

RESEARCH

Open Access



A biologically informed framework for instrument selection in dietary Mendelian randomization using chemosensory genetics

Liang-Dar Hwang^{1,2*}, Cailu Lin², David M Evans^{1,3,4}, Nicholas G Martin⁵, Danielle R. Reed^{2*} and Paule V. Joseph^{6*}

Abstract

Background Mendelian randomization (MR) is increasingly used to strengthen causal inference in nutritional epidemiology. However, dietary MR studies often rely on instruments selected from genome-wide association studies of self-reported intake based solely on statistical significance, increasing vulnerability to pleiotropy and reverse causation and potentially violating key MR assumptions. We aimed to develop and evaluate a biologically informed framework for selecting valid genetic instruments for dietary exposures, leveraging genes encoding taste and olfactory receptors that influence chemosensory perception and shape food preferences and dietary behaviour.

Methods We prioritised 1,214 nonsynonymous variants (minor allele frequency $\geq 1\%$) across 30 taste and 295 olfactory receptor genes. Associations with 140 food-liking traits were tested in UK Biobank participants aged 37 to 73 years. Candidate variants were evaluated using a multi-stage filtering pipeline comprising replication in an independent younger cohort (Avon Longitudinal Study of Parents and Children, age 25), concordance between food liking and intake, exclusion of associations with socioeconomic status, assessment of food specificity accounting for linkage disequilibrium and co-consumption, and directionality testing to reduce reverse causation. Retained variants were used as instruments in MR analyses of cardiometabolic outcomes.

Results We identified 268 variants within 101 olfactory and 16 taste receptor genes associated with 96 food-liking traits. Filtering yielded 24 candidate instruments for 20 foods. The instrument for onion liking satisfied all pre-defined criteria for classification as high confidence. As an illustrative MR application, genetically proxied onion liking was associated with lower systolic and diastolic blood pressure and reduced risk of type 2 diabetes, with no evidence of effects on body mass index, glycaemic traits, or serum lipid levels.

Conclusions Instrument selection guided by chemosensory genetics provides a scalable strategy for dietary MR that can improve instrument credibility and reduce susceptibility to pleiotropy and reverse causation. However, this approach prioritises biological specificity at the cost of fewer available instruments. This framework supports more

*Correspondence:

Liang-Dar Hwang
d.hwang@uq.edu.au
Danielle R. Reed
reed@monell.org
Paule V. Joseph
paule.joseph@nih.gov

Full list of author information is available at the end of the article



© The Author(s) 2026. **Open Access** This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, 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 changes were made. 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/4.0/>.

robust causal evaluation of diet–disease relationships and strengthens inference in nutritional epidemiology and public health research.

Keywords Mendelian randomization, Taste receptor, Olfactory receptor, Food liking, Dietary exposure, Blood pressure, Type 2 diabetes, UK Biobank, ALSPAC

Background

Unhealthy diets are a major contributor to the global disease burden, accounting for an estimated 11 million premature deaths annually, primarily from cardiovascular disease, followed by cancer, and type 2 diabetes [1]. Rates of early-onset obesity, cancers [2], and diabetes [3] continue to rise, placing increasing pressure on healthcare systems. Despite extensive observational evidence linking diet to chronic disease, robust causal evidence for specific foods, dietary patterns, or nutrients remains limited. Randomized controlled trials (RCTs), the gold standard for causal inference, are rarely feasible in nutrition due to long follow-up periods, high costs, logistical constraints, and poor adherence [4]. Observational studies are frequently confounded or affected by reverse causation; for example, while observational studies suggested that vitamin E supplementation lowers coronary artery disease risk, large-scale RCTs failed to replicate this effect [5]. Such discrepancies underscore the need for rigorous methods that can reliably identify the causal effects of diet.

Mendelian randomization (MR) provides an alternative framework for causal inference. By using genetic variants as instrumental variables (IVs), MR mimics the randomization process of RCTs through Mendel's Laws of Segregation and Independent Assortment, thereby reducing confounding [6]. MR has reshaped the understanding of cardiometabolic disease by confirming the causal role of low-density lipoprotein cholesterol in cardiovascular disease, while demonstrating that associations with high-density lipoprotein cholesterol are likely non-causal. These findings align with RCTs and have influenced clinical practice and drug development [7]. Because MR reflects lifelong exposure effects, it offers practical advantages over short-term, resource-intensive RCTs.

The application of MR in nutrition research is expanding [8], with studies clarifying the causal effects of coffee [9], alcohol [10], and milk intake [11]. Despite these advances, MR analyses of foods and dietary patterns remain challenging [8], mainly because it is difficult to identify valid IVs. The MR exclusion restriction assumption requires that genetic variants affect outcomes solely through the exposure (i.e., no horizontal pleiotropy; Fig. 1, blue arrow). Many dietary instruments are selected solely based on statistical significance in genome-wide association studies (GWAS), thereby increasing susceptibility to pleiotropy. This problem is compounded by the fact that most diet GWAS have been conducted in older

adults [12–21], whose food choices may reflect pre-existing health conditions (Fig. 1, orange arrow). For example, *FTO* variants associated with higher body mass index [22] are counterintuitively related to decreased sugar intake [13], and *APOE* variants linked to increased risk of cardiovascular [23] and neurological disorders [24] are likewise associated with reduced preference for fried foods [19]. These patterns are unlikely to reflect primary effects on diet; instead, they suggest reverse causation, in which individuals modify their diet after diagnosis [14, 15]. Using such loci as IVs in MR would introduce substantial bias and exacerbate inconsistency in nutritional epidemiology.

Genetic variants that influence taste and olfactory perception show promise as candidate IVs for dietary exposures. The chemical senses play a central role in shaping food preferences and choices, and variants associated with chemosensory perception [25–30] have consistently been linked to differences in dietary behavior [31–34]. For example, variants in bitter taste receptor genes influence the bitterness of quinine [26] and caffeine [28] and predict the consumption of bitter beverages [31]. Large-scale GWAS of food liking and intake have also identified variants within taste and olfactory receptor genes [14, 18, 19]. Because chemosensory perception influences eating behaviour before the onset of most chronic disease [35], associations between these variants and dietary intake are less likely to reflect disease-driven behavioral changes and may be less prone to health-related confounding. These features make genetic variants in chemosensory pathways biologically interpretable and potentially credible candidates for use as IVs in dietary MR.

Here, we evaluated whether genetic variants in taste and olfactory receptor genes can serve as IVs for dietary exposures. We prioritized nonsynonymous variants within chemosensory receptor genes (i.e., the taste receptor type 1 [T1Rs] and type 2 [T2Rs] families, and the olfactory receptor [OR] gene families), as these variants alter amino acid sequences and are therefore more likely to affect receptor structure and potentially ligand binding. Using UK Biobank food-liking traits [36], we identified candidate variants and applied a pre-specified multi-stage filtering framework comprising replication in a younger cohort (Avon Longitudinal Study of Parents and Children [ALSPAC] [37, 38]), concordance with intake of the same foods, exclusion of socioeconomic associations, and assessment of food specificity accounting for linkage disequilibrium (LD). We then used

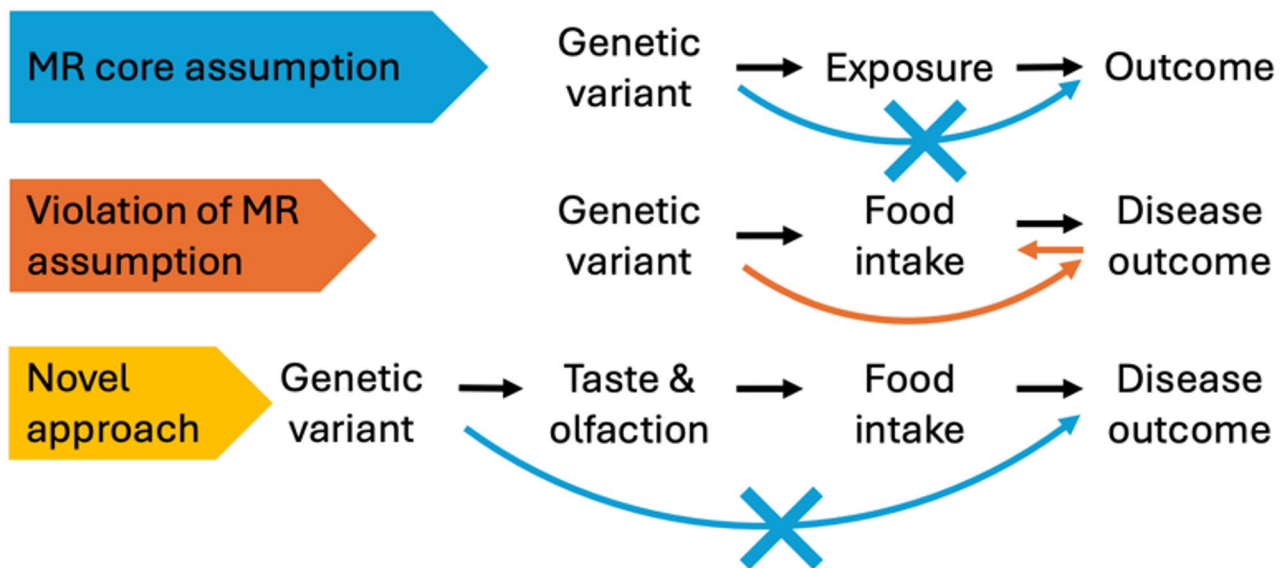


Fig. 1 Genetic variants associated with chemosensory perception as novel instrumental variable for dietary exposures in Mendelian randomization (MR) for causal inference. MR assumes that a genetic variant affects an outcome solely through the exposure of interest (i.e., no pleiotropy; **Top**, blue arrow). However, most variants associated with dietary intake have been identified in older individuals, where food choices often reflect pre-existing health conditions (**Middle**, orange arrow) rather than cause them. The effect of genetic variants associated with chemosensory perception on diet is less likely to be influenced by health status, making them superior instruments for dietary exposures (**Bottom**)

