

Dietary preference and susceptibility to type 2 diabetes mellitus: a wide-angle Mendelian randomization study

Mia D. Lee¹, Benjamin F. Voight^{2,3,4,*}

¹ Pharmacology Graduate Group, University of Pennsylvania – Perelman School of Medicine, Philadelphia PA

² Department of Systems Pharmacology and Translational Therapeutics, University of Pennsylvania – Perelman School of Medicine, Philadelphia PA

³ Department of Genetics, University of Pennsylvania – Perelman School of Medicine, Philadelphia PA

⁴ Institute of Translational Medicine and Therapeutics, University of Pennsylvania – Perelman School of Medicine, Philadelphia PA

*Correspondence to:

Benjamin F. Voight, PhD

University of Pennsylvania Perelman School of Medicine

Smilow Center for Translational Research, Room 10-126

3400 Civic Center Boulevard

Philadelphia, PA 19104

bvoight@upenn.edu

Keywords: Type 2 diabetes mellitus, diet, Mendelian randomization, obesity, education

ABSTRACT

1 **Background:** Susceptibility to type 2 diabetes mellitus (T2D) is driven by genetic and
2 environmental risk factors. Dietary preferences are a modifiable and largely environmental risk
3 factor for T2D. The role of diet in disease liability has been limited to observational and
4 epidemiologic studies with mixed findings.

5 **Objective:** To clarify the role of diet on susceptibility to T2D using genetic variants associated
6 dietary preferences.

7 **Methods:** We collected genome-wide association data for 38 dietary preference traits plus T2D
8 and 21 related cardiometabolic traits. We performed Mendelian randomization (MR) using
9 genetic variants to test causal hypotheses between diet as the exposure and T2D or
10 cardiometabolic traits as outcomes using univariable and multivariable methods along with the
11 MR Robust Adjusted Profile Score (MR-RAPS) approach to increase power. We performed
12 mediation analyses to evaluate the effects of dietary preferences on T2D to elucidate potential
13 causal graphs and estimate the effects of dietary preferences mediated by potential mediators.

14 **Results:** We report 17 significant relationships between dietary preferences and T2D or a
15 cardiometabolic risk factor (Bonferroni-corrected $P < 5.99 \times 10^{-5}$), including that higher intake of
16 cheese, dried fruit, muesli, or fat-based spreads protected against T2D. We detected 7 additional
17 associations (Bonferroni-corrected $P < 1 \times 10^{-4}$), with inclusion of additional genetic variants in
18 MR-RAPS analysis. In multivariable MR, we discovered that body mass index (BMI) was a
19 common, shared mediator for many of these observed associations. In mediation analysis, we
20 confirmed that substantial proportions of the protective effects of cheese, dried fruit and muesli
21 intakes on T2D were mediated by BMI. We further observed that educational attainment was an
22 additional mediator exclusively for muesli intake-T2D association.

23 **Conclusions:** Our results provide genetic evidence supporting a link between diet and body
24 weight, and are in line with observation of obesity and T2D in individuals and their specific
25 preferences for food.

26 **ABBREVIATIONS**

27 ALP, Alkaline phosphatase; ALT, Alanine transaminase; ASAT, Abdominal subcutaneous adipose
28 tissue; AST, Aspartate aminotransferase; BMI, Body mass index; CORNFLAK, Corn
29 flakes/frosties; DBP, Diastolic blood pressure; DRIEFRU, Dried fruit; EA, Educational
30 attainment; FG, Fasting glucose adjusted for BMI; FI, Fasting insulin adjusted for BMI; GGT,
31 Gamma-glutamyl transferase; GWAS - Genome-wide association study; HbA1c, Hemoglobin
32 A1c; HDL, High-density lipoprotein; INFO, Information score (quality measurement of
33 imputation); IV, Instrumental variables; IVW, Inverse variance weighted; LDL, Low-density
34 lipoprotein; LDSC, Linkage disequilibrium score regression; Liverfat, Liver fat content;
35 Liveriron, Liver iron level; Livervol, Liver volume; MAF, Minor allele frequency; MR,
36 Mendelian Randomization; MR-RAPS, Mendelian Randomization Robust Adjusted Profile
37 Score; MVMR, Multivariable Mendelian Randomization; PAL, Highly palatable; Pancfat,
38 Pancreas fat content; Panciron, Pancreas iron level; Pancvol, Pancreas volume; PHY, Physical
39 activity; SAVCAL, Savory/ caloric; SBP, Systolic blood pressure; SED, Sedentary behavior at
40 work; STR, Strong flavored; T2D, Type 2 diabetes mellitus; TG, Triglyceride; UVMR,
41 Univariable Mendelian Randomization; VAT, Visceral adipose tissue; WHR, Wait-hip ratio;
42 WHRadjBMI, Wait-hip ratio adjusted for BMI; WM, Weighted median

43

44 **INTRODUCTION**

45 Type 2 diabetes mellitus (T2D) is a complex metabolic disease with individual risk influenced by
46 both genetic and environmental factors [1,2]. The growing incidence of T2D in parallel with the
47 obesity epidemic suggests the essential role of environment to disease risk [3]. While genome-
48 wide association studies (GWAS) have elucidated the polygenic nature of the disease with
49 discovery of hundreds of associated genetic loci [4,5], the spectrum of the environmental
50 components that cause T2D is poorly understood. Improvements in this understanding could
51 facilitate personalized interventions for patients who are at the most risk of cardiovascular,
52 neurological and renal complications of diabetes based on these factors [6].

53 Despite observational data supporting the role of diet as a major but modifiable risk factor
54 associated with T2D [7,8], a causal role of dietary factors and cardiometabolic disease is far less
55 established with only a handful of examples in evidence [9,10]. Previous work supporting strong
56 association between unhealthy diet and T2D has been largely accrued through observational
57 studies in which a formal assessment of causality is difficult if not impossible. Moreover,
58 previous reports on the effects of diet on susceptibility to T2D have been inconsistent and
59 inconclusive [11-13], compounding the challenge of establishing a robust assessment of
60 causality. Thus, a key challenge is to elucidate how specific dietary risk factors – or potentially
61 how established risk factors like obesity – mediate or cause diabetes, and subsequently how to
62 use these insights to action downstream interventional studies.

63 While non-genetic contributions to dietary preference are large and central, some food intake
64 patterns have been observed to be heritable [14,15] and interact with T2D-relevant genes [16],
65 suggesting that genetics may contribute a small component to interindividual variability in
66 dietary preferences. Recently, GWAS have reported genetic associations with macronutrient
67 intake, dietary habits, and food preferences [17-19]. Excitingly, these data offer the opportunity
68 to examine evidence of causal relationships between diet and cardiometabolic outcomes using
69 statistical approaches.

70 In this report, we utilized the framework of Mendelian randomization (MR) using genetic
71 summary data related to diet and 42 dietary preferences to understand the role of these factors for
72 T2D susceptibility and other cardiometabolic traits (**Figure 1**). MR is one statistical approach
73 designed to infer causal effects between an exposure and an outcome. This method leverages
74 genetic variants associated with an exposure (here, dietary preference) as instrumental variables

75 (IVs) to estimate the effect of that exposure on an outcome (e.g., liability to T2D) [20]. MR has
76 been used to confirm or refute a range of risk factors for T2D [22-23], including other diseases
77 [24-26], as well as using dietary instrumental variables as exposures [27]. We developed
78 instruments for 42 dietary preferences as exposures and performed univariable and multivariable
79 experiments using T2D and 21 cardiometabolic risk factors as outcomes.

80 81 **METHODS**

82 *Acquisition of dietary trait summary genetic data*

83 For dietary traits, we selected 42 dietary preferences, which were composed of: (i) 4 relative
84 macronutrient intake measures, (ii) 26 dietary habits, and (iii) 12 food-liking traits. The summary
85 statistic association data for these dietary traits were obtained from 3 studies on diet composition
86 [17], dietary habits [18] and food liking [19] (**Supplementary Table 1**). These studies were
87 conducted mainly on groups of individuals similar to the 1000 Genomes Project Continental
88 group from Europe as a reference.

89 Based on 24h dietary food recall questionnaire, which contains ≥ 70 food items assessed in UK
90 Biobank participants, the summary association data for diet composition, which was adjusted for
91 sex, year of birth, and the first 10 genetic principal components, identified genetic associations
92 with relative intake of carbohydrate, sugar, fat and protein [17]. The carbohydrate intake
93 included a consumption of all saccharides, while the sugar intake mono- and disaccharides only.
94 The measurements for 4 macronutrient intakes were corrected for total energy intake. The energy
95 values of the macronutrients were obtained using the conversion factor of 4 kcal/gram for
96 protein, sugar and carbohydrate, and 9 kcal/gram for fat.

97 The summary association data for dietary habits, in which covariates were age in months, sex
98 and 10 genetic principal components, was based on UK Biobank food frequency questionnaire
99 (FFQ) on 85 single food items on daily and/or weekly basis [18]. The measurement of the intake
100 of each food item was in either quantitative continuous variables (e.g., cups per day, pieces per
101 day, bowls per week) or ordinal non-quantitative variables (e.g., number of intakes per day or
102 week). The GWAS for dietary habits also estimated SNP heritability (h^2_g) of those single food
103 intake using BOLT-lmm software (v.2.3.2) pseudo-heritability measurement representing the
104 fraction of phenotypic variance explained by the estimated relatedness matrix [18]. We selected
105 26 specific dietary habits that were in the top 25 percentile of the observed SNP heritability (h^2_g
106 > 0.028).

107 The summary association data obtained for food liking conducted an assessment for likings for
108 139 specific foods and beverages using the food preferences/liking questionnaire. This
109 questionnaire uses the 9-point hedonic scale, where 1 corresponds to “Extremely dislike” and 9
110 to “Extremely like”. The GWAS for food liking included age, sex and the first 10 genetic
111 principal components as covariates for the analysis. Because these foods and beverages are not
112 generalized and overlap with some of the dietary habits, we selected 3 main dimensions and 9
113 sub-dimensions of food likings that were categorized by taste and food type in the multi-level
114 hierarchical manner [19].

115 *Acquisition of T2D and related traits summary genetic data*

116 For T2D, we obtained the latest summary statistics from Million Veteran Program (MVP) study
117 [4], which included age, sex and ten genetic principal components as covariates. We selected the
118 following 21 cardiometabolic risk factors related to T2D: Body mass index (BMI) [28], Waist-

119 hip ratio (WHR) [28], Waist-hip ratio adjusted for BMI (WHRadjBMI) [28], Fasting glucose
120 adjusted for BMI (FG) [29], Fasting insulin adjusted for BMI (FI) [29], Glycated hemoglobin
121 (HbA1c) [29], High-density lipoprotein cholesterol (HDL) [30], Low-density lipoprotein
122 cholesterol (LDL) [30], Triglyceride (TG) [30], Abdominal subcutaneous adipose tissue volume
123 (ASAT) [31], Visceral adipose tissue volume (VAT) [31], Pancreas fat (Pancfat) [34128465],
124 Pancreas iron content (Panciron) [31], Pancreas volume (Pancvol) [31], Liver fat (Liverfat) [31],
125 Liver iron content (Liveriron) [31], Liver volume (Livervol) [31], Alkaline phosphatase (ALP)
126 [32], Alanine aminotransferase (ALT) [32], Aspartate transaminase (AST) [33], and Gamma-
127 glutamyl Transferase (GGT) [32]. For reasons of statistical power, we focused primarily on
128 European ancestry data, which had the largest sample sizes. The locations, sample sizes and
129 other information for these summary data are detailed in **Supplementary Table 2**.

130 ***GWAS data processing and quality control***

131 Prior to selection of genetic instruments for dietary traits, we first removed rare variants (MAF <
132 0.01) and, if applicable, variants with low imputation quality (INFO < 0.5) in the summary-level
133 data sets for dietary traits (**Supplementary Figure 2**). As additional quality control, we also
134 eliminated duplicate variants. Finally, we obtained consensus rsIDs for all variants (dbSNP build
135 155) and aligned them to GRCh37 for all traits (see **Code Availability**).

