

1  
2  
3  
4  
5  
6  
7  
8  
9  
10  
11  
12  
13  
14  
15  
16  
17  
18  
19  
20  
21  
22  
23  
24  
25  
26  
27  
28  
29  
30  
31  
32  
33  
34  
35  
36  
37

## Genome-wide Association Study of Traumatic Brain Injury in U.S. Military Veterans Enrolled in the VA Million Veteran Program

Victoria C. Merritt<sup>1,2,3,\*</sup>, Adam X. Maihofer<sup>1,2</sup>, Marianna Gasperi<sup>1,2,3</sup>, Elizabeth Ketema<sup>1,2</sup>, Catherine Chanfreau-Coffinier<sup>4</sup>, Murray B. Stein<sup>1,2,5</sup>, Matthew S. Panizzon<sup>1,2,6</sup>, Richard L. Hauger<sup>1,2,3,6</sup>, Mark W. Logue<sup>7,8,9,10</sup>, Lisa Delano-Wood<sup>1,3,2</sup>, & Caroline M. Nievergelt<sup>1,2,3</sup>,  
on behalf of the VA Million Veteran Program

<sup>1</sup>VA San Diego Healthcare System (VASDHS), San Diego, CA, United States. <sup>2</sup>Department of Psychiatry, University of California San Diego, La Jolla, CA, United States. <sup>3</sup>Center of Excellence for Stress and Mental Health, VASDHS, San Diego, CA, United States. <sup>4</sup>VA Informatics and Computing Infrastructure (VINCI), VA Salt Lake City Health Care System, Salt Lake City, UT, United States. <sup>5</sup>School of Public Health, University of California San Diego, La Jolla, CA, United States. <sup>6</sup>Center for Behavior Genetics of Aging, University of California, San Diego, La Jolla, CA, United States. <sup>7</sup>National Center for PTSD, Behavioral Sciences Division, VA Boston Healthcare System, Boston, MA, United States. <sup>8</sup>Boston University School of Medicine, Department of Psychiatry, Boston, MA, United States. <sup>9</sup>Boston University School of Medicine, Biomedical Genetics, Boston, MA, United States. <sup>10</sup>Boston University School of Public Health, Department of Biostatistics, Boston, MA, United States.

\*Address correspondence to:  
Victoria C. Merritt, Ph.D.  
VA San Diego Healthcare System (151B)  
3350 La Jolla Village Drive  
San Diego, CA 92161  
Phone: (858) 552-8585 (x2670)  
Email: [victoria.merritt@va.gov](mailto:victoria.merritt@va.gov)

## Abstract

38  
39  
40 Large-scale genetic studies of traumatic brain injury (TBI) are lacking; thus, our  
41 understanding of the influence of genetic factors on TBI risk and recovery is incomplete.  
42 This study aimed to conduct a genome-wide association study (GWAS) of TBI in VA  
43 Million Veteran Program enrollees. Participants included a multi-ancestry cohort  
44 (European, African, and Hispanic ancestries; N=304,485; 111,494 TBI cases, 192,991  
45 controls). TBI was assessed using MVP survey data and ICD codes from the Veterans  
46 Health Administration's electronic health record. GWAS was performed using logistic  
47 regression in PLINK, and meta-analyzed in METAL. FUMA was used for post-GWAS  
48 analysis. Genomic structural equation modeling (gSEM) was conducted to investigate  
49 underlying genetic associations with TBI, and bivariate MiXeR was used to estimate  
50 phenotype specific and shared polygenicity. SNP-based heritability was 0.060  
51 (SE=0.004,  $p=7.83 \times 10^{-66}$ ). GWAS analysis identified 15 genome-wide significant (GWS)  
52 loci at  $p < 5 \times 10^{-8}$ . Gene-based analyses revealed 14 gene-wide significant genes,  
53 including *NCAM1*, *APOE*, *FTO*, and *FOXP2*. Gene tissue expression analysis identified  
54 the brain as significantly enriched, particularly in the frontal cortex, anterior cingulate  
55 cortex, and nucleus accumbens. Genetic correlations with TBI were significant for risk-  
56 taking behaviors and psychiatric disorders, but generally not significant for the  
57 neurocognitive variables investigated. gSEM analysis revealed stronger associations  
58 with risk-taking traits than with psychiatric traits. Finally, the genetic architecture of TBI  
59 was similar to polygenic psychiatric disorders. Neurodegenerative disorders including  
60 Alzheimer's and Parkinson's disease showed much less polygenicity, however, the  
61 proportion of shared variance with TBI was high. This first well-powered GWAS of TBI  
62 identified 15 loci including genes relevant to TBI biology, and showed that TBI is a  
63 heritable trait with comparable genetic architecture and high genetic correlation with  
64 psychiatric traits. Our findings set the stage for future TBI GWASs that focus on injury  
65 severity and diversity and chronicity of symptom sequelae.

66  
67

68

## 69 Introduction

70 Military personnel serving in Iraq and Afghanistan have sustained traumatic brain  
71 injuries (TBIs) at a higher frequency than prior conflicts<sup>1</sup>, prompting widespread interest  
72 in the assessment and management of TBI and its associated sequelae. Within military  
73 populations, the overwhelming majority of TBIs are considered “closed” head injuries;  
74 among these, approximately 80% are classified as mild in severity<sup>2</sup>. Although  
75 designated as “mild,” it has become well established that a host of clinically meaningful  
76 sequelae (i.e., neurobehavioral symptoms, psychiatric distress, and problems with  
77 thinking and cognition) develop acutely post-injury<sup>3-5</sup>. Despite the expectation of full  
78 recovery within days to weeks following mild TBI (mTBI)<sup>6</sup>, these difficulties can persist  
79 for many years for some individuals<sup>7-10</sup>. Moreover, long-term outcome studies, even in  
80 mTBI, have demonstrated an association between TBI and increased risk of dementia,  
81 including Alzheimer’s disease (AD), as well as chronic symptoms of posttraumatic  
82 stress disorder (PTSD) and depression<sup>8,11</sup>.

83 Accumulating evidence also suggests that Veterans with a history of TBI—  
84 especially when combined with psychiatric comorbidities—experience poor health-  
85 related quality of life<sup>12-14</sup>, increased difficulties with psychosocial functioning<sup>15</sup>, and  
86 greater global disability<sup>16</sup>. The costs associated with TBI assessment and management  
87 are also substantial, with one study estimating that the median annual healthcare cost  
88 per patient was approximately four times higher for Veterans with a history of TBI  
89 compared to Veterans without a history of TBI<sup>17</sup>. It is thus crucial to deepen our  
90 understanding of the risk factors associated with *sustaining* a TBI as well as the  
91 underlying factors that contribute to poor outcome and recovery *following* TBI.

92 It has been hypothesized that there are several mechanisms by which genetics  
93 may influence risk for and response to neurotrauma, including pre-injury risk factors for  
94 sustaining a TBI (e.g., risk-taking behaviors, impulsivity, substance use); the acute  
95 response to neurotrauma (i.e., extent of injury); the repair or recovery processes  
96 following injury (i.e., capacity for recovery); and pre- and post-injury cognitive capacity  
97 and cognitive reserve<sup>18</sup>. However, large-scale genetic studies of TBI are lacking and  
98 thus our understanding of the influence of genetic factors on TBI risk and recovery  
99 remains incomplete. Examining the role of genetics in TBI risk and recovery may offer  
100 novel mechanistic insights and lead to important discoveries that can inform preventive  
101 efforts and innovative treatments for affected individuals.

102 Among the published studies that have examined genetic associations with TBI  
103 risk and recovery, findings are considerably disparate, likely due to varied  
104 methodologies as well as inadequate sample sizes and therefore low power to detect  
105 meaningful differences in TBI samples<sup>18-23</sup>. Furthermore, existing studies have largely  
106 adopted a “candidate gene” approach, focusing on a specific gene of interest—most  
107 commonly, the apolipoprotein E (*APOE*) gene<sup>24-28</sup>, though it is widely believed that  
108 genetic predisposition to complex traits or conditions is highly polygenic. Thus, not only  
109 are adequately powered studies needed to better understand the influence of genetic  
110 markers on TBI risk and recovery but, given the likelihood that multiple genes may be  
111 involved, critical next steps are to apply the concept of polygenic risk to TBI and  
112 capitalize on recent advances in genetics and genomics research—namely, the  
113 implementation of genome-wide association studies.

114           Until recently, it had been very challenging to amass a sufficient number of cases  
115 to adequately carry out a genome-wide association study (GWAS) in TBI. However, the  
116 VA Million Veteran Program, a nationwide research initiative that examines how genes  
117 influence health and behaviors<sup>29</sup>, affords a unique opportunity to explore a large cohort  
118 of U.S. military Veterans with and without a history of TBI. In this GWAS of TBI using  
119 Veterans Health Administration’s (VHA) electronic health record data, we assessed TBI  
120 history among military Veterans. We examined genome-wide significant (GWS) variants  
121 associated with TBI in a multi-ancestry sample, and separately examined Veterans of  
122 European ancestry (EA), African ancestry (AA), and Hispanic ancestry (HA). We  
123 examined replication of results with data from another biobank, FinnGen (which  
124 included individuals of EA only), and assessed SNP-based heritability. We then  
125 examined genetic correlations with various health domains, including risk-taking  
126 behaviors, psychiatric disorders, neurocognition, and brain morphometrics, and  
127 evaluated phenotype-specific and shared polygenicity among these traits.

128

## 129 **Methods**

130

### 131 *Procedures*

132           This research was conducted as part of project “MVP026” within the Million  
133 Veteran Program (MVP). The MVP, sponsored by the Department of Veterans Affairs  
134 Office of Research and Development, is a nationwide initiative offering military Veterans  
135 the opportunity to participate in research that seeks to better understand how genetic  
136 factors influence health. Upon consenting to participate in MVP, participants are asked  
137 to complete two comprehensive surveys—the “Baseline Survey” and “Lifestyle  
138 Survey”—and provide a blood sample for genetic analysis. Participants also consent to  
139 provide MVP investigators access to their VHA electronic health record (EHR). MVP  
140 study procedures have been previously described by Gaziano and colleagues<sup>29</sup>. The  
141 overarching MVP project was approved by the VA Central Institutional Review Board  
142 (IRB) in 2010 and this specific project (“MVP026”) received VA Central IRB approval in  
143 2019.

144

### 145 *Participants*

146           Veterans (N=304,485; 111,494 TBI cases; 192,991 controls) were included in  
147 this study if they had both phenotype and genotyping data available (described below in  
148 more detail). The EA cohort included 249,785 Veterans (85,613 cases; 164,172  
149 controls); the AA cohort included 35,470 Veterans (15,714 cases; 19,756 controls); and  
150 the HA cohort included 19,230 Veterans (10,167 cases; 9,063 controls). See Table 1 for  
151 overall (multi-ancestry) sample characteristics and Supplemental Table 1 for EA, AA,  
152 and HA sample characteristics.

153

### 154 *MVP Data Sources*

155           The following data sources (from MVP v20\_1 data release) were used to gather  
156 information pertaining to TBI history: MVP Baseline Survey, MVP Lifestyle Survey, and  
157 *International Classification of Diseases, 9<sup>th</sup> Revision Clinical Modification* (ICD-9) and  
158 *International Classification of Diseases, 10<sup>th</sup> Revision Clinical Modification* (ICD-10)  
159 codes gleaned from Veterans’ EHRs.

160 MVP Baseline Survey: The MVP Baseline Survey gathers information pertaining  
161 to sociodemographics, military history, health status and habits, and medical history and  
162 health care usage. Within the “Medical History and Health Care Usage” section of the  
163 survey, Veterans are provided a list of health conditions and are asked to self-report  
164 whether they have been diagnosed with any of the health conditions. Two questions  
165 pertaining to TBI were extracted from the MVP Baseline Survey and utilized in our TBI  
166 phenotype: “Concussion or loss of consciousness” and “Traumatic brain injury”  
167 (hereafter labeled “BL-CC” and “BL-TBI,” respectively).

