# Genetics & Molecular Medicine

# Genome-wide Association Study of Caffeine Consumption Using Coriell Personalized Medicine Collaborative Data

## Dara M. Kusic<sup>1\*</sup>, Stefan C. Zajic<sup>1,2</sup>, Neda Gharani<sup>1,3</sup>, Erynn S. Gordon<sup>1,4</sup>, Tara Schmidlen<sup>1,5</sup>, Pan Zhang<sup>1,6</sup>, Norman Gerry<sup>1,7</sup> and Laura B. Scheinfeldt<sup>1</sup>

<sup>1</sup>*Faculty Research, Coriell Institute for Medical Research, 403 Haddon Avenue, Camden, New Jersey.* 

<sup>2</sup>Clinical Pharmacology Modeling & Simulation – Oncology, GSK, 1250 S. Collegeville Road, Collegeville, Pennsylvania.

<sup>3</sup>Director, Gharani Consulting, Weybridge, Surrey, United Kingdom.

<sup>4</sup>Director, Ripple Genetics Consulting, Mt. Laurel, New Jersey.

<sup>5</sup>Clinical Content Lead, Nest Genomics, 1887 Whitney Mesa Dr. #9919, Henderson, NV 89014.

<sup>6</sup>*Principal Research Scientist, Amazon Lab126, 1100 Enterprise Way, Sunnyvale, California.* 

<sup>7</sup>Assistant Clinical Laboratory Director, Advanced Biomedical Laboratories, 1605 Industrial Hwy, Cinnaminson, New Jersey.

#### \*Correspondence:

Dara M. Kusic, Faculty Research, Coriell Institute for Medical Research, 403 Haddon Avenue, Camden, New Jersey.

Received: 09 Apr 2023; Accepted: 21 May 2023; Published: 26 May 2023

**Citation:** Kusic DM, Zajic SC, Gharani N et al. Genome-wide Association Study of Caffeine Consumption Using Coriell Personalized Medicine Collaborative Data. Genet Mol Med. 2023; 5(1): 1-10.

#### ABSTRACT

More than 90% of Americans consume caffeine daily, but the amount consumed varies widely. Previous studies have shown that genes related to caffeine metabolism and receptor binding may influence caffeine consumption. A better understanding of the relationship between genotype and caffeine consumption has the potential to inform mechanisms of this common dietary behavior.

We performed a genome-wide association study (GWAS) of self-reported consumption data combining a wide range of caffeinated beverages including coffee, tea, soda, and energy drinks in the Coriell Personalized Medicine Collaborative (N=2,830) using a generalized linear model. Our analysis identified several candidate loci implicated in caffeine consumption. One previously identified SNP associated with caffeine consumption and replicated in the current study, rs2472297 (P-value=3.8x10-6), is located in an intergenic region between CYP1A1 and CYP1A2 (the gene that encodes the enzyme primarily responsible for caffeine metabolism), and has been associated with caffeine consumption in prior studies. Our results further support follow-up functional studies focused on the role of CYP1A1 and CYP1A2 on caffeine consumption habits, metabolism, and health status and outcomes.

#### Keywords

Caffeine, Coffee, CPMC, GWAS, CYP1A1, CYP1A2, rs2472297.

#### Introduction

Caffeine is a widely consumed compound with metabolic and psychoactive effects upon the body [1-3]. There is significant interindividual variability in the consumption of and pharmacodynamic effects of caffeine [4-7]. Published findings have reported a range of both beneficial and deleterious health effects that often depend on the amount of caffeine consumed, accounting for related factors such as age, gender, and smoking status [8-18]. Many of the health studies evaluating the impact of caffeine have focused on coffee consumption. Some of the health conditions that appear to be modulated by coffee intake include hypertension, cardiovascular disease (CVD), type 2 diabetes (T2D), hepatocellular carcinoma (HCC), endometrial cancer, melanoma, and nonmelanoma skin cancer [15].

In a recent systematic review by Kanbay and colleagues (2021), coffee consumption was associated with a decrease in risk for chronic kidney disease as well as a lower risk of end stage kidney disease [19]. Studies that have evaluated the impact of coffee consumption on hypertension have identified an inverse association whereby the risk of hypertension decreases in response to increased daily coffee [20]. An inverse linear relationship between risk and coffee consumption has been found across multiple diseases including type 2 diabetes and malignant melanoma [21,22].

Unlike other conditions where a consistent linear relationship has been seen, a recent study of the association between coffee consumption and atrial fibrillation (afib) found that afib risk was decreased in individuals with intermediate consumption of coffee (1-7 cups per week) but not for higher levels of coffee consumption [23].

Genome-wide association studies (GWAS) have implicated several single-nucleotide- polymorphisms (SNPs) in caffeine consumption [24-27]. However, a more complete understanding of the genetic contributions to caffeine consumption is needed. To address this need, we performed a GWAS using self-reported consumption data combining a wide range of caffeinated beverages including coffee, tea, soda, and energy drinks collected from 2,830 participants in the Coriell Personalized Medicine Collaborative.

#### Materials and Methods Subjects

The CPMC was a longitudinal prospective study investigating the impact of personalized genetic risk reports for common complex diseases and drug metabolism on health behavior and outcomes [28,29]. As part of the CPMC, participants completed a variety of initial and follow-up questionnaires, which allowed researchers to evaluate not only the impact of personalized genomic reports but to identify novel genetic associations with phenotypes collected through participant surveys. The CPMC study [28-35] comprises several cohorts that were included in this caffeine consumption analysis. The 2,830 participants included in this study consisted of

a cohort recruited through the United States Air Force (n = 2,537) and a CPMC community cohort from the general population (n = 293), both recruited during the active study period from 2008 through 2017. All participants were adults (at least 18 years old) that have given written informed consent to participate in the study. No participants were excluded based on comorbidities including any health conditions related to chronic and/or preexisting conditions.