136 ***Selection of variants for genetic instruments for dietary traits***

137 We used PLINK (v1.9 software) [34] and European 1000 Genomes Project Phase 3 data (EUR,
138 GRCh37) for LD reference to group loci together using the ‘LD clumping’ procedure to select
139 genome-wide significant ($P < 5 \times 10^{-8}$) and independent genetic instruments (instrumental
140 variables) with low pairwise linkage with one another ($r^2 < 0.001$) within a window of 500kb

141 [21,35] (Supp. Figure 2). After clumping, we excluded four dietary traits from the MR analysis
142 because they had relatively few IVs contributing (less than $n = 5$): Flora spread intake, low-fat
143 spread intake, low-fat milk intake and liking for healthy foods.

144 We removed palindromic (A/T or C/G) genetic variants that are present in the dietary traits to
145 ensure robustness of our instruments (**Supplementary Table 3**). The genetic variants that were
146 present in the dietary traits but not the outcomes of interest (e.g., T2D and related traits) were
147 removed from MR analysis and not replaced with proxy variants. With this approach, we
148 identified 931 genetic instruments (803 unique SNPs) in total across 38 dietary traits. The final
149 genetic instruments used for MR are shown in the **Supplementary Tables 14 and 15**, and the
150 SNP selection process is described in **Supplementary Figure 1**.

151 For each genetic instrument, we estimated R^2 , the proportion of variance explained of exposure
152 by created IV, assuming a strictly additive model [36]. For a single SNP R^2 given by: $2 \times p \times (1$
153 $- p) \times \beta^2$, where p is the minor allele frequency and β is the estimated effect size of the SNP for
154 the exposure trait [36]. The total variance explained then is simply the sum across all SNPs,
155 R^2_{total} . Once we obtained R^2_{total} , for each exposure, we calculated the F-statistic (F) to assess the
156 strength of the instrumental variables. The formula for F is $(N - K - 1)/K \times R^2_{\text{total}} / (1 - R^2_{\text{total}})$,
157 where N is the sample size, and K is the number of SNPs used for the IV [37]. All of our
158 generated instruments exceeded an $F > 10$, suggesting minimal effects from weak instrument
159 bias (**Supplementary Table 4**).

160 We used the LD-clumping procedure described above using 1000 Genomes Project Phase 3
161 (EUR) as the LD reference to identify the best LD proxy ($r^2 \geq 0.8$ with given index variants) for
162 index variants that were present with dietary genetic instruments but not present in the summary
163 association trait data used in multivariable MR (MVMR) analyses.

164 For the analysis using MR-RAPS (Robust Adjusted Profile Score) [38], we used the same
165 approach described above (LD-clumping via PLINK, [34]), and applied the same parameters to
166 obtain independent, sub-genome wide ($P < 1 \times 10^{-4}$) significant genetic variants as additional
167 instruments. As described above for the primary MR analyses, we removed palindromic genetic
168 variants in the MR-RAPS analysis.

169 *Mendelian randomization methods*

170 MR is a statistical causal inference analysis that employs genetic variants associated with an
171 exposure (here, dietary preference) as instrumental variables (IVs) to estimate the effect of that
172 exposure on an outcome (e.g., liability to T2D) [20]. Specific assumptions for instrumental
173 variables are required, namely that: (i, Relevance assumption) genetic variants are associated
174 with the exposure, (ii, Independence assumption) genetic variants are not associated with any
175 potential confounders in the exposure-outcome association, and that (iii, Exclusion restriction
176 assumption) genetic variants can influence the outcome only through the exposure
177 (Supplementary Figure 1). Given Mendel's law of segregation and independent assortment at
178 meiosis, the design of an MR experiment is analogous to the design of the randomized control
179 trial via random allocation of genetic variants [21] and can address issues of confounding and
180 reverse causality in assessment of causality.

181 We populated STROBE-MR checklist as a guideline to conduct all our MR analyses and report
182 the results clearly (**Supplementary Data 1**).

183 For primary analysis, we performed two-sample univariable MR (UVMR) using summary
184 statistic association data from 38 dietary traits as exposures and from T2D plus 21 T2D-related
185 cardiometabolic traits as outcomes. Though partial sample overlap exists between dietary traits

186 and some outcome traits, the bias due to the observed sample overlap were estimated to be small
187 in the analysis, especially when the strength of genetic instruments estimated by F-statistic (F) is
188 strong ($F > 10$) and the datasets are of different sample size [39] (**Supplementary Table 21**).
189 Using TwoSampleMR package (v.0.5.8) [40,41], we employed three standard UVMR methods to
190 estimate causal effects: Inverse variance weighted under the random effects model (IVW),
191 Weighted median (WM), and Egger regression. Each method has different sensitivity for the IV
192 assumptions. IVW has the greatest statistical power because it assumes all variants do not violate
193 any IV assumptions; WM assumes at least half of IVs are valid instruments, providing robustness
194 to Assumptions 2 and 3; Egger regression estimates the causal effect after accounting for
195 horizontal pleiotropy in IVs, providing robustness to Assumption 3 [42]. We used the Egger bias
196 intercept test as a quantitative indicator of bias due to horizontal pleiotropy (**Supplementary**
197 **Table 5**). After exclusion for instrument size (see above) and checking the presence of exposure
198 IVs in the endpoint data, we performed UVMR that tested a total of 835 exposure-outcome
199 pairwise associations (Supplementary Table 5). We also performed MR-Steiger test [41] to
200 examine whether assumption that exposure causes outcome is valid and determine the
201 directionality of those associations by using the same input data that were used to estimate
202 effects via the 3 standard UVMR methods.

203 To address the burden of multiple testing, we considered Bonferroni correction to determine the
204 adjusted nominal p-value threshold for significance ($n=835$ tests with a 5% error rate: $P < 5.99 \times$
205 10^{-5}). We considered Bonferroni-corrected exposure-outcome associations further if two
206 additional criteria were met: (i) they were associated across at least two MR analyses, and (ii) at
207 least 5 IVs were used in the sensitivity analyses.

208 *Analysis using MR-RAPS (Robust Adjusted Profile Score)*

209 17 nominally but not multi-test corrected significant (i.e., MR $P < 1 \times 10^{-3}$ but $P > 5.99 \times 10^{-5}$)
210 exposure-outcome associations across at least two sensitivity analyses were further analyzed
211 using the MR-Robust Adjusted Profile Score (MR-RAPS) method. MR-RAPS minimizes
212 measurement error in SNP-exposure effects due to weak instruments bias and is robust to
213 systematic pleiotropy [38]. In addition to genome-wide significant variants ($P < 5 \times 10^{-8}$), sub-
214 genome wide significant ($P < 1 \times 10^{-4}$) genetic variants as weak instruments were included to
215 increase the power of MR for causal inference and re-assess those associations via this method.
216 For statistical stringency, we applied the following criteria to determine whether the associations
217 are significant and additional associations were identified by increased power from MR-RAPS:
218 (i) when estimated using the combined set of both weak ($P < 1 \times 10^{-4}$) and strong ($P < 5 \times 10^{-8}$)
219 instruments, the association exceeds the threshold p-value that was applied for the primary
220 UVMR analysis (e.g., $P < 5.99 \times 10^{-5}$); (ii) when estimated using only weak ($P < 1 \times 10^{-4}$)
221 instruments, the association exceeds the threshold p-value that was applied for the primary
222 UVMR analysis ($P < 5.99 \times 10^{-5}$); and (iii) the effect direction of the association that meets those
223 two criteria above was concordant with that observed from the initial UVMR methods described
224 above.

225 *Multivariable MR experimental methods*

226 To further evaluate exposure-outcome associations that emerged as significant from UVMR, we
227 used MVMR package (v.0.4) [43] along with the TwoSampleMR package (v.0.5.8) and
228 performed multivariable MR (MVMR) using IVW to estimate the direct effects of dietary
229 preferences by accounting for the indirect effects mediated through selected potential mediators
230 (**Supplementary Figure 3**). Since T2D is a multi-factorial, complex disease, we relied on
231 biological prior and literature-based observations in T2D to narrow the list of potential traits that

232 could possibly mediate the causal effects of exposures on outcomes. Then, we conducted
233 Linkage Disequilibrium Score Regression (LDSC) analysis [44] on those traits to ensure that
234 they are genetically correlated with T2D and dietary exposures of interest (**Supplementary**
235 **Tables 17-18**). Based on these factors, we selected 8 potential mediators: Body mass index
236 (BMI) [28], Fasting insulin (FI) [29], Waist-hip ratio (WHR) [28], Diastolic blood pressure
237 (DBP) [45], Systolic blood pressure (SBP) [45], Physical activity (PHY) [46] and Sedentary
238 behavior at work (SED) [46], and Educational Attainment (EA, Supplementary Table 2) [47].
239 Our MVMR analysis consisted of two parts: single-mediator MVMR and multiple-mediator
240 MVMR. For both single-mediator and multiple-mediator MVMR, the genetic instruments
241 selected for potential mediators and outcomes were based on the genetic variants that were
242 considered strong, viable genetic instruments for the primary exposures (here, dietary
243 preferences), namely exposure-centric instruments. In other words, all the exposure-centric
244 instruments were the same instruments utilized for UVMR.

245 In single-mediator MVMR, each of the 8 selected potential mediators was applied in a pairwise
246 manner to estimate the direct effects of the primary exposures on T2D and related
247 cardiometabolic traits as outcomes. In the multiple-mediator MVMR, at least 2 potential
248 mediators were included when assessing the effect of the primary exposures on the outcomes.

249 ***Mediation analysis***

250 We employed a two-step MR approach to conduct mediation analysis, which facilitates
251 estimation of the indirect effect of a primary exposure on an outcome via a mediating trait and,
252 consequently, the proportion mediated. For this analysis, the two steps are: (i) estimation of the

253 effect of a primary exposure on a mediating trait, and (ii) estimation of the direct effect of a
254 mediating trait on an outcome, after adjusting for a primary exposure [48,49].

255 We applied this approach to 3 exposure-outcome pairs of our main interest which were
256 associations of cheese, muesli and dried fruits intake with T2D. Spreads intake association with
257 T2D was excluded due to a very small number of genetic instruments available for spreads
258 intake. For mediation analysis, BMI, DBP and EA were the mediating traits of interest because
259 they attenuated those 3 associations in single-mediator MVMR.

260 In the first step, we used “exposure-centric” instruments and obtained the effects of the 3 dietary
261 preferences on BMI, DBP and EA individually via UVMR. In the second step, we estimated the
262 direct effects of BMI, DBP and EA on T2D, accounting the effects of the dietary preferences, via
263 MVMR. Because the 3 mediators were considered the primary exposures in the second step, we
264 utilized all the genetic variants that serve as strong genetic instruments for each mediator, namely
265 mediator-centric instruments, to obtain estimates of the second step with increased statistical
266 power. We chose IVW as a main method for both steps.

267 We then employed product of coefficients method to multiply the estimates (i.e., the effect sizes
268 in beta coefficients) from the two steps and obtained the indirect effects of the 3 dietary
269 preferences on T2D. These indirect effects were divided by the total effects of the dietary
270 preferences on T2D estimated in the primary UVMR to calculate the proportions mediated by
271 BMI, DBP and EA, respectively.

272 Finally, we used RMediation package (v. 1.2.2) [50] to estimate indirect effects and approximate
273 standard errors and 95% confidence intervals, using 3 different methods: Monte Carlo,
274 distribution of product and asymptotic distribution (i.e., delta). The package uses the effect sizes

275 in beta coefficients and standard errors of the two steps including a parameter ρ , which is a
276 correlation ($-1 < \rho < 1$) between the two steps which are exposure-to-mediator and mediator-
277 to-outcome pathways. We assumed the default value of 0 for this parameter for the 3 methods
278 and they yielded similar confidence intervals for indirect effects. The indirect effect estimates
279 from these 3 methods are comparable to the ones computed via product of coefficients method.
280 Of the 3 methods, we relied on the estimates mainly from delta method, which is relatively more
281 suitable for summary-level data in two-sample MR [51,52].