retained instruments in MR analyses of cardiometabolic outcomes. An overview of the study design is presented in **Additional file 1: Fig S1**.

Methods

Sample and genotyping

UK Biobank is a prospective cohort study of ~ 500,000 participants (aged 37–73 years; 54.4% female; ~5% of those invited) recruited across England, Wales, and Scotland between 2006 and 2010. Baseline questionnaires, clinical assessments and biospecimens have been described previously [36]. This project used the resource under Application 57,780. Genotyping was performed using the Affymetrix UK BiLEVE Axiom or Affymetrix UK Biobank Axiom arrays. Imputation was performed using IMPUTE2 software with the UK10K haplotype and Haplotype Reference Consortium reference panels [36]. Single nucleotide polymorphisms (SNPs) with a call rate < 90%, minor allele frequency < 0.005, imputation score < 0.3, and Hardy–Weinberg equilibrium score of $P < 1.0 \times 10^{-6}$ were excluded. Ancestry was assigned using a distance-based genetic relationship and fingerprinting (Graf) method [39]; individuals clustering with the European group were retained. Relatedness was defined as having a KING kinship coefficient > 0.0884 (i.e., third-degree relatives or closer) [40]; one individual from each pair was removed.

ALSPAC is a longitudinal birth cohort study that enrolled pregnant women residing in and around the city of Bristol in the South West of England with expected

delivery dates between April 1, 1991, and December 31, 1992 [37, 38]. The enrolled cohort included 15,247 pregnancies resulting in 14,775 live-born babies. The mothers and their children have been followed up through postal questionnaires and at clinics. A searchable variable dictionary is available online (<http://www.bristol.ac.uk/alspac/researchers/our-data/>). The ALSPAC Ethics and Law Committee and local Research Ethics Committees granted ethical approval. This study used the children's cohort. Genotyping was performed using the Illumina HumanHap550 quad chip array. SNPs with a call rate < 90%, minor allele frequency < 0.01, and Hardy–Weinberg equilibrium score of $P < 1.0 \times 10^{-6}$ were excluded. Imputation was performed using Impute V2.2.2 with the 1000 Genomes Phase 1 Version 3 reference panel [38]. Ancestry was assessed using multidimensional analysis and compared with HapMap II (release 22) European descent (CEU), Han Chinese, Japanese, and Yoruba reference populations; all individuals with non-European ancestry were removed. Relatedness was defined using a genome-wide identity-by-descent estimated from PLINK [41]; one individual from each pair with PI_HAT > 0.10 was removed.

Common missense, nonsense, and frameshift variants in chemosensory receptor genes

We downloaded the complete list of human genes (GRCh37/hg19) from the GENCODE project [42] and extracted taste receptor genes (i.e., gene symbol starting with *TAS1R* or *TAS2R*) and olfactory receptor genes

(i.e., gene symbol starting with *OR*). We excluded pseudogenes and genes without chromosome or base-pair information. Based on the chromosome locations of each gene, we extracted all variants within each gene and excluded those with a minor allele frequency < 0.01 in the UK Biobank. The remaining genetic variants were annotated using Haploreg (v4.2) [43] and dbSNP [44] for functional annotations. Only variants annotated as missense, nonsense, or frameshift mutations were included in the analysis.

Food liking and food intake

Food-liking traits from UK Biobank were collected through an online questionnaire comprising 152 items, including 140 food and drink items and additional non-food items that captured liking for health-related behaviors such as physical activity (<https://biobank.ndph.ox.ac.uk/showcase/showcase/docs/foodpref.pdf>). Participants rated their liking on a 9-point Hedonic scale, with 1 corresponding to “Extremely dislike” and 9 to “Extremely like”. The questionnaire was administered in 2019 to all UK Biobank participants who had agreed to be recontacted by the study. Participants who answered “Never tried” or “Do not wish to answer” were excluded. Data from up to 182,165 individuals were available for analysis. For significant SNP-food-liking associations, we further examined the associations between the same SNPs and the intake of the same food items. Dietary intake in the UK Biobank was assessed using i) a touchscreen dietary frequency questionnaire from all participants during their visits to the assessment center and ii) a follow-up online 24-hour recall dietary questionnaire [45] from up to 214,950 individuals. Answers of “None” from the 24-hour recall were recoded to 0. Data collected from multiple instances were averaged before analysis.

Data on food liking in ALSPAC were collected from children of pregnant women recruited to the study when they were 25 years old using the Life at 25 + Questionnaire. Preferences for 97 food items were measured using a 9-point Hedonic scale similar to the one used in UK Biobank, with 1 corresponding to “Extremely dislike”, 9 to “Extremely like”, and 0 to “Never tasted”. From October 2017, participants received an email containing a username/password for the online version of the questionnaire. Participants without email addresses or who did not respond to the email were sent a paper questionnaire. The questionnaire was sent out to 10,001 live-born children within the ALSPAC cohort. As of 1st October 2018, 4398 completed questionnaires had been returned. Study data were collected and managed using REDCap electronic data capture tools hosted at the University of Bristol [46]. REDCap (Research Electronic Data Capture) is a secure, web-based software platform designed to support data capture for research studies. Only unrelated

individuals of European ancestry with genotyping data, age, and sex information, and who provided an answer between 1 and 9 were included, leaving a total of up to 2,779 responses per food item, with 64% female. The ALSPAC participants included in this study did not overlap with those from the UK Biobank.

Statistical analysis

We first calculated the effect of each SNP on each food-liking trait in the UK Biobank using a linear regression model. Covariates included sex, age, and the first 10 genetic principal components (PCs) (Formula 1).

Formula 1: Food liking ~ SNP + sex + age + PCs.

A prior significance threshold was set at an FDR-adjusted p-value of 0.05 to account for multiple testing. Four downstream analyses were performed following the significant associations.

First, for significantly associated SNPs, we assessed their associations with socioeconomic status, proxied by the Townsend Deprivation Index, in the UK Biobank using GWAS summary statistics from the OpenGWAS database [47].

Second, for SNPs identified in the UK Biobank, we investigated their association with their corresponding food-liking traits in the younger ALSPAC cohort using the same linear regression model [48]. For SNPs unavailable in ALSPAC, their proxies with a LD of r^2 (square of the Pearson correlation coefficient between allele dosages of two SNPs) > 0.8 in the British in England and Scotland (GBR) population were identified using the LDproxy Tool [49]. We performed a heterogeneity test to compare the difference in SNP effects between the UK Biobank and ALSPAC.

Third, we investigated associations with intake of the corresponding food items in the UK Biobank using the same linear regression model.

Fourth, we quantified the variance explained for each food-liking trait in UK Biobank and ALSPAC, and for the corresponding food intake trait in UK Biobank, using the SNPs that were significantly associated with the food-liking trait in UK Biobank. We constructed a base model with only covariates (Formula 2) and a full model with covariates and SNPs associated with a trait (Formula 3). The variance explained for each trait was calculated as the difference between the R^2 of the base model and that of the full model.

Formula 2: Food liking ~ sex + age + PCs.

Formula 3: Food liking ~ SNP₁ + SNP₂ + ... + SNP_n + sex + age + PCs.

Where the SNPs were those significantly associated with a food-liking trait identified using formula 1.

Next, we selected SNPs to be used as IVs for dietary exposures. To ensure that a SNP only affects the outcome through one dietary exposure of interest, each



Fig. 2 (See legend on next page.)

(See figure on previous page.)

