

# Genome-Wide Association Studies of Coffee Intake in UK/US Participants of European Ancestry Uncover Gene-Cohort Influences

Hayley H A Thorpe<sup>1,2</sup>, Pierre Fontanillas<sup>3</sup>, Benjamin K Pham<sup>4</sup>, John J Meredith<sup>4</sup>, Mariela V Jennings<sup>4</sup>, Natasia S Courchesne-Krak<sup>4</sup>, Laura Vilar-Ribó<sup>5</sup>, Sevim B Bianchi<sup>4</sup>, Julian Mutz<sup>6</sup>, 23andMe Research Team<sup>2</sup>, Sarah L Elson<sup>2</sup>, Jibrán Y Khokhar<sup>1,2</sup>, Abdel Abdellaoui<sup>7</sup>, Lea K Davis<sup>8,9,10</sup>, Abraham A Palmer<sup>4,11</sup>, & Sandra Sanchez-Roige<sup>4,9</sup>

<sup>1</sup>Department of Anatomy and Cell Biology, Schulich School of Medicine and Dentistry, Western University, London, ON, Canada.

<sup>2</sup>Department of Biomedical Sciences, Ontario Veterinary College, University of Guelph, Guelph, ON, Canada

<sup>3</sup>23andMe, Inc., Sunnyvale, CA, USA

<sup>4</sup>Department of Psychiatry, University of California San Diego, La Jolla, CA, USA

<sup>5</sup>Psychiatric Genetics Unit, Group of Psychiatry, Mental Health and Addiction, Vall d'Hebron Research Institute (VHIR), Universitat Autònoma de Barcelona, Barcelona, Spain

<sup>6</sup>Social, Genetic and Developmental Psychiatry Centre, Institute of Psychiatry, Psychology & Neuroscience, King's College London, London, UK

<sup>7</sup>Department of Psychiatry, Amsterdam UMC, University of Amsterdam, Amsterdam, The Netherlands

<sup>8</sup>Department of Biomedical Informatics, Vanderbilt University Medical Center, Nashville, TN, USA

<sup>9</sup>Department of Medicine, Vanderbilt University Medical Center, Nashville, TN, USA

<sup>10</sup>Department of Psychiatry and Behavioral Sciences, Vanderbilt University Medical Center, Nashville, TN, USA

<sup>11</sup>Institute for Genomic Medicine, University of California San Diego, La Jolla, CA, USA

25     **Corresponding author:** Sandra Sanchez-Roige, PhD, Department of Psychiatry, University of  
26                   California San Diego, San Diego, CA, USA. Email: [sanchezroige@ucsd.edu](mailto:sanchezroige@ucsd.edu)

27

## 28 **Abstract**

29 Coffee is one of the most widely consumed beverages. We performed a genome-wide  
30 association study (**GWAS**) of coffee intake in US-based 23andMe participants ( $N=130,153$ ) and  
31 identified 7 significant loci, with many replicating in three multi-ancestral cohorts. We examined  
32 genetic correlations and performed a phenome-wide association study across thousands of  
33 biomarkers and health and lifestyle traits, then compared our results to the largest available  
34 GWAS of coffee intake from UK Biobank (UKB;  $N=334,659$ ). The results of these two GWAS were  
35 highly discrepant. We observed positive genetic correlations between coffee intake and  
36 psychiatric illnesses, pain, and gastrointestinal traits in 23andMe that were absent or negative in  
37 UKB. Genetic correlations with cognition were negative in 23andMe but positive in UKB. The only  
38 consistent observations were positive genetic correlations with substance use and obesity. Our  
39 study shows that GWAS in different cohorts could capture cultural differences in the relationship  
40 between behavior and genetics.

## Introduction

Coffee is a leading global food commodity that has psychoactive properties that are largely due to the presence of caffeine<sup>1</sup>. While rates of use and daily intake varies widely by geographic region, it is estimated that approximately 60-85% of adults in Europe and the United States consume between 0.6 to 5.5 cups of coffee daily<sup>2-4</sup>. Intake of coffee and its bioactive constituents is associated with benefits on cognitive function<sup>5</sup> and lower risk of liver disease<sup>6,7</sup> (but see<sup>8</sup>), Parkinson's and other neurodegenerative diseases<sup>6,7,9</sup>, cardiovascular disease<sup>6,7</sup>, type II diabetes<sup>6,7</sup>, and certain cancers<sup>6,7,10</sup>. However, coffee intake is also associated with higher risks for some adverse outcomes, including increased risk of other substance use and misuse<sup>11-14</sup>, some cancers (e.g., lung cancer<sup>7,10,15</sup>), poor lipid profile<sup>6,7</sup>, pregnancy loss<sup>6,7</sup>, gastrointestinal maladies<sup>16</sup>, and worse cardiovascular outcomes following excessive intake<sup>17</sup>. Given the widespread and regular intake of coffee across the globe, addressing the full spectrum of correlations with health and disease is an important but challenging task.

Genetic studies offer a compelling avenue to investigate the relationships between coffee intake and other complex traits. Twin studies that calculate genetic contributions to daily coffee intake estimate it to be 36-56% heritable, suggesting that coffee intake should be amenable to genetic analysis. Whereas phenotypic correlations, which depend on measuring two or more traits in the same cohort, can arise from genetic and environmental factors, genetic correlations assess genetically driven relationships using the results from genome-wide association studies (**GWAS**) and can therefore examine correlations between two or more traits, even if they were measured in entirely non-overlapping cohorts. In the past decade, over a dozen GWAS ( $N=1,207-407,072$ ) have examined coffee intake<sup>18-34</sup>. Several of these GWAS have found associations with single nucleotide polymorphisms (**SNPs**) within or near genes that metabolize caffeine (**Supplementary Table 1**), such as *CYP1A1* and *CYP1A2*<sup>18-20,23-26,30,32,33</sup>. Some of these loci are also associated with other complex traits, including liver disease<sup>35-37</sup>, cancers<sup>38-41</sup>, and alcohol consumption<sup>42-44</sup>.



This pleiotropy could suggest that these other associations are mediated by coffee intake or that these loci also influence these traits via alternative independent mechanisms. Genetic correlations have also been reported between coffee intake with other substance use<sup>45,46</sup>, reduced gray matter volumes<sup>18</sup>, psychiatric illness<sup>45</sup>, osteoarthritis<sup>47</sup>, sleep<sup>48</sup>, body mass index (**BMI**)<sup>49</sup>, type II diabetes<sup>49</sup>, and migraine<sup>50</sup>. However, some genetic correlations were conducted under a *priori* justification (e.g., other substance use traits, sleep) and as such may fail to capture the full scope of genetic correlations between coffee with other traits. Thus, a data-driven examination of trait associations with coffee intake remains unexplored.

While coffee is the primary source of caffeine for many, other common dietary sources of caffeine include tea, soft drinks, and chocolate. Consequently, when we refer to coffee intake, we mean explicit measures of coffee intake (e.g., measured as cups/day) and not caffeine intake unless otherwise specified. Intake of other caffeine sources also varies by geographic region based on beverage sales<sup>2</sup>. For example, tea (rather than coffee) is the preferred source of caffeine in the United Kingdom (UK; tea vs. coffee: ~50% vs. 20%) compared to the United States (US; ~10% vs. 30%)<sup>2</sup>. As some genetic studies used data from the UK Biobank (**UKB**) only<sup>18,47,48,51-53</sup> or combined cohorts across regions with different patterns of caffeinated beverage intake (**Supplementary Table 1**)<sup>32,33,46</sup>, this distinction may limit generalizability or introduce environmental and cultural confounds that affect the genetic associations between coffee intake and other traits.

In this study, we used survey responses from US-based 23andMe, Inc. research participants of European ancestry ( $N=130,153$ ) and performed a GWAS of a single item “How many 5-ounce (cup-sized) servings of caffeinated coffee do you consume each day?”. Using genetic correlations and phenome- and laboratory-wide association studies (**PheWAS**, **LabWAS**), we explored the relationships between coffee intake and thousands of biomarkers, health features, and lifestyle traits to provide a fuller inventory of genetic correlations with coffee

intake. We compared our findings from the 23andMe cohort to those from the UKB using publicly available GWAS summary statistics of coffee intake (“How many cups of coffee do you drink each day? (Include decaffeinated coffee)”,  $N=334,659$ , <http://www.nealelab.is/uk-biobank/>). Although we had originally intended to perform a meta-analysis, our results revealed a lower-than-expected genetic correlation between coffee intake in the two cohorts; therefore, we instead used these datasets to explore cohort differences in coffee intake across these two distinct populations.

## Results

### GWAS in the 23andMe US-based cohort replicated seven loci implicated in coffee intake

Participant demographics of the 23andMe cohort are described in **Supplementary Table 2**. The cohort was 65% male, had a mean age of  $52.8 \pm 16.9$  years old, and an average BMI of  $28.38 \pm 6.54$  (range: 14.0-69.1), similar to the US average of 27.5 (95% CI: 25.5-29.4)<sup>54</sup>. The average coffee intake in the cohort was  $1.98 (\pm 2.35 \text{ SD})$  cups per day, similar to the coffee intake distributions in UKB ( $2.14 \pm 2.09 \text{ SD}$ ; see **Supplementary Figure 1** and **Supplementary Table 3** for distributions).

We conducted a GWAS of 14,274,006 imputed genetic variants assuming an additive genetic model that included age, sex, the first five genetic principal components, and indicator variables for genotype platforms as covariates (**Supplementary Table 4**; **Supplementary Material** for additional genotyping and GWAS details). The genomic control inflation factor of the GWAS was  $\lambda=1.09$ , suggesting no substantial inflation due to population stratification. SNP-heritability of coffee intake via Linkage Disequilibrium score regression (**LDSC**) was  $7.57\% \pm 0.59$  (**Supplementary Table 5**).

We identified seven genome-wide significant ( $p < 5.00 \times 10^{-8}$ ) independent ( $r^2 < 0.1$ ) loci that were associated with coffee intake (**Figure 1**, **Table 1**; **Supplementary Figures 2-8** for locus zoom plots). These associations replicated prior coffee or caffeine GWAS findings

(**Supplementary Table 6**). For example, rs2472297 ( $p=3.60E-65$ , chr15q24.1) is in the intergenic region between *CYP1A1* and *CYP1A2*, and has been previously associated with coffee and caffeine intake<sup>18,20,24,26,28,30,32,33,55</sup>. *CYP1A1* and *CYP1A2* encode members of the cryptochrome P450 superfamily of enzymes involved in xenobiotic metabolism<sup>22</sup>. rs2472297 has also been previously associated with traits like alcohol consumption<sup>56,57</sup>, clozapine pharmacokinetics<sup>58</sup>, kidney function<sup>59-62</sup>, and the concentration of biomarkers in urine<sup>63-68</sup>. We also identified rs4410790 ( $p=5.20E-55$ , chr7p21.1), which is located upstream of the *AHR* gene encoding a transcription factor that regulates *CYP1A1/CYP1A2* and is activated by polycyclic aromatic hydrocarbons, which are present in coffee<sup>22,69</sup>. Prior studies associated rs4410790 and caffeine intake from tea<sup>20</sup>, as well as with traits like caffeine metabolism<sup>28</sup>, bitter beverage intake<sup>26</sup>, and urine biomarkers<sup>64,66-68,70</sup>. Lastly, rs199612805 ( $p=1.80E-10$ , chr22q11.23), which is located near *ADORA2A*, was also implicated in coffee intake. This variant was recently associated with caffeine intake from tea and coffee in the UKB<sup>20</sup>. *ADORA2A* encodes an adenosine G-protein coupled receptor that is inhibited by caffeine to produce stimulating effects<sup>71</sup>. The remaining four SNPs – rs34645063, rs28634426, rs117824460, and rs11474881 – were in linkage disequilibrium (LD) with SNPs previously identified by other coffee or caffeine GWAS<sup>18,20,24,26,28,30,55</sup>. rs34645063 ( $p=3.30E-09$ , chr6q16.1) is a deletion/insertion polymorphism between *MMS22L* and *POU3F2*. rs34645063 is in LD ( $R^2=0.74$ ) with rs754177720 and is also associated with caffeine intake from coffee or tea<sup>20</sup>. rs28634426 ( $p=2.10E-10$ , chr7q11.23) is an intronic variant of *STYXL1* in LD with rs17685 ( $R^2=0.78$ ) and rs1057868 ( $R^2=0.76$ ), which were previously implicated by coffee GWAS<sup>18,20,24,26,30,55</sup>. rs117824460 ( $p=1.70E-08$ , chr19q13.2) is an intronic variant of *CYP2A6*, and is in LD ( $R^2=0.05$ ) with rs56113850, which was implicated in coffee intake<sup>20,26</sup> and caffeine metabolism<sup>28</sup>. *CYP2A6* encodes a cryptochrome P450 superfamily enzyme member that metabolizes nicotine<sup>72</sup>; rs117824460 has also been associated with smoking traits<sup>57,73</sup> and serum albumin<sup>74</sup>, C-reactive protein<sup>75</sup>, and liver alkaline phosphatase levels<sup>66</sup>. The final significant variant we identified, rs11474881 (chr20q13.33,  $p=1.10E-08$ ), is an intronic variant of the

*PCMTD2* gene; rs11474881 is in LD ( $R^2=0.98$ ) with rs6062679, which was previously implicated in coffee and tea intake and bitter beverage consumption<sup>26</sup>. We used three additional multi-ancestral cohorts to replicate these findings (**Table 1; Supplementary Table 3**). Of the SNPs that passed QC, all replicated in a larger sample of 23andMe research participants of European ancestry ( $N=689,661$ ), all replicated in those with African American ancestry ( $N=32,312$ ), and one replicated in those of Latin American ancestry ( $N=124,155$ ).

