Phenome-wide associations of coffee intake in the human phenotype project

Jin Dai, Wen Dai, Yoriko Heianza, Lu Qi

PII: S0026-0495(25)00281-1

DOI: https://doi.org/10.1016/j.metabol.2025.156412

Reference: YMETA 156412

To appear in: *Metabolism*

Received date: 8 July 2025

Accepted date: 29 September 2025

Please cite this article as: J. Dai, W. Dai, Y. Heianza, et al., Phenome-wide associations of coffee intake in the human phenotype project, *Metabolism* (2025), https://doi.org/10.1016/j.metabol.2025.156412

Title: Phenome-wide associations of coffee intake in the Human Phenotype Project

Running title: Phenome-wide associations of coffee intake

Authors: Jin Dai, PhD[1]; Wen Dai[1,2] MPH; Yoriko Heianza[1,3], PhD; Lu Qi[1,3], PhD

Affiliations:

1. Department of Epidemiology, School of Public Health and Tropical Medicine, Tulane University, 1440 Canal Street, Suite 1724, New Orleans, LA 70112, USA

2. Bureau of Family Health, Louisiana Department of Health, 1450 Poydras Street, New Orleans, LA 70112, USA

3. Department of Nutrition, Harvard T.H. Chan School of Public Health, 677 Huntington Ave, Boston, MA 02115, USA

Corresponding author: Lu Qi, email: lqi1@tulane.edu (https://orcid.org/0000-0002-8041-7791)

Word count (main text): 4,618; **Word count (abstract)**: 250; **Number of Figures**: 5

ABSTRACT

Objective: Coffee is one of the most widely consumed beverages globally and has been linked to favorable health outcomes. However, its system-wide relationships with human biology and the underlying mechanisms remain poorly characterized. This study aimed to investigate the relationship between coffee consumption and continuous glucose monitoring (CGM) metrics and other biological systems in healthy adults.

Research Design and Methods: In the Human Phenotype Project, 8,666 generally healthy Israeli adults provided two weeks of real-time dietary logs, from which coffee intake was estimated. Participants wore CGM devices throughout this period, and multimodal data spanning 11 additional systems (e.g., gut microbiome, serum lipidomics, and body composition) were collected. We employed machine learning approaches to quantify the extent to which each system reflected coffee intake. We performed linear regression to identify individual traits associated with coffee intake, with false discovery rates <0.05 considered significant.

Results: This cross-sectional study identified continuously-monitored glucose regulation and gut microbial composition as the most reflective systems of coffee intake, with further analyses revealing favorable glycemic profiles spanning diverse aspects of glucose regulation with increasing coffee intake, and *Clostridium phoceensis* (i.e., *Lawsonibacter asaccharolyticus*) as the most significant species positively associated with coffee intake. Additionally, coffee intake was favorably associated with traits across body composition, serum lipidomics, and hepatic, hematopoietic, and renal systems.

Conclusions: This study found that habitual coffee intake was linked to multifaceted favorable glucose control captured by CGM and favorable profiles across multiple biological systems,

providing mechanistic insights that may guide precision nutrition strategies for diabetes prevention.

1 INTRODUCTION

Coffee is one of the most widely consumed beverages globally, with numerous observational studies linking its habitual intake to favorable health outcomes, such as a lower risk of type 2 diabetes, liver disease, and all-cause mortality [1, 2]. However, the underlying biological mechanisms remain incompletely understood [2]. Moreover, prior research has predominantly focused on isolated disease outcomes [3, 4] or limited biomarker panels [5, 6], thus precluding a holistic understanding of how habitual coffee intake influences human health.

Phenome-wide association studies (PheWAS) extend the genome-wide association paradigm by systematically testing an exposure against a broad spectrum of phenotypes [7]. Compared with traditional hypothesis-driven approaches, PheWAS provides a system-level perspective that better reflects the multidimensional nature of human health, emphasizing overall homeostatic balance rather than isolated disease outcomes [8]. To date, only two studies have adopted a PheWAS framework to investigate coffee intake, but their findings were contradictory, partly due to reliance on weak genetic instruments and assessment in a low-consumption population [9, 10]. Moreover, both studies focused primarily on low-resolution disease outcomes, leaving the question of how habitual coffee intake influences human biology at a high-resolution, system-wide level largely unexplored. Indeed, coffee's global ubiquity across cultures [11], habitual consumption patterns (typically consumed daily or not at all) [12], and rich bioactive compound profile [13] make it an ideal model for investigating how habitual dietary exposures shape human biology across systems.

To address this gap, we conducted the first high-resolution PheWAS of habitual coffee intake using data from the Human Phenotype Project (HPP), a deeply phenotyped cohort of 8,666 generally healthy adults. We integrated multimodal data spanning 12 biological systems,

including continuous glucose monitoring (CGM), gut microbiome, serum lipidomics, dual-energy X-ray Absorptiometry (DXA) scan, liver ultrasound, and home sleep apnea testing, among others. Leveraging this comprehensive phenotyping, our study provides novel mechanistic insights into how coffee may affect multiple domains of human biology.

2 METHODS

2.1 Study design and population

HPP is an ongoing, deeply phenotyped, prospective cohort of generally healthy adults established in Israel [14]. The study design has been previously described in detail [14-16]. In brief, the primary objective of the HPP is to develop prediction models for and identify molecular biomarkers of disease onset and progression. To achieve this goal, HPP collected extensive multimodal data, such as DXA, electrocardiography, liver and carotid ultrasound, and multi-omics profiles. Participants were also requested to wear a CGM device for two weeks, while simultaneously logging their dietary intake through a smartphone App. HPP was approved by the Institutional Review Board of the Weizmann Institute of Science (reference number: 1719-1). Our cross-sectional investigation of the HPP study was approved by the Institutional Review Board of Tulane University (reference number: 2025-618). All participants signed an informed consent form at the research site.