The Coriell Institute Institutional Review Board reviewed and approved protocols for each of the above-mentioned cohorts. In addition, the Institutional Review Boards of Coriell Institute and the Air Force approved their respective cohort-specific protocols (Coriell Community Cohort protocol R144, and Air Force protocol numbers R153 and F-WR-2011-0046, respectively). This work was conducted in accordance with the Helsinki Declaration of 1975, as revised in 2013.

Consented participants provided a saliva sample (Oragene method (DNA Genotek,Inc. Ottawa, Ontario, Canada)), set up an account in a secure CPMC online portal, and completed required questionnaires that captured baseline information on demographics, lifestyle, medical history, medication history, family medical history, and baseline disease risk perception. Upon completion of the questionnaires, Coriell's Clinical Laboratory Improvement Amendments (CLIA) certified genotyping lab processed DNA for genetic analysis used for personalized risk reporting. No identifying information about the participants was accessible to the study's authors during or after data collection.

#### **Genetic Data Collection**

Upon completion of the questionnaires, Coriell's Clinical Laboratory Improvement Amendments (CLIA) certified genotyping lab extracted DNA from saliva submissions. Genotyping was completed using the Affymetrix SNP 6.0 GeneChip array to obtain calls for 909,622 SNPs per sample [36]. In total, 681,624 SNPs passed internal quality control standards and analysis filters for inclusion in this caffeinated drink and coffee consumption analysis. Analysis filters included not more than 5% missing data for any given marker, no marker with a minor allele frequency < 5%, and a minimum of 90% of genotyping data available for each participant DNA sample. The allele more commonly found in the 1000 Genomes Project Phase 3 whole genome sequencing dataset corresponds to the allele we refer to as "major allele" in Table 1 [37].

Table	1:	Participa	nt self	-reporte	d demo	ographic	and	caffeinated	drink
consun	npti	ion data c	ollecte	d from C	CPMC p	articipa	nts in	this study.	

	Women	Men
	1,441 (51%)	1,389 (49%)
Hispanic or Latino	149 (10%)	103 (7%)
Not Hispanic or Latino	1,283 (89%)	1,273 (92%)
Unknown	9 (1%)	13 (1%)
Asian	60 (4%)	41 (3%)
Black or African-American	125 (9%)	70 (5%)
Native American or Alaska Native	4 (< 1%)	4 (0%)
Native Hawaiian or other Pacific Islander	4 (< 1%)	6 (0%)
White(Caucasian)	1117 (78%)	1171 (84%)

More than one race	131 (9%)	97 (7%)
Current Smoker = No	1342 (93%)	1321 (95%)
Current Smoker = Yes	97 (7%)	63 (5%)
Current Smoker = Don't Know	0 (0%)	3 (< 1%)
Average BMI (kg/m <sup>2</sup> )	$25.8\pm5$	$26.6\pm4$
Mean Age at Survey (Years)	$38.5\pm11$	$40.3\pm12$
Air Force Cohort	1,277 (89%)	1,260 (91%)
CPMC Community Cohort	164 (11%)	129 (9%)
Incidence of Regular Coffee Consumption Mean Regular Coffee Drinks per Week	1,026 (71%) 7.0 ± 9	972 (70%) 8.5 ± 10
Incidence of Regular Tea Consumption Mean Regular Tea Drinks per Week	904 (63%) 2.6 ± 6	719 (52%) 1.9 ± 5
Incidence of Sugared Soda Consumption Mean Sugared Soda Drinks per Week	557 (39%) 1.2 ± 4	648 (47%) 1.3 ± 3
Incidence of Diet Soda Consumption Mean Diet Soda Drinks per Week	533 (37%) 1.8 ± 5	581 (42%) 2.87 ± 7
Incidence of Energy Drink Consumption Mean Energy Drinks per Week	281 (20%) 0.4 ± 2	422 (30%) 0.8 ± 3
Incidence of All Caffeine Consumption Mean All Caffeine Drinks per Week	1,376 (95%) 13 ± 13	1,320 (95%) 15 ± 14

#### **Nongenetic Data Collection**

Participants completed online questionnaires including the collection of demographic information (including race, and ethnicity), medical history, family history, lifestyle, and personal health information (including: height, weight (which were used to calculate BMI), gender, cohort, age-at-survey, smoking status, and caffeinated beverage consumption habits). For each question, participants were given the option to select 'do not want to answer'. Participants were asked to select one or more of the following: White (Caucasian), Black or African-American, Native American or Alaska Native, or Asian, Native Hawaiian and/or other Pacific Islander. Participants were additionally asked to select from among Hispanic or Latino, Not Hispanic or Latino, or Unknown.

Participants were also asked about their current smoking status with choices of Yes, No, and Don't Know as possible answers. The survey asked for self-reported height in feet and inches and weight in pounds, which after conversion to SI units were used to calculate body-mass index (BMI) according to the formula,  $BMI = kg/m^2$ . Cohort was coded as a binary value indicating whether participants enrolled in the Air Force cohort or the CPMC Community cohort, the two cohorts included in this study.

The survey asked for the number of days per week and the number of drinks per day each participant consumed for each of the following beverages: regular (caffeinated) coffee, regular tea, regular (sugared) soda, diet soda, and energy drinks. The soda drinks both regular and diet were assumed to contain caffeine for the purposes of this study, as participants were asked separately about decaffeinated regular and diet soda consumption, as well as about decaffeinated coffee consumption. For this study, the reported number of days each type of caffeinated drink was consumed per week was multiplied by the number of drinks per day and summed over all types to compute the total number of caffeinated drinks per week. Our study employed association-testing models for the total number of caffeinated drinks per week, as well as the subset of coffee drinks only.

#### **Statistical Methods**

We applied a generalized linear regression model, consistent with the approach employed in previous genetic association studies of continuous measures of coffee intake [24,25,38,39], with total caffeinated drinks per week as the dependent variable and number of minor alleles at a particular SNP as the independent variable. Covariates of age-at-reporting, cohort, smoking status, BMI, and gender were included in the generalized linear model, selected for inclusion by the step Akaike Information Criteria (AIC) function from the MASS statistical package in R [40,41]. The generalized linear model also included the first ten principal components from a principal component analysis (PCA) of the genome-wide SNP data described above to correct for population stratification [42-43]. The glm() function from the stats package version 3.6.0 in the R computing platform [45] was used for the following model: caffeinated Drinks Per Week  $\sim$  genotype +age + cohort + gender + bmi + smoking + PC1:PC10. To comparatively assess the results within the same data set using a sub-phenotype, we also evaluated the association of caffeinated coffee-only drinks per week, using the glm() function in R for the following model: coffee Drinks Per Week  $\sim$  genotype +age + cohort + gender + bmi + smoking + PC1:10.