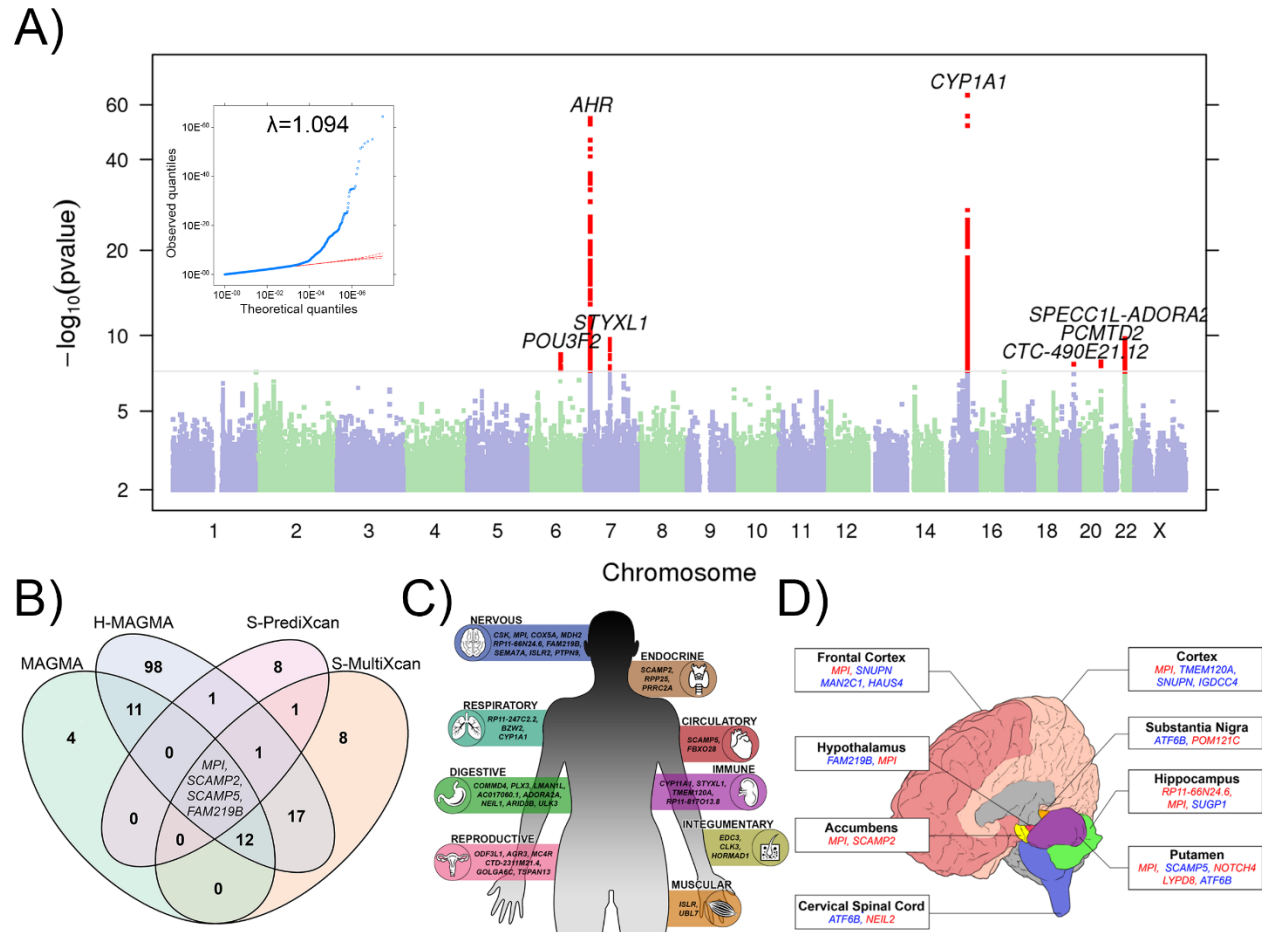
We also report several notable nominal associations with coffee intake ( $p<1.00E-06$ , **Supplementary Table 6**). rs72790130 ( $p=5.50E-08$ , chr16q23.3) and rs2155645 ( $p=9.80E-07$ , chr11q23.2) are intronic variants of two cell adhesion molecule genes, *CDH13* and *NCAM1*, respectively. Both genes have been previously implicated in substance use traits by other GWAS<sup>20,57,76-82</sup> and candidate gene studies in humans and animal models<sup>83-88</sup>. rs2155645 is also in LD with rs2298527 ( $R^2=0.43$ ), which was previously implicated in daily caffeine intake from coffee<sup>20</sup>. rs11204734 ( $p=2.90E-07$ , chr1q21.3) is an intronic variant of *ARNT*; its protein heterodimerizes with *AHR* and binds to xenobiotic response elements to regulate transcription of *CYP1A1* and *CYP1A2*<sup>22</sup>. Finally, rs340019 ( $p=2.10E-07$ , chr15q22.2) is an intronic variant of *RORA*, which is involved in circadian rhythm and metabolic regulation, among other functions<sup>89</sup>. rs340019 is in LD with rs12591786 ( $R^2=0.25$ ), which was implicated in daily caffeine intake from cups of coffee and tea<sup>20</sup>.

159 **Table 1.** Significant ( $p < 5.00 \times 10^{-8}$ ) GWAS results for coffee intake from 23andMe research participants ( $N = 130,153$ ) of European  
160 ancestry (**EA**). Replication (**EA Rep**) was conducted in an additional cohort of 23andMe participants of EA ( $N = 689,661$ ), and in those  
161 with African American ancestry (**AA**;  $N = 32,312$ ) and Latin American ancestry (**LA**;  $N = 124,155$ ); \*SNPs that did not pass QC in  
162 replication. See **Supplementary Table 6** for additional information.

SNP	BP	Alleles	Cytoband	$p$ value	EA EAF	Effect	EA Rep $p$ value	EA Rep EAF	AA $p$ value	AA EAF	LA $p$ value	LA EAF	Nearest gene(s)
rs2472297	75027880	C/T	chr15q24.1	3.60E-65	0.23	0.08	1.28E-234*	0.24	3.09E-05	0.07	5.47E-47*	0.14	CYP1A1, CYP1A2
rs4410790	17284577	C/T	chr7p21.1	5.20E-55	0.38	-0.06	7.58E-212*	0.38	4.52E-15	0.52	9.89E-60*	0.55	AGR3, AHR
rs199612805	24843991	D/I	chr22q11.23	1.80E-10	0.01	-0.10	2.48E-29	0.015	1.28E-07	0.08	6.32E-13	0.02	ADORA2A, UPB1
rs28634426	75675594	G/T	chr7q11.23	2.10E-10	0.20	0.03	3.08E-16	0.24	0.08	0.32	8.17E-06	0.27	STYXL1

rs34645063	9859107 5	D/I	chr6q16.1	3.30E -09	0.47	-0.02	5.23E-16	0.48	0.05	0.6 5	4.35E-06*	0.5 8	<i>MMS22L, POU3F2</i>
rs11474881	6289295 6	D/I	chr20q13.3 3	1.10E -08	0.55	-0.02	2.36E-12	0.55	.21	0.4 5	0.02*	0.6 2	<i>PCMTD2</i>
rs11782446 0	4137148 0	A/G	chr19q13.2	1.70E -08	0.03	-0.06	2.00e-07	0.03	0.08	0.0 1	0.08	0.0 2	<i>CTC- 490E21.1 2</i>

163



**Figure 1. GWAS and secondary analyses of coffee intake from the 23andMe cohort. A)** Manhattan plot displays seven genome-wide significant loci for coffee intake in the 23andMe cohort ( $N=130,153$ ). The horizontal line represents the threshold for significance ( $p=5.00E-08$ ). Nearest protein-coding genes ( $<1\text{Mb}$ ) to significant loci are labeled. Quantile-quantile plot shown in upper left corner. For more details, see **Table 1** and **Supplementary Table 6**. **B)** Overlap of genes identified by MAGMA, H-MAGMA, S-PrediXcan, and S-MultiXcan. Genes identified by all four methods are displayed. **C)** Genes predicted to affect coffee intake identified by S-MultiXcan according to the most significantly associated biological systems. For more details, see **Supplementary Table 9**. **D)** Genes implicated in coffee intake by S-PrediXcan according to brain regions. Upregulated genes are shown in red, downregulated shown in blue. For more detail, see **Supplementary Table 10**.

## Gene-based and tissue enrichment analyses suggest coffee intake is primarily associated with gene expression in the brain

We used gene- and transcriptome-based analyses (MAGMA, H-MAGMA, S-MultiXcan/S-PrediXcan) and identified 165 target candidate genes that may be most relevant to coffee intake. MAGMA identified 31 genes implicated in coffee intake in physical proximity to GWAS loci (**Supplementary Table 7**). H-MAGMA, which maps SNPs to genes via chromatin interaction from human brain tissue, implicated 143 unique gene-tissue pairs showing expression specific to cell type (75.16% neuron [31.30% cortical neuron, 33.04% iPSC derived neurons; 35.65% midbrain dopamine neurons], 24.83% astrocyte) and developmental (48.00% fetal, 52.00% adult) (**Supplementary Table 8**). Finally, S-MultiXcan predicted significant transcriptional regulation of 40 genes implicated in coffee intake dispersed across 20 tissues (**Figure 1C**; **Supplementary Table 10**). Of the top biological systems implicated by S-MultiXcan, nine were attributed to the nervous system (brain  $N=5$ ; tibial nerve  $N=4$ ), eight to the digestive system (esophagus  $N=6$ ; pancreas  $N=1$ ; small intestine  $N=1$ ), and six to the reproductive system (testis  $N=4$ ; prostate  $N=2$ ; **Figure 1C**). Fifty percent of these genes were predicted to be downregulated in the digestive and reproductive systems, whereas 66.67% of nervous system genes were predicted to be upregulated. Cortical enrichment was further supported by S-PrediXcan (**Figure 1D**), showing that SNPs associated with coffee intake most frequently correlated with predicted gene expression in overall cortical and frontal cortical regions ( $N=4$ /tissue), as well as the putamen ( $N=5$ ). Overall, four genes (*SCAMP2*, *SCAMP5*, *MPI*, and *FAM219B*) were identified by all four methods, and six of the 165 discovered genes (*FBXO28*, *NEIL2*, *HAUS4*, *IGDCC4*, *RP11-298/3.5*, *RP11-298/3.5*) were not within 1Mb of SNPs identified by prior GWAS of coffee or caffeine traits (**Supplementary Table 11**; **Figure 1B**; **Table 2**). These novel genes have been associated with substance use (e.g., *HAUS4* and smoking initiation<sup>73</sup>), educational outcomes



(e.g., *HAUS4* and educational attainment<sup>90</sup>), and biomarkers (e.g., *FBXO28* and mean platelet volume<sup>91</sup>; *IGDCC4* and mean corpuscular volume<sup>92</sup>).

Next, we used MAGMA gene-set analysis to identify biological pathways that may be most strongly associated with coffee intake. This analysis revealed significant enrichment ( $p=4.75E-07$ ) in pathways related to the metabolism of xenobiotics or foreign substances (i.e., chemicals) (**Supplementary Table 12**).

MAGMA tissue-based enrichment analyses suggested that coffee intake was only significantly associated with brain tissue (**Supplementary Figure 9A**). More specifically, differential expression by coffee intake was enriched ( $p<9.25E-04$ ) in the frontal cortex, overall cortex, cerebellum, and cerebellar hemispheres (**Supplementary Figure 9B**; **Supplementary Table 12**), consistent with the S-PrediXcan findings (**Supplementary Table 9**).

# **Genetic correlation and polygenic score analyses of coffee intake in US- and UK-based cohorts reveal discrepant associations with health and psychiatric traits, but consistent positive associations with substance use and obesity**

To boost statistical power and identify novel genes associated with coffee intake, we sought to meta-analyze our data (**metaGWAS**) with those from the UKB using METAL<sup>93</sup>. As a preliminary step to determine the appropriateness of a meta-analysis, we examined the genetic correlation between coffee intake in the 23andMe and UKB cohorts. Surprisingly, the two datasets were only moderately correlated ( $r_g=0.63$ ,  $p=3.54E-43$ ), although all top loci ( $p<5.0E-05$ ) shared direction of effect and had similar effect strengths (**Supplementary Figure 10**). In addition, the estimated LDSC  $SNP_{h^2}$  heritability of coffee intake of our metaGWAS was slightly lower than for both the univariate GWAS (metaGWAS  $SNP_{h^2}=4.09\% \pm 0.26$  vs. 23andMe  $SNP_{h^2}=7.57\% \pm 0.59$  vs. UKB  $SNP_{h^2}=4.85\% \pm 0.33$ ; **Supplementary Table 5**). We interpreted these results as an

223 indication of cohort heterogeneity and proceeded to analyze genetic associations with coffee  
224 intake in each cohort independently.



**Figure 2. Discordant genetic and phenotypic associations with genetic disposition to coffee intake in US and UK cohorts. A)** Comparison of genetic correlations across psychiatric (light gray), anthropologic (medium gray), and health (dark gray) traits between 23andMe (lanes 1 and 2) and UKB (lanes 3 and 4). Lanes 1 and 3 show  $r_g$  values calculated by LDSC, and lanes 2 and 4 show FDR-corrected  $p$  values. Only traits for which at least one cohort was FDR-significant are displayed. For a full list of correlations and trait names, see **Supplementary Table 14**. Most signals persisted after conditioning for dietary sugar, cigarettes per day, and Alcohol Use Disorder Identification (**AUDIT**) Consumption scores using mtCOJO<sup>94</sup> (**Supplementary Tables 15-16; Supplementary Figures 11-12**). Genetic correlations for traits denoted with \* could not be calculated in both cohorts; \*\* denotes reverse coding. **B)** Phenomic associations (panel 1: PheWAS ( $p < 3.62E-05$ ), panel 2: LabWAS ( $p < 1.57E-04$ )) identified from PGS of coffee intake from 23andMe and UKB summary statistics. Only traits for which at least one cohort was FDR-significant are displayed (saturated bars=FDR significant; desaturated bars=FDR non-significant). neuro.=neurological; gen.=genitourinary; neopl.=neoplasms; sense=sense organs; derma.=dermatologic; imm.=immune. For full trait names and more detail, see **Supplementary Table 18-19**.

To further understand these discrepancies, we performed a series of genetic correlation and polygenic score analyses. First, we examined the genetic architecture of coffee intake measured in 23andMe and UKB by comparing patterns of LDSC genetic correlation ( $r_g$ ) with 317 traits across 20 health, psychiatric, and anthropologic categories from publicly available GWAS summary statistics (**Figure 2A; Supplementary Table 14**). After accounting for multiple testing, 75 traits were genetically correlated with coffee intake in the 23andMe cohort and 74 traits in the UKB cohort. These associations could be underpinned by other unmeasured factors, like sugar intake from coffee sweeteners or smoking and alcohol use<sup>95</sup>. However, these patterns of genetic correlations persisted after conditioning on dietary sugar intake, cigarettes smoked per day, and alcohol consumption measured by the Alcohol Use Disorder Identification Test (**AUDIT; Supplementary Tables 11-12; Supplementary Tables 15-16**). Strikingly, of the traits significant in at least one cohort, only 34 (29.57%) were significant in both datasets, and only 58.82% of the traits significant in both datasets shared the same direction of correlation.

