>1,2,\*</sup>

1. Department of Psychiatry, Yale School of Medicine, West Haven, CT 06516, USA

2. VA CT Healthcare System, West Haven, CT, 06516, USA

3. Department of Anthropology, University of Toronto, Toronto, Ontario, Canada

4. Division of Biostatistics, Dalla Lana School of Public Health, University of Toronto, Toronto, Ontario, Canada

5. Institute for Molecular Bioscience, The University of Queensland, Brisbane, Queensland, 4072, Australia

\*Correspondence: Renato Polimanti, PhD. Yale University School of Medicine, Department of Psychiatry, VA CT 116A2, 950 Campbell Avenue, West Haven, CT 06516, USA. Phone: +1 (203) 932-5711 x5745; Fax: +1 (203) 937-3897; E-mail: [renato.polimanti@yale.edu](mailto:renato.polimanti@yale.edu).

## Abstract (148/150 w)

To investigate assortative mating (AM), participation bias, and socioeconomic status (SES) with respect to the genetics of behavioral and psychiatric traits, we analyzed gametic phase disequilibrium (GPD), within-spouses and within-siblings polygenic risk score (PRS) correlation, performing a SES conditional analysis. We observed genetic signatures of AM across multiple methods for traits related to substance use with SES conditioning increasing the within-spouses PRS correlation for *Frequency of drinking alcohol* (2.5% to 6%), *Maximum habitual alcohol intake* (1.33% to 4.43%), and *Ever taken cannabis* (1.5% to 5.3%). Comparing UK Biobank mental health questionnaire responders vs. non-responders, major depressive disorder PRS showed significant GPD in both groups when based on the Million Veteran Program (3.2% vs. 3%), but only in responders when based on the Psychiatric Genomics Consortium (3.8% vs. 0.2%). These results highlight the impact of AM, participation bias, and SES on the polygenic risk of behavioral and psychiatric traits.

## Introduction

Psychiatric disorders and traits are highly polygenic, i.e. they are influenced by several thousands of genetic variants, each having a small effect on disease risk<sup>1</sup>. Large-scale genome-wide studies have demonstrated that our ability to investigate their polygenic architecture could be influenced by several factors such as assortative mating (AM; i.e., mate choice driven by phenotypic similarity) and participation bias (i.e., individuals with a certain phenotype are more likely to enter a study)<sup>2-4</sup>. For example, AM increases the genetic variance in a population because it induces a systematic positive correlation between trait-increasing alleles across the genome<sup>5</sup>. Therefore, AM can result in inflated genetic effects as compared to those estimated in randomly mating population or using a family-based design. On the other hand, participation bias might exacerbate differences between sub-groups in a study, thereby reducing the generalizability of genetic effects estimated from the whole sample<sup>6</sup>.

As exemplified hereafter, both AM and participation bias have been linked to behavioral traits and psychiatric disorders. A previous study investigating more than 700,000 individuals reported evidence of mate resemblance within and across eleven psychiatric disorders, including attention-deficit/hyperactivity disorder (ADHD), autism spectrum disorder (ASD), schizophrenia, bipolar disorder, major depressive disorder (MDD), generalized anxiety disorder (GAD), agoraphobia, social phobia, obsessive-compulsive disorder (OCD), anorexia, and substance abuse<sup>7</sup>. A subsequent analysis provided an estimate of the genetic consequences of AM for psychiatric traits, suggesting a modest impact on their heritability but this may be considerable for the population prevalence of rare disorders with a high heritability<sup>8</sup>. Recently, an analysis of cross-phenotype AM highlighted how cross-phenotype mate correlations may bias estimates of genetic correlation between pairs of psychiatric disorders<sup>9</sup>.

With respect to participation bias<sup>10-15</sup>, certain psychiatric traits (e.g., MDD, anxiety, and alcohol consumption) can be associated with the likelihood of becoming or remaining as study participants<sup>16,17</sup>. Genome-wide investigations showed that the non-participation in health surveys (including mental health assessments) is genetically correlated with several behavioral traits, such as educational attainment and neuroticism, and neuropsychiatric disorders, such as schizophrenia and Alzheimer's disease<sup>10,13</sup>. Additionally, differential participation bias between sexes (i.e., participation bias where the determinants of study participation affect women and men to differing extents) is genetically correlated with behavioral and psychiatric traits, including educational attainment, risk-taking behaviors, cannabis use, loneliness, MDD, ASD, schizophrenia, and ADHD<sup>12</sup>.

Factors responsible for AM (e.g., opportunities or exposure to potential intimate partners, individual preferences, and third-party constraints) can be strongly affected by socioeconomic status (SES)<sup>18</sup>. Similarly, the proportions of non-participation are typically not uniform across sociodemographic groups, with those from deprived backgrounds often under-represented in health surveys<sup>15</sup>. To our knowledge, no previous study systematically investigated the potential impact of SES on the genetic signatures of AM and participation bias across multiple behavioral and psychiatric traits. Accordingly, we evaluated whether the SES association with the polygenic risk of psychiatric and behavioral traits contributes to the genetic signatures of AM and participation bias. Specifically, we leveraged findings from large-scale genome-wide association studies generated by the Psychiatric Genomics Consortium (PGC)<sup>19</sup> and the Million Veteran Program (MVP)<sup>20</sup> together with individual-level data from UK Biobank

(UKB) participants that completed the UKB Mental Health Questionnaire (MHQ) and comparing to those that did not complete this assessment.

## Methods

### *UK Biobank*

The UKB is a general population-based cohort comprising approximately 502,000 participants. This sample was recruited between 2006 and 2010 in 22 assessment centers across the UK<sup>21</sup>. UKB received ethical approval from the NHS National Research Ethics Service Northwest (reference: 11/NW/0382). UKB obtained informed written consent from all participants. A self-reported detailed account of sociodemographic, lifestyle, mental, and physical health information was collected from all UKB participants<sup>21</sup>.

The collection and processing of UKB genetic data have been described previously<sup>22</sup>. Briefly, genome-wide genotype data were obtained from all UKB participants using the UKB Axiom array. UKB genotypic data were imputed using the Haplotype Reference Consortium reference panel. In this study, we analyzed a sample of 362,132 unrelated individuals of European descent (EUR) with available genotype data. Because of the limited sample size available, we were not able to analyze other ancestry groups. Ancestry and relatedness information of each UKB participant were obtained from the Pan-ancestry genetic analysis of the UKB (Pan-UKB)<sup>23</sup>. Briefly, genetic relatedness among UKB participants was estimated with PC-Relate, a principal component analysis (PCA)-based method<sup>24</sup>. While ancestry assignment was performed using a combined reference panel including both the 1000 Genomes Project<sup>25</sup> and Human Genome Diversity Panel (HGDP)<sup>26</sup>. Then, a random forest classifier trained with the top six principal components (PCs) from the reference data was applied to the UKB PCs data. UKB participants were assigned to an ancestry group (African, Admixed American, Central/South Asian, East Asian, EUR, or Middle Eastern) based on a classifier probability >50%. A detailed description of the Pan-UKB methods is available at <https://pan.ukbb.broadinstitute.org>.

### *UKB Mental Health Questionnaire*

As part of the UKB assessment, behavioral and psychiatric outcomes, including mood disorders, anxiety, mental distress, self-harm, traumatic events, substance use, and post-traumatic stress disorder (PTSD), were evaluated with an online follow-up assessment including the UKB MHQ (Resource 22 on <http://biobank.ctsu.ox.ac.uk>)<sup>27</sup>. The MHQ was completed by 157,366 UKB participants (31% of total UKB participants) aged 45–82 years, 57% of them were female and had a higher SES (i.e., higher income and higher educational attainment) compared with UKB participants who did not complete this assessment<sup>27,28</sup>.

Also, diagnostic criteria were evaluated for MDD, hypomania or mania, GAD, alcohol use disorder (AUD), and PTSD<sup>27</sup>. Addiction to substances other than alcohol and/or behavior was defined based only on self-report. The MHQ is based on previously existing and validated measures, including the Composite International Diagnostic Interview Short Form (CIDI-SF) to assess lifetime mental disorders in general<sup>29</sup>, as well as instruments for specific mental disorders and trauma exposures, i.e., the PTSD Check List - Civilian Short version (PCL-6), the Childhood Trauma Screener – 5 item (CTS-5)<sup>30,31</sup>, and validated instruments developed specifically for the UKB such as an adult trauma screener<sup>27</sup>.

## *Large-scale Genome-wide Association Studies of Behavioral and Psychiatric Traits*

To investigate additional behavioral and psychiatric traits, we investigate AM genetic signatures using GWAS statistics that were generated from samples that did not include UKB. These included ADHD<sup>32</sup>, ASD<sup>33</sup>, anorexia nervosa<sup>34</sup>, anxiety disorder<sup>35</sup>, bipolar disorder<sup>36</sup>, bipolar disorder type I<sup>36</sup>, bipolar disorder type II<sup>36</sup>, MDD<sup>37</sup>, schizophrenia<sup>38</sup>, panic disorder<sup>39</sup>, PTSD<sup>40</sup>, Tourette syndrome<sup>41</sup>, and OCD<sup>42</sup> obtained from the PGC (available at <https://pgc.unc.edu/for-researchers/download-results/>); and AUD<sup>43</sup>, Alcohol Use Disorder Identification Test-Consumption (AUDIT-C)<sup>43</sup>, maximum habitual alcohol intake<sup>44</sup>, MDD<sup>45</sup>, opioid use disorder (OUD)<sup>46</sup>, PTSD<sup>47</sup>, and anxiety disorder<sup>48</sup> obtained from the MVP (available at [https://www.ncbi.nlm.nih.gov/projects/gap/cgi-bin/study.cgi?study\\_id=phs001672.v8.p1](https://www.ncbi.nlm.nih.gov/projects/gap/cgi-bin/study.cgi?study_id=phs001672.v8.p1)). Briefly, PGC data was generally generated from meta-analyses of genome-wide genetic data derived from many cohorts with different characteristics and assessed with different instruments<sup>49</sup>. Conversely, MVP data were obtained from a single observational cohort study of US veterans<sup>50</sup>. The genetic correlation between PGC and MVP overlapping traits, i.e., MDD, PTSD, and anxiety was calculated with the Linkage Disequilibrium Score Regression method (LDSC)<sup>51</sup>. A description of these datasets is shown in Supplementary Table 1. Due to the limited sample size available for other ancestry groups in UKB, PGC, and MVP, our analyses were limited only to EUR.

## *Gametic Phase Disequilibrium Analysis*

We estimated AM genetic signatures across behavioral and psychiatric traits using a method proposed by Yengo et al.,<sup>5</sup>. Briefly, this method is based on the fact that AM signatures of a specific trait can be quantified as the directional correlation between trait-increasing alleles, also referred as gametic phase disequilibrium (GPD). This can be estimated as the correlation between trait-specific PRS based on variants located on odd and even chromosomes<sup>5</sup>.