We applied the Bonferroni procedure [46] to adjust the alpha threshold to evaluate each association at the 0.05 level of significance to control for multiple hypothesis testing. The Bonferroni- adjusted alpha threshold was calculated using the number of SNPs included in the GWAS (N=681,624) to a value of  $7.3 \times 10^{-8}$ .

We additionally used PANTHER to perform pathway analysis of the top 1,000 associated SNP related genes in each GWAS (caffeine and coffee, respectively) [47-49]. In particular, we extracted the closest Ensembl gene id associated with each of the 2,000 SNPs from the annotation file associated with the Affymetrix SNP 6.0 GeneChip array, used the PANTHER Gene List tool, selected the Homo sapiens as our organism, and performed a statistical overrepresentation test of PANTHER pathways with the default settings of Fisher's Exact test type and False Discovery Rate (FDR) correction for multiple testing.

#### Results

Table 1 includes CPMC participant self-reported demographic and caffeinated drink consumption data. The median age of participants at the time of survey was 37 years, with a minimum age of 18 years and a maximum age of 84 years. Participants included 1,441 (51%) women and 1,389 (49%) men. There were 2,288 (81%) participants who self-reported race as white only, and 542 (19%) reported another or multiple races. There were 252 (9%) participants who self-identified as Hispanic or Latino, and 2,578 (92%) responded as Not Hispanic or Latino or Unknown. A total of 160 participants (5.6%) reported currently smoking. The mean BMI among participants in the study was 26 kg/m<sup>2</sup> with a standard error of 4 kg/m<sup>2</sup>.

In total, 95% of participants reported drinking some type of caffeinated beverage per week. The mean number of self-reported

caffeinated drinks per week was 14 with a standard error of 13. The most prevalent caffeinated drink consumed was coffee, with 71% of participants reporting some consumption of regular coffee. The mean number of self-reported coffee drinks per week was 7.8 with a standard error of 9.6. After coffee, the reported caffeinated drinks consumed in order of prevalence was regular tea, sugared soda, diet soda, followed by energy drinks. Across all categories of caffeinated beverage, men reported consuming a greater number of drinks per week than women.

Figure 1 displays the Manhattan plot of the  $-\log_{10}$ (P-value) of all SNPs evaluated in the generalized linear model to test for association with caffeinated drinks per week [50]. Supporting Fig. S1 shows the quartile-quartile (QQ) plot of residuals of the  $-\log_{10}$ (P-values) evaluated with the generalized linear model against what would be expected in a normal distribution (Q-Q plot  $\lambda$ =1.01). Caffeine GWAS aggregate summary statistics for all tested SNPs can be found in Supporting Table 1. In addition, coffee GWAS aggregate summary statistics for all tested SNPs can be found in Supporting Table 2. We tested the top 1,000 most associated SNPs with caffeine and coffee, respectively, for pathway statistical overrepresentation using the PANTHER Gene List tool, and did not find any significant overrepresentation after FDR correction for either gene set [47-49].

#### Manhattan Plot Affy 6.0 SNPs for Caffeine Drinks Per Week CaffDrinksPerWk~geno+age+cohort+gender+bmi+smoke+PC1:10



**Figure 1:** Manhattan plot of P-values from a generalized linear model highlighting a top association of rs2472297 with caffeine drink consumption per week. The red line depicts the genome-wide significant threshold commonly employed in GWAS, and the blue line is the genome-wide significant threshold for significance based on a Bonferroni correction using the number of tested SNPs.

We identified one locus that has been previously reported in other studies of caffeine consumption [27,51], rs2472297, as having a

suggestive association with caffeinated beverages consumed per week (P-value =  $3.8 \times 10^{-6}$ ) (Figure 1, Table 2). Rs2472297 is located between the *CYP1A1* and *CYP1A2* genes on chromosome 15. Figure 2 shows self-reported caffeine consumption in number of caffeinated drinks per week by genotype for the SNP rs2472297 (*CYP1A1/CYP1A2*) (intergenic). There is a trend toward consumption of more caffeinated drinks per week with each additional minor allele.

Caffeine Drinks Consumed per Week by rs2472297 Genotype chr.15, MAF=0.20, p-value=3.8e-06



Figure 2: Caffeine drinks consumed per week by rs2472297 genotype.

Other suggestive associations with caffeinated drink consumption in our study (P-value  $< 10^{-5}$ ) (Table 2), include a cluster of SNPs in linkage disequilibrium [37,52] on chromosome 2 (rs6546165, rs3815554, rs897875, rs12614613) that is 170K base pairs downstream of the SPRED2 gene (smallest P-value =  $1.5 \times 10^{-7}$ ), and another cluster also on chromosome 2 (rs10173369, rs13018057) that is 104K base pairs downstream of the GULP1 gene (smallest P-value =  $1.7 \times 10^{-6}$ ). Two SNPs on chromosome 18, not in linkage disequilibrium, rs6507907, 100K base pairs upstream of the LAMA3 gene, and rs572614, 77K base pairs downstream of the LAMA1 gene, associated with caffeine consumption with P-values of 1.3 x 10<sup>-6</sup> and 1.6 x 10<sup>-6</sup>, respectively. Rs1814653, 64K base pairs upstream of the MIR5087 gene, associated with a P-value of 4.0 x 10<sup>-6</sup>. Two SNPS in linkage disequilibrium on chromosome 5, approximately 200K base pairs upstream of the ENC1 gene, associated with a P-value of 4.4 x 10<sup>-6</sup>. Lastly, rs17718841 on chromosome 6, 6K base pairs upstream of the MOXD1 gene, rs2055559 on chromosome 3, 267K downstream of the ALCAM gene, two SNPs in linkage disequilibrium, rs13125107 and rs11947461, on chromosome 4 approximately 100K base pairs downstream of the RAPGEF2 gene, and rs2869682 32K base pairs downstream of the DMP1 gene on chromosome 4, associated with caffeine consumption with P-values on the order of 10<sup>-6</sup>.