Among the traits that were significant and consistent in direction for both cohorts, we observed positive genetic correlations between coffee intake and substance use phenotypes. For example, we identified positive genetic correlations with smoking initiation (23andMe:  $r_g=0.50$ ,  $p=4.74E-47$ ; UKB:  $r_g=0.12$ ,  $p=1.89E-06$ ), drinks per week (23andMe:  $r_g=0.39$ ,  $p=3.38E-28$ ; UKB:  $r_g=0.21$ ;  $p=1.39E-14$ ), and cannabis initiation (23andMe:  $r_g=0.28$ ,  $p=1.34E-08$ ; UKB:  $r_g=0.09$ ,  $p=5.61E-03$ ). The strength of genetic correlations for substance use and misuse traits significant in at least one cohort was stronger in 23andMe compared to the UKB ( $0.30\pm0.03$  vs.  $0.09\pm0.02$ ; Welch's  $t(51.97)=5.96$ ,  $p=2.23E-07$ ). For example, associations with substance use disorder and dependence traits were mostly observed in the 23andMe cohort and were weaker or not observed in the UKB, such as for tobacco use disorder, opioid use disorder, cannabis use disorder, nicotine dependence, and alcohol dependence ( $r_g=0.24$  to  $0.44$ ,  $p=6.54E-23$  to  $2.12E-03$ ), as well as externalizing (23andMe:  $r_g=0.48$ ,  $p=7.21E-41$ ; UKB:  $r_g=0.07$ ,  $p=4.37E-03$ ), which is highly correlated with substance use and misuse<sup>96</sup>. Cluster analysis showed that genetic correlations for coffee intake in both cohorts aligned more with general substance use than misuse (Supplementary Figure 13).

Metabolic traits were largely congruent in their positive genetic correlations with coffee intake in both cohorts. For example, BMI (23andMe:  $r_g=0.19$ ,  $p=1.61E-11$ ; UKB:  $r_g=0.25$ ,  $p=7.85E-26$ ) and waist-to-hip ratio (23andMe:  $r_g=0.12$ ,  $p=4.33E-04$ ; UKB:  $r_g=0.13$ ,  $p=3.96E-07$ ) were positively genetically correlated with coffee intake in both datasets. Also consistent across cohorts were the lack of significant genetic correlations with most cardiovascular and cancer traits.

The majority of traits were only significant in one cohort or showed discrepancies in the direction of association. For example, coffee intake measured in the 23andMe dataset was positively genetically correlated with anxiety-related traits ( $r_g=0.22$  to  $0.44$ ,  $p=1.41E-05$  to  $8.53E-03$ ). In contrast, all significant genetic correlations between coffee intake and anxiety-related traits in the UKB were negative ( $r_g=-0.33$  to  $-0.12$ ,  $p=5.49E-06$  to  $8.12E-03$ ), except clinically diagnosed

anxiety ( $r_g=0.17$ ,  $p=1.39E-05$ ). We also identified significant positive genetic correlations with cross-disorder, attention deficit hyperactivity disorder, schizophrenia, and anorexia ( $r_g=0.12$  to  $0.27$ ,  $p=1.00E-07$  to  $0.01$ ) that were exclusive to the 23andMe dataset, whereas these associations were not apparent or were negatively genetically correlated in the UKB ( $r_g=-0.13$  to  $0.02$ ,  $p=1.01E-04$  to  $0.55$ ). Significant positive correlations with cognitive variables, such as executive function and intelligence, were found in the UKB ( $r_g=0.13$  to  $0.23$ ,  $p=4.55E-08$  to  $8.04E-23$ ), though these were negatively genetically correlated in 23andMe ( $r_g=-0.17$  to  $-0.10$ ,  $p=7.83E-08$  to  $2.06E-03$ ). Certain correlations with physical health traits also differed between cohorts. While correlations with most gastrointestinal traits in the 23andMe cohort were positive, such as a positive genetic correlation with gastric ulcers ( $r_g=0.41$ ,  $p=3.58E-03$ ), the corresponding genetic correlations observed in the UKB were either non-significant or negative ( $r_g=-0.22$  to  $0.12$ ,  $p=1.34E-06$  to  $0.88$ ). Positive genetic correlations with chronic pain as well as back, hip, and knee pain were observed in the 23andMe dataset ( $r_g=0.12$  -  $0.26$ ,  $p=9.02E-08$  -  $3.58E-03$ ), yet only negative genetic correlations with pain traits were reported in the UKB ( $r_g=-0.22$  to  $-0.12$ ,  $p=6.23E-04$  -  $2.54E-06$ ). Across all health and psychiatric traits that were significant within each cohort, all traits showed a positive genetic correlation with coffee intake in 23andMe participants. Only 41.3% of correlations were positive in the UKB.

We observed similar discrepancies when we extended our results to a health-care system population (**Figure 2B**). We conducted PheWAS and LabWAS by testing the association between polygenic scores (**PGS**) for coffee intake derived from 23andMe or the UKB with 1,655 medical traits and biomarkers. We identified 31 PheWAS and LabWAS traits that met the 5% FDR significance threshold using the 23andMe PGS, and 24 using the UKB PGS (**Supplementary Tables 17 and 18**). Only two endocrine traits (i.e., obesity and morbid obesity) and two biomarkers related to red blood cells were consistent in significance and direction of association. Otherwise, all significant associations were observed when testing PGS generated from one cohort but not



the other. For instance, when coffee intake PGS were derived from 23andMe, among the top positive PheWAS and LabWAS associations were substance use disorders and respiratory conditions (e.g., chronic airway obstruction, emphysema, and respiratory failure) and absolute monocyte count. Among the top negative associations derived from 23andMe PGS were those with sense organs, neoplasms, certain respiratory conditions (i.e., allergic rhinitis, acute and chronic tonsillitis, chronic tonsillitis and adenoiditis), and urea nitrogen serum/plasma. When coffee intake PGS were derived from UKB, among the top positive PheWAS and LabWAS associations were endocrine and musculoskeletal disorders, as well as the two metabolic biomarkers, glycated hemoglobin A1c and glucose. The only significant negative PheWAS and LabWAS associations from UKB-derived PGS were with anxiety disorders, and biomarkers related to blood (mean corpuscular hemoglobin concentration) and metabolic (cholesterol and triglycerides in serum or blood) traits.

## Discussion

In this study, we contributed to the existing GWAS literature of coffee intake by analyzing a US population of 130,153 participants. We uncovered seven loci associated with coffee intake, most of which were in genes implicated in metabolic processes. Coffee related variants were significantly enriched in the central nervous system. Despite prior evidence that coffee intake confers health benefits, we found genetic correlations mostly with adverse outcomes in both cohorts, particularly substance use disorders and obesity-related traits, in both cohorts. Relationships with other medical, anthropologic, and psychiatric traits were inconsistent in the US and UK cohorts, suggesting that differences between populations may affect coffee intake GWAS results and its genetic relationships with other traits.

Our GWAS replicated prior associations with genes and variants implicated in coffee and caffeine intake as well as other metabolic and xenobiotic processes<sup>28</sup>, including rs2472297 near *CYP1A1/CYP1A2*<sup>18,24,26,33,46</sup> and rs4410790 near *AHR*<sup>18,23,24,26,27,46,97</sup>, even though our study

sample was smaller compared to other GWAS ( $N=125,776-373,522^{20,24,26}$ ). Gene-based analyses uncovered 165 candidate genes, including four genes that overlapped across all four analyses: *MPI*, *SCAMP2*, *SCAMP5*, and *FAM219B*, all of which have been implicated in a prior coffee GWAS<sup>18</sup>. These overlapping genes have other associations with substance use and medical biomarkers including blood pressure, hypertension, and LDL cholesterol<sup>91,98-104</sup>. We identified gene enrichment in brain tissues across the frontal cortex, putamen, and hippocampus, consistent with prior GWAS showing enrichment for SNPs associated with coffee and caffeine in the central nervous system<sup>18,20,26</sup>. This is supported by brain imaging studies across cortical and subcortical areas showing morphological<sup>105-108</sup> and functional<sup>109,110</sup> differences between those who habitually drink coffee compared to those who do not.

One of the most striking observations of this study is the breadth and magnitude of positive associations between coffee intake with substance use. It is widely believed that use of one substance heightens risk for use of other substances and that there are common genetic risk factors for any substance use<sup>111,112</sup>; coffee, which is not generally considered a drug of misuse, does not appear to be exempt from this. We identified positive genetic correlations between coffee intake and other substances (i.e., tobacco, alcohol, cannabis and opioid use), as well as relevant personality traits like externalizing behavior. The genetics of coffee intake aligned with substance consumption phenotypes, corroborating prior GWAS and twin studies<sup>113-115</sup> (but see<sup>23</sup>), but not with substance misuse. This is perhaps unsurprising because the phenotype probed by the 23andMe and UKB cohorts focuses on quantity rather than clinically-defined dependence. We and others previously demonstrated that the genetic architectures of other substance intake versus problematic use are unique<sup>43,111,116-119</sup>, and this is likely also true for coffee.

We found consistent positive genetic correlations with BMI and obesity in both 23andMe and the UKB. This is in contrast to meta-analyses of randomized control trials and epidemiological studies that found unclear effects by any coffee or decaffeinated coffee intake on waist circumference and BMI-defined obesity, and a modest inverse relationship between coffee intake



and BMI<sup>120,121</sup>. Results for these studies are highly heterogeneous, likely due to interindividual variability in the inclusion of sugary coffee additives, cultivation, roasting, and brewing conditions affecting its chemical makeup<sup>9,122</sup>, and other habits surrounding coffee intake (e.g., concurrent food intake or appetite suppression by nicotine if smoking concurrently<sup>123</sup>). This contentious relationship may also be explained by the amount of coffee intake, as greater coffee intake seems to attenuate the genetic associations with BMI and obesity<sup>49</sup>, possibly due to the appetite suppressant effects of caffeine<sup>124</sup>. Alongside accounting for other dietary intake, detailed accounting of coffee preparation, and consumptive habits formed with coffee intake, future subgroup analyses may help explain discrepant associations between the genetics and prevalence of coffee intake with BMI-related traits.

We did not recapitulate the beneficial phenotypic relationships between coffee intake and a variety of health outcomes that are generally reported by health association studies<sup>6-8,10,125-137</sup>, perhaps because our study focused on the genetic relationship between coffee intake and other medical outcomes, or because our study focused on coffee intake and not caffeine intake. At the genetic level, we find no evidence of a common genetic background that could explain the beneficial effects of coffee on 29 cancers, Alzheimer's disease/dementia/cognitive impairments, Parkinson's disease, diabetes, cirrhosis, most cardiovascular conditions, or gout. In fact, some of these associations (e.g., cardiovascular traits and type II diabetes) were positive in the 23andMe cohort but showed no significant associations in the UKB cohort. Similarly, phenome-wide analysis did not support prior cancerous, metabolic, cardiovascular, or neurological health advantages of coffee intake<sup>6-8,10,126-136,138</sup>. Although this may seem discrepant to phenotypic associations that generally report health benefits of coffee intake, recent meta-analysis of over 100 phenotypic studies on coffee intake health outcomes suggest high levels of heterogeneity across cohorts<sup>6,7</sup>, especially across geographically separated populations<sup>6</sup>.

We found many opposing relationships with the genetics of coffee intake between 23andMe and UKB. For example, genetic correlations with pain, psychiatric illnesses, and

gastrointestinal traits were positively genetically correlated with coffee intake in 23andMe, but these associations were negative in the UKB. Inversely, the UKB analysis revealed that coffee intake was positively genetically correlated with cognitive traits, such as executive function and intelligence, corroborating prior evidence<sup>139-142</sup>, yet genetic correlations with these two traits were negative in 23andMe. Multiple PheWAS associations were also discordant. When PGS were derived from 23andMe, we observed heightened odds between genetic liability for coffee intake and respiratory illnesses, ischemic heart disease, infection, and alcohol-related disorders. Higher odds for musculoskeletal and sleep conditions were mostly associated with coffee PGS generated from the UKB. Only 11 out of the 42 phenotypes associated with coffee intake PGS showed negative associations, and none of these purported health “benefits” were consistently observed in both cohorts. Whereas the coffee intake PGS from 23andMe was associated with lower odds for ear conditions, skin neoplasms, allergic rhinitis, and tonsillitis, the PGS of coffee intake from the UKB was associated with a lower risk of anxiety disorders. Also of note is that the number of positive genetic correlations and PGS associations between coffee intake and these other traits was greater when analyzed using data from the 23andMe cohort versus from the UKB, and the strength of these associations was usually stronger. Partially consistent with this, one meta-analysis of mortality found an inverse relationship between coffee intake and all-cause mortality in European but not US studies<sup>143</sup>.

Our study shows that cultural, cohort, or geographic influences could affect the inferred genetic architecture of coffee intake and its associations with other health and lifestyle outcomes. Geographic regions may have an observable influence on GWAS results<sup>144</sup>. We observed no significant differences in subtle geographic differences on coffee intake correlations using location data available in the UKB (**Supplementary Figure 14**), suggesting cultural differences may contribute more to the cohort variations we report here. There is considerable variation in how or with whom one may consume coffee that could be subject to cultural influence. Caffeinated beverage sales, for instance, suggest that coffee and carbonated caffeinated beverages are more

preferred in the US than the UK<sup>2</sup>, whereas tea is the preferred source of caffeine in the UK and may modify coffee intake<sup>2</sup> (**Supplementary Figure 15**). Higher levels of coffee intake or caffeine from high caloric beverages in the US cohort may partially explain the higher number and magnitude of negative health associations observed in the 23andMe analysis. Even across coffee beverage subtypes, the concentration of caffeine, other coffee chemical constituents, and manufacturing byproducts (e.g., plastics and metals from packaging) varies and thus may be important parameters in health associations<sup>122,145,146</sup>. A recent investigation revealed the volume of ground or instant coffee is important to the potential health effects of its intake<sup>147</sup>; instant coffee (~60 mg of caffeine per cup) is more commonplace in the UK whereas fresh brewed coffee (~85 mg of caffeine per cup<sup>20</sup>) is preferred in the US<sup>2</sup>. Cultural differences in coffee intake could help explain the divergent patterns of health and lifestyle associations between UK and US participants, though the relative contributions of culture, geography, and their interactions to these differences will need further exploration.