Initially, we used GWAS statistics generated from the analysis of MHQ traits assessed in EUR MHQ responders to calculate PRS with respect to 243,476 EUR unrelated UKB participants who did not respond to the MHQ (MHQ non-responders). MHQ GWAS statistics were derived from the Pan-UKB analysis that used a generalized mixed model association testing framework available from the Scalable and Accurate Implementation of GEneralized (SAIGE) software<sup>52</sup>. A detailed description of these GWAS is available at <https://pan.ukbb.broadinstitute.org>. To reduce the multiple-testing burden, we tested only MHQ traits with SAIGE heritability estimates > 0.03 and single nucleotide polymorphisms-based heritability (SNP-h<sup>2</sup>) p-value < 0.05. For each MHQ trait, SNP-h<sup>2</sup> was estimated using the LDSC approach<sup>51</sup> as described in the Supplemental Methods.

Before PRS calculation, quality control was performed on GWAS summary statistics and UK individual genotypic data using PLINK 1.9<sup>53</sup>. SNPs were included if they meet the following criteria: i) Hardy-Weinberg equilibrium (HWE) p values > 1 × 10<sup>-6</sup>, ii) missingness < 0.05, iii) minor allele frequencies ≥ 0.01. PRS analysis included only LD-independent SNPs, selected with a clump r<sup>2</sup> < 0.1 for SNPs < 1Mb apart using 1000 Genomes Project EUR populations as reference<sup>25</sup>.

Even- and odd-chromosomes PRS were calculated using the software package PRSice<sup>54,55</sup>. We included 20 PCs from SNPs in even and odd chromosomes when calculating the PRS for odd and even chromosomes, respectively to correct for population stratification. PCs were calculated from LD-independent SNPs in even and odd chromosomes separately using the fast PCA approach implemented

in PLINK version 2.0<sup>56,57</sup>. LD pruning was performed in PLINK ( $r^2 < 0.1$  for SNPs  $< 1\text{Mb}$  apart) using HapMap3 (Utah residents with ancestry from northern and western EUR (CEU) as reference<sup>58</sup>.

As suggested by the GPD method developers<sup>5</sup>, a P-value threshold of 0.005 was applied to select SNPs included for PRS calculation. Thus, AM was estimated as the coefficient from a linear regression model of the even-chromosomes PRS ( $\text{PRS}_{\text{even}}$ ) onto odd-chromosomes PRS ( $\text{PRS}_{\text{odd}}$ ) and 20 PCs from the SNPs in odd chromosomes:

$$\text{PRS}_{\text{even}} \sim \text{PRS}_{\text{odd}} + 20\text{PC}_{\text{odd}} \text{ or } \text{PRS}_{\text{odd}} \sim \text{PRS}_{\text{even}} + 20\text{PC}_{\text{even}}$$

AM estimate of each MHQ trait was obtained from the regression onto the PRS with the larger variance. A false discovery rate (FDR  $q < 0.05$ ) was applied to correct the GPD results for the number of phenotypes evaluated in each analysis. Then, we calculated GPD estimates for psychiatric traits in unrelated EUR UKB participants ( $N=362,132$ ), and to assess the effect of SES on participation bias in the UKB, we compared GPD estimates for psychiatric traits between MHQ-responders ( $N= 118,656$ ) and MHQ-non-responders ( $N= 243,476$ ) using a z-test. Also, we assessed PRS-distribution differences between both groups using a t-test.

#### *Within-spouses and within-siblings polygenic risk correlation assessment*

In addition to using the GPD approach, we also estimated AM testing PRS correlation within spouses (WSps) and within siblings (WSib) available in the UKB cohort. The first analysis can be informative of the AM in the current generation, while the second one is informative of the AM in the previous generations.

Putative spouses were identified using a method described previously<sup>59</sup>. First, we selected only UKB EUR participants identified as unrelated by kinship coefficients ( $N=362,132$ ). Then, we selected pairs of individuals who were of opposite sex that reported identical and complete information for the following fields: (a) living with their spouse (UKB field ID: 6141), (b) length of time living in the house (UKB field ID: 699), (c) number of occupants in the household (UKB field ID: 709), (d) number of vehicles (UKB field ID: 728), (e) accommodation type and rental status (UKB field IDs: 670, 680), (f) home coordinates (UKB field IDs: 20074, 20075) and (g) registered in the same UKB recruitment center (UKB field ID: 54) and (h) available genotype data. When more than two individuals shared identical information, then these individuals were removed. To confirm the lack of relatedness in the selected sample, we recalculated the kinship coefficients using Kinship-based INference for GWAS (KING) toolset<sup>60</sup>. Three closely related pairs (identical by descent  $> 0.1$ ) were removed and only those individuals classified as unrelated by KING were further analyzed. A total of 45,570 putative spouse pairs were identified and included in our analysis.

The WSps-PRS correlation was calculated as the coefficient of the regression of PRS of spouse 1 onto PRS of spouse 2 and 20 PCs of spouse 2 to adjust by population stratification:

$$\text{PRS}_{\text{Spouse1}} \sim \text{PRS}_{\text{Spouse2}} + 20\text{PC}_{\text{Spouse2}}$$

Evidence of possible AM in the current generation was considered when the WSps-PRS correlation was statistically significant after FDR multiple testing correction (FDR  $q < 0.05$ ).

To estimate AM in previous generations, we estimated the WSib-PRS correlation within siblings in the UKB. We selected only UKB EUR participants identified as related by kinship coefficients ( $n = 64,304$ ). To estimate relatedness in the selected sample, we calculated the kinship coefficients using the KING toolset<sup>60</sup>, and further analyzed only those individuals classified as full siblings by this algorithm. Only two siblings per family were selected for further analysis. A total of 17,911 sibling pairs were included in our analysis.

The WSib-PRS correlation was calculated as the coefficient of the regression of PRS of sibling 1 onto PRS of sibling 2 and 20 PCs of sibling 2 to adjust by population stratification:

$$\text{PRS}_{\text{Sibling1}} \sim \text{PRS}_{\text{Sibling2}} + 20\text{PC}_{\text{Sibling2}}$$

Evidence of possible AM in previous generations was considered significant when the WSib-PRS correlation was statistically different from 0.5 (WSib-PRS  $\rho = 0.5\Delta$ ) after FDR multiple testing correction (FDR  $q < 0.05$ ).

### *Conditional analysis to account for the effect of socioeconomic status on polygenic risk of behavioral and psychiatric traits*

To evaluate whether the estimates of AM genetic signatures were affected by SES, we conducted a conditional analysis adjusting the GWAS summary statistics. Specifically, we used the multi-trait-based conditional and joint analysis (mtCOJO)<sup>61</sup> to adjust UKB-MHQ, PGC, and MVP GWAS statistics by the effect of two SES-related variables: household income (HI, UKB Data-Field 738; i.e., the combined gross income of all members of a household) and the Townsend deprivation index (TDI, UKB Data-Field 189; i.e., a measure of material deprivation based on four variables: unemployment, non-car ownership, non-home ownership, and household overcrowding aggregated for postcodes of residence)<sup>62</sup>. The p-value threshold to select SNPs for clumping in mtCOJO was 0.05. We generated SES-adjusted GWAS summary statistics considering three models: i) HI-adjusted, ii) TDI-adjusted, and iii) adjusted for both HI and TDI. The HI and TDI GWAS statistics were generated by analyzing unrelated EUR UKB-MHQ responders ( $N = 118,656$ ) as described in the Supplementary Methods.

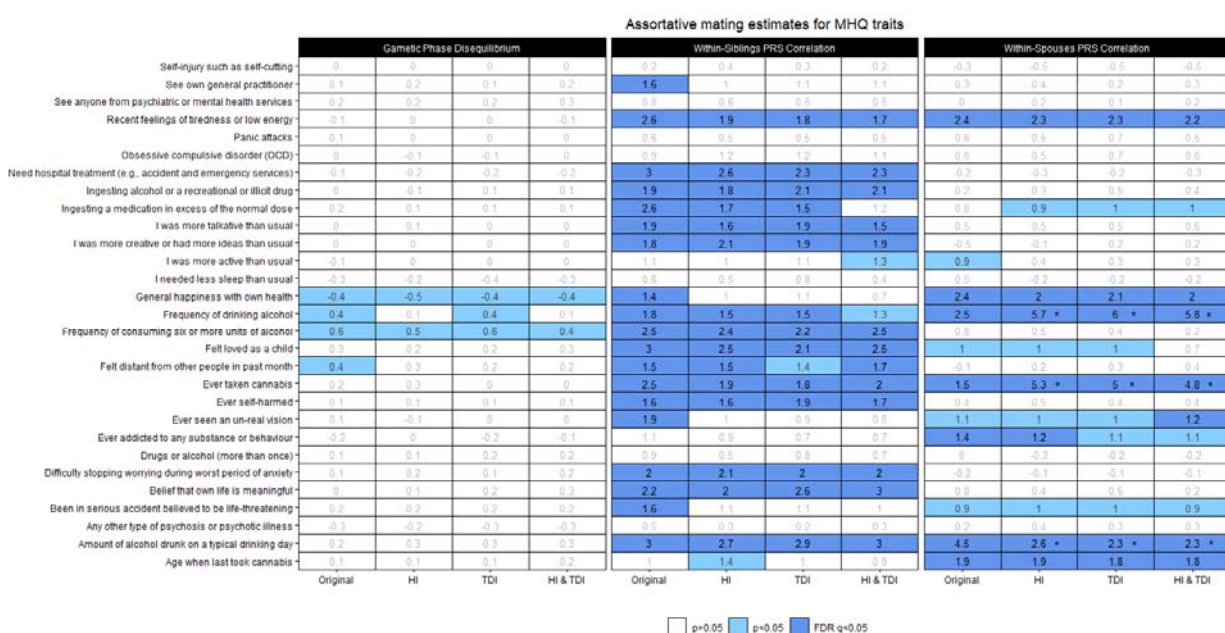
Leveraging the SES-adjusted GWAS summary statistics, the estimates of AM genetic signatures for the behavioral and psychiatric traits were re-estimated and statistical differences between original and SES-adjusted estimates were tested with z-tests. A false discovery rate (FDR  $q < 0.05$ ) was applied to correct the results for the number of phenotypes evaluated. Furthermore, we evaluated the effect of SES-adjustment on heritability by estimating  $\text{SNP-}h^2$  for the SES-adjusted GWAS summary statistics of the included traits using LDSC<sup>51</sup>.

## **Results**

### *Genetic signatures of assortative mating in MHQ-derived traits*

We investigated genetic signatures of AM on MHQ-derived traits (Supplementary Table 2) using three different approaches: GPD<sup>5</sup>, WSib-PRS correlation<sup>63</sup>, and WSps-PRS correlation<sup>59</sup> (Figure 1). WSib- and WSps-PRS correlation analyses identified multiple AM genetic signatures surviving FDR multiple testing correction, while GPD analysis showed only nominally significant findings. However, although they

model different aspects of AM, we observed consistency across the three methods. Interestingly, the MHQ-derived traits showing evidence of AM in at least two analyses were all related to substance use (mostly alcohol consumption) and emotional well-being. The two traits that showed significant estimates in all analyses (WSib- and WSps-PRS correlation FDR  $q < 0.05$  and GPD  $p < 0.05$ ) were *Frequency of drinking alcohol* (UKB Field ID: 20414) and *General happiness with own health* (UKB Field ID: 20459). Consistency between WSib- and WSps-PRS correlation analyses (FDR  $q < 0.05$  in both) was also observed for *Amount of alcohol drunk on a typical drinking day* (UKB Field ID: 20403), *Ever taken cannabis* (UKB Field ID: 20453), and *Recent feelings of tiredness or low energy*, UKB Field ID: 20519). The WSps-PRS correlation and GPD analyses showed significant results (FDR  $q < 0.05$  and  $p < 0.05$ , respectively) for *Frequency of consuming six or more units of alcohol* (UKB Field ID: 20416) and *Felt distant from other people in past month* (UKB Field ID: 20496). Below, we described the results obtained in each analysis and the differences observed after adjusting for SES variables.