282

283 **RESULTS**

284 *Univariable MR detected 17 associations of dietary preferences with T2D and cardiometabolic*
285 *risk factors.*

286 To examine the relationship between diet and T2D, we first performed two-sample MR on 38
287 dietary traits as exposures and T2D plus 21 additional cardiometabolic risk factors as outcomes
288 (**Methods**, Supplementary Tables 1-2). **Supplementary Table 3** shows the number of genetic
289 instruments our for 38 dietary traits, and we did not observe evidence of potential weak
290 instrument bias from our generated genetic IVs as calculated by F -statistics (**Supplementary**
291 **Table 4**). Of the 835 exposure-outcome pairs we tested, we considered only those significant
292 after Bonferroni multiple test correction (nominal $P < 5.99 \times 10^{-5}$, **Methods**). Of 835
293 associations, we excluded associations where less than 5 genetic instruments were used in the
294 analyses. We applied three univariable MR (UVMR) statistics for our initial approach: Inverse
295 variance weighted (IVW, random effects model), Weighted median (WM) and MR-Egger,
296 requiring that at least two achieved a Bonferroni corrected p-value to claim significance
297 (**Methods**). After filtering, we observed non-zero causal effect estimates for 17 exposure-

298 outcome relationship (**Figure 2**). While 16 of them were robust specifically through IVW and
299 WM (Supplementary Table 5), only one exceeded the given threshold p-value in all 3 sensitivity
300 analyses.

301 We observed that four dietary preferences had inverse associations with T2D. A one-SD
302 increment in cheese intake (CHEESE), which was measured in a number of consumptions per
303 day and per week, lowered T2D risk by 57% (IVW, $\beta = -0.84$; OR, 0.43 per 1 SD; 95%
304 confidence (CI), 0.32 to 0.58; $P = 2.91 \times 10^{-8}$). A one-SD higher intake of dried fruit
305 (DRIEDFRU), which was measured in a number of pieces per day, also lower T2D risk (IVW, β
306 $= -1.23$; OR, 0.29; 95% CI, 0.19 to 0.45; $P = 2.49 \times 10^{-8}$). Intake of muesli (MUESLI), which is
307 primarily composed of rolled oats, was also observed to reduce the risk of T2D (IVW, $\beta = -0.93$;
308 OR, 0.40; 95% CI, 0.36 to 0.68; $P = 2.58 \times 10^{-9}$). As a type of cereal, muesli intake was measured
309 in a number of bowls per week. Lastly, intake of spreads (SPREADS), which consist of number
310 of uses of various fat-based spreads, was shown to have the largest reduction in T2D risk (IVW β
311 $= -2.60$; OR, 0.07; 95% CI, 0.02 to 0.21; $P = 1.41 \times 10^{-6}$), though we note that only five genetic
312 instruments were used to assess the effect of spread intake against T2D.

313 *Intake of dried fruits, muesli, spreads and corn flakes/frosties are causally associated with*
314 *cardiometabolic risk factors.*

315 Consistent with the above effect of diet on T2D susceptibility, dried fruit, muesli and spread
316 intake also were related to several cardiometabolic risk factors. The protective effect of dried
317 fruit intake on T2D was reflected through lower body mass index (BMI) (IVW, $\beta = -0.51$; OR,
318 3.21; 95% CI, 0.60 to 0.72; $P = 1.27 \times 10^{-8}$) and lower waist-hip ratio (WHR) (IVW, $\beta = -0.46$;
319 OR, 0.63; 95% CI, 0.53 to 0.74; $P = 5.88 \times 10^{-8}$). Also, dried fruit intake was shown to have a

320 mild reducing effect on the liver function measure of gamma-glutamyl transferase (GGT) (IVW,
321 $\beta = -0.08$; OR, 0.92; 95% CI, 0.90 to 0.95; $P = 5.25 \times 10^{-7}$). Likewise, muesli intake lowered
322 volumes of two adiposity-relevant organs: abdominal subcutaneous adipose tissue (ASAT)
323 (IVW, $\beta = -0.77$; OR, 0.46 ; 95% CI, 0.32 to 0.66; $P=2.45 \times 10^{-5}$) and visceral adipose tissue
324 (VAT) (IVW, $\beta = -0.70$; OR, 0.50; 95% CI, 0.37 to 0.67; $P = 2.64 \times 10^{-6}$). Intake of spreads
325 elevated high-density lipoprotein (HDL) (IVW $\beta = 1.00$; OR, 2.73; 95% CI, 1.97 to 3.79; $P=1.91$
326 $\times 10^{-9}$) and lower triglycerides (TG) (IVW $\beta = -0.73$; OR, 0.48; 95% CI, 0.37 to 0.63; $P=1.17 \times$
327 10^{-7}). However, we note that only six genetic instruments were used to assess the relationship
328 between spreads intake and these two lipids.

329 In addition, intake of corn flakes/frosties (CORNFLAK), which was based on a number of bowls
330 per week, increased ASAT by 89% (IVW, $\beta = 0.64$; OR, 1.89; 95% CI, 1.62 to .21; $P = 6.55 \times$
331 10^{-15}) and BMI by 49% (IVW, $\beta = 0.40$; OR, 1.49; 95% CI, 1.23 to 1.80; $P = 4.65 \times 10^{-5}$). Intake
332 of Corn flakes/frosties also elevated Pancreas volume (Pancvol) (IVW, $\beta = 0.50$; OR, 1.66; 95%
333 CI, 1.41 to 1.95; $P=8.94 \times 10^{-10}$) but lowered Liver iron content (Liveriron) (IVW, $\beta = -1.48$; OR,
334 0.23; 95% CI, 0.16 to 0.32; $P = 1.70 \times 10^{-17}$). Corn flakes/frosties intake to Liver iron content
335 was the only association tested to be significant across all 3 sensitivity analyses. The MR-Egger
336 causal estimate of this association was consistent with that of IVW and WM, and here the MR-
337 Egger intercept was 0.04 ($P = 2.28 \times 10^{-3}$, Supplementary Table 5), suggesting a suppressive
338 effect. Corn flakes and frosties are processed, refined carbohydrate-based foods with a high
339 glycemic index. Hence, the associations of corn flakes/frosties intake with these two organs
340 suggest the presence of dietary role in affecting the health of diabetes-relevant organs in the
341 development of T2D. Together, the relationships between these four dietary traits and

342 cardiometabolic risk factors agree with observational data that suggests both central adiposity
343 and dyslipidemia are strongly linked to T2D susceptibility [53,54].

344 *Preferences for Savory/Caloric foods and Strong-flavored foods are causally linked with*
345 *cardiometabolic risk factors.*

346 Finally, UVMR revealed two food preference traits related with cardiometabolic risk factors.
347 Preference for savory/caloric (SAVCAL) food increased VAT (IVW, $\beta = 0.13$; OR, 1.14; 95%
348 CI, 1.09 to 1.19; $P = 2.70 \times 10^{-9}$). SAVCAL falls under one of main food liking categories,
349 namely 'highly-palatable', which consists of high caloric, energy-dense foods, such as meat and
350 deep-fried foods. Preference for strong flavored (STR) food increased HDL (IVW, $\beta = 0.06$; OR,
351 1.06; 95% CI, 1.04 to 1.08; $P = 7.42 \times 10^{-12}$), though the effect size was relatively modest. STR
352 is a part of the other main food liking category, 'acquired', which consists of bitter, spicy, sharp
353 flavored foods, such as coffee and cheese, where preferences are acquired throughout life.

354 To evaluate the directionality of the causal effect estimated between dietary traits and outcomes,
355 we conducted the MR-Steiger test (**Supplementary Table 6**) [41]. We observed that 14 of the 17
356 exposure-outcome associations were predicted to have the causal effect direction indicated by
357 UVMR analyses. For the remaining three, the test showed significant reverse causality (corn
358 flakes/frosties intake to Liver iron content, Steiger $P = 3.23 \times 10^{-19}$) or was inconclusive on this
359 test (muesli intake to ASAT and VAT, Steiger $P > 0.05$). The possible explanations for the
360 opposite direction of causality could be due to the presence of pleiotropy in genetic instruments
361 and unmeasured confounding, which can be examined through inclusion of potential mediators
362 in multivariable MR (MVMR).

363 *MR-RAPS included weak genetic instruments and detected 7 additional associations.*

364 To identify additional exposure-outcome associations that could provide more information about
365 the Diet-T2D relationship, we employed MR-RAPS (Robust Adjusted Profile Score), an
366 alternative to the 3 standard MR methods, to re-assess 17 nominally significant ($P < 1 \times 10^{-3}$)
367 (**Supplementary Table 7**) associations in at least 2 sensitivity analyses. Thus, in addition to
368 strong instruments selected for UVMR, we included weak instruments, which are sub-genome
369 wide significant ($P < 1 \times 10^{-4}$), for MR-RAPS analysis.

370 Out of 17 associations, we identified seven additional exposure-outcome effects that met our
371 criteria for significance (**Methods, Figure 3**). These associations surpassed the Bonferroni-
372 adjusted p-value ($P < 5.99 \times 10^{-5}$) applied in the primary MR analysis (**Supplementary Table**
373 **8**) and were in the correct direction of causality based on Steiger's test (**Supplementary Table**
374 **9**). The directionality of these seven associations estimated by MR-RAPS and standard MR
375 methods was concordant, although as expected the effect size estimated from in MR-RAPs based
376 on a larger panel of SNP ($P < 1 \times 10^{-3}$) was substantially lower in many cases, relative to
377 instruments from genome-wide significant variants ($P < 5 \times 10^{-8}$).

378 In MR-RAPS, elevated preference for cheese consumption lowered BMI ($\beta = -0.18$; OR, 0.84
379 per 1 SD; 95% CI, 0.82 to 0.86; $P = 1.35 \times 10^{-33}$) and higher preference for muesli also lowered
380 BMI ($\beta = -0.16$; OR, 0.85 per 1 SD; 95% CI, 0.83 to 0.88; $P = 6.15 \times 10^{-28}$). These 2 associations
381 were in line with the protective effects of cheese and muesli intake on T2D observed in the
382 primary analysis. Also, they support that BMI is a potential mediator of the associations of these
383 2 dietary traits with T2D. Conversely, preference for coffee intake was shown to elevate BMI (β
384 $= 0.09$; OR, 1.09 per 1 SD; 95% CI, 1.06 to 1.13; $P = 1.19 \times 10^{-9}$). Additionally, MR-RAPS

385 suggested a negative link between alcohol intake and triglycerides ($\beta = -0.11$; OR, 0.90 per 1 SD;
386 95% CI, 0.88 to 0.92; $P = 7.51 \times 10^{-28}$).

387 Preference for STR was linked to slightly lower ASAT ($\beta = -0.06$; OR, 0.94 per 1 SD; 95% CI,
388 0.93 to 0.96; $P = 3.30 \times 10^{-15}$). Preference for SAVCAL, which was shown to have increasing
389 effect on VAT in the primary analysis, elevated liver fat content (Liverfat) ($\beta = 0.07$; OR, 1.07
390 per 1 SD; 95% CI, 1.05 to 1.09; $P = 7.11 \times 10^{-17}$). Preference for highly palatable foods (PAL),
391 whose sub-category is SAVCAL, increased liver fat content ($\beta = 0.17$; OR, 1.18 per 1 SD; 95%
392 CI, 1.12 to 1.24; $P = 5.91 \times 10^{-11}$).

393 We also explored MR-RAPS with spreads intake and HDL, TG and T2D, given only 5-6 genetic
394 instruments were utilized in UVMR. With inclusion of weak instruments, the relationship
395 between spreads intake and these traits did not reject the null hypothesis, and the strong causal
396 effects, particularly on HDL and TG, estimated in UVMR were no longer observed
397 (**Supplementary Table 10**). Given the high p-value, the extremely large effect of spreads intake
398 on T2D was greatly reduced with additional instruments ($\beta = -0.083$; OR, 0.92 per 1 SD; 95%
399 CI, 0.82 to 1.03; $P = 0.15$).