<b>Table 2:</b> Top SNP associations with all caffeinated drinks consumed per week (P-value $< 10^{-5}$ ) and with coffee drinks per week (with P-value $< 10^{-5}$ ).									
dbSNP ID	Chr.	Position (HG19)	Nearest Gene/s (Distance, bp)	Alleles	MAF	P-value (all caffeine drinks per week)	P-value (coffee drinks per week)	β (all caffeine drinks per week)	β (coffee drinks per week)
rs6546165	2	65829712	SPRED2 (+170K)	G>A	0.07	1.5x10 <sup>-7</sup>	1.0x10 <sup>-</sup> 5	5.3	4.4
rs6507907	18	21353351	LAMA3 (-100K)	C>T	0.13	1.3x10 <sup>-6</sup>	6.6x10 <sup>-</sup> 4	4.8	3.4
rs572614	18	7006871	LAMA1 LINC00668 (+77K)	C>T	0.08	1.6x10 <sup>-6</sup>	1.2x10 <sup>-</sup> 2	4.8	2.5
rs10173369	2	189564693	LOC105373790 GULP1 (+104K)	A>G	0.10	1.7x10 <sup>-6</sup>	1.2x10 <sup>-</sup> 3	4.8	3.2
rs3815554	2	65818332	SPRED2 (+159K)	T>C	0.09	1.9x10 <sup>-6</sup>	1.3x10 <sup>-</sup> 4	4.8	3.8
rs897875	2	65818446	SPRED2 (+159K)	G>T	0.09	2.0x10 <sup>-6</sup>	1.4x10 <sup>-</sup> 4	4.8	3.8
rs12614613	2	65824440	SPRED2 (+165K)	G>C	0.09	3.2x10 <sup>-6</sup>	1.9x10 <sup>-</sup> 4	4.7	3.7
rs2472297	15	75027880	<i>СҮР1А1</i> (+10К) <i>СҮР1А2</i> (-13К)	G>A	0.20	3.8x10 <sup>-6</sup>	1.3x10 <sup>-</sup> 4	4.6	3.8
rs1814653	1	147743035	MIR5087 (-64K)	G>A	0.15	4.0x10 <sup>-6</sup>	1.7x10 <sup>-</sup> 2	4.6	2.4
rs463328	5	73720455	ENC1 (-203K)	C>T	0.13	4.4x10 <sup>-6</sup>	2.1x10 <sup>-</sup> 3	4.6	3.1
rs13018057	2	189572285	<i>LOC105373790 GULP1</i> (+112K)	A>G	0.10	4.7x10 <sup>-6</sup>	2.5x10 <sup>-</sup> 3	4.6	3.0
rs212280	5	73749181	ENC1 (-174K)	T>C	0.14	5.6x10 <sup>-6</sup>	2.0x10 <sup>-</sup> 3	4.5	3.1
rs17718841	6	132611452	MOXD1 (-6K)	A>G	0.06	5.6x10 <sup>-6</sup>	4.4x10 <sup>-5</sup>	4.5	4.1
rs2055559	3	105562507	ALCAM (267K)	T>C	0.05	6.8x10 <sup>-6</sup>	7.8x10 <sup>-</sup> 3	4.5	2.7
rs13125107	4	160390380	RAPGEF2 (+109K)	T>A	0.26	7.1x10 <sup>-6</sup>	2.0x10 <sup>-</sup> 3	4.5	3.1
rs11947461	4	160374255	RAPGEF2 (+93K)	G>A	0.34	8.8x10 <sup>-6</sup>	1.7x10 <sup>-</sup> 4	4.5	3.8
rs2869682	4	88620742	DMP1 (+32K)	C>A	0.29	8.9x10 <sup>-6</sup>	3.5x10 <sup>-3</sup>	-4.4	-2.9
rs7031274	9	3012276	<i>KIAA0020</i> (+168K)	G>A	0.37	3.1x10 <sup>-5</sup>	9.6x10 <sup>-</sup> 6	4.2	4.4
rs10894623	11	132795569	NTM (+589K)	C>A	0.25	1.7x10 <sup>-4</sup>	3.3x10 <sup>-</sup> 7	-3.8	-5.1
rs11906093	20	54456232	CBLN4 (- 116K)	G>A	0.34	2.1x10 <sup>-4</sup>	5.1x10 <sup>-</sup> 6	3.7	4.6
rs381324	1	110429022	<i>CSF1</i> (-24K)	C>G	0.06	7.3x10 <sup>-4</sup>	4.7x10 <sup>-</sup> 6	-3.4	-4.6
rs11620007	13	53492990	РСДН8 (+70К)	G>T	0.11	1.2x10-3	6.7x10 <sup>-</sup> 7	3.2	5.0
rs2031838	13	74239494	<i>KLF12</i> (-21K)	A>T	0.11	1.4x10 <sup>-2</sup>	3.1x10 <sup>-</sup> 6	-2.5	-4.7

#### Discussion

The CPMC study queried participants on caffeinated drink consumption from a wide range of sources, including tea, coffee, soda, and energy drinks. Studies of associations between *CYP1A2* variation, caffeine consumption, and links to diseases that have not examined all sources of caffeine besides coffee have been criticized for misclassifying the exposure of interest [53]. As we found in our study, and as is shown in Table 2, the strength of association between genetic variants and caffeine consumption varies from that with coffee consumption only, although the directionality of the association remains consistent between all caffeinated drinks and coffee only.

