





The health impacts and genetic architecture of food liking in cardio-metabolic diseases

Received: 16 October 2024

Accepted: 5 May 2025

Published online: 23 May 2025

 Check for updates

Wenbo Jiang^{1,2,8}, Hang Wang^{1,3,8}, Yiding Geng^{1,4}, Meijuan Guo^{1,4}, Yingdong Zuo¹, Sijia Han^{1,4}, Zijie Liu^{1,4}, Shuaijun Chen¹, Shuzhen Fan^{1,4}, Shangying Li^{1,4}, Conghui Qiao¹, Qianzhu Li¹, Bai Li⁵, Yunpeng Zhang⁶ , Wei Wei¹  & Tianshu Han^{1,4,7} 

We evaluated temporal and genetic relationships between 176 food-liking-traits and cardio-metabolic diseases using data from the UK Biobank (N = 182,087) for observational analyses and summary-level GWAS data from FinnGen and other consortia (N = 406,565–977,323) for genetic analyses. Integrating observational and genetic results, we identified two detrimental food-liking-traits (bacon and diet-fizzy-drinks) and three protective food-liking-traits (broccoli, pizza, and lentils/beans). These food-liking-traits are associated with habitual food intake and influence cardio-metabolic proteins and biological processes. Notably, we found three genetic links: diet-fizzy-drinks with heart-failure, bacon with type-2-diabetes, and lentils/beans with type-2-diabetes, identifying 54 pleiotropic single-nucleotide-variants, impacting both phenotypes. Our data show the diet-fizzy-drinks and heart-failure link maybe not direct, as diet-fizzy-drinks liking correlates with sweet food consumption and shares variants linked to BMI, adiposity, platelet count and cardio-metabolic traits. The pleiotropic single-nucleotide-variants map to 251 tissue-specific genes, with four showing high druggability potential, highlighting personalized dietary strategies for cardio-metabolic diseases.

The increasing availability of food in the Western world, driven by a mass food supply system, provides humans with more autonomous food choices. Instead of simply eating what is accessible, people are now driven by personal choices in their food selection and intake. Food liking, which reflects an individual's hedonic responses to foods¹, has emerged as a significant driver of food choice and intake behavior^{2,3}. The food liking is shaped by a complex interplay of genetic inheritance, biology, psychology, and environmental factors^{4–7}. Alongside efforts to understand the impact of food intake on health,

there is growing interest in examining the potential role of food liking in relation to human health. This understanding is critical as it could guide the development of more targeted dietary interventions and the creation of nutritious foods that are more acceptable to consumers based on their personal food liking, which ultimately promotes healthier food intake through a deeper understanding of food liking for reducing the burden of various diseases.

The close association between food intake and cardiometabolic diseases (CMDs) has been well established in numerous studies^{8,9}.

¹Key Laboratory of Precision Nutrition and Health, Ministry of Education. Department of Nutrition and Food Hygiene, the National Key Discipline, School of Public Health, Harbin Medical University, Harbin, China. ²Department of Cardiology, The First Affiliated Hospital of Harbin Medical University, Harbin, China. ³Department of Colorectal surgery, Harbin Medical University Cancer Hospital, Harbin, China. ⁴Department of Endocrinology, The Second Affiliated Hospital of Harbin Medical University, Harbin, China. ⁵Department of Biology, University of Ottawa, 30 Marie-Curie Private, Gendron Hall, Ottawa, ON K1N 9B4, Canada. ⁶College of Bioinformatics Science and Technology, Harbin Medical University, Harbin, China. ⁷State Key Laboratory of Frigid Zone Cardiovascular Diseases, Harbin Medical University, Harbin, China. ⁸These authors contributed equally: Wenbo Jiang, Hang Wang. ✉ e-mail: zhangyp@hrbmu.edu.cn; 102594@hrbmu.edu.cn; 102403@hrbmu.edu.cn

However, the temporal relationship between food liking, an important determinant factor of food intake, and the development of CMDs largely remains unexplored. Moreover, beyond the temporal relationship, it is likely that genetic factors also play a substantial role in shaping these traits and their connection, given that both food-liking traits (FLT) and CMDs exhibit high heritability^{10–12}. Previous research indicates that genetically determined physiological factors in the brain exert a greater influence on individual FLT than environmental factors¹³. In addition to regulating FLT, the brain also plays a central role in neuro-metabolic regulation and inter-organ genetic communication¹⁴, including with the heart and vasculature¹⁵. Such mechanisms suggest that genetic variations influencing FLT contribute to metabolic and cardiovascular processes, thereby supporting a potential genetic link between FLT and CMDs. Furthermore, recent GWAS findings have reported significant genetic correlations between FLT and key metabolic risk factors for CMDs, including body fat, blood glucose, and lipids¹⁰. Since these metabolic traits are well-established contributors to CMD development and share common genetic determinants with CMDs^{16–18}, this further supports the hypothesis that FLT and CMDs may share overlapping genetic architecture. Identifying these pleiotropic variants is crucial for understanding the genetic basis of dietary behaviors and CMD susceptibility. For example, candidate gene studies have shown that variations in genes encoding taste receptors¹⁹, such as *TAS2R43*, are expressed in the heart and are associated with inflammation, oxidative stress, and cardiac rhythm^{20,21}. Therefore, a genome-wide investigation to examine the genetic architecture between FLT and CMDs is essential. Although this approach may not reflect a direct causal relationship between FLT and CMDs, it can strengthen our understanding of the health impacts of FLT on CMDs and can help identify pleiotropic variants that could serve as potential intervention targets.

This study utilized a large-scale dataset from the UK Biobank (UKB) and summary-level GWAS data to evaluate the temporal, genetic, and pleiotropic relationships between FLT and CMDs. Specifically, we investigated the associations between various FLT and CMDs using longitudinal data. We evaluated the consistency between FLT and food intake patterns, and their effects on cardiometabolic proteins to uncover their interconnections and shared biological mechanisms that influence CMD development. In addition, we used Mendelian randomization analysis and examined genetic correlations and overlapping genetic factors between FLT and CMDs. This included identifying pleiotropic variants, genetic intervention targets, enriched biological pathways, and tissue-specific expression patterns to further understand their interconnected mechanisms.

Results

Cox proportional hazards model analysis for food liking and CMDs

The overview of the epidemiological design is depicted in Fig. 1a. A total of 182,087 participants from the UK Biobank, who provided information on food liking via questionnaires, blood samples, and medical records, were included, with a median follow-up period of 10.4 years. The 176 specific FLT were categorized into 10 major food-liking groups: alcohol/beverage, coffee/tea, dairy/cheese, desserts, fruits, grains, processed products, animal foods, vegetables, and condiments. Further details of the FLT within these 10 major food groups are shown in Fig. 1b. Baseline characteristics of the study population, stratified by the baseline status of CMDs, are presented in Supplementary Table 1. Compared to the overall sample, participants with CMDs were more likely to be older, men, and have lower levels of education and household income. In contrast, they had higher BMI, smoking, and drinking rates (all $P < 0.05$).

As shown in Fig. 1c, positive associations between FLT and a higher risk of CMDs were primarily observed for traits related to alcohol/beverage, processed products, and animal products.

Conversely, negative associations were mainly identified for traits related to fruits and vegetables/beans, after adjusting for potential covariates in the Cox proportional hazards model (all P -values with false discovery rates [FDR] adjustments < 0.05 , Supplementary Data 1). Moreover, the results of sensitivity analyses showed that excluding participants with high CMD risk or whose self-reported sex did not align with their genetically determined sex did not significantly change the associations between FLT and CMDs. This suggests that the potential reverse causation effect due to dietary behavior changes in high-risk populations or disagreement in sex and gender is unlikely to significantly affect these associations (Supplementary Fig. 1 and Data 2 and Supplementary Fig. 2 and Data 3), and no significant modification effects of sex on the associations between FLT and CMDs were detected (Supplementary Fig. 3).

Univariable Mendelian randomization (MR) analysis for food liking and CMDs

To provide more comprehensive evidence on the relationship between FLT and CMDs, we conducted univariate MR analysis. Although this approach cannot provide a direct causal relationship, it could strengthen the results of the temporal relationship between FLT and CMDs documented above. Summary-level GWAS data for 176 specific FLT were obtained from the UKB¹⁰. For CMD-related outcomes, we used large-scale summary-level GWAS data from the FinnGen²², MegaStroke²³, and Hermes consortia²⁴, ensuring that exposure and outcome datasets were independent. Detailed information on the GWAS data for each outcome is provided in Supplementary Table 2. The process is outlined in Fig. 2a, with significant causal effects on different types of CMDs, adjusted for FDR, shown in Fig. 2b. All univariate MR results were validated through at least one sensitivity analysis (Supplementary Data 4). Although most MR-Egger results indicated null associations, the estimates were generally comparable to those from inverse-variance weighted (IVW) results, albeit with wider 95% confidence intervals (CIs). Only minimal heterogeneity was observed between instrumental variables for the two FLT, and the instrumental validity test suggested sufficient instrument strength (all F -statistics > 59 ; Supplementary Data 4). No horizontal pleiotropy was detected (all P -values for Egger intercept > 0.20), and no reverse causality was detected between instrumental variables (all P -values for Steiger test < 0.05) (Supplementary Data 5).

Genetic correlation and univariate MR analysis for covariates and CMDs

It has been documented that household income, physical activity, and educational attainment influence an individual's food liking⁷, and these factors are also associated with CMD development^{25,26}. Therefore, we considered these variables as potential confounders in the causal association between FLT and CMDs. Linkage disequilibrium score (LDSC) and high-definition likelihood regression identified 54 and 63 significant genetic correlations, respectively, between the three confounders and FLT. The absolute genetic correlation values $****(|r_g|)$ ranged from 0.089 to 0.52 ($P = 3.91 \times 10^{-2}$ to 2.60×10^{-6}), and correlation values of high-definition likelihood regression ($r_{\text{high-definition likelihood}}$) ranged from 0.033 to 0.47 ($P = 4.32 \times 10^{-2}$ to 5.28×10^{-168}). All of these confounders were also genetically correlated with CMDs $****(|r_g| = 0.15$ to 0.35 , $P = 7.25 \times 10^{-5}$ to 7.78×10^{-41} ; $|r_{\text{high-definition likelihood}}| = 0.16$ to 0.37 , $P = 7.17 \times 10^{-5}$ to 1.76×10^{-45}) (Supplementary Data 6), further supporting that they may act as potential confounders in the genetically causal association between FLT and CMDs. Meanwhile, the genetic covariance intercept ranged from -0.1 to 0.1 , suggesting minimal inflation due to sample overlap or population stratification (Supplementary Data 6). Moreover, we observed that genetically determined higher household income and educational attainment were associated with a lower risk

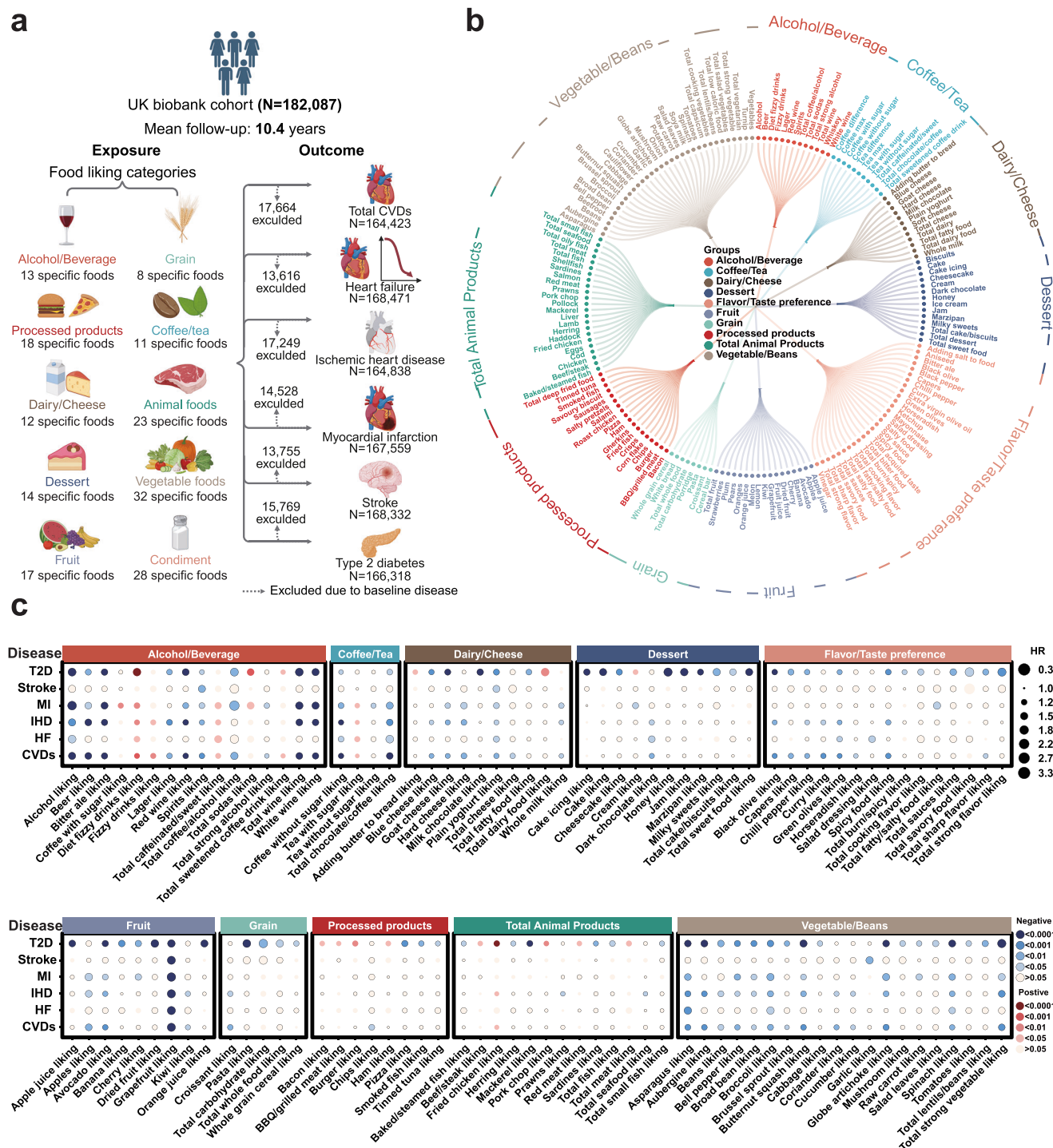
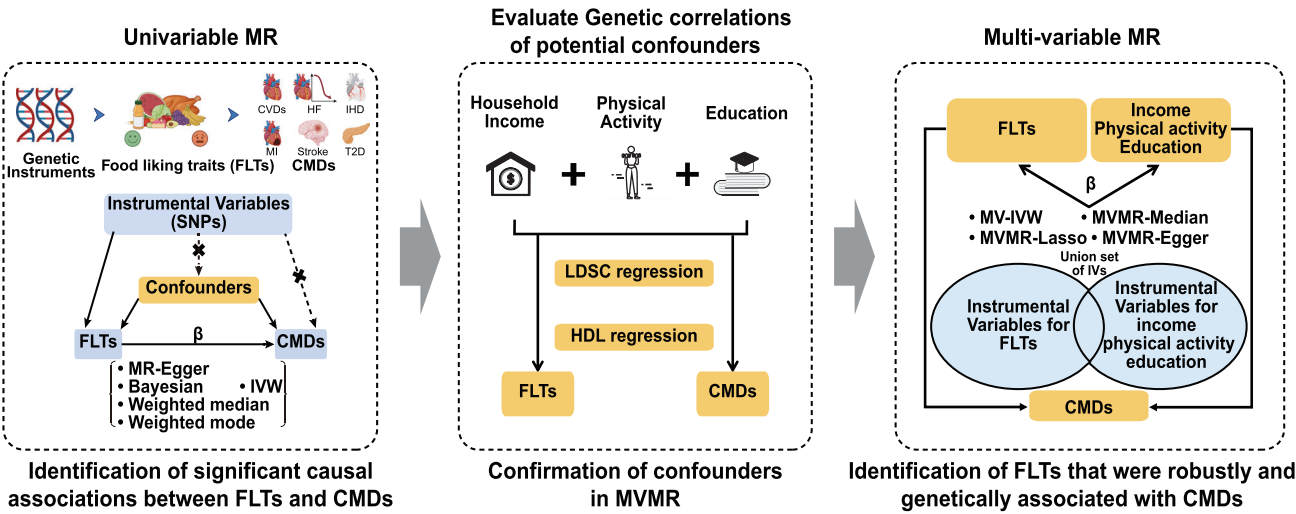