168 MVP Lifestyle Survey: The MVP Lifestyle Survey gathers information pertaining  
169 to psychosocial health and activities, men’s and women’s health, exercise and nutrition,  
170 well-being, and military and environmental experiences. Within the “Military and  
171 Environmental Experiences” section of the survey, Veterans are asked specifically  
172 about deployment-related experiences. If a Veteran indicates that they have been  
173 deployed and served in a combat or war zone, they are asked additional questions to  
174 assess for history of deployment-related TBI. These questions, based on the Brief  
175 Traumatic Brain Injury Screen<sup>30</sup>, were designed to assess mechanism of injury, signs or  
176 symptoms associated with TBI, and current symptoms. Specific questions were as  
177 follows: (1) “Did you have any injury(ies) during your deployment from any of the  
178 following?”, with response options including “fragment,” “bullet,” “vehicular,” “fall,”  
179 “blast,” “other,” and “none”; (2) “Did any injury received while you were deployed result  
180 in any of the following?”, with response options including “being dazed, confused or  
181 seeing stars,” “not remembering the injury,” “losing consciousness [knocked out] for less  
182 than a minute,” “losing consciousness for 1-20 minutes,” “losing consciousness for  
183 longer than 20 minutes,” “having any symptoms of concussion afterward (such as  
184 headache, dizziness, irritability, etc.),” “head injury,” and “none of the above”; and (3)  
185 “Are you currently experiencing any of the following problems that you think might be  
186 related to a possible head injury or concussion?”, with response options including  
187 “headache,” “dizziness,” “memory problems,” “balance problems,” “ringing in the ears,”  
188 “irritability,” “sleep problems,” and “other”. Responses to the second question pertaining  
189 to signs and symptoms associated with TBI (hereafter labeled “LS-Q39”) were extracted  
190 from the MVP Lifestyle Survey and utilized in our TBI phenotype.

191 ICD-9/ICD-10 Codes: ICD-9/10 codes pertaining to TBI were extracted from the  
192 EHR. TBI-specific ICD-9/10 codes were established using surveillance case definitions  
193 for TBI set forth by the Armed Forces Health Surveillance Branch of the Military Health  
194 System.<sup>31</sup> The list of ICD-9 and ICD-10 codes utilized in this study is provided in  
195 Supplemental Tables 2 and 3, respectively.

### 196 197 *TBI Phenotyping*

198 TBI Case Definition: To be considered a TBI case, there must have been  
199 evidence of TBI as determined by: (1) endorsement of BL-CC *or* (2) endorsement of BL-  
200 TBI *or* (3) endorsement of at least one sign/symptom associated with TBI on LS-Q39 *or*  
201 (4) at least one inpatient or outpatient ICD-9/10 code for TBI. In other words, Veterans  
202 with evidence of *any one or more* of these criteria were classified as a TBI case.

203 TBI Control Definition: To be considered a TBI control, there must have been no  
204 evidence of TBI as determined by: (1) no endorsement of BL-CC *and* (2) no  
205 endorsement of BL-TBI *and* (3) no endorsement of any signs/symptoms associated with



206 TBI on LS-Q39 *and* (4) at least one inpatient or outpatient ICD-9/10 diagnosis in the  
207 EHR but no ICD-9/10 codes for TBI at any point in the EHR. In other words, Veterans  
208 must have completed the MVP Baseline Survey and MVP Lifestyle Survey and not  
209 reported any event consistent with TBI on either survey. Additionally, there must have  
210 been evidence of VHA usage as determined by the presence of an inpatient or  
211 outpatient ICD-9/10 diagnosis, but no TBI ICD codes in the EHR.

212

### 213 *MVP Genotyping Procedures*

214 The MVP Bioinformatics Core completed the data cleaning and processing for all  
215 genetics data. All genetic analyses for this study utilized MVP Release 4 data. MVP  
216 genotyping procedures have been described in detail elsewhere<sup>29</sup>; briefly, the  
217 “Affymetrix Axiom Biobank Array” was used for genotyping and standard quality control  
218 procedures were applied. Phasing occurred using Eagle v2.4 and imputation occurred  
219 through Minimac v4 using the Haplotype Reference Consortium panel<sup>32,33</sup>.

220

### 221 *Ancestry Assignment & Principal Components*

222 For this study, the harmonized ancestry and race/ethnicity (HARE) method<sup>34</sup> was  
223 used to inform genetic ancestry for GWAS; ancestry groups included EA, AA, and HA.  
224 KING<sup>35</sup> was used to estimate relatedness (defined as kinship coefficient > 0.0884);  
225 related individuals were removed from analysis by prioritizing cases and removing  
226 controls. When related pairs were both cases (or both controls), one individual was  
227 removed at random. FlashPCA2<sup>36</sup> was used to compute principal components (PCs)  
228 within each ancestry group.

229

### 230 *GWAS and Meta-Analysis*

231 Separate GWASs were conducted for each ancestry group (EA, AA, and HA).  
232 Logistic regression analyses were performed to evaluate associations between  
233 genotype dosage data and TBI case-control status using PLINK<sup>37</sup>, including the first 10  
234 PCs as covariates. The standard genome-wide significance threshold was used ( $p < 5 \times$   
235  $10^{-8}$ ). SNPs with MAF < 1% or an imputation information score < 0.6 were excluded  
236 from the results. Multi-ancestry meta-analysis was conducted in METAL<sup>38</sup> using SNPs  
237 present in all datasets. To produce regional visualizations of GWS loci, LocusZoom  
238 1.4<sup>39</sup> was used. The 1000 Genomes phase 3 reference data (1KGPp3)<sup>32</sup> was utilized to  
239 calculate LD, relying on ancestry-specific reference genotypes.

240

### 241 *Replication*

242 GWS SNPs in the EA MVP cohort were compared to the largest available  
243 external TBI data repository—FinnGen (<https://www.finnngen.fi/en>). Within FinnGen  
244 (Release 7), two TBI phenotypes were examined, one for “concussion” (15,787 cases,  
245 184,565 controls) and one for “severe TBI” (4,927 cases, 304,227 controls).

246

### 247 *Functional Mapping and Annotation*

248 Functional Mapping and Annotation (FUMA) software<sup>40</sup> was used for functional  
249 mapping and annotation of GWAS results (default settings were used unless otherwise  
250 specified), with annotations derived from human genome assembly GRCh37 (hg19). To  
251 delineate independent genomic risk loci (defined by  $r^2 > 0.6$ ) and variants in LD with

252 lead SNPs, the SNP2Gene module was applied using ancestry-appropriate reference  
253 genotypes. To map risk loci to protein-coding genes, a 10kb window size was used.  
254 eQTL mapping was completed for significant SNP-gene pairs (FDR < 0.05) using GTEx  
255 v8 brain tissue expression data.

#### 256 *Gene-Based Results and Gene-Pathway and Tissue-Enrichment Analyses with* 257 *MAGMA*

259 Multi-Marker Analysis of GenoMic Annotation (MAGMA<sup>41</sup>) was used to conduct  
260 analyses pertaining to genes and tissue enrichment (i.e., gene-based, gene-pathway,  
261 and tissue-enrichment analyses). Regarding gene-based analyses, SNPs were mapped  
262 to 18,873 protein coding genes, and Bonferroni correction was used to establish the  
263 GWS threshold for these analyses ( $p=0.05/18,873=2.65 \times 10^{-6}$ ). To determine whether  
264 specific biological pathways were associated with TBI, 15,485 curated gene sets and  
265 GO terms (from MsigDB) were used in a pathway analysis; Bonferroni correction was  
266 again applied ( $p=0.05/15,485=3.23 \times 10^{-6}$ ). Finally, tissue-enrichment analyses were  
267 carried out using GTEx v8 RNA-seq and BrainSpan RNA-seq expression data.

#### 268 *SNP-Based Heritability Estimation and Genetic Correlations*

269 SNP-based heritability (SNP- $h^2$ ) was computed using linkage disequilibrium  
270 score regression (LDSC)<sup>42</sup> in the EA cohort. The 1KGp3 EA reference sample was  
271 used to calculate input LD scores. Both observed scale and liability scale SNP- $h^2$  are  
272 reported; liability scale SNP- $h^2$  was computed using a population prevalence of 2% and  
273 20%. Genetic correlations ( $r_g$ ) were also computed using LDSC.

#### 274 *Univariate and Bivariate Gaussian Mixer Model (MiXeR) Analysis*

275 MiXeR v1.3<sup>43</sup> was used to estimate polygenicity and discoverability of causal  
276 variants for TBI. Polygenicity estimates reflect the number of loci needed to explain 90%  
277 of SNP- $h^2$ . Bivariate MiXeR<sup>44</sup> was utilized to estimate phenotype specific and shared  
278 polygenicity. Akaike information criteria (AIC) values were used to evaluate model fit.

#### 281 *Genomic Structural Equation Modeling (gSEM)*

282 Genomic structural equation modeling (gSEM) was conducted using GWAS  
283 summary statistics in R<sup>45</sup> (using the genomicSEM package<sup>46</sup>). LDSC<sup>42</sup> was carried out  
284 in the EA cohort to estimate the genetic covariance matrix and the sampling covariance  
285 matrix using the 1KGp3 EA reference for the following variables: TBI, general risk  
286 tolerance (GRT), major depressive disorder (MDD), attention-deficit/hyperactivity  
287 disorder (ADHD), alcohol dependence (ALCH), posttraumatic stress disorder (PTSD),  
288 reaction time (RT), verbal numeric reasoning (VNR), neuroticism (NEUR), Parkinson's  
289 disease (PD), and Alzheimer's disease (AD). Additionally, we evaluated the following  
290 brain morphometric variables: nucleus accumbens (NCA), amygdala (AM), brainstem  
291 (BS), caudate nucleus (CN), globus pallidus (GP), putamen (PU), thalamus (TH),  
292 hippocampal volume (HV), intracranial volume (ICV), full surface area (FSA), and full  
293 average thickness (FT). SNPs within the major histocompatibility region were removed  
294 in the calculation of genetic covariance.

295 Exploratory factor analysis (EFA) was used to assess the number of latent  
296 factors among variables pertaining to TBI, risk-taking behaviors, and psychiatric  
297

298 disorders. Using the R *factanal* function, EFA was carried out on the odd chromosomes  
299 and then confirmatory factor analyses (CFAs) were carried out on the covariance matrix  
300 from the even chromosomes. A factor loading threshold of 0.35 or greater was required  
301 from the EFA for variables to be assigned to factors in the CFA. The weighted least  
302 squares estimator was used to fit the data. Model fit was evaluated using the chi-square  
303 test *p*-value, comparative fit index (CFI), AIC, and standardized root mean square  
304 residual (SRMR).

305

### 306 *Phenome-wide Association Study (PheWAS)*

307 A PheWAS for GWS SNPs from the TBI GWAS was conducted using GWAS  
308 Atlas (<https://atlas.ctglab.nl/PheWAS>)<sup>47</sup>. SNP rs-ID's were entered and traits were  
309 sorted by "Domain and P-value" and the max P-value was set to 0.05. Data were  
310 downloaded from the GWAS Atlas website and plotted in R. Bonferroni correction was  
311 applied ( $p=0.05/3,302=1.51 \times 10^{-5}$ ).

312

## 313 **Results**

314

### 315 *SNP-Based Results*

316 The multi-ancestry meta-analysis GWAS of TBI (case-control) was conducted  
317 using a sample of 304,485 Veterans (111,494 cases; 192,991 controls). A Manhattan  
318 plot of the complete results is shown in Figure 1. We identified 15 genome-wide  
319 significant (GWS) independent loci (Table 2). Regional annotation plots of GWS loci are  
320 presented in Supplemental Figure 1. The leading SNP of the most significant locus  
321 (rs1940701, locus 13) was located in *NCAM1* on chromosome 11. The leading SNP of  
322 the second most significant locus (rs729053, locus 17) was in *TCF4* on chromosome  
323 18. The leading SNP of the third most significant locus (rs429358, locus 19) was in  
324 *APOE* on chromosome 19. Notably, rs429358 is one of the two polymorphisms  
325 (alongside rs7412) that determines the *APOE* haplotype<sup>48</sup>.

326 When examining Veterans of EA only (N=249,785; 85,613 cases, 164,172  
327 controls), we similarly identified 15 GWS independent loci. As in the multi-ancestry  
328 analysis, the most significant locus was in *NCAM1*, with the same leading variant  
329 (rs1940701) identified. Of the 15 loci, 4 were distinct from the multi-ancestry analysis  
330 (Table 2, Supplemental Figure 2), but all had been suggestive of significance ( $p < 1 \times 10^{-6}$ )  
331 in the multi-ancestry meta-analysis. The leading SNP of the most significant of these 4  
332 distinct loci was in *ARHGAP15* on chromosome 2 (rs34799259, locus 5). Supplemental  
333 Table 4 shows functional mapping and annotation of the TBI GWAS results in the EA  
334 cohort. No GWS loci were identified in Veterans of AA (N=35,470; 15,714 cases, 19,756  
335 controls) or HA (N=19,230; 10,167 cases, 9,063 controls; Table 2, Supplemental  
336 Figures 3-4).