Past studies [28,39,51,54-58] have found associations with caffeine consumption in several genes, including ABCG2, AHR, POR, BDNF, CYP1A2, and SLC6A4, although our study only replicated a finding consistently reported near the CYP1A2 gene [51,55,56]. Rs2472297 is located at 15q24 between the CYP1A1 (OMIM 108330) and CYP1A2 (OMIM 124060) genes that share a promoter region for aryl hydrocarbon receptor (AHR) implicated in pharmacogenetics [59,60]. CYP1A2 encodes the cytochrome P450 1A2 enzyme that is responsible for partial to predominant metabolism of many drugs, including antipsychotic medications such as olanzapine and clozapine as well as caffeine [61-65]. Cigarette smoking has been shown to increase CYP1A2 activity, and CYP1A2 is more active in men than in women [6,39,66,67]. Rs2472297 has been previously associated with coffee and caffeine consumption in previous genome-wide association studies and is used as a reporting locus in the 23andMe consumer genomic reporting service [51,54-57,68]. Differential expression of the *CYP1A1* and *CYP1A2* genes after caffeine administration in lymphoblastoid cell lines, and association of SNPs in the 5' flanking regions between the two genes strongly implicates the *CYP1A1/CYP1A2* locus in response to coffee drinking [25].

Caffeine is primarily metabolized in the liver by cytochrome P450 enzymes encoded by the CYP1A1 and CYP1A2 genes. CYP1A1 metabolizes polycyclic aromatic hydrocarbons found in coffee, and CYP1A2 metabolizes caffeine by 3-N-demethylation [54,69]. CYP1A1 has been found to have increased gene expression in epithelial cells exposed to caffeine [70]. The -163 C>A CYP1A2 variant, rs762551, has been shown to affect the inducibility of CYP1A2 in heavy consumers of coffee and smokers [65], and significant differences in individual CYP1A2 enzyme activity have been associated with variation in the 5'-flanking region of CYP1A2 [71]. Rs2472297, located 14 kb upstream and implicated in our GWAS and in previous studies of caffeine consumption, lies within a 23-kb segment of the 5' flanking region between CYP1A1 and CYP1A2 that has no open reading frames [54,72]. Activated hydrocarbons that bind to xenobiotic elements of the 5' flanking region have been found to induce both CYP1A1 and CYP1A2 enzymes [72,73]. Health outcomes that have been linked to coffee, caffeine consumption, and variants in CYP1A2 include hypertension, fasting glucose, and blood pressure [74-76]. Variants in CYP1A1 and CYP1A2, in combination with variants in other xenobiotic genes, have been associated with increased risk of congenital heart disease [77].

Another suggestive association implicates rs572614 on chromosome 18, which is located in an intron region 77K base pairs downstream of the *LAMA1* gene (OMIM 150320) that encodes for the laminin A glycoprotein involved in structural scaffolding of many tissues including salivary Glands [78]. Although *LAMA1* has not been previously associated with caffeine consumption in the literature, it has been associated with metabolism more generally, specifically the isorder of Type 2 diabetes [79,80].

The relationship between coffee and human health remains unclear in its presence and extent, and it is also unclear whether any potential relationship is attributable to the caffeine or non- caffeine components of coffee. CYP1A1 and CYP1A2 are responsible for metabolizing both the caffeine and non-caffeine components of coffee and may have a modifying effect on consumption habits and effects of the compounds; our results lend support for an association between caffeine consumption and a SNP in the *CYP1A1/CYP1A2* locus [51] and support he need for a more translational understanding of the relationship between coffee and caffeine consumption dietary habits, genetic variation, and human health.

#### Conclusions

These new and replication findings add support to a growing body of literature characterizing the suggestive influence of genetics on caffeine consumption habits that, particularly when combined with a better functional metabolic and health outcome understanding, has the potential to inform lifestyle choices about consumption behavior. Further research with larger cohorts into the role of genetics in mediating the relationship between coffee and caffeine consumption and health status is of merit [81,82]. Such work has the potential to eventually inform personalized interventions aimed at improving cardiovascular, metabolic, and quality of life outcomes that may be linked to common consumption habits.

### Acknowledgements

We would like to express our great appreciation to the CPMC participants as well as the CPMC team at Coriell. We would finally like to acknowledge all of the contributions that the late Dr. Michael F. Christman has made to the CPMC.

## Funding

This work was funded in part by the United States Air Force Cooperative Agreement FA8650-14-2-6533 to Dr. Christman, Principal Investigator.

## **Conflicts of Interest**

The authors declare no conflict of interest. The funders had no role in the design of the study; in the collection, analyses, or interpretation of data; in the writing of the manuscript, or in the decision to publish the results. The content is solely the responsibility of the authors and does not necessarily represent the official views of the United States Air Force.

### **Institutional Review Board Statement**

The Coriell Institute Institutional Review Board reviewed and

approved protocols for each of the abovementioned cohorts. In addition, the Institutional Review Boards of Coriell Institute and the Air Force approved their respective cohort-specific protocols (Coriell Community Cohort protocol R144, and Air Force protocol numbers R153 and F-WR-2011-0046, respectively). This work was conducted in accordance with the Helsinki Declaration of 1975, as revised in 2013.