400 *Multivariable MR suggests that BMI is a common mediating risk factor for dietary exposure-*
401 *outcome relationships.*

402 To evaluate the 17 associations observed in the UVMR and better understand the underlying
403 causal pathways on outcomes, we next conducted multivariable MR (MVMR) experiments. We
404 employed 8 well-established non-dietary T2D risk factors as potential mediators (**Methods**),
405 including BMI, WHR, Fasting insulin adjusted for BMI (FI), Educational attainment (EA),

406 Diastolic blood pressure (DBP), Systolic blood pressure (SBP), Physical activity (PHY) and
407 Sedentary behavior at work (SED, Supplementary Table 2). We found that the causal effects
408 observed for 15 exposure-outcome associations were either substantially or fully attenuated after
409 including BMI in single-mediator MVMR (**Figure 4, Supplementary Table 11**), which assessed
410 the effect of an exposure on an outcome after adjustment for a single mediating trait in a pairwise
411 manner. For example, BMI appeared to be responsible for mediating the protective effect of
412 cheese intake against T2D, as no other trait attenuated this association (Supplementary Table
413 11). In addition, BMI completely attenuated the association between SAVCAL and VAT as well
414 as the associations of corn flakes/frosties intake with ASAT, Liver iron content, and Pancreas
415 volume. Furthermore, BMI yielded the largest attenuation in spreads intake to T2D and spreads
416 intake to HDL associations. This observation shows that the preference of various spreads is
417 closely linked with adiposity and therefore could contribute to cardiometabolic health.

418 Since the 4 dietary associations with T2D were our primary interest, we also checked whether
419 there were additional factors that attenuated any of those associations beyond BMI
420 (Supplementary Table 11). The single-mediator MVMR suggested that DBP could be an
421 additional driver for the observed effect of dried fruit intake on BMI, GGT, T2D and WHR and
422 the observed effect of muesli intake on ASAT, T2D and VAT. However, the effect of DBP as a
423 potential mediator on these associations were not clearly determinable ($P > 0.05$) and segregated
424 relatively wide 95% confidence intervals. Surprisingly, there was residual direct effect of dried
425 fruit intake on T2D even after accounting the mediating effect of BMI in the single-mediator
426 MVMR. We observed that EA was an additional factor that attenuated the associations of muesli
427 intake with T2D, ASAT and VAT, to a degree that comparable to those by BMI (**Figure 5**).

428 Followed by single-mediator MVMR, we then conducted additional MVMR experiments to re-
429 estimate the effect of the exposures in joint with multiple mediators on the outcomes. We carried
430 out 2 different multiple-mediator MVMR analyses: non-BMI mediators MVMR and all-inclusive
431 MVMR, which includes all selected potential mediating traits. The purpose of multiple-mediator
432 MVMR was to verify whether BMI is the primary mediators of the associations amongst the
433 selected potential mediating traits. In non-BMI mediators MVMR, we checked whether non-
434 BMI mediators together as a group can yield greater attenuation in the absence of BMI. In all-
435 inclusive MVMR, we examined the direct effect of the exposures on the outcomes by accounting
436 the summative effect of all mediators and compared with that of non-BMI mediators MVMR.

437 Overall, the multiple-mediator MVMR showed that BMI was a major mediator for most of the
438 associations observed (**Supplementary Figure 4, Supplementary Table 12**). For few
439 associations, it was not immediately clear if BMI was the strongest mediator. The multiple-
440 mediator MVMR further supported that BMI was a definitive primary risk factor for mediating
441 the protective effect of cheese intake against T2D. These analyses also indicated that summative
442 effect of multiple mediators does not necessarily result in greater attenuation.

443 *BMI was a common mediator for the 3 dietary associations identified by MR-RAPS.*

444 For further assessment, we performed single-mediator MVMR on the seven additional
445 associations detected in MR-RAPS (**Supplementary Table 13**). The same eight potential
446 mediators were employed to determine whether same, non-dietary risk factors mediate these
447 associations. Notably, none of the potential mediators attenuated the Cheese Intake-BMI and
448 ALC-TG associations. This underscores our observation that BMI is the primary mediator for the
449 effect of cheese intake on T2D. BMI was shown to attenuate PAL-Liver fat content, SAVCAL-

450 Liver fat content and STR-ASAT association (**Figure 6**). EA was also a mediator for muesli
451 intake-BMI, SAVCAL-Liver fat content, and STR-ASAT associations (**Figure 7**). These two
452 findings were congruent with the observations in the MVMR on the dietary associations from
453 UVMR.

454 *BMI was a major contributor to the dietary associations with T2D in the mediation analysis.*

455 In addition to MVMR, we took two-step MR approach (**Methods**) to conduct mediation analysis
456 and estimate the indirect effects of cheese, dried fruit and muesli intake on T2D after accounting
457 for BMI, DBP and EA individually as mediating traits (**Supplementary Tables 19-20**). Spreads
458 intake-T2D association was excluded due to a small number of strong genetic instruments (n=5)
459 used to test the association. Whereas MVMR estimates the direct effect of a primary exposure on
460 outcome, mediation analysis quantifies the indirect effect mediated by a secondary exposure. We
461 compared the results of mediation analysis to those of MVMR and evaluated those 3 mediating
462 traits as potential mediators for the dietary associations with T2D. **Figure 8** displays the
463 proportion of the effect of each dietary preference on T2D explained by each mediator
464 separately.

465 Only BMI explained 44.97% (95% CI 24.85%, 65.09%) of the total effect of cheese intake on
466 T2D, confirming the finding of MVMR that BMI is the only mediator for the protective effect of
467 cheese intake on T2D. After adjustment for cheese intake, elevated BMI increased risk of T2D
468 (IVW, $\beta = 0.86$; OR 2.36; 95% CI, 2.21 to 2.52; $P = 1.20 \times 10^{-144}$) but the associations of DBP
469 and EA were not significant. The BMI-mediated effect constituted 35.77% (95% CI 22.76%,
470 47.97%) and 58.13% (95% CI 27.99%, 88.27%) of the total effects of dried fruit and muesli
471 intake, respectively, on T2D.

472 The mediation analysis also helped clarify that DBP does not mediate the associations of dried
473 fruit and muesli intake with T2D, because the DBP-mediate effects of dried fruit (1.54%, 95% CI
474 -0.02%, 3.25%) and muesli intake (4.52%, 95% CI 0.29%, 8.72%) on T2D were minimal.

475 Consistent with MVMR, EA was another mediator for muesli intake-T2D association. Given that
476 higher EA reduces the risk of T2D (IVW, $\beta = -0.49$; OR 0.61; 95% CI, 0.57 to 0.65; $P = 1.20 \times$
477 10^{-46}), EA explained 54.9% (95% CI 36.6%, 73.2%) of the total effect of muesli intake on T2D.

478 In agreement with MVMR, the mediation analysis demonstrated that BMI is a common factor
479 that substantially mediates the associations of cheese, dried fruit and muesli intake with T2D.

480 **DISCUSSION**

481 Type 2 diabetes mellitus (T2D) is a chronic disease that requires pharmacologic and lifestyle
482 interventions for effective disease management. In particular, diet plays an important role in
483 modulating the risk of T2D and influencing disease progression and onset of complications.
484 Though past observational studies have shown that certain diets can promote good metabolic
485 health or confer impaired glucose homeostasis [55,56], they cannot directly test causality.

486 Hence, we conducted Mendelian randomization (MR) study to examine genetic effects of various
487 dietary preferences on T2D and 21 related cardiometabolic risk factors in order to determine
488 which, if any, dietary exposures exerted causal effects on susceptibility to T2D or related traits.

489 In univariable MR (UVMR), the most notable associations were between intake of cheese, dried
490 fruits, muesli and spreads and T2D. These four inverse associations correspond to previously
491 reported observational findings. Some prospective cohort studies found that total dairy
492 consumption was associated with reduced T2D risk [57,58]. Cheese, a nutrient-dense dairy
493 product with many bioactive compounds, has been observed to lower cardiovascular mortality

494 and incidence of T2D [59]. Furthermore, the association between cheese intake and T2D could
495 be partly explained by the interaction between lactase *LCT* gene and milk in the susceptibility of
496 T2D [60,61]. High intake of whole grains, not refined grains, was shown to lower T2D risk [62].
497 The dose-response meta-analysis provided weak-to-moderate evidence that increasing servings
498 of fiber-rich fruits and vegetables can help reduce incidence of T2D and related mortality
499 [63,64]. Dietary fat is a low-glycemic source of energy and bioactive fatty acids that affect
500 pancreatic beta cell metabolism, which is critical for glucose homeostasis [65]. Though total
501 dietary fat consumption was not associated with T2D, some dietary fats, particularly regular
502 dairy fats, have been shown to have benefits of reducing risk of T2D [66].

503 The follow-up multivariable MR experiments showed that socioeconomic and cardiometabolic
504 risk factors play a mediating role in the observed associations. Importantly, body mass index
505 (BMI) [67,68] was the factor that attenuated most of the dietary associations. These results
506 reflect both the long-standing observation that obesity is a strong etiological risk factor for T2D,
507 but also the close relationship between diet and adiposity. These results explain, in part, why
508 some dietary preferences were shown to be associated with BMI including other adiposity
509 related traits, such as WHR and MRI-based volumes of central, abdominal adipose tissues in
510 UVMR. Surprisingly, none of the associations were affected by WHR, which is an alternative
511 measure of adiposity. This observation might be due to difference in statistical power because
512 GWAS data for BMI had a slightly larger sample size than for WHR. Alternatively, it might
513 suggest that BMI and WHR are implicated in different pathways. In addition, we identified that
514 EA was an additional mediating trait for muesli intake associations.

515 In the two-step MR mediation analysis, we were able to further interpret our initial observations
516 from MVMR. The analysis further supported the specific diet preferences, particularly cheese

517 intake, mediate susceptibility to T2D partly via BMI. It also concurred with MVMR that DBP
518 was not a potential mediating factor in association between diet and T2D, aside from BMI and
519 EA. Such observations suggest that adiposity and education are factors that diet interacts with to
520 play an essential role in cardiometabolic health.

521 However, the study has limitations. Other than the three GWAS data sets for various dietary
522 preference traits used for this study, there are no additional genetic data that are available in
523 comparable sample size to replicate or further examine the findings. Also, the GWAS data for
524 dietary preferences were specifically based on participants with European ancestry background
525 that our results lack generalizability to other populations or ancestries. Because these dietary
526 preferences are qualitatively measured via self-reported questionnaires from participants, they
527 could be prone to bias. In the 2-step MR mediation analysis, the mediated (indirect) effect
528 estimates should be taken with caution as the outcome of our interest was T2D, which is binary,
529 that they were defined in noncollapsible odds ratio and therefore may not be accurate [69,70].
530 Lastly, because T2D is a multifactorial and heterogeneous disease, there could be unmeasured
531 risk factors that were not tested in this study to further explain diet-T2D association.

532 Collectively, our findings show that diet is not an independent risk factor for T2D but dependent
533 of established T2D risk factors, particularly obesity and low education level, to causally
534 influence liability to T2D. These imply that diet plays a role in the maintenance of healthy
535 weight and making food choices could be affected by education status and therefore is important
536 for reducing T2D risk. The study also highlights that genetic data associated with diet should be
537 integrated in tailoring effective dietary interventions for individuals at risk for poor metabolic
538 health. Though further investigation is needed to confirm these findings, the current study

539 provides positive genetic evidence that diet is one of the factors that contribute to susceptibility
540 to T2D.

541 **ACKNOWLEDGEMENTS**

542 Summary association data for type 2 diabetes obtained from the MVP project (phs001672) was
543 obtained from dbGAP through data access request (approval #27398). BFV is grateful for
544 support for the work from the NIH/NIDDK (DK126194). MDL is grateful for support for the
545 work from the NIH/NHGRI (T32 HG000046).

546 **AUTHOR CONTRIBUTIONS**

547 The authors' responsibilities were as follows— MDL and BFV designed research and had primary
548 responsibility for the final content; MDL conducted the analyses and wrote the manuscript; BFV
549 supervised the project and contributed to the methodology used in the paper; MDL and BFV:
550 interpreted the results, critically reviewed the manuscript, and approved the final manuscript.