Fig. 2 Associations between 96 food-liking traits and 268 nonsynonymous single-nucleotide polymorphisms (SNPs) within 101 olfactory (left panel) and 16 taste (right panel) receptor genes identified in UK Biobank (FDR-adjusted p-value < 0.05). Food items are clustered and colored according to their properties. A single SNP can be associated with up to 17 food-liking traits (i.e., *OR4K17* rs8005245), and 117 variants were only associated with one specific trait. Liking for fish, dairy, and carbohydrates is only related to variants within olfactory receptor genes, whereas liking for caffeinated and most alcoholic beverages is only associated with variants within taste receptor genes

food-liking associated SNP was screened based on: (i) it is only associated with one food-liking trait, and (ii) all food-liking associated SNPs in LD ($r^2 \geq 0.1$) with the selected SNP are also only associated with the same food-liking trait. Given that MR is a method to assess the causal effect of long-term exposure, IVs were then tiered based on whether they have consistent effects across young (ALSPAC) and middle-aged adults (UK Biobank). To reflect the potential causal effect of food consumption, IVs were further tiered based on their association with food intake. Tier 1 IVs were SNPs with the same direction of effect on food liking in the ALSPAC (p-value < 0.05) and on food intake in the UK Biobank (p-value < 0.05); tier 2 SNPs were those with the same direction of effect on food liking in ALSPAC (p-value < 0.05); the remaining SNPs belonged to tier 3. See **Additional file 1: Fig S2** for a pipeline illustrating the selection of IVs and **Additional file 2: Table S1** for LD between SNPs.

Finally, we used SNPs significantly associated with food-liking traits as IVs in MR to assess the potential causal influences of food liking on cardiometabolic traits. SNP associations with cardiometabolic traits were extracted from published GWAS summary results statistics, including systolic blood pressure, diastolic blood pressure [50], high-density lipoprotein cholesterol, low-density lipoprotein cholesterol, total lipid cholesterol [51], fasting glucose, fasting insulin, hemoglobin A1C [52], BMI [22], coronary artery disease [23], and type 2 diabetes [53]. All associations were extracted from GWAS primarily of European individuals. We prioritized using data without UK Biobank participants to avoid bias toward observational associations. See **Additional file 2: Table S2** for details and data sources of cardiometabolic traits.

MR analyses were performed using the Wald estimator when an exposure only had one IV, and the causal effect was calculated as $\hat{\beta}_{wald} = \hat{\beta}_{outcome} / \hat{\beta}_{exposure}$ with the standard error being estimated using the Delta method as $\widehat{SE}_{wald} = \widehat{SE}_{outcome} / \hat{\beta}_{exposure} \cdot \hat{\beta}_{exposure}$ and $\hat{\beta}_{outcome}$ are the estimated effect of the IV on food liking and the cardiometabolic outcome, respectively; $\widehat{SE}_{outcome}$ is the standard error of $\hat{\beta}_{outcome}$. When an exposure had two IVs, the causal effect was further calculated using the inverse variance-weighted (IVW) method that meta-analyzes the $\hat{\beta}_{wald}$ of each IV weighted by the inverse of their variance [54]. Steiger filtering was applied

to ensure that SNPs were consistent with the exposure-outcome direction; SNPs explaining more variance in the outcome than in the exposure would suggest reverse causation [55]. For analyses with two IVs, we report Cochran's Q as a heterogeneity check (df = 1). Pleiotropy-robust estimators that require three or more IVs (e.g., MR Egger [56], weighted median [57]) were not applicable. For significant MR results, if the outcome datasets contained UK Biobank participants, we performed sensitivity analyses to assess whether the effects were driven by UK Biobank by testing the IV-outcome association in UK Biobank using GWAS summary statistics results from the Neale lab (<http://www.nealelab.is/uk-biobank/>). See **Additional file 1: Supplementary Note** for assumptions underlying Mendelian randomization and sensitivity analyses.

We conducted a direct comparison between the biologically informed IV selection approach and a conventional GWAS-based approach. For the conventional approach, IVs were selected from GWAS summary statistics (May-Wilson et al., 2021) using standard criteria (p-value = 5×10^{-8} , clumping window = 10,000 kb, and $r^2 = 0.001$), followed by Steiger filtering. MR analyses were performed using IVW, MR-Egger, and weighted median estimators implemented in the TwoSampleMR package (v0.6.19). To evaluate instrument validity, we conducted a phenome-wide association scan of the GWAS-derived IVs using ieugwasr (v1.0.2) [58], querying OpenGWAS (accessed 27 March 2026) [47] and applying a threshold of p-value < 1×10^{-5} to identify potential associations with confounders.

All analyses were performed using R version 4.4.1 and RStudio version 2024.4.1.748 (RStudio Team, 2020).

Results

Associations with food liking and intake

Of the 883 chemosensory receptor genes in the human genome (GRCh37/hg19), 30 taste receptor genes and 395 olfactory receptor genes are non-pseudo and have known chromosomal locations. Within these 425 genes, 7,400 single-nucleotide polymorphisms (SNPs) have a minor allele frequency $\geq 1\%$, of which 1,158 are missense, 29 are nonsense, and 27 are frameshift variants (**Additional file 2: Table S3**).

We tested the association between these 1,214 nonsynonymous SNPs and 140 food-liking traits in the UK Biobank. We identified 700 associations with an FDR-adjusted p-value < 0.05 across 268 SNPs within 117 genes

(16 taste receptor genes and 101 olfactory receptor genes) and 96 food-liking traits (Fig. 2; **Additional file 1: Fig S3**; **Additional file 2: Table S4**). The strongest associations were with liking garlic (p-value = 1.53×10^{-69} with *OR4K17* rs8005245), grapefruit (p-value = 1.88×10^{-54} with *TAS2R19* rs10772420), onions (p-value = 5.38×10^{-41} with *OR2T6* rs6587467), horseradish/wasabi (p-value = 2.28×10^{-32} with *TAS2R38* rs713598), adding salt to foods (p-value = 2.85×10^{-24} with *TAS2R38* rs10246939), and broad beans (p-value = 1.30×10^{-19} with *TAS2R16* rs860170). Of the 268 SNPs associated with food liking, 37 (13.8%) showed a nominal association with Townsend Deprivation Index in UK Biobank (p-value < 0.05), 12 (4.5%) at p-value < 0.01, and 3 (1.1%) at p-value < 0.001. These SNPs were associated with liking butternut squash, marzipan, plain yogurt, cream, strawberries, liver, pork chop, herring, and aniseed (**Additional file 2: Table S4**).

Among the 268 SNPs and 96 associated food-liking traits identified in UK Biobank, data for 223 SNPs or their proxy SNPs (i.e., SNPs in LD with $r^2 > 0.8$; **Additional file 2: Table S5**) and 66 corresponding food-liking traits are available in ALSPAC (**Additional file 2: Table S6**). The majority of associations found in ALSPAC (410 out of 540 associations being tested) had the same direction of effect as observed in UK Biobank, among which 45 associations had a p-value < 0.05 (Fig. 3, left panel); **Additional file 2: Table S4**). The strongest association identified in the UK Biobank (for liking garlic) was not significant in ALSPAC (p-value = 0.60 with *OR4K17* rs8005245).

We observed a trend that the variants identified in UK Biobank had a larger effect on the same food items in ALSPAC (Fig. 3, left panel) and explained more variance (**Additional file 1: Fig S4**; **Additional file 2: Table S7**) in ALSPAC; however, more than 95% of associations (514/540) showed no significant heterogeneity between cohorts (p-value > 0.05; **Additional file 2: Table S4**). Among the 96 food-liking traits identified in UK Biobank, 60 had corresponding food-intake or proxied food-intake traits available in UK Biobank (**Additional file 2: Table S6**). Amongst the SNPs that were significantly associated with food liking, we identified 121 associations for intake of the same food items (p-value < 0.05), with 120 exhibiting the same direction of effect (Fig. 3, right panel); **Additional file 2: Table S8**).

Instrumental variable selection and Mendelian randomization

Among the 268 food-liking-associated SNPs, each was associated with 1 to 17 food-liking traits. After excluding SNPs (and SNPs in LD) with pleiotropic effects on multiple food-liking traits, we identified 24 IVs for 20 food-liking traits, including a tier 1 IV for liking onions (Table 1;

see **Additional file 2: Table S9** for full results). The SNP *OR2T6* rs6587467 for liking onions had the same direction of effect on liking onions in UK Biobank ($\beta = 0.087$, p-value = 5.38×10^{-41} , $r^2 = 0.11\%$) and ALSPAC ($\beta = 0.23$, p-value = 4.78×10^{-4} , $r^2 = 0.48\%$), suggesting its long-lasting effect across adulthood. Further, the SNP was also associated with onion intake in the predicted direction ($\beta = 0.002$, p-value = 5.43×10^{-9} , $r^2 = 0.02\%$).