The HPP enrolled Israeli adults aged 40 to 70 years who did not have major chronic diseases (e.g., cardiovascular disease, dementia, or cancer), were not pregnant, had no recent or chronic antibiotic exposure, and had no recent unexplained weight loss (**Supplementary Figure 1**). Our study included 10,434 Israeli participants enrolled in HPP between December 2018 and December 2022. To improve the representativeness of participants' habitual dietary intake, we

excluded those with less than 7 days of valid dietary intake logging. Logging of energy intake <500 or >4000 kcal was considered invalid. Additionally, we excluded participants with self-reported physician-diagnosed diabetes to eliminate the impact of antidiabetic medications on CGM metrics. Finally, to mitigate the impact of extreme values on the phenome-wide associations of coffee intake, we excluded participants reporting coffee intake greater than 99% of daily coffee intake at the population level. A total of 8,666 participants were retained as our analytic sample.

2.2 Assessment of coffee intake

Participants were instructed to record their dietary intake in real time over the two weeks of wearing CGM using a bespoke smartphone app ("Project 10K app"). The app's food composition database was built on the Israeli Ministry of Health database and expanded with additional certified food items to include over 7,000 items. The app allows participants to record food items along with their weights or portion sizes and submit those entries to their profile, and has supported multiple studies [17-19]. The dietary data have been found to be correlated with diet-related serum metabolites, which offers an objective validation [20]. For each participant, we calculated average daily coffee intake (g/d) using all valid dietary logging days.

2.3 Assessment of biological systems

We grouped most measures into 12 categories representing major biological systems (Figure 1). All measures in each system, along with the number of non-missing values, are presented in Supplementary Tables 1.

Glucose regulation was phenotyped using the *FreeStyle Libre Pro Flash* continuous glucose monitoring (CGM) system (Abbott), worn for 14 consecutive days, with interstitial glucose readings recorded every 15 minutes. Based on the raw CGM data, we computed 49

CGM metrics capturing diverse dimensions of glycemic patterns, including general glucose, euglycemia, hyperglycemia, hypoglycemia, variability, and composite measures and risk scores, using the R package *iglu* [21, 22]. Given the substantial intrapersonal variability in fasting glucose revealed by a recent study [23], we computed the mean fasting glucose and its standard deviation (SD) as previously reported [23], and derived the coefficient of variation (CV) as SD/mean × 100% [23].

Fecal sample collection, DNA extraction, library preparation, and metagenome sequencing were conducted as previously described [24]. Taxonomic profiling was conducted using MetaPhlAn v4.0.6. We calculated α-diversity as species richness and Shannon index) and assessed β-diversity with Bray-Curtis distance using R package *vegen*. We obtained the first two axes by performing principal coordinate analysis (PCoA) based on the Bray-Curtis dissimilarity matrix. We excluded species with an average relative abundance of <0.01% or present in <10% of samples, yielding 311 prevalent taxa. Finally, we performed centered log-ratio (CLR) transformation on the taxonomic data for downstream analyses.

Body composition was evaluated by whole-body dual-energy X-ray absorptiometry (DXA) using a GE Lunar Prodigy Advance system (GE Healthcare) with CoreScan software. Participants underwent supine scans to quantify percentage fat, fat mass, and lean mass across five regions (arms, legs, trunk, android, and gynoid), yielding 108 regional measures. CoreScan additionally estimated visceral and subcutaneous adipose tissue area, mass, and volume within the android region. Body weight, standing height, and waist and hip circumferences were measured. Given that body mass index (BMI) is a well-established risk factor for multiple health outcomes [25], we included BMI as a covariate in our statistical models.

Bone mineral density was quantified during the whole-body DXA scan. We obtained 182 site-specific measures at bilateral femoral necks and lumbar spine vertebrae L1–L4.

Liver health was assessed using abdominal ultrasound (Supersonic Aixplorer MACH; Supersonic Imagine) combined with two-dimensional shear-wave elastography, which yielded quantitative measures of tissue viscosity, elasticity, acoustic attenuation, and sound-speed propagation. Additionally, we included serum biomarkers of hepatic function, including alkaline phosphatase, alanine aminotransferase, aspartate aminotransferase, total protein, total bilirubin, platelet count, and albumin, to capture biochemical indices of liver status.

Cardiovascular system was phenotyped using multiple complementary modalities: (1) a 12-lead resting electrocardiogram (PC-ECG1200; NORAV); (2) brachial blood pressure measured after five minutes of seated rest with an OMRON oscillometric monitor; (3) ankle–brachial index (ABI) measured with a Vasionix Falcon device; (3) carotid–femoral pulse-wave velocity (PWV) assessed via the same Vasionix system; (4) bilateral carotid intima–media thickness (cIMT) quantified by Doppler ultrasound (Supersonic Aixplorer MACH30; Hologic); and (6) retinal vascular parameters extracted from retinal imaging using the AutoMorph software [26].

Blood lipid profiling integrated untargeted lipidomics with routine clinical assays. Fasting plasma samples were analyzed on a Waters ACQUITY UPLC system coupled to a Vion IMS QToF mass spectrometer (Waters Corporation), while serum HDL-C, LDL-C, and triglycerides were quantified by standard enzymatic assays. Lipidomic features missing in > 20% of samples were removed, and remaining missing values were imputed as half the minimum observed value per feature. We then performed the rank-based inverse-normal transformation for the lipidomic features, yielding 133 unique lipids.

Renal function was assessed by standard clinical assays of serum creatinine, blood urea, and electrolytes (i.e., sodium and potassium).

Hematopoietic system profiling comprised serum ferritin, hemoglobin, hematocrit, mean corpuscular volume, mean corpuscular hemoglobin, mean corpuscular hemoglobin concentration, and red blood cell count and distribution width.