Fig. 1 | Temporal relationships between FLT and CMDs. **a** Overview of the epidemiological design of the prospective cohort in the UK Biobank; Created in BioRender. <https://BioRender.com/udq2dks>; **b** Categories of 176 FLT by ten major food categories; **c** Heatmap plot showing the association between FLT and CMDs. Each circle represents the HRs estimating the association between FLT and CMDs from Cox proportional hazards regression models, adjusted for age, sex, smoking status, drinking status, BMI, education level, income, TDI, MET, SBP, DBP, HDL-C, LDL-C, TC, and TG, with color indicating the direction of the association; The circle

size reflects the effect size of FLT on each CMD, and a two-sided P -FDR < 0.05 was considered statistically significant. FLT food-liking traits, CMDs cardiometabolic diseases, HR hazards ratio, T2D type 2 diabetes, MI myocardial infarction, IHD ischemic heart disease, HF heart failure, CVDs cardiovascular diseases, BMI body mass index, TDI Townsend deprivation index, MET metabolic equivalent of task, SBP systolic blood pressure, DBP diastolic blood pressure, HDL-C high density lipoprotein cholesterol, LDL-C low density lipoprotein cholesterol, TC total cholesterol, TG triglycerides, P -FDR P -values with false discovery rates.

a



b

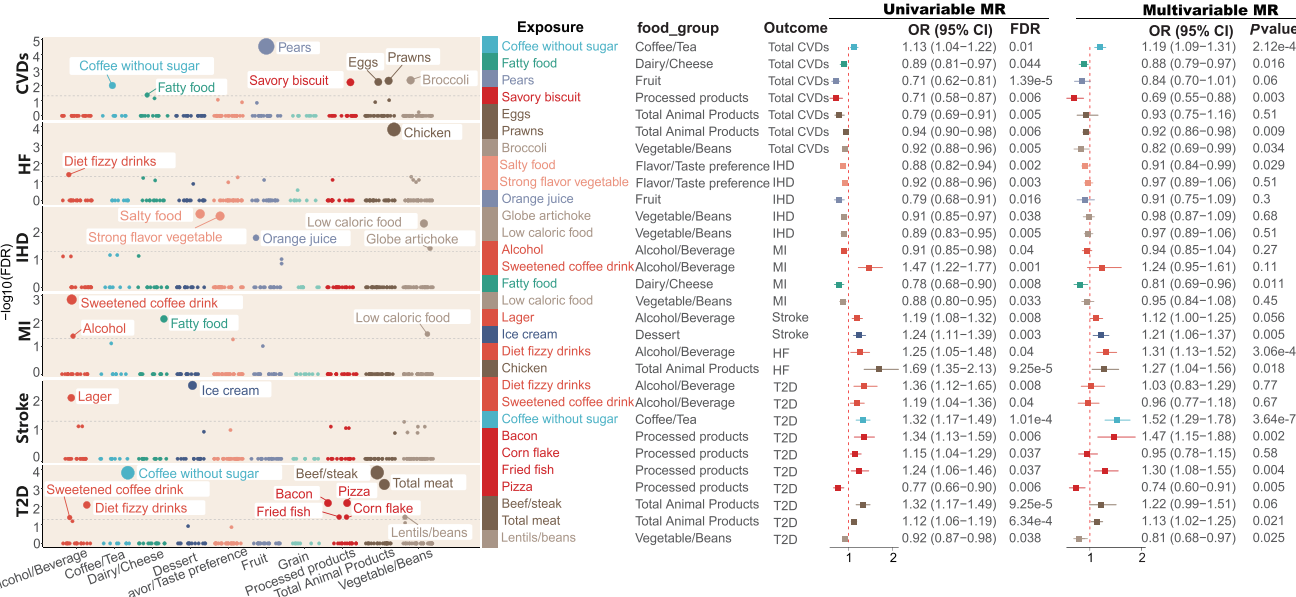


Fig. 2 | Genetic associations between FLT and CMDs. a Overview of the design of the MR analysis; Created in BioRender. <https://BioRender.com/r07vihs>. **b** UVMR and MVMR for FLT and CMDs. Left panel: Manhattan plot showing the P -FDR values in the UVMR for 10 categories of FLT. The height of each point represents the negative logarithm of the P -FDR value for the IVW, with the color bar indicating different FLT categories. The black dashed line indicates the P -FDR threshold (a two-sided P -FDR for UVMR-IVW < 0.05), and the specific FLT showing significant differences after FDR correction are annotated with text labels. Right panel: Forest plots for the ORs and 95% CIs estimated by UV-IVW and MV-IVW, adjusted for household income, educational attainment, and physical activity levels. A two-

sided P -FDR for UVMR-IVW and P -value for MVMR < 0.05 was considered statistically significant. Sample size [diseases (case/N)]: Total CVDs (221,781/231,952); IHD (75,592/378,141); MI (28,546/378,019); Stroke (67,162/454,450); HF (47,309/930,014); T2D (71,728/369,007). FLT food-liking traits, CMDs cardiometabolic diseases, CVDs cardiovascular diseases, IHD ischemic heart disease, HF heart failure, MI myocardial infarction, T2D type 2 diabetes, LDSC linkage disequilibrium score regression, HDL high-definition likelihood, IVW inverse-variance weighted, UVMR univariate Mendelian randomization, MVMR multi-variable Mendelian randomization, OR odds ratio, 95% CI 95% confidence interval, MR Mendelian randomization.

of CMDs, including total cardiovascular diseases (CVDs), ischemic heart disease (IHD), myocardial Infarction (MI), stroke, heart failure (HF), and type 2 diabetes (T2D) (Supplementary Data 7). No horizontal pleiotropy was detected (all P -values for Egger intercept > 0.10) (Supplementary Table 3). These results suggest that these factors may act as potential confounders in the genetically causal association between FLT and CMDs.

Multi-variable MR analysis for FLT and CMDs

The FLT identified in the univariate MR were retained in the multi-variable MR analysis, which was adjusted for the identified confounders. A comparison of the results between univariable and multi-variable MR is shown in Fig. 2b. As indicated by the odds ratios (ORs) and 95% CIs estimated using multi-variable IVW analysis, genetically determined liking for coffee without sugar was associated with a

higher risk of total CVDs (OR: 1.19, 95% CI: 1.09–1.31) and T2D (OR: 1.52, 95% CI: 1.29–1.78). Similar associations were observed for liking biscuits (OR: 1.14, 95% CI: 1.00–1.30) and ice cream (OR: 1.21, 95% CI: 1.06–1.37) with stroke, and for diet fizzy drink liking (OR: 1.31, 95% CI: 1.13–1.52) and porridge liking (OR: 1.21, 95% CI: 1.09–1.34) with HF. Additionally, bacon liking (OR: 1.47, 95% CI: 1.15–1.88), fried fish liking (OR: 1.30, 95% CI: 1.08–1.55), and total meat liking (OR: 1.13, 95% CI: 1.02–1.25) were associated with T2D.

In contrast, genetically determined liking for certain foods showed protective causal associations with a lower risk of CMD. For example, liking total fatty foods (OR: 0.88, 95% CI: 0.79–0.97), total salty foods (OR: 0.90, 95% CI: 0.84–0.97), orange juice (OR: 0.72, 95% CI: 0.61–0.85), savory biscuits (OR: 0.69, 95% CI: 0.55–0.88), prawns (OR: 0.92, 95% CI: 0.86–0.98), and broccoli (OR: 0.82, 95% CI: 0.69–0.99) were associated with a lower risk of total CVDs. Similar protective associations were observed between liking total salty foods and lower risk of IHD (OR: 0.91, 95% CI: 0.84–0.99), liking total fatty foods and lower risk of MI (OR: 0.81, 95% CI: 0.69–0.96), liking lentils/beans and pizza and lower risk of T2D (OR: 0.81, 95% CI: 0.68–0.97 and OR: 0.74, 95% CI: 0.60–0.91). All multi-variable MR results were validated through at least one sensitivity analysis (Supplementary Data 8). Although heterogeneity was observed between some instrumental variables, the multi-variable MR-Egger method provided more robust causal estimates in the presence of heterogeneity, further supporting the results from the IVW analysis (Supplementary Data 8). No evidence of horizontal pleiotropy was detected (all P -values for Egger intercept >0.07 ; Supplementary Data 9).

Among the FLT-CMD pairs identified in multi-variable MR, we performed bidirectional MR analysis to examine if genetically predicted CMDs influence FLTs. We found that only genetically predicted total CVDs showed a bidirectional causal relationship with prawn liking, suggesting that the unidirectional genetic relationships between other FLTs and CMDs remain stable (Supplementary Data 10 and Supplementary Data 11).

Potential biological mechanisms underlying identified FLTs and CMDs

Based on the results from the Cox proportional hazards models, univariate and multi-variable MR analyses, five pairwise associations between FLTs and CMDs were identified. These associations included liking broccoli and CVDs, liking diet fizzy drinks and HF, and liking bacon, pizza, and lentils/beans and T2D (Fig. 3a). Given these findings, we focused on investigating potential biological mechanisms linking these FLTs to their corresponding CMDs.

The detailed analysis process for exploring these potential biological mechanisms is illustrated in Fig. 3b. These FLTs resulted in distinct habitual dietary intake patterns (Fig. 3c, d). Both genetically determined liking for bacon and its reflection by actual measurement were associated with increased consumption of red meats, alcoholic beverages, and potatoes/starchy foods, while being negatively associated with the intake of seafood, fruits, and nuts/seeds. Liking for diet fizzy drinks, indicated in two aspects, showed similar dietary intake patterns but was strongly positively associated with the intake of sugar and low-sugar sweetened beverages. On the other hand, liking for broccoli and lentils/beans showed similar dietary intake patterns, with both positively associated with the intakes of vegetables, fruits, and nuts/seeds, while negatively correlated with sugar and low-sugar sweetened beverages.

Since the UKB cohort did not measure all corresponding intakes of these FLTs, we evaluated the correlations between bacon liking and processed meat intake ($r = 0.23$, $P < 0.001$), diet fizzy drinks liking and low-sugar sweetened beverages ($r = 0.39$, $P < 0.001$), broccoli liking and green leafy vegetable intake ($r = 0.15$, $P < 0.001$), lentils/beans liking and legumes intake ($r = 0.13$, $P < 0.001$), and pizza liking and pizza intake ($r = 0.11$, $P < 0.001$). These results suggest that FLTs are

modestly associated with actual food intake, indicating that an individual's food liking can modestly reflect their actual dietary habits. Moreover, we also observed that these FLTs were associated with social factors and anthropometric indices. Liking for bacon and diet fizzy drinks was associated with increased levels of body fat and fat-free mass, whereas liking for pizza was only linked to increased fat-free mass. Additionally, a liking for pizza was associated with higher income and lower Townsend deprivation index levels (all $P < 0.05$) (Supplementary Fig. 4).