MR showed that each point increased in the liking onions on a 9-point Hedonic scale was associated with lower systolic blood pressure ($\beta = -1.256$ mmHg, p-value = 0.001), diastolic blood pressure ($\beta = -0.716$ mmHg, p-value = 0.001), and the risk of type 2 diabetes (odds ratio [95% confidence interval] = 0.856 [0.781, 0.939], p-value = 0.001) (Fig. 4). There was no strong evidence for a causal effect of liking onions on serum lipid cholesterol, blood glucose, BMI, or coronary artery disease. See MR results, Steiger filtering, and Cochran's Q tests for all food-liking traits in **Additional file 2: Table S10**.

When using the conventional significance-based approach, six IVs for onion liking were identified. MR analyses using these instruments yielded inconsistent results across estimators for BMI, with effect estimates differing in both magnitude and direction (IVW: $\beta = 0.140$, p-value = 0.141; weighted median: $\beta = 0.172$, p-value = 4.59×10^{-3} ; MR-Egger: $\beta = -0.335$, $p = 0.041$), alongside evidence of directional pleiotropy (MR-Egger intercept $p = 0.01$). For blood pressure, the protective effect observed using the *OR2T6* instrument was replicated only in the weighted median estimator (p-value < 0.001) but not in IVW or MR-Egger analyses. No association with type 2 diabetes was observed across any method (**Additional file 2: Table S11**). Phenome-wide association scans further indicated that, except for the IV (rs7511986) in high LD with the nonsynonymous *OR2T6* variant ($r^2 = 0.9$), the remaining five IVs were associated with potential confounders (p-value < 1×10^{-5}), indicating a high likelihood of horizontal pleiotropy (**Additional file 2: Table S12**).

Discussion

This study sought to address a central methodological challenge in dietary MR: the identification of biologically credible IVs for dietary exposures. Many existing dietary instruments are derived from significance-based GWAS conducted in older adults, in whom self-reported diet may reflect underlying health status rather than inherent preferences. We proposed that genetic variants for chemosensory perception—biological determinants of food liking established prior to most chronic disease—could provide more interpretable and potentially more credible instruments. Applying a pre-defined multi-stage

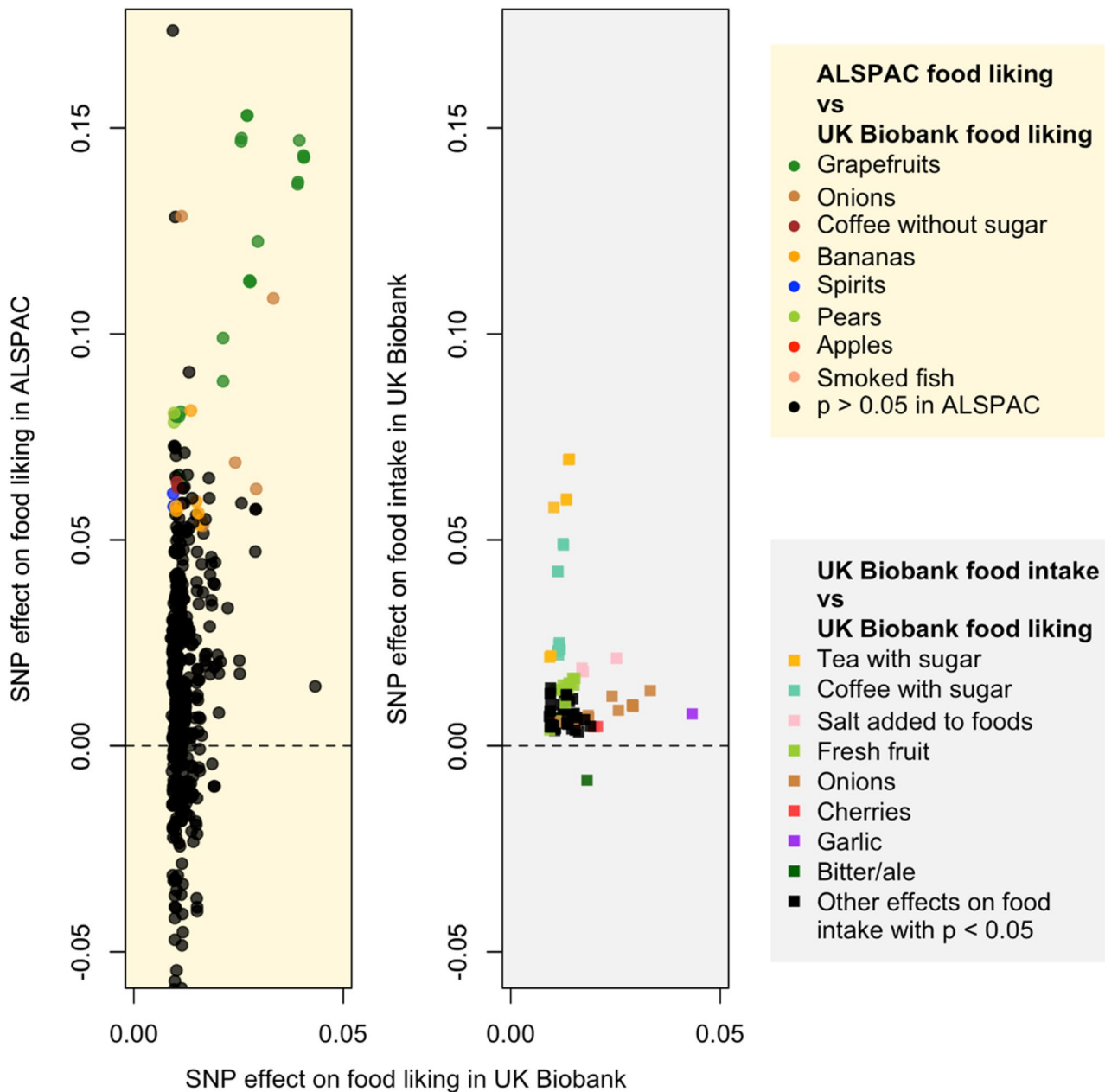


Fig. 3 Comparison of standardized SNP effects on food-liking and intake traits in UK Biobank and Avon Longitudinal Study of Parents and Children (ALSPAC). **Left panel:** the most substantial concordant food-liking effects between UK Biobank and ALSPAC are for liking grapefruit and missense variants within a cluster of bitter taste receptors on chromosome 12 (*TAS2R31* rs10845293). Other concordant effects include variants for liking onions (*OR2T6* rs6587467), pears (*OR2A25* rs6951485), and apples (*OR2F2* rs2240359). There is a trend that the SNPs identified in the UK Biobank have a larger effect on the same food items in ALSPAC. **Right panel:** all food-liking SNPs affect their corresponding intake traits in the same direction; for example, the *OR4K17* rs8005245 C allele was associated with increased liking and higher intake of garlic ($p\text{-value} = 7.17 \times 10^{-4}$). The only exception is *TAS2R1* rs2234233 and bitter/ale, where the G allele is associated with increased liking and lower intake. This was probably because liking was measured for bitter or ale (both bitter), while intake was measured for beer or cider (with the latter being sweet). Only food-liking-associated SNPs associated with intake ($p < 0.05$) are plotted. Alleles are oriented to have a positive effect on food-liking traits

filtering framework, we identified 24 instruments for 20 dietary exposures, including a high-confidence instrument for onion liking (*OR2T6* rs6587467) that demonstrated food-specificity, absence of socioeconomic association, and consistent effects across young and

middle-aged adulthood. MR analyses using this variant suggested associations between genetically proxied onion liking and lower blood pressure and reduced risk of type 2 diabetes, without evidence of effects on BMI, glycaemic traits, or lipid levels. These estimates likely reflect

Table 1 Instrumental variables for food-liking traits identified in the UK Biobank