337

### 338 *Replication*

339 We performed replication analysis of the 15 EA GWS loci in the FinnGen  
340 concussion phenotype GWAS (15,787 cases, 184,565 controls; Supplemental Table 5).  
341 rs71149745 (locus 10, *FOXP2*) remained significant after adjustment for multiple  
342 comparisons. Three additional loci were at least nominally significant (at  $p < .05$ ): (1)  
343 rs1940701, nearest to *NCAM1* (locus 13); (2) rs12705715 (locus 9); and (3) rs1320139



344 (a proxy variant for rs34799259, locus 5), nearest to *ARHGAP15*. A sign test was  
345 performed to determine whether the direction of effects corresponded across studies at  
346 a rate beyond chance (50%). The direction of effects matched for 13 of the 15 variants,  
347 a significantly greater rate of correspondence than chance ( $p=.001$ ). When comparing  
348 the MVP EA case-control TBI phenotype to the FinnGen severe TBI phenotype (4,927  
349 cases, 304,227 controls; Supplemental Table 4), two of the 15 possible SNPs nominally  
350 replicated at  $p<.05$ : (1) rs71149745, nearest to *FOXP2* (locus 10) and (2) rs4933752 (a  
351 proxy variant for rs34350548, locus 12), nearest to *EXOC6*.

352

### 353 *Gene-Based Analyses*

354 In the genome-wide gene-based association study in the multi-ancestry cohort,  
355 14 genes were determined to be gene-wide significant following Bonferroni correction  
356 for multiple comparisons (Table 3, Supplemental Table 6, Supplemental Figure 5a). The  
357 top gene identified was *NCAM1* ( $p=4.10 \times 10^{-12}$ ), which was already named in positional  
358 mapping of the SNP-based analysis results, followed by *APOE* ( $p=1.30 \times 10^{-9}$ ), *FTO*  
359 ( $p=6.94 \times 10^{-9}$ ), and *FOXP2* ( $p=2.68 \times 10^{-8}$ ), also all positionally mapped to significant loci  
360 from the SNP-based analysis. In contrast, five of the significant genes from this analysis  
361 (*MSRA*, *FAM120A*, *NAV3*, *RBFOX1*, and *DCC*) were not identified based on positional  
362 mapping of loci from the SNP-based analysis. *MSRA* is located on the chromosome 8  
363 inversion region<sup>49</sup> where LD patterns are such that the association may be tagging the  
364 inversion itself rather than *MSRA* specifically.

365 We also examined GWS genes in Veterans of EA, AA, and HA separately. In the  
366 EA cohort, 18 genes were determined to be GWS following Bonferroni correction (Table  
367 3, Supplemental Table 7, Supplemental Figure 5b). Eight of these genes (*AKAP10*,  
368 *BRWD3*, *FAM46D*, *FBXL17*, *PHF2*, *RCN2*, *TCF12*, and *TSSK1B*) were distinct from the  
369 multi-ancestry gene-based analysis. Among these, the leading association was with  
370 *PHF2*. In the AA and HA cohorts, there were no gene-wide significant genes identified  
371 (data not shown).

372

### 373 *Gene Tissue Expression*

374 We used FUMA software<sup>40</sup> to evaluate gene-tissue expression in the multi-  
375 ancestry cohort. Results showed that among 30 general tissues, the brain was the only  
376 significantly enriched tissue, particularly in the frontal cortex, anterior cingulate cortex,  
377 and nucleus accumbens (tissues aggregated: Figure 2a, Supplemental Table 8;  
378 stratified by tissue subtype: Figure 2b, Supplemental Table 9). Additionally, gene-set  
379 analyses were conducted to understand potential gene pathways of importance;  
380 however, results did not identify any significant gene sets in the multi-ancestry cohort  
381 (Supplemental Table 10). When evaluating Veterans of EA separately, there were no  
382 significant gene-tissue expression results and no significant gene sets (Supplemental  
383 Tables 11-12; Supplemental Figures 6a-6b).

384

### 385 *SNP-Based Heritability*

386 Linkage disequilibrium score regression (LDSC)<sup>42</sup> was used to estimate SNP-  
387 based heritability ( $SNP-h^2$ ) of the case-control TBI phenotype in the EA cohort.  
388 Observed  $SNP-h^2$  was estimated to be 0.060 (SE=0.004,  $p=7.83 \times 10^{-66}$ ). On the liability  
389 scale, at 2% population prevalence,  $SNP-h^2$  was 0.044 (SE=0.003); at 20% population

390 prevalence,  $SNP-h^2$  was 0.087 (SE=0.005). To obtain further insights into the genetic  
391 architecture of TBI, we used univariate MiXeR analysis to decompose heritability into  
392 discoverability and polygenicity (number of influential variants) subcomponents. An  
393 estimated 10,470 (SE=848) influential variants explain 90% of the  $SNP-h^2$  and the  
394 discoverability was  $9.2 \times 10^{-6}$  (SE= $6.7 \times 10^{-7}$ ). The 15 variants identified in the EA GWAS  
395 explained an estimated 0.2% of the heritability of TBI.

396

### 397 *Genetic Overlap*

398 We examined genetic correlations ( $r_g$ ) between the MVP TBI variable and a  
399 variety of *a priori* selected health domains including risk-taking behaviors, psychiatric  
400 disorders, neurocognitive variables, and brain morphometrics in the EA cohort (Figure  
401 3, Supplemental Tables 13-16, Supplemental Figures 7-8). The strongest genetic  
402 correlation was with PTSD ( $r_g=0.69$ , SE=0.04,  $p=2.5 \times 10^{-75}$ ). Broadly, genetic  
403 correlations with TBI were moderate-to-large for risk-taking behaviors and psychiatric  
404 disorders. TBI was also negatively correlated with verbal numeric reasoning ( $r_g=-0.35$ ,  
405 SE=0.03,  $p=4.2 \times 10^{-42}$ ). Genetic correlations with reaction time and neurodegenerative  
406 disorders (i.e., AD and PD) were not significantly different from 0 after adjustment for  
407 multiple comparisons. With regard to brain morphometrics, TBI was weakly genetically  
408 correlated to ICV ( $r_g=-0.12$ , SE=0.04,  $p=0.002$ ), but all other brain morphometric  
409 correlations were not significantly different from 0.

410 We used univariate and bivariate MiXeR to perform a more detailed evaluation of  
411 the genetic overlap between TBI and these phenotypes. Univariate analyses suggest  
412 that the genetic architecture of the TBI phenotype (i.e., the polygenicity and  
413 discoverability of risk variants) was similar to polygenic psychiatric disorders  
414 (Supplemental Tables 17-18). In contrast, neurodegenerative disorders such as AD  
415 were characterized by relatively low polygenicity but high discoverability. In bivariate  
416 analyses, for five phenotypes investigated (AD, neuroticism, risky behaviors, reaction  
417 time (RT), and verbal numeric reasoning), there was evidence of polygenic overlap with  
418 TBI beyond what was explained by a genetic correlation model (best vs min AIC > 0).  
419 Notably, while the genetic correlations of TBI with AD and RT were not significantly  
420 different from 0, MiXeR reported that 60% (number of shared influential variants=57,  
421 SE=23) of the influential variation in AD and 94% (number of shared influential  
422 variants=8,902, SE=502) of the influential variation in RT were shared with TBI  
423 (Supplemental Figures 9-10). MiXeR also reported the existence of genetic variation  
424 specific to PD (best vs max AIC = 0.57), where the  $r_g$  estimated among only shared  
425 influential variants was -0.83 (SE=0.14).

426

### 427 *Genomic Structural Equation Modeling (gSEM)*

428 Given the strong genetic correlations between TBI and risk-taking behaviors and  
429 psychiatric disorders, genomic structural equation models<sup>46</sup> were conducted within the  
430 EA cohort to better understand whether TBI is more genetically associated with a latent  
431 factor representing risk-taking behaviors versus a latent factor representing psychiatric  
432 disorders. Pairwise genetic correlations were calculated between all phenotypes to  
433 produce a genetic correlation matrix (Supplemental Figure 7-8) for use in factor  
434 analysis. In genomic confirmatory factor analysis, the TBI variable loaded onto its own  
435 factor while the psychiatric disorders loaded onto one factor and the risk-taking

436 behaviors loaded onto another factor (CFI=0.908) (Figure 4). The correlation between  
437 the TBI and psychiatric disorders factors was 0.55 (SE=0.04), and the correlation  
438 between the TBI and risk-taking behaviors factors was 0.73 (SE=0.06), indicating that  
439 TBI risk has a stronger genetic correlation with risk-taking behaviors than with the  
440 psychiatric disorders. When examining genetic associations between TBI and individual  
441 psychiatric disorders, PTSD, alcohol dependence, and ADHD showed the strongest  
442 associations with TBI (Supplemental Figure 7). Given this observation, a final set of  
443 analyses was run to examine the genetic correlation between these individual disorders  
444 and TBI, controlling for the genetic association with risk-taking. These multiple  
445 regression models are presented in Figure 5; the model with PTSD continued to  
446 showed the strongest relationship with TBI even after accounting for risk-taking.

#### 447 448 *PheWAS*

449 Our PheWAS examined associations of leading variants in TBI risk loci with  
450 phenotypes across 28 domains (Supplemental Figure 11, Supplemental Table 19). The  
451 most common associations were with phenotypes in psychiatric (11 loci), activities (9  
452 loci), metabolic (8 loci), and environment (7 loci) domains (Supplemental Figure 12).  
453 Considering the degree of pleiotropy of the loci, the (median) average locus was  
454 associated with phenotypes across three different domains. Loci 15 (*FTO*) and 19  
455 (*APOE*) had significant associations with the largest number of domains (16 and 15  
456 domains, respectively). In contrast, six loci were only associated with phenotypes within  
457 a single domain (e.g., locus 7, which was only associated with a phenotype in the  
458 skeletal domain [osteoarthritis]).

#### 459 460 461 **Discussion**

462 Examining the role of genetics in TBI risk and recovery may offer novel  
463 mechanistic insights and lead to important discoveries that can inform preventive efforts  
464 and innovative treatments for patients who have sustained TBIs. In this study, we  
465 present the largest GWAS to date for TBI using data from the VA Million Veteran  
466 Program. Our results showed GWS variants associated with a TBI case-control  
467 phenotype in a multi-ancestry cohort of Veterans. These associations were primarily  
468 driven by the EA cohort, which comprised the majority of the overall sample (82%), and  
469 findings replicated in the FinnGen cohort (notably, the direction of effects matched for  
470 13 of the 15 variants). We also independently evaluated Veterans of African and  
471 Hispanic ancestry, but found no independent GWS loci, likely owing to reduced sample  
472 size and therefore lower power in those groups.

473 In the multi-ancestry cohort, our SNP-based results revealed 15 GWS loci, and  
474 our gene-based results revealed 14 gene-wide significant genes, many of which have  
475 presumed roles in psychiatric and neurologic functioning. The top GWS SNPs within the  
476 multi-ancestry sample were in and around *NCAM1*, followed by *APOE*, *FTO*, and  
477 *FOXP2*. Notably, many of these genes have emerged as significant genes in prior  
478 studies of neuropsychiatric and neurodegenerative conditions. For example, *NCAM1*  
479 (neural cell adhesion molecule 1) is a glycoprotein expressed on the surface of neurons,  
480 glia, and skeletal muscle, and has been shown to have a role in several brain-related  
481 biological processes such as neuronal development and maintenance as well as

482 synaptic plasticity<sup>50-52</sup>. The *NCAM1* gene has been associated with cognitive  
483 functioning, particularly learning and memory, and has been implicated in a range of  
484 neuropsychiatric and neurodegenerative disorders including schizophrenia, mood and  
485 anxiety disorders, and AD<sup>51</sup>.