**Figure 1.** Genetic signatures of assortative mating across Mental Health Questionnaire (MHQ) traits. Cell shades correspond to the significance strength of each estimate, from white ( $p > 0.05$ , non-significant), light blue ( $p < 0.05$ , nominally-significant), to dark blue (FDR  $q < 0.05$ , FDR-significant). The estimate reported as a percentage is shown in the center of each cell. An asterisk in the cell indicates a significant difference of the SES-conditioned estimate with respect to the original estimate (difference- $p < 0.05$ ). Abbreviations: Household income (HI); Townsend Deprivation Index (TDI).

In the WSib-PRS correlation analysis, we identified FDR-significant results with respect to 19 MHQ-derived traits (FDR  $q < 0.05$ ; Figure 1, Supplemental Table 3). These included substance use (e.g., *Frequency of consuming six or more units of alcohol*: WSib-PRS  $\rho - 0.5\Delta = 0.03$ ,  $p = 7.05 \times 10^{-5}$ , UKB Field ID: 20416; *Ever taken cannabis*: WSib-PRS  $\rho - 0.5\Delta = 0.03$ ,  $p = 9.14 \times 10^{-5}$ , UKB Field ID: 20453), self-harm behaviors (e.g., *Ingesting a medication in excess of the normal dose*: WSib-PRS  $\rho - 0.5\Delta = 0.03$ ,  $p = 5.41 \times 10^{-5}$ , UKB Field ID: 20553, coding:4; *Need hospital treatment following self-harm*: WSib-PRS  $\rho - 0.5\Delta = 0.03$ ,  $p = 2.14 \times 10^{-6}$ , UKB Field ID: 20554, coding:3), negative emotions (e.g., *Recent feelings of tiredness or low energy*: WSib-PRS  $\rho - 0.5\Delta = 0.03$ ,  $p = 5.55 \times 10^{-5}$ , UKB Field ID: 20519), positive emotions (e.g., *Belief that own life is meaningful*: WSib-PRS  $\rho - 0.5\Delta = 0.02$ ,  $p = 5.31 \times 10^{-4}$ , UKB Field ID: 20460),



social support (*Felt loved as a child*: Wsib-PRS  $\rho=0.5\Delta=0.03$ ,  $p=2.48 \times 10^{-6}$ , UKB Field ID:20489), mania manifestations (e.g., *I was more talkative than usual*: Wsib-PRS  $\rho=0.5\Delta=0.02$ ,  $p=0.003$ , UKB Field ID:20548, coding:1), and other psychiatric symptoms (e.g., *Difficulty stopping worrying during worst period of anxiety*: Wsib-PRS  $\rho=0.5\Delta=0.02$ ,  $p=0.002$ , UKB Field ID:20541). Wsib-PRS  $\rho=0.5\Delta$  estimates remained virtually unchanged after conditioning for SES variables (Figure 1, Supplemental Table 3).

Considering the WSps-PRS correlation, we observed significant results surviving multiple testing correction (FDR  $q < 0.05$ ) with respect to seven MHQ-derived traits (Figure 1, Supplemental Table 4). These included *Ever addicted to any substance or behaviour* (WSps-PRS  $\rho=0.014$ ,  $p=0.003$ ; UKB Field ID:20401), *Amount of alcohol drunk on a typical drinking day* (WSps-PRS  $\rho=0.045$ ,  $p=1.29 \times 10^{-21}$ ; UKB Field ID:20403), *Frequency of drinking alcohol* (WSps-PRS  $\rho=0.025$ ,  $p=1.52 \times 10^{-7}$ ; UKB Field ID:20414), *Ever taken cannabis* (WSps-PRS  $\rho=0.042$ ,  $p=0.002$ ; UKB Field ID:20453), *Age when last took cannabis* (WSps-PRS  $\rho=0.019$ ,  $p=4.11 \times 10^{-5}$ ; UKB Field ID: 20455), *General happiness with own health* (WSps-PRS  $\rho=0.024$ ,  $p=3.04 \times 10^{-5}$ ; UKB Field ID:20459), and *Recent feelings of tiredness or low energy* (WSps-PRS  $\rho=0.024$ ,  $p=2.85 \times 10^{-7}$ ; UKB Field ID:20519). Considering the WSps-PRS correlation conditioned with respect to SES variables, four of these traits showed statistically significant differences when compared to the unconditioned estimates (Figure 1; Supplementary Table 4). Statistically significant increases of the WSps-PRS correlation were observed for *Frequency of drinking alcohol* (original estimate 2.5% vs. TDI-adjusted estimate 6%, difference- $p=9.44 \times 10^{-8}$ ) and *Ever taken cannabis* (original estimate 1.5% vs. HI-adjusted estimate 5.3%, difference- $p=7.88 \times 10^{-9}$ ). The *Amount of alcohol drunk on a typical drinking day* WSps-PRS correlation from 4.5% in the original estimate changed to 2.3% in the estimate adjusted for both HI and TDI (difference- $p=9.13 \times 10^{-4}$ ). While these changes were consistent across the different SES conditioning performed (i.e., HI, TDI, and HI+TDI), a significant reduction in *General happiness with own health* WSps-PRS correlation was observed, decreasing from 2.4% in the original estimate to 0.6% when adjusted for HI (difference- $p=0.005$ ), but not for the other SES-adjusted traits (difference- $p > 0.05$ ).

GPD estimates with respect to MHQ-derived traits were not significant after FDR correction (Figure 1, Supplementary Table 5).

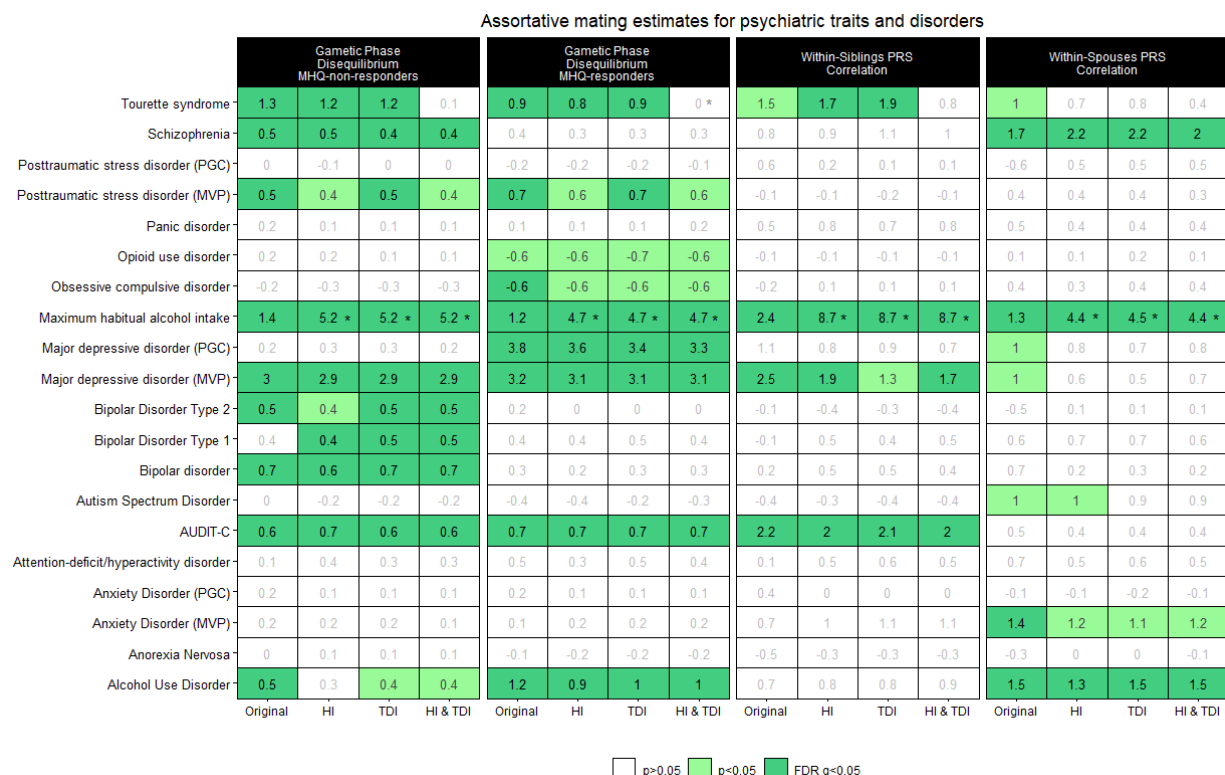
With respect to LDSC  $h^2$  estimates, we observed nominally-significant differences after adjusting by HI and both HI and TDI for *Frequency of drinking alcohol* (original LDSC  $h^2=8.3\%$  vs. LDSC  $h^2$  HI-adjusted=6.7%; difference- $p=0.028$ ; original LDSC  $h^2=8.3\%$  vs. LDSC  $h^2$  HI-TDI-adjusted=6.7%; difference- $p=0.028$ ; UKB Field ID:20414) and *Ever taking cannabis* (original LDSC  $h^2=7.29\%$  vs. LDSC  $h^2$  TDI-adjusted=5.6%; difference- $p=0.021$ ; UKB Field ID:20453) (Supplemental Table 2).

### *Genetic signatures of assortative mating in psychiatric traits and disorders*

To further expand the breadth of our study, we investigated genetic signatures of AM on psychiatric traits and disorders previously analyzed by large-scale PGC and MVP GWAS (Supplementary Table 1). For three phenotypes, we had information from both MVP and PGC GWAS. The genetic correlation between traits assessed in both PGC and MVP datasets was 1.01 (se=0.17) for anxiety, 0.50 (se=0.15) for PTSD, and 0.92 (se=0.01) for MDD. The limited genetic correlation between PGC-PTSD (excluding UKB) and MVP-PTSD is in line with what previously reported<sup>64</sup>.

We investigated GPD estimates in psychiatric traits and disorders in the whole sample combining MHQ responders and non-responders (Supplementary Table 6; Supplementary Figure 1), as well as differences

between MHQ responders vs. non-responders with respect to PRS (Supplementary Table 7; Supplementary Figure 2) and GPD for psychiatric disorders. Comparing GPD, WSib-PRS correlation, and WSps-PRS correlation results (Figure 2), we observed FDR-significant AM genetic signatures in at least two different methods for seven phenotypes. In particular, maximum habitual alcohol intake showed FDR-significant results in all approaches and in both MHQ responders and non-responders. Similar consistency was observed for MDD assessed in MVP and Tourette syndrome where FDR-significant evidence was observed in the GPD analysis (both MHQ responders and non-responders) and WSib-PRS correlation with only nominally significant WSps-PRS correlation. Below, we described the results obtained in each analysis and the differences observed after adjusting for SES variables.