Trait	Gene	SNP	EA/NEA	EAF	Beta	SE	P-value	R ²	F-statistics	Tier*
Onions	<i>OR2T6</i>	rs6587467	G/T	0.69	0.087	0.007	5.38×10^{-41}	0.11%	179.627	1
Bananas	<i>OR5H6</i>	rs9853906	G/A	0.69	0.029	0.007	3.91×10^{-5}	0.01%	16.839	2
Broad beans	<i>TAS2R16</i>	rs860170	T/C	0.68	-0.08	0.009	1.30×10^{-19}	0.05%	81.029	3
Grapefruit	<i>TAS2R7</i>	rs619381	T/C	0.13	0.075	0.014	2.78×10^{-8}	0.02%	30.740	3
Tomato ketchup	<i>OR2B11</i>	rs12070953	C/T	0.14	0.057	0.012	2.74×10^{-6}	0.01%	21.446	3
Avocado	<i>OR10C1</i>	rs17177674	A/C	0.01	-0.186	0.043	1.46×10^{-5}	0.01%	18.344	3
Vinegar	<i>OR4A16</i>	rs117538213	C/T	0.04	0.078	0.019	3.17×10^{-5}	0.01%	17.161	3
Cauliflower	<i>OR51I2</i>	rs12577167	G/A	0.12	0.038	0.009	3.90×10^{-5}	0.01%	16.518	3
Goats cheese	<i>OR1A2</i>	rs2469791	T/C	0.26	-0.044	0.011	5.54×10^{-5}	0.01%	16.236	3
Salty pretzels	<i>OR4E2</i>	rs2874103	A/G	0.82	-0.045	0.011	4.28×10^{-5}	0.01%	16.226	3
Garlic	<i>OR11G2</i>	rs4981088	A/G	0.52	0.028	0.007	1.59×10^{-4}	0.01%	16.032	3
Bananas	<i>OR2F2</i>	rs61739648	A/G	0.02	0.092	0.023	6.99×10^{-5}	0.01%	15.867	3
Black pepper	<i>OR8D4</i>	rs10790610	C/T	0.33	0.03	0.007	7.01×10^{-5}	0.01%	15.800	3
Green olives	<i>TAS1R1</i>	rs34160967	A/G	0.13	-0.062	0.016	7.59×10^{-5}	0.01%	15.595	3
Fresh tomatoes	<i>OR56B1</i>	rs62621167	T/C	0.05	0.054	0.013	6.78×10^{-5}	0.01%	15.578	3
Aniseed	<i>OR13D1</i>	rs13294411	G/C	0.09	-0.062	0.016	8.68×10^{-5}	0.01%	15.439	3
Goats cheese	<i>OR10J1</i>	rs12048482	G/A	0.64	-0.039	0.01	8.92×10^{-5}	0.01%	15.317	3
Apple juice	<i>OR2C1</i>	rs62000975	A/G	0.01	0.132	0.034	1.01×10^{-4}	0.01%	14.87	3
Pollock	<i>OR52K2</i>	rs61997232	C/A	0.03	0.098	0.026	1.52×10^{-4}	0.01%	14.228	3
Oranges	<i>OR5H15</i>	rs13082608	T/C	0.18	-0.03	0.008	1.71×10^{-4}	0.01%	14.085	3
Shellfish other than prawns	<i>OR51I2</i>	rs35301588	TCA/T	0.38	-0.041	0.011	1.62×10^{-4}	0.01%	14.040	3
Biscuits	<i>OR2T4</i>	rs57795102	G/A	0.01	0.129	0.035	1.86×10^{-4}	0.01%	13.921	3
Broad beans	<i>OR52I1</i>	rs146674516	C/T	0.02	0.116	0.031	2.03×10^{-4}	0.01%	13.691	3
Garlic	<i>OR4K1</i>	rs34394400	T/C	0.31	-0.037	0.01	1.86×10^{-4}	0.01%	13.626	3

*Tier 1: The effect on food liking is replicated in ALSPAC (p-value < 0.05), and the SNP affects the intake of the same food in the UK Biobank (p-value < 0.05)

Tier 2: The effect on food liking is replicated in the Avon Longitudinal Study of Parents and Children

Tier 3: others. SNP, single nucleotide polymorphism; EA/NEA, effect allele/non-effect allele; EAF, effect allele frequency; SE, standard error

Units for beta-coefficients are "point" on a 9-point liking scale. R² represents the variance explained in food liking by the SNP

broader biological and behavioural pathways captured by the genetic instrument rather than a direct effect of a unit change in food liking. They should be interpreted cautiously in light of potential residual pleiotropy and the trade-off between maximising biological specificity and limiting the number of available instruments. More broadly, this framework aligns with recent calls to strengthen the credibility of MR by improving instrument validity, interpretability, and resistance to reverse causation [59].

Our approach, focusing on nonsynonymous variation in taste and olfactory receptor genes, identified 268 SNPs across 16 taste and 101 olfactory receptor genes associated with 96 food-liking traits. These included 9 of the 10 non-pseudo chemosensory receptor genes previously identified in the GWAS of food liking using the same data set from UK Biobank [19]. Many of these associations would not have been identified in a hypothesis-free GWAS using the conventional genome-wide significance threshold (i.e., $p = 5 \times 10^{-8}$). While some SNPs were linked to a single food, most variants were associated with multiple foods, consistent with the idea that individual receptors respond to shared chemical compounds (e.g. volatiles, nutrients, or flavonoids) across items [60,

61]. This pattern highlights the challenge of mapping receptor-level detection of chemical compounds onto self-reported food items, which often aggregate multiple sensory components. Most liking-associated SNPs also influenced intake of the corresponding foods in the same direction. Fifty-four associations were replicated in young adults from ALSPAC. This replication is important because the relationship between a genetic variant and food liking may change across the lifespan. The persistence of these associations from young adulthood into mid/late adulthood supports their validity as IVs for dietary exposures. Associations that did not replicate (e.g., *OR4K17* rs8005245 and liking garlic) may reflect limited statistical power in ALSPAC and/or the effect of some variants may not generalize uniformly across the life course.

We intentionally restricted the screening to G protein-coupled receptors, i.e., the *TAS1R/TAS2R* families and *OR* repertoire, because these seven-transmembrane receptors directly mediate detection of tastants and odorants in the oral and nasal cavities [62, 63]. The *OR2T6* instrument is particularly notable because *OR2T6* expression is largely restricted to the olfactory epithelium [64, 65], providing biological coherence for its role in