486 Another notable gene to emerge as significant was *APOE*. *APOE* is a protein  
487 expressed in peripheral tissues and the central nervous system and its primary function  
488 is facilitating the transportation and distribution of lipids throughout the body. *APOE*  
489 plays a role in neuronal maintenance, neural transmission, neuroinflammation, and  
490 synaptic plasticity within the brain and has been associated with response and repair  
491 processes following neurotrauma<sup>53,54</sup>. The *APOE*  $\epsilon$ 4 allele is a major risk factor for AD<sup>55-</sup>  
492 <sup>57</sup>, and other research has shown associations between *APOE* and medical conditions  
493 such as vascular dementia, diabetes mellitus, and cardiovascular disease<sup>58,59</sup>, as well  
494 as with cognitive functioning<sup>60</sup>. Furthermore, our findings of an association between  
495 *APOE* and TBI align with previous candidate gene studies in TBI<sup>18,19,61-64</sup> and may help  
496 to explain the heterogeneous neurocognitive outcomes observed chronically following  
497 TBI. However, it is important to note that our association was observed with a single  
498 *APOE* SNP (and not the *APOE* haplotype which defines the *APOE*  $\epsilon$ 4 allele, which was  
499 not examined), and the effect of the haplotype may be different. Follow-up studies by  
500 our group are planned to more comprehensively address the role of the *APOE*  
501 haplotype in the context of TBI.

502 As for *FTO* (Fat and Obesity-Associated), less is known about the physiological  
503 function of this gene in the brain; however, *FTO* has been found to have important  
504 actions in metabolic and cardiovascular systems and may be associated with body  
505 mass index, obesity risk, and type 2 diabetes<sup>65,66</sup>. *FTO* gene variants have also been  
506 causally associated with risk of AD<sup>67</sup> and have been reported to promote  
507 phosphorylation of tau by activating the mTor pathway<sup>68</sup>, which has been implicated in  
508 AD pathophysiology. Finally, *FOXP2* (forkhead box P2) is a protein that is active in  
509 several tissues including the brain, and was initially implicated in speech and language  
510 development<sup>69,70</sup>. *FOXP2* has also been shown to play a role in other biological  
511 processes including brain development, cell differentiation and proliferation, and  
512 neurodegeneration<sup>69,70</sup>, and has previously been implicated as a genetic variant  
513 associated with ADHD, risk-taking behaviors, and psychiatric disorders such as  
514 PTSD<sup>70,71</sup>. Taken together, these observations offer critical insights into the genes that  
515 are associated with TBI. Although the functional significance of these genes within the  
516 context of TBI will still need to be determined, these initial findings are promising and  
517 offer targets for future research.

518 To better understand the role of TBI-associated genes in higher-order systems,  
519 we examined gene tissue expression. As expected, our results showed that the brain  
520 was the only significantly enriched region, with TBI-associated genes specifically  
521 implicated in the cortex, frontal cortex, anterior cingulate cortex (ACC), nucleus  
522 accumbens, caudate, hippocampus, amygdala, putamen, cerebellum, cerebellar  
523 hemisphere, and hypothalamus. These findings converge with a recent GWAS  
524 examining risk tolerance that similarly showed significant gene expression in brain  
525 regions such as the frontal cortex, ACC, and cortex<sup>72</sup>. Of note, the ACC is well-known to  
526 have connections with both the limbic system and the prefrontal cortex and has been



527 implicated in decision-making and impulse control<sup>73,74</sup>, underscoring the biological link  
528 between TBI and risk-taking.

529 With regard to SNP-based heritability of the TBI phenotype, we found an  
530 observed heritability of 6%, which is generally consistent with other heritability statistics  
531 reported for risk-taking behaviors (e.g., general risk tolerance:  $\text{SNP-}h^2=0.05^{72}$ ),  
532 psychopathology (e.g., PTSD [MVP]:  $\text{SNP-}h^2=0.05\text{-}0.07^{71}$ ), and neurocognition (e.g.,  
533 AD,  $\text{SNP-}h^2=0.07^{75}$ ). Notably, there are no adequately powered studies reporting  $\text{SNP-}$   
534  $h^2$  for TBI for comparison. Future studies are thus needed to further explore the  
535 heritability of TBI and to specifically tease apart heritability associated with TBI  
536 acquisition versus TBI sequelae.

537 Finally, we examined genetic correlations and our LDSC results suggest that  
538 there is significant genetic overlap between the TBI case-control phenotype and risk-  
539 taking behaviors and psychopathology. In contrast, there was weak genetic correlation  
540 between the TBI case-control phenotype and cognition and brain morphometrics, with  
541 the exception of verbal numeric reasoning. gSEM analyses further confirmed underlying  
542 genetic associations between TBI and both risk-taking behaviors and psychiatric  
543 disorders, though the association was stronger between TBI and risk-taking. Still, when  
544 evaluating individual traits/conditions, PTSD had the strongest genetic overlap with TBI,  
545 followed by ADHD and alcohol use when controlling for general risk tolerance, providing  
546 evidence of shared genetic etiology. Finally, MiXeR results showed that the genetic  
547 architecture of TBI was similar to polygenic psychiatric disorders and risk-taking traits.  
548 In contrast, neurodegenerative disorders such as AD showed much less polygenicity  
549 but had a high degree of shared variance with TBI. The discrepancies between the  
550 architectures of TBI and neurodegenerative disorders suggest that results need to be  
551 interpreted with caution; however, they are at least indicative of shared genetics.

552 This is the first large-scale GWAS of TBI in the VHA and findings significantly  
553 advance understanding of the genetic basis of TBI. Nevertheless, there are caveats to  
554 consider when interpreting our findings. To maximize sample size and ensure sufficient  
555 power, we defined a TBI case as any positive indication of TBI on either the MVP  
556 Baseline Survey or MVP Lifestyle Survey, or at least one inpatient or outpatient ICD  
557 code for TBI. In theory, this approach allowed us to capture lifetime history of TBI.  
558 However, the diverse methods of data collection, as well as the lack of precise  
559 information on the timing between injury event and reporting or documentation of the  
560 event, may have impacted results. Of note, the MVP Baseline Survey specifically  
561 assessed for lifetime history of concussion/TBI whereas the MVP Lifestyle Survey  
562 evaluated deployment-related TBI. Furthermore, the ICD codes were limited to the data  
563 available within the Veterans' VA EHRs, which may omit historical TBIs (i.e., TBI  
564 sustained prior to military service/enrollment in VA). Given the data, our TBI phenotype  
565 specifically focused on whether or not Veterans had a history of TBI, though it is  
566 possible that our phenotype also captured aspects of TBI recovery. Future MVP studies  
567 are planned to further understand the genetics associated with TBI-associated *sequelae*  
568 (e.g., "post-concussive syndrome" or "persistent post-concussive symptoms") and other  
569 aspects of TBI *recovery*, as it is likely that different genes may be associated with TBI  
570 acquisition/susceptibility vs. TBI outcome. Moreover, given that we did not evaluate or  
571 consider TBI severity in our analyses, future research may also consider evaluating TBI



572 or TBI-associated variables on a continuum—for example, examining injury severity or  
573 injury characteristics such as LOC or PTA.

574 Final study caveats relate to sample characteristics. Notably, the MVP cohort is  
575 predominately male; given our findings and the strong associations observed with risk-  
576 taking behaviors and psychiatric traits, it will be important for future research to  
577 determine whether genetic associations differ as a function of biological sex. The MVP  
578 cohort may be unique in other ways (e.g., high frequency of medical comorbidities in  
579 this population) and possibly distinct from other major biobanks; thus, it will be  
580 necessary for other biobanks to also examine the genetics associated with TBI.

## 581 **Conclusions**

583 Learning more about the genetics associated with TBI risk and recovery is  
584 necessary for developing a better understanding of the pathophysiology of this injury  
585 and could aid in developing novel therapeutics. The results of this first large-scale  
586 GWAS examining TBI in military Veterans identified 15 loci in the multi-ancestry cohort,  
587 including genes previously known to be relevant to TBI biology that will be important to  
588 investigate in future studies. Findings also showed that TBI is a heritable trait with  
589 comparable genetic architecture and high genetic correlation with psychiatric and risk-  
590 taking traits. Results set the stage for future TBI GWAS studies within MVP that focus  
591 on diversity and chronicity of symptom sequelae as well as severity of injury.

592  
593  
594

## 595 References

- 596 1. Helmick KM, Spells CA, Malik SZ, Davies CA, Marion DW, Hinds SR. Traumatic  
597 brain injury in the US military: Epidemiology and key clinical and research programs.  
598 *Brain Imaging and Behavior*. 2015;9(3):358-366.
- 599 2. DoD Numbers for Traumatic Brain Injury Worldwide - Totals (Defense Health  
600 Agency ) (2021).
- 601 3. Karr JE, Areshenkoff CN, Garcia-Barrera MA. The neuropsychological outcomes  
602 of concussion: a systematic review of meta-analyses on the cognitive sequelae of mild  
603 traumatic brain injury. *Neuropsychology*. 2014;28(3):321-336.
- 604 4. McCrea M, Guskiewicz K, Doncevic S, et al. Day of injury cognitive performance  
605 on the Military Acute Concussion Evaluation (MACE) by US military service members in  
606 OEF/OIF. *Military Medicine*. 2014;179(9):990-997.
- 607 5. Nelson LD, Temkin NR, Dikmen S, et al. Recovery after mild traumatic brain  
608 injury in patients presenting to US level I trauma centers: A transforming research and  
609 clinical knowledge in traumatic brain injury (TRACK-TBI) study. *JAMA Neurology*.  
610 2019;76(9):1049-1059.
- 611 6. Chapman JC, Diaz-Arrastia R. Military traumatic brain injury: A review.  
612 *Alzheimer's & Dementia*. 2014;10:S97-S104.
- 613 7. Schwab K, Terrio HP, Brenner LA, et al. Epidemiology and prognosis of mild  
614 traumatic brain injury in returning soldiers: A cohort study. *Neurology*.  
615 2017;88(16):1571-1579.
- 616 8. Perry DC, Sturm VE, Peterson MJ, et al. Traumatic brain injury is associated with  
617 subsequent neurologic and psychiatric disease: A meta-analysis. *Journal of*  
618 *Neurosurgery* 2016;124(2):511-526.
- 619 9. Stein MB, Ursano RJ, Campbell-Sills L, et al. Prognostic indicators of persistent  
620 post-concussive symptoms after deployment-related mild traumatic brain injury: A  
621 prospective longitudinal study in US Army soldiers. *Journal of Neurotrauma*.  
622 2016;33(23):2125-2132.
- 623 10. Stein MB, Jain S, Giacino JT, et al. Risk of posttraumatic stress disorder and  
624 major depression in civilian patients after mild traumatic brain injury: A TRACK-TBI  
625 study. *JAMA Psychiatry*. 2019;76(3):249-258.
- 626 11. Barnes DE, Byers AL, Gardner RC, Seal KH, Boscardin WJ, Yaffe K. Association  
627 of mild traumatic brain injury with and without loss of consciousness with dementia in  
628 US military veterans. *JAMA Neurology*. 2018;75(9):1055-1061.
- 629 12. Brickell TA, Lange RT, French LM. Health-related quality of life within the first 5  
630 years following military-related concurrent mild traumatic brain injury and polytrauma.  
631 *Military Medicine*. 2014;179(8):827-838.
- 632 13. Sakamoto MS, Delano-Wood L, Schiehser DM, Merritt VC. Predicting veteran  
633 health-related quality of life following mild traumatic brain injury. *Rehabilitation*  
634 *Psychology*. 2021;doi:10.1037/rep0000392
- 635 14. Agtarap SD, Campbell-Sills L, Jain S, et al. Satisfaction with life after mild  
636 traumatic brain injury: A TRACK-TBI study. *Journal of Neurotrauma*. 2021;38(5):546-  
637 554.
- 638 15. Pietrzak RH, Johnson DC, Goldstein MB, Malley JC, Southwick SM.  
639 Posttraumatic stress disorder mediates the relationship between mild traumatic brain  
640 injury and health and psychosocial functioning in veterans of Operations Enduring