**Figure 2.** Genetic signatures of assortative mating across psychiatric traits and disorders. Cell shades correspond to the significance strength of each estimate, from white ( $p > 0.05$ , non-significant), light green ( $p < 0.05$ , nominally significant), to dark green ( $FDR q < 0.05$ , FDR-significant). The estimate reported as a percentage is shown in the center of each cell. An asterisk in the cell indicates a significant difference of the SES-conditioned estimate with respect to the original estimate (difference- $p < 0.05$ ). Abbreviations: Psychiatric Genomics Consortium (PGC); Million Veteran Program (MVP); Mental Health Questionnaire (MHQ); Household income (HI); Townsend Deprivation Index (TDI).

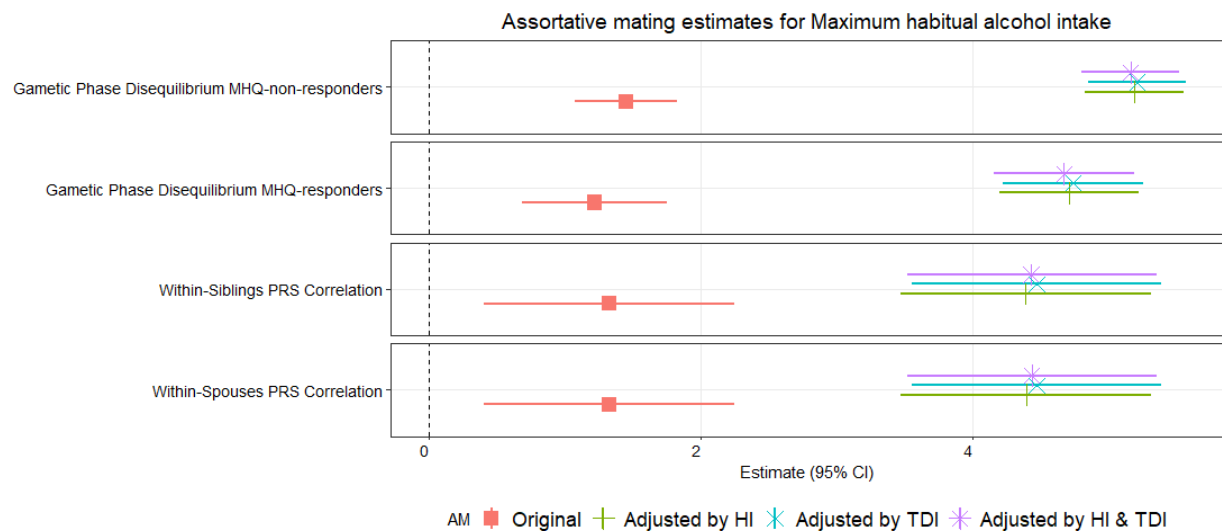
As mentioned above, the GPD analysis was conducted in MHQ responders and non-responders, separately (Figure 2; Supplementary Tables 8 and 9, respectively). We found significant GPD estimates ( $FDR q < 0.05$ ) for six psychiatric traits in both groups: AUD ( $GPD_{\text{Responders}} = 1.2\%$ ,  $p = 7.09 \times 10^{-6}$ ;  $GPD_{\text{Non-Responders}} = 0.45\%$ ,  $p = 0.016$ ), AUDIT-C ( $GPD_{\text{Responders}} = 0.7\%$ ,  $p = 0.008$ ;  $GPD_{\text{Non-Responders}} = 0.62\%$ ,  $p = 7.92 \times 10^{-4}$ ), MVP-MDD ( $GPD_{\text{Responders}} = 3.23\%$ ,  $p = 9.44 \times 10^{-33}$ ;  $GPD_{\text{Non-Responders}} = 2.98\%$ ,  $p = 7.04 \times 10^{-55}$ ), maximum habitual alcohol intake ( $GPD_{\text{Responders}} = 1.22\%$ ,  $p = 8.43 \times 10^{-6}$ ;  $GPD_{\text{Non-Responders}} = 1.45\%$ ,  $p = 3.99 \times 10^{-14}$ ), MVP-PTSD ( $GPD_{\text{Responders}} = 0.66\%$ ,  $p = 0.018$ ;  $GPD_{\text{Non-Responders}} = 0.52\%$ ,  $p = 0.008$ ), and Tourette syndrome

( $GPD_{\text{Responders}}=0.94\%$ ,  $p=3.22 \times 10^{-4}$ ;  $GPD_{\text{Non-Responders}}=1.26\%$ ,  $p=4.56 \times 10^{-12}$ ). Also, we found FDR-significant GPD estimates related to BD ( $GPD_{\text{Non-Responders}}=0.68\%$ ,  $p=1.69 \times 10^{-4}$ ), BD1 ( $GPD_{\text{Non-Responders}}=0.39\%$ ,  $p=0.036$ ), BD2 ( $GPD_{\text{Non-Responders}}=0.46\%$ ,  $p=0.016$ ), and schizophrenia ( $GPD_{\text{Non-Responders}}=0.51\%$ ,  $p=9.03 \times 10^{-4}$ ) in MHQ-non-responders. While there were significant GPD estimates with respect to PGC-MDD ( $GPD_{\text{Responders}}=3.83\%$ ,  $p=7.5 \times 10^{-50}$ ) and OCD ( $GPD_{\text{Responders}}=-0.64\%$ ,  $p=0.018$ ) in MHQ-responders. Although there was some variation in the GDP estimates between the two samples investigated, we found significant differences between MHQ-responders and MHQ-non-responders only for PGC-MDD ( $GPD_{\text{Responders}}=3.83\%$  vs.  $GPD_{\text{Non-Responders}}=0.23\%$ , difference- $p=2.27 \times 10^{-30}$ ) and AUD ( $GPD_{\text{Responders}}=1.2\%$  vs.  $GPD_{\text{Non-Responders}}=0.45\%$ , difference- $p=0.02$ ). Also, we found a significant difference for PGC-MDD GPD estimates in MHQ-responders ( $GPD_{\text{EUR}}=1.44\%$  vs.  $GPD_{\text{Responders}}=3.83\%$ , difference- $p=8.36 \times 10^{-16}$ ) and MHQ-non-responders ( $GPD_{\text{EUR}}=1.44\%$  vs.  $GPD_{\text{Non-Responders}}=0.23\%$ , difference- $p=1.79 \times 10^{-7}$ ) with respect to those in EUR (Supplementary Table 10).

After conditioning on SES variables, we observed statistically significant changes in the GPD estimates only for maximum habitual alcohol intake and Tourette syndrome. These changes were observable in both MHQ responders and non-responders (Figure 2; Supplementary Tables 8 and 9). For TDI adjustment of maximum habitual alcohol intake, GPD increased from 1.22% and 1.45% to 4.74% and 5.21% in MHQ-responders (difference- $p=1.61 \times 10^{-20}$ ) and MHQ-non-responders (difference- $p=3.68 \times 10^{-45}$ ), respectively. Conversely, HI-TDI adjustment reduced the GPD estimates for Tourette syndrome from 0.94% and 1.26% in MHQ-responders down to <0.001% and 0.14% (difference- $p=0.018$ ) in MHQ-non-responders (difference- $p=2.47 \times 10^{-5}$ ).

We observed significant WSps-PRS correlation (FDR  $q<0.05$ ) with respect to four traits (Figure 2, Supplementary Table 11): AUD (WSps-PRS  $\rho=1.47\%$ ,  $p=0.002$ ), MVP-assessed anxiety disorder (WSps-PRS  $\rho=1.4\%$ ,  $p=0.003$ ), maximum habitual alcohol intake (WSps-PRS  $\rho=1.32\%$ ,  $p=0.005$ ), and schizophrenia (WSps-PRS  $\rho=1.74\%$ ,  $p=2.11 \times 10^{-4}$ ). Conditioning for SES variables, we observed significant changes in the WSps-PRS correlation only for maximum habitual alcohol intake where the estimate increased from 1.32% to 4.47% after accounting for TDI (difference- $p=2.24 \times 10^{-6}$ ). A similar SES effect was also present in WSib-PRS correlation analysis where the WSib-PRS  $\rho-0.5\Delta$  estimate changed from 2.4% ( $p=1.39 \times 10^{-4}$ ) to 5.87% ( $p=1.97 \times 10^{-45}$ ) after accounting for TDI (difference- $p=8.32 \times 10^{-13}$ ). FDR-significant WSib-PRS  $\rho-0.5\Delta$  estimates were observed also for other two traits (Figure 2; Supplementary Table 9): AUDIT-C (WSib-PRS  $\rho-0.5\Delta=0.022$ ,  $p=0.001$ ), and MVP-MDD (WSib-PRS  $\rho-0.5\Delta=0.025$ ,  $p=9.61 \times 10^{-5}$ ). The conditioning for SES variables did not change the estimates observed for the other traits (difference- $p>0.05$ ; Figure 2, Supplementary Table 12).

Considering the results of the SES conditioning across the three methods applied, we observed a strong and statistically significant increase of the AM genetic signature in the polygenic risk of maximum habitual alcohol intake. Specifically, the adjustment for HI, TDI, and both HI and TDI increased GPD, WSib-PRS  $\rho-0.5\Delta$ , and WSps-PRS  $\rho$  estimates more than three times (Figure 3).



**Figure 3.** Genetic signatures of assortative mating for Maximum habitual alcohol intake before and after SES conditioning. 95% confidence intervals (CI) are reported for each estimate. Abbreviations: Mental Health Questionnaire (MHQ); Household income (HI); Townsend Deprivation Index (TDI).

## Discussion

In the present study, we used multiple methods to detect genetic signatures of AM across behavioral and psychiatric traits. We also quantified differences in AM genetic signatures due to i) the effect of SES factors, ii) participation bias by comparing UKB MHQ responders and non-responders, and iii) related to the genetic effect estimates detected by GWAS conducted in samples with different characteristics (i.e., PGC and MVP).

In the UKB MHQ-based analysis, we found consistent evidence of AM genetic signatures across multiple methods for traits related to substance use and emotional well-being. With respect to substance use, three outcomes were related to alcohol consumption (i.e., *Frequency of drinking alcohol*, *Amount of alcohol drunk on a typical drinking day*, and *Frequency of consuming six or more units of alcohol*) and one to cannabis use (i.e., *Ever taken cannabis*). While several studies have demonstrated phenotypic resemblance between spouses pairs regarding alcohol drinking behaviors<sup>65-67</sup>, limited information is available regarding AM and cannabis use. It has been hypothesized that individuals pick companions compatible with, and supportive of, their substance use, leading to high levels of similarity between romantic partners<sup>68</sup>. Similar mechanisms could be responsible also for the genetic signatures observed with respect to individual feelings (i.e., *General happiness with own health*, *Recent feelings of tiredness or low energy*, and *Felt distant from other people in past month*). Indeed, evidence of the effect of negative emotions on mating preferences has been previously reported<sup>69</sup>. Interestingly, *General happiness with own health* showed significant evidence of AM across the three methods applied, but with a nominally-significant negative GDP. This suggests that factors other than AM may affect the genetics of this trait in UKB. With respect to possible temporal changes, our results showed mostly evidence of AM both in current and past generations (i.e., significant PRS correlation within spouses and within siblings, respectively). However, considering the two domains identified (i.e., substance use and

emotions), *Frequency of consuming six or more units of alcohol*, *Felt distant from other people in past month*, *Difficulty stopping worrying during worst period of anxiety*, and *Belief that own life is meaningful* showed genetic signatures for past-generations AM but not for current generation AM. A similar pattern was present also for other MHQ traits, including those related to self-harm (e.g., *Ever self-harmed*) and mania manifestation (e.g., *I was more talkative than usual*). Conversely, the only two traits with evidence of current-generation AM and not to past-generations AM were related to substance use: *Age when last took cannabis* and *Ever addicted to any substance or behaviour*. These different AM patterns may be related to temporal changes in demographic phenomena in the UK populations. For instance, the increased availability of cannabis<sup>70</sup> may influence mating preferences more in the current generation than in previous ones.