Immune system profiling consisted of measures from complete blood counts, including total white blood cell count, and absolute and relative (percentage) values for neutrophils, lymphocytes, monocytes, eosinophils, and basophils.

Frailty was evaluated by combining hand grip strength with appendicular lean mass (arms and legs) measured by DXA.

Sleep health phenotyping was conducted at home using the WatchPAT 300 device (Itamar Medical), which integrates peripheral arterial tonometry, finger pulse oximetry, acoustic snoring, and body-position sensors. Each participant completed three consecutive days of monitoring. Proprietary software processed the raw data to generate metrics of oxygen saturation, respiratory events, heart rate, snoring intensity, and sleep-stage architecture. A total of 57 parameters were retained for data analysis.

All measurements mentioned above were standardized unless explicitly stated otherwise.

2.4 System-level modeling of biological responses to coffee intake

We assessed each biological system's ability to capture the biological footprint of habitual coffee intake (g/d) by fitting both elastic-net and XGBoost regression models in a five-fold cross-validation framework repeated ten times. Features consisted of centered log-ratio transformed (gut microbiome), inverse-normal transformed (serum lipidomics), or z-score standardized (for the other biological systems) system-specific measures, together with standardized age and a

binary sex indicator. In each fold, we computed the Spearman correlation (ρ) between predicted and observed habitual coffee intake on the validation dataset. Within each repeat, we obtained the median ρ across the five folds and summarized overall model performance by the median ρ across all ten repeats.

2.5 System-specific traits associated with coffee intake

To evaluate the associations of habitual coffee intake with each biological system measure, we performed multivariable linear regression analyses, adjusting for age, sex, educational attainment, household income, BMI, smoking status, smoking cessation time, current smoking number, alcohol intake status, moderate-to-vigorous physical activity, healthy dietary pattern score, total energy intake, and tea intake. The dietary pattern score was based on seven components: (1) vegetables and fruits, (2) whole grains, (3) nuts, (4) legumes, (5) red and processed meat, (6) fish, and (7) sodium. For each component, participants with intakes above the sex-specific median were assigned a score of 1, and those below the median a score of 0, except for red and processed meat and sodium, where scoring was reversed (i.e., intake above the median = 0, below the median = 1). Within each system, we adjusted *P*-values for the multiple comparisons using the Benjamini-Hochberg procedure, with a two-sided false discovery rate (FDR)-adjusted *P*-value < 0.05 as statistically significant [15]. Missing values were imputed as medians for continuous variables and missing categories for categorical variables.

We performed several sensitivity analyses. First, we examined the potential of non-linearity of the significant associations by comparing models with and without cubic spline terms (with knots placed at 5^{th}, 50^{th}, and 95^{th} percentiles), using the likelihood-ratio test. Second, we evaluated whether associations between coffee intake and biological system measures varied by coffee type by including an interaction term between coffee

intake and coffee type (non-coffee drinkers, instant coffee drinkers, ground/filtered coffee drinkers, decaffeinated coffee drinkers, and other coffee drinkers), using the likelihood ratio test. Third, to mitigate the potential of reverse causation between coffee intake and continuously monitored glucose regulation, we evaluated the associations between daily average coffee intake during the first half of the CGM wear (median: 7 days) and the CGM metrics derived from the remaining period (median: 7 days). Fourth, to assess the robustness of our findings to potential dietary confounding, we adjusted for the individual dietary components underlying the dietary pattern used in the primary analysis. Fifth, to partially mitigate potential confounding due to coffee sweetening status, we re-examined the associations between coffee intake and biological system measures after excluding coffee entries labeled as "sweet" or "diet".

We analyzed all data using R 4.4.3 (R Foundation for Statistical Computing, Vienna, Austria).

3 RESULTS

3.1 Cohort characteristics

Our study population included 8,666 generally healthy Israeli adults (Supplementary Table 2). Participants had a median of 14 days (IQR, 13-14) of valid dietary logs. The mean age was 51.8 years (SD, 7.8), and 52.4% were female (n=4,537). Participants had a mean of 20.2 years (SD, 2.9) of education, 44.3% had normal body weight, 62.3% were never smokers, and 71.1% reported current alcohol consumption. Mean daily energy intake was 1724.7 kcal (SD, 425.9). Mean daily coffee intake was 110.5 g (SD, 146.8), with few participants classified as heavy consumers (i.e., 73 participants reported >600 g/day of coffee intake)[27]. Coffee intake declined slightly with aging (Supplementary Figure 2).

3.2 System-level links with habitual coffee intake

We employed machine learning approaches (i.e., elastic net regression and XGBoost algorithms) to assess how well each biological system captures the phenotypic footprint of habitual coffee intake and ranked them based on their predictive performance (Methods). Across the 12 biological systems, elastic net outperformed XGBoost in nine systems, whereas XGBoost achieved slightly higher correlations for liver health, renal system, and sleep health, suggesting that associations between habitual coffee intake and system-specific features were predominantly linear, as captured by the regularized linear model (Figure 2). Notably, gut microbial species relative abundances exhibited the highest correlation with habitual coffee intake (median ρ=0.28 from elastic net regression), indicating that interindividual differences in microbial composition captured a meaningful portion of variability in habitual coffee intake. Glucose regulation metrics derived from CGM ranked second (median ρ=0.17 from elastic net regression) and exhibited the highest stability in predictive performance across iterations. The cardiovascular system and body composition reflected coffee intake to a moderate extent, whereas the other biological systems showed only marginal correlations.