551 **CONFLICT OF INTEREST**

552 The authors report no conflicts of interest.

553 **SOFTWARE**

554 Software Analyses were conducted in R software version 4.3.2, with the use of packages
555 TwoSampleMR, MVMR [40,41,43] and RMediation [50].

556 **DATA AVAILABILITY**

557 Data described in the manuscript, code book, and analytic code will be made publicly and freely
558 available without restriction at [https://github.com/mdayeon/Mendelian-randomization-on-diet-](https://github.com/mdayeon/Mendelian-randomization-on-diet-and-T2D)
559 [and-T2D.](https://github.com/mdayeon/Mendelian-randomization-on-diet-and-T2D)

560 **FIGURE LEGENDS**

561 **Figure 1:** Schematic workflow of the Mendelian Randomization study on diet and T2D.

562 This figure describes how the MR study on diet and T2D was conducted in 4 steps to investigate
563 their relationship and evaluate the causality. After selecting viable genetic variants that meet the
564 given criteria, univariable MR analysis estimated the total effects of dietary preferences on T2D
565 and 21 related cardiometabolic risk factors. Then, MR-RAPS, which is an alternative method to
566 standard MR, included weak instruments to re-evaluate the nominally significant
567 exposure:outcome associations and detect additional associations. Finally, the significant
568 associations from UVMR and MR-RAPS were further assessed by estimating the direct effects
569 of dietary preferences on the outcomes after accounting for potential mediators in MVMR.

570 Abbreviations – MR, Mendelian Randomization; MR-RAPS, Mendelian Randomization Robust
571 Adjusted Profile Score; T2D, Type 2 diabetes mellitus; UVMR, Univariable Mendelian
572 randomization.

573 **Figure 2:** Forest plot for 17 associations from UVMR.

574 This figure is a forest plot for 17 associations from UVMR-IVW and UVMR-WM in descending
575 order of OR for each dietary preference trait. Given the Bonferroni-corrected p-value, these
576 associations passed the significance threshold ($P < 5.99 \times 10^{-5}$) in MR using IVW and WM
577 methods. Abbreviations – IVW, Inverse variance weighted; MR, Mendelian Randomization; OR,
578 Odds ratio; UVMR, Univariable Mendelian randomization; WM, Weighted median.

579 **Figure 3:** Forest plot for 7 associations from MR-RAPS.

580 This figure is a forest plot for 7 associations identified via MR-RAPS. In comparison, the results
581 from MR-RAPS analysis using only weak instruments and using both strong and weak
582 instruments are consistent in direction with that of IVW method in UVMR. Abbreviations –

583 IVW, Inverse variance weighted; MR-RAPS, Mendelian Randomization Robust Adjusted Profile
584 Score; UVMR, Univariable Mendelian Randomization.

585 **Figure 4:** Forest plot for single-mediator MVMR using BMI vs. UVMR-IVW.

586 This figure is a forest plot that displays the results of both single-mediator MVMR using BMI as
587 a potential mediator and UVMR-IVW on 15 associations. Compared to the results of univariable
588 MR using IVW, the strengths of 15 associations were either partially or completely attenuated by
589 BMI. Abbreviations – BMI, Body mass index; IVW, Inverse variance weighted; MVMR,
590 Multivariable Mendelian Randomization; UVMR, Univariable Mendelian Randomization.

591 **Figure 5:** Forest plot for single-mediator MVMR using BMI and EA on Muesli intake
592 associations. This figure is a forest plot for Muesli-ASAT, T2D and VAT associations in UVMR
593 and single-mediator MVMR. Aside from BMI, EA was an additional factor that attenuated the
594 muesli intake associations, to a degree that comparable to those by BMI. Abbreviations – ASAT,
595 Abdominal subcutaneous adipose tissue; BMI, Body mass index; EA, Educational attainment;
596 MVMR, Multivariable Mendelian Randomization; T2D; UVMR, Univariable Mendelian
597 Randomization; VAT, Visceral adipose tissue.

598 **Figures 6:** Forest plot for single-mediator MVMR using BMI on the associations from MR-
599 RAPS. This figure is a forest plot for single-mediator MVMR using BMI on the following
600 associations detected in MR-RAPS: PAL-Liver fat content, SAVCAL-Liver fat content and
601 STR-ASAT associations. As observed in single-mediator MVMR on the associations in UVMR,
602 BMI was a common mediating trait that attenuated the effects of these 3 dietary preferences on
603 Liver fat content and ASAT. Abbreviations – ASAT, Abdominal subcutaneous adipose tissue;
604 BMI, Body mass index; MR-RAPS, Mendelian Randomization Robust Adjusted Profile Score;

605 MVMR, Multivariable Mendelian Randomization; PAL, Highly palatable; SAVCAL, Savory/
606 caloric; STR, Strong flavored; UVMR, Univariable Mendelian Randomization.
607 **Figures 7:** Forest plot for single-mediator using EA on the associations from MR-RAPS. This
608 figure is a forest plot for single-mediator MVMR using EA on the following associations
609 detected in MR-RAPS: SAVCAL-Liver fat content, STR-ASAT and muesli intake-BMI
610 associations. Forest plot for single-mediator MVMR on 3 associations in MR-RAPS. Consistent
611 with the observations in MVMR on the associations, particularly with muesli intake, in UVMR,
612 EA was another mediating factor for dietary associations with cardiometabolic risk factors.
613 Abbreviations – ASAT, Abdominal subcutaneous adipose tissue; BMI, Body mass index; EA,
614 Educational attainment ; MR-RAPS, Mendelian Randomization Robust Adjusted Profile Score;
615 MVMR, Multivariable Mendelian Randomization; SAVCAL, Savory/ caloric; STR, Strong
616 flavored.

617 **Figure 8:** Forest plot for two-step MR mediation analysis. This figure is a forest plot for two-step
618 MR mediation analysis that includes 95% confidence interval for estimated indirect effect and
619 proportion mediated. It displays the indirect effects of 3 dietary preferences on T2D mediated
620 individually by BMI, DBP and EA via RMediation package. Delta method (“asymptotic
621 distribution”) was used to estimate the mediated effects by these 3 factors and calculate 95%
622 confidence intervals. Consistent with MVMR, BMI was a major mediator for the 3 dietary
623 associations with T2D and EA was additional mediator exclusively for the effect of muesli intake
624 on T2D. Abbreviations – BMI, Body mass index; DBP, Diastolic blood pressure; EA,
625 Educational attainment; MR, Mendelian Randomization; MVMR, Multivariable Mendelian
626 Randomization; T2D, Type 2 diabetes mellitus.

627 **SUPPLEMENTARY FIGURE AND TABLE LEGENDS**

- 628 **Supplementary Table 1:** GWAS data for dietary preferences used in MR.
- 629 **Supplementary Table 2:** GWAS data for T2D and 21 cardiometabolic risk factors used in MR.
- 630 **Supplementary Table 3:** Number of genetic instruments selected for each exposure.
- 631 **Supplementary Table 4:** *F*-statistics of the genetic instruments for exposures.
- 632 **Supplementary Table 5:** 17 significant associations ($P < 5.99 \times 10^{-5}$) in UVMR.
- 633 **Supplementary Table 6:** MR-Steiger test for the directions of the 17 associations in UVMR.
- 634 **Supplementary Table 7:** 17 Nominally significant ($P < 1 \times 10^{-3}$) associations in UVMR.
- 635 **Supplementary Table 8:** 7 significant associations ($P < 2.94 \times 10^{-3}$) in MR-RAPS.
- 636 **Supplementary Table 9:** MR-Steiger test for the directions of the 17 nominally significant
637 associations in UVMR.
- 638 **Supplementary Table 10:** Re-assessment of spreads intake associations in MR-RAPS.
- 639 **Supplementary Table 11:** Single-mediator MVMR using 8 potential mediators.
- 640 **Supplementary Table 12:** Multiple-mediator MVMR (all-inclusive vs. non-BMI).
- 641 **Supplementary Table 13:** Single-mediator MVMR on 7 significant associations in MR-RAPS.
- 642 **Supplementary Table 14:** A table of genetic instruments for exposures, outcomes and potential
643 mediators for UVMR and MVMR.
- 644 **Supplementary Table 15:** A table of genetic instruments for exposures, outcomes and potential
645 mediators for MR-RAPS.
- 646 **Supplementary Table 16:** A summary table of UVMR on 835 exposure-outcome associations.
- 647 **Supplementary Table 17:** A table of genetic correlations between potential mediators and
648 exposures.
- 649 **Supplementary Table 18:** A table of genetic correlations between potential mediators and
650 outcomes.

651 **Supplementary Table 19:** 2-step MR mediation analysis on 3 dietary associations with T2D.

652 **Supplementary Table 20:** Indirect effect, proportion mediated and 95% confidence intervals via
653 RMediation package.

654 **Supplementary Table 21:** Estimates of bias and Type 1 error rates due to sample overlap for the
655 associations in UVMR and MR-RAPS.

656 **Supplementary Figure 1: Framework of MR**

657 Mendelian randomization employs genetic variants (e.g., SNPs) associated with an exposure as
658 instrumental variables (IVs). Given that the selected IVs meet the 3 assumptions for IVs, the
659 effect of an exposure on an outcome of interest is estimated.

660 **Supplementary Figure 2: Genetic instrument selection via PLINK**

661 The genetic instrument process was taken in 3 steps. First, variants that are rare ($MAF < 0.01$) or
662 have low imputation quality ($INFO < 0.5$) were eliminated. Also, the datasets were screened for
663 duplicate variants, if any. Second, a tool PLINK was used to perform clumping procedure and
664 select genome-wide significant variants with low linkage disequilibrium. Last, only non-
665 palindromic variants were taken into estimation of instrument strength using F-statistics formula
666 to determine whether they are strong instruments that can be used in MR to test exposure-
667 outcome associations.

668 **Supplementary Figure 3: Framework of Multivariable MR (MVMR)**

669 Multivariable MR (MVMR) assumes that genetic instruments selected are pleiotropic and
670 therefore also associated with other traits (e.g., secondary exposure). The secondary exposure
671 can be a mediator or confounder. It assesses the exposure-outcome association with inclusion of
672 the secondary exposure. Whereas, univariable MR estimates the total effect of the exposure on

673 the outcome, MVMR estimates the direct effect of exposure on outcome after adjusting for the
674 effect of a potential confounder.

675 **Supplementary Figure 4: Forest plot for multiple-mediator MVMR**

676 Forest plot for multiple-mediator multivariable MR for comparison. The plot compares the
677 summative effects of all mediators on the associations to that of all non-BMI mediators to
678 determine whether BMI is the main driver for the causal effects of the dietary traits on T2D and
679 cardiometabolic risk factors.

REFERENCES

1. Ahlqvist E, Ahluwalia TS, Groop L. Genetics of type 2 diabetes. *Clin Chem*. 2011;57(2):241-54.
2. Ardisson Korat AV, Willett WC, Hu FB. Diet, lifestyle, and genetic risk factors for type 2 diabetes: a review from the Nurses' Health Study, Nurses' Health Study 2, and Health Professionals' Follow-up Study. *Curr Nutr Rep*. 2014;3(4):345-354.
3. Zimmet PZ. Diabetes and its drivers: the largest epidemic in human history? *Clin Diabetes Endocrinol*. 2017;3:1.
4. Vujkovic M, Keaton JM, Lynch JA, Miller DR, Zhou J, Tcheandjieu C, 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*. 2020;52(7):680-691.
5. Mahajan A, Spracklen CN, Zhang W, Ng MCY, Petty LE, Kitajima H, et al. Multi-ancestry genetic study of type 2 diabetes highlights the power of diverse populations for discovery and translation. *Nat Genet*. 2022;54(5):560-572.
6. Cole JB, Florez JC. Genetics of diabetes mellitus and diabetes complications. *Nat Rev Nephrol*. 2020;16(7):377-390.