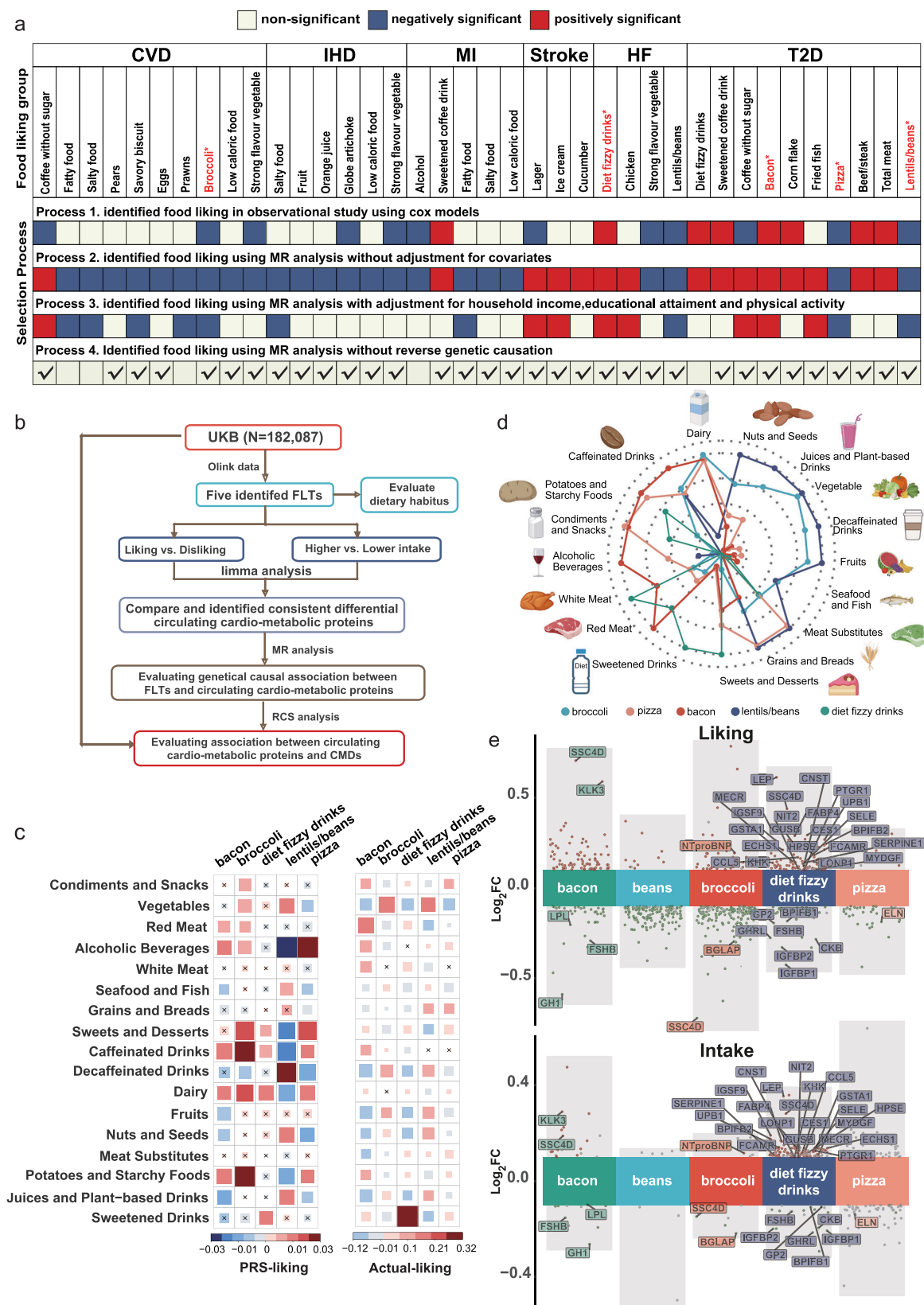
To evaluate whether and how food intake partially mediated the association between FLTs and CMDs, we further examined the influence of FLTs and their corresponding food intakes on circulating cardiometabolic proteins (Fig. 3e). We identified a total of 35 circulating proteins that overlapped between liking and intake with consistent changes in direction (all P -FDR < 0.05 , Fig. 4a). These proteins are mainly involved in glucose and lipid metabolism, inflammation, oxidative stress, and immunoregulation. Additionally, we observed that, after adjusting for potential covariates, most of these protein expressions were correlated with increasing hedonic responses to the corresponding FLTs (Fig. 4b). Furthermore, we examined the association between these circulating proteins and the disease outcomes related to their specific FLTs (Fig. 4c, d). As indicated by Cox proportional hazards regression models, elevations in most of these proteins were associated with a greater risk of T2D, CVD, and HF, showing a linear relationship. Only higher levels of lipoprotein lipase (LPL) were associated with a lower risk of T2D. To strengthen the evidence linking circulating proteins to CMDs, we also performed summary data-based MR analysis with reference protein quantitative trait locus (pQTL) data to identify genetic causal associations. We observed significant associations between LPL and T2D ($\beta = 0.10$, $P = 0.028$ & Heterogeneity in Dependent Instrument (HEIDI) = 0.42), as well as Ketohexokinase (KHK) ($\beta = 0.06$, $P = 0.003$ & HEIDI = 0.96), Insulin Like Growth Factor Binding Protein 2 (IGFBP2) ($\beta = 0.50$, $P = 0.009$ & HEIDI = 0.42) and Glutathione S-transferase A1 (GSTA1) ($\beta = -0.06$, $P = 0.028$ & HEIDI = 0.59) and HF (Supplementary Data 12).

Genetic correlations and shared genetic loci between FLTs and CMDs

The overview of the genetic correlation assessment, identification of shared genetic loci, and confirmation of pleiotropic loci and genes is presented in Fig. 5a. To further investigate the health impacts of FLTs, we explored whether the pairwise associations between FLTs and CMDs share a common genetic basis. Using LDSC and high-definition likelihood analyses, we identified significant genetic correlations between HF and diet fizzy drinks, T2D and bacon, as well as T2D and lentils/beans. The absolute genetic correlation values $****(|r_g|)$ ranged from 0.14 to 0.32, and the correlation values of high-definition likelihood ($r_{\text{high-definition likelihood}}$) ranged from 0.10 to 0.37 (Fig. 5b). Additionally, based on the pleiotropic analysis under the composite null hypothesis (PLACO) and the Functional Mapping and Annotation of genetic associations (FUMA) platform, a total of 54 pleiotropic genomic risk single-nucleotide variants (SNVs) were identified (HF-diet fizzy drinks: 35; T2D-bacon: 6; T2D-lentils/beans: 13) ($P < 1.00 \times 10^{-5}$, Supplementary Data 13 and Data 14). Notably, 41 of the 54 SNVs (70.8%) demonstrated concordant associations with both traits, suggesting that these variants may simultaneously influence both FLTs and CMDs (Supplementary Data 15).

Phenome-wide associations of the genomic risk SNVs in FLTs and CMDs pairs

We investigated phenotypic associations of the identified lead SNVs (including genomic risk SNVs) using existing GWAS literature. A phenome-wide association query revealed a total of 1737 significant associations (1678 unique) between these pleiotropic lead SNVs and various clinical phenotypes (HF-diet fizzy drinks: 1172; T2D-bacon: 173;



T2D-lentils/beans: 392 ($P < 1.00 \times 10^{-5}$, Supplementary Data 16). For the HF-diet fizzy drinks pair trait, associations were predominantly linked to indicators of body mass index (BMI), body composition, brain structure and function, as well as various CMD indicators. In the T2D-bacon pair trait, associations were primarily connected to diabetes, body composition, glycolipid metabolism indicators, physical

activity, and dietary habits. In addition to the previously mentioned phenotypes, we also found that these lead SNVs were associated with liver function, oxidative stress and inflammation indicators (Fig. 5c). Notably, we identified seven new loci (*rs11052302*, *rs75571852*, *rs80169769*, *rs4525110*, *rs1051909*, *rs111411461*, *rs35297435*) that have not been previously associated with any clinical phenotypes.

Fig. 3 | Potential biological mechanisms underlying identified FLT and CMDs. **a** Selection process for FLT and their corresponding diseases based on the consistency of temporal relationships and genetic associations. Four criteria were applied to screen for FLT that were both temporally and genetically associated with CMDs: (1) The FLT should be temporally associated with CMDs; (2) and (3) The FLT should be genetically associated with CMDs, as indicated by UVMR and MVMR (adjusted for household income; educational attainment, and physical activity); (4) The genetic association between FLT and CMDs should be unidirectional. **b** Overview of the analysis framework for potential biological mechanisms underlying the five identified FLT and their corresponding CMDs. **c** Heatmap plot showing the correlation between the five identified FLT and food intake, as indicated by PRS ($N = 206,451$) and questionnaire measurements ($N = 124,352$). Each cell represents Spearman correlation coefficients, and a two-sided $P < 0.05$ was considered statistically significant; **d** Radar chart depicting the dietary intake of the 17 food categories by the five identified FLT ($N = 124,352$). Created in BioRender.

<https://BioRender.com/x3i0573>. **e** Visualization of differential circulating cardiometabolic proteins between individuals with extremely liking and extremely disliking for specific foods ($N = 6061$ for bacon; $N = 7419$ for broccoli; $N = 8432$ for diet fizzy drinks; $N = 4792$ for lentils/beans; $N = 3163$ for pizza), and between the highest and lowest intake ($N = 7370$). Significantly up-regulated proteins are shown as red points, significantly down-regulated proteins are shown as green points, and non-significant proteins are shown as gray points. Overlapping proteins with the same direction of change are labeled with text annotations. A two-sided $P\text{-FDR} < 0.05$ was considered statistically significant. FLT food-like traits, CMDs cardiometabolic diseases, CVDs cardiovascular diseases, TDI Townsend deprivation index, MET metabolic equivalent of task, IHD ischemic heart disease, HF heart failure, MI myocardial infarction, T2D type 2 diabetes, UVMR univariate Mendelian randomization, MVMR multi-variable Mendelian randomization, PRS polygenic risk score, $P\text{-FDR}$ P -values with false discovery rates.

Gene mapping and druggability assessment

A total of 240 genes were identified through gene mapping analysis, 89 genes associated with HF-diet fizzy drinks, 17 genes associated with T2D-bacon, and 33 genes associated with T2D-lentils/beans, all located near the genomic risk SNVs (Fig. 6a, Supplementary Data 17). Additionally, 151 genes were identified through gene expression Quantitative Trait Locus (eQTL) analysis, including 143 for HF-diet fizzy drinks and 8 for T2D-lentils/beans ($P\text{-FDR} < 0.05$, Fig. 6a, Supplementary Data 17). Using gene-level multi-marker analysis of GenoMic annotation (MAGMA), 9 genes were identified as being associated with HF-diet fizzy drinks, 9 genes were associated with T2D-bacon, and 28 genes with T2D-lentils/beans ($P\text{-FDR} < 0.05$, Supplementary Data 18). Summary data-based MR analysis further revealed 152 genes for HF-diet fizzy drinks, 194 for T2D-bacon, and 171 for T2D-lentils/beans, suggesting potential pleiotropic drug targets for regulating both phenotypes in each trait pair ($P < 0.05$ & HEIDI > 0.05 , Supplementary Data 19 to Supplementary Data 21).

Figure 6b presents 16 genes identified by at least two of the methods mentioned above. Upon searching DrugCentral, the Drug Gene Interaction Database, and DrugBank, we found that genes such as *neutrophil cytosolic factor 1* (*NCF1*), *sterol regulatory element binding transcription factor 1* (*SREBF1*), *phenylethanolamine N-methyltransferase* (*PNMT*), and *erb-b2 receptor tyrosine kinase 2* (*ERBB2*) are already used as drug targets for diabetes complications and heart-related diseases with Food and Drug Administration (FDA) approved drugs (Supplementary Data 22). Although genes such as *AC007382.1*, *Zinc Finger Protein 131* (*ZNF131*), *Pleckstrin Homology Domain Interacting Protein* (*PHIP*), and *chromosome 18 open reading frame 8* (*C18orf8*) were simultaneously identified by gene mapping, summary data-based MR, and MAGMA methods, emphasizing their important role in regulating both FLT and CMDs, these genes currently do not have FDA-approved drugs.

Gene-set enrichment analysis (GSEA) for biological pathway

The identified pleiotropic genomic risk loci prompted us to investigate the pleiotropic biological process in each FLT and CMD pair (Fig. 7a), which could enhance our understanding of the connection between FLT and CMDs. MAGMA gene-set analysis revealed a total of 2637 significantly enriched pathways ($P < 0.05$, Supplementary Data 23). Since immune-inflammatory responses and metabolic regulation have been widely reported to be closely related to heart disease^{27,28}, and since brain development and energy metabolism play key roles in determining individual FLT^{29,30}, we specifically focused on the top 10 pathways associated with these biological processes in each trait pair. For the HF-diet fizzy drinks trait pair, genes were notably enriched in pathways related to T lymphocyte and NK progenitor activity, regulation of synaptic vesicle transport, and muscle hypertrophy in response to stress (top 10 pathways, $P < 0.05$, Fig. 7b). For the T2D-bacon trait pair, gene sets were significantly involved in pathways

including amygdala development, sulfur transferase activity, and IL2 responsive FOXP3 targets (top 10 pathways, $P < 0.05$, Fig. 7c). The T2D-lentils/beans trait pair showed enrichment in pathways related to the MHC protein complex, FOXP3 targets clusters, the caspase cascade, and catabolic processes (top 10 pathways, $P < 0.05$, Fig. 7d). Moreover, Metascape analysis identified a total of 459 enriched pathways for the gene sets derived from genomic risk SNVs mapping analysis and the MAGMA platform. These pathways were largely involved in immune-inflammatory responses, energy metabolism, hormone regulation of glucose and lipid metabolism, metal ion homeostasis and response, and the development and differentiation signaling of the nervous system across the three pair traits ($P < 0.05$, Fig. 8a–c, Supplementary Data 24).

Tissue-specific gene expression analysis (TSEA) for pleiotropic gene expression patterns

To explore the pleiotropic gene expression patterns of the three trait pairs, we conducted TSEA using the MAGMA platform, imputed with the Genotype-Tissue Expression project (GTEx) RNA-sequencing dataset. Among 54 human tissues, genes associated with the HF-diet fizzy drinks trait pair demonstrated tissue specificity in various brain regions ($P < 0.05$). Additionally, genes related to the T2D-bacon trait pair showed significant over-expression in cultured fibroblasts and EBV-transformed lymphocytes ($P < 0.05$). For the T2D-lentils/beans trait pair, genes exhibited significant over-expression in the pancreas ($P < 0.05$, Fig. 8d).