3.3 Coffee intake and continuously monitored glucose regulation

We prioritized glucose regulation for in-depth analyses, as it exhibited one of the strongest signals related to habitual coffee intake in our previous analytic step. A deeper understanding of glucose regulation requires a comprehensive collection of metrics derived from CGM. Here, we derived 49 metrics from the literature [21, 23], and organized them into seven domains, reflecting general glucose (e.g., mean glucose), fasting state (e.g., continuously-monitored fasting glucose), euglycemia (e.g., percentage of time in range between 63 and 140 mg/dL), hyperglycemia (e.g., percentage of time above 140 mg/dL), hypoglycemia (e.g., low blood

glucose index (LBGI)), glucose variability (e.g., SD), and composite measurements and risk scores (e.g., J-index) (Figure 3A). Overall, general glycemia measures were strongly positively correlated with each other ($\rho \geq 0.92$). Continuously-monitored fasting glucose exhibited strong positive correlations with general glycemia measures ($\rho \geq 0.85$) but weakly correlated with fasting glucose variability ($|\rho| \leq 0.20$). Among the euglycemia measures, GRADE_eugly strongly correlated with all time in range measurements ($\rho \geq 0.79$). Of the hyperglycemia measurements, HBGI, Hyper index, and GRADE_hyper strongly positively correlated with each other ($\rho \geq 0.95$), while their associations with percentage time above threshold declined at higher cut-offs. Findings were similar for hypoglycemia measures. Variability metrics correlated positively with each other overall, and composite measures and risk scores exhibited more heterogeneous interrelationships.

To elucidate how habitual coffee intake shape glucose regulation, we conducted linear regression analyses adjusting for a broad range of potential confounders, including age, sex, educational attainment, household income, body mass index (BMI), smoking status, smoking cessation time, current smoking number, alcohol intake status, moderate-to-vigorous physical activity, dietary pattern, daily energy intake, and tea intake. Coffee intake was associated with 23 of the 49 CGM metrics (FDR < 0.05) across six domains of glucose regulation (except for euglycemia), with 21 showing inverse associations (Figure 3B). Specifically, habitual coffee intake was inversely associated with all six general glucose measures and continuously monitored fasting glucose levels. Additionally, coffee intake was inversely associated with two hyperglycemia metrics and nine variability measurements featuring diverse variability, including overall variability (IQR), within-day variability (SDw), and between-day variability (SDbdm). Moreover, coffee intake was inversely associated with three composite measures and risk scores,

including AUC, J index, and COGI. By contrast, coffee intake was positively associated with two hypoglycemia measurements, including GRADE hypo and LBGI.

3.4 Coffee intake and gut microbiome

Building on our prior analytical step that revealed the highest correlation between coffee intake and gut microbiome, we next focused on the gut microbiome for further analyses. Overall, coffee intake was positively associated with gut microbial α-diversity (species richness ($P < 0.001$) and Shannon index ($P = 0.03$)) and associated with β-diversity (the first two principal coordinates of the Bray-Curtis distance matrix ($P \leq 0.02$)) (Figures 4A-C). To identify specific taxa driving these patterns, we performed multivariable linear regression adjusted for the aforementioned covariates, with species-level relative abundances processed with centered log-ratio transformation (CLR) to account for compositionality. Of the 311 prevalent species, habitual coffee intake was associated with 42 microbial species relative abundances, with 18 showing positive associations (Figure 4D). The strongest associations were observed for *Clostridium phoceensis* (FDR < 10^{-24}) (recently reclassified as *Lawsonibacter asaccharolyticus*),[27] followed by the so far uncharacterized taxon 'GGB9522 SGB14921' (FDR < 10^{-20}), both exhibiting non-linear relationships with coffee (Figure 4E). In addition, the associations for five species reached FDR < 10^{-4}, including *Bifidobacterium psedocatenulatum*, *Clostridium* species AM22 11AC, *Blautia stercoris*, *Faecalibacterium prausnitzii*, and uncharacterized 'GGB9635 SGB15106'.

3.5 Coffee intake and multiple biological systems

Of the 118 body composition metrics, coffee intake was associated with 22 metrics after adjusting for BMI and other covariates (Figure 5A). Notably, all the associations were inverse, suggesting that higher coffee intake lowers multiple body composition features. The top ten

metrics were all related to body fatness, including trunk (left, right, and total) fat mass; visceral adipose tissue mass, area, and volume; total, left, and right fat mass, and android fat mass. Additionally, habitual coffee intake was inversely associated with waist circumference after adjusting for BMI.

Among the 133 lipids, coffee intake was associated with six lipids, with five exhibiting positive and one exhibiting inverse associations after adjustment for age, sex, BMI, and other potential confounders (Figure 5B). Specifically, coffee intake was positively associated with lysophosphatidylethanolamine (LysoPE) 18:0, LysoPE 16:0, lysophosphatidylcholine (LysoPC) 20:4, phosphatidylethanolamine (PE) 34:1, and high-density lipoprotein cholesterol (HDL-C). By contrast, coffee intake was inversely associated with phosphatidylcholine (PC) 32:1.

Among the 11 liver measures, coffee intake was inversely associated with alkaline phosphatase activity (Figure 5C). No significant associations were observed for liver ultrasound measures or enzymes. Of the eight circulating hematopoietic measures, coffee intake was inversely associated with red blood cell count (Figure 5D). Among the four renal system measures, coffee intake was positively associated with serum potassium (Figure 5E). All significant associations were at FDR < 0.05.

3.6 Sensitivity analyses

The association between coffee intake and each biological system measure showed no evidence of non-linearity (FDR > 0.05), did not vary by coffee type (FDR > 0.05; Supplementary Table 3), and remained robust after adjusting for individual food component that comprised the dietary pattern controlled for in the primary analysis (Supplementary Table 4) and after excluding coffee entries labeled as “sweet” or “diet” (Supplementary Table 5). Additionally, the associations between coffee intake and CGM metrics remained materially unchanged when evaluating coffee

intake during the first half of the CGM wear (median: 7 days) in relation to CGM metrics derived from the remaining period (median: 7 days) (Supplementary Table 6).