- 641 Freedom and Iraqi Freedom. *The Journal of nervous and mental disease*.  
642 2009;197(10):748-753.
- 643 16. MacDonald CL, Barber J, Jordan M, et al. Early clinical predictors of 5-year  
644 outcome after concussive blast traumatic brain injury. *JAMA Neurology*.  
645 2017;74(7):821-829.
- 646 17. Taylor BC, Hagel EM, Carlson KF, et al. Prevalence and costs of co-occurring  
647 traumatic brain injury with and without psychiatric disturbance and pain among  
648 Afghanistan and Iraq War Veteran VA users. *Medical Care*. 2012:342-346.
- 649 18. McAllister TW. Genetic factors modulating outcome after neurotrauma. *PM&R*.  
650 2010;2(12):S241-S252.
- 651 19. Jordan BD. Genetic influences on outcome following traumatic brain injury.  
652 *Neurochemical Research*. 2007;32(4):905-915.
- 653 20. Weaver SM, Chau A, Portelli JN, Grafman J. Genetic polymorphisms influence  
654 recovery from traumatic brain injury. *The Neuroscientist*. 2012;18(6):631-644.
- 655 21. Dardiotis E, Fountas KN, Dardioti M, et al. Genetic association studies in patients  
656 with traumatic brain injury. *Neurosurgical Focus*. 2010;28(1):E9.
- 657 22. Zeiler FA, McFadyen C, Newcombe VF, et al. Genetic influences on patient-  
658 oriented outcomes in traumatic brain injury: A living systematic review of non-  
659 apolipoprotein E single-nucleotide polymorphisms. *Journal of Neurotrauma*.  
660 2021;38(8):1107-1123.
- 661 23. Panenka WJ, Gardner AJ, Dretsch MN, Crynen GC, Crawford FC, Iverson GL.  
662 Systematic review of genetic risk factors for sustaining a mild traumatic brain injury.  
663 *Journal of Neurotrauma*. 2017;34(13):2093-2099.
- 664 24. Lawrence DW, Comper P, Hutchison MG, Sharma B. The role of apolipoprotein  
665 E epsilon ( $\epsilon$ )-4 allele on outcome following traumatic brain injury: A systematic review.  
666 *Brain Injury*. 2015;29(9):1018-1031.
- 667 25. Sun X-C, Jiang Y. Genetic susceptibility to traumatic brain injury and  
668 apolipoprotein E gene. *Chinese Journal of Traumatology*. 2008;11(4):247-252.
- 669 26. Zhou W, Xu D, Peng X, Zhang Q, Jia J, Crutcher KA. Meta-analysis of APOE 4  
670 allele and outcome after traumatic brain injury. *Journal of Neurotrauma*. 2008;25(4):279-  
671 290.
- 672 27. Zeng S, Jiang J-X, Xu M-H, et al. Prognostic value of apolipoprotein E epsilon4  
673 allele in patients with traumatic brain injury: A meta-analysis and meta-regression.  
674 *Genetic Testing and Molecular Biomarkers*. 2014;18(3):202-210.
- 675 28. McFadyen CA, Zeiler FA, Newcombe V, et al. Apolipoprotein E4 polymorphism  
676 and outcomes from traumatic brain injury: A living systematic review and meta-analysis.  
677 *Journal of Neurotrauma*. 2021;38(8):1124-1136.
- 678 29. Gaziano JM, Concato J, Brophy M, et al. Million Veteran Program: A mega-  
679 biobank to study genetic influences on health and disease. *Journal of Clinical*  
680 *Epidemiology*. 2016;70:214-223.
- 681 30. Schwab K, Ivins B, Cramer G, et al. Screening for traumatic brain injury in troops  
682 returning from deployment in Afghanistan and Iraq: initial investigation of the usefulness  
683 of a short screening tool for traumatic brain injury. *The Journal of Head Trauma*  
684 *Rehabilitation*. 2007;22(6):377-389.

- 685 31. U.S. Department of Defense. Traumatic Brain Injury (TBI): DoD Standard  
686 Surveillance Case Definition for TBI Adapted for AFHSD Use. U.S. Department of  
687 Defense.
- 688 32. The 1000 Genomes Project Consortium. A global reference for human genetic  
689 variation. *Nature*. 2015;526(7571):68-74.
- 690 33. Loh P-R, Danecek P, Palamara PF, et al. Reference-based phasing using the  
691 Haplotype Reference Consortium panel. *Nature Genetics*. 2016;48(11):1443-1448.
- 692 34. Fang H, Hui Q, Lynch J, et al. Harmonizing genetic ancestry and self-identified  
693 race/ethnicity in genome-wide association studies. *The American Journal of Human*  
694 *Genetics*. 2019;105(4):763-772.
- 695 35. Manichaikul A, Mychaleckyj JC, Rich SS, Daly K, Sale M, Chen W-M. Robust  
696 relationship inference in genome-wide association studies. *Bioinformatics*.  
697 2010;26(22):2867-2873.
- 698 36. Abraham G, Qiu Y, Inouye M. FlashPCA2: Principal component analysis of  
699 Biobank-scale genotype datasets. *Bioinformatics*. 2017;33(17):2776-2778.
- 700 37. Chang CC, Chow CC, Tellier LC, Vattikuti S, Purcell SM, Lee JJ. Second-  
701 generation PLINK: rising to the challenge of larger and richer datasets. *Gigascience*.  
702 2015;4(7)
- 703 38. Willer CJ, Li Y, Abecasis GR. METAL: Fast and efficient meta-analysis of  
704 genomewide association scans. *Bioinformatics*. 2010;26(17):2190-2191.
- 705 39. Pruim RJ, Welch RP, Sanna S, et al. LocusZoom: Regional visualization of  
706 genome-wide association scan results. *Bioinformatics*. 2010;26(18):2336-2337.
- 707 40. Watanabe K, Taskesen E, Van Bochoven A, Posthuma D. Functional mapping  
708 and annotation of genetic associations with FUMA. *Nature Communications*.  
709 2017;8(1):1826.
- 710 41. Lee SH, Wray NR, Goddard ME, Visscher PM. Estimating missing heritability for  
711 disease from genome-wide association studies. *The American Journal of Human*  
712 *Genetics*. 2011;88(3):294-305.
- 713 42. Bulik-Sullivan BK, Loh P-R, Finucane HK, et al. LD Score regression  
714 distinguishes confounding from polygenicity in genome-wide association studies. *Nature*  
715 *Genetics*. 2015;47(3):291-295.
- 716 43. Holland D, Frei O, Desikan R, et al. Beyond SNP heritability: Polygenicity and  
717 discoverability of phenotypes estimated with a univariate Gaussian mixture model.  
718 *PLoS Genetics*. 2020;16(5):e1008612.
- 719 44. Frei O, Holland D, Smeland OB, et al. Bivariate causal mixture model quantifies  
720 polygenic overlap between complex traits beyond genetic correlation. *Nature*  
721 *Communications*. 2019;10:2417.
- 722 45. R Core Team. R: A language and environment for statistical computing.  
723 <https://www.R-project.org/>
- 724 46. Grotzinger AD, Rhemtulla M, de Vlaming R, et al. Genomic structural equation  
725 modelling provides insights into the multivariate genetic architecture of complex traits.  
726 *Nature Human Behaviour*. 2019;3(5):513-525.
- 727 47. Watanabe Kea. A global view of genetic architecture in human complex traits.  
728 [under preparation].
- 729 48. Seripa D, D'Onofrio G, Panza F, Cascavilla L, Masullo C, Pilotto A. The genetics  
730 of the human APOE polymorphism. *Rejuvenation Research*. 2011;14(5):491-500.

- 731 49. Broman KW, Matsumoto N, Giglio S, et al. Common long human inversion  
732 polymorphism on chromosome 8p. *Lecture Notes-Monograph Series*. 2003;237-245.
- 733 50. Schmid RS, Maness PF. *Cell recognition molecules and disorders of*  
734 *neurodevelopment*. International Handbook on Brain and Behaviour in Human  
735 Development. Kluwer Academic Publishers; 2001:199-218.
- 736 51. Brennaman LH, Maness PF. *NCAM in neuropsychiatric and neurodegenerative*  
737 *disorders*. vol 663. Structure and Function of the Neural Cell Adhesion Molecule NCAM:  
738 Advances in Experimental Medicine and Biology Springer; 2010:299-317.
- 739 52. Vukojevic V, Mastrandreas P, Arnold A, et al. Evolutionary conserved role of  
740 neural cell adhesion molecule-1 in memory. *Translational Psychiatry*. 2020;10:217.
- 741 53. Mahley RW, Huang Y. Apolipoprotein e sets the stage: Response to injury  
742 triggers neuropathology. *Neuron*. 2012;76(5):871-885.
- 743 54. Mahley RW, Weisgraber KH, Huang Y. Apolipoprotein E4: A causative factor and  
744 therapeutic target in neuropathology, including Alzheimer's disease. *Proceedings of the*  
745 *National Academy of Sciences*. 2006;103(15):5644-5651.
- 746 55. Farrer LA, Cupples LA, Haines JL, et al. Effects of age, sex, and ethnicity on the  
747 association between apolipoprotein E genotype and Alzheimer disease: A meta-  
748 analysis. *JAMA*. 1997;278(16):1349-1356.
- 749 56. Horsburgh K, McCarron MO, White F, Nicoll JA. The role of apolipoprotein E in  
750 Alzheimer's disease, acute brain injury and cerebrovascular disease: Evidence of  
751 common mechanisms and utility of animal models. *Neurobiology of Aging*.  
752 2000;21(2):245-255.
- 753 57. Verghese PB, Castellano JM, Holtzman DM. Apolipoprotein E in Alzheimer's  
754 disease and other neurological disorders. *The Lancet Neurology*. 2011;10(3):241-252.
- 755 58. Reinvang I, Espeseth T, Westlye LT. APOE-related biomarker profiles in non-  
756 pathological aging and early phases of Alzheimer's disease. *Neuroscience &*  
757 *Biobehavioral Reviews*. 2013;37(8):1322-1335.
- 758 59. Van Giau V, Bagyinszky E, An SSA, Kim SY. Role of apolipoprotein E in  
759 neurodegenerative diseases. *Neuropsychiatric Disease and Treatment*. 2015;11:1723-  
760 1737.
- 761 60. Wisdom NM, Callahan JL, Hawkins KA. The effects of apolipoprotein E on non-  
762 impaired cognitive functioning: A meta-analysis. *Neurobiology of Aging*. 2011;32(1):63-  
763 74.
- 764 61. McAllister TW, Flashman LA, Harker Rhodes C, et al. Single nucleotide  
765 polymorphisms in ANKK1 and the dopamine D2 receptor gene affect cognitive outcome  
766 shortly after traumatic brain injury: A replication and extension study. *Brain Injury*.  
767 2008;22(9):705-714.
- 768 62. Merritt VC, Clark AL, Sorg SF, et al. Apolipoprotein E  $\epsilon$ 4 genotype is associated  
769 with elevated psychiatric distress in veterans with a history of mild to moderate  
770 traumatic brain injury. *Journal of Neurotrauma*. 2018;35(19):2272-2282.
- 771 63. Merritt VC, Clark AL, Sorg SF, et al. Apolipoprotein E (APOE)  $\epsilon$ 4 genotype is  
772 associated with reduced neuropsychological performance in military veterans with a  
773 history of mild traumatic brain injury. *Journal of Clinical and Experimental*  
774 *Neuropsychology*. 2018;40(10):1050-1061.



- 775 64. Merritt VC, Lange RT, Lippa SM, et al. Apolipoprotein e (APOE)  $\epsilon$ 4 genotype  
776 influences memory performance following remote traumatic brain injury in US military  
777 service members and veterans. *Brain and cognition*. 2021;154:105790.
- 778 65. Zhao X, Yang Y, Sun B-F, Zhao Y-L, Yang Y-G. FTO and obesity: Mechanisms  
779 of association. *Current Diabetes Reports*. 2014;14(5):1-9.
- 780 66. McCarthy MI. Genomics, type 2 diabetes, and obesity. *New England Journal of*  
781 *Medicine*. 2010;363(24):2339-2350.
- 782 67. Reitz C, Tosto G, Mayeux R, Luchsinger JA, Group N-LNFS, Initiative tAsDN.  
783 Genetic variants in the Fat and Obesity Associated (FTO) gene and risk of Alzheimer's  
784 disease. *PloS One*. 2012;7(12):e50354.
- 785 68. Li H, Ren Y, Mao K, et al. FTO is involved in Alzheimer's disease by targeting  
786 TSC1-mTOR-Tau signaling. *Biochemical and Biophysical Research Communications*.  
787 2018;498(1):234-239.
- 788 69. Co M, Anderson AG, Konopka G. FOXP transcription factors in vertebrate brain  
789 development, function, and disorders. *Wiley Interdisciplinary Reviews: Developmental*  
790 *Biology*. 2020;9(5):e375.
- 791 70. Den Hoed J, Devaraju K, Fisher SE. Molecular networks of the FOXP2  
792 transcription factor in the brain. *EMBO Reports*. 2021;22(8):e52803.
- 793 71. Stein MB, Levey DF, Cheng Z, et al. Genome-wide association analyses of post-  
794 traumatic stress disorder and its symptom subdomains in the Million Veteran Program.  
795 *Nature Genetics*. 2021;53(2):174-184.
- 796 72. Linnér RK, Biroli P, Kong E, et al. Genome-wide association analyses of risk  
797 tolerance and risky behaviors in over 1 million individuals identify hundreds of loci and  
798 shared genetic influences. *Nature Genetics*. 2019;51(2):245-257.
- 799 73. Bush G, Luu P, Posner MI. Cognitive and emotional influences in anterior  
800 cingulate cortex. *Trends in Cognitive Sciences*. 2000;4(6):215-222.
- 801 74. Stevens FL, Hurley RA, Taber KH. Anterior cingulate cortex: unique role in  
802 cognition and emotion. *The Journal of Neuropsychiatry and Clinical Neurosciences*.  
803 2011;23(2):121-125.
- 804 75. Kunkle BW, Grenier-Boley B, Sims R, et al. Genetic meta-analysis of diagnosed  
805 Alzheimer's disease identifies new risk loci and implicates A $\beta$ , tau, immunity and lipid  
806 processing. *Nature Genetics*. 2019;51(3):414-430.
- 807  
808  
809  
810  
811  
812  
813  
814  
815  
816  
817  
818  
819  
820

821 **Acknowledgements**

822 The authors sincerely thank the Veterans who volunteered to participate in the  
823 Million Veteran Program. This research is based on data from the Million Veteran  
824 Program (Project MVP026), Office of Research and Development, Veterans Health  
825 Administration. This publication does not represent the views of the Department of  
826 Veteran Affairs or the United States Government.