There are multiple caveats to consider when interpreting our findings. Firstly, our study does not address causality between coffee intake and other health and lifestyle traits. Mendelian randomization (**MR**) studies have attempted to address the exposure-outcome relationships between two traits by using genetic instruments (i.e., SNPs identified by GWAS) as proxies for exposure and associating them with an outcome of interest. For example, MR using genetic markers associated with coffee intake suggest that coffee consumption has no causal effect on obesity and endocrine disorders, despite observational studies suggesting protective effects of coffee<sup>148</sup>. Similarly, MR studies of coffee and other substance use (e.g., tobacco, alcohol, cannabis) are also contentious<sup>149,150</sup>, with evidence that inconsistencies may be driven by gene-cohort confounds such as those we found in this study<sup>151</sup>. Secondly, the phenotype examined by 23andMe was exclusively caffeinated coffee intake, with one cup defined as 5 ounces, whereas the UKB also included decaffeinated coffee and did not explicitly define the volume of one cup. The caffeine content within coffee was also not directly measured. However, secondary analysis

using summary statistics of estimated caffeine intake from any coffee subtype in the UKB<sup>20</sup> yielded remarkably similar patterns of genetic correlations as those derived from our GWAS of cups of coffee consumed (**Supplementary Figure 16, Supplemental Table 14**). This analysis presumably mitigated the relative contribution of decaffeinated coffee (3mg of caffeine per cup versus 60 to 85mg per cup of caffeinated coffee<sup>20</sup>) to the revealed genetic associations, so we do not believe the cohort discrepancies are driven by the inclusion of decaffeinated coffee drinkers in the UKB. Another consideration is the possible health effects of non-caffeine coffee components, which are comparatively under-investigated<sup>9</sup>, such as other coffee bean phytochemical and drink additives. Furthermore, while it is unlikely that the discrepancies in genetic associations are driven by age, which is similar between cohorts (approximately 53 years old in 23andMe versus 57<sup>20</sup> years old in UKB), these cohorts skew older than the population average. They are also of above average socioeconomic status<sup>152</sup> and are of European descent, limiting generalizability of our findings to a larger population. Some studies also show sex-dependent differences in coffee and caffeine metabolism and health associations with intake<sup>138,153,154</sup>, which was not examined in our study.

Overall, we present striking differences in genetic associations of coffee intake across two large cohorts of European ancestry. While some genetic signals replicate across diverse cohorts, such as our GWAS findings and the associations between coffee intake with substance use and obesity traits, other associations may be obscured by cohort or cultural differences related to the phenotype in question. Our study provides a cautionary perspective on combining large cohort datasets gathered from unique geo-cultural populations.

## Acknowledgements

We would like to thank the research participants and employees of 23andMe for making this work possible. The following members of the 23andMe Research Team contributed to this study: Stella Aslibekyan, Adam Auton, Elizabeth Babalola, Robert K. Bell, Jessica Bielenberg, Katarzyna Bryc, Emily Bullis, Daniella Coker, Gabriel Cuellar Partida, Devika Dhamija, Sayantan Das, Teresa Filshtein, Kipper Fletez-Brant, Will Freyman, Karl Heilbron, Pooja M. Gandhi, Karl Heilbron, Barry Hicks, David A. Hinds, Ethan M. Jewett, Yunxuan Jiang, Katelyn Kukar, Keng-Han Lin, Maya Lowe, Jey C. McCreight, Matthew H. McIntyre, Steven J. Micheletti, Meghan E. Moreno, Joanna L. Mountain, Priyanka Nandakumar, Elizabeth S. Noblin, Jared O'Connell, Aaron A. Petrakovitz, G. David Poznik, Morgan Schumacher, Anjali J. Shastri, Janie F. Shelton, Jingchunzi Shi, Suyash Shringarpure, Vinh Tran, Joyce Y. Tung, Xin Wang, Wei Wang, Catherine H. Weldon, Peter Wilton, Alejandro Hernandez, Corinna Wong, Christophe Toukam Tchakouté.

We would also like to thank The Externalizing Consortium for sharing the GWAS summary statistics of externalizing. The Externalizing Consortium: Principal Investigators: Danielle M. Dick, Philipp Koellinger, K. Paige Harden, Abraham A. Palmer. Lead Analysts: Richard Karlsson Linnér, Travis T. Mallard, Peter B. Barr, Sandra Sanchez-Roige. Significant Contributors: Irwin D. Waldman. The Externalizing Consortium has been supported by the National Institute on Alcohol Abuse and Alcoholism (R01AA015416 -administrative supplement), and the National Institute on Drug Abuse (R01DA050721). Additional funding for investigator effort has been provided by K02AA018755, U10AA008401, P50AA022537, as well as a European Research Council Consolidator Grant (647648 EdGe to Koellinger). The content is solely the responsibility of the authors and does not necessarily represent the official views of the above funding bodies. The Externalizing Consortium would like to thank the following groups for making the research possible: 23andMe, Add Health, Vanderbilt University Medical Center's BioVU, Collaborative Study on the Genetics of Alcoholism (COGA), the Psychiatric Genomics Consortium's Substance

Use Disorders working group, UK10K Consortium, UK Biobank, and Philadelphia Neurodevelopmental Cohort.

MVJ, SBB, and SSR are supported by funds from the California Tobacco-Related Disease Research Program (TRDRP; Grant Number T29KT0526 and T32IR5226). SBB and AAP were also supported by P50DA037844. BP, JJM, and SSR are supported by NIH/NIDA DP1DA054394. HHAT is funded through a Natural Science and Engineering Research Council PGS-D scholarship and Canadian Institutes of Health Research (CIHR) Fellowship. JYK is supported by a CIHR Canada Research Chair in Translational Neuropsychopharmacology. LKD is supported by R01 MH113362. NSC is funded through an Interdisciplinary Research Fellowship in NeuroAIDs (Grant Number R25MH081482). JM is funded by the National Institute for Health and Care Research (NIHR) Maudsley Biomedical Research Centre at South London and Maudsley NHS Foundation Trust and King's College London. The content is solely the responsibility of the authors and does not necessarily represent the official views of the National Institutes of Health.

The datasets used for the PheWAS and LabWAS analyses described were obtained from Vanderbilt University Medical Center's BioVU which is supported by numerous sources: institutional funding, private agencies, and federal grants. These include the NIH funded Shared Instrumentation Grant S10RR025141; and CTSA grants UL1TR002243, UL1TR000445, and UL1RR024975. Genomic data are also supported by investigator-led projects that include U01HG004798, R01NS032830, RC2GM092618, P50GM115305, U01HG006378, U19HL065962, R01HD074711; and additional funding sources listed at <https://victr.vumc.org/biovu-funding/>. PheWAS and LabWAS analyses used CTSA (SD, Vanderbilt Resources). This project was supported by the National Center for Research Resources, Grant UL1 RR024975-01, and is now at the National Center for Advancing Translational Sciences, Grant 2 UL1 TR000445-06.

## Author Contributions

SSR and AAP conceived the idea. PF and SLE contributed formal analyses and curation of 23andMe data. HHAT contributed to formal analyses, investigation, and data visualization. BP, AA, and NCK contributed to formal data analysis and data visualization. JJM and LVR contributed to formal analyses. JM contributed to data visualization. HHAT and SSR wrote the manuscript. All authors reviewed and edited the manuscript.

### **Competing interests**

PF and SLE are employees of 23andMe, Inc., and hold stock or stock options in 23andMe. AAP is on the scientific advisory board of Vivid Genomics for which he receives stock options.

### **Materials and Correspondence**

Correspondence and material requests should be directed to Sandra Sanchez-Roige, [sanchezroige@ucsd.edu](mailto:sanchezroige@ucsd.edu).

### **Data availability**

We will provide 23andMe summary statistics for the top 10,000 SNPs upon publication. Full GWAS summary statistics will be made available through 23andMe to qualified researchers under an agreement with 23andMe that protects the privacy of the 23andMe participants. Please visit (<https://research.23andme.com/collaborate/#dataset-access/>) for more information and to apply to access the data.



## Methods

### Study cohorts, coffee intake and univariate GWAS

23andMe. Univariate GWAS was conducted in a sample of 130,153 male and female research participants of the genetics testing company 23andMe, Inc, as previously described<sup>155</sup>. Participants provided informed consent and volunteered to participate in the research online, under a protocol approved by the external AAHRPP-accredited IRB, Ethical & Independent (E&I) Review Services. As of 2022, E&I Review Services is part of Salus IRB (<https://www.versiticlinicaltrials.org/salusirb>). During 4 months in 2015 and 14 months between 2018-2020, participant responses to the question “How many 5-ounce (cup-sized) servings of caffeinated coffee do you consume each day?” were collected as part of a larger survey. Participants categorized as of European descent by genotype data were included in the univariate GWAS (see **Supplementary Material**)<sup>156</sup>. Participant demographics are presented in **Supplementary Table 2**.

DNA extraction and genotyping were performed from saliva samples by clinical laboratories CLIA-certified and CAP-accredited by the Laboratory Corporation of America. 23andMe, Inc. conducted all quality control, imputation, and univariate genome-wide analyses as previously described (see **Supplementary Table 4** for SNPs analyzed following quality control and imputation)<sup>157,158</sup>. Variants were imputed based on an imputation panel combining 1000 Genomes Phase 3, UK10K and the Human Reference Consortium. The 23andMe pipeline removes variants of low genotyping quality (failed a Mendelian transmission test in trios ( $p < 1.00E-20$ ), failed Hardy-Weinberg test ( $p < 1.02E-20$ ), failed batch effects test (ANOVA  $p < 1.00E-20$ ), or had a call rate  $< 90\%$ ) or imputation quality ( $r^2 < 0.50$  averages across genotyping arrays or a minimum  $r^2 < 0.3$  or variants with apparent batch effects ( $p < 1.00E-50$ ))<sup>157,159</sup>. The 23andMe GWAS pipeline performs linear regression and assumes an additive model for allelic effects<sup>43,155,160,161</sup>. Only unrelated participants were included in the GWAS, which was determined using a segmental identity-by-



descent (**IBD**) estimation algorithm. Two individuals that shared more than 700 cM IBD at one or both genomic segments (~20% of the genome) were classified as related. This is the minimum threshold expected to identify first cousins in an outbred population. Age (inverse-normal transformed), sex, the top five principal genotype components, and indicator variables for genotype platforms were applied as covariates and *p*-values were corrected for genomic control.

We conducted replication in three multi-ancestral cohorts (European ancestry *N*=689,661; African American ancestry *N*=32,312; Latin American ancestry *N*=124,155; daily mg of caffeine from coffee, transformed by log<sub>10</sub>(*x*+75)). Demographic information on these cohorts is shown in **Supplementary Table 3**. Ancestry was determined by analyzing local ancestry (see **Supplementary Material**)<sup>156</sup>.

UK Biobank. Summary statistics of coffee intake (*N*=334,659) were generated from UK Biobank (**UKB**) participants. Participants provided informed consent, were of White British descent, and answered the question “How many cups of coffee do you drink each day? (Include decaffeinated coffee)”. Other previously published GWAS of coffee intake with publicly available summary statistics were not included in our meta-analysis due to differences in the way that coffee intake was measured (e.g., “How often do you drink coffee?”, “How much coffee do you consume per year?”<sup>30</sup>), or differences in ascertainment (e.g., Parkinson’s disease only<sup>31</sup>). Secondary analysis was also conducted with GWAS summary statistics of caffeine intake from coffee (*N*=373,522) in the UKB that was calculated based on the number of cups of caffeinated and decaffeinated coffee consumed<sup>20</sup>. For further information about the UKB data collection and GWAS summary statistics for coffee intake and caffeine consumed from coffee, see <http://www.nealelab.is/uk-biobank/> (field 1498, both sexes) and Said et al. 2020<sup>20</sup>, respectively.

GWAS meta-analysis. We performed sample size weighted meta-analysis of the 23andMe and UKB cohorts using METAL (version 2020-05-05)<sup>93</sup> as previously described<sup>43</sup>. A total of

491,347 participants of European ancestry and 9,551,852 SNPs passing quality control were included in this meta-analysis.

### **Gene-based analyses (MAGMA, H-MAGMA, SPrediXcan/S-MultiXcan)**

The web-based platform Functional Mapping and Annotation of Genome-Wide Association Studies (**FUMA** v1.3.8) was used to further explore the functional consequences of lead SNPs and identify prior associations in the literature. GWA significant lead SNPs of coffee cups per day consumed from the UKB were identified using FUMA. SNPs were annotated based on ANNOVAR categories, Combined Annotation Dependent Depletion scores, RegulomeDB scores, expression quantitative trait loci (**eQTLs**), and chromatin state predicted by ChromHMM. Novel coffee intake candidate genes were identified as genes not in linkage disequilibrium or within 1Mb of GWAS-significant SNPs uncovered by other GWAS of coffee and caffeine traits (e.g., coffee/caffeine intake or caffeine metabolism). These SNPs were identified using the EBI GWAS Catalog (<https://www.ebi.ac.uk/gwas/>).

*MAGMA gene-based and pathway analyses.* We used “Multi-marker Analysis of GenoMic Annotation” (MAGMA, v1.08) to conduct gene-based associations on the 23andMe GWAS summary statistics of coffee intake. SNPs were annotated to protein-coding genes using FUMA and Ensembl build v92, which accounts for SNP LD using multiple regression methods. The default settings were used, and LD was estimated using the 1000 Genomes European reference sample. The significance of associations across 19,773 genes were adjusted using Bonferroni correction for multiple testing (one-sided  $p < 2.53E-06$ ; **Supplementary Table 7**). Gene-set analysis was subsequently conducted on 10,678 gene-sets and Gene Ontology terms curated from the Molecular Signatures Database (MsigDB v7.0). Tissue-specific gene expression profiles were also assessed in 54 tissue types and 30 general tissue types across the body with average gene expression in each tissue type used as a covariate in the analysis (**Supplementary Table**

13). Using Genome-Tissue Expression (**GTEx**, v8) RNA-seq data, gene expression values were  $\log_2$  transformed values of the average Reads Per Kilobase Million (RPKM) for each tissue type (RPKM>50 were listed as 50). Significance was determined following Bonferroni correction (one-sided  $p < 9.26 \times 10^{-4}$  for 54 tissue types; one-sided  $p < 1.67 \times 10^{-3}$  for 30 general tissue types).

**H-MAGMA**. To identify neurobiologically relevant target genes, we incorporated coffee intake GWAS data with chromatin interaction profiles from human brain tissue using Hi-C coupled MAGMA (**H-MAGMA**)<sup>162</sup>.

**S-PrediXcan and S-MultiXcan**. We performed a transcriptome-wide association study (**TWAS**) using the MetaXcan package (ver0.7.5)<sup>163,164</sup> consisting of S-PrediXcan and S-MultiXcan to identify specific eQTL-linked genes associated with coffee intake. eQTLs are genomic loci that contribute to heritable variation in mRNA levels that might influence the expression of a particular gene or its neighbors. This approach uses genetic information to predict gene expression levels in various brain tissues and tests whether the predicted gene expression correlates with coffee intake. S-PrediXcan uses precomputed tissue weights from the GTEx project database (<https://www.gtexportal.org/>) as the reference transcriptome dataset via Elastic net models. As input data, we included our GWAS summary statistics, transcriptome tissue data, and covariance matrices of the SNPs within each gene model (based on HapMap SNP set; available to download at the PredictDB Data Repository) from all available tissues ( $N=49$ ). We applied a Bonferroni correction for multiple testing across all tissues ( $N=21,565$ ; **Supplementary Table 10**).