4 DISCUSSION

In this high-resolution, phenome-wide association study, we demonstrated that habitual coffee intake was systematically associated with multiple domains of human biology in a deeply phenotyped, generally healthy adult population. By integrating multi-system physiological, biochemical, and omics data, we identified gut microbial composition and continuously monitored glucose regulation as the biological systems most strongly linked to habitual coffee intake. Further analyses revealed favorable glycemic profiles across diverse aspects of glucose regulation with increasing coffee intake and identified *Clostridium phoceensis*, recently reclassified as *Lawsonibacter asaccharolyticus*, as the most significant microbial species positively associated with coffee intake. Beyond these leading systems, coffee intake was associated with a broad range of features across multiple biological systems, including body composition, serum lipids, hematopoietic, liver, and renal systems, with most associations indicating favorable profiles with increasing coffee intake.

To our knowledge, this is the first study to showcase that habitual coffee intake is favorably associated with multiple dimensions of glucose regulation, as continuously assessed using a comprehensive panel of CGM-derived metrics. These associations remained robust after adjusting for BMI, smoking-related variables, and other potential confounders. To date, only two studies have employed CGM to investigate the glycemic effects of coffee intake in the general population [28, 29]. However, both focused exclusively on short-term glycemic responses, which are unlikely to reflect the long-term effects of habitual intake, particularly given the progressive

development of caffeine tolerance [30]. While cohort and Mendelian randomization studies have linked habitual coffee intake or its genetic proxy with a lower risk of type 2 diabetes, no study has explored its influence on continuous glycemic dynamics [31]. Our study advances this literature by examining how habitual coffee intake shapes real-world glucose regulation. We found consistent associations with lower glycemic variability, improved composite measures and risk scores, and more favorable profiles across general glucose, fasting glucose, and hyperglycemia domains. It is noteworthy that habitual coffee intake was positively associated with hypoglycemic metrics, including GRADE_hypo and LBGI. While these associations may reflect tighter glycemic regulation in the general population, they warrant caution in individuals with diabetes, particularly those on glucose-lowering medications, in whom coffee intake may potentiate hypoglycemic episodes.

To date, only a handful of studies have explored the relationship between habitual coffee intake and the human gut microbiome [27, 32, 33], with most of them limited by small sample sizes and low taxonomic resolution [27, 33]. In our study, gut microbial composition emerged as a biological system most strongly linked to habitual coffee intake, outperforming other domains such as glucose regulation and body composition. This suggests that the gut microbiome may be particularly sensitive to coffee intake and could partly mediate its systemic biological effects. This observation is further supported by a recent multi-cohort metagenomic study conducted predominantly in the UK and US cohorts [27]. That study identified coffee as the strongest dietary correlate of gut microbial composition among 150 food items, reinforcing the central role of the gut microbiome in response to habitual coffee intake [27]. Additionally, our findings of increased alpha diversity and significant shifts in beta diversity with increasing coffee intake align with the existing evidence on the coffee-microbiome relationship [27, 33], and support

prior evidence linking increased gut microbial diversity to favorable cardiometabolic health profiles [34, 35]. Moreover, leveraging in-depth metagenomic sequencing, we identified over 40 microbial species whose relative abundances were associated with habitual coffee intake. Notably, *Clostridium phoceensis*, which showed the strongest and positive association in our dataset, was also the species most significantly associated with habitual coffee intake in the multi-cohort study [27]. Although that study annotated this species as *Lawsonibacter asaccharolyticus*, taxonomic analysis confirms both names refer to the same species-level genome bin, with *L. asaccharolyticus* now recognized as the correct taxonomic assignment. This robust association has also been validated experimentally in vitro using both caffeinated and decaffeinated coffee, indicating the biological effects of coffee components beyond caffeine [27]. The reproducibility of *L.asaccharolyticus (C. phoceensis*) across geographically distinct populations further supports its generalizability as a key microbial indicator of habitual coffee intake.

We observed inverse associations between habitual coffee intake and multiple adiposity-related measures, including visceral, trunk, and total fatness, which remained significant after adjustment for BMI. These associations align with existing evidence indicating that caffeine may reduce appetite and elevate basal metabolic rate and food-induced thermogenesis via activating the sympathetic nervous system and brown adipose tissue [36, 37]. Our findings extend beyond general measures of adiposity by implicating a potential role for coffee intake in shifting fat distribution toward a more favorable metabolic profile. Given that visceral fat and waist circumference provide additive predictive value for cardiometabolic risk beyond BMI [38], our findings highlight the potential for coffee to exert protective effects on cardiometabolic health through lowering central adiposity.

Our serum lipidomic analyses identified lysoPE 18:1 as the lipid most strongly and positively associated with habitual coffee intake. To our knowledge, no prior studies have reported this, indicating lysoPE 18:1 as a potential novel lipidomic biomarker of coffee consumption. Although the functionality of lysoPE 18:1 remains to be established, it has previously been reported to be elevated in patients with systemic lupus erythematosus, an autoimmune condition linked to insulin resistance and atherosclerosis [39, 40]. In addition, habitual coffee intake was inversely associated with HDL-C, a classic cardiometabolic biomarker [41-43], and linked to lower serum alkaline phosphatase, red blood cell count, and higher serum potassium levels. Collectively, these associations across lipidomic, liver, hematopoietic, and renal systems suggest that habitual coffee intake may exert multi-system biological effects beyond glycemic, gut microbial, and adiposity domains.

As the primary component of coffee, caffeine may contribute to the observed associations through several molecular pathways, primarily by acting as a non-selective adenosine receptor antagonist that enhances alertness, improves insulin sensitivity, and attenuates inflammatory responses [2]. Beyond caffeine, coffee contains diverse bioactive compounds that may underlie its health benefits [2]. For instance, chlorogenic acids and related phenolic compounds exhibit antioxidant and anti-inflammatory properties, potentially improving glucose regulation and modulating gut microbial composition [2]. Moreover, melanoidins formed during the roasting process have been shown to exert prebiotic effects, promoting beneficial microbial taxa [44]. Collectively, these mechanisms suggest that the favorable profiles observed in our study likely reflect the combined effects of multiple coffee constituents across diverse biological systems. Further in-depth mechanistic studies are warranted to fully elucidate how coffee and its bioactive components systematically influence human biology.