With respect to the SES conditioning in the MHQ analyses, changes in the AM genetic signatures were observed only for traits related to substance use with respect to the WSps-PRS correlation. However, they showed different patterns: increased WSps-PRS correlation for *Frequency of drinking alcohol* and *Ever taken cannabis* and reduced WSps-PRS correlation for *Frequency of consuming six or more units of alcohol* after SES conditioning. This could be due to the known genetic differences between substance use and substance use disorders observed for alcohol, cannabis, and other addictive substances<sup>71-73</sup>. Additionally, misreports and longitudinal changes have been reported to bias genetic associations of alcohol-drinking behaviors<sup>74</sup>. These factors could contribute to the effect of SES conditioning on the WSps-PRS correlation of MHQ-derived alcohol consumption phenotypes together with the complex association of SES observed across the spectrum of alcohol use traits<sup>75</sup>. The analyses based on PGC and MVP GWAS data also showed AM genetic signatures across multiple alcohol-drinking phenotypes: AUD (GPD<sub>responders</sub>; GPD<sub>non-responders</sub>; WSps-PRS), AUDIT-C (GPD<sub>responders</sub>; GPD<sub>non-responders</sub>; WSib-PRS), and maximum habitual alcohol intake (GPD<sub>responders</sub>; GPD<sub>non-responders</sub>; WSib-PRS; WSps-PRS). Interestingly, while significant GDP estimates were observed across these traits, the possible AM temporal scale appears to be different. AUD, an indicator of alcohol addiction, showed AM genetic signatures only in the current generation (i.e., significant estimates for WSps-PRS and not for WSib-PRS). Conversely, AUDIT-C, an indicator of alcohol consumption, showed AM genetic signatures only in the past generations (i.e., significant estimates for WSib-PRS and not for WSps-PRS). In line with the fact that it is less related to addiction than AUD and is more genetically correlated with AUDIT-C than AUD in UKB<sup>76</sup>, maximum habitual alcohol intake showed AM genetic signatures in both current and past generations. Additionally, SES conditioning showed inflated AM genetic signatures of maximum habitual alcohol intake across all methods used (Figure 3). Accordingly, the same population phenomena and/or confounders contributing to the SES-conditioned inflation of AM genetic signature observed in UKB MHQ substance-use traits (i.e., *Frequency of drinking alcohol* and *Ever taken cannabis*) could also be involved in the SES-conditioned inflation of AM genetic signatures observed in MVP maximum habitual alcohol intake. Unfortunately, the limited availability of large-scale GWAS of substance use disorders<sup>77</sup> did not permit us to fully explore patterns of fully AM genetic differences across different substances.

The analysis of PGC and MVP GWAS data (not including UKB) allowed us to explore GDP differences between UKB MHQ responders and non-responders. While most of the GDP estimates were not statistically different between these UKB subsamples, we observed higher GDP estimates for AUD and PGC-assessed MDD in MHQ responders than in non-responders. As previously described<sup>28,78</sup>, UKB MHQ responders have higher socioeconomic status, are healthier, and report less severe internalizing symptoms. Accordingly, the differences observed may be due to the MHQ participation bias, which may

not only affect the generalizability of the prevalence of the MHQ-assessed traits but also the characterization of the polygenic architecture.

Because we had access to PGC and MVP GWAS data related to the same phenotypes (anxiety disorder, MDD, and PTSD), we were able to consider this additional layer of variability. As previously shown<sup>28,79,80</sup> and confirmed in this study, there is a moderate to high PGC-MVP genetic correlation with respect to these traits. As mentioned, PGC-MDD showed different GPD estimates between MHQ responders and non-responders (3.8% vs. 0.2%). Conversely, MVP-MDD showed FDR-significant and similar GPD estimates in both groups (3.2% vs. 3%). With respect to PTSD, while there was no difference with respect to MHQ participation, FDR-significant GPD was observed when testing MVP-PTSD PRS (responders=0.66%; non-responders=0.52%) but not when considering PGC-PTSD PRS (responders=-0.02%; non-responders=-0.01%). Instead, anxiety disorder showed only FDR-significant WSPs-PRS correlation when testing MVP data but not when considering PGC data. These differences are likely due to the characteristics of the PGC and MVP cohorts. PGC GWAS are based on the meta-analysis of many cohorts including participants from different countries that were assessed with different instruments and were enrolled with different recruiting strategies<sup>19,81</sup>. Conversely, MVP GWAS are based on a single cohort that includes only US veterans that were assessed with the same instruments and were enrolled through the same recruiting strategy<sup>20</sup>. Based on the differences observed between PGC- vs. MVP-based analyses, we hypothesize that the analysis of AM genetic signatures based on PRS generated from GWAS meta-analyses may be less affected by the specific characteristics of a cohort (e.g., target population group, assessment, and recruitment strategy). Instead, the analysis of AM genetic signatures based on PRS generated from GWAS conducted in a single cohort may be more affected by the characteristics of that cohort. For example, MVP and UKB cohorts likely present population dynamics that are specific to US and UK demographic histories, respectively<sup>82,83</sup>. Although PRS differences are plausible contributors to GPD differences between both groups, the widespread presence of such differences across the analyzed phenotypes i.e., not being limited to phenotypes where we found significant AM estimates, and the opposite direction of such differences to GPD estimates make them unlikely to significantly influence our results. Accordingly, the AM genetic signatures generated from the analysis of GWAS generated from MVP and UKB-MHQ data could be influenced by cohort-specific characteristics.

Although our study provides new insights into the impact of AM, participation bias, and SES on the polygenic risk of behavioral and psychiatric traits, we acknowledge several limitations. First, due to confidentiality, the identified spouse pairs cannot be confirmed. Thus, it is possible that to an unknown extent the identified pairs do not correspond to actual spouse pairs. Second, the statistical power of the GWAS used to generate the PRS may have contributed to the differences observed across methods and datasets. Third, we cannot discard that observed changes in AM estimates after controlling for SES-related variables may be partially influenced by a potential collider bias<sup>84</sup>. Fourth, the limited availability of large-scale GWAS representative of diverse ancestry groups limited the present study only to data generated from participants of European descent.

In conclusion, we provide evidence of the possible interplay among AM, participation bias, and SES in the polygenic risk of multiple behavioral and psychiatric disorders. Our findings indicate that population phenomena and cohort-specific characteristics could influence our ability to model the polygenicity of traits related to mental health. This highlights the need to model more accurately different aspects that could influence the generalizability of genetic effects detected across cohorts and study designs.

## Data availability

All results used to make conclusions discussed in this study are provided as Supplementary Material. All GWAS data are publicly available on their respective websites.

UK Biobank, <https://www.ukbiobank.ac.uk/enable-your-research/apply-for-access>

Psychiatric Genomics Consortium, <https://pgc.unc.edu/for-researchers/download-results/>

Million Veteran Program, [https://www.ncbi.nlm.nih.gov/projects/gap/cgi-bin/study.cgi?study\\_id=phs001672.v8.p1](https://www.ncbi.nlm.nih.gov/projects/gap/cgi-bin/study.cgi?study_id=phs001672.v8.p1)

## Acknowledgements

We thank the participants and the investigators involved in the UK Biobank, Million Veteran Program, and the Psychiatric Genomics Consortium for making their data publicly available. This research was conducted using the UK Biobank Resource (application reference no. 58146). The authors acknowledge support from the National Institutes of Health (R21 DC018098, R33 DA047527, RF1 MH132337, and K99 AG078503), One Mind, and the American Foundation for Suicide Prevention (PDF-1-022-21).

## References

1. Wendt, F. R., Pathak, G. A., Overstreet, C., Tylee, D. S., Gelernter, J., Atkinson, E. G., & Polimanti, R. (2021). Characterizing the effect of background selection on the polygenicity of brain-related traits. *Genomics*, 113(1 Pt 1), 111–119. <https://doi.org/10.1016/j.ygeno.2020.11.032>
2. Plomin, R., & von Stumm, S. (2022). Polygenic scores: Prediction versus explanation. *Molecular Psychiatry*, 27(1), 49–52. <https://doi.org/10.1038/s41380-021-01348-y>
3. Polygenic Risk Score Task Force of the International Common Disease Alliance, Adeyemo, A., Balaconis, M. K., Darnes, D. R., Fatumo, S., Granados Moreno, P., Hodonsky, C. J., Inouye, M., Kanai, M., Kato, K., Knoppers, B. M., Lewis, A. C. F., Martin, A. R., McCarthy, M. I., Meyer, M. N., Okada, Y., Richards, J. B., Richter, L., Ripatti, S., ... Zhou, A. (2021). Responsible use of polygenic risk scores in the clinic: Potential benefits, risks and gaps. *Nature Medicine*, 27(11), 1876–1884. <https://doi.org/10.1038/s41591-021-01549-6>
4. Kerminen, S., Martin, A. R., Koskela, J., Ruotsalainen, S. E., Havulinna, A. S., Surakka, I., Palotie, A., Perola, M., Salomaa, V., Daly, M. J., Ripatti, S., & Pirinen, M. (2019). Geographic variation and bias in the polygenic scores of complex diseases and traits in Finland. *The American Journal of Human Genetics*, 104(6), 1169–1181. <https://doi.org/10.1016/j.ajhg.2019.05.001>
5. Yengo, L., Robinson, M. R., Keller, M. C., Kemper, K. E., Yang, Y., Trzaskowski, M., Gratten, J., Turley, P., Cesarini, D., Benjamin, D. J., Wray, N. R., Goddard, M. E., Yang, J., & Visscher, P. M. (2018). Imprint of assortative mating on the human genome. *Nature Human Behaviour*, 2(12), 948–954. <https://doi.org/10.1038/s41562-018-0476-3>
6. Taylor, A. E., Jones, H. J., Sallis, H., Euesden, J., Stergiakouli, E., Davies, N. M., Zammit, S., Lawlor, D. A., Munafò, M. R., Davey Smith, G., & Tilling, K. (2018). Exploring the association of genetic factors with participation in the Avon Longitudinal Study of Parents and Children. *International Journal of Epidemiology*, 47(4), 1207–1216. <https://doi.org/10.1093/ije/dyy060>
7. Nordsletten, A. E., Larsson, H., Crowley, J. J., Almqvist, C., Lichtenstein, P., & Mataix-Cols, D. (2016). Patterns of Nonrandom Mating Within and Across 11 Major Psychiatric Disorders. *JAMA Psychiatry*, 73(4), 354–361. <https://doi.org/10.1001/jamapsychiatry.2015.3192>