## **LDSC heritability and genetic correlations**

Linkage Disequilibrium Score regression (**LDSC**; <https://github.com/bulik/ldsc>) was used to calculate heritability ( **$SNP-h^2$** ) and genetic correlations between habitual coffee intake and other phenotypes<sup>165</sup>.  $SNP-h^2$  was calculated from publicly available, pre-computed LD scores ("eur\_w\_ld\_chr"). LDSC was also used to calculate genetic correlations ( $r_g$ ) between habitual

coffee intake and 317 selected traits informed by prior literature across the following categories: substance use and misuse, anxiety, bipolar disorder & depression, cancer, cardiovascular, diet, gastrointestinal, lifestyle, metabolic, neurological, pain, personality, other anthropologic, other health, and other psychiatric traits.

## mtCOJO

We conditioned summary statistics on traits that are correlated with coffee intake using mtCOJO<sup>94</sup> to determine if genetic associations are retained after controlling for their effects. Conditioning on GWAS summary statistics for dietary sugar intake<sup>166</sup>, Alcohol Use Disorder Identification Test (**AUDIT**) consumption<sup>43</sup>, and cigarettes smoked per day<sup>73</sup> were examined at a  $p < 1.0 \times 10^{-5}$  and clump- $r^2$  threshold of 0.10.

## Phenome and laboratory-wide association studies

We conducted phenome-wide association analyses (**PheWAS**) and Laboratory-wide association analyses (**LabWAS**) to test the association between polygenic scores (**PGS**) for coffee intake and liability across thousands of medical conditions from hospital-based cohorts. These analyses were conducted using data from the Vanderbilt University Medical Center (**VUMC**). The project was approved by the VUMC Institutional Review Board (IRB #160302, #172020, #190418). VUMC is an integrated health system with individual-level health data from electronic health record (**EHR**) data for about 3.2 million patients. The VUMC biobank contains clinical data from EHR as well as biomarkers obtained from laboratory assessments. A portion of the individuals from VUMC also have accompanying array genotyping data. This cohort, with over 72,821 patients, is called BioVU<sup>167,168</sup>.

For each of the unrelated genotyped individuals of European ancestry from BioVU, we computed polygenic scores for coffee intake using the PRS-CS “auto” version<sup>167</sup>. Genotyping and quality control for this cohort have been extensively described<sup>168,169</sup>.

Phenome-wide association analyses (PheWAS). To identify associations between the PGS for coffee and clinical phenotypes, we performed a PheWAS. We fitted a logistic regression model to each of 1,338 case/control disease phenotypes (“phecodes”) to estimate the odds of each diagnosis given the coffee PGS, while adjusting for sex, median age of the longitudinal EHR, and the first 10 PCs. Analyses were conducted using the PheWAS v0.12 R package<sup>170</sup>. We required the presence of at least two International Disease Classification codes mapped to a PheWAS disease category (Phecode Map 1.2; <https://phewascatalog.org/phecodes>) and a minimum of 100 cases for inclusion of a phecode. The disease phenotypes included 145 circulatory system, 120 genitourinary, 119 endocrine/metabolic, 125 digestive, 117 neoplasms, 91 musculoskeletal, 85 sense organs, 76 injuries & poisonings, 65 dermatological, 76 respiratory, 69 neurological, 64 mental disorders, 42 infectious diseases, 42 hematopoietic, 34 congenital anomalies, 37 symptoms, and 31 pregnancy complications.

Laboratory-wide association analyses (LabWAS). We also examined laboratory results in BioVU, which we refer to as LabWAS. We implemented the pipeline already established by Dennis, et al.<sup>169</sup>. Broadly, LabWAS uses the median, inverse normal quantile transformed age-adjusted values from the QualityLab pipeline in a linear regression to determine the association with the input coffee intake PGS variable. We controlled for the same covariates as for the PheWAS analyses, excluding median age because the pipeline corrects for age using cubic splines with 4 knots.

## **Cluster analysis**

Previous studies have shown that consumption and misuse/dependence phenotypes have a distinct genetic architecture<sup>43,111,116-119</sup>. To explore whether the coffee intake analysis clustered closer to substance intake or misuse/dependence phenotypes, we used an unsupervised machine learning hierarchical clustering algorithm known as agglomerative nesting (AGNES)<sup>167</sup> on a genetic correlation matrix of all traits. AGNES initially forms single-item clusters

that are fused together into intermediate groups until all traits are included in a single cluster<sup>171</sup>. Clusters are formed with Ward's method such that the total within cluster variance is minimized while maintaining the fewest number of clusters based on cluster dissimilarity. Dissimilarity is assessed through Euclidean Distance of each pairwise genetic correlation with another trait. The product of AGNES is a dendrogram, formed with multiple brackets called "branches". AGNES was implemented in R using the *cluster* package (ver2.1.4)<sup>167</sup>.

Clustering was conducted with summary statistics of cigarettes per day<sup>73</sup>, former smoker<sup>73</sup>, smoking initiation<sup>73</sup>, problematic opioid use<sup>161</sup>, ICD10 F17 nicotine dependence<sup>172</sup>, alcohol dependence<sup>173</sup>, AUDIT consumption<sup>116</sup>, AUDIT problems<sup>116</sup>, cannabis initiation<sup>79</sup>, cannabis use disorder<sup>118</sup>, drinks per week<sup>73</sup>, externalizing psychopathology<sup>96</sup>, Fagerström Test for Nicotine Dependence (FTND)<sup>174</sup>, general risk tolerance<sup>175</sup>, age of smoking initiation<sup>73</sup>, and opioid use disorder<sup>161</sup>. The genetic correlations of cigarettes per day, former smoker, and smoking initiation were reverse coded to show the intuitive effects against the other traits in the dendrogram.

## References

- 1 International Coffee Organization. Annual Review Coffee Year 2019/2020. (International Coffee Organization, London, 2021).
- 2 Reyes, C. M. & Cornelis, M. C. Caffeine in the Diet: Country-Level Consumption and Guidelines. *Nutrients* **10**, doi:10.3390/nu10111772 (2018).
- 3 Landais, E. *et al.* Coffee and Tea Consumption and the Contribution of Their Added Ingredients to Total Energy and Nutrient Intakes in 10 European Countries: Benchmark Data from the Late 1990s. *Nutrients* **10**, doi:10.3390/nu10060725 (2018).
- 4 Rehm, C. D., Ratliff, J. C., Riedt, C. S. & Drewnowski, A. Coffee Consumption among Adults in the United States by Demographic Variables and Purchase Location: Analyses of NHANES 2011-2016 Data. *Nutrients* **12**, doi:10.3390/nu12082463 (2020).
- 5 Camandola, S., Plick, N. & Mattson, M. P. Impact of Coffee and Cacao Purine Metabolites on Neuroplasticity and Neurodegenerative Disease. *Neurochem Res* **44**, 214-227, doi:10.1007/s11064-018-2492-0 (2019).
- 6 Grosso, G., Godos, J., Galvano, F. & Giovannucci, E. L. Coffee, Caffeine, and Health Outcomes: An Umbrella Review. *Annu Rev Nutr* **37**, 131-156, doi:10.1146/annurev-nutr-071816-064941 (2017).
- 7 Poole, R. *et al.* Coffee consumption and health: umbrella review of meta-analyses of multiple health outcomes. *BMJ* **359**, j5024, doi:10.1136/bmj.j5024 (2017).
- 8 Kositamongkol, C. *et al.* Coffee Consumption and Non-alcoholic Fatty Liver Disease: An Umbrella Review and a Systematic Review and Meta-analysis. *Front Pharmacol* **12**, 786596, doi:10.3389/fphar.2021.786596 (2021).
- 9 Socala, K., Szopa, A., Serefko, A., Poleszak, E. & Wlaz, P. Neuroprotective Effects of Coffee Bioactive Compounds: A Review. *Int J Mol Sci* **22**, doi:10.3390/ijms22010107 (2020).



711 10 Zhao, L. G. *et al.* Coffee drinking and cancer risk: an umbrella review of meta-analyses  
712 of observational studies. *BMC Cancer* **20**, 101, doi:10.1186/s12885-020-6561-9 (2020).

713 11 Freedman, N. D., Park, Y., Abnet, C. C., Hollenbeck, A. R. & Sinha, R. Association of  
714 coffee drinking with total and cause-specific mortality. *N Engl J Med* **366**, 1891-1904,  
715 doi:10.1056/NEJMoa1112010 (2012).

716 12 Svikis, D. S. *et al.* Coffee and energy drink use patterns in college freshmen:  
717 associations with adverse health behaviors and risk factors. *BMC Public Health* **22**, 594,  
718 doi:10.1186/s12889-022-13012-3 (2022).

719 13 Swan, G. E., Carmelli, D. & Cardon, L. R. Heavy consumption of cigarettes, alcohol and  
720 coffee in male twins. *J Stud Alcohol* **58**, 182-190, doi:10.15288/jsa.1997.58.182 (1997).

721 14 Treur, J. L. *et al.* Associations between smoking and caffeine consumption in two  
722 European cohorts. *Addiction* **111**, 1059-1068, doi:10.1111/add.13298 (2016).

723 15 Tang, N., Wu, Y., Ma, J., Wang, B. & Yu, R. Coffee consumption and risk of lung cancer:  
724 a meta-analysis. *Lung Cancer* **67**, 17-22, doi:10.1016/j.lungcan.2009.03.012 (2010).

725 16 Nehlig, A. Effects of Coffee on the Gastro-Intestinal Tract: A Narrative Review and  
726 Literature Update. *Nutrients* **14**, doi:10.3390/nu14020399 (2022).

727 17 Butt, M. S. & Sultan, M. T. Coffee and its consumption: benefits and risks. *Crit Rev Food*  
728 *Sci Nutr* **51**, 363-373, doi:10.1080/10408390903586412 (2011).

729 18 Kang, J. *et al.* Increased brain volume from higher cereal and lower coffee intake:  
730 shared genetic determinants and impacts on cognition and metabolism. *Cereb Cortex*  
731 **32**, 5163-5174, doi:10.1093/cercor/bhac005 (2022).

732 19 Cornelis, M. C. & van Dam, R. M. Genetic determinants of liking and intake of coffee and  
733 other bitter foods and beverages. *Sci Rep* **11**, 23845, doi:10.1038/s41598-021-03153-7  
734 (2021).

735 20 Said, M. A., van de Vegte, Y. J., Verweij, N. & van der Harst, P. Associations of  
736 Observational and Genetically Determined Caffeine Intake With Coronary Artery



737 Disease and Diabetes Mellitus. *J Am Heart Assoc* **9**, e016808,  
738 doi:10.1161/JAHA.120.016808 (2020).

739 21 Jin, T. *et al.* Interactions of Habitual Coffee Consumption by Genetic Polymorphisms  
740 with the Risk of Prediabetes and Type 2 Diabetes Combined. *Nutrients* **12**,  
741 doi:10.3390/nu12082228 (2020).

742 22 Vogel, C. F. A., Van Winkle, L. S., Esser, C. & Haarmann-Stemmann, T. The aryl  
743 hydrocarbon receptor as a target of environmental stressors - Implications for pollution  
744 mediated stress and inflammatory responses. *Redox Biol* **34**, 101530,  
745 doi:10.1016/j.redox.2020.101530 (2020).

746 23 Matoba, N. *et al.* GWAS of 165,084 Japanese individuals identified nine loci associated  
747 with dietary habits. *Nat Hum Behav* **4**, 308-316, doi:10.1038/s41562-019-0805-1 (2020).

748 24 Kennedy, O. J. *et al.* Coffee Consumption and Kidney Function: A Mendelian  
749 Randomization Study. *Am J Kidney Dis* **75**, 753-761, doi:10.1053/j.ajkd.2019.08.025  
750 (2020).

751 25 Jia, H. *et al.* GWAS of habitual coffee consumption reveals a sex difference in the  
752 genetic effect of the 12q24 locus in the Japanese population. *BMC Genet* **20**, 61,  
753 doi:10.1186/s12863-019-0763-7 (2019).

754 26 Zhong, V. W. *et al.* A genome-wide association study of bitter and sweet beverage  
755 consumption. *Hum Mol Genet* **28**, 2449-2457, doi:10.1093/hmg/ddz061 (2019).

756 27 Nakagawa-Senda, H. *et al.* A genome-wide association study in the Japanese  
757 population identifies the 12q24 locus for habitual coffee consumption: The J-MICC  
758 Study. *Sci Rep* **8**, 1493, doi:10.1038/s41598-018-19914-w (2018).

759 28 Cornelis, M. C. *et al.* Genome-wide association study of caffeine metabolites provides  
760 new insights to caffeine metabolism and dietary caffeine-consumption behavior. *Hum*  
761 *Mol Genet* **25**, 5472-5482, doi:10.1093/hmg/ddw334 (2016).

762 29 Pirastu, N. *et al.* Non-additive genome-wide association scan reveals a new gene  
763 associated with habitual coffee consumption. *Sci Rep* **6**, 31590, doi:10.1038/srep31590  
764 (2016).

765 30 Coffee and Caffeine Genetics Consortium *et al.* Genome-wide meta-analysis identifies  
766 six novel loci associated with habitual coffee consumption. *Mol Psychiatry* **20**, 647-656,  
767 doi:10.1038/mp.2014.107 (2015).

768 31 Hamza, T. H. *et al.* Genome-wide gene-environment study identifies glutamate receptor  
769 gene GRIN2A as a Parkinson's disease modifier gene via interaction with coffee. *PLoS*  
770 *Genet* **7**, e1002237, doi:10.1371/journal.pgen.1002237 (2011).

771 32 Amin, N. *et al.* Genome-wide association analysis of coffee drinking suggests  
772 association with CYP1A1/CYP1A2 and NRCAM. *Mol Psychiatry* **17**, 1116-1129,  
773 doi:10.1038/mp.2011.101 (2012).

774 33 Sulem, P. *et al.* Sequence variants at CYP1A1-CYP1A2 and AHR associate with coffee  
775 consumption. *Hum Mol Genet* **20**, 2071-2077, doi:10.1093/hmg/ddr086 (2011).

776 34 Yang, A., Palmer, A. A. & de Wit, H. Genetics of caffeine consumption and responses to  
777 caffeine. *Psychopharmacology (Berl)* **211**, 245-257, doi:10.1007/s00213-010-1900-1  
778 (2010).

779 35 Raja, A. M. *et al.* Genetic Susceptibility to Chronic Liver Disease in Individuals from  
780 Pakistan. *Int J Mol Sci* **21**, doi:10.3390/ijms21103558 (2020).

781 36 Gao, H. *et al.* Association of GCKR Gene Polymorphisms with the Risk of Nonalcoholic  
782 Fatty Liver Disease and Coronary Artery Disease in a Chinese Northern Han Population.  
783 *J Clin Transl Hepatol* **7**, 297-303, doi:10.14218/JCTH.2019.00030 (2019).