## References

- Ren X, Chen JF. Caffeine and Parkinson's Disease: Multiple Benefits and Emerging Mechanisms. Front Neurosci. 2020; 14: 602697.
- Nehlig A, Daval JL, Debry G. Caffeine and the central nervous system: mechanisms of action, biochemical, metabolic and psychostimulant effects. Brain Res Brain Res Rev. 1992; 17: 139-170.
- 3. Evans J, Richards JR, Battisti AS. Caffeine. StatPearls. 2021.
- Aranda JV, Beharry KD. Pharmacokinetics, pharmacodynamics and metabolism of caffeine in newborns. Semin Fetal Neonatal Med. 2020; 25: 101183.
- 5. Grant DM, Tang BK, Kalow W. Variability in caffeine metabolism. Clin Pharmacol Ther. 1983; 33: 591-602.
- 6. Rasmussen BB, Brix TH, Kyvik KO, et al. The interindividual differences in the 3-demthylation of caffeine alias CYP1A2 is determined by both genetic and environmental factors. Pharmacogenetics. 2002; 12: 473-478.
- Yang A, Palmer AA, de Wit H. Genetics of caffeine consumption and responses to caffeine. Psychopharmacology (Berl). 2010; 211: 245-257.
- Loftfield E, Cornelis MC, Caporaso N, et al. Association of Coffee Drinking With Mortality by Genetic Variation in Caffeine Metabolism: Findings From the UK Biobank. JAMA Intern Med. 2018; 178: 1086-1097.
- 9. Tsioufis C, Dimitriadis K, Vasiliadou C, et al. Heavy coffee consumption in conjunction with smoking is accompanied by increased inflammatory processes and impaired thrombosis/ fibrinolysis system in essential hypertensive subjects. J Hum Hypertens. 2006; 20: 470-472.
- McCarty MF. A chlorogenic acid-induced increase in GLP-1 production may mediate the impact of heavy coffee consumption on diabetes risk. Med Hypotheses. 2005; 64: 848-853.
- 11. Zhou A, Hypponen E. Long-term coffee consumption, caffeine metabolism genetics, and risk of cardiovascular disease: a prospective analysis of up to 347,077 individuals and 8368 cases. Am J Clin Nutr. 2019; 109: 509-516.
- 12. Wijarnpreecha K, Panjawatanan P, Mousa OY, et al. Heavy Coffee Consumption and Risk of Pancreatitis: A Systematic Review and Meta-Analysis. Dig Dis Sci. 2018; 63: 3134-3140.
- 13. Djordjevic N, Radmanovic B, Cukic J, et al. Cigarette smoking and heavy coffee consumption affecting response to olanzapine: The role of genetic polymorphism. World J Biol Psychiatry. 2020; 21: 29-52.

- 14. Kim JH, Park YS. Light coffee consumption is protective against sarcopenia, but frequent coffee consumption is associated with obesity in Korean adults. Nutr Res. 2017; 41: 97-102.
- Di Maso M, Boffetta P, Negri E, et al. Caffeinated Coffee Consumption and Health Outcomes in the US Population: A Dose-Response Meta-Analysis and Estimation of Disease Cases and Deaths Avoided. Adv Nutr. 2021; 12: 1160-1176.
- 16. Poole R, Kennedy OJ, Roderick P, et al. Coffee consumption and health: umbrella review of meta-analyses of multiple health outcomes. BMJ. 2017; 359: j5024.
- Grosso G, Godos J, Galvano F, et al. Coffee, Caffeine, and Health Outcomes: An Umbrella Review. Annu Rev Nutr. 2017; 37: 131-156.
- Visser E, Geleijnse JM, de Roos B. Inter-Individual Variation in Cancer and Cardiometabolic Health Outcomes in Response to Coffee Consumption: A Critical Review. Mol Nutr Food Res. 2020; 64: e1900479.
- 19. Kanbay M, Siriopol D, Copur S, et al. Effect of Coffee Consumption on Renal Outcome: A Systematic Review and Meta-Analysis of Clinical Studies. J Ren Nutr. 2021; 31: 5-20.
- 20. Xie C, Cui L, Zhu J, et al. Coffee consumption and risk of hypertension: a systematic review and dose-response metaanalysis of cohort studies. J Hum Hypertens. 2018; 32: 83-93.
- Ding M, Bhupathiraju SN, Chen M, et al. Caffeinated and decaffeinated coffee consumption and risk of type 2 diabetes: a systematic review and a dose-response meta-analysis. Diabetes Care. 2014; 37: 569-586.
- 22. Liu J, Shen B, Shi M, et al. Higher Caffeinated Coffee Intake Is Associated with Reduced Malignant Melanoma Risk: A Meta-Analysis Study. PLoS One. 2016; 11: e0147056.
- 23. Bazal P, Gea A, Navarro AM, et al. Caffeinated coffee consumption and risk of atrial fibrillation in two Spanish cohorts. Eur J Prev Cardiol. 2021; 28: 648-657.
- 24. Cornelis MC, Byrne EM, Esko T, et al. Genome-wide metaanalysis identifies six novel loci associated with habitual coffee consumption. Molecular Psychiatry. 2015; 20: 647-656.
- Amin N, Byrne E, Johnson J, et al. Genome-wide association analysis of coffee drinking suggests association with CYP1A1/CYP1A2 and NRCAM. Mol Psychiatry. 2012; 17: 1116-1129.
- 26. Jia H NS, Kawafune K, Hachiya T, et al. GWAS of habitual coffee consumption reveals a sex difference in the genetic effect of the 12q24 locus in the Japanese population. BMC Genetics. 2019; 20: 61.
- 27. Josse AR, Da Costa LA, Campos H, et al. Associations between polymorphisms in the AHR and CYP1A1-CYP1A2 gene regions and habitual caffeine consumption. Am J Clin Nutr. 2012; 96: 665-671.
- 28. Keller MA, Gordon ES, Stack CB, et al. Coriell Personalized Medicine Collaborative ((R)): a prospective study of the

utility of personalized medicine. Per Med. 2010; 7: 301-317.