8. Peyrot, W. J., Robinson, M. R., Penninx, B. W. J. H., & Wray, N. R. (2016). Exploring boundaries for the genetic consequences of assortative mating for psychiatric traits. *JAMA Psychiatry*, 73(11), 1189–1195. <https://doi.org/10.1001/jamapsychiatry.2016.2566>
9. Border, R., Athanasiadis, G., Buil, A., Schork, A. J., Cai, N., Young, A. I., Werge, T., Flint, J., Kendler, K. S., Sankararaman, S., Dahl, A. W., & Zaitlen, N. A. (2022). Cross-trait assortative mating is widespread and inflates genetic correlation estimates. *Science*, 378(6621), 754–761. <https://doi.org/10.1126/science.abo2059>
10. Adams, M. J., Hill, W. D., Howard, D. M., Dashti, H. S., Davis, K. A. S., Campbell, A., Clarke, T.-K., Deary, I. J., Hayward, C., Porteous, D., Hotopf, M., & McIntosh, A. M. (2020). Factors associated with sharing e-mail information and mental health survey participation in large population cohorts. *International Journal of Epidemiology*, 49(2), 410–421. <https://doi.org/10.1093/ije/dyz134>
11. Lee, H., & Han, B. (2022). A theory-based practical solution to correct for sex-differential participation bias. *Genome Biology*, 23(1), 138. <https://doi.org/10.1186/s13059-022-02703-0>
12. Pirastu, N., Cordioli, M., Nandakumar, P., Mignogna, G., Abdellaoui, A., Hollis, B., Kanai, M., Rajagopal, V. M., Parolo, P. D. B., Baya, N., Carey, C. E., Karjalainen, J., Als, T. D., Van der Zee, M. D., Day, F. R., Ong, K. K., FinnGen Study, 23andMe Research Team, Agee, M., ... Ganna, A. (2021). Genetic analyses identify widespread sex-differential participation bias. *Nature Genetics*, 53(5), 663–671. <https://doi.org/10.1038/s41588-021-00846-7>
13. Tyrrell, J., Zheng, J., Beaumont, R., Hinton, K., Richardson, T. G., Wood, A. R., Davey Smith, G., Frayling, T. M., & Tilling, K. (2021). Genetic predictors of participation in optional components of UK Biobank. *Nature Communications*, 12(1), 886. <https://doi.org/10.1038/s41467-021-21073-y>
14. Russell, G., Mandy, W., Elliott, D., White, R., Pittwood, T., & Ford, T. (2019). Selection bias on intellectual ability in autism research: A cross-sectional review and meta-analysis. *Molecular Autism*, 10(1), 9. <https://doi.org/10.1186/s13229-019-0260-x>
15. McMinn, M. A., Martikainen, P., Gorman, E., Rissanen, H., Härkänen, T., Tolonen, H., Leyland, A. H., & Gray, L. (2019). Validation of non-participation bias methodology based on record-linked Finnish register-based health survey data: A protocol paper. *BMJ Open*, 9(4), e026187. <https://doi.org/10.1136/bmjopen-2018-026187>
16. Lamers, F., Hoogendoorn, A. W., Smit, J. H., van Dyck, R., Zitman, F. G., Nolen, W. A., & Penninx, B. W. (2012). Sociodemographic and psychiatric determinants of attrition in the Netherlands study of depression and anxiety (Nesda). *Comprehensive Psychiatry*, 53(1), 63–70. <https://doi.org/10.1016/j.comppsy.2011.01.011>
17. Gorman, E., Leyland, A. H., McCartney, G., White, I. R., Katikireddi, S. V., Rutherford, L., Graham, L., & Gray, L. (2014). Assessing the representativeness of population-sampled health surveys through linkage to administrative data on alcohol-related outcomes. *American Journal of Epidemiology*, 180(9), 941–948. <https://doi.org/10.1093/aje/kwu207>
18. Lichter, D. T., & Qian, Z. (2019). The study of assortative mating: Theory, data, and analysis. In R. Schoen (Ed.), *Analytical Family Demography* (Vol. 47, pp. 303–337). Springer International Publishing. [https://doi.org/10.1007/978-3-319-93227-9\\_13](https://doi.org/10.1007/978-3-319-93227-9_13)
19. Sullivan, P. F., Agrawal, A., Bulik, C. M., Andreassen, O. A., Børghlum, A. D., Breen, G., Cichon, S., Edenberg, H. J., Faraone, S. V., Gelernter, J., Mathews, C. A., Nievergelt, C. M., Smoller, J. W., O'Donovan, M. C., & Psychiatric Genomics Consortium. (2018). Psychiatric genomics: An update and an agenda. *The American Journal of Psychiatry*, 175(1), 15–27. <https://doi.org/10.1176/appi.ajp.2017.17030283>
20. 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. *Journal of Clinical Epidemiology*, 70, 214–223. <https://doi.org/10.1016/j.jclinepi.2015.09.016>
21. 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 Medicine*, 12(3), e1001779. <https://doi.org/10.1371/journal.pmed.1001779>
22. 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>
23. Pan-UKB team. <https://pan.ukbb.broadinstitute.org>. 2020.
24. Conomos, M. P., Reiner, A. P., Weir, B. S., & Thornton, T. A. (2016). Model-free estimation of recent genetic relatedness. *American Journal of Human Genetics*, 98(1), 127–148. <https://doi.org/10.1016/j.ajhg.2015.11.022>
25. 1000 Genomes Project Consortium, 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>
26. Li, J. Z., Absher, D. M., Tang, H., Southwick, A. M., Casto, A. M., Ramachandran, S., Cann, H. M., Barsh, G. S., Feldman, M., Cavalli-Sforza, L. L., & Myers, R. M. (2008). Worldwide human relationships inferred from genome-wide patterns of variation. *Science (New York, N.Y.)*, 319(5866), 1100–1104. <https://doi.org/10.1126/science.1153717>
27. Davis, K. A. S., Coleman, J. R. I., Adams, M., Allen, N., Breen, G., Cullen, B., Dickens, C., Fox, E., Graham, N., Holliday, J., Howard, L. M., John, A., Lee, W., McCabe, R., McIntosh, A., Pearsall, R., Smith, D. J., Sudlow, C., Ward, J., ... Hotopf, M. (2020). Mental health in UK Biobank – development, implementation and results from an online questionnaire completed by 157 366 participants: A reanalysis. *BJPsych Open*, 6(2), e18. <https://doi.org/10.1192/bjo.2019.100>
28. Wendt, F. R., Pathak, G. A., Deak, J. D., De Angelis, F., Koller, D., Cabrera-Mendoza, B., Lebovitch, D. S., Levey, D. F., Stein, M. B., Kranzler, H. R., Koenen, K. C., Gelernter, J., Huckins, L. M., & Polimanti, R. (2022). Using phenotype risk scores to enhance gene discovery for generalized anxiety disorder and posttraumatic stress disorder. *Molecular Psychiatry*, 27(4), 2206–2215. <https://doi.org/10.1038/s41380-022-01469-y>
29. Kessler, R. C., Andrews, G., Mroczek, D., Ustun, B., & Wittchen, H.-U. (1998). The world health organization composite international diagnostic interview short-form(CIDI-SF). *International Journal of Methods in Psychiatric Research*, 7(4), 171–185. <https://doi.org/10.1002/mpr.47>
30. Wilkins, K. C., Lang, A. J., & Norman, S. B. (2011). Synthesis of the psychometric properties of the PTSD checklist (Pcl) military, civilian, and specific versions. *Depression and Anxiety*, 28(7), 596–606. <https://doi.org/10.1002/da.20837>
31. Glaesmer, H., Schulz, A., Häuser, W., Freyberger, H., Brähler, E., & Grabe, H.-J. (2013). Der childhood trauma screener (Cts)—Entwicklung und validierung von schwellenwerten zur klassifikation. *Psychiatrische Praxis*, 40(04), 220–226. <https://doi.org/10.1055/s-0033-1343116>
32. Demontis, D., Walters, R. K., Martin, J., Mattheisen, M., Als, T. D., Agerbo, E., Baldursson, G., Belliveau, R., Bybjerg-Grauholm, J., Bækvad-Hansen, M., Cerrato, F., Chambert, K., Churchhouse, C., Dumont, A., Eriksson, N., Gandal, M., Goldstein, J. I., Grasby, K. L., Grove, J.,

- ... Neale, B. M. (2019). Discovery of the first genome-wide significant risk loci for attention deficit/hyperactivity disorder. *Nature Genetics*, 51(1), 63–75. <https://doi.org/10.1038/s41588-018-0269-7>
33. Grove, J., Ripke, S., Als, T. D., Mattheisen, M., Walters, R. K., Won, H., Pallesen, J., Agerbo, E., Andreassen, O. A., Anney, R., Awasthi, S., Belliveau, R., Bettella, F., Buxbaum, J. D., Bybjerg-Grauholm, J., Bækvad-Hansen, M., Cerrato, F., Chambert, K., Christensen, J. H., ... Børglum, A. D. (2019). Identification of common genetic risk variants for autism spectrum disorder. *Nature Genetics*, 51(3), 431–444. <https://doi.org/10.1038/s41588-019-0344-8>
34. Duncan, L., Yilmaz, Z., Gaspar, H., Walters, R., Goldstein, J., Anttila, V., Bulik-Sullivan, B., Ripke, S., Eating Disorders Working Group of the Psychiatric Genomics Consortium, Thornton, L., Hinney, A., Daly, M., Sullivan, P. F., Zeggini, E., Breen, G., & Bulik, C. M. (2017). Significant locus and metabolic genetic correlations revealed in genome-wide association study of anorexia nervosa. *The American Journal of Psychiatry*, 174(9), 850–858. <https://doi.org/10.1176/appi.ajp.2017.16121402>
35. Otowa, T., Hek, K., Lee, M., Byrne, E. M., Mirza, S. S., Nivard, M. G., Bigdeli, T., Aggen, S. H., Adkins, D., Wolen, A., Fanous, A., Keller, M. C., Castelao, E., Kutalik, Z., Van der Auwera, S., Homuth, G., Nauck, M., Teumer, A., Milaneschi, Y., ... Hettema, J. M. (2016). Meta-analysis of genome-wide association studies of anxiety disorders. *Molecular Psychiatry*, 21(10), 1391–1399. <https://doi.org/10.1038/mp.2015.197>
36. Stahl, E. A., Breen, G., Forstner, A. J., McQuillin, A., Ripke, S., Trubetskoy, V., Mattheisen, M., Wang, Y., Coleman, J. R. I., Gaspar, H. A., de Leeuw, C. A., Steinberg, S., Pavlides, J. M. W., Trzaskowski, M., Byrne, E. M., Pers, T. H., Holmans, P. A., Richards, A. L., Abbott, L., ... Bipolar Disorder Working Group of the Psychiatric Genomics Consortium. (2019). Genome-wide association study identifies 30 loci associated with bipolar disorder. *Nature Genetics*, 51(5), 793–803. <https://doi.org/10.1038/s41588-019-0397-8>
37. Coleman, J. R. I., Gaspar, H. A., Bryois, J., Bipolar Disorder Working Group of the Psychiatric Genomics Consortium, Major Depressive Disorder Working Group of the Psychiatric Genomics Consortium, & Breen, G. (2020). The genetics of the mood disorder spectrum: Genome-wide association analyses of more than 185,000 cases and 439,000 controls. *Biological Psychiatry*, 88(2), 169–184. <https://doi.org/10.1016/j.biopsych.2019.10.015>
38. Lam, M., Chen, C.-Y., Li, Z., Martin, A. R., Bryois, J., Ma, X., Gaspar, H., Ikeda, M., Benyamin, B., Brown, B. C., Liu, R., Zhou, W., Guan, L., Kamatani, Y., Kim, S.-W., Kubo, M., Kusumawardhani, A. A. A., Liu, C.-M., Ma, H., ... Huang, H. (2019). Comparative genetic architectures of schizophrenia in East Asian and European populations. *Nature Genetics*, 51(12), 1670–1678. <https://doi.org/10.1038/s41588-019-0512-x>
39. Forstner, A. J., Awasthi, S., Wolf, C., Maron, E., Erhardt, A., Czamara, D., Eriksson, E., Lavebratt, C., Allgulander, C., Friedrich, N., Becker, J., Hecker, J., Rambau, S., Conrad, R., Geiser, F., McMahon, F. J., Moebus, S., Hess, T., Buerfent, B. C., ... Schumacher, J. (2021). Genome-wide association study of panic disorder reveals genetic overlap with neuroticism and depression. *Molecular Psychiatry*, 26(8), 4179–4190. <https://doi.org/10.1038/s41380-019-0590-2>
40. Duncan, L. E., Ratanatharathorn, A., Aiello, A. E., Almli, L. M., Amstadter, A. B., Ashley-Koch, A. E., Baker, D. G., Beckham, J. C., Bierut, L. J., Bisson, J., Bradley, B., Chen, C.-Y., Dalvie, S., Farrer, L. A., Galea, S., Garrett, M. E., Gelernter, J. E., Guffanti, G., Hauser, M. A., ... Koenen, K. C. (2018). Largest GWAS of PTSD (N=200,070) yields genetic overlap with schizophrenia and sex differences in heritability. *Molecular Psychiatry*, 23(3), 666–673. <https://doi.org/10.1038/mp.2017.77>