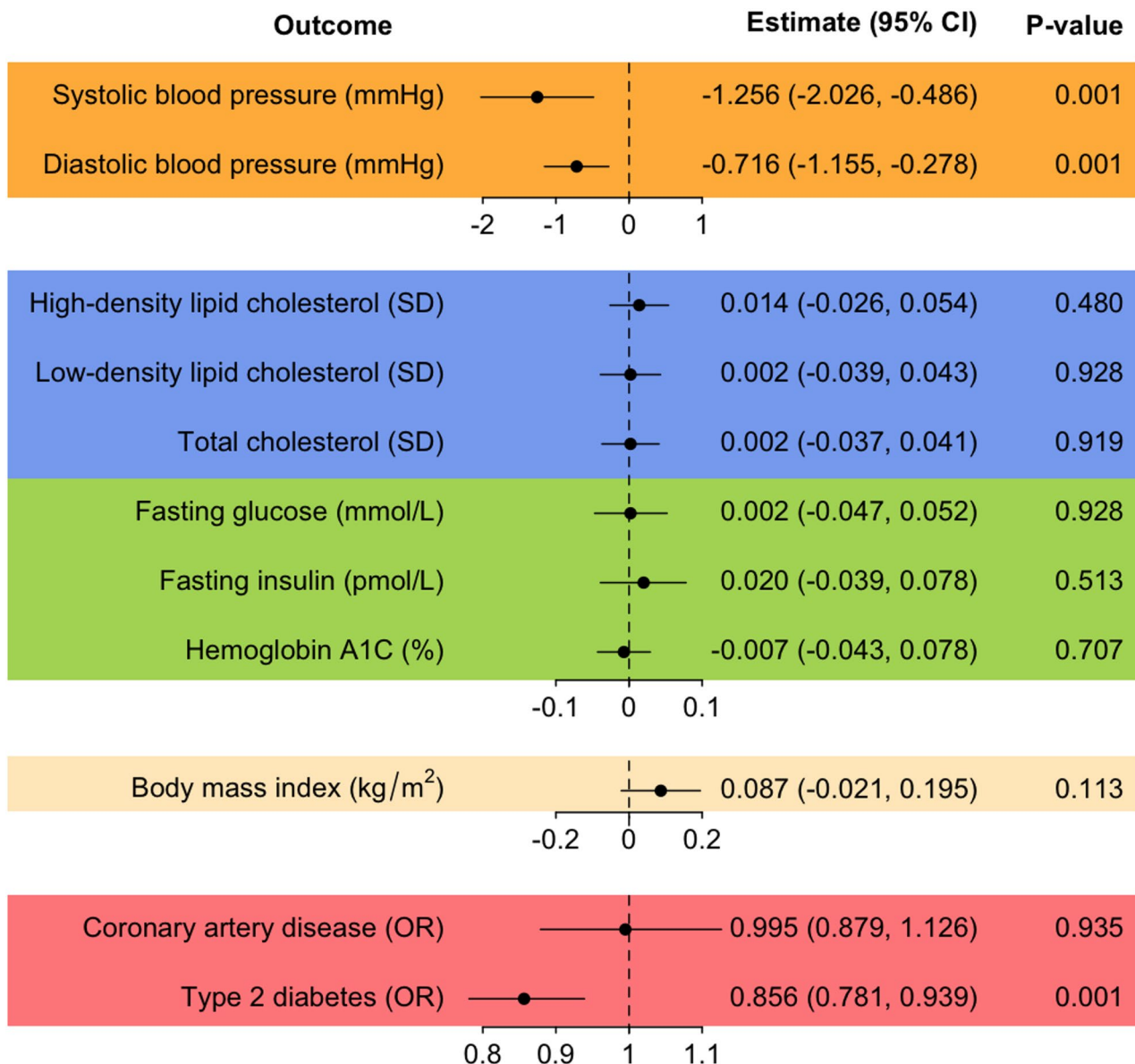


Fig. 4 Causal effects of liking onions on cardiometabolic outcomes estimated using Mendelian randomization. Using *OR2T6* rs6587467 and the Wald estimator, we found that higher genetically predicted liking onions (i.e., one point increase on a 9-point Hedonic scale) was associated with lower systolic blood pressure ($\beta = -1.256$ mmHg, p -value=0.001), diastolic blood pressure ($\beta = -0.716$ mmHg, p -value=0.001), and lower risk of type 2 diabetes (odds ratio [95% confidence interval]=0.856 [0.781, 0.939], p -value=0.001). There was no evidence of a causal relationship between onion intake and serum lipid cholesterol, blood glucose, body mass index, or coronary artery disease. SD, standard deviation; OR, odds ratio. The unit for hemoglobin A1C is expressed as the percentage (%) of total hemoglobin that is glycated

onion perception, preference, and intake. If the observed association between genetically proxied onion liking and cardiometabolic outcomes reflects a causal relationship, plausible mechanisms may involve bioactive compounds in onions, including flavonoids and organosulfur compounds [66]. Small clinical studies have reported hypoglycaemic effects of onion consumption in individuals with diabetes [67], and experimental studies suggest that quercetin, a flavonoid found in onions, may influence pancreatic islet function and insulin secretion [68].

However, these observations should be interpreted cautiously. The present findings are intended to illustrate the feasibility of the instrument-selection framework rather than to support specific dietary recommendations. Replication in independent cohorts and mechanistic investigation will be required to establish causality and clinical relevance.

The present study has limitations. The *TAS1R*, *TAS2R*, and *OR* gene families do not include genes underlying the salty or sour taste modalities, which are mediated

predominantly by ion channels rather than G protein-coupled receptors [69]. While our focus on taste and olfactory receptor genes enabled a systematic and biologically interpretable framework, it likely underestimates the full set of potential instruments, as other genes involved in downstream chemosensory signalling and neural processing may also influence food preferences. Prioritizing food-specific instruments also reduced the number of available IVs, limiting sensitivity analyses; variants associated with multiple foods may alternatively be modelled using multivariable MR [70], although this requires well-characterised associations and careful specification of correlated exposures. Some chemosensory receptors are also expressed outside the oral-nasal cavity and may have additional physiological functions [71], potentially introducing pleiotropic effects. Selection bias is another potential issue, as the UK Biobank has a response rate of ~ 5% and participants differ from the underlying population in socioeconomic status, health, and survival [72]. The 24-hour recall and food-liking subsamples are more restricted, and the same applies to the ALSPAC food-liking subsample. If study participation depends jointly on genotype and phenotype, MR estimates may be biased. Therefore, replication in larger, more diverse, and more representative cohorts will be necessary for validation.

Conclusions

Despite these limitations, our findings demonstrate the potential of chemosensory genetics to improve instrument selection in dietary MR. Chemosensory variants provide biologically grounded IVs that may reduce susceptibility to pleiotropy and reverse causation, although their limited number and modest effect sizes constrain statistical power and generalizability. The instrument for onion liking illustrates the feasibility of this approach, while also highlighting the challenges inherent in applying MR to complex dietary exposures, including modest variance explained, limited instrument availability, and the complexity of eating behaviour. Importantly, causal interpretation of these findings should be supported through triangulation with independent lines of evidence, including experimental and observational studies. Complementary approaches have been used to examine broader sensory-driven dietary patterns (e.g. bitter or pungent foods, such as onion and garlic) [12, 19], as shared chemical properties across foods may be detected by overlapping chemosensory receptor systems. Future work should extend this framework to additional chemosensory pathways and develop analytical methods that jointly model multiple, correlated foods to improve causal estimation of dietary patterns.

Abbreviations

ALSPAC	Avon Longitudinal Study of Parents and Children
BMI	body mass index
GWAS	genome-wide association studies
IV	instrumental variable
LD	linkage disequilibrium
MR	Mendelian randomization
PC	principal component
SNP	single nucleotide polymorphism

Supplementary Information

The online version contains supplementary material available at <https://doi.org/10.1186/s12916-026-04966-x>.

Supplementary Material 1: Additional file 1: Figures S1-S4

Supplementary Material 2: Additional file 2: Tables S1-S12

Acknowledgements

We are extremely grateful to all the families who participated in this study, the midwives for their help recruiting them, and the ALSPAC team, including interviewers, computer and laboratory technicians, clerical workers, research scientists, volunteers, managers, receptionists, and nurses. We thank those who contributed to the survey design and data collection of food preferences and intake from UK Biobank. We thank Phoebe Hui for performing sensitivity analyses using the conventional MR approach.

Author contributions

L.D.H., D.R.R., and P.V.J. designed research; L.D.H. and P.V.J. accessed data; L.D.H. and C.L. led data analysis; D.M.E., N.G.M., D.R.R., and P.V.J. contributed to data analysis; L.D.H. wrote the initial manuscript; all authors critically reviewed the manuscript. L.D.H. had primary responsibility for final content. All authors read and approved the final manuscript.

Funding

This work and L.D.H. were supported by an Australian Research Council Discovery Early Career Researcher Award (DE240100014). D.M.E. was supported by an NHMRC Leadership Fellowship (2017942). P.V.J. is supported by the National Institute of Alcohol Abuse and Alcoholism under award number, Z01AA000135. P.V.J. was supported (in part) by the Intramural Research Program of the NIH, National Institute on Deafness and Other Communication Disorders, the Rockefeller University Heilbrunn Nurse Scholar Award and National Institutes of Health Distinguished Scholar Program. The UK Medical Research Council and Wellcome (Grant ref: MR/Z505924/1) and the University of Bristol provide core support for ALSPAC. Genome-wide genotyping data from the ALSPAC was generated by Sample Logistics and Genotyping Facilities at Wellcome Sanger Institute and LabCorp (Laboratory Corporation of America) using support from 23andMe. A comprehensive list of grants funding is available on the ALSPAC website. (<http://www.bristol.ac.uk/alspac/external/documents/grant-acknowledgements.pdf>); this research was specifically funded by Wellcome Trust and MRC (Grant ref: **102215/2/13/2**). This publication is the work of the authors, and Liang-Dar Hwang will serve as guarantor for the contents of this paper.

Data availability

The UK Biobank data (<https://www.ukbiobank.ac.uk/>) were accessed with application ID 57780. The informed consent obtained from ALSPAC participants does not allow the data to be made available through any third party maintained public repository. Supporting data are available from ALSPAC on request under the approved proposal number, B3544. Full instructions for applying for data access can be found here: <http://www.bristol.ac.uk/alspac/researchers/access/>. GWAS summary statistics results of cardiometabolic traits are available at: <https://www.ebi.ac.uk/gwas/studies/GCST006624> for systolic blood pressure and <https://www.ebi.ac.uk/gwas/studies/GCST006630> for diastolic blood pressure; <https://csg.sph.umich.edu/willer/public/glgc-lipids2021/> for high-density lipid cholesterol, low-density lipid cholesterol, and total lipid cholesterol; <https://www.magicinvestigators.org/> for fasting glucose, fasting insulin, and hemoglobin A1C; https://giant-consortium.web.broadinstitute.org/GIANT_consortium_data_files for BMI; <https://cardiogramplus4d.org/data-downloads/> for coronary artery disease; <https://>

[/www.diagram-consortium.org/downloads.html](http://www.diagram-consortium.org/downloads.html) for type 2 diabetes. GWAS summary statistics results of the Townsend Deprivation Index are available in the OpenGWAS database (<https://gwas.mrcieu.ac.uk/>). GWAS summary statistics results from the Neale lab were accessed at (<http://www.nealelab.is/uk-biobank/>). The analysis pipeline code is available at https://github.com/Cailu086Lin/UKBioBank_Data_Processing_TR_OR_Food_Liking and https://github.com/daniellhwang/ALSPAC_Food_Liking.

Declarations

Ethics approval and consent to participate

UK Biobank received ethical approval from the National Information Governance Board for Health and Social Care and the National Health Service North West Centre for Research Ethics Committee (Ref: 21/NW/0157). Ethical approval for the study was obtained from the ALSPAC Ethics and Law Committee and the Local Research Ethics Committees. Informed consent for the use of all data collected was obtained from participants following the recommendations of the ALSPAC Ethics and Law Committee at the time. Participants can contact the study team at any time to retrospectively withdraw consent for their data to be used. Study participation is voluntary and during all data collection sweeps, information was provided on the intended use of data.