- Stack CB, Gharani N, Gordon ES, et al. Genetic risk estimation in the Coriell Personalized Medicine Collaborative. Genet Med. 2011; 13: 131-139.
- 30. Delaney SK, Brenner R, Schmidlen TJ, et al. Precision Military Medicine: Conducting a multi-site clinical utility study of genomic and lifestyle risk factors in the United States Air Force. NPJ Genom Med. 2017; 2: 2.
- 31. Gharani N, Keller MA, Stack CB, et al. The Coriell personalized medicine collaborative pharmacogenomics appraisal, evidence scoring and interpretation system. Genome Med. 2013; 5: 93.
- 32. Scheinfeldt LB, Gharani N, Kasper RS, et al. Using the Coriell Personalized Medicine Collaborative Data to conduct a genome-wide association study of sleep duration. Am J Med Genet B Neuropsychiatr Genet. 2015; 168: 697-705.
- Schmidlen TJ, Scheinfeldt L, Zhaoyang R, et al. Genetic Knowledge Among Participants in the Coriell Personalized Medicine Collaborative. J Genet Couns. 2016; 25: 385-394.
- 34. Zajic SC, Jarvis JP, Zhang P, et al. Individuals with CYP2C8 and CYP2C9 reduced metabolism haplotypes self-adjusted ibuprofen dose in the Coriell Personalized Medicine Collaborative. Pharmacogenet Genomics. 2019; 29: 49-57.
- Kusic DM, Roberts WN, Jarvis JP, et al. rs11670527 Upstream of ZNF264 Associated with Body Mass Index in the Coriell Personalized Medicine Collaborative. Mil Med. 2020; 185: 649-655.
- Affymetrix. Genome-Wide Human SNP Array 6.0 Support Materials. Affymetrix (ThermoFisher Scientific). Accessed Oct. 21, 2021, 2021.
- 37. Auton A, Brooks LD, Durbin RM, et al. A global reference for human genetic variation. Nature. 2015; 526: 68-74.
- 38. Jia H, Nogawa S, Kawafune K, et al. GWAS of habitual coffee consumption reveals a sex difference in the genetic effect of the 12q24 locus in the Japanese population. BMC Genet. 2019; 20: 61.
- 39. Rodenburg EM, Eijgelsheim M, Geleijnse JM, et al. CYP1A2 and coffee intake and the modifying effect of sex, age, and smoking. Am J Clin Nutr. 2012; 96: 182-187.
- Stoica P, Selen Y. Model-order selection: a review of information criterion rules. IEEE Signal Processing Magazine. 2004; 21: 36-47.
- 41. https://www.rdocumentation.org/packages/MASS/ versions/7.3-57/topics/stepAIC.
- 42. Zhao H, Mitra N, Kanetsky PA, et al. A practical approach to adjusting for population stratification in genome-wide association studies: principal components and propensity scores (PCAPS). Stat Appl Genet Mol Biol. 2018; 17:
- 43. Hinrichs AL, Larkin EK, Suarez BK. Population stratification and patterns of linkage disequilibrium. Genet Epidemiol. 2009; 33: S88-92.
- 44. Price AL, Patterson NJ, Plenge RM, et al. Principal components analysis corrects for stratification in genome-wide association studies. Nat Genet. 2006; 38: 904-909.

- 45. https://www.rdocumentation.org/packages/stats/versions/ 3.6.0/topics/glm.
- 46. Holm S. A Simple Sequentially Rejective Multiple Test Procedure. Scandinavian Journal of Statistics. 1979; 6: 65-70.
- Mi H, Ebert D, Muruganujan A, et al. PANTHER version 16: a revised family classification, tree-based classification tool, enhancer regions and extensive API. Nucleic Acids Res. 2021; 49: D394-D403.
- Thomas PD, Kejariwal A, Guo N, et al. Applications for protein sequence-function evolution data: mRNA/protein expression analysis and coding SNP scoring tools. Nucleic Acids Res. 2006; 34: W645-W650.
- 49. Mi H, Thomas P. PANTHER pathway: an ontology-based pathway database coupled with data analysis tools. Methods Mol Biol. 2009; 563: 123-140.
- 50. Turner SD. qqman: an R package for visualizing GWAS results using Q-Q and manhattan plots. Journal of Open Source Software. 2018; 2.
- 51. Sulem P, Gudbjartsson DF, Geller F, et al. Sequence variants at CYP1A1-CYP1A2 and AHR associate with coffee consumption. Hum Mol Genet. 2011; 20: 2071-2077.
- 52. https://ldlink.nci.nih.gov/?tab=ldmatrix.
- 53. El-Sohemy A. Coffee and health: what we still don't know. Am J Clin Nutr. 2019; 109: 489-490.
- Amin N, Byrne E, Johnson J, et al. Genome-wide association analysis of coffee drinking suggests association with CYP1A1/CYP1A2 and NRCAM. Mol Psychiatry. 2012; 17: 1116-1129.
- 55. Cornelis MC, Byrne EM, Esko T, et al. Genome-wide metaanalysis identifies six novel loci associated with habitual coffee consumption. Mol Psychiatry. 2015; 20: 647-656.
- 56. Cornelis MC, Kacprowski T, Menni C, et al. Genome-wide association study of caffeine metabolites provides new insights to caffeine metabolism and dietary caffeine-consumption behavior. Hum Mol Genet. 2016; 25: 5472-5482.
- 57. McMahon G, Taylor AE, Davey Smith G, et al. Phenotype refinement strengthens the association of AHR and CYP1A1 genotype with caffeine consumption. PLoS One. 2014; 9: e103448.
- 58. Nakagawa-Senda H, Hachiya T, Shimizu A, et al. A genomewide association study in the Japanese population identifies the 12q24 locus for habitual coffee consumption: The J-MICC Study. Sci Rep. 2018; 8: 1493.
- 59. Jorge-Nebert LF, Jiang Z, Chakraborty R, et al. Analysis of human CYP1A1 and CYP1A2 genes and their shared bidirectional promoter in eight world populations. Hum Mutat. 2010; 31: 27-40.
- 60. Thorn CF, Aklillu E, Klein TE, et al. PharmGKB summary: very important pharmacogene information for CYP1A2. Pharmacogenet Genomics. 2012; 22: 73-77.
- 61. Eiermann B, Engel G, Johansson I, et al. The involvement of CYP1A2 and CYP3A4 in the metabolism of clozapine. Br J Clin Pharmacol. 997; 44: 439-446.