A major strength of our study lies in the integration of high-resolution, multimodal data across multiple biological systems in a large, well-characterized, generally healthy population. This study design enabled a comprehensive systems-level investigation of how habitual coffee intake maps onto diverse domains of human biology. Leveraging real-time dietary intake logging further enhances the accuracy of exposure assessment. In addition, we carefully adjusted for a broad range of potential confounders, including multiple smoking-related variables [4], which alleviates residual confounding. Nevertheless, our study has several limitations. First, coffee intake and biological system measures were collected concurrently at baseline. This cross-sectional design raises the possibility of reverse causation, although our findings remained robust in a sensitivity analysis that evaluated coffee intake during the first half of the CGM monitoring period and CGM metrics from the second half. Moreover, several high-impact cross-sectional investigations from the HPP, such as those linking fasting glucose variability with cardiovascular disease risk [23] and sleep characteristics with hypertension risk [15], have yielded findings consistent with large prospective cohorts [45, 46], thereby supporting the validity of HPP-based cross-sectional investigations. Also, causality cannot be established due to the observational nature of the study. Second, although a sensitivity analysis excluding coffee entries labeled as "sweet" or "diet" produced consistent results, the lack of systematic documentation of coffee sweetening status limited our ability to thoroughly evaluate its health effects [47]. This limitation highlights the need for future studies with more granular data on coffee sweetening practices. Finally, our cohort included only Israeli adults, and coffee preparation methods were relatively uniform, which may limit the generalizability of our findings to other populations and cultural contexts. Future studies in diverse ancestries, with variation in coffee preparation practices and long-term interventional designs, will be critical to validate our findings.

In summary, our phenome-wide association analysis demonstrates the capacity of coffee, a globally consumed beverage, to influence multiple dimensions of human biology. By integrating multimodal deep phenotyping data with systems-level analysis, our study provides a generalizable framework to investigate how habitual exposures systematically shape human biology in real-world settings. Moreover, our study underscores the potential of dietary exposures, such as coffee, to serve as potent drivers in developing personalized, systems-informed nutritional strategies.

ACKNOWLEDGEMENTS

The authors thank the investigators, the staff, and the participants of the HPP study for their valuable contributions.

Conflict of interests

The authors declare no conflict of interest.

Funding

The study was supported by grants from the National Heart, Lung, and Blood Institute (HL071981, HL034594, HL126024), the National Institute of Diabetes and Digestive and Kidney Diseases (DK115679, DK091718, DK100383, DK078616), the Fogarty International Center (TW010790), and Tulane Research Centers of Excellence Awards. Lu Qi is also supported by P30DK072476 and NIGMS P20GM109036.

Author contributions

J.D. and L.Q. designed the study concept. J.D. and W.D., conducted data curation. J.D. performed data analysis and drafted the first version of the manuscript. All authors interpreted the data and revised the manuscript critically for important intellectual content. L.Q. is the guarantor of this work and, as such, has full access to all the data in the study and takes responsibility for the integrity of the data and the accuracy of the data analysis. L.Q. attests that all listed authors meet authorship criteria and that no others meeting the criteria have been omitted.

Guarantor statement

L.Q. affirms that the manuscript is an honest, accurate, and transparent account of the study being reported; that no important aspects of the study have been omitted.

Copyright/License for publication

Data sharing

Data in this paper are part of the Human Phenotype Project (HPP) and are accessible to researchers from universities and other research institutions at https://humanphenotypeproject.org/data-access. The HPP data includes personal information and, in compliance with institutional review board regulations, cannot be publicly available.

Supplementary data

Supplementary material

REFERENCES

[1] Poole R, Kennedy OJ, Roderick P, Fallowfield JA, Hayes PC, Parkes J. Coffee consumption and health: umbrella review of meta-analyses of multiple health outcomes. Bmj. 2017;359:j5024.

[2] van Dam RM, Hu FB, Willett WC. Coffee, Caffeine, and Health. N Engl J Med. 2020;383:369-78.

[3] van Dam RM, Feskens EJ. Coffee consumption and risk of type 2 diabetes mellitus. Lancet. 2002;360:1477-8.

[4] Freedman ND, Park Y, Abnet CC, Hollenbeck AR, Sinha R. Association of coffee drinking with total and cause-specific mortality. N Engl J Med. 2012;366:1891-904.

[5] Alperet DJ, Rebello SA, Khoo EY, Tay Z, Seah SS, Tai BC, et al. The effect of coffee consumption on insulin sensitivity and other biological risk factors for type 2 diabetes: a randomized placebo-controlled trial. Am J Clin Nutr. 2020;111:448-58.

[6] Larsen SC, Mikkelsen ML, Frederiksen P, Heitmann BL. Habitual coffee consumption and changes in measures of adiposity: a comprehensive study of longitudinal associations. Int J Obes (Lond). 2018;42:880-6.

[7] Bastarache L, Denny JC, Roden DM. Phenome-Wide Association Studies. Jama. 2022;327:75-6.

[8] Yurkovich JT, Evans SJ, Rappaport N, Boore JL, Lovejoy JC, Price ND, et al. The transition from genomics to phenomics in personalized population health. Nat Rev Genet. 2024;25:286-302.

[9] Nicolopoulos K, Mulugeta A, Zhou A, Hyppönen E. Association between habitual coffee consumption and multiple disease outcomes: A Mendelian randomisation phenome-wide association study in the UK Biobank. Clin Nutr. 2020;39:3467-76.