784 37 Cai, W. *et al.* Genetic polymorphisms associated with nonalcoholic fatty liver disease in  
785 Uyghur population: a case-control study and meta-analysis. *Lipids Health Dis* **18**, 14,  
786 doi:10.1186/s12944-018-0877-3 (2019).

787 38 Xiao, X. *et al.* Functional POR A503V is associated with the risk of bladder cancer in a  
788 Chinese population. *Sci Rep* **5**, 11751, doi:10.1038/srep11751 (2015).

789 39 Menashe, I. *et al.* Large-scale pathway-based analysis of bladder cancer genome-wide  
790 association data from five studies of European background. *PLoS One* **7**, e29396,  
791 doi:10.1371/journal.pone.0029396 (2012).

792 40 Rotunno, M. *et al.* Phase I metabolic genes and risk of lung cancer: multiple  
793 polymorphisms and mRNA expression. *PLoS One* **4**, e5652,  
794 doi:10.1371/journal.pone.0005652 (2009).

795 41 Aldrich, M. C. *et al.* CYP1A1/2 haplotypes and lung cancer and assessment of  
796 confounding by population stratification. *Cancer Res* **69**, 2340-2348, doi:10.1158/0008-  
797 5472.CAN-08-2576 (2009).

798 42 Thompson, A. *et al.* Functional validity, role, and implications of heavy alcohol  
799 consumption genetic loci. *Sci Adv* **6**, eaay5034, doi:10.1126/sciadv.aay5034 (2020).

800 43 Sanchez-Roige, S. *et al.* Genome-Wide Association Study Meta-Analysis of the Alcohol  
801 Use Disorders Identification Test (AUDIT) in Two Population-Based Cohorts. *Am J*  
802 *Psychiatry* **176**, 107-118, doi:10.1176/appi.ajp.2018.18040369 (2019).

803 44 Clarke, T. K. *et al.* Genome-wide association study of alcohol consumption and genetic  
804 overlap with other health-related traits in UK Biobank (N=112 117). *Mol Psychiatry* **22**,  
805 1376-1384, doi:10.1038/mp.2017.153 (2017).

806 45 Yin, B., Wang, X., Huang, T. & Jia, J. Shared Genetics and Causality Between  
807 Decaffeinated Coffee Consumption and Neuropsychiatric Diseases: A Large-Scale  
808 Genome-Wide Cross-Trait Analysis and Mendelian Randomization Analysis. *Front*  
809 *Psychiatry* **13**, 910432, doi:10.3389/fpsy.2022.910432 (2022).

810 46 Coffee *et al.* Genome-wide meta-analysis identifies six novel loci associated with  
811 habitual coffee consumption. *Mol Psychiatry* **20**, 647-656, doi:10.1038/mp.2014.107  
812 (2015).

813 47 Xu, J. *et al.* Assessing the Association between Important Dietary Habits and  
814 Osteoporosis: A Genetic Correlation and Two-Sample Mendelian Randomization Study.  
815 *Nutrients* **14**, doi:10.3390/nu14132656 (2022).

816 48 Qi, X. *et al.* Evaluating the Effects of Diet-Gut Microbiota Interactions on Sleep Traits  
817 Using the UK Biobank Cohort. *Nutrients* **14**, doi:10.3390/nu14061134 (2022).

818 49 Wang, T. *et al.* Habitual coffee consumption and genetic predisposition to obesity: gene-  
819 diet interaction analyses in three US prospective studies. *BMC Med* **15**, 97,  
820 doi:10.1186/s12916-017-0862-0 (2017).

821 50 Yuan, S., Daghlis, I. & Larsson, S. C. Alcohol, coffee consumption, and smoking in  
822 relation to migraine: a bidirectional Mendelian randomization study. *Pain* **163**, e342-  
823 e348, doi:10.1097/j.pain.0000000000002360 (2022).

824 51 Laaboub, N. *et al.* Associations Between High Plasma Methylxanthine Levels, Sleep  
825 Disorders and Polygenic Risk Scores of Caffeine Consumption or Sleep Duration in a  
826 Swiss Psychiatric Cohort. *Front Psychiatry* **12**, 756403, doi:10.3389/fpsyt.2021.756403  
827 (2021).

828 52 Song, M. Y. & Park, S. Association of Polygenetic Risk Scores Related to Immunity and  
829 Inflammation with Hyperthyroidism Risk and Interactions between the Polygenetic  
830 Scores and Dietary Factors in a Large Cohort. *J Thyroid Res* **2021**, 7664641,  
831 doi:10.1155/2021/7664641 (2021).

832 53 Kim, E. J. *et al.* Coffee Consumption and Incident Tachyarrhythmias: Reported Behavior,  
833 Mendelian Randomization, and Their Interactions. *JAMA Intern Med* **181**, 1185-1193,  
834 doi:10.1001/jamainternmed.2021.3616 (2021).

835 54 Ellison-Barnes, A., Johnson, S. & Gudzone, K. Trends in Obesity Prevalence Among  
836 Adults Aged 18 Through 25 Years, 1976-2018. *JAMA* **326**, 2073-2074,  
837 doi:10.1001/jama.2021.16685 (2021).

838 55 Cole, J. B., Florez, J. C. & Hirschhorn, J. N. Comprehensive genomic analysis of dietary  
839 habits in UK Biobank identifies hundreds of genetic associations. *Nat Commun* **11**, 1467,  
840 doi:10.1038/s41467-020-15193-0 (2020).

841 56 Zhou, H. *et al.* Genome-wide meta-analysis of problematic alcohol use in 435,563  
842 individuals yields insights into biology and relationships with other traits. *Nat Neurosci*  
843 **23**, 809-818, doi:10.1038/s41593-020-0643-5 (2020).

844 57 Liu, M. *et al.* Association studies of up to 1.2 million individuals yield new insights into  
845 the genetic etiology of tobacco and alcohol use. *Nat Genet* **51**, 237-244,  
846 doi:10.1038/s41588-018-0307-5 (2019).

847 58 Pardinas, A. F. *et al.* Pharmacogenomic Variants and Drug Interactions Identified  
848 Through the Genetic Analysis of Clozapine Metabolism. *Am J Psychiatry* **176**, 477-486,  
849 doi:10.1176/appi.ajp.2019.18050589 (2019).

850 59 Liu, H. *et al.* Epigenomic and transcriptomic analyses define core cell types, genes and  
851 targetable mechanisms for kidney disease. *Nat Genet* **54**, 950-962, doi:10.1038/s41588-  
852 022-01097-w (2022).

853 60 Stanzick, K. J. *et al.* Discovery and prioritization of variants and genes for kidney  
854 function in >1.2 million individuals. *Nat Commun* **12**, 4350, doi:10.1038/s41467-021-  
855 24491-0 (2021).

856 61 Hellwege, J. N. *et al.* Mapping eGFR loci to the renal transcriptome and phenome in the  
857 VA Million Veteran Program. *Nat Commun* **10**, 3842, doi:10.1038/s41467-019-11704-w  
858 (2019).

859 62 Wuttke, M. *et al.* A catalog of genetic loci associated with kidney function from analyses  
860 of a million individuals. *Nat Genet* **51**, 957-972, doi:10.1038/s41588-019-0407-x (2019).

861 63 Yin, X. *et al.* Genome-wide association studies of metabolites in Finnish men identify  
862 disease-relevant loci. *Nat Commun* **13**, 1644, doi:10.1038/s41467-022-29143-5 (2022).

863 64 Casanova, F. *et al.* A genome-wide association study implicates multiple mechanisms  
864 influencing raised urinary albumin-creatinine ratio. *Hum Mol Genet* **28**, 4197-4207,  
865 doi:10.1093/hmg/ddz243 (2019).

866 65 Tin, A. *et al.* Target genes, variants, tissues and transcriptional pathways influencing  
867 human serum urate levels. *Nat Genet* **51**, 1459-1474, doi:10.1038/s41588-019-0504-x  
868 (2019).

869 66 Pazoki, R. *et al.* GWAS for urinary sodium and potassium excretion highlights pathways  
870 shared with cardiovascular traits. *Nat Commun* **10**, 3653, doi:10.1038/s41467-019-  
871 11451-y (2019).

872 67 Zanetti, D. *et al.* Identification of 22 novel loci associated with urinary biomarkers of  
873 albumin, sodium, and potassium excretion. *Kidney Int* **95**, 1197-1208,  
874 doi:10.1016/j.kint.2018.12.017 (2019).

875 68 Haas, M. E. *et al.* Genetic Association of Albuminuria with Cardiometabolic Disease and  
876 Blood Pressure. *Am J Hum Genet* **103**, 461-473, doi:10.1016/j.ajhg.2018.08.004 (2018).

877 69 Jimenez, A., Adisa, A., Woodham, C. & Saleh, M. Determination of polycyclic aromatic  
878 hydrocarbons in roasted coffee. *J Environ Sci Health B* **49**, 828-835,  
879 doi:10.1080/03601234.2014.938552 (2014).

880 70 Teumer, A. *et al.* Genome-wide association meta-analyses and fine-mapping elucidate  
881 pathways influencing albuminuria. *Nat Commun* **10**, 4130, doi:10.1038/s41467-019-  
882 11576-0 (2019).

883 71 Fredholm, B. B., Chen, J. F., Masino, S. A. & Vaugeois, J. M. Actions of adenosine at its  
884 receptors in the CNS: insights from knockouts and drugs. *Annu Rev Pharmacol Toxicol*  
885 **45**, 385-412, doi:10.1146/annurev.pharmtox.45.120403.095731 (2005).

886 72 Tanner, J. A. & Tyndale, R. F. Variation in CYP2A6 Activity and Personalized Medicine.  
887 *J Pers Med* **7**, doi:10.3390/jpm7040018 (2017).

888 73 Saunders, G. R. B. *et al.* Genetic diversity fuels gene discovery for tobacco and alcohol  
889 use. *Nature* **612**, 720-724, doi:10.1038/s41586-022-05477-4 (2022).

890 74 Sakaue, S. *et al.* A cross-population atlas of genetic associations for 220 human  
891 phenotypes. *Nat Genet* **53**, 1415-1424, doi:10.1038/s41588-021-00931-x (2021).

892 75 Koskeridis, F. *et al.* Pleiotropic genetic architecture and novel loci for C-reactive protein  
893 levels. *Nat Commun* **13**, 6939, doi:10.1038/s41467-022-34688-6 (2022).

894 76 Deak, J. D. *et al.* Genome-wide association study in individuals of European and African  
895 ancestry and multi-trait analysis of opioid use disorder identifies 19 independent  
896 genome-wide significant risk loci. *Mol Psychiatry* **27**, 3970-3979, doi:10.1038/s41380-  
897 022-01709-1 (2022).

898 77 Pasman, J. A. *et al.* Genetic Risk for Smoking: Disentangling Interplay Between Genes  
899 and Socioeconomic Status. *Behav Genet* **52**, 92-107, doi:10.1007/s10519-021-10094-4  
900 (2022).

901 78 Xu, K. *et al.* Genome-wide association study of smoking trajectory and meta-analysis of  
902 smoking status in 842,000 individuals. *Nat Commun* **11**, 5302, doi:10.1038/s41467-020-  
903 18489-3 (2020).

904 79 Pasman, J. A. *et al.* GWAS of lifetime cannabis use reveals new risk loci, genetic  
905 overlap with psychiatric traits, and a causal influence of schizophrenia. *Nat Neurosci* **21**,  
906 1161-1170, doi:10.1038/s41593-018-0206-1 (2018).

907 80 Schumann, G. *et al.* KLB is associated with alcohol drinking, and its gene product beta-  
908 Klotho is necessary for FGF21 regulation of alcohol preference. *Proc Natl Acad Sci U S*  
909 *A* **113**, 14372-14377, doi:10.1073/pnas.1611243113 (2016).

910 81 Stringer, S. *et al.* Genome-wide association study of lifetime cannabis use based on a  
911 large meta-analytic sample of 32 330 subjects from the International Cannabis  
912 Consortium. *Transl Psychiatry* **6**, e769, doi:10.1038/tp.2016.36 (2016).



913 82 Wain, L. V. *et al.* Novel insights into the genetics of smoking behaviour, lung function,  
914 and chronic obstructive pulmonary disease (UK BiLEVE): a genetic association study in  
915 UK Biobank. *Lancet Respir Med* **3**, 769-781, doi:10.1016/S2213-2600(15)00283-0  
916 (2015).

917 83 King, C. P. *et al.* Cdh13 and AdipoQ gene knockout alter instrumental and Pavlovian  
918 drug conditioning. *Genes Brain Behav* **16**, 686-698, doi:10.1111/gbb.12382 (2017).

919 84 Drgonova, J. *et al.* Cadherin 13: human cis-regulation and selectively-altered addiction  
920 phenotypes and cerebral cortical dopamine in knockout mice. *Mol Med* **22**, 537-547,  
921 doi:10.2119/molmed.2015.00170 (2016).

922 85 Yang, J. *et al.* The contribution of rare and common variants in 30 genes to risk nicotine  
923 dependence. *Mol Psychiatry* **20**, 1467-1478, doi:10.1038/mp.2014.156 (2015).

924 86 Weber, M. *et al.* Increased polysialic acid neural cell adhesion molecule expression in  
925 human hippocampus of heroin addicts. *Neuroscience* **138**, 1215-1223,  
926 doi:10.1016/j.neuroscience.2005.11.059 (2006).

927 87 Barker, J. M., Torregrossa, M. M. & Taylor, J. R. Low prefrontal PSA-NCAM confers risk  
928 for alcoholism-related behavior. *Nat Neurosci* **15**, 1356-1358, doi:10.1038/nn.3194  
929 (2012).

930 88 Mackowiak, M. *et al.* Cocaine decreases the expression of PSA-NCAM protein and  
931 attenuates long-term potentiation via glucocorticoid receptors in the rat dentate gyrus.  
932 *Eur J Neurosci* **27**, 2928-2937, doi:10.1111/j.1460-9568.2008.06255.x (2008).

933 89 Kojetin, D. J. & Burris, T. P. REV-ERB and ROR nuclear receptors as drug targets. *Nat*  
934 *Rev Drug Discov* **13**, 197-216, doi:10.1038/nrd4100 (2014).

935 90 Lee, J. J. *et al.* Gene discovery and polygenic prediction from a genome-wide  
936 association study of educational attainment in 1.1 million individuals. *Nat Genet* **50**,  
937 1112-1121, doi:10.1038/s41588-018-0147-3 (2018).



938 91 Chen, M. H. *et al.* Trans-ethnic and Ancestry-Specific Blood-Cell Genetics in 746,667  
939 Individuals from 5 Global Populations. *Cell* **182**, 1198-1213 e1114,  
940 doi:10.1016/j.cell.2020.06.045 (2020).