7. Parillo M, Riccardi G. Diet composition and the risk of type 2 diabetes: epidemiological and clinical evidence. *Br J Nutr.* 2004;92(1):7-19.
8. Guo Y, Huang Z, Sang D, Gao Q, Li Q. The Role of Nutrition in the Prevention and Intervention of Type 2 Diabetes. *Front Bioeng Biotechnol.* 2020;8:575442.
9. Koeth RA, Wang Z, Levison BS, Buffa JA, Org E, Sheehy BT, et al. Intestinal microbiota metabolism of L-carnitine, a nutrient in red meat, promotes atherosclerosis. *Nat Med.* 2013;19(5):576-85.
10. Laeger T, Henagan TM, Albarado DC, Redman LM, Bray GA, Noland RC, et al. FGF21 is an endocrine signal of protein restriction. *J Clin Invest.* 2014;124(9):3913-22.
11. Basu S, Yoffe P, Hills N, Lustig RH. The relationship of sugar to population-level diabetes prevalence: an econometric analysis of repeated cross-sectional data. *PLoS One.* 2013;8(2):e57873.
12. Luukkonen PK, Sädevirta S, Zhou Y, Kayser B, Ali A, Ahonen L, et al. Saturated Fat Is More Metabolically Harmful for the Human Liver Than Unsaturated Fat or Simple Sugars. *Diabetes Care.* 2018;41(8):1732-1739.
13. Hu FB, van Dam RM, Liu S. Diet and risk of Type II diabetes: the role of types of fat and carbohydrate. *Diabetologia.* 2001;44(7):805-17.
14. Tholin S, Rasmussen F, Tynelius P, Karlsson J. Genetic and environmental influences on eating behavior: the Swedish Young Male Twins Study. *Am J Clin Nutr.* 2005;81(3):564-9.
15. Hasselbalch AL, Heitmann BL, Kyvik KO, Sørensen TI. Studies of twins indicate that genetics influence dietary intake. *J Nutr.* 2008;138(12):2406-12.

16. Ortega Á, Berná G, Rojas A, Martín F, Soria B. Gene-Diet Interactions in Type 2 Diabetes: The Chicken and Egg Debate. *Int J Mol Sci.* 2017;18(6):1188.
17. 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.
18. 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):2056-2069.
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. Burgess S, Small DS, Thompson SG. A review of instrumental variable estimators for Mendelian randomization. *Stat Methods Med Res.* 2017;26(5):2333-2355.
21. Sanderson E, Glymour MM, Holmes MV, Kang H, Morrison J, Munafò MR, et al. Mendelian randomization. *Nat Rev Methods Primers.* 2022;2:6.
22. Yuan S, Larsson SC. An atlas on risk factors for type 2 diabetes: a wide-angled Mendelian randomisation study. *Diabetologia.* 2020;63(11):2359-2371.
23. Ference BA, Robinson JG, Brook RD, Catapano AL, Chapman MJ, Neff DR, et al. Variation in PCSK9 and HMGCR and Risk of Cardiovascular Disease and Diabetes. *N Engl J Med.* 2016;375(22):2144-2153.
24. Holmes MV, Dale CE, Zuccolo L, Silverwood RJ, Guo Y, Ye Z, et al. Association between alcohol and cardiovascular disease: Mendelian randomisation analysis based on individual participant data. *BMJ.* 2014;349:g4164.

25. Larsson SC, Carter P, Kar S, Vithayathil M, Mason AM, Michaëlsson K, et al. Smoking, alcohol consumption, and cancer: A mendelian randomisation study in UK Biobank and international genetic consortia participants. *PLoS Med.* 2020;17(7):e1003178.
26. Cohen JC, Boerwinkle E, Mosley TH Jr, Hobbs HH. Sequence variations in PCSK9, low LDL, and protection against coronary heart disease. *N Engl J Med.* 2006;354(12):1264-72.
27. Yao S, Zhang M, Dong SS, Wang JH, Zhang K, Guo J, et al. Bidirectional two-sample Mendelian randomization analysis identifies causal associations between relative carbohydrate intake and depression. *Nat Hum Behav.* 2022;6(11):1569-1576.
28. Pulit SL, Stoneman C, Morris AP, Wood AR, Glastonbury CA, Tyrrell J, et al. Meta-analysis of genome-wide association studies for body fat distribution in 694 649 individuals of European ancestry. *Hum Mol Genet.* 2019;28(1):166-174.
29. 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):840-860.
30. Graham SE, Clarke SL, Wu KH, Kanoni S, Zajac GJM, Ramdas S, et al. The power of genetic diversity in genome-wide association studies of lipids. *Nature.* 2021;600(7890):675-679.
31. Liu Y, Bastý N, Whitcher B, Bell JD, Sorokin EP, van Bruggen N, et al. Genetic architecture of 11 organ traits derived from abdominal MRI using deep learning. *Elife.* 2021;10:e65554.
32. Pazoki R, Vujkovic M, Elliott J, Evangelou E, Gill D, Ghanbari M, et al. Genetic analysis in European ancestry individuals identifies 517 loci associated with liver enzymes. *Nat Commun.* 2021;12(1):2579.

33. Chen VL, Du X, Chen Y, Kuppa A, Handelman SK, Vohnoutka RB, et al. Genome-wide association study of serum liver enzymes implicates diverse metabolic and liver pathology. *Nat Commun.* 2021;12(1):816.
34. Purcell S, Neale B, Todd-Brown K, Thomas L, Ferreira MA, Bender D, et al. PLINK: a tool set for whole-genome association and population-based linkage analyses. *Am J Hum Genet.* 2007;81(3):559-75.
35. Burgess S, Davey Smith G, Davies NM, Dudbridge F, Gill D, Glymour MM, et al. Guidelines for performing Mendelian randomization investigations: update for summer 2023. *Wellcome Open Res.* 2023;4:186.
36. Brion MJ, Shakhbazov K, Visscher PM. Calculating statistical power in Mendelian randomization studies. *Int J Epidemiol.* 2013;42(5):1497-501.
37. Burgess S, Thompson SG. Bias in causal estimates from Mendelian randomization studies with weak instruments. *Stat Med.* 2011;30(11):1312-23.
38. Zhao Q, Wang J, Hemani G, Bowden J, Small DS. Statistical inference in two-sample summary-data Mendelian randomization using robust adjusted profile score. *Ann Statist.* 2020;48(3):1742-1769.
39. Burgess S, Davies NM, Thompson SG. Bias due to participant overlap in two-sample Mendelian randomization. *Genet Epidemiol.* 2016;40(7):597-608.
40. Hemani G, Zheng J, Elsworth B, Wade KH, Haberland V, Baird D, et al. The MR-Base platform supports systematic causal inference across the human phenome. *Elife.* 2018;7:e34408.

41. 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.
42. Slob EAW, Burgess S. A comparison of robust Mendelian randomization methods using summary data. *Genet Epidemiol.* 2020;44(4):313-329.
43. Sanderson E, Spiller W, Bowden J. Testing and correcting for weak and pleiotropic instruments in two-sample multivariable Mendelian randomization. *Stat Med.* 2021;40(25):5434-5452.
44. Bulik-Sullivan BK, Loh PR, Finucane HK, Ripke S, Yang J, Schizophrenia Working Group of the Psychiatric Genomics Consortium; et al. LD Score regression distinguishes confounding from polygenicity in genome-wide association studies. *Nat Genet.* 2015;47(3):291-5.
45. Papaioannou TG, Oikonomou E, Lazaros G, Christoforatu E, Vogiatzi G, Tsalamandris S, et al. Arterial stiffness and subclinical aortic damage of reclassified subjects as stage 1 hypertension according to the new 2017 ACC/AHA blood pressure guidelines. *Vasa.* 2019;48(3):236-243.
46. Wang Z, Emmerich A, Pillon NJ, Moore T, Hemerich D, Cornelis MC, et al. Genome-wide association analyses of physical activity and sedentary behavior provide insights into underlying mechanisms and roles in disease prevention. *Nat Genet.* 2022;54(9):1332-1344.
47. Okbay A, Wu Y, Wang N, Jayashankar H, Bennett M, Nehzati SM, et al. Polygenic prediction of educational attainment within and between families from genome-wide association analyses in 3 million individuals. *Nat Genet.* 2022;54(4):437-449.

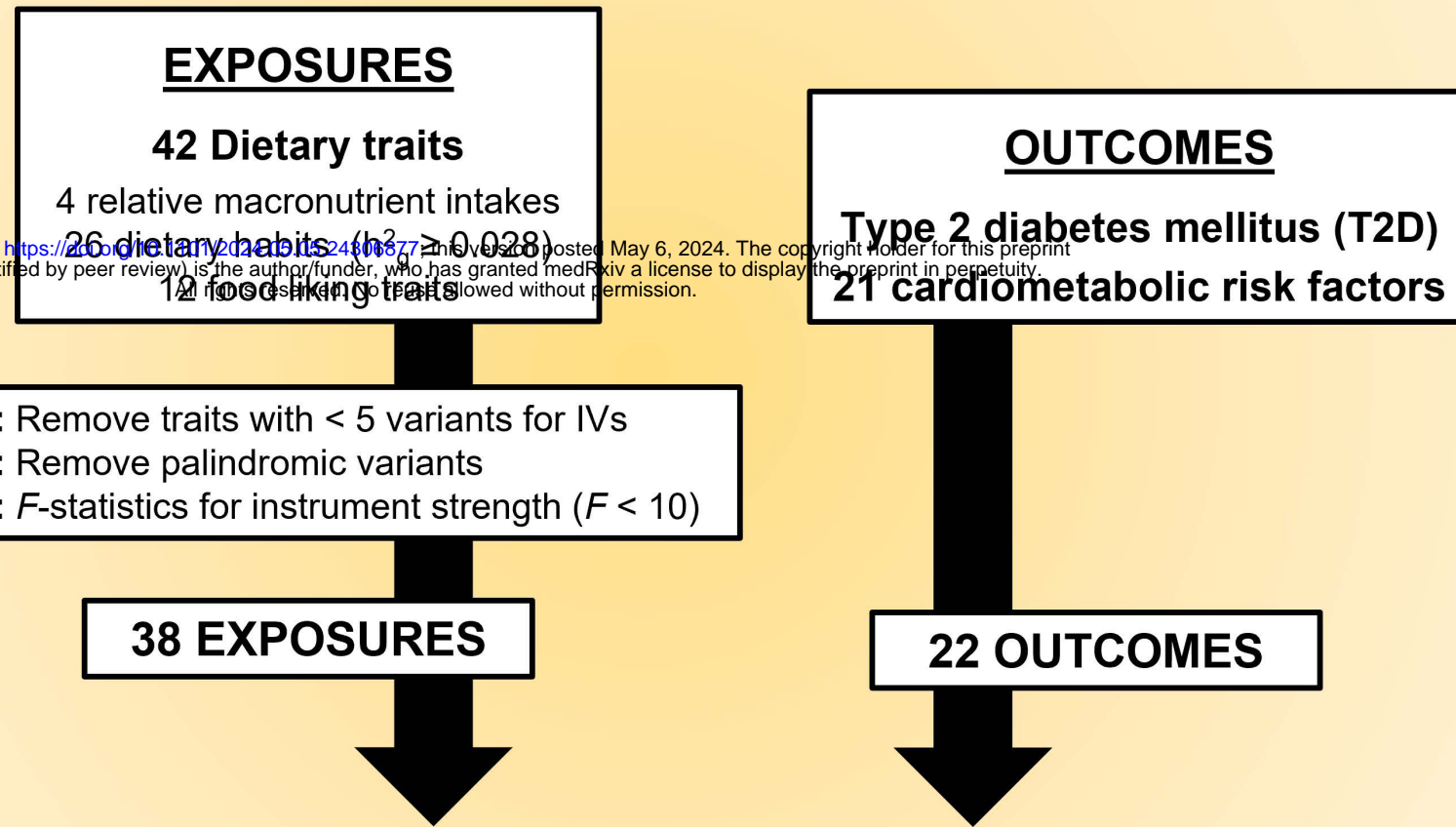
48. Carter AR, Sanderson E, Hammerton G, Richmond RC, Davey Smith G, Heron J, et al. Mendelian randomisation for mediation analysis: current methods and challenges for implementation. *Eur J Epidemiol*. 2021;36(5):465-478.
49. Burgess S, Daniel RM, Butterworth AS, Thompson SG, EPIC-InterAct Consortium. Network Mendelian randomization: using genetic variants as instrumental variables to investigate mediation in causal pathways. *Int J Epidemiol*. 2015;44(2):484-95.
50. Tofighi D, MacKinnon DP. RMediation: an R package for mediation analysis confidence intervals. *Behav Res Methods*. 2011;43(3):692-700.
51. Thompson JR, Minelli C, Del Greco M F. Mendelian Randomization using Public Data from Genetic Consortia. *Int J Biostat*. 2016;12(2):/j/ijb.2016.12.issue-2/ijb-2015-0074/ijb-2015-0074.xml.
52. Carter AR, Gill D, Davies NM, Taylor AE, Tillmann T, Vaucher J, et al. Understanding the consequences of education inequality on cardiovascular disease: mendelian randomisation study. *BMJ*. 2019;365:l1855.
53. Sung KC, Seo MH, Rhee EJ, Wilson AM. Elevated fasting insulin predicts the future incidence of metabolic syndrome: a 5-year follow-up study. *Cardiovasc Diabetol*. 2011;10:108.
54. ter Horst KW, Gilijamse PW, Koopman KE, de Weijer BA, Brands M, Kootte RS, et al. Insulin resistance in obesity can be reliably identified from fasting plasma insulin. *Int J Obes (Lond)*. 2015;39(12):1703-9.
55. Martín-Peláez S, Fito M, Castaner O. Mediterranean Diet Effects on Type 2 Diabetes Prevention, Disease Progression, and Related Mechanisms. A Review. *Nutrients*. 2020;12(8):2236.