[10] Mak JKL, Chau YP, Tan KC, Kung AW, Cheung CL. Phenome-Wide Analysis of Coffee Intake on Health over 20 Years of Follow-Up Among Adults in Hong Kong Osteoporosis Study. Nutrients. 2024;16.

[11] Cicco S. Diet as a cornerstone of cardiovascular prevention: the key role of coffee consumption. Eur J Prev Cardiol. 2022;29:979-81.

[12] Salvini S, Hunter DJ, Sampson L, Stampfer MJ, Colditz GA, Rosner B, et al. Food-based validation of a dietary questionnaire: the effects of week-to-week variation in food consumption. Int J Epidemiol. 1989;18:858-67.

[13] Karagöz MF, Koçyiğit E, Koçak T, Özturan Şirin A, Icer MA, Ağagündüz D, et al. Decoding coffee cardiometabolic potential: Chemical composition, nutritional, and health relationships. Compr Rev Food Sci Food Saf. 2024;23:e13414.

[14] Shilo S, Bar N, Keshet A, Talmor-Barkan Y, Rossman H, Godneva A, et al. 10 K: a large-scale prospective longitudinal study in Israel. Eur J Epidemiol. 2021;36:1187-94.

[15] Kohn S, Diament A, Godneva A, Dhir R, Weinberger A, Reisner Y, et al. Phenome-wide associations of sleep characteristics in the Human Phenotype Project. Nat Med. 2025;31:1026-37.

[16] Reicher L, Bar N, Godneva A, Reisner Y, Zahavi L, Shahaf N, et al. Phenome-wide associations of human aging uncover sex-specific dynamics. Nature Aging. 2024;4:1643-55.

[17] Zeevi D, Korem T, Zmora N, Israeli D, Rothschild D, Weinberger A, et al. Personalized Nutrition by Prediction of Glycemic Responses. Cell. 2015;163:1079-94.

[18] Korem T, Zeevi D, Zmora N, Weissbrod O, Bar N, Lotan-Pompan M, et al. Bread Affects Clinical Parameters and Induces Gut Microbiome-Associated Personal Glycemic Responses. Cell Metab. 2017;25:1243-53.e5.

[19] Ben-Yacov O, Godneva A, Rein M, Shilo S, Kolobkov D, Koren N, et al. Personalized Postprandial Glucose Response-Targeting Diet Versus Mediterranean Diet for Glycemic Control in Prediabetes. Diabetes Care. 2021;44:1980-91.

[20] Bar N, Korem T, Weissbrod O, Zeevi D, Rothschild D, Leviatan S, et al. A reference map of potential determinants for the human serum metabolome. Nature. 2020;588:135-40.

[21] Keshet A, Shilo S, Godneva A, Talmor-Barkan Y, Aviv Y, Segal E, et al. CGMap: Characterizing continuous glucose monitor data in thousands of non-diabetic individuals. Cell Metab. 2023;35:758-69.e3.

[22] Broll S, Urbanek J, Buchanan D, Chun E, Muschelli J, Punjabi NM, et al. Interpreting blood GLUcose data with R package iglu. PLoS One. 2021;16:e0248560.

[23] Shilo S, Keshet A, Rossman H, Godneva A, Talmor-Barkan Y, Aviv Y, et al. Continuous glucose monitoring and intrapersonal variability in fasting glucose. Nat Med. 2024;30:1424-31.

[24] Keshet A, Segal E. Identification of gut microbiome features associated with host metabolic health in a large population-based cohort. Nat Commun. 2024;15:9358.

[25] González-Muniesa P, Mártinez-González M-A, Hu FB, Després J-P, Matsuzawa Y, Loos RJF, et al. Obesity. Nature Reviews Disease Primers. 2017;3:17034.

[26] Zhou Y, Wagner SK, Chia MA, Zhao A, Woodward-Court P, Xu M, et al. AutoMorph: Automated Retinal Vascular Morphology Quantification Via a Deep Learning Pipeline. Transl Vis Sci Technol. 2022;11:12.

[27] Manghi P, Bhosle A, Wang K, Marconi R, Selma-Royo M, Ricci L, et al. Coffee consumption is associated with intestinal Lawsonibacter asaccharolyticus abundance and prevalence across multiple cohorts. Nat Microbiol. 2024;9:3120-34.

[28] Marcus GM, Rosenthal DG, Nah G, Vittinghoff E, Fang C, Ogomori K, et al. Acute Effects of Coffee Consumption on Health among Ambulatory Adults. N Engl J Med. 2023;388:1092-100.

[29] Okada J, Osaki A, Shimoda Y, Yamada E, Saito T, Ozawa A, et al. Differences in the effects of Kenyan, Tanzanian, and Ethiopian coffee intake on interstitial glucose levels measured by FreeStyle Libre: A pilot case study. Curr Ther Res Clin Exp. 2020;93:100606.

[30] Nehlig A. Interindividual Differences in Caffeine Metabolism and Factors Driving Caffeine Consumption. Pharmacol Rev. 2018;70:384-411.

[31] Ungvari Z, Kunutsor SK. Coffee consumption and cardiometabolic health: a comprehensive review of the evidence. Geroscience. 2024;46:6473-510.

[32] González S, Salazar N, Ruiz-Saavedra S, Gómez-Martín M, de Los Reyes-Gavilán CG, Gueimonde M. Long-Term Coffee Consumption is Associated with Fecal Microbial Composition in Humans. Nutrients. 2020;12.

[33] Dai A, Hoffman K, Xu AA, Gurwara S, White DL, Kanwal F, et al. The Association between Caffeine Intake and the Colonic Mucosa-Associated Gut Microbiota in Humans-A Preliminary Investigation. Nutrients. 2023;15.

[34] Chen Z, Radjabzadeh D, Chen L, Kurilshikov A, Kavousi M, Ahmadizar F, et al. Association of Insulin Resistance and Type 2 Diabetes With Gut Microbial Diversity: A Microbiome-Wide Analysis From Population Studies. JAMA Netw Open. 2021;4:e2118811.