827

828 **Funding**

829 This work was supported by a Career Development Award awarded to Dr. Merritt  
830 from the VA Clinical Science Research & Development Service (IK2 CX001952).

831

832 **Author Contributions**

833 Overall study coordination: VCM, CMN. Study concept and design: VCM, CMN,  
834 AXM, LDW. Data curation: CCC. Phenotype analysis: CCC (lead), VCM, LDW.  
835 Statistical analysis: AXM (lead), MG, EK. Data interpretation: AXM, MG, CMN, MBS,  
836 MSP, RLH, MWL, VCM. Writing: VCM (lead), AXM, CMN. All authors provided edits,  
837 feedback, and approved the final version of the paper for submission.

838

839 **Conflicts of Interest**

840 The authors have no conflicts of interest to declare.

841

## 842 **Figure Legends**

843

844 **Figure 1.** Manhattan plot of TBI risk for the **multi-ancestry cohort**. The GWAS was  
845 conducted using logistic regression, including first 10 principal components of ancestry  
846 as covariates. The x-axis displays the chromosome position, and the y-axis shows the  
847 GWAS  $p$  value on a  $-\log_{10}$  scale. The red dashed line reflects genome-wide significance  
848 at  
849  $p < 5 \times 10^{-8}$ .

850

851 **Figure 2a.** Gene tissue expression (tissues aggregated) in the **multi-ancestry cohort**.

852 **Figure 2b.** Gene tissue expression (stratified by tissue subtype) in the **multi-ancestry**  
853 cohort.

854

855 **Figure 3.** Genetic correlation between TBI and other phenotypes.

856 The x-axis denotes the magnitude of genetic correlation. The blue squares represent  
857 the genetic correlation between TBI and each given phenotype, with lines representing  
858 95% confidence intervals.

859

860 **Figure 4.** Genomic structural equation modeling (gSEM) results. Correlated factors  
861 model with standardized estimates and standard error estimates in parentheses.  
862 Genetic components of disorders are in circles and latent factors are represented by  
863 ovals. Single-headed arrows represent regression relationships and parameter  
864 estimates that can be squared to estimate the proportion of variance accounted for by  
865 components. Double-headed arrows represent correlations. Residual variance  
866 represented by double-headed arrow connecting the variable to itself. Each phenotype  
867 loads onto at least one factor. The TBI risk variable is loaded onto its own factor ( $T_g$ )  
868 while the psychiatric disorders loaded onto one factor ( $F1_g$ ) and the risk-taking  
869 behaviors loaded onto another factor ( $F2_g$ ); CFI=0.908. Abbreviations: TBI = traumatic  
870 brain injury; MDD = major depressive disorder; NEUR = neuroticism; PTSD =  
871 posttraumatic stress disorder; ALCH = alcohol dependence; GRT = general risk  
872 tolerance; ADHD = attention-deficit/hyperactivity disorder.

873

874 **Figure 5.** Multiple regression results. Genetic components of disorders are in circles.  
875 Single-headed arrows represent regression relationships and parameter estimates that  
876 can be squared to estimate the proportion of variance accounted for. Double-headed  
877 arrows represent correlations. Residual variance represented by double-headed arrow  
878 connecting the variable to itself. The models reflect the genetic association between TBI  
879 and (A) PTSD, (B) alcohol dependence, and (C) ADHD while controlling for general risk  
880 taking. Abbreviations: PTSD = posttraumatic stress disorder; GRT = general risk  
881 tolerance; TBI = traumatic brain injury; ALCH = alcohol dependence; ADHD = attention-  
882 deficit/hyperactivity disorder.