Furthermore, we performed a TSEA of genes identified from genomic risk SNVs mapping and the MAGMA analyses using the decode tissue specificity (deTS) platform. The results indicated that genes associated with the HF-diet fizzy drinks trait pair showed higher tissue specificity in the minor salivary gland, stomach, gastroesophageal sphincter, breast epithelium, ovary, uterus, and prostate gland ($P < 0.05$). For the T2D-bacon trait pair, genes were significantly expressed in the lung and visceral adipose tissue (omentum) ($P < 0.05$). Similarly, genes related to the T2D-lentils/beans trait pair exhibited high expression in the minor salivary gland, breast tissue and breast epithelium ($P < 0.05$, Fig. 8d). The overall TSEA findings of pleiotropic genes expression patterns for the three trait pairs are presented in Fig. 8e. These results suggested that, in addition to regulating FLT and CMDs, these shared pleiotropic genes also regulate the development of these organs.

Discussion

This study evaluates the temporal relationship and genetic causal associations between 176 FLT and the development of CMDs, as well as the underlying biological mechanisms and shared genetic architecture. By integrating observational and genetic results, we identified five robust pairwise associations between FLT and CMDs, including two detrimental FLT (bacon and diet fizzy drinks) and three

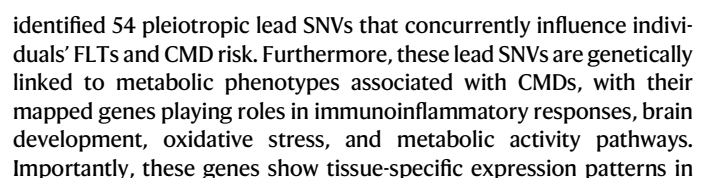


Fig. 4 | Association between identified circulating plasma proteins and incidence of CMDs. **a** Summary of the overlapping proteins. The arrows indicate the direction of change. **b** β coefficients estimated by linear regression models for the association between FLTs and circulating cardiometabolic proteins, adjusted for age, sex, BMI, TDI, and MET. **c** Forest plot showing the HRs and 95% CIs for the association between circulating proteins and corresponding diseases related to specific FLTs. HRs and 95% CIs were estimated by Cox proportional hazards regression models, adjusted for age, sex, race, smoking rate, education level, drinking rate, TDI, MET, BMI, SBP, DBP, HDL, LDL, TC, and TG. The dots and lines represent HRs and 95% CIs. The black dashed line indicates the P threshold (two-sided $P < 0.05$). **d** Restricted cubic spline for the association between circulating

cardiometabolic proteins and their corresponding CMDs, adjusted for age, sex, race, smoking rate, education level, drinking rate, TDI, MET, BMI, SBP, DBP, HDL, LDL, TC, and TG. The P -value for the linear regression test was assessed by ANOVA, and a two-sided $P < 0.05$ was considered statistically significant. The lines represent HRs, and the shaded area around the red line represents the 95% CIs. FLTs food-liking traits, CMDs cardiometabolic diseases, CVDs cardiovascular diseases, IHD ischemic heart disease, HF heart failure, MI myocardial infarction, T2D type 2 diabetes, TDI Townsend deprivation index, MET metabolic equivalent of task, BMI body mass index, SBP systolic blood pressure, DBP diastolic blood pressure, HDL-C high density lipoprotein cholesterol, LDL-C low density lipoprotein cholesterol, TC total cholesterol, TG triglycerides, HRs hazards ratio.

the brain, lymphocytes, visceral adipose tissue, and pancreas, among other tissues.

Previous epidemiological studies indicate that higher consumption of bacon, grilled meats, beef, and fried chicken correlates with an increased risk of T2D³¹, whereas higher intakes of coffee, tea, fruits, and vegetables are associated with a reduced risk^{31–33}. Our findings align with these observations, revealing similar associations from the aspects of food liking. Interestingly, while some food items can be grouped together as expected (e.g., fruits and vegetables), others reveal significant variability within seemingly homogeneous categories. Unlike previous studies that examined broader food categories, our research focuses on specific items, providing nuanced insights and complete supplements. For instance, few studies have assessed the impact of sweetened coffee or tea on CMD risk, and the health effects of diet fizzy drinks remain contentious. Our analysis shows that a liking for sweetened coffee or tea and diet fizzy drinks is associated with an increased risk of CMDs, including CVDs, HF, IHD, MI, and T2D. Conversely, a liking for certain fruits and vegetables, such as grapefruit, asparagus, broccoli, butternut squash, and globe artichoke, correlates with a lower risk of these conditions. These results emphasize the importance of specific food items in preventing and managing CMDs, providing more precise and practical strategies for dietary intervention in the prevention and management of CMDs.

Research has demonstrated that FLTs possess relatively high heritability, with physiological factors exerting a stronger influence than behavioral factors^{4,13,34}. Recent GWAS have identified 1401 genetic loci associated with various FLTs, providing a unique opportunity to examine the genetic causal relationships between FLTs and CMDs¹⁰. Using MR for genetic causal analysis, we established a genetic causal relationship between FLTs and CMDs. We observed that genetically determined liking for vegetables or low-calorie food was associated with a lower risk of IHD, while a genetically determined liking for red meat and bacon was associated with a higher risk of T2D, which further strengthened prior nutritional knowledge related to CMDs prevention. Our results also reveal several intriguing findings. First, a genetically determined liking for coffee is linked to a higher risk of total CVDs and T2D. The negative impacts of excessive coffee intake (over 8 cups/day) on CMDs have been documented³⁵, suggesting that a strong liking for coffee may elevate future CMD risk. Our study additionally provides genetic evidence for extremely high coffee intake and higher CMD risk. Second, a genetically determined liking for salty foods is associated with a lower risk of IHD, which contrasts with previous observational studies³⁶. Some research suggests that the body can metabolically adapt to a high-salt diet to maintain electrolyte homeostasis and regulate gene expression, such as *FoxP2*^{37–39}. We also found that the *rs2049603* variant on *FoxP2* is closely linked to salty food liking. Although the involvement of *FoxP2* in maintaining electrolyte homeostasis is still unknown, studies have shown that FLTs have evolutionary origins^{10,13,19,40,41}. Thus, we hypothesize that long-term high salt intake could shift an individual's salt preference, enabling better adaptation to high-salt diets and improving survival chances. More targeted evolutionary genetics studies are necessary to test this

hypothesis. Third, current studies have suggested that dairy is a heterogeneous food group with divergent health effects depending on its specific type, and a recent meta-analysis study has shown that cheese consumption, but not milk and yogurt, was inversely associated with CVD risk^{42,43}. Consistent with these studies, we did not observe a temporal relationship between total fatty foods derived from cheese, butter, and dairy products with the incidence of total CVDs. However, we found a genetically inverse causal association between liking for total fatty foods and total CVDs. Specifically, we found that the *rs429358* variant on *APOE* is associated with liking for total fatty foods, and individuals carrying this variant had lower levels of circulating *APOE*. Previous studies have found that circulating *APOE* levels are positively associated with the risk of CVDs⁴⁴.

Overall, based on the consistency of epidemiological research and MR analyses, we identified five pairwise associations: liking for broccoli is linked to a lower risk of total CVDs, while a liking for diet fizzy drinks is associated with a higher risk of HF. Similarly, a liking for bacon, lentils/beans, or pizza correlates with a lower risk of T2D, prompting us to investigate the biological mechanisms behind these associations. Our results indicate that individuals with these FLTs exhibit distinct dietary intake patterns. Participants who like diet fizzy drinks and bacon consume fewer plant-based foods and more animal-based foods. In contrast, those who like broccoli and lentils/beans consume more plant-based foods and less animal-based foods. The modest association between FLTs and corresponding food intake patterns led us to examine the differences in circulating cardiometabolic proteins influenced by FLTs and food intake to evaluate whether FLTs affect the pathological processes underlying CMD occurrence. Furthermore, we identified a total of 32 circulating cardiometabolic proteins simultaneously influenced by both these FLTs and corresponding food intakes, showing a dose-response relationship with increasing hedonic response to food liking. The consistency of circulating protein expression patterns also suggests that the association between FLTs and CMDs may be partially mediated by food intake. For example, we observed that a liking for bacon was associated with higher intakes of red and processed meats, probably contributing to lower levels of LPL. We also identified a significant genetic causal association between lower *LPL* expression and a higher risk of T2D, further suggesting that this dietary liking pattern is linked to an increased risk of T2D, as described in previous studies⁴⁵. Moreover, we observed that these FLTs were associated with distinct social factors and anthropometric indices, which may further explain the relationship between specific FLTs and corresponding CMDs. For instance, although we did not find overlapped circulating proteins between pizza liking and pizza intake, we observed that individuals who like pizza had lower levels of body fat and higher levels of fat-free mass in arms and legs. They also had higher incomes and lower Townsend Deprivation Index levels. These factors are all established risk factors of T2D^{46–48}, which may partially explain the negative association between pizza liking and T2D risk.

We also explored whether these paired associations share a genetic architecture to enhance the understanding of the impact of diet on CMDs. In our study, we identified three significant genetic



Fig. 5 | Genetic correlation analysis and identification of shared genomic risk loci and genes in different FLT and CMDs pairs. a Workflow for genetic correlation analysis and identification of pleiotropic genomic risk loci and genes for each FLT and CMD pair. **b** Genetic correlations between FLT and CMD pairs, as estimated using LDSC and HDL methods. A two-sided P -FDR < 0.05 was considered statistically significant. **c** Left panel: Manhattan plots of shared lead and INSIG SNVs in different FLT and CMD pairs, identified by PLACO. The height of each point represents the negative logarithm of the P -value estimated by PLACO, with the color bar indicating different chromosomes. The black dashed line represents the P -value threshold (a two-sided $P < 1 \times 10^{-5}$), and significant SNVs are labeled with text annotations. Right panel: Word cloud plots showing phenotype-wide

associations of these SNVs with clinical phenotypes. Sample Sizes: [diseases (case/ N)]: Total CVDs (221,781/231,952); IHD (75,592/378,141); MI (28,546/378,019); Stroke (67,162/454,450); HF (47,309/930,014); T2D (71,728/369,007). FLT's food-like traits, CMDs cardiometabolic diseases, LDSC linkage disequilibrium score regression, HDL high-definition likelihood, HF heart failure, T2D type 2 diabetes, PLACO the pleiotropic analysis under composite null hypothesis, INSIG independent significant, SNVs single-nucleotide variants, eQTL expression quantitative trait loci, MAGMA the gene-level multi-marker analysis of GenoMic annotation, SMR summary data-based Mendelian randomization, HEIDI heterogeneity in dependent instruments, rg genetic correlation, CVDs cardiovascular diseases, IHD ischemic heart disease, HF heart failure, MI myocardial infarction, T2D type 2 diabetes.

correlations: diet fizzy drinks with HF, bacon with T2D, and lentils/beans with T2D, leading to the discovery of 54 pleiotropic SNVs that concurrently regulate these FLT's and their corresponding disease outcomes. These pleiotropic SNVs were also associated with food intake and various cardiometabolic phenotypes, including body composition, inflammation, and oxidative stress, as highlighted in existing GWAS literature. This genetic evidence suggests that these pleiotropic SNVs collectively influence the development of CMDs through metabolic clinical characteristics, indicating that targeted regulation of these loci or related genes could affect CMD occurrence on four levels: food-like phenotypes, food intake phenotypes,

physiological metabolic phenotypes, and disease phenotypes. Therefore, identifying potential intervention targets that can simultaneously regulate FLT's and CMDs is crucial, as combining dietary behavior modification with drug treatment could serve as an effective strategy for preventing CMDs.

Therefore, we mapped these pleiotropic SNVs to genes and evaluated whether these pleiotropic genes have druggability. Among these identified genes, *ERRB2* has been widely used as a drug target for treating diabetes complications, metabolic disorders, and CVDs. Moreover, the combination of gene mapping, summary data-based MR, and MAGMA analysis consistently identified *ACO07382.1*, *ZNF131*,