41. Yu, D., Sul, J. H., Tsetsos, F., Nawaz, M. S., Huang, A. Y., Zelaya, I., Illmann, C., Osiecki, L., Darrow, S. M., Hirschtritt, M. E., Greenberg, E., Muller-Vahl, K. R., Stuhmann, M., Dion, Y., Rouleau, G., Aschauer, H., Stamenkovic, M., Schlögelhofer, M., Sandor, P., ... Tourette Association of America International Consortium for Genetics, the Gilles de la Tourette GWAS Replication Initiative, the Tourette International Collaborative Genetics Study, and the Psychiatric Genomics Consortium Tourette Syndrome Working Group. (2019). Interrogating the genetic determinants of Tourette's syndrome and other tic disorders through genome-wide association studies. *The American Journal of Psychiatry*, 176(3), 217–227. <https://doi.org/10.1176/appi.ajp.2018.18070857>
42. International Obsessive Compulsive Disorder Foundation Genetics Collaborative (IOCDF-GC) and OCD Collaborative Genetics Association Studies (OC GAS). (2018). Revealing the complex genetic architecture of obsessive-compulsive disorder using meta-analysis. *Molecular Psychiatry*, 23(5), 1181–1188. <https://doi.org/10.1038/mp.2017.154>
43. Kranzler, H. R., Zhou, H., Kember, R. L., Vickers Smith, R., Justice, A. C., Damrauer, S., Tsao, P. S., Klarin, D., Baras, A., Reid, J., Overton, J., Rader, D. J., Cheng, Z., Tate, J. P., Becker, W. C., Concato, J., Xu, K., Polimanti, R., Zhao, H., & Gelernter, J. (2019). Genome-wide association study of alcohol consumption and use disorder in 274,424 individuals from multiple populations. *Nature Communications*, 10(1), 1499. <https://doi.org/10.1038/s41467-019-09480-8>
44. Gelernter, J., Sun, N., Polimanti, R., Pietrzak, R. H., Levey, D. F., Lu, Q., Hu, Y., Li, B., Radhakrishnan, K., Aslan, M., Cheung, K.-H., Li, Y., Rajeevan, N., Sayward, F., Harrington, K., Chen, Q., Cho, K., Honerlaw, J., Pyarajan, S., ... Million Veteran Program. (2019). Genome-wide association study of maximum habitual alcohol intake in >140,000 u. S. European and african american veterans yields novel risk loci. *Biological Psychiatry*, 86(5), 365–376. <https://doi.org/10.1016/j.biopsych.2019.03.984>
45. Levey, D. F., Stein, M. B., Wendt, F. R., Pathak, G. A., Zhou, H., Aslan, M., Quaden, R., Harrington, K. M., Nuñez, Y. Z., Overstreet, C., Radhakrishnan, K., Sanacora, G., McIntosh, A. M., Shi, J., Shringarpure, S. S., 23andMe Research Team, Million Veteran Program, Concato, J., Polimanti, R., & Gelernter, J. (2021). Bi-ancestral depression GWAS in the Million Veteran Program and meta-analysis in >1.2 million individuals highlight new therapeutic directions. *Nature Neuroscience*, 24(7), 954–963. <https://doi.org/10.1038/s41593-021-00860-2>
46. Zhou, H., Rentsch, C. T., Cheng, Z., Kember, R. L., Nunez, Y. Z., Sherva, R. M., Tate, J. P., Dao, C., Xu, K., Polimanti, R., Farrer, L. A., Justice, A. C., Kranzler, H. R., Gelernter, J., & Veterans Affairs Million Veteran Program. (2020). Association of *oprm1* functional coding variant with opioid use disorder: A genome-wide association study. *JAMA Psychiatry*, 77(10), 1072. <https://doi.org/10.1001/jamapsychiatry.2020.1206>
47. Gelernter, J., Sun, N., Polimanti, R., Pietrzak, R., Levey, D. F., Bryois, J., Lu, Q., Hu, Y., Li, B., Radhakrishnan, K., Aslan, M., Cheung, K.-H., Li, Y., Rajeevan, N., Sayward, F., Harrington, K., Chen, Q., Cho, K., Pyarajan, S., ... Department of Veterans Affairs Cooperative Studies Program (#575B) and Million Veteran Program. (2019). Genome-wide association study of post-traumatic stress disorder reexperiencing symptoms in >165,000 US veterans. *Nature Neuroscience*, 22(9), 1394–1401. <https://doi.org/10.1038/s41593-019-0447-7>
48. Levey, D. F., Gelernter, J., Polimanti, R., Zhou, H., Cheng, Z., Aslan, M., Quaden, R., Concato, J., Radhakrishnan, K., Bryois, J., Sullivan, P. F., Million Veteran Program, & Stein, M. B. (2020). Reproducible genetic risk loci for anxiety: Results from ~200,000 participants in the million veteran program. *The American Journal of Psychiatry*, 177(3), 223–232. <https://doi.org/10.1176/appi.ajp.2019.19030256>

49. Sullivan, P. F. (2010). The psychiatric GWAS consortium: Big science comes to psychiatry. *Neuron*, 68(2), 182–186. <https://doi.org/10.1016/j.neuron.2010.10.003>
50. 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. *Journal of Clinical Epidemiology*, 70, 214–223. <https://doi.org/10.1016/j.jclinepi.2015.09.016>
51. Schizophrenia Working Group of the Psychiatric Genomics Consortium, Bulik-Sullivan, B. K., Loh, P.-R., Finucane, H. K., Ripke, S., Yang, J., Patterson, N., Daly, M. J., Price, A. L., & Neale, B. M. (2015). LD Score regression distinguishes confounding from polygenicity in genome-wide association studies. *Nature Genetics*, 47(3), 291–295. <https://doi.org/10.1038/ng.3211>
52. Zhou, W., Nielsen, J. B., Fritsche, L. G., Dey, R., Gabrielsen, M. E., Wolford, B. N., LeFaive, J., VandeHaar, P., Gagliano, S. A., Gifford, A., Bastarache, L. A., Wei, W.-Q., Denny, J. C., Lin, M., Hveem, K., Kang, H. M., Abecasis, G. R., Willer, C. J., & Lee, S. (2018). Efficiently controlling for case-control imbalance and sample relatedness in large-scale genetic association studies. *Nature Genetics*, 50(9), 1335–1341. <https://doi.org/10.1038/s41588-018-0184-y>
53. Chang, C. C., Chow, C. C., Tellier, L. C., Vattikuti, S., Purcell, S. M., & Lee, J. J. (2015). Second-generation PLINK: Rising to the challenge of larger and richer datasets. *GigaScience*, 4(1), 7. <https://doi.org/10.1186/s13742-015-0047-8>
54. Choi, S. W., & O'Reilly, P. F. (2019). PRSice-2: Polygenic Risk Score software for biobank-scale data. *GigaScience*, 8(7), giz082. <https://doi.org/10.1093/gigascience/giz082>
55. Choi, S. W., Heng Mak, T. S., & O'Reilly, P. F. (2018). *A guide to performing Polygenic Risk Score analyses* [Preprint]. Genomics. <https://doi.org/10.1101/416545>
56. Galinsky, K. J., Bhatia, G., Loh, P.-R., Georgiev, S., Mukherjee, S., Patterson, N. J., & Price, A. L. (2016). Fast principal-component analysis reveals convergent evolution of *adh1b* in Europe and East Asia. *American Journal of Human Genetics*, 98(3), 456–472. <https://doi.org/10.1016/j.ajhg.2015.12.022>
57. Purcell, S., Neale, B., Todd-Brown, K., Thomas, L., Ferreira, M. A. R., Bender, D., Maller, J., Sklar, P., de Bakker, P. I. W., Daly, M. J., & Sham, P. C. (2007). PLINK: A tool set for whole-genome association and population-based linkage analyses. *American Journal of Human Genetics*, 81(3), 559–575. <https://doi.org/10.1086/519795>
58. International HapMap Consortium. (2003). The international hapmap project. *Nature*, 426(6968), 789–796. <https://doi.org/10.1038/nature02168>
59. Howe, L. J., Battram, T., Morris, T. T., Hartwig, F. P., Hemani, G., Davies, N. M., & Smith, G. D. (2021). Assortative mating and within-spouse pair comparisons. *PLOS Genetics*, 17(11), e1009883. <https://doi.org/10.1371/journal.pgen.1009883>
60. Manichaikul, A., Mychaleckyj, J. C., Rich, S. S., Daly, K., Sale, M., & Chen, W.-M. (2010). Robust relationship inference in genome-wide association studies. *Bioinformatics*, 26(22), 2867–2873. <https://doi.org/10.1093/bioinformatics/btq559>
61. Yang, J., Ferreira, T., Morris, A. P., Medland, S. E., Genetic Investigation of ANthropometric Traits (GIANT) Consortium, DIAbetes Genetics Replication And Meta-analysis (DIAGRAM) Consortium, Madden, P. A. F., Heath, A. C., Martin, N. G., Montgomery, G. W., Weedon, M. N., Loos, R. J., Frayling, T. M., McCarthy, M. I., Hirschhorn, J. N., Goddard, M. E., & Visscher, P. M. (2012). Conditional and joint multiple-SNP analysis of GWAS summary statistics identifies additional variants influencing complex traits. *Nature Genetics*, 44(4), 369–375, S1-3. <https://doi.org/10.1038/ng.2213>
62. Townsend, P., Phillimore, P., & Beattie, A. (1988). *Health and deprivation: Inequality and the North*. Croom Helm