941 92 Thompson, A., King, K., Morris, A. P. & Pirmohamed, M. Assessing the impact of alcohol  
942 consumption on the genetic contribution to mean corpuscular volume. *Hum Mol Genet*  
943 **30**, 2040-2051, doi:10.1093/hmg/ddab147 (2021).

944 93 Willer, C. J., Li, Y. & Abecasis, G. R. METAL: fast and efficient meta-analysis of  
945 genomewide association scans. *Bioinformatics* **26**, 2190-2191,  
946 doi:10.1093/bioinformatics/btq340 (2010).

947 94 Zhu, Z. *et al.* Causal associations between risk factors and common diseases inferred  
948 from GWAS summary data. *Nat Commun* **9**, 224, doi:10.1038/s41467-017-02317-2  
949 (2018).

950 95 Kendler, K. S., Schmitt, E., Aggen, S. H. & Prescott, C. A. Genetic and environmental  
951 influences on alcohol, caffeine, cannabis, and nicotine use from early adolescence to  
952 middle adulthood. *Arch Gen Psychiatry* **65**, 674-682, doi:10.1001/archpsyc.65.6.674  
953 (2008).

954 96 Karlsson Linner, R. *et al.* Multivariate analysis of 1.5 million people identifies genetic  
955 associations with traits related to self-regulation and addiction. *Nat Neurosci* **24**, 1367-  
956 1376, doi:10.1038/s41593-021-00908-3 (2021).

957 97 Wojcik, G. L. *et al.* Genetic analyses of diverse populations improves discovery for  
958 complex traits. *Nature* **570**, 514-518, doi:10.1038/s41586-019-1310-4 (2019).

959 98 Vuckovic, D. *et al.* The Polygenic and Monogenic Basis of Blood Traits and Diseases.  
960 *Cell* **182**, 1214-1231 e1211, doi:10.1016/j.cell.2020.08.008 (2020).

961 99 Klarin, D. *et al.* Genetics of blood lipids among ~300,000 multi-ethnic participants of the  
962 Million Veteran Program. *Nat Genet* **50**, 1514-1523, doi:10.1038/s41588-018-0222-9  
963 (2018).

964 100 Feitosa, M. F. *et al.* Novel genetic associations for blood pressure identified via gene-  
 965 alcohol interaction in up to 570K individuals across multiple ancestries. *PLoS One* **13**,  
 966 e0198166, doi:10.1371/journal.pone.0198166 (2018).

967 101 Sung, Y. J. *et al.* A Large-Scale Multi-ancestry Genome-wide Study Accounting for  
 968 Smoking Behavior Identifies Multiple Significant Loci for Blood Pressure. *Am J Hum*  
 969 *Genet* **102**, 375-400, doi:10.1016/j.ajhg.2018.01.015 (2018).

970 102 Astle, W. J. *et al.* The Allelic Landscape of Human Blood Cell Trait Variation and Links to  
 971 Common Complex Disease. *Cell* **167**, 1415-1429 e1419, doi:10.1016/j.cell.2016.10.042  
 972 (2016).

973 103 Meeks, K. A. C. *et al.* Genome-wide analyses of multiple obesity-related cytokines and  
 974 hormones informs biology of cardiometabolic traits. *Genome Med* **13**, 156,  
 975 doi:10.1186/s13073-021-00971-2 (2021).

976 104 German, C. A., Sinsheimer, J. S., Klimentidis, Y. C., Zhou, H. & Zhou, J. J. Ordered  
 977 multinomial regression for genetic association analysis of ordinal phenotypes at Biobank  
 978 scale. *Genet Epidemiol* **44**, 248-260, doi:10.1002/gepi.22276 (2020).

979 105 Pham, K. *et al.* High coffee consumption, brain volume and risk of dementia and stroke.  
 980 *Nutr Neurosci* **25**, 2111-2122, doi:10.1080/1028415X.2021.1945858 (2022).

981 106 Haller, S., Montandon, M. L., Rodriguez, C., Herrmann, F. R. & Giannakopoulos, P.  
 982 Impact of Coffee, Wine, and Chocolate Consumption on Cognitive Outcome and MRI  
 983 Parameters in Old Age. *Nutrients* **10**, doi:10.3390/nu10101391 (2018).

984 107 Araujo, L. F. *et al.* Association of Coffee Consumption with MRI Markers and Cognitive  
 985 Function: A Population-Based Study. *J Alzheimers Dis* **53**, 451-461, doi:10.3233/JAD-  
 986 160116 (2016).

987 108 Perlaki, G. *et al.* Coffee consumption may influence hippocampal volume in young  
 988 women. *Brain Imaging Behav* **5**, 274-284, doi:10.1007/s11682-011-9131-6 (2011).

989 109 Magalhaes, R. *et al.* Habitual coffee drinkers display a distinct pattern of brain functional  
990 connectivity. *Mol Psychiatry* **26**, 6589-6598, doi:10.1038/s41380-021-01075-4 (2021).  
991 110 Laurienti, P. J. *et al.* Dietary caffeine consumption modulates fMRI measures.  
992 *Neuroimage* **17**, 751-757 (2002).  
993 111 Deak, J. D. & Johnson, E. C. Genetics of substance use disorders: a review. *Psychol*  
994 *Med* **51**, 2189-2200, doi:10.1017/S0033291721000969 (2021).  
995 112 Vanyukov, M. M. *et al.* Common liability to addiction and "gateway hypothesis":  
996 theoretical, empirical and evolutionary perspective. *Drug Alcohol Depend* **123 Suppl 1**,  
997 S3-17, doi:10.1016/j.drugalcdep.2011.12.018 (2012).  
998 113 Chang, L. H. *et al.* Investigating the genetic and causal relationship between initiation or  
999 use of alcohol, caffeine, cannabis and nicotine. *Drug Alcohol Depend* **210**, 107966,  
1000 doi:10.1016/j.drugalcdep.2020.107966 (2020).  
1001 114 Hettema, J. M., Corey, L. A. & Kendler, K. S. A multivariate genetic analysis of the use of  
1002 tobacco, alcohol, and caffeine in a population based sample of male and female twins.  
1003 *Drug Alcohol Depend* **57**, 69-78, doi:10.1016/s0376-8716(99)00053-8 (1999).  
1004 115 Swan, G. E., Carmelli, D. & Cardon, L. R. The consumption of tobacco, alcohol, and  
1005 coffee in Caucasian male twins: a multivariate genetic analysis. *J Subst Abuse* **8**, 19-31,  
1006 doi:10.1016/s0899-3289(96)90055-3 (1996).  
1007 116 Mallard, T. T. & Sanchez-Roige, S. Dimensional Phenotypes in Psychiatric Genetics:  
1008 Lessons from Genome-Wide Association Studies of Alcohol Use Phenotypes. *Complex*  
1009 *Psychiatry* **7**, 45-48, doi:10.1159/000518863 (2021).  
1010 117 Gelernter, J. & Polimanti, R. Genetics of substance use disorders in the era of big data.  
1011 *Nat Rev Genet* **22**, 712-729, doi:10.1038/s41576-021-00377-1 (2021).  
1012 118 Johnson, E. C. *et al.* A large-scale genome-wide association study meta-analysis of  
1013 cannabis use disorder. *Lancet Psychiatry* **7**, 1032-1045, doi:10.1016/S2215-  
1014 0366(20)30339-4 (2020).

1015 119 Sanchez-Roige, S., Palmer, A. A. & Clarke, T. K. Recent Efforts to Dissect the Genetic  
1016 Basis of Alcohol Use and Abuse. *Biol Psychiatry* **87**, 609-618,  
1017 doi:10.1016/j.biopsych.2019.09.011 (2020).

1018 120 Ramli, N. N. S., Alkhaldy, A. A. & Mhd Jalil, A. M. Effects of Caffeinated and  
1019 Decaffeinated Coffee Consumption on Metabolic Syndrome Parameters: A Systematic  
1020 Review and Meta-Analysis of Data from Randomised Controlled Trials. *Medicina*  
1021 *(Kaunas)* **57**, doi:10.3390/medicina57090957 (2021).

1022 121 Lee, A. *et al.* Coffee Intake and Obesity: A Meta-Analysis. *Nutrients* **11**,  
1023 doi:10.3390/nu11061274 (2019).

1024 122 Ludwig, I. A. *et al.* Variations in caffeine and chlorogenic acid contents of coffees: what  
1025 are we drinking? *Food Funct* **5**, 1718-1726, doi:10.1039/c4fo00290c (2014).

1026 123 Schwartz, A. & Bellissimo, N. Nicotine and energy balance: A review examining the  
1027 effect of nicotine on hormonal appetite regulation and energy expenditure. *Appetite* **164**,  
1028 105260, doi:10.1016/j.appet.2021.105260 (2021).

1029 124 Schubert, M. M. *et al.* Caffeine, coffee, and appetite control: a review. *Int J Food Sci Nutr*  
1030 **68**, 901-912, doi:10.1080/09637486.2017.1320537 (2017).

1031 125 Cornelis, M. C., Bennett, D. A., Weintraub, S., Schneider, J. A. & Morris, M. C. Caffeine  
1032 Consumption and Dementia: Are Lewy Bodies the Link? *Ann Neurol* **91**, 834-846,  
1033 doi:10.1002/ana.26349 (2022).

1034 126 Di Maso, M., Boffetta, P., Negri, E., La Vecchia, C. & Bravi, F. Caffeinated Coffee  
1035 Consumption and Health Outcomes in the US Population: A Dose-Response Meta-  
1036 Analysis and Estimation of Disease Cases and Deaths Avoided. *Adv Nutr* **12**, 1160-  
1037 1176, doi:10.1093/advances/nmaa177 (2021).

1038 127 Mentis, A. A., Dardiotis, E., Efthymiou, V. & Chrousos, G. P. Non-genetic risk and  
1039 protective factors and biomarkers for neurological disorders: a meta-umbrella systematic  
1040 review of umbrella reviews. *BMC Med* **19**, 6, doi:10.1186/s12916-020-01873-7 (2021).

1041 128 Chen, X., Zhao, Y., Tao, Z. & Wang, K. Coffee consumption and risk of prostate cancer:  
1042 a systematic review and meta-analysis. *BMJ Open* **11**, e038902, doi:10.1136/bmjopen-  
1043 2020-038902 (2021).

1044 129 Hong, C. T., Chan, L. & Bai, C. H. The Effect of Caffeine on the Risk and Progression of  
1045 Parkinson's Disease: A Meta-Analysis. *Nutrients* **12**, doi:10.3390/nu12061860 (2020).

1046 130 Sartini, M. *et al.* Coffee Consumption and Risk of Colorectal Cancer: A Systematic  
1047 Review and Meta-Analysis of Prospective Studies. *Nutrients* **11**,  
1048 doi:10.3390/nu11030694 (2019).

1049 131 Loftfield, E. *et al.* Association of Coffee Drinking With Mortality by Genetic Variation in  
1050 Caffeine Metabolism: Findings From the UK Biobank. *JAMA Intern Med* **178**, 1086-1097,  
1051 doi:10.1001/jamainternmed.2018.2425 (2018).

1052 132 Wu, L., Sun, D. & He, Y. Coffee intake and the incident risk of cognitive disorders: A  
1053 dose-response meta-analysis of nine prospective cohort studies. *Clin Nutr* **36**, 730-736,  
1054 doi:10.1016/j.clnu.2016.05.015 (2017).

1055 133 Liu, Q. P. *et al.* Habitual coffee consumption and risk of cognitive decline/dementia: A  
1056 systematic review and meta-analysis of prospective cohort studies. *Nutrition* **32**, 628-  
1057 636, doi:10.1016/j.nut.2015.11.015 (2016).

1058 134 Kennedy, O. J. *et al.* Systematic review with meta-analysis: coffee consumption and the  
1059 risk of cirrhosis. *Aliment Pharmacol Ther* **43**, 562-574, doi:10.1111/apt.13523 (2016).

1060 135 Crippa, A., Discacciati, A., Larsson, S. C., Wolk, A. & Orsini, N. Coffee consumption and  
1061 mortality from all causes, cardiovascular disease, and cancer: a dose-response meta-  
1062 analysis. *Am J Epidemiol* **180**, 763-775, doi:10.1093/aje/kwu194 (2014).

1063 136 Ding, M., Bhupathiraju, S. N., Satija, A., van Dam, R. M. & Hu, F. B. Long-term coffee  
1064 consumption and risk of cardiovascular disease: a systematic review and a dose-  
1065 response meta-analysis of prospective cohort studies. *Circulation* **129**, 643-659,  
1066 doi:10.1161/CIRCULATIONAHA.113.005925 (2014).

1067 137 Abalo, R. Coffee and Caffeine Consumption for Human Health. *Nutrients* **13**,  
1068 doi:10.3390/nu13092918 (2021).

1069 138 Hou, C. *et al.* Medical conditions associated with coffee consumption: Disease-trajectory  
1070 and comorbidity network analyses of a prospective cohort study in UK Biobank. *Am J*  
1071 *Clin Nutr* **116**, 730-740, doi:10.1093/ajcn/nqac148 (2022).

1072 139 Paz-Graniel, I. *et al.* Association between coffee consumption and total dietary caffeine  
1073 intake with cognitive functioning: cross-sectional assessment in an elderly  
1074 Mediterranean population. *Eur J Nutr* **60**, 2381-2396, doi:10.1007/s00394-020-02415-w  
1075 (2021).

1076 140 Dong, X., Li, S., Sun, J., Li, Y. & Zhang, D. Association of Coffee, Decaffeinated Coffee  
1077 and Caffeine Intake from Coffee with Cognitive Performance in Older Adults: National  
1078 Health and Nutrition Examination Survey (NHANES) 2011-2014. *Nutrients* **12**,  
1079 doi:10.3390/nu12030840 (2020).

1080 141 Haskell-Ramsay, C. F. *et al.* The Acute Effects of Caffeinated Black Coffee on Cognition  
1081 and Mood in Healthy Young and Older Adults. *Nutrients* **10**, doi:10.3390/nu10101386  
1082 (2018).

1083 142 Johnson-Kozlow, M., Kritz-Silverstein, D., Barrett-Connor, E. & Morton, D. Coffee  
1084 consumption and cognitive function among older adults. *Am J Epidemiol* **156**, 842-850,  
1085 doi:10.1093/aje/kwf119 (2002).

1086 143 Kim, Y., Je, Y. & Giovannucci, E. Coffee consumption and all-cause and cause-specific  
1087 mortality: a meta-analysis by potential modifiers. *Eur J Epidemiol* **34**, 731-752,  
1088 doi:10.1007/s10654-019-00524-3 (2019).