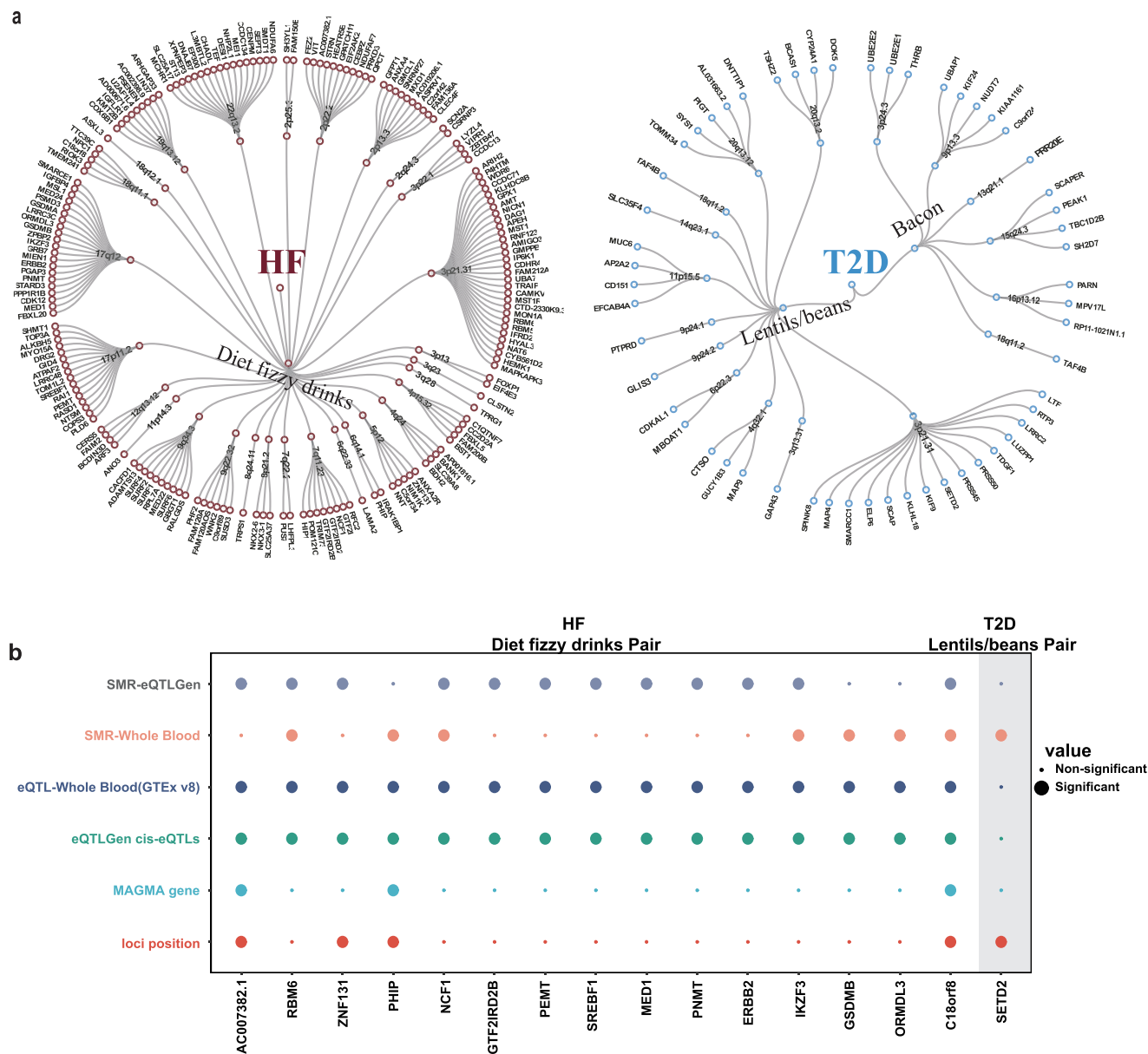


Fig. 6 | Gene mapping and druggability assessment. a Circular dendrograms showing the mapped genes of pleiotropic genomic risk SNVs in different FLT and CMD pairs. The inner points represent the disease; the first circle represents related food liking; the second circle shows the pleiotropic genomic SNV regions; the third circle displays the mapped genes of pleiotropic genomic risk SNVs in each FLT and CMD pair. **b** Pleiotropic genes in different FLT and CMD pairs, identified through loci positions, MAGMA gene analysis, eQTLGen cis-eQTLs, eQTL-Whole Blood (GTEx v8), SMR-Whole Blood, SMR-eQTLGen analyses. FLTs food-liking traits,

CMDs cardiometabolic diseases, SNVs single-nucleotide variants, MAGMA the gene-level multi-marker analysis of GenoMic annotation, eQTLGen cis-eQTLs the Cis-acting Expression Quantitative Trait Loci data generated by the eQTLGen Consortium, eQTL-whole blood (GTEx v8) expression quantitative trait loci for whole blood (from the Genotype-Tissue Expression Project version 8); SMR-eQTLGen, Summary data-based Mendelian Randomization with Expression Quantitative Trait Loci as reference data.

PHIP, and *C18orf8*, which show potential druggability in regulating individuals' liking for diet fizzy drinks and the risk of HF. While the exact function of *AC007382.1* remains unclear, its expression has been shown to be associated with several clinically relevant biomarkers, such as mean platelet volume, glycated hemoglobin, and C-reactive protein⁴⁹. Recent studies also show that *AC007382.1* expression is associated with brain structures, such as amygdala volume and brain connectivity^{50,51}, which are involved in stress responses and cognitive functions, potentially influencing food preferences and cardiovascular risk through neuroendocrine pathways. *ZNF131* (also known as *ZBTB35*) is a transcription factor that plays a role in immune cell coordination and central nervous system function. Its role in

regulating inflammatory responses and immune cell activity links it to conditions like CVD and diabetes^{52,53}. *PHIP* encodes a protein that regulates glucose transporter translocation in skeletal muscle and supports pancreatic beta-cell growth and survival processes. These processes are critical for maintaining glucose homeostasis and preventing metabolic disorders^{54,55}. *C18orf8* is involved in endosome morphology and LDL trafficking⁵⁶ and has recently been identified to be related to sudden cardiac death⁵⁷.

Our findings have potential clinical and public health significance. For example, despite the fact that diet fizzy drinks contain no calories and are often considered healthier than traditional sugary beverages, recent observational studies have linked higher consumption of diet

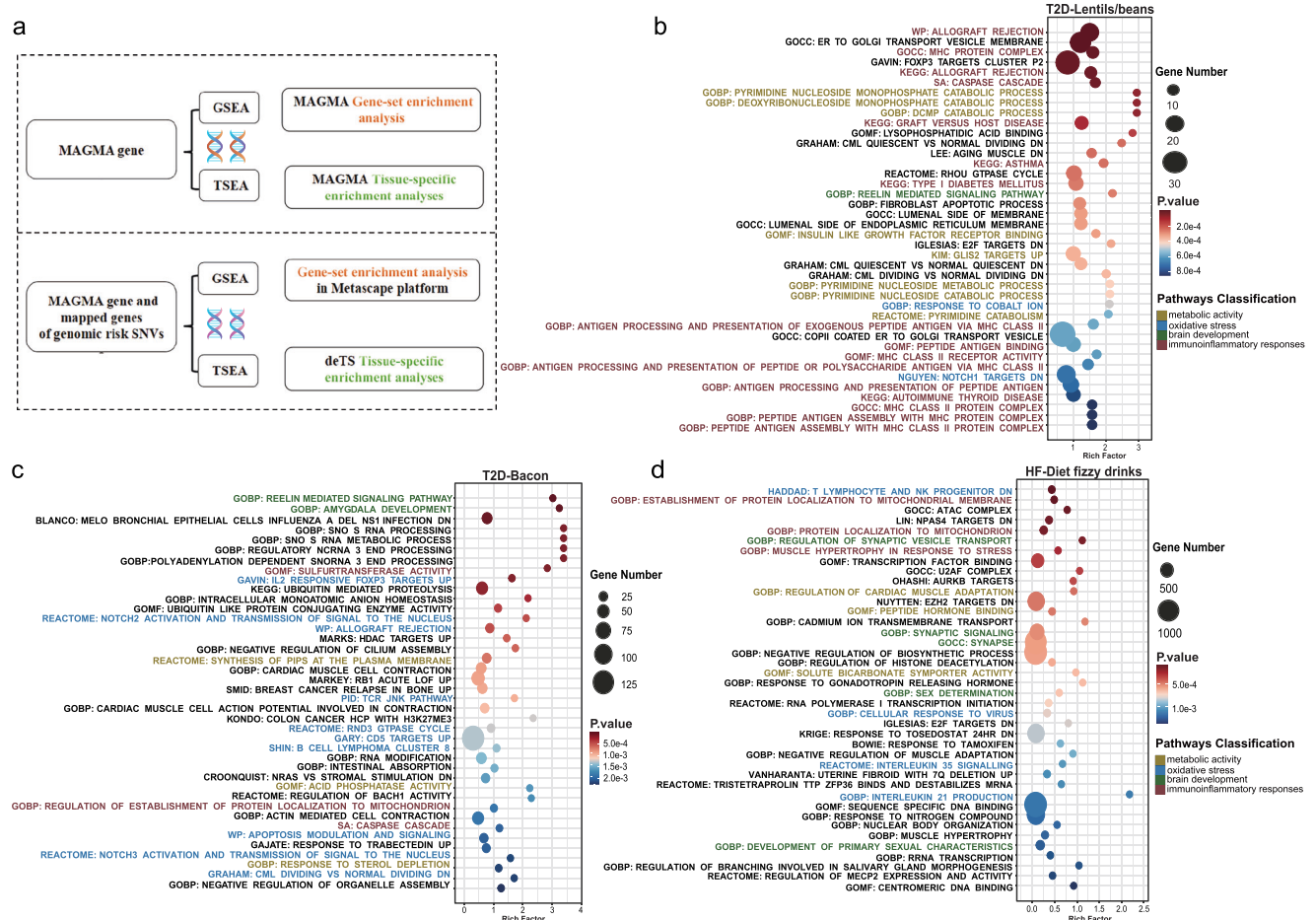


Fig. 7 | Shared biological pathways in different FLT and CMDs pairs. a Workflow for identifying shared biological pathways and tissue specificity in FLT and CMD pairs. **b–d** shared enriched biological pathways in different FLT and CMD pairs, identified by MAGMA gene-set analysis. The top pathways are presented for HF-Diet fizzy drinks, T2D-Bacon, and T2D-Lentils/beans. The red, green, blue, and brown pathways are respectively associated with immunoinflammatory responses, brain

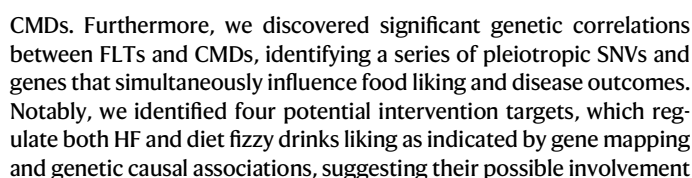
development, oxidative stress, and metabolic activity. Significantly enriched pathways, estimated by GSEA using Permutation Testing, were determined using a nominal threshold of two-sided $P < 0.05$. FLT food-like traits, CMDs cardiometabolic diseases, MAGMA the gene-level multi-marker analysis of GenoMic annotation, GSEA gene-set enrichment analysis, TSEA tissue-specific enrichment analysis.

fizzy drinks to an increased risk of heart disease or T2D^{58–60}. Consistent with these studies, we also observed that a liking for diet fizzy drinks was temporally and genetically associated with a higher risk of HF, sharing a genetic basis with HF. However, our findings also indicated that a liking for diet fizzy drinks was correlated with higher intake of sweets, dessert foods, and sugar beverages, which are well-established risk factors for CMDs^{61,62}. This suggests that diet fizzy drinks liking may act as a proxy for certain lifestyle factors that contribute to an elevated risk of CMDs. Our research also provides biological insights into the associations between diet fizzy drinks liking and HF, demonstrating that these traits share a genetic basis. Moreover, our genetic analysis identified 35 pleiotropic SNVs shared between diet fizzy drinks liking and HF, and many of which have been previously associated with key cardiometabolic traits, including BMI, body fat percentage, platelet count, and lung function. This suggests that diet fizzy drinks liking may be associated with a higher risk of HF through genetic mechanisms. Lastly, we find that certain genetic risk factors identified in our study could serve as potential targets for future therapeutic interventions aimed at addressing the physiological factors associated with diet fizzy drinks liking and the risk of HF.

Finally, to further investigate the common biological processes underlying the significant genetic correlations between these FLT and their corresponding diseases, we found that these genes are enriched in processes related to brain development, immune-inflammatory

responses, oxidative stress, and metabolic activities. This suggests that these biological processes play a critical role in shaping liking for diet fizzy drinks, bacon, and lentils/beans, as well as in the development of HF and T2D. Additionally, we observed that these pleiotropic genes exhibit tissue-specific expression patterns, primarily in the brain, lymphocytes, minor salivary glands, stomach, and pancreas. Previous research has shown that genes regulating FLT are mainly enriched in the brain^{10,63}, and the brain has been shown to be closely linked to the function and structure of peripheral organs, such as the heart and liver. These findings suggest that the brain plays a pivotal role in regulating both FLT and CMDs. Moreover, other identified tissues, such as lymphocytes, the stomach, and the pancreas, are also crucial in the development of CMDs. This implies that individuals with these FLT characteristics may be more susceptible to influences on the development of these organs and the occurrence of CMDs. Therefore, they should prioritize protecting these organs to better prevent or treat CMDs related to dietary intake.

Our research comprehensively evaluates the health impacts of FLT on CMDs by leveraging large-scale datasets and conducting MR analyses. Our findings reveal that a genetic predisposition toward certain FLT is associated with various types of CMDs, contributing valuable insights beyond current research. This emphasizes the potential for targeted dietary interventions based on specific food likings, which could enhance health outcomes in populations at risk for



13

Fig. 8 | Enriched biological pathways and tissue-specific enrichment analysis. **a–c** shared enriched biological pathways in different FLT and CMD pairs, identified by Metascape analysis. The top pathways are presented for HF-Diet fizzy drinks, T2D-Bacon, and T2D-Lentils/beans. The red, green, blue, and brown pathways are associated with immunoinflammatory responses, brain development, oxidative stress, and metabolic activity. Significantly enriched pathways were determined using a nominal threshold of two-sided $P < 0.05$. **d** deTS tissue-specific enrichment analyses (TSEA) of pleiotropic genes using two different reference panels: GTEx and

ENCODE were conducted (upper: MAGMA platform; lower left: deTS method with GTEx panel; lower right: deTS method with ENCODE panel). * Indicates a significant two-sided $P < 0.05$ in the analyses, estimated by Permutation Testing of TSEA. **e** Schematic diagram of the overall tissue-specific information in different FLT and CMD pairs. FLT's food-like traits, CMDs cardiometabolic diseases, MAGMA the gene-level multi-marker analysis of genoMic annotation, GTEx Genotype-Tissue Expression project, ENCODE Encyclopedia of DNA Elements project, T2D type 2 diabetes. Created in BioRender. <https://BioRender.com/viv2kjr>.

targeted regulation of pleiotropic genes or pathways holds potential for future therapeutic interventions to modulate these risk factors and slow CMD progression in these groups. These findings provide new insights for developing personalized nutrition and health strategies.

However, our study has certain limitations. First, while we establish temporal and genetic associations, the inferred causal relationships may not fully account for unmeasured variables, limiting definitive conclusions. Second, reliance on data from the UK Biobank and individuals of European ancestry may restrict the generalizability of our findings to other populations with different dietary behaviors and genetic backgrounds. Third, although we controlled for a range of covariates in the epidemiological and genetic analyses, food preferences are influenced by various environmental, cultural, and psychological factors that may not have been fully captured. Lastly, although the food-liking questionnaire demonstrates good statistical properties, its self-reported nature may introduce biases or inaccuracies, potentially affecting the robustness of our findings.