- 62. Faber MS, Jetter A, Fuhr U. Assessment of CYP1A2 activity in clinical practice. Basic Clin Pharmacol Toxicol. 2005; 97: 125-134.
- 63. Pirmohamed M, Williams D, Madden S, et al. Metabolism and bioactivation of clozapine by human liver in vitro. J Pharmacol Exp Ther. 1995; 272: 984-990.
- 64. Prior TI, Baker GB. Interactions between the cytochrome P450 system and the second generation antipsychotics. Psych Neurosci. 2003; 28: 99-112.
- Sachse C, Brockmoller J, Bauer S, et al. Functional significance of a C-->A polymorphism in intron 1 of the cytochrome P450 CYP1A2 gene tested with caffeine. Br J Clin Pharmacol. 1999; 47: 445-449.
- 66. Chu T. Gender Differences in Pharmacokinetics. US Pharm. 2014; 39: 40-43.
- Ou-Yang DS, Huang SL, Wang W, et al. Phenotypic polymorphism and gender-related differences of CYP1A2 activity in a Chinese population. Br J Clin Pharmacol. 2000; 49: 145-151.
- 68. Nordestgaard AT, Nordestgaard BG. Coffee intake, cardiovascular disease and all-cause mortality: observational and Mendelian randomization analyses in 95 000-223 000 individuals. Int J Epidemiol. 2016; 45: 1938-1952.
- 69. Butler MA, Iwasaki M, Guengerich FP, et al. Human cytochrome P-450PA (P-450IA2), the phenacetin Odeethylase, is primarily responsible for the hepatic 3demethylation of caffeine and Noxidation of carcinogenic arylamines. Proc Natl Acad Sci U S A. 1989; 86: 7696-7700.
- 70. Toydemir G, Loonen LMP, Venkatasubramanian PB, et al. Coffee induces AHR- and Nrf2-mediated transcription in intestinal epithelial cells. Food Chem. 2020; 341: 128261.
- 71. Castorena-Torres F, Mendoza-Cantu A, de Leon MB, et al. CYP1A2 phenotype and genotype in a population from the Carboniferous Region of Coahuila, Mexico. Toxicol Lett. 2005; 156: 331-339.
- 72. Corchero J, Pimprale S, Kimura S, et al. Organization of the CYP1A cluster on human chromosome 15: implications for gene regulation. Pharmacogenetics. 2001; 11: 1-6.
- 73. Kapelyukh Y, Henderson CJ, Scheer N, et al. Defining the Contribution of CYP1A1 and CYP1A2 to Drug Metabolism Using Humanized CYP1A1/1A2 and Cyp1a1/Cyp1a2 Knockout Mice. Drug Metab Dispos. 2019; 47: 907-918.
- 74. Palatini P, Benetti E, Mos L, et al. Association of coffee consumption and CYP1A2 polymorphism with risk of impaired fasting glucose in hypertensive patients. Eur J Epidemiol. 2015; 30: 209-217.
- 75. Palatini P, Ceolotto G, Ragazzo F, et al. CYP1A2 genotype modifies the association between coffee intake and the risk of hypertension. J Hypertens. 2009; 27: 1594-1601.
- 76. Soares RN, Schneider A, Valle SC, et al. The influence of CYP1A2 genotype in the blood pressure response to caffeine ingestion is affected by physical activity status and caffeine consumption level. Vascul Pharmacol. 2018; 106: 67-73.

- 77. Tsepokina A, Shmulevich S, Ponasenko A, et al. Genetic predisposition to the development of congenital heart diseases: Role of xenobiotic biotransformation genes. Birth Defects Res. 2021; 113: 579-588.
- 78. Rebustini IT, Patel VN, Stewart JS, et al. Laminin alpha5 is necessary for submandibular gland epithelial morphogenesis and influences FGFR expression through beta1 integrin signaling. Dev Biol. 2007; 308: 15-29.
- 79. Geusz RJ, Wang A, Chiou J, et al. Pancreatic progenitor epigenome maps prioritize type 2 diabetes risk genes with roles in development. Elife. 2021; 10: e59067.
- Perry JR, Voight BF, Yengo L, et al. Stratifying type 2 diabetes cases by BMI identifies genetic risk variants in LAMA1 and enrichment for risk variants in lean compared to obese cases. PLoS Genet. 2012; 8: e1002741.
- Jiang N, Lee S, Park T. Hierarchical structural component model for pathway analysis of common variants. BMC Med Genomics. 2020; 13: 26.
- Chen X, Wang L, Hu B, et al. Pathway-based analysis for genome-wide association studies using supervised principal components. Genet Epidemiol. 2010; 34: 716-724.



Figure S1: QQ-plot of the residuals of a generalized linear model for caffDrinksPerWk  $\sim$  geno+age+cohort+gender+bmi+smoking+PC1:10, Lambda = 1.01.



#### Manhattan Plot Affy 6.0 SNPs for Coffee Drinks Per Week CoffDrinksPerWk~geno+age+cohort+gender+bmi+smoke+PC1:10

Figure S2: Manhattan plot of P-values from a generalized linear model highlighting a top association of rs2472297 with coffee only drink consumption per week.



Coffee Drinks Consumed per Week by rs2472297 Genotype

Figure S3: Coffee only drinks consumed per week by rs2472297 genotype.

QQ-plot of GLM for Coffee Drinks per Week, λ=0.99 CoffDrinksPerWeek~geno+age+cohort+gend+bmi+smoke+PC1:10



Figure S4: QQ-plot of the residuals of a generalized linear model for coffDrinksPerWk  $\sim$  geno+age+cohort+gender+bmi+smoking+PC1:10, Lambda = 0.99.

**Table S1:** Caffeine GWAS aggregate summary statistics for all tested SNPs. The fields in order are: variant\_id (rsid of the SNP), p\_value (of the SNP association with caffeinated drinks per week), chromosome (location of variant), base\_pair\_location (of the variant within the chromosome), effect\_ allele (associated with beta), other\_allele (associated with (0 –beta)), effect\_allele\_frequency (proportion of effect allele at SNP locus within sample), beta (cups of caffeinated drink per week per effect allele), standard\_error (of beta).

**Table S2:** Coffee GWAS aggregate summary statistics for all tested SNPs. The fields in order are: variant\_id (rsid of the SNP), p\_value (of the SNP association with coffee drinks per week), chromosome (location of variant), base\_pair\_location (of the variant within the chromosome), effect\_allele (associated with beta), other\_allele (associated with (0 –beta)), effect\_allele\_frequency (proportion of effect allele at SNP locus within sample), beta (cups of coffee drink per week per effect allele), standard\_error (of beta).

© 2023 Kusic DM, et al. This article is distributed under the terms of the Creative Commons Attribution 4.0 International License