1089 144 Abdellaoui, A., Dolan, C. V., Verweij, K. J. H. & Nivard, M. G. Gene-environment  
1090 correlations across geographic regions affect genome-wide association studies. *Nat*  
1091 *Genet* **54**, 1345-1354, doi:10.1038/s41588-022-01158-0 (2022).

1092 145 De Toni, L. *et al.* Phthalates and heavy metals as endocrine disruptors in food: A study  
1093 on pre-packed coffee products. *Toxicol Rep* **4**, 234-239,  
1094 doi:10.1016/j.toxrep.2017.05.004 (2017).

1095 146 Sakaki, J. R., Melough, M. M., Provatas, A. A., Perkins, C. & Chun, O. K. Evaluation of  
1096 estrogenic chemicals in capsule and French press coffee using ultra-performance liquid  
1097 chromatography with tandem mass spectrometry. *Toxicol Rep* **7**, 1020-1024,  
1098 doi:10.1016/j.toxrep.2020.08.015 (2020).

1099 147 Chieng, D. *et al.* The impact of coffee subtypes on incident cardiovascular disease,  
1100 arrhythmias, and mortality: long-term outcomes from the UK Biobank. *Eur J Prev Cardiol*  
1101 **29**, 2240-2249, doi:10.1093/eurjpc/zwac189 (2022).

1102 148 Nordestgaard, A. T. Causal relationship from coffee consumption to diseases and  
1103 mortality: a review of observational and Mendelian randomization studies including  
1104 cardiometabolic diseases, cancer, gallstones and other diseases. *Eur J Nutr* **61**, 573-  
1105 587, doi:10.1007/s00394-021-02650-9 (2022).

1106 149 Treur, J. L. *et al.* Smoking and caffeine consumption: a genetic analysis of their  
1107 association. *Addict Biol* **22**, 1090-1102, doi:10.1111/adb.12391 (2017).

1108 150 Verweij, K. J. H., Treur, J. L. & Vink, J. M. Investigating causal associations between use  
1109 of nicotine, alcohol, caffeine and cannabis: a two-sample bidirectional Mendelian  
1110 randomization study. *Addiction* **113**, 1333-1338, doi:10.1111/add.14154 (2018).

1111 151 Ware, J. J. *et al.* Does coffee consumption impact on heaviness of smoking? *Addiction*  
1112 **112**, 1842-1853, doi:10.1111/add.13888 (2017).

1113 152 Sanchez-Roige, S. & Palmer, A. A. Emerging phenotyping strategies will advance our  
1114 understanding of psychiatric genetics. *Nat Neurosci* **23**, 475-480, doi:10.1038/s41593-  
1115 020-0609-7 (2020).



1116 153 Jee, H. J., Lee, S. G., Bormate, K. J. & Jung, Y. S. Effect of Caffeine Consumption on  
1117 the Risk for Neurological and Psychiatric Disorders: Sex Differences in Human. *Nutrients*  
1118 **12**, doi:10.3390/nu12103080 (2020).

1119 154 Nehlig, A. Interindividual Differences in Caffeine Metabolism and Factors Driving  
1120 Caffeine Consumption. *Pharmacol Rev* **70**, 384-411, doi:10.1124/pr.117.014407 (2018).

1121 155 Sanchez-Roige, S. *et al.* CADM2 is implicated in impulsive personality traits by genome-  
1122 and phenome-wide association studies in humans, with further support from studies of  
1123 Cadm2 mutant mice. *Translational Psychiatry* **13**, 167, doi:10.1038/s41398-023-02453-y  
1124 (2023).

1125 156 Durand, E. Y., Do, C. B., Mountain, J. L. & Macpherson, J. M. Ancestry Composition: A  
1126 Novel, Efficient Pipeline for Ancestry Deconvolution. *bioRxiv*. doi:10.1101/010512  
1127 (2014).

1128 157 Hyde, C. L. *et al.* Identification of 15 genetic loci associated with risk of major depression  
1129 in individuals of European descent. *Nat Genet* **48**, 1031-1036, doi:10.1038/ng.3623  
1130 (2016).

1131 158 Eriksson, N. *et al.* Web-based, participant-driven studies yield novel genetic associations  
1132 for common traits. *PLoS Genet* **6**, e1000993, doi:10.1371/journal.pgen.1000993 (2010).

1133 159 Lam, M. *et al.* Large-Scale Cognitive GWAS Meta-Analysis Reveals Tissue-Specific  
1134 Neural Expression and Potential Nootropic Drug Targets. *Cell Rep* **21**, 2597-2613,  
1135 doi:10.1016/j.celrep.2017.11.028 (2017).

1136 160 Sanchez-Roige, S. *et al.* Genome-wide association study of delay discounting in 23,217  
1137 adult research participants of European ancestry. *Nat Neurosci* **21**, 16-18,  
1138 doi:10.1038/s41593-017-0032-x (2018).

1139 161 Sanchez-Roige, S. *et al.* Genome-wide association study of problematic opioid  
1140 prescription use in 132,113 23andMe research participants of European ancestry. *Mol*  
1141 *Psychiatry* **26**, 6209-6217, doi:10.1038/s41380-021-01335-3 (2021).



1142 162 Sey, N. Y. A. *et al.* A computational tool (H-MAGMA) for improved prediction of brain-  
1143 disorder risk genes by incorporating brain chromatin interaction profiles. *Nat Neurosci*  
1144 **23**, 583-593, doi:10.1038/s41593-020-0603-0 (2020).

1145 163 Barbeira, A. N. *et al.* Exploring the phenotypic consequences of tissue specific gene  
1146 expression variation inferred from GWAS summary statistics. *Nat Commun* **9**, 1825,  
1147 doi:10.1038/s41467-018-03621-1 (2018).

1148 164 Barbeira, A. N. *et al.* Integrating predicted transcriptome from multiple tissues improves  
1149 association detection. *PLoS Genet* **15**, e1007889, doi:10.1371/journal.pgen.1007889  
1150 (2019).

1151 165 Bulik-Sullivan, B. K. *et al.* LD Score regression distinguishes confounding from  
1152 polygenicity in genome-wide association studies. *Nat Genet* **47**, 291-295,  
1153 doi:10.1038/ng.3211 (2015).

1154 166 Meddens, S. F. W. *et al.* Genomic analysis of diet composition finds novel loci and  
1155 associations with health and lifestyle. *Mol Psychiatry* **26**, 2056-2069,  
1156 doi:10.1038/s41380-020-0697-5 (2021).

1157 167 Ge, T., Chen, C. Y., Ni, Y., Feng, Y. A. & Smoller, J. W. Polygenic prediction via  
1158 Bayesian regression and continuous shrinkage priors. *Nat Commun* **10**, 1776,  
1159 doi:10.1038/s41467-019-09718-5 (2019).

1160 168 Roden, D. M. *et al.* Development of a large-scale de-identified DNA biobank to enable  
1161 personalized medicine. *Clin Pharmacol Ther* **84**, 362-369, doi:10.1038/clpt.2008.89  
1162 (2008).

1163 169 Dennis, J. *et al.* Genetic risk for major depressive disorder and loneliness in sex-specific  
1164 associations with coronary artery disease. *Mol Psychiatry* **26**, 4254-4264,  
1165 doi:10.1038/s41380-019-0614-y (2021).

1166 170 Carroll, R. J., Bastarache, L. & Denny, J. C. R PheWAS: data analysis and plotting tools  
1167 for phenome-wide association studies in the R environment. *Bioinformatics* **30**, 2375-  
1168 2376, doi:10.1093/bioinformatics/btu197 (2014).

1169 171 Kaufman, L. & Rousseeuw, P. J. *Finding groups in data : an introduction to cluster*  
1170 *analysis*. (Wiley, 1990).

1171 172 Watanabe, K. *et al.* A global overview of pleiotropy and genetic architecture in complex  
1172 traits. *Nat Genet* **51**, 1339-1348, doi:10.1038/s41588-019-0481-0 (2019).

1173 173 Walters, R. K. *et al.* Transancestral GWAS of alcohol dependence reveals common  
1174 genetic underpinnings with psychiatric disorders. *Nat Neurosci* **21**, 1656-1669,  
1175 doi:10.1038/s41593-018-0275-1 (2018).

1176 174 Quach, B. C. *et al.* Expanding the genetic architecture of nicotine dependence and its  
1177 shared genetics with multiple traits. *Nat Commun* **11**, 5562, doi:10.1038/s41467-020-  
1178 19265-z (2020).

1179 175 Karlsson Linner, R. *et al.* Genome-wide association analyses of risk tolerance and risky  
1180 behaviors in over 1 million individuals identify hundreds of loci and shared genetic  
1181 influences. *Nat Genet* **51**, 245-257, doi:10.1038/s41588-018-0309-3 (2019).

1182 176 Barkley-Levenson, A. M., Lagarda, F. A. & Palmer, A. A. Glyoxalase 1 (GLO1) Inhibition  
1183 or Genetic Overexpression Does Not Alter Ethanol's Locomotor Effects: Implications for  
1184 GLO1 as a Therapeutic Target in Alcohol Use Disorders. *Alcohol Clin Exp Res* **42**, 869-  
1185 878, doi:10.1111/acer.13623 (2018).

1186 177 Distler, M. G. *et al.* Glyoxalase 1 increases anxiety by reducing GABAA receptor agonist  
1187 methylglyoxal. *J Clin Invest* **122**, 2306-2315, doi:10.1172/JCI61319 (2012).

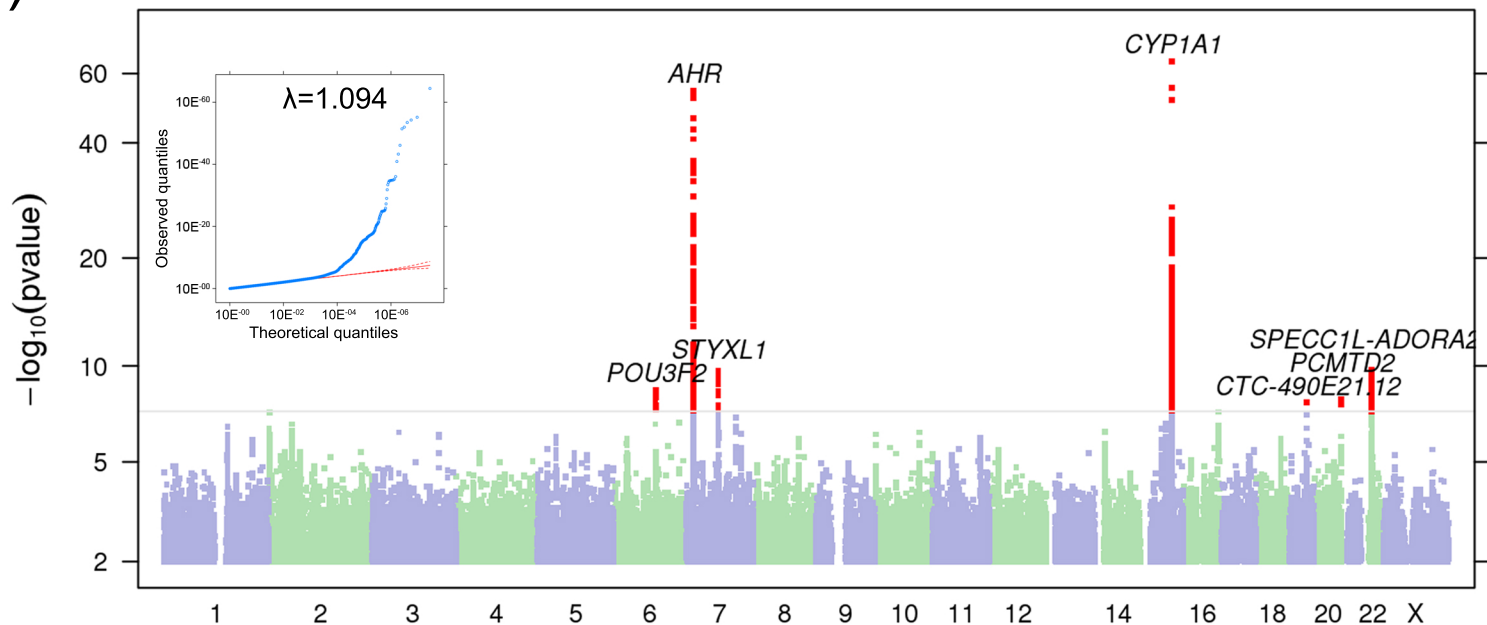
1188 178 Dawson, G. R. & Tricklebank, M. D. Use of the elevated plus maze in the search for  
1189 novel anxiolytic agents. *Trends Pharmacol Sci* **16**, 33-36, doi:10.1016/s0165-  
1190 6147(00)88973-7 (1995).

1191 179 Henn, B. M. *et al.* Cryptic distant relatives are common in both isolated and  
1192 cosmopolitan genetic samples. *PLoS One* **7**, e34267, doi:10.1371/journal.pone.0034267  
1193 (2012).

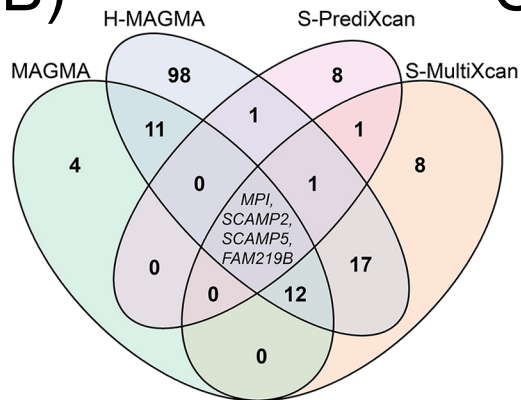
1194 180 Fuchsberger, C., Abecasis, G. R. & Hinds, D. A. minimac2: faster genotype imputation.  
1195 *Bioinformatics* **31**, 782-784, doi:10.1093/bioinformatics/btu704 (2015).

1196 181 Pruim, R. J. *et al.* LocusZoom: regional visualization of genome-wide association scan  
1197 results. *Bioinformatics* **26**, 2336-2337, doi:10.1093/bioinformatics/btq419 (2010).

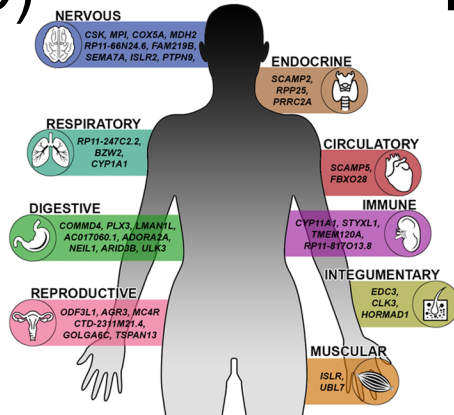
A)



B)



C)



D)