In conclusion, our study elucidates the temporal and genetic relationships between FLT's and the development of CMDs. We identified five significant associations between FLT's and CMDs, highlighting two harmful preferences (bacon and diet fizzy drinks) and three protective ones (broccoli, pizza, and lentils/beans). Our findings reveal that FLT's not only influence dietary habits but also correlate with circulating cardiometabolic proteins. Notably, we established genetic links between FLT's and CMDs, uncovering pleiotropic variants that affect both dietary intakes and disease outcomes. This research underscores the importance of specific food items in CMDs prevention and management, and suggests potential genetic targets for dietary interventions. Future studies should further explore the biological mechanisms underlying these associations to inform public health strategies.

Methods

Epidemiological study

Studying population. This study used the epidemiological data from the UKB under the projects 90232 and 203867, which were approved by the Northwest Multicenter Research Ethics Committee in the United Kingdom, the National Information Governance Board for Health and Social Care in England and Wales, and the Community Health Index Advisory Group in Scotland. All participants provided written informed consent. From March 2006 to October 2010, a total of 502,507 participants aged 40–69 years across the UK were enrolled^{10,63}. In this study, we included 182,087 participants who completed the food-liking questionnaire and provided valid responses. Our study adhered to the STROBE (Strengthening the Reporting of Observational Studies in Epidemiology) guidelines to ensure transparent and comprehensive reporting of observational studies.

Food-liking measurement. Food-liking data was collected via an online questionnaire (<https://biobank.ndph.ox.ac.uk/showcase/showcase/docs/foodpref.pdf>) from 182,176 participants. Liking was assessed using a 9-point hedonic scale, where 1 represents extremely dislike and 9 represents extremely like. This widely used scale has good statistical properties, discriminates well between different points, and exhibits linearity⁶⁴. In this study, we included a total of 176 FLT's,

consisting of 140 FLT's from the questionnaire and 36 FLT's constructed based on food similarity attributes (Fig. 1B, Supplementary Method 1 and Supplementary Data 25)¹⁰. We excluded 186 participants (0.1%) who responded never tried or do not wish to answer on more than 30% of the 176 food and beverage items, resulting in a final sample of 182,087 participants.

Food intake measurement. Dietary intake data was collected using the Oxford WebQ, a validated, web-based 24-h self-administered dietary assessment tool designed for large observational studies³⁶. Previous large-scale cohort studies have shown that this 24-h dietary assessment method provides a reliable estimate of usual dietary intake⁶⁵. Detailed information on food intake measurement can be found at <https://biobank.ndph.ox.ac.uk/showcase/label.cgi?id=100118>.

Measurement of circulating cardiometabolic related proteins. Plasma cardiometabolic protein levels were measured using proteomic profiling on blood plasma samples with the antibody-based Olink Explore 3072 PEA, which detects 2941 protein analytes. A total of 2923 unique proteins were captured, including 732 cardiometabolic-related proteins. Detailed information for Olink technology is presented in Supplementary Method 2.

Covariates and outcome assessment. Standard covariates of no interest is used in the epidemiological study, including age (years), self-reported sex (men/women), current smoking status (yes/no), current drinking status (yes/no), BMI (kg/m²), education level, Townsend deprivation index, Metabolic Equivalent of Task (MET), systolic blood pressure (SBP), diastolic blood pressure (DBP), high density lipoprotein cholesterol (HDL-C), low density lipoprotein cholesterol (LDL-C), total cholesterol (TC), and triglycerides (TG).

The main outcomes in the epidemiological study were CMDs, including total CVDs, IHD, MI, stroke, HF, and T2D. Information on the incidence and timing of these diseases was collected through the cumulative medical records of hospital diagnoses. The International Classification of Diseases, 10th edition (ICD-10) codes were used to categorize these conditions. Total CVDs were defined as ICD-10 codes I20–I25, I46, I60, I61, I63 and I64, including IHD (I20–I25), MI (I21–I23, I24.1 and I25.1), stroke (I60, I61, I63 and I64), and HF (I50). T2D were defined as ICD-10 codes E11.

MR analysis

Data source in MR

Exposures. The summary-level GWAS data for more than 176 specific food likings were obtained from a recent study encompassing 161,625 individuals of European ancestry in the UK Biobank study¹⁰. The GWAS data for food-liking phenotypes were measured using a 9-point hedonic scale. These FLT's were validated through genetic correlation with magnetic resonance imaging brain traits or corresponding food consumption traits, which showed substantial agreement in the direction of effects across 11 independent cohorts.

The genetic instruments for each exposure were selected based on genome-wide significance ($P < 5.00 \times 10^{-8}$) and were independent of each other linkage disequilibrium ($r^2 < 0.001$ within 10,000 kb). In the multi-variable MR analyses, the genetic instruments consisted of

SNVs that were genome-wide significant in the GWAS for each exposure and independent of each other.

Covariates. This study selected educational attainment, household income, and physical activity as covariates because these factors have been documented to be associated with FLT and CMDs^{7,15,16}. These three covariates were all obtained from the UK Biobank. Household income was estimated as self-reported average total household income in 2006–2010 before tax reduction ($N=311,028$)⁶⁶. Educational attainment was measured by the age at which participants finished full-time education ($N=283,749$)⁶⁷. Physical activity was measured using average acceleration from a wrist-worn accelerometer (Axivity AX3) ($N=91,084$)⁶⁸.

Outcomes. The summary-level GWAS data for T2D ($N=440,735$), IHD ($N=453,733$), and MI ($N=406,565$) were sourced from FinnGen, released in 2024²². Previous studies have shown that while the Finnish population has some unique characteristics, it shares the greatest genetic similarity with other European populations^{22,69}. The summary-level GWAS data for HF was sourced from a recent meta-analysis of GWAS data, including 977,323 individuals of European ancestry across 26 studies conducted by the Heart Failure Molecular Epidemiology for Therapeutic Targets Consortium²⁴. Data for stroke was sourced from a meta-analysis of GWAS data, including 446,696 individuals of European ancestry, conducted by the MEGASTROKE Consortium²³. Detailed information on the GWAS data for each outcome is shown in Supplementary Table 2.

Our study followed the STROBE (Strengthening the Reporting of Observational Studies in Epidemiology by MR) guidelines to ensure transparent and comprehensive reporting of observational studies.

Statistical analysis

All epidemiological and MR analyses were conducted in R software (version 4.3.1) using the R packages *Survival* (version 3.5.5), *rms* (version 6.7.0), and *limma* (version 3.6.4). MR analyses were conducted using the R packages *TwoSampleMR* (version 0.5.7), *MVMR* (version 0.4), *MRPRESSO* (version 1.0), and *MRLap* (version 0.0.3.0). All genetic architecture analyses were performed after excluding SNVs in the major histocompatibility complex region (chromosome 6: 25–35 megabase [Mb]) due to its complex linkage disequilibrium structure. Analyses were restricted to biallelic SNVs with a minor allele frequency larger than 0.01.

Descriptive analysis

The baseline characteristics, including demographic characteristics and anthropometric measures, were presented as mean (standard deviation, SD) for continuous variables and Numbers (proportion, %) for categorical variables, which were respectively compared by *t*-test or χ^2 test according to the baseline status of cardiometabolic diseases. A 2-sided *P*-value < 0.05 was considered statistically significant.

Cox proportional hazards regression models

The Cox proportional hazards models were divided into two sections. Section 1: the models were used to evaluate the survival relationship between FLT and the incidence of CMDs after excluding participants who had CMDs at the baseline survey. Survival time was defined as the period from participation in the UKB assessment center to the diagnosis of the main outcomes or up to February 1, 2024. A series of covariates were controlled for, including age, sex, smoking status, drinking status, BMI, education level, income, Townsend deprivation index, MET, SBP, DBP, HDL-C, LDL-C, TC, and TG. Section 2: the models were performed to examine whether circulating cardiometabolic proteins, influenced by specific FLT, were associated with the risk of CMDs after adjustment for covariates in section 1, and a two-sided *P*-FDR < 0.05 was

considered statistically significant. Moreover, we conducted a sensitivity analysis to evaluate the potential for reverse causation between FLT and CMDs. Specifically, participants classified as high-risk for CMDs, based on QRISK3 risk prediction models⁷⁰, were excluded. The second sensitivity analysis excluded participants whose self-reported sex did not align with their genetically determined sex, to evaluate whether disaggregating by sex and gender would influence our results. The third sensitivity analysis examined whether sex has modifying effects on the associations between FLT and CMDs.

Identification of differentially expressed protein

We used the *limma* Bioconductor package to identify differentially expressed proteins by comparing groups based on their liking for a particular food: those who extremely like it (scored 9 on the hedonic scale) versus those who extremely dislike it (scored 1 on the hedonic scale). The statistical significance threshold was set at FDR-adjusted two-sided *P* < 0.05.

Restricted cubic spline

Restricted cubic spline analysis was used to visualize the survival dose-response relationship, the identified differential circulating proteins and CMDs by setting 5 knots at the 5th, 25th, 50th, 75th, and 95th percentiles. Analysis of variance was used to examine the linear or non-linear relationship of the spline. The covariates controlled in the restricted cubic spline analysis were similar to those in the Cox proportional hazards regression models.

Univariate MR and multi-variable MR analysis

We applied univariate MR to assess the genetic association of each FLT with each type of CMDs. Significant pairs for FLT and CMD were selected. Then, we performed multi-variable MR to evaluate these identified pairs with adjustment for household income, educational attainment, and physical activity to determine whether FLT had independent genetic association on CMDs.

MR analysis employs genetic variants as instrumental variables to infer causal relationships between exposures and outcomes. Genetic variants, being randomly allocated at conception, are independent of self-selected behaviors and predate disease onset, reducing confounding and reverse causality. The genetic instrument effect is estimated through genotype randomization during meiosis, similar to blinding in trials, ensuring robust association with the exposure, independence from confounders, and no alternative pathways to the disease outcome.

Instrument variables used for MR need to satisfy 3 assumptions⁷¹. First, instrument variables must be robustly associated with CMDs. Second, instrumental variables should not be related to confounding factors. Third, instrumental variables should affect the outcome only through the exposure factor. The 176 FLT were categorized into 10 major food groups, and we performed univariate MR analysis separately for each group. We calculated an FDR-adjusted two-sided *P*-value to test statistical significance. Then, statistically significant FLT and disease pairs were then tested using multi-variable MR analysis, with a *P*-value < 0.05 considered statistically significant. The IVWs were considered genetic associations only if they had the same direction and statistical significance as at least one sensitivity analysis. MR estimates were presented as ORs with corresponding 95% CIs.

We used the random-effect IVW method as the main analysis in univariable MR⁷², and performed the multi-variable IVW as the main analysis in multi-variable MR⁷³. The IVW method produces the most precise, unbiased, and efficient causal estimates if instrument variables satisfy MR assumptions⁷⁴. The random-effects IVW method accounts for heterogeneity among instrument variables, while the multi-variable IVW accounts for correlations between multiple exposures.

MR sensitivity analysis

In univariable MR, we performed the Bayesian weighted, weighted median, weighted mode, and MR-Egger methods to validate the robustness of the IVW results based on different assumptions (Supplementary Method 3)^{75–79}. In multi-variable MR, we performed the multi-variable MR-median, multi-variable MR-Egger, and multi-variable MR-Lasso methods to validate the robustness of the multi-variable IVW results (Supplementary Method 4)⁸⁰. The F-statistics, Egger intercept, Cochran's Q statistics, and Steiger test were applied to assess the validity and heterogeneity of the instrument variables as well as reverse causation, respectively^{81–83}. Additionally, we performed bidirectional univariate MR analysis for the FLT and CMD pairs identified in the multi-variable MR analysis. In this analysis, CMDs were treated as exposures and FLTs as outcomes to assess whether genetically predicted CMDs could influence FLTs for detecting reverse causation.

Analysis for genome-wide genetic correlations

The LDSC and high-definition likelihood method were used to estimate genome-wide genetic correlations between FLTs and CMDs^{84,85}. For the LDSC analysis, we used European 1000 Genomes Phase 3 data as the reference and set a filtering threshold for minor allele frequencies above 0.01 to enhance statistical power and minimize the influence of rare variants. The standard error was estimated using the leave-one-out method in LDSC. Population overlap was also estimated to provide unbiased genetic correlation estimates for improving the reliability of the results. Compared with the LDSC method, the high-definition likelihood method makes more efficient use of GWAS data, reducing the variance in genetic association estimates by about 60%, which further validates the reliability of our findings.

Identification of the genomic risk SNVs and related mapped genes

We used the PLACO method to identify shared pleiotropic SNVs between trait pairs with significant genetic correlations⁸⁶. The PLACO summary statistics were processed using the FUMA SNP2GENE platform (v1.5.0) (<https://fuma.ctglab.nl/snp2gene>) to identify lead and insignificant SNVs, genomic risk SNVs, gene mapping, and exploration of biological pathways, including MAGMA gene analysis, MAGMA gene-set analysis, and TSEA. We set European 1000 Genomes Phase 3 data as the reference, and a $P < 1.00 \times 10^{-5}$ as the threshold value. Lead SNVs were defined as those with linkage disequilibrium $r^2 \leq 0.1$ with a distance of <250 kb. Independent significant SNVs were defined as those with linkage disequilibrium $r^2 \leq 0.6$ with a distance of <250 kb⁸⁷. Based on the linkage disequilibrium information of lead SNVs, genomic risk loci were further characterized and annotated. Detailed information is presented in Supplementary Data 13 and 14.

To confirm the biological relevance of the genomic risk SNVs, we performed functional gene mapping using two methods: (1) positional mapping, assigning SNVs to genes based on physical distances (default window of 10 kb), and (2) eQTL mapping, assigning SNVs to genes based on gene expression levels using cis-eQTLGen and GTEx Whole Blood v8 data. The P -value threshold of significant SNV-gene pairs was set at $P\text{-FDR} < 0.05$. The MHC region was excluded from both annotations due to its complicated linkage disequilibrium structure, defined as between the MOG and COL11A2 genes.