56. Neuenschwander M, Ballon A, Weber KS, Norat T, Aune D, Schwingshackl L, et al. Role of diet in type 2 diabetes incidence: umbrella review of meta-analyses of prospective observational studies. *BMJ*. 2019;366:12368.
57. Choi HK, Willett WC, Stampfer MJ, Rimm E, Hu FB. Dairy consumption and risk of type 2 diabetes mellitus in men: a prospective study. *Arch Intern Med*. 2005;165(9):997-1003.
58. Pittas AG, Dawson-Hughes B, Li T, Van Dam RM, Willett WC, Manson JE, et al. Vitamin D and calcium intake in relation to type 2 diabetes in women. *Diabetes Care*. 2006;29(3):650-6.
59. Zhang M, Dong X, Huang Z, Li X, Zhao Y, Wang Y, et al. Cheese consumption and multiple health outcomes: an umbrella review and updated meta-analysis of prospective studies. *Adv Nutr*. 2023;14(5):1170-1186.
60. Luo K, Chen GC, Zhang Y, Moon JY, Xing J, Peters BA, et al. Variant of the lactase LCT gene explains association between milk intake and incident type 2 diabetes. *Nat Metab*. 2024;6(1):169-186.
61. Littleton SH, Grant SFA. Metabolic links among milk, genes and gut. *Nat Metab*. 2024;6(1):12-13.
62. Aune D, Norat T, Romundstad P, Vatten LJ. Whole grain and refined grain consumption and the risk of type 2 diabetes: a systematic review and dose-response meta-analysis of cohort studies. *Eur J Epidemiol*. 2013;28(11):845-58.
63. Wang DD, Li Y, Bhupathiraju SN, Rosner BA, Sun Q, Giovannucci EL, et al. Fruit and Vegetable Intake and Mortality: Results From 2 Prospective Cohort Studies of US Men

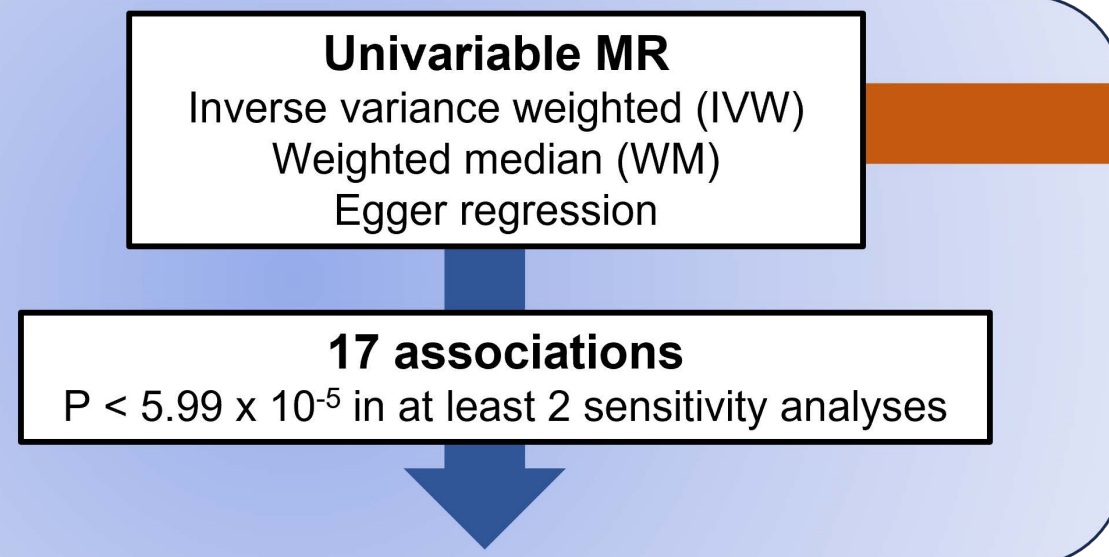
- and Women and a Meta-Analysis of 26 Cohort Studies. *Circulation*. 2021;143(17):1642-1654.
64. Li M, Fan Y, Zhang X, Hou W, Tang Z. Fruit and vegetable intake and risk of type 2 diabetes mellitus: meta-analysis of prospective cohort studies. *BMJ Open*. 2014;4(11):e005497.
65. Acosta-Montaña P, García-González V. Effects of Dietary Fatty Acids in Pancreatic Beta Cell Metabolism, Implications in Homeostasis. *Nutrients*. 2018;10(4):393.
66. Rice Bradley BH. Dietary Fat and Risk for Type 2 Diabetes: a Review of Recent Research. *Curr Nutr Rep*. 2018;7(4):214-226.
67. Narayan KM, Boyle JP, Thompson TJ, Gregg EW, Williamson DF. Effect of BMI on lifetime risk for diabetes in the U.S. *Diabetes Care*. 2007;30(6):1562-6.
68. Gray N, Picone G, Sloan F, Yashkin A. Relation between BMI and diabetes mellitus and its complications among US older adults. *South Med J*. 2015;108(1):29-36.
69. Pang M, Kaufman JS, Platt RW. Studying noncollapsibility of the odds ratio with marginal structural and logistic regression models. *Stat Methods Med Res*. 2016;25(5):1925-1937.
70. Vanderweele TJ, Vansteelandt S. Odds ratios for mediation analysis for a dichotomous outcome. *Am J Epidemiol*. 2010;172(12):1339-48.