63. Torvik, F. A., Eilertsen, E. M., Hannigan, L. J., Cheesman, R., Howe, L. J., Magnus, P., Reichborn-Kjennerud, T., Andreassen, O. A., Njølstad, P. R., Havdahl, A., & Ystrom, E. (2022). Modeling assortative mating and genetic similarities between partners, siblings, and in-laws. *Nature Communications*, 13(1), 1108. <https://doi.org/10.1038/s41467-022-28774-y>
64. Nievergelt, C. M., Maihofer, A. X., Klengel, T., Atkinson, E. G., Chen, C.-Y., Choi, K. W., Coleman, J. R. I., Dalvie, S., Duncan, L. E., Gelernter, J., Levey, D. F., Logue, M. W., Polimanti, R., Provost, A. C., Ratanatharathorn, A., Stein, M. B., Torres, K., Aiello, A. E., Almli, L. M., ... Koenen, K. C. (2019). International meta-analysis of PTSD genome-wide association studies identifies sex- and ancestry-specific genetic risk loci. *Nature Communications*, 10(1), 4558. <https://doi.org/10.1038/s41467-019-12576-w>
65. Agrawal, A., Heath, A. C., Grant, J. D., Pergadia, M. L., Statham, D. J., Bucholz, K. K., Martin, N. G., & Madden, P. A. F. (2006). Assortative Mating for Cigarette Smoking and for Alcohol Consumption in Female Australian Twins and their Spouses. *Behavior Genetics*, 36(4), 553–566. <https://doi.org/10.1007/s10519-006-9081-8>
66. Grant, J. D., Heath, A. C., Bucholz, K. K., Madden, P. A. F., Agrawal, A., Statham, D. J., & Martin, N. G. (2007). Spousal concordance for alcohol dependence: Evidence for assortative mating or spousal interaction effects? *Alcoholism: Clinical and Experimental Research*, 31(5), 717–728. <https://doi.org/10.1111/j.1530-0277.2007.00356.x>
67. Reynolds, C. A., Barlow, T., & Pedersen, N. L. (2006). Alcohol, tobacco and caffeine use: Spouse similarity processes. *Behavior Genetics*, 36(2), 201–215. <https://doi.org/10.1007/s10519-005-9026-7>
68. Rhule-Louie, D. M., & McMahon, R. J. (2007). Problem behavior and romantic relationships: Assortative mating, behavior contagion, and desistance. *Clinical Child and Family Psychology Review*, 10(1), 53–100. <https://doi.org/10.1007/s10567-006-0016-y>
69. Segrin, C. (2004). Concordance on negative emotion in close relationships: Transmission of emotion or assortative mating? *Journal of Social and Clinical Psychology*, 23(6), 836–856. <https://doi.org/10.1521/jscp.23.6.836.54802>
70. Manthey, J., Freeman, T. P., Kilian, C., López-Pelayo, H., & Rehm, J. (2021). Public health monitoring of cannabis use in Europe: Prevalence of use, cannabis potency, and treatment rates. *The Lancet Regional Health. Europe*, 10, 100227. <https://doi.org/10.1016/j.lanepe.2021.100227>
71. Johnson, E. C., Demontis, D., Thorgeirsson, T. E., Walters, R. K., Polimanti, R., Hatoum, A. S., Sanchez-Roige, S., Paul, S. E., Wendt, F. R., Clarke, T.-K., Lai, D., Reginsson, G. W., Zhou, H., He, J., Baranger, D. A. A., Gudbjartsson, D. F., Wedow, R., Adkins, D. E., Adkins, A. E., ... Agrawal, A. (2020). A large-scale genome-wide association study meta-analysis of cannabis use disorder. *The Lancet. Psychiatry*, 7(12), 1032–1045. [https://doi.org/10.1016/S2215-0366\(20\)30339-4](https://doi.org/10.1016/S2215-0366(20)30339-4)
72. Kranzler, H. R., Zhou, H., Kember, R. L., Vickers Smith, R., Justice, A. C., Damrauer, S., Tsao, P. S., Klarin, D., Baras, A., Reid, J., Overton, J., Rader, D. J., Cheng, Z., Tate, J. P., Becker, W. C., Concato, J., Xu, K., Polimanti, R., Zhao, H., & Gelernter, J. (2019). Genome-wide association study of alcohol consumption and use disorder in 274,424 individuals from multiple populations. *Nature Communications*, 10(1), 1499. <https://doi.org/10.1038/s41467-019-09480-8>
73. Polimanti, R., Walters, R. K., Johnson, E. C., McClintick, J. N., Adkins, A. E., Adkins, D. E., Bacanu, S.-A., Bierut, L. J., Bigdeli, T. B., Brown, S., Bucholz, K. K., Copeland, W. E., Costello, E. J., Degenhardt, L., Farrer, L. A., Foroud, T. M., Fox, L., Goate, A. M., Gruzca, R., ... Gelernter, J. (2020). Leveraging genome-wide data to investigate differences between opioid use vs.

- Opioid dependence in 41,176 individuals from the Psychiatric Genomics Consortium. *Molecular Psychiatry*, 25(8), 1673–1687. <https://doi.org/10.1038/s41380-020-0677-9>
74. Xue, A., Jiang, L., Zhu, Z., Wray, N. R., Visscher, P. M., Zeng, J., & Yang, J. (2021). Genome-wide analyses of behavioural traits are subject to bias by misreports and longitudinal changes. *Nature Communications*, 12(1), 20211. <https://doi.org/10.1038/s41467-020-20237-6>
75. Grant, J. D., Heath, A. C., Bucholz, K. K., Madden, P. A. F., Agrawal, A., Statham, D. J., & Martin, N. G. (2007). Spousal concordance for alcohol dependence: Evidence for assortative mating or spousal interaction effects? *Alcoholism: Clinical and Experimental Research*, 31(5), 717–728. <https://doi.org/10.1111/j.1530-0277.2007.00356.x>
76. Deak, J. D., Levey, D. F., Wendt, F. R., Zhou, H., Galimberti, M., Kranzler, H. R., Gaziano, J. M., Stein, M. B., Polimanti, R., The Million Veteran Program, & Gelernter, J. (2022). *Genome-wide investigation of maximum habitual alcohol intake (Maxalc) in 247,755 European and African Ancestry U.S. Veterans informs the relationship between habitual alcohol consumption and alcohol use disorder* [Preprint]. Genetic and Genomic Medicine. <https://doi.org/10.1101/2022.05.02.22274580>
77. Gelernter, J., & Polimanti, R. (2021). Genetics of substance use disorders in the era of big data. *Nature Reviews. Genetics*, 22(11), 712–729. <https://doi.org/10.1038/s41576-021-00377->
78. Davis, K. A. S., Coleman, J. R. I., Adams, M., Allen, N., Breen, G., Cullen, B., Dickens, C., Fox, E., Graham, N., Holliday, J., Howard, L. M., John, A., Lee, W., McCabe, R., McIntosh, A., Pearsall, R., Smith, D. J., Sudlow, C., Ward, J., ... Hotopf, M. (2020). Mental health in UK Biobank - development, implementation and results from an online questionnaire completed by 157 366 participants: A reanalysis. *BJPsych Open*, 6(2), e18. <https://doi.org/10.1192/bjo.2019.100>
79. Levey, D. F., Stein, M. B., Wendt, F. R., Pathak, G. A., Zhou, H., Aslan, M., Quaden, R., Harrington, K. M., Nuñez, Y. Z., Overstreet, C., Radhakrishnan, K., Sanacora, G., McIntosh, A. M., Shi, J., Shringarpure, S. S., 23andMe Research Team, Million Veteran Program, Concato, J., Polimanti, R., & Gelernter, J. (2021). Bi-ancestral depression GWAS in the Million Veteran Program and meta-analysis in >1.2 million individuals highlight new therapeutic directions. *Nature Neuroscience*, 24(7), 954–963. <https://doi.org/10.1038/s41593-021-00860-2>
80. Nievergelt, C. M., Maihofer, A. X., Klengel, T., Atkinson, E. G., Chen, C.-Y., Choi, K. W., Coleman, J. R. I., Dalvie, S., Duncan, L. E., Gelernter, J., Levey, D. F., Logue, M. W., Polimanti, R., Provost, A. C., Ratanatharathorn, A., Stein, M. B., Torres, K., Aiello, A. E., Almli, L. M., ... Koenen, K. C. (2019). International meta-analysis of PTSD genome-wide association studies identifies sex- and ancestry-specific genetic risk loci. *Nature Communications*, 10(1), 4558. <https://doi.org/10.1038/s41467-019-12576-w>
81. Lam, M., Awasthi, S., Watson, H. J., Goldstein, J., Panagiotaropoulou, G., Trubetskoy, V., Karlsson, R., Frei, O., Fan, C.-C., De Witte, W., Mota, N. R., Mullins, N., Brügger, K., Lee, S. H., Wray, N. R., Skarabis, N., Huang, H., Neale, B., Daly, M. J., ... Ripke, S. (2020). Ricopili: Rapid imputation for consortias pipeline. *Bioinformatics (Oxford, England)*, 36(3), 930–933. <https://doi.org/10.1093/bioinformatics/btz633>
82. Wendt, F. R., Pathak, G. A., Vahey, J., Qin, X., Koller, D., Cabrera-Mendoza, B., Haeny, A., Harrington, K. M., Rajeevan, N., Duong, L. M., Levey, D. F., De Angelis, F., De Lillo, A., Bigdeli, T. B., Pyarajan, S., VA Million Veteran Program, Gaziano, J. M., Gelernter, J., Aslan, M., ... Department of Veteran Affairs Cooperative Study Program (#2006). (2022). *Modeling the longitudinal changes of ancestry diversity in the Million Veteran Program* [Preprint]. Genomics. <https://doi.org/10.1101/2022.01.24.477583>

83. Colbert, S. M. C., Wendt, F. R., Pathak, G. A., Helmer, D. A., Hauser, E. R., Keller, M. C., Polimanti, R., & Johnson, E. C. (2022). *Declining autozygosity over time: An exploration in over 1 million individuals from three diverse cohorts* [Preprint]. Genomics. <https://doi.org/10.1101/2022.10.13.512166>
84. Munafò, M. R., Tilling, K., Taylor, A. E., Evans, D. M., & Davey Smith, G. (2018). Collider scope: When selection bias can substantially influence observed associations. *International Journal of Epidemiology*, 47(1), 226–235. <https://doi.org/10.1093/ije/dyx206>