Phenome-wide association with the lead SNVs

We investigated the lead SNVs identified in each pair of FLT and corresponding disease to determine whether these loci had been previously associated with any clinical phenotype. Three major platforms: the EMBL-EBI GWAS Catalog (EMBL-EBI 2024) (<https://www.ebi.ac.uk/gwas/home>)⁸⁸, IEU Open GWAS (v8.9.1) (<https://gwas.mrcieu.ac.uk/>)⁸⁹, and GWAS Atlas (v2.0) (<https://atlas.ctglab.nl/PheWAS>)⁹⁰ were utilized.

In the EMBL-EBI GWAS Catalog, we examined whether each lead SNV was linked to any clinical phenotype (e.g., rs11052302, <https://www.ebi.ac.uk/gwas/search?query=rs11052302>, query date: September 20, 2024).

For IEU Open GWAS and GWAS Atlas, we conducted comprehensive searches for lead SNVs and directly extracted significantly associated clinical phenotypes (query date: September 20, 2024). We applied a uniform P threshold of $<1 \times 10^{-5}$ as the search criterion across all platforms¹⁴. To manage linkage disequilibrium and remove redundant associations among independent significant SNVs, we followed the procedure: If a lead SNV had clinical associations, it was considered the primary locus. If the lead SNV did not show any clinical phenotype associations, we examined independent significant SNVs that were closely correlated with the lead SNV, starting with the most significant ones, until an established association was identified¹⁴.

For visualization purposes, we categorized clinical phenotypes into 548 items, which were primarily related to CMDs and associated clinical traits, brain development and neurological disorders, inflammation indices, as well as diet and lifestyle behaviors. Keyword cloud plots were created based on the frequency of these clinical phenotypes within each FLT and CMD pair. The size of each rectangle in the keyword cloud plots reflects the total number of associations between the lead SNVs identified in our study and the clinical phenotypes reported in the literature.

MAGMA gene annotation

We performed genome-wide gene-level association analysis by the MAGMA gene analysis method. Genes located in SNVs were annotated and mapped using reference European 1000 Genomes Phase 3 data as the reference, employing the SNV-wide mean model. Gene-level P -values were obtained for traits inputted against curated protein-coding genes containing valid SNVs. The significance of pleiotropic genes was declared at $P\text{-FDR} < 0.05$ ⁹¹.

Summary data-based MR analysis using reference pQTL or eQTL

The summary data-based MR analysis consisted of two sections. Section 1: a summary data-based MR analysis was conducted to investigate the potential causal relationship between plasma circulating protein levels associated with FLTs and corresponding types of CMDs. For this, we used reference pQTL data from the UKB PPP project⁹². Section 2: we conducted SMR analysis to examine whether the SNVs associated with both traits in each FLT-CMD pair were driven by the same gene expression with reference to eQTL data from whole blood gene expression and eQTLGen datasets. The summary data-based MR method integrates an MR framework to infer causality by testing associations between GWAS and eQTL or pQTL data⁹³. The HEIDI test was applied to distinguish causality (or pleiotropy) from linkage, where possible, based on the available data⁹⁴. This approach deepened our understanding of the underlying genetic mechanisms and provided valuable insights for drug target discovery for each trait pair. We selected significant probes common to both phenotypes from the HEIDI analysis, using the threshold of summary data-based MR estimated P -value < 0.05 and HEIDI test P -value > 0.05 .

Druggability assessment

For the genes we identified, we first searched for corresponding drugs and indications in Drugcentral (<https://drugcentral.org/>), DGIdb (<https://dgidb.org/>), DrugBank (<https://go.drugbank.com/>), and Therapeutic Target Database (<https://db.idrblab.net/ttd/>). We then supplemented this information with clinical trial data from ClinicalTrials.gov (<https://clinicaltrials.gov/>) and verified whether the identified drugs have been approved by the FDA through the official FDA website (<https://www.accessdata.fda.gov/scripts/cder/daf/index.cfm>) to assess their potential druggability.

GSEA

GSEA was performed to identify potential shared biological pathways using two approaches. First, we utilized pleiotropic genes identified

from the MAGMA gene analysis to explore the biological functions of shared variants⁹⁵. The MAGMA gene sets were obtained from MsigDB v6.2, including 10,678 gene sets (curated gene sets: 4761, GO terms: 5917). Second, we employed Metascape Web Tools (v3.5) (<https://metascape.org/gp/index.html>) to analyze gene sets derived from both the MAGMA gene analysis and mapped genes of the genomic risk SNVs. Significantly enriched pathways were determined using a nominal threshold of two-sided $P < 0.05$.

TSEA

We used the MAGMA platform to perform gene property analyses for each FLT and CMD pair, using reference GTEx V8 data, which encompassed 54 tissue types. This gene property analysis converted gene-level association P -values to z scores and tested the expression value of a specific tissue's gene against the average expression value across all tissues in a regression model. To further examine the gene properties of these pleiotropic loci, we also conducted TSEA to deTS analysis (<https://github.com/bsml320/deTS>) using two different reference panels⁹⁵: GTEx and Encyclopedia of DNA Elements project (ENCODE). This methodology highlighted the tissue specificity of pleiotropic genes derived from both the MAGMA gene analysis and the mapped genes of the genomic risk SNVs. Significantly enriched pathways were determined using a nominal threshold of two-sided $P < 0.05$.

Ethics statement

The UK biobank was approved by the Northwest Multicenter Research Ethics Committee in the United Kingdom, the National Information Governance Board for Health and Social Care in England and Wales, and the Community Health Index Advisory Group in Scotland. All the summary-level GWAS data used in the analyses are publicly available, and therefore, this study was exempt from ethics review. Ethical approval for the GWASs can be found in the corresponding GWAS publications cited in the manuscript.

Reporting summary

Further information on research design is available in the Nature Portfolio Reporting Summary linked to this article.

Data availability

All GWAS summary statistics analyzed in this study are publicly available as shown in Supplementary Table 2. The UKB data can be obtained from <https://ams.ukbiobank.ac.uk/ams/resApplications>. The GWAS data for FLTs can be obtained from the GWAS catalog (<https://www.ebi.ac.uk/gwas/publications/35585065>). The GWAS data for household income can be obtained from <https://gwas.mrcieu.ac.uk/datasets/ukb-b-7408/>, and data for education can be obtained from <https://gwas.mrcieu.ac.uk/datasets/ukb-b-6134/>, and data for physical activity can be obtained from <https://www.ebi.ac.uk/gwas/studies/GCST006099>. The GWAS data for CVDs, IHD, MI, and T2D traits can be obtained from the FinnGen website (<https://r11.finnngen.fi/>). The GWAS data for stroke and HF traits can be obtained from the GWAS catalog (<https://www.ebi.ac.uk/gwas/studies/GCST005838> and <https://www.ebi.ac.uk/gwas/studies/GCST009541>). MAGMA gene-set analysis reference was obtained from the Molecular Signatures Database (MsigDB v6.2) (<https://www.gsea-msigdb.org/gsea/>). For TSEA, MAGMA used data from GTEx V8 (<http://gtexportal.org/home/>), while deTS utilized data from both GTEx and ENCODE. Summary data-based based MR analysis used the reference whole blood gene expression and eQTLGen datasets (<https://molgenis26.gcc.rug.nl/downloads/eqtlgen/>). All the data generated during the study are available in <https://github.com/jiangwenbo0721/Nature-Communications> and Code Ocean: <https://doi.org/10.24433/CO.8587531.v1>.

Code availability

Custom code in our study is available at <https://github.com/jiangwenbo0721/Nature-Communications> and Code Ocean: <https://doi.org/10.24433/CO.8587531.v1>.

References

- Berridge, K. C., Robinson, T. E. & Aldridge, J. W. Dissecting components of reward: 'liking', 'wanting', and learning. *Curr. Opin. Pharmacol.* **9**, 65–73 (2009).
- Eertmans, A., Baeyens, F. & Van den Bergh, O. Food likes and their relative importance in human eating behavior: review and preliminary suggestions for health promotion. *Health Educ. Res.* **16**, 443–456 (2001).
- Livingstone, K. M., Phosamy, H., Riddell, L. J. & Cicerale, S. Demographic, behavioural and anthropometric correlates of food liking: a cross-sectional analysis of young adults. *Nutrients* **12**, 3078 (2020).
- Breen, F. M., Plomin, R. & Wardle, J. Heritability of food preferences in young children. *Physiol. Behav.* **88**, 443–447 (2006).
- Fildes, A. et al. Nature and nurture in children's food preferences. *Am. J. Clin. Nutr.* **99**, 911–917 (2014).
- Morales, I. & Berridge, K. C. Liking' and 'wanting' in eating and food reward: Brain mechanisms and clinical implications. *Physiol. Behav.* **227**, 113152 (2020).
- Ventura, A. K. & Worobey, J. Early influences on the development of food preferences. *Curr. Biol.* **23**, R401–R408 (2013).
- Mozaffarian, D. et al. "Food is medicine" strategies for nutrition security and cardiometabolic health equity: JACC state-of-the-art review. *J. Am. Coll. Cardiol.* **83**, 843–864 (2024).
- Food as medicine: translating the evidence. *Nat. Med.* **29**, 753–754 (2023).
- May-Wilson, S. et al. Large-scale GWAS of food liking reveals genetic determinants and genetic correlations with distinct neurophysiological traits. *Nat. Commun.* **13**, 2743 (2022).
- Zdravkovic, S. et al. Heritability of death from coronary heart disease: a 36-year follow-up of 20 966 Swedish twins. *J. Intern. Med.* **252**, 247–254 (2002).
- Vujkovic, M. et al. Discovery of 318 new risk loci for type 2 diabetes and related vascular outcomes among 1.4 million participants in a multi-ancestry meta-analysis. *Nat. Genet.* **52**, 680–691 (2020).
- Smith, A. D. et al. Genetic and environmental influences on food preferences in adolescence. *Am. J. Clin. Nutr.* **104**, 446–453 (2016).
- Wen, J. et al. The genetic architecture of biological age in nine human organ systems. *Nat. Aging* **4**, 1290–1307 (2024).
- Zhao, B. et al. Heart-brain connections: phenotypic and genetic insights from magnetic resonance images. *Science* **380**, abn6598 (2023).
- Zhuang, Z., Yao, M., Wong, J., Liu, Z. & Huang, T. Shared genetic etiology and causality between body fat percentage and cardiovascular diseases: a large-scale genome-wide cross-trait analysis. *BMC Med.* **19**, 100 (2021).
- Burgess, S. et al. Dose-response relationship between genetically proxied average blood glucose levels and incident coronary heart disease in individuals without diabetes mellitus. *Diabetologia* **64**, 845–849 (2021).
- Cadby, G. et al. Heritability of 596 lipid species and genetic correlation with cardiovascular traits in the Busselton Family Heart Study. *J. Lipid Res.* **61**, 537–545 (2020).
- Pirastu, N. et al. Genetics of food preferences: a first view from silk road populations. *J. Food Sci.* **77**, S413–S418 (2012).
- Hamdard, E. et al. Denatonium benzoate induces oxidative stress in the heart and kidney of Chinese fast yellow chickens by regulating apoptosis, autophagy, antioxidative activities and bitter taste receptor gene expressions. *Animals* **9**, 701 (2019).
- Yuan, G., Jing, Y., Wang, T., Fernandes, V. S. & Xin, W. The bitter taste receptor agonist-induced negative chronotropic effects on