Consent for publication

Not applicable.

Competing interests

L.D.H. is a member of the BMC Medicine Editorial Board. L.D.H. has no role in editorial handling. All other authors declare no competing interests.

Author details

¹Institute for Molecular Bioscience, The University of Queensland, Brisbane, QLD, Australia

²Monell Chemical Senses Center, Philadelphia, PA, USA

³Frazer Institute, The University of Queensland, Brisbane, QLD, Australia

⁴MRC Integrative Epidemiology Unit, University of Bristol, Bristol, UK

⁵QIMR Berghofer Medical Research Institute, Brisbane, QLD, Australia

⁶Section of Sensory Science and Metabolism (SenSMet), National Institute on Alcohol Abuse and Alcoholism & National Institute on Deafness and Other Communication Disorders, National Institutes of Health, Bethesda, MD, USA

Received: 20 February 2026 / Accepted: 27 May 2026

Published online: 01 June 2026

References

1. GBD 2017 Diet Collaborators. Health effects of dietary risks in 195 countries, 1990–2017: a systematic analysis for the Global Burden of Disease Study 2017. *Lancet*. 2019;393(10184):1958–72.
2. Markey W, Srinath H. The alarming rise of early-onset colorectal cancer. *Aust J Gen Pract*. 2025;54(6):392–9.
3. Australian Institute of Health and Welfare. 2025 [cited 2025 Jun 25]. Reports & data. Available from: <https://www.aihw.gov.au/reports-data>
4. Brown AW, Aslibekyan S, Bier D, Ferreira da Silva R, Hoover A, Klurfeld DM, et al. Toward more rigorous and informative nutritional epidemiology: The rational space between dismissal and defense of the status quo. *Crit Rev Food Sci Nutr*. 2023;63(18):3150–67.
5. Eidelman RS, Hollar D, Hebert PR, Lamas GA, Hennekens CH. Randomized trials of vitamin E in the treatment and prevention of cardiovascular disease. *Arch Intern Med*. 2004;164(14):1552–6.
6. Sanderson E, Glymour MM, Holmes MV, Kang H, Morrison J, Munafo MR, et al. Mendelian randomization. *Nat Rev Methods Primers*. 2022;2:6.
7. Evans DM, Davey Smith G. Mendelian Randomization: New Applications in the Coming Age of Hypothesis-Free Causality. *Annu Rev Genomics Hum Genet*. 2015;16:327–50.
8. Sutton KJ, Gervis J, Jatoi M, Hwang LD, Hendricks A, Ghosh D et al. Dietary Intake Mendelian Randomization: Assessment and Development of Methods for Instrument Selection and Robust Inference. medRxiv. 2025;2025.06.26.25330002.
9. Cornelis MC, Munafo MR. Mendelian Randomization Studies of Coffee and Caffeine Consumption. *Nutrients*. 2018;10(10):1343.
10. Zuccolo L, Holmes MV, Commentary. Mendelian randomization-inspired causal inference in the absence of genetic data. *Int J Epidemiol*. 2017;46(3):962–5.
11. Sacerdote C, Guarrera S, Smith GD, Grioni S, Krogh V, Masala G, et al. Lactase persistence and bitter taste response: instrumental variables and mendelian randomization in epidemiologic studies of dietary factors and cancer risk. *Am J Epidemiol*. 2007;166(5):576–81.
12. Zhong VW, Kuang A, Danning RD, Kraft P, van Dam RM, Chasman DI, et al. A genome-wide association study of bitter and sweet beverage consumption. *Hum Mol Genet*. 2019;28(14):2449–57.
13. Hwang LD, Lin C, Gharahkhani P, Cuellar-Partida G, Ong JS, An J, et al. New insight into human sweet taste: a genome-wide association study of the perception and intake of sweet substances. *Am J Clin Nutr*. 2019;109(6):1724–37.
14. Cole JB, Florez JC, Hirschhorn JN. Comprehensive genomic analysis of dietary habits in UK Biobank identifies hundreds of genetic associations. *Nat Commun*. 2020;11(1):1467.
15. Niarchou M, Byrne EM, Trzaskowski M, Sidorenko J, Kemper KE, McGrath JJ, et al. Genome-wide association study of dietary intake in the UK biobank study and its associations with schizophrenia and other traits. *Transl Psychiatry*. 2020;10(1):51.
16. Matoba N, Akiyama M, Ishigaki K, Kanai M, Takahashi A, Momozawa Y, et al. GWAS of 165,084 Japanese individuals identified nine loci associated with dietary habits. *Nat Hum Behav*. 2020;4(3):308–16.
17. Meddens SFW, de Vlaming R, Bowers P, Burik CAP, Linnér RK, Lee C, et al. Genomic analysis of diet composition finds novel loci and associations with health and lifestyle. *Mol Psychiatry*. 2021;26(6):2056–69.
18. Pirastu N, McDonnell C, Grzeszkowiak EJ, Mounier N, Imamura F, Merino J, et al. Using genetic variation to disentangle the complex relationship between food intake and health outcomes. *PLoS Genet*. 2022;18(6):e1010162.
19. May-Wilson S, Matoba N, Wade KH, Hottenga JJ, Concas MP, Mangino M, et al. Large-scale GWAS of food liking reveals genetic determinants and genetic correlations with distinct neurophysiological traits. *Nat Commun*. 2022;13(1):2743.
20. Mompeo O, Freidin MB, Gibson R, Hysi PG, Christofidou P, Segal E, et al. Genome-Wide Association Analysis of Over 170,000 Individuals from the UK Biobank Identifies Seven Loci Associated with Dietary Approaches to Stop Hypertension (DASH) Diet. *Nutrients*. 2022;14(20):4431.
21. Merino J, Dashti HS, Sarnowski C, Lane JM, Todorov PV, Udler MS, et al. Genetic analysis of dietary intake identifies new loci and functional links with metabolic traits. *Nat Hum Behav*. 2022;6(1):155–63.
22. Locke AE, Kahali B, Berndt SI, Justice AE, Pers TH, Day FR, et al. Genetic studies of body mass index yield new insights for obesity biology. *Nature*. 2015;518(7538):197–206.
23. Aragam KG, Jiang T, Goel A, Kanoni S, Wolford BN, Atri DS, et al. Discovery and systematic characterization of risk variants and genes for coronary artery disease in over a million participants. *Nat Genet*. 2022;54(12):1803–15.
24. Bellenguez C, Küçükali F, Jansen IE, Kleindam L, Moreno-Grau S, Amin N, et al. New insights into the genetic etiology of Alzheimer's disease and related dementias. *Nat Genet*. 2022;54(4):412–36.
25. Kim Ukyung, Jorgenson E, Coon H, Leppert M, Risch N, Drayna D. Positional cloning of the human quantitative trait locus underlying taste sensitivity to phenylthiocarbamide. *Science*. 2003;299(5610):1221–5.
26. Reed DR, Zhu G, Breslin PAS, Duke FF, Henders AK, Campbell MJ, et al. The perception of quinine taste intensity is associated with common genetic variants in a bitter receptor cluster on chromosome 12. *Hum Mol Genet*. 2010;19(21):4278–85.
27. Knaapila A, Hwang LD, Lysenko A, Duke FF, Fesi B, Khoshnevisan A, et al. Genetic analysis of chemosensory traits in human twins. *Chem Senses*. 2012;37(9):869–81.
28. Ledda M, Kutalik Z, Souza Destito MC, Souza MM, Cirillo CA, Zamboni A, et al. GWAS of human bitter taste perception identifies new loci and reveals additional complexity of bitter taste genetics. *Hum Mol Genet*. 2014;23(1):259–67.
29. Hwang LD, Gharahkhani P, Breslin PAS, Gordon SD, Zhu G, Martin NG, et al. Bivariate genome-wide association analysis strengthens the role of bitter receptor clusters on chromosomes 7 and 12 in human bitter taste. *BMC Genomics*. 2018;19(1):678.
30. Kan M, Saraiva LR, Hwang LD, Lowenthal ED, Himes BE, Mennella JA. Genome-wide association study of the taste and hedonic ratings of the low-calorie sweetener acesulfame potassium. *Sci Rep*. 2025;15(1):22121.