883

884

Figure 1

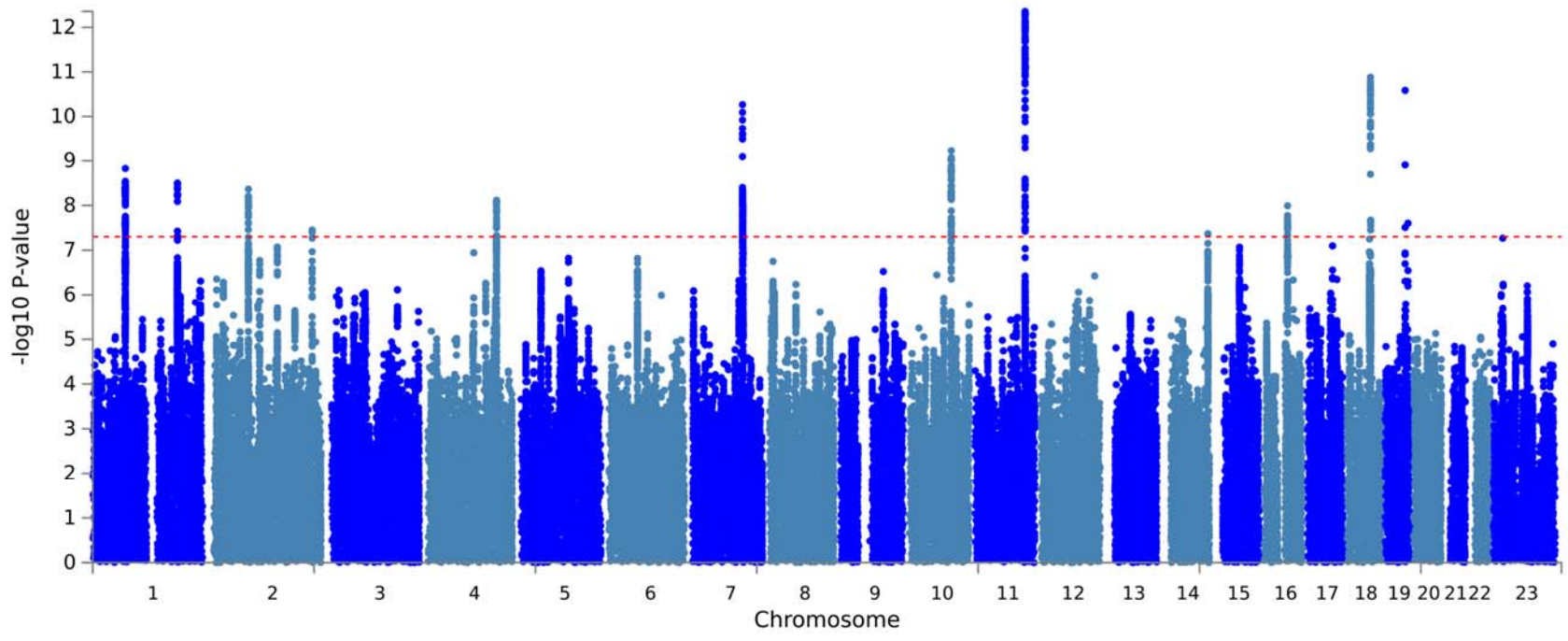


Figure 2a

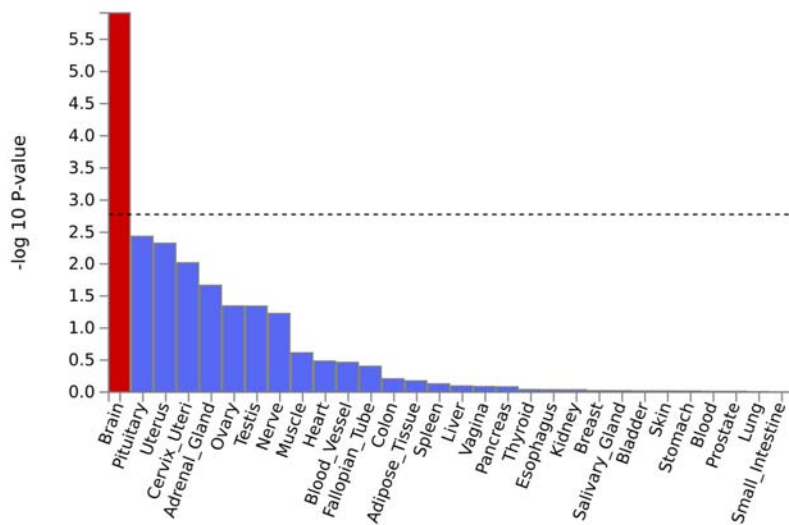


Figure 2b

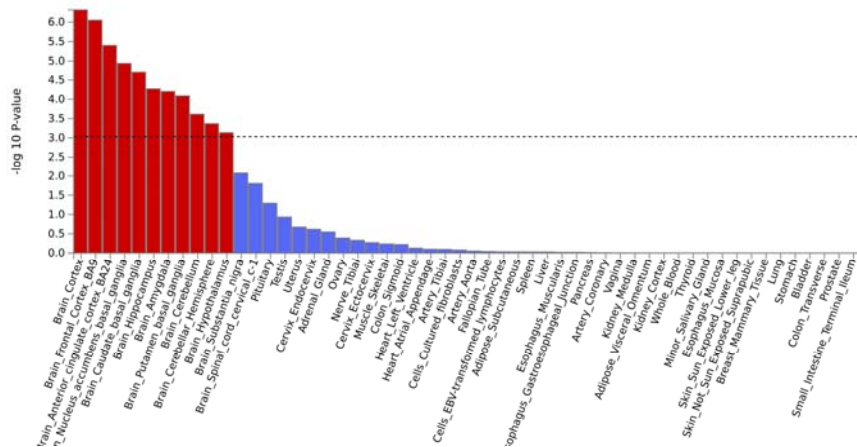




Figure 3

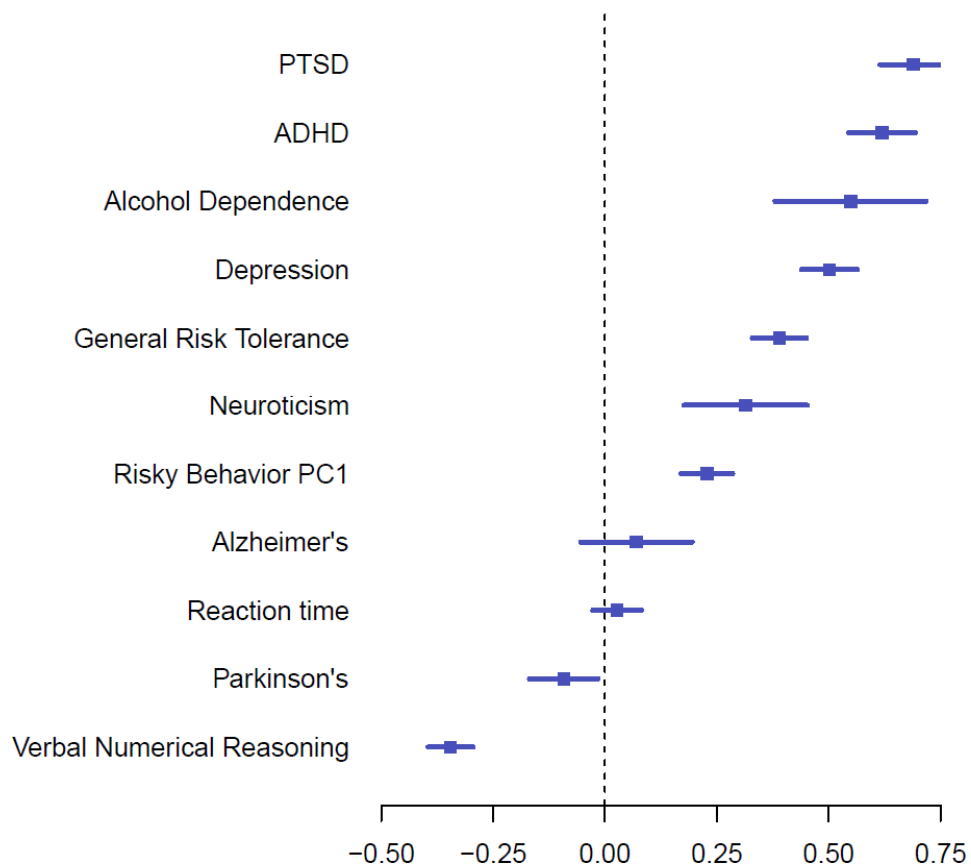


Figure 4

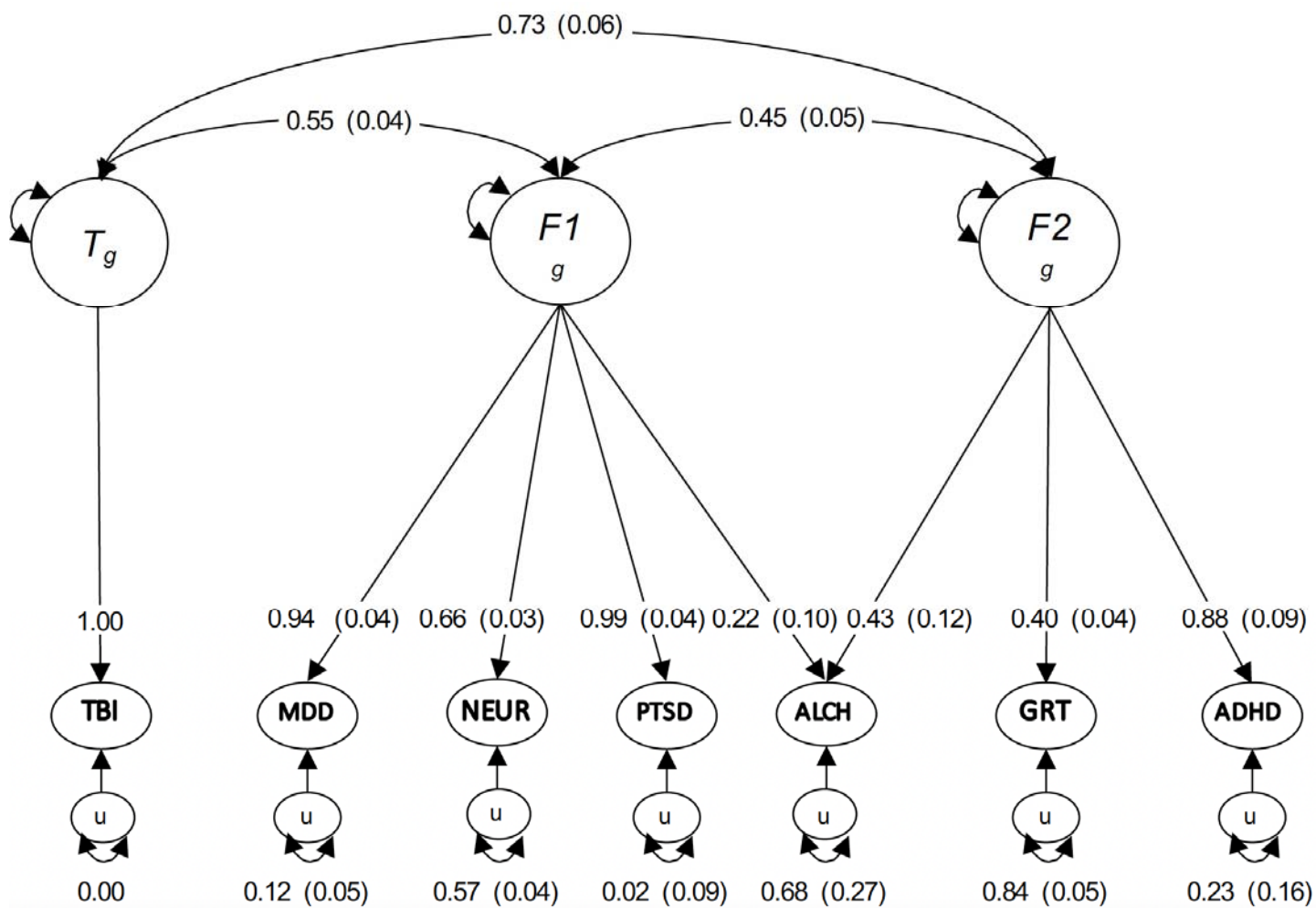
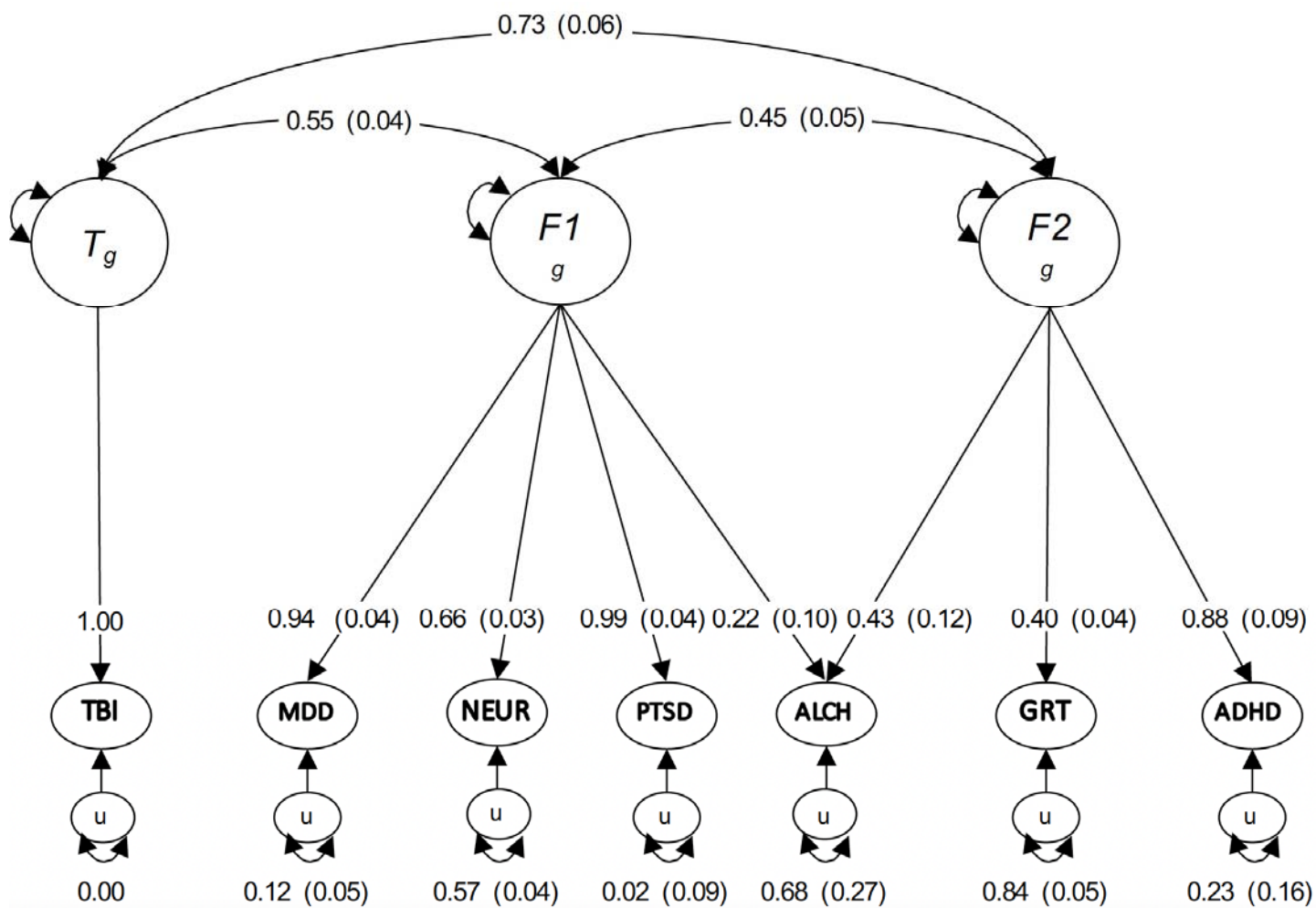


Figure 5



**Table 1. Multi-ancestry participant characteristics.**

	<b>Overall N=304,485</b>	<b>TBI Case n=111,494</b>	<b>Control n=192,991</b>	<b>p</b>
Age, Mean (SD)	64.36 (13.16)	59.92 (14.55)	66.92 (11.54)	<.001
Sex, N (%)				
Male	281,108 (92.3)	103,378 (92.7)	177,730 (92.1)	<.001
Female	23,370 (7.7)	8,109 (7.3)	15,261 (7.9)	
Missing	7 (0.0)	7 (0.0)	0 (0.0)	
Ancestry, N (%)				
European	249,785 (82.0)	85,613 (76.8)	164,172 (85.1)	<.001
African	35,470 (11.6)	15,714 (14.1)	19,756 (10.2)	
Hispanic	19,230 (6.3)	10,167 (9.1)	9,063 (4.7)	
Data Source, N (%)				
BL	9,039 (3.0)	9,039 (8.1)		--
LS	4,101 (1.3)	4,101 (3.7)		
ICD	23,776 (7.8)	23,776 (21.3)		
BL + LS	45,270 (14.8)	44,544 (40.0)	--	
BL + ICD	11,557 (3.8)	11,557 (10.4)		
LS + ICD	2,558 (0.8)	2,558 (2.3)		
BL + LS + ICD	208,184 (68.4)	15,919 (14.3)		
TBI Source, N (%)				
BL-CC	6,623 (2.2)	6,623 (5.9)		--
BL-TBI	1,251 (0.4)	1,251 (1.1)		
LS	4,101 (1.3)	4,101 (3.7)		
ICD	58,606 (19.2)	57,880 (51.9)		
BL-CC + BL-TBI	1,165 (0.4)	1,165 (1.0)		
BL-CC + LS	13,982 (4.6)	13,982 (12.5)		
BL-CC + ICD	1,399 (0.5)	1,399 (1.3)	--	
BL-TBI + ICD	3,138 (1.0)	3,138 (2.8)		
LS + ICD	205,154 (67.4)	12,889 (11.6)		
BL-CC + BL-TBI + ICD	3,478 (1.1)	3,478 (3.1)		
BL-CC + LS + ICD	2,106 (0.7)	2,106 (1.9)		
BL-TBI + LS + ICD	1,419 (0.5)	1,419 (1.3)		
All 4 Sources	2,063 (0.7)	2,063 (1.9)		

Abbreviations: BL = MVP Baseline Survey; LS = MVP Lifestyle Survey; ICD = International Statistical Classification of Diseases; BL-CC = MVP Baseline Survey, Concussion or Loss of Consciousness; BL-TBI = MVP Baseline Survey, Traumatic Brain Injury

Table 2. Genome-wide significant loci associated with TBI (case-control) for multi-ancestry, EA, AA, and HA cohorts.