- the Langendorff-perfused isolated rat hearts. *Eur. J. Pharmacol.* **876**, 173063 (2020).
22. Kurki, M. I. et al. FinnGen provides genetic insights from a well-phenotyped isolated population. *Nature* **613**, 508–518 (2023).
 23. Malik, R. et al. Multiancestry genome-wide association study of 520,000 subjects identifies 32 loci associated with stroke and stroke subtypes. *Nat. Genet.* **50**, 524–537 (2018).
 24. Shah, S. et al. Genome-wide association and Mendelian randomisation analysis provide insights into the pathogenesis of heart failure. *Nat. Commun.* **11**, 163 (2020).
 25. Khaing, W., Vallibhakara, S. A., Attia, J., McEvoy, M. & Thakkinian, A. Effects of education and income on cardiovascular outcomes: a systematic review and meta-analysis. *Eur. J. Prev. Cardiol.* **24**, 1032–1042 (2017).
 26. Blodgett, J. M. et al. Device-measured physical activity and cardiometabolic health: the Prospective Physical Activity, Sitting, and Sleep (ProPASS) consortium. *Eur. Heart J.* **45**, 458–471 (2024).
 27. Fernández-Ruiz, I. Immune system and cardiovascular disease. *Nat. Rev. Cardiol.* **13**, 503 (2016).
 28. Lopaschuk, G. D., Karwi, Q. G., Tian, R., Wende, A. R. & Abel, E. D. Cardiac energy metabolism in heart failure. *Circ. Res.* **128**, 1487–1513 (2021).
 29. Schiff, H. C. et al. Experience-dependent plasticity of gustatory insular cortex circuits and taste preferences. *Sci. Adv.* **9**, eade6561 (2023).
 30. Berthoud, H. R., Morrison, C. D., Ackroff, K. & Sclafani, A. Learning of food preferences: mechanisms and implications for obesity & metabolic diseases. *Int. J. Obes.* **45**, 2156–2168 (2021).
 31. Li, C. et al. Meat consumption and incident type 2 diabetes: an individual-participant federated meta-analysis of 1·97 million adults with 100 000 incident cases from 31 cohorts in 20 countries. *Lancet Diabetes Endocrinol.* **12**, 619–630 (2024).
 32. Yang, J., Mao, Q. X., Xu, H. X., Ma, X. & Zeng, C. Y. Tea consumption and risk of type 2 diabetes mellitus: a systematic review and meta-analysis update. *BMJ Open* **4**, e005632 (2014).
 33. Carter, P., Gray, L. J., Troughton, J., Khunti, K. & Davies, M. J. Fruit and vegetable intake and incidence of type 2 diabetes mellitus: systematic review and meta-analysis. *BMJ* **341**, c4229 (2010).
 34. Keskitalo, K. et al. Same genetic components underlie different measures of sweet taste preference. *Am. J. Clin. Nutr.* **86**, 1663–1669 (2007).
 35. Loftfield, E. et al. Association of coffee drinking with mortality by genetic variation in caffeine metabolism: findings from the UK Biobank. *JAMA Intern. Med.* **178**, 1086–1097 (2018).
 36. Ma, H., Wang, X., Li, X., Heianza, Y. & Qi, L. Adding salt to foods and risk of cardiovascular disease. *J. Am. Coll. Cardiol.* **80**, 2157–2167 (2022).
 37. Torres-Pinzon, D. L., Ralph, D. L., Veiras, L. C. & McDonough, A. A. Sex-specific adaptations to high-salt diet preserve electrolyte homeostasis with distinct sodium transporter profiles. *Am. J. Physiol. Cell Physiol.* **321**, C897–C909 (2021).
 38. Geerling, J. C. et al. FoxP2 expression defines dorsolateral pontine neurons activated by sodium deprivation. *Brain Res.* **1375**, 19–27 (2011).
 39. Shin, J. W., Geerling, J. C., Stein, M. K., Miller, R. L. & Loewy, A. D. FoxP2 brainstem neurons project to sodium appetite regulatory sites. *J. Chem. Neuroanat.* **42**, 1–23 (2011).
 40. Rees, J., Wood, A. M. & Burgess, S. Extending the MR-Egger method for multivariable Mendelian randomization to correct for both measured and unmeasured pleiotropy. *Stat. Med.* **36**, 4705–4718 (2017).
 41. Pallister, T. et al. Food preference patterns in a UK Twin Cohort. *Twin Res. Hum. Genet.* **18**, 793–805 (2015).
 42. Zhuang, P. et al. A global analysis of dairy consumption and incident cardiovascular disease. *Nat. Commun.* **16**, 437 (2025).
 43. Van Parys, A. et al. The association between dairy intake and risk of cardiovascular disease and mortality in patients with stable angina pectoris. *Eur. J. Prev. Cardiol.* **30**, 219–229 (2023).
 44. Rasmussen, K. L., Tybjaerg-Hansen, A., Nordestgaard, B. G. & Frikke-Schmidt, R. Plasma levels of apolipoprotein E, APOE genotype, and all-cause and cause-specific mortality in 105 949 individuals from a white general population cohort. *Eur. Heart J.* **40**, 2813–2824 (2019).
 45. Krauss, R. M. Lipids and lipoproteins in patients with type 2 diabetes. *Diabetes Care* **27**, 1496–1504 (2004).
 46. Gupta, P. et al. The association between body composition using dual energy X-ray absorptiometry and type-2 diabetes: a systematic review and meta-analysis of observational studies. *Sci. Rep.* **9**, 12634 (2019).
 47. Agardh, E., Allebeck, P., Hallqvist, J., Moradi, T. & Sidorchuk, A. Type 2 diabetes incidence and socio-economic position: a systematic review and meta-analysis. *Int. J. Epidemiol.* **40**, 804–818 (2011).
 48. Nagar, S. D., Nápoles, A. M., Jordan, I. K. & Mariño-Ramírez, L. Socioeconomic deprivation and genetic ancestry interact to modify type 2 diabetes ethnic disparities in the United Kingdom. *EclinicalMedicine* **37**, 100960 (2021).
 49. Legault, M. A., Perreault, L. L., Tardif, J. C. & Dubé, M. P. ExPheWas: a platform for cis-Mendelian randomization and gene-based association scans. *Nucleic Acids Res.* **50**, W305–W311 (2022).
 50. Tissink, E. et al. The genetic architectures of functional and structural connectivity properties within cerebral resting-state networks. *eNeuro* **10**, ENEURO.0242-0222.2023 (2023).
 51. Mufford, M. S. et al. The genetic architecture of amygdala nuclei. *Biol. Psychiatry* **95**, 72–84 (2024).
 52. Cheng, Z. Y., He, T. T., Gao, X. M., Zhao, Y. & Wang, J. ZBTB transcription factors: key regulators of the development, differentiation and effector function of T cells. *Front. Immunol.* **12**, 713294 (2021).
 53. Trappe, R. et al. The murine BTB/POZ zinc finger gene Znf131: predominant expression in the developing central nervous system, in adult brain, testis, and thymus. *Biochem. Biophys. Res. Commun.* **296**, 319–327 (2002).
 54. Farhang-Fallah, J. et al. The pleckstrin homology (PH) domain-interacting protein couples the insulin receptor substrate 1 PH domain to insulin signaling pathways leading to mitogenesis and GLUT4 translocation. *Mol. Cell Biol.* **22**, 7325–7336 (2002).
 55. Podcheko, A. et al. Identification of a WD40 repeat-containing isoform of PHIP as a novel regulator of beta-cell growth and survival. *Mol. Cell Biol.* **27**, 6484–6496 (2007).
 56. van den Boomen, D. et al. A trimeric Rab7 GEF controls NPC1-dependent lysosomal cholesterol export. *Nat. Commun.* **11**, 5559 (2020).
 57. Zhao, W. et al. Novel indel variation of NPC1 gene associates with risk of sudden cardiac death. *Front. Genet.* **13**, 869859 (2022).
 58. Vyas, A. et al. Diet drink consumption and the risk of cardiovascular events: a report from the Women’s Health Initiative. *J. Gen. Intern. Med.* **30**, 462–468 (2015).
 59. Torres-Ibarra, L. et al. Regular consumption of soft drinks is associated with type 2 diabetes incidence in Mexican adults: findings from a prospective cohort study. *Nutr. J.* **19**, 126 (2020).
 60. Nettleton, J. A. et al. Diet soda intake and risk of incident metabolic syndrome and type 2 diabetes in the Multi-Ethnic Study of Atherosclerosis (MESA). *Diabetes Care* **32**, 688–694 (2009).
 61. Yang, Q. et al. Added sugar intake and cardiovascular diseases mortality among US adults. *JAMA Intern. Med.* **174**, 516–524 (2014).
 62. Lara-Castor, L. et al. Burdens of type 2 diabetes and cardiovascular disease attributable to sugar-sweetened beverages in 184 countries. *Nat. Med.* <https://doi.org/10.1038/s41591-024-03345-4> (2025).
 63. Zhang, R. et al. Associations of dietary patterns with brain health from behavioral, neuroimaging, biochemical and genetic analyses. *Nat. Ment. Health* **2**, 535–552 (2024).

64. Peryam, D. R. & Pilgrim, F. J. Hedonic scale method of measuring food preferences. *Food Technol.* 10.1007/BF02638783 (1957).
65. Freese, J. et al. Development and evaluation of a short 24-h food list as part of a blended dietary assessment strategy in large-scale cohort studies. *Eur. J. Clin. Nutr.* **68**, 324–329 (2014).
66. Hemani, G. et al. The MR-Base platform supports systematic causal inference across the human phenotype. *eLife* **7**, e34408 (2018).
67. Lee, J. J. et al. Gene discovery and polygenic prediction from a genome-wide association study of educational attainment in 1.1 million individuals. *Nat. Genet.* **50**, 1112–1121 (2018).
68. Schoeler, T. et al. Participation bias in the UK Biobank distorts genetic associations and downstream analyses. *Nat. Hum. Behav.* **7**, 1216–1227 (2023).
69. 1000 Genomes Project Consortium A global reference for human genetic variation. *Nature* **526**, 68–74 (2015).
70. Hippisley-Cox, J., Coupland, C. & Brindle, P. Development and validation of QRISK3 risk prediction algorithms to estimate future risk of cardiovascular disease: prospective cohort study. *BMJ* **357**, j2099 (2017).
71. Emdin, C. A., Khera, A. V. & Kathiresan, S. Mendelian randomization. *JAMA* **318**, 1925–1926 (2017).
72. Bowden, J. et al. A framework for the investigation of pleiotropy in two-sample summary data Mendelian randomization. *Stat. Med.* **36**, 1783–1802 (2017).
73. Grant, A. J. & Burgess, S. Pleiotropy robust methods for multivariable Mendelian randomization. *Stat. Med.* **40**, 5813–5830 (2021).
74. Burgess, S. et al. Guidelines for performing Mendelian randomization investigations: update for summer 2023. *Wellcome Open Res.* **4**, 186 (2019).
75. Burgess, S. et al. Using published data in Mendelian randomization: a blueprint for efficient identification of causal risk factors. *Eur. J. Epidemiol.* **30**, 543–552 (2015).
76. Zhao, J. et al. Bayesian weighted Mendelian randomization for causal inference based on summary statistics. *Bioinformatics* **36**, 1501–1508 (2020).
77. Bowden, J., Davey Smith, G., Haycock, P. C. & Burgess, S. Consistent Estimation In Mendelian Randomization With Some Invalid Instruments Using A Weighted Median Estimator. *Genet. Epidemiol.* **40**, 304–314 (2016).
78. Hartwig, F. P., Davey Smith, G. & Bowden, J. Robust inference in summary data Mendelian randomization via the zero modal pleiotropy assumption. *Int. J. Epidemiol.* **46**, 1985–1998 (2017).
79. Bowden, J., Davey Smith, G. & Burgess, S. Mendelian randomization with invalid instruments: effect estimation and bias detection through Egger regression. *Int. J. Epidemiol.* **44**, 512–525 (2015).
80. Burgess, S. & Thompson, S. G. Multivariable Mendelian randomization: the use of pleiotropic genetic variants to estimate causal effects. *Am. J. Epidemiol.* **181**, 251–260 (2015).
81. Burgess, S. & Thompson, S. G. CRP CHD Genetics Collaboration. Avoiding bias from weak instruments in Mendelian randomization studies. *Int. J. Epidemiol.* **40**, 755–764 (2011).
82. Burgess, S., Butterworth, A. & Thompson, S. G. Mendelian randomization analysis with multiple genetic variants using summarized data. *Genet. Epidemiol.* **37**, 658–665 (2013).
83. Hemani, G., Tilling, K. & Davey Smith, G. Orienting the causal relationship between imprecisely measured traits using GWAS summary data. *PLoS Genet.* **13**, e1007081 (2017).
84. Bulik-Sullivan, B. K. et al. LD score regression distinguishes confounding from polygenicity in genome-wide association studies. *Nat. Genet.* **47**, 291–295 (2015).
85. Ning, Z., Pawitan, Y. & Shen, X. High-definition likelihood inference of genetic correlations across human complex traits. *Nat. Genet.* **52**, 859–864 (2020).
86. Ray, D. & Chatterjee, N. A powerful method for pleiotropic analysis under composite null hypothesis identifies novel shared loci between type 2 diabetes and prostate cancer. *PLoS Genet.* **16**, e1009218 (2020).
87. Watanabe, K., Taskesen, E., van Bochoven, A. & Posthuma, D. Functional mapping and annotation of genetic associations with FUMA. *Nat. Commun.* **8**, 1826 (2017).
88. Buniello, A. et al. The NHGRI-EBI GWAS Catalog of published genome-wide association studies, targeted arrays and summary statistics 2019. *Nucleic Acids Res.* **47**, D1005–D1012 (2019).
89. Elsworth, B. et al. The MRC IEU OpenGWAS data infrastructure. Preprint at *bioRxiv* <https://doi.org/10.1101/2020.08.10.244293> (2020).
90. Watanabe, K. et al. A global overview of pleiotropy and genetic architecture in complex traits. *Nat. Genet.* **51**, 1339–1348 (2019).
91. de Leeuw, C. A., Mooij, J. M., Heskes, T. & Posthuma, D. MAGMA: generalized gene-set analysis of GWAS data. *PLoS Comput. Biol.* **11**, e1004219 (2015).
92. Sun, B. B. et al. Plasma proteomic associations with genetics and health in the UK Biobank. *Nature* **622**, 329–338 (2023).
93. Zhu, Z. et al. Integration of summary data from GWAS and eQTL studies predicts complex trait gene targets. *Nat. Genet.* **48**, 481–487 (2016).
94. Restuadi, R. et al. Functional characterisation of the amyotrophic lateral sclerosis risk locus GPX3/TNIP1. *Genome Med.* **14**, 7 (2022).
95. Subramanian, A. et al. Gene set enrichment analysis: a knowledge-based approach for interpreting genome-wide expression profiles. *Proc. Natl Acad. Sci. USA* **102**, 15545–15550 (2005).

Acknowledgements

We wish to extend our heartfelt thanks to the UK Biobank team for their essential role in furthering clinical research in our area. We are also grateful to the participants of all the GWASs used in this manuscript and the investigators who made these GWAS data publicly available. This work was funded by the grants from the National Natural Science Foundation of China (82173498 and 82373561 to T.H., 82204017 to W.J., 82204016 to W.W., 62472131 and 62172131 to Y.Z.).

Author contributions

T.H., W.W. and Y.Z. developed the idea of the research. W.J., H.W., Y.G., Q.L., M.G. and S.C. carried out the main data analyses. T.H., W.W. and Y.Z. interpreted the data. S.L., D.Z. and C.Q. prepared and revised the figures. M.G., Z.L., B.L., S.F. and S.H. performed the data cleaning and applied for the UK Biobank data. All authors reviewed the manuscript drafts, critically revised the manuscript, and approved the final manuscript.

Competing interests

The authors declare no competing interests.

Additional information

Supplementary information The online version contains supplementary material available at <https://doi.org/10.1038/s41467-025-59945-2>.

Correspondence and requests for materials should be addressed to Yunpeng Zhang, Wei Wei or Tianshu Han.

Peer review information *Nature Communications* thanks Matthias Heinig, Daniel Levey, and the other anonymous reviewer(s) for their contribution to the peer review of this work. A peer review file is available.

Reprints and permissions information is available at <http://www.nature.com/reprints>

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Open Access This article is licensed under a Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 International License, which permits any non-commercial use, sharing, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if you modified the licensed material. You do not have permission under this licence to share adapted material derived from this article or parts of it. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by-nc-nd/4.0/>.

© The Author(s) 2025