[35] Menni C, Lin C, Cecelja M, Mangino M, Matey-Hernandez ML, Keehn L, et al. Gut microbial diversity is associated with lower arterial stiffness in women. Eur Heart J. 2018;39:2390-7.

[36] Harpaz E, Tamir S, Weinstein A, Weinstein Y. The effect of caffeine on energy balance. J Basic Clin Physiol Pharmacol. 2017;28:1-10.

[37] Velickovic K, Wayne D, Leija HAL, Bloor I, Morris DE, Law J, et al. Caffeine exposure induces browning features in adipose tissue in vitro and in vivo. Sci Rep. 2019;9:9104.

[38] Ross R, Neeland IJ, Yamashita S, Shai I, Seidell J, Magni P, et al. Waist circumference as a vital sign in clinical practice: a Consensus Statement from the IAS and ICCR Working Group on Visceral Obesity. Nat Rev Endocrinol. 2020;16:177-89.

[39] Yafasova A, Fosbøl EL, Schou M, Baslund B, Faurschou M, Docherty KF, et al. Long-Term Cardiovascular Outcomes in Systemic Lupus Erythematosus. J Am Coll Cardiol. 2021;77:1717-27.

[40] Tso TK, Huang HY, Chang CK, Liao YJ, Huang WN. Clinical evaluation of insulin resistance and beta-cell function by the homeostasis model assessment in patients with systemic lupus erythematosus. Clin Rheumatol. 2004;23:416-20.

[41] Di Angelantonio E, Sarwar N, Perry P, Kaptoge S, Ray KK, Thompson A, et al. Major lipids, apolipoproteins, and risk of vascular disease. Jama. 2009;302:1993-2000.

[42] Chang HC, Nfor ON, Ho CC, Chen PH, Kung YY, Hsu SY, et al. Changes in High-Density Lipoprotein Cholesterol Levels in Relation to Coffee Consumption Among Taiwanese Adults. J Multidiscip Healthc. 2020;13:1427-32.

[43] Gebeyehu GM, Feleke DG, Molla MD, Admasu TD. Effect of habitual consumption of Ethiopian Arabica coffee on the risk of cardiovascular diseases among non-diabetic healthy adults. Heliyon. 2020;6:e04886.

[44] Iriondo-DeHond A, Rodríguez Casas A, Del Castillo MD. Interest of Coffee Melanoidins as Sustainable Healthier Food Ingredients. Front Nutr. 2021;8:730343.

[45] Hsu JC, Yang YY, Chuang SL, Huang KC, Lee JK, Lin LY. Long-Term Visit-to-Visit Glycemic Variability as a Predictor of Major Adverse Limb and Cardiovascular Events in Patients With Diabetes. J Am Heart Assoc. 2023;12:e025438.

[46] Knutson KL, Van Cauter E, Rathouz PJ, Yan LL, Hulley SB, Liu K, et al. Association between sleep and blood pressure in midlife: the CARDIA sleep study. Arch Intern Med. 2009;169:1055-61.

[47] Liu D, Li ZH, Shen D, Zhang PD, Song WQ, Zhang WT, et al. Association of Sugar-Sweetened, Artificially Sweetened, and Unsweetened Coffee Consumption With All-Cause and Cause-Specific Mortality : A Large Prospective Cohort Study. Ann Intern Med. 2022;175:909-17.

Figure 1. Illustration of the high-resolution multimodal data from the Human Phenotype Project, showing (A) 12 biological systems and (B) their collection timeline.

Figure 2. Ranks of the biological systems by their capacity to reflect habitual coffee intake-related variations. The performance for each system was assessed based on the Pearson correlation coefficient between the actual and predicted coffee consumption using five-fold cross-validation with elastic net regression or XGBoost (whichever performed better), trained on system-specific features along with age and sex, and repeated ten times.

Figure 3. Coffee consumption and glucose regulation captured by continuous glucose monitoring. (A) Pearson correlation matrix of CGM-derived metabolic metrics. (B) Associations between habitual coffee intake and individual CGM metrics. Multivariable linear regression models were performed, adjusted for age, sex, educational attainment, household income, smoking status, alcohol intake, moderate-to-vigorous physical activity, dietary pattern, daily energy intake, tea consumption, and percentage of coffee consumption from decaffeinated coffee. Benjamini–Hochberg False Discovery Rate (FDR)-adjusted P-value < 0.05 was considered statistically significant. Measurements were standardized before data analyses.

Figure 4. Habitual coffee intake and gut microbial composition. (A) Associations with species-level richness. (B) Associations with species-level Shannon diversity index. (C) Associations with the first two principal coordinates (PCoA) derived from Bray–Curtis dissimilarity. (D) Associations with relative abundances of individual gut microbial species (centered log-ratio transformed). (E) Mean (±SE) of CLR-transformed relative abundance of two microbial species exhibiting non-linear associations with coffee intake (1 cup = 200 g of coffee).

Figure 5. Associations of habitual coffee intake with measures from (A) body composition, (B) blood lipids, (C) liver system, (D) hematopoietic system, and (E) renal system. Serum lipidomics were rank-based inverse-normal transformed, and the other measurements were z-score standardized. Multivariable linear regression models were performed, adjusted for age, sex, educational attainment, household income, smoking status, alcohol intake, moderate-to-vigorous physical activity, dietary pattern, daily energy intake, tea consumption, and percentage of coffee consumption from decaffeinated coffee. Benjamini–Hochberg False Discovery Rate (FDR)-adjusted P-value < 0.05 was considered statistically significant.

Highlights

- Coffee intake was favorably associated with multiple human biological systems
- CGM and gut microbiome are the most reflective systems of coffee intake
- Multifaceted favorable glycemic profiles were associated with higher coffee intake