31. Ong JS, Hwang LD, Zhong VW, An J, Gharahkhani P, Breslin PAS, et al. Understanding the role of bitter taste perception in coffee, tea and alcohol consumption through Mendelian randomization. *Sci Rep.* 2018;8(1):16414.
32. Bobowski N, Reed DR, Mennella JA. Variation in the TAS2R31 bitter taste receptor gene relates to liking for the nonnutritive sweetener Acesulfame-K among children and adults. *Sci Rep.* 2016;6:39135.
33. Feeney EL, McGuinness L, Hayes JE, Nolden AA. Genetic variation in sensation affects food liking and intake. *Curr Opin Food Sci.* 2021;42:203–14.
34. Brito Nunes C, Lim AWY, Dy Q, Ong JS, Hwang LD. Phenome-wide investigation of dietary and health outcomes associated with bitter taste receptor gene TAS2R38. *Eur J Nutr.* 2025;64(5):218.
35. Daly AN, O'Sullivan EJ, Kearney JM. Considerations for health and food choice in adolescents. *Proc Nutr Soc.* 2022;81(1):75–86.
36. Bycroft C, Freeman C, Petkova D, Band G, Elliott LT, Sharp K, et al. The UK Biobank resource with deep phenotyping and genomic data. *Nature.* 2018;562(7726):203–9.
37. Boyd A, Golding J, Macleod J, Lawlor DA, Fraser A, Henderson J, et al. Cohort Profile: the 'children of the 90s'—the index offspring of the Avon Longitudinal Study of Parents and Children. *Int J Epidemiol.* 2013;42(1):111–27.
38. Fraser A, Macdonald-Wallis C, Tilling K, Boyd A, Golding J, Davey Smith G, et al. Cohort Profile: the Avon Longitudinal Study of Parents and Children: ALSPAC mothers cohort. *Int J Epidemiol.* 2013;42(1):97–110.
39. Jin Y, Schaffer AA, Feolo M, Holmes JB, Kattman BL. GRAF-pop: A Fast Distance-Based Method To Infer Subject Ancestry from Multiple Genotype Datasets Without Principal Components Analysis. *G3 (Bethesda).* 2019;9(8):2447–61.
40. Manichaikul A, Mychaleckyj JC, Rich SS, Daly K, Sale M, Chen WM. Robust relationship inference in genome-wide association studies. *Bioinformatics.* 2010;26(22):2867–73.
41. Chang CC, Chow CC, Tellier LC, Vattikuti S, Purcell SM, Lee JJ. Second-generation PLINK: rising to the challenge of larger and richer datasets. *Gigascience.* 2015;4:7.
42. Harrow J, Frankish A, Gonzalez JM, Tapanari E, Diekhans M, Kokocinski F, et al. GENCODE: the reference human genome annotation for The ENCODE Project. *Genome Res.* 2012;22(9):1760–74.
43. Ward LD, Kellis M. HaploReg v4: systematic mining of putative causal variants, cell types, regulators and target genes for human complex traits and disease. *Nucleic Acids Res.* 2016;44(D1):D877–881.
44. Smigielski EM, Sirotkin K, Ward M, Sherry ST. dbSNP: a database of single nucleotide polymorphisms. *Nucleic Acids Res.* 2000;28(1):352–5.
45. Bradbury KE, Young HJ, Guo W, Key TJ. Dietary assessment in UK Biobank: an evaluation of the performance of the touchscreen dietary questionnaire. *J Nutr Sci.* 2018;7:e6.
46. Harris PA, Taylor R, Thielke R, Payne J, Gonzalez N, Conde JG. Research electronic data capture (REDCap)—a metadata-driven methodology and workflow process for providing translational research informatics support. *J Biomed Inf.* 2009;42(2):377–81.
47. Lyon MS, Andrews SJ, Elsworth B, Gaunt TR, Hemani G, Marcora E. The variant call format provides efficient and robust storage of GWAS summary statistics. *Genome Biol.* 2021;22(1):32.
48. Hui PS, Zhang J, Hwang LD. Genetic influences on food liking and food preference patterns in young adults: a genome-wide association study. *bioRxiv.* 2026;2026.03.25.714302.
49. Machiela MJ, Chanock SJ. LDlink: a web-based application for exploring population-specific haplotype structure and linking correlated alleles of possible functional variants. *Bioinformatics.* 2015;31(21):3555–7.
50. Evangelou E, Warren HR, Mosen-Ansorena D, Mifsud B, Pazoki R, Gao H, et al. Genetic analysis of over 1 million people identifies 535 new loci associated with blood pressure traits. *Nat Genet.* 2018;50(10):1412–25.
51. Graham SE, Clarke SL, Wu KHH, Kanoni S, Zajac GJM, Ramdas S, et al. The power of genetic diversity in genome-wide association studies of lipids. *Nature.* 2021;600(7890):675–9.
52. Chen J, Spracklen CN, Marenne G, Varshney A, Corbin LJ, Luan J, et al. The trans-ancestral genomic architecture of glycemic traits. *Nat Genet.* 2021;53(6):840–60.
53. Suzuki K, Hatzikotoulas K, Southam L, Taylor HJ, Yin X, Lorenz KM, et al. Genetic drivers of heterogeneity in type 2 diabetes pathophysiology. *Nature.* 2024;627(8003):347–57.
54. Burgess S, Butterworth A, Thompson SG. Mendelian randomization analysis with multiple genetic variants using summarized data. *Genet Epidemiol.* 2013;37(7):658–65.
55. Hemani G, Tilling K, Davey Smith G. Orienting the causal relationship between imprecisely measured traits using GWAS summary data. *PLoS Genet.* 2017;13(11):e1007081.
56. Bowden J, Davey Smith G, Burgess S. Mendelian randomization with invalid instruments: effect estimation and bias detection through Egger regression. *Int J Epidemiol.* 2015;44(2):512–25.
57. Bowden J, Davey Smith G, Haycock PC, Burgess S. *Genet Epidemiol.* 2016;40(4):29686387.
58. Hermani G, Elsworth B, Palmer T, Rasteiro R. ieugwas: Interface to the OpenGWAS Database API. Available from: <https://github.com/MRCIEU/ieugwas>
59. Burgess S, Woolf B, Mason AM, Ala-Korpela M, Gill D. Addressing the credibility crisis in Mendelian randomization. *BMC Med.* 2024;22(1):374.
60. Tong T, Wang Y, Kang SG, Huang K. Ectopic Odorant Receptor Responding to Flavor Compounds: Versatile Roles in Health and Disease. *Pharmaceutics.* 2021;13(8):1314.
61. Osakabe N, Shimizu T, Fujii Y, Fushimi T, Calabrese V. Sensory Nutrition and Bitterness and Astringency of Polyphenols. *Biomolecules.* 2024;14(2):234.
62. Bachmanov AA, Beauchamp GK. Taste receptor genes. *Annu Rev Nutr.* 2007;27:389–414.
63. Barwich AS. What Makes a Discovery Successful? The Story of Linda Buck and the Olfactory Receptors. *Cell.* 2020;181(4):749–53.
64. Uhlén M, Fagerberg L, Hallström BM, Lindskog C, Oksvold P, Mardinoglu A, et al. Proteomics. Tissue-based map of the human proteome. *Science.* 2015;347(6220):1260419.
65. Durante MA, Kurtenbach S, Sargi ZB, Harbour JW, Choi R, Kurtenbach S, et al. Single-cell analysis of olfactory neurogenesis and differentiation in adult humans. *Nat Neurosci.* 2020;23(3):323–6.
66. Zhao XX, Lin FJ, Li H, Li HB, Wu DT, Geng F, et al. Recent Advances in Bioactive Compounds, Health Functions, and Safety Concerns of Onion (*Allium cepa* L). *Front Nutr.* 2021;8:669805.
67. Taj Eldin IM, Ahmed EM, Elwahab HMA. Preliminary Study of the Clinical Hypoglycemic Effects of *Allium cepa* (Red Onion) in Type 1 and Type 2 Diabetic Patients. *Environ Health Insights.* 2010;4:71–7.
68. Vessal M, Hemmati M, Vasei M. Antidiabetic effects of quercetin in streptozocin-induced diabetic rats. *Comp Biochem Physiol C Toxicol Pharmacol.* 2003;135 C(3):357–64.
69. Taruno A, Nomura K, Kusakizako T, Ma Z, Nureki O, Foskett JK. Taste transduction and channel synapses in taste buds. *Pflugers Arch.* 2021;473(1):3–13.
70. Burgess S, Thompson SG. Multivariable Mendelian Randomization: The Use of Pleiotropic Genetic Variants to Estimate Causal Effects. *Am J Epidemiol.* 2015;181(4):251–60.
71. Dalesio NM, Barreto Ortiz SF, Pluznick JL, Berkowitz DE. Olfactory, Taste, and Photo Sensory Receptors in Non-sensory Organs: It Just Makes Sense. *Front Physiol.* 2018;9:1673.
72. Schoeler T, Speed D, Porcu E, Pirastu N, Pingault JB, Kutalik Z. Participation bias in the UK Biobank distorts genetic associations and downstream analyses. *Nat Hum Behav.* 2023;7(7):1216–27.

Publisher's note

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