Locus	SNP	Chr	Start	Stop	A1	A2	Multi-Ancestry (311,494 cases, 192,991 controls)				European Ancestry (85,613 cases, 164,172 controls)				African Ancestry (15,714 cases, 19,756 controls)				Hispanic Ancestry (10,167 cases, 9,063 controls)				Genes
							A1 Freq	Z score	p-value	HEPVal	A1 Freq	OR	SE	p-value	A1 Freq	OR	SE	p-value	A1 Freq	OR	SE	p-value	
1	rs140930533*	1	73,766,431	73,991,651	CAG	C	0.570	-0.035	<b>1.49E-09</b>	0.680	0.573	0.958	0.006	<b>4.09E-07</b>	0.641	0.956	0.017	0.007	0.406	0.955	0.022	0.037	ZFRANB2, LRRCS3, HKR8
2	rs523111*	1	190,886,930	191,418,368	A	C	0.404	0.034	<b>3.32E-09</b>	0.242	0.373	1.038	0.006	<b>1.70E-09</b>	0.670	1.008	0.017	0.649	0.340	1.027	0.022	0.235	BRINP2, RGS21
2	rs11332270*	1	191,154,894	191,418,368	G	GA	0.403	0.034	<b>3.16E-09</b>	0.309	0.372	1.038	0.006	<b>2.28E-09</b>	0.671	1.010	0.017	0.548	0.339	1.026	0.022	0.246	RGS21
3	rs4557006*	2	22,430,795	22,493,637	A	G	0.441	0.028	<b>5.96E-07</b>	0.063	0.442	1.034	0.006	<b>4.87E-08</b>	0.342	1.009	0.016	0.568	0.611	0.986	0.023	0.542	NA
4	rs138580110*	2	78,623,429	78,838,751	A	ATTAT	0.216	-0.036	<b>4.33E-09</b>	0.751	0.333	0.966	0.006	<b>1.04E-07</b>	0.122	0.948	0.024	0.027	0.277	0.968	0.024	0.170	NA
5	rs34799299*	2	144,151,161	144,248,905	A	AT	0.579	-0.030	<b>1.35E-07</b>	0.191	0.565	0.966	0.006	<b>2.90E-08</b>	0.612	0.995	0.017	0.753	0.717	0.988	0.024	0.611	ARHGAP15, GTDC1, ZEB2
6	rs17394072*	2	222,869,560	222,872,562	A	G	0.507	-0.031	<b>4.05E-08</b>	0.078	0.476	0.965	0.006	<b>2.95E-09</b>	0.819	1.002	0.020	0.928	0.547	0.997	0.021	0.903	FAM124B
6	rs562684999*	2	222,869,560	222,921,656	CT	C	0.535	-0.032	<b>3.54E-08</b>	0.770	0.512	0.970	0.006	<b>8.60E-07</b>	0.668	0.959	0.017	0.017	0.608	0.978	0.023	0.324	PAX3, CCDC140, SCPP2, FAM124B
7	rs41377571*	4	154,650,106	154,768,011	A	T	0.454	0.032	<b>1.89E-08</b>	0.073	0.464	1.037	0.006	<b>7.22E-09</b>	0.411	0.996	0.017	0.821	0.403	1.040	0.022	0.069	TMEM154, TIGD4, ARFIP1, FHDCL1, MND1, TLR2, RNF175, SFRP2, PLRG1, FGA
7	rs566208700*	4	154,650,106	154,768,011	G	GGTGTGTGGTGT	0.564	-0.033	<b>7.64E-09</b>	0.193	0.539	0.966	0.006	<b>1.58E-08</b>	0.738	0.998	0.019	0.913	0.642	0.952	0.023	0.030	TMEM154, TIGD4, ARFIP1, FHDCL1, MND1, TLR2, RNF175, SFRP2, PLRG1, FGA
8	rs147355048*	5	107,402,938	107,597,787	A	AGGATGGGATG	0.555	-0.030	<b>1.51E-07</b>	0.144	0.529	0.967	0.006	<b>4.57E-08</b>	0.825	0.975	0.022	0.253	0.620	1.011	0.022	0.617	EFNA5, FBXL17, MAN2A1, TMEM232, SLCS2A46
9	rs12705715*	7	110,267,021	110,284,071	A	G	0.490	-0.028	<b>7.19E-07</b>	0.057	0.466	0.967	0.006	<b>3.87E-08</b>	0.786	1.004	0.020	0.851	0.430	1.005	0.022	0.826	JRBN3, TMEM168
10	rs71149145*	7	113,858,363	114,290,415	A	AAATTCAT	0.597	0.038	<b>5.46E-11</b>	0.848	0.576	1.040	0.006	<b>3.20E-10</b>	0.839	1.028	0.022	0.207	0.613	1.034	0.022	0.134	PPP1R3A, FOP2, MDPC
11	rs201720501*	7	114,940,147	115,113,279	T	TG	0.561	0.034	<b>1.02E-08</b>	0.812	0.538	1.034	0.006	<b>2.91E-07</b>	0.744	1.031	0.019	0.114	0.575	1.049	0.023	0.036	WNT2
12	rs34350548*	10	94,577,629	94,750,745	T	TA	0.362	-0.038	<b>1.96E-09</b>	0.179	0.380	0.958	0.007	<b>4.31E-10</b>	0.204	0.999	0.023	0.978	0.275	0.981	0.028	0.487	KIF11, EXDC5, CYP26C1, CYP26A1, MYOF, CEP350, LGI1
12	rs4933752*	10	94,570,602	94,756,363	T	C	0.307	-0.039	<b>6.00E-10</b>	0.937	0.323	0.962	0.007	<b>1.33E-09</b>	0.084	0.954	0.032	0.141	0.191	0.969	0.028	0.252	KIF11, EXDC5, CYP26C1, CYP26A1, MYOF, CEP350, LGI1
13	rs1940701*	11	113,826,311	113,048,466	T	C	0.595	-0.040	<b>4.38E-13</b>	0.303	0.601	0.961	0.006	<b>1.35E-10</b>	0.609	0.973	0.016	0.078	0.566	0.934	0.021	0.001	IL8, TEX12, APOD8B4L3, RGD4, NCAM1, TM6SS5
14	rs12891288*	14	103,849,715	104,053,764	T	G	0.277	0.035	<b>4.29E-08</b>	0.758	0.275	1.036	0.007	<b>4.65E-07</b>	0.277	1.039	0.018	0.032	0.306	1.018	0.024	0.452	ANKRD9, TNFAIP2, EIF5, MARK3, CBX, TRMT61A, BAG5, KLIC1, RPL11-73M18.2, APOPT1, XRCC3, ALDH8B40.1, ZFYVE21, PPP1R13B, C14orf2, TOR1A, KIF26A
15	rs972653*	16	53,797,908	53,845,487	T	G	0.395	-0.030	<b>4.54E-08</b>	0.284	0.398	0.966	0.006	<b>2.81E-08</b>	0.418	0.991	0.015	0.567	0.289	0.979	0.023	0.350	RPGRP1L, FTO, RPL11-324D17.1, IRX3, IRX5, IRX6, MMP2, CES1, GNAO1, MT18, MT1F, MT1G, MT1H, MT1X
15	rs62033400*	16	53,797,908	53,845,487	A	G	0.628	0.033	<b>1.02E-08</b>	0.858	0.604	1.035	0.006	<b>2.37E-08</b>	0.897	1.026	0.025	0.317	0.744	1.024	0.024	0.333	RPGRP1L, FTO, RPL11-324D17.1, IRX3, IRX5, IRX6, MMP2, CES1, GNAO1, MT18, MT1F, MT1G, MT1H, MT1X
16	rs79940062*	17	4,708,126	4,879,323	A	G	N/A	N/A	N/A	N/A	0.023	1.129	0.020	<b>9.11E-10</b>	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A	UBE2G1, MED11, VMG1, GLTPD2, PSMB6, PLD2, MINK1, CHRNE, C17orf107, SLCS25A11, RNF167, PFN1, ENO3, SPAG7, CAMTA2, INCA1, KIF1C, SCMP, NLRP1
17	rs4608411*	18	52,765,283	52,827,668	T	C	0.322	0.041	<b>2.70E-11</b>	0.210	0.334	1.041	0.007	<b>8.44E-10</b>	0.111	1.010	0.026	0.688	0.343	1.073	0.023	0.002	RAB27B, TCF4, TXNL1, WDR7, NARS, NEDD4L
17	rs729053*	18	52,765,283	52,827,668	C	G	0.324	0.040	<b>1.32E-11</b>	0.302	0.335	1.040	0.006	<b>1.04E-09</b>	0.097	1.022	0.027	0.412	0.342	1.073	0.022	0.001	RAB27B, TCF4, TXNL1, WDR7, NARS, NEDD4L
18	rs1818869*	18	53,108,836	53,456,801	T	C	N/A	N/A	N/A	N/A	0.145	0.951	0.009	<b>3.50E-08</b>	N/A	N/A	N/A	N/A	0.080	0.970	0.043	0.844	STARD6, C18orf54
18	rs1651350*	18	53,210,302	53,464,917	A	G	0.876	0.049	<b>2.14E-08</b>	0.825	0.871	1.051	0.009	<b>5.29E-08</b>	0.975	1.071	0.052	0.186	0.388	1.028	0.044	0.529	STARD6, C18orf54
19	rs429358*	19	45,392,254	45,424,351	T	C	0.851	-0.051	<b>2.60E-11</b>	0.335	0.862	0.949	0.009	<b>1.00E-09</b>	0.787	0.945	0.018	0.002	0.890	0.996	0.033	0.905	PVR12, TOMM40, APOE, APOC1, APOC4-APOC2, APOC4, APOC2, C7B-129P6.11, CLPTM1, CLASRP, TRAPP6A, BLOC1S3, AC005779.2
20	rs14810808*	19	51,781,455	51,815,081	A	G	0.053	-0.079	<b>2.50E-08</b>	0.857	0.055	0.925	0.015	<b>1.13E-07</b>	0.010	0.907	0.086	0.299	0.019	0.884	0.088	0.161	NARS4, C19orf48, SIGLEC1, IGLONS, VSG10L, ETFB, CTD-2616/11.11, CLDN22, HNG7, ZNF577

Note: For multi-ancestry analysis, a SNP had to be present in all 3 ancestries (EA, AA, and HA) There were no independently significant hits in the African or Hispanic ancestries.

\*top SNP (within each locus) in the multi-ancestry cohort.

\*top SNP (within each locus) in the European ancestry cohort.



**Table 3. Gene-based loci analysis (Bonferroni significant) in the multi-ancestry and European ancestry cohorts.**

SYMBOL	CHR	START	STOP	Multi-Ancestry				European Ancestry			
				NSNPS	NPARAM	ZSTAT	P	NSNPS	NPARAM	ZSTAT	P
<i>SFRP2</i> *^	4	154701744	154710272	16	4	4.761	9.65E-07	18	5	4.579	2.34E-06
<i>RNF175</i> *^	4	154631277	154681387	208	12	4.951	3.70E-07	216	13	5.040	2.33E-07
<i>TSSK1B</i> ^	5	112768251	112770728	-	-	-	-	15	4	4.605	2.06E-06
<i>FBXL17</i> ^	5	107194736	107717799	-	-	-	-	1390	38	4.703	1.28E-06
<i>FOXP2</i> *^	7	113726382	114333827	889	34	5.439	2.68E-08	1122	49	5.794	3.43E-09
<i>MSRA</i> *	8	9911778	10286401	1471	59	4.707	1.25E-06	-	-	-	-
<i>FAM120A</i> *^	9	96214004	96328397	300	15	4.704	1.28E-06	354	19	5.054	2.17E-07
<i>PHF2</i> ^	9	96338689	96441869	-	-	-	-	491	26	5.270	6.84E-08
<i>EXOC6</i> *^	10	94590935	94819250	525	19	4.775	8.97E-07	652	26	4.663	1.56E-06
<i>NCAM1</i> *^	11	112831997	113149158	848	38	6.835	4.10E-12	999	57	6.195	2.92E-10
<i>NAV3</i> *^	12	78224685	78606790	1251	63	5.152	1.29E-07	1463	78	5.097	1.73E-07
<i>TRMT61A</i> *	14	103995521	104003410	18	3	4.488	3.59E-06	-	-	-	-
<i>RCN2</i> ^	15	77223960	77242601	-	-	-	-	34	12	4.499	3.42E-06
<i>TCF12</i> ^	15	57210821	57591479	-	-	-	-	1009	18	4.515	3.16E-06
<i>FTO</i> *^	16	53737875	54155853	1303	93	5.675	6.94E-09	1503	116	5.617	9.72E-09
<i>RBFOX1</i> *	16	6069095	7763340	9543	385	4.937	3.98E-07	-	-	-	-
<i>AKAP10</i> ^	17	19807615	19881656	-	-	-	-	123	8	4.674	1.48E-06
<i>DCC</i> *^	18	49866542	51057784	4415	85	4.993	2.97E-07	5169	101	4.926	4.19E-07
<i>APOE</i> *^	19	45409011	45412650	5	2	5.955	1.30E-09	6	3	5.682	6.66E-09
<i>APOC1</i> *	19	45417504	45422606	11	3	4.621	1.91E-06	-	-	-	-
<i>FAM46D</i> ^	X	79591003	79700810	-	-	-	-	58	3	5.055	2.15E-07
<i>BRWD3</i> ^	X	79926353	80065187	-	-	-	-	180	10	4.540	2.81E-06

Notes: N=304,485 for the multi-ancestry cohort; N=249,785 for the European ancestry cohort.

\*top gene in the multi-ancestry cohort.

^top gene in the European ancestry cohort.