Figure 1

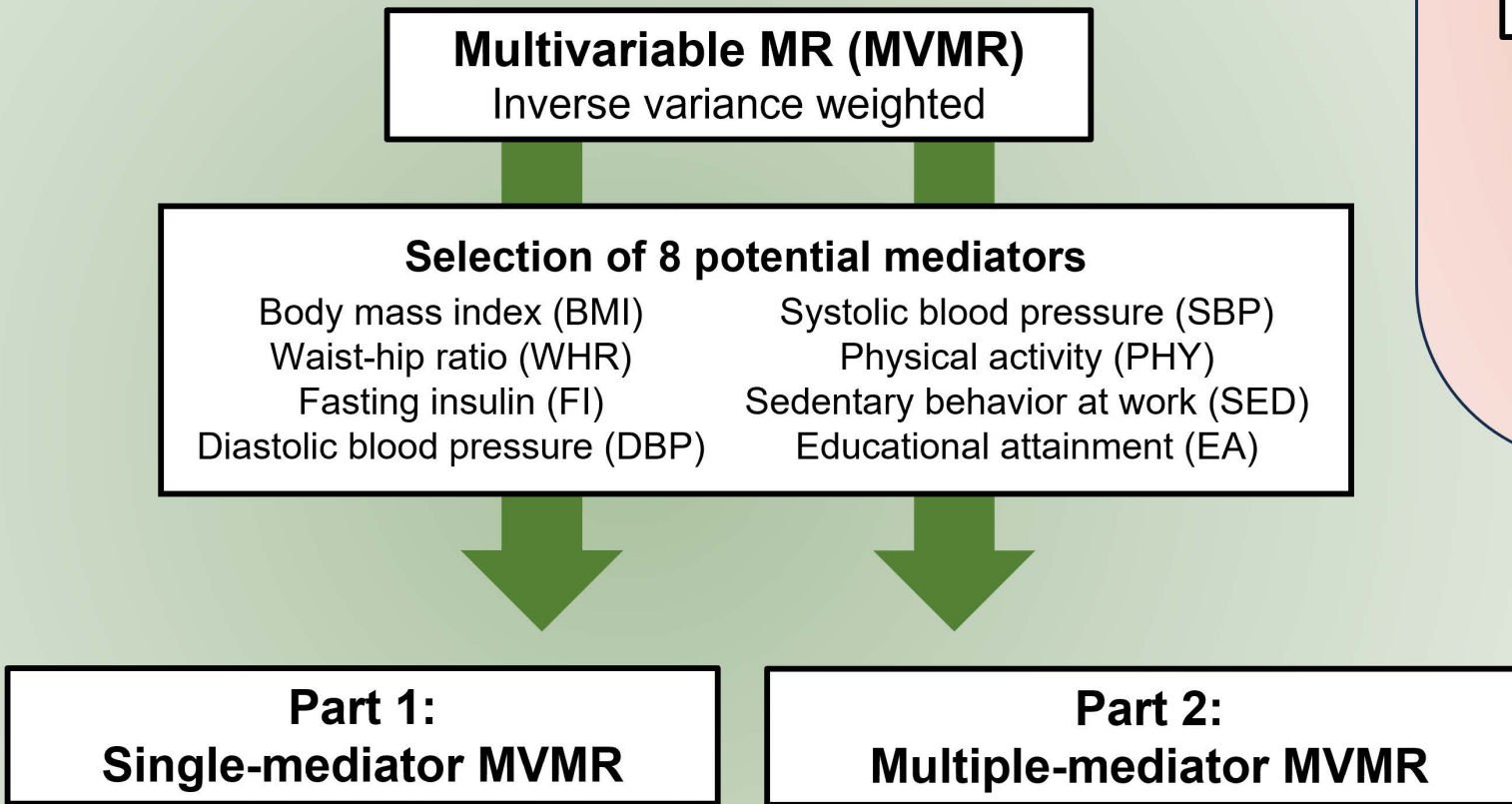
Step 1



Step 2



Step 3



Step 4

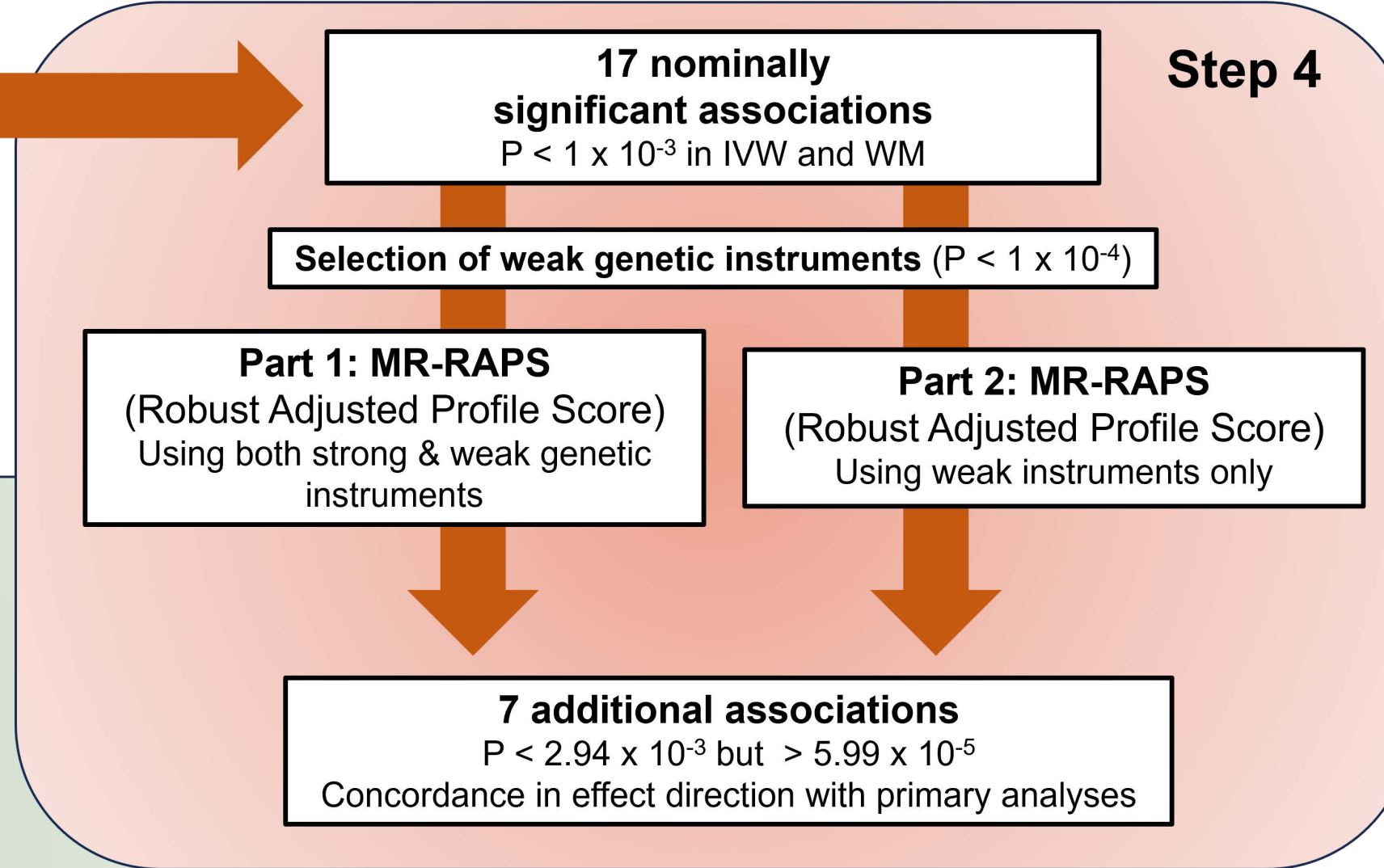
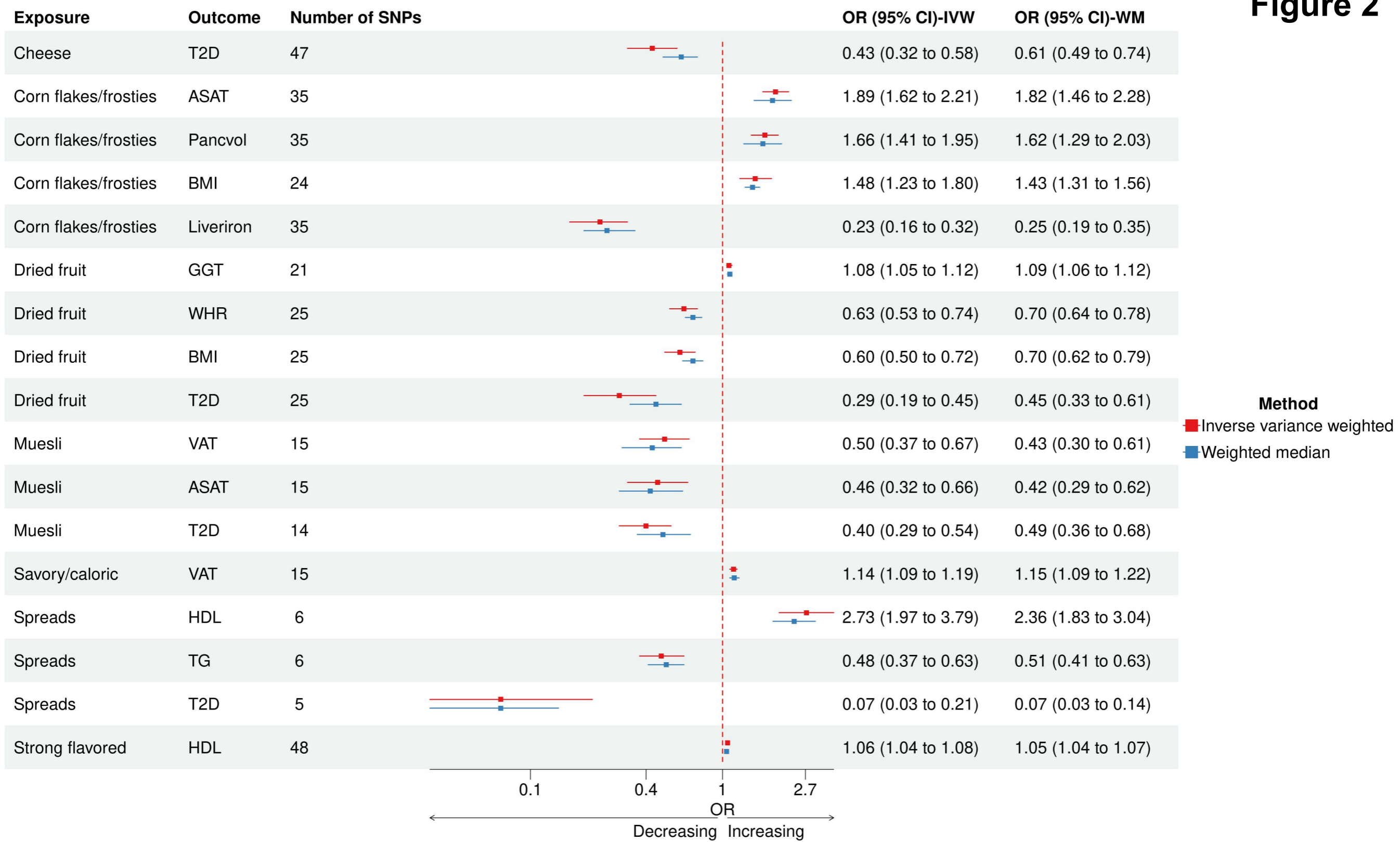
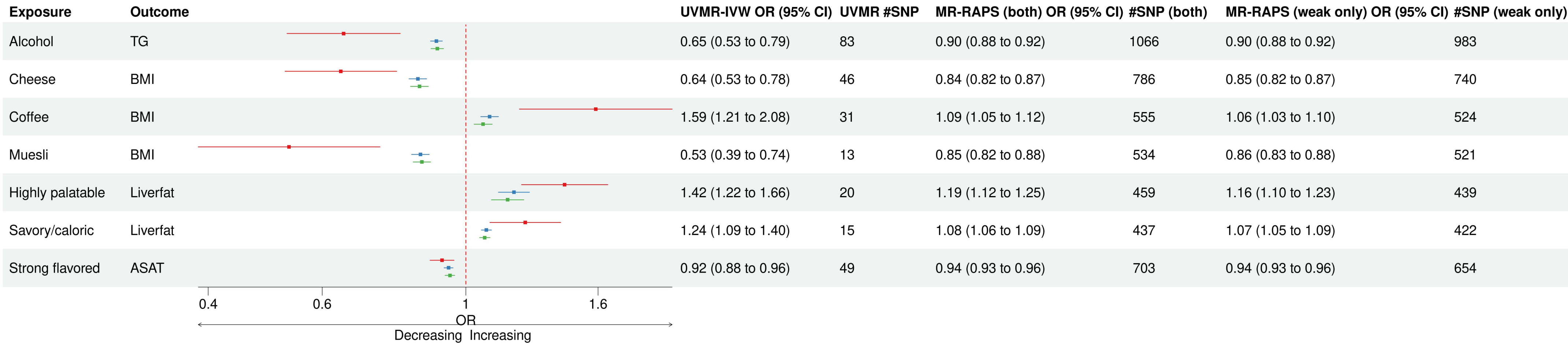


Figure 2





Method
■ UVMR-IVW
■ MR-RAPS (both)
■ MR-RAPS (weak only)

Figure 3

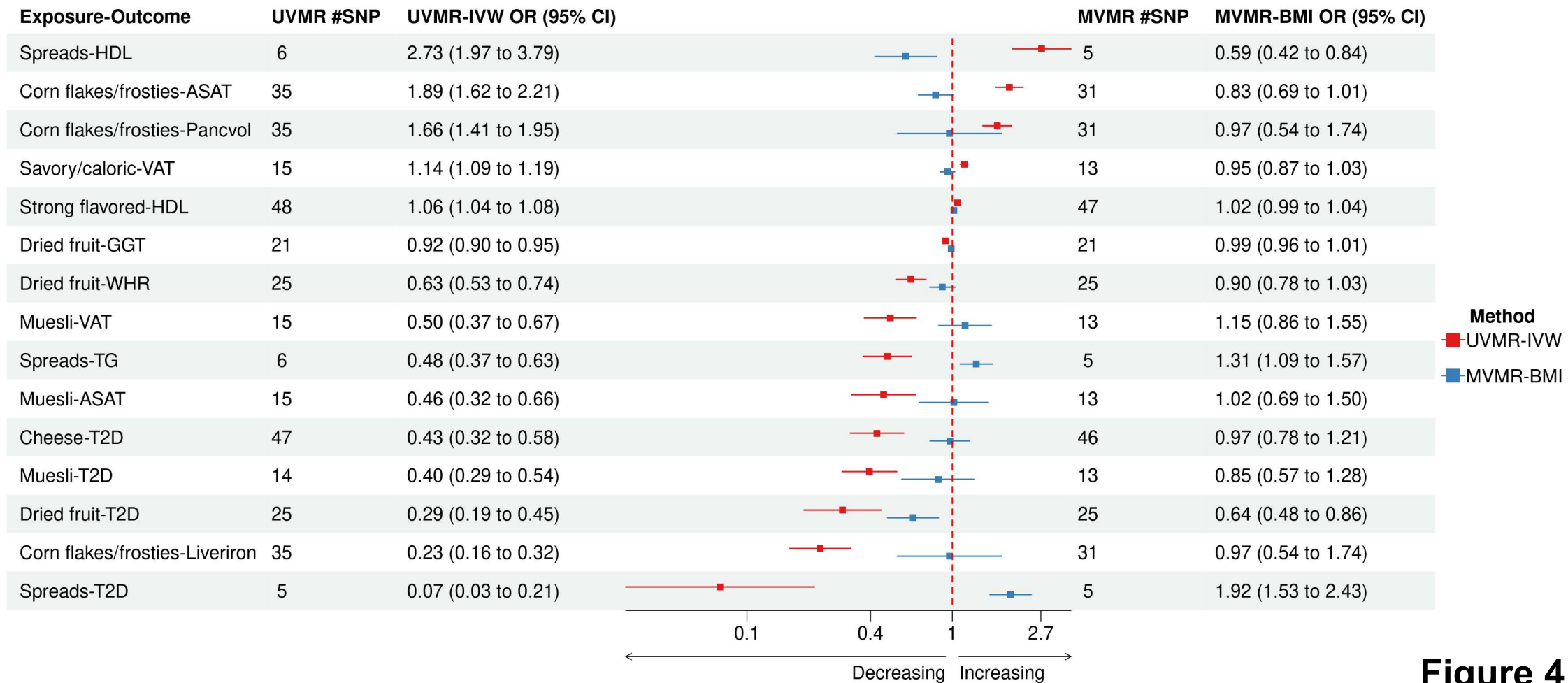


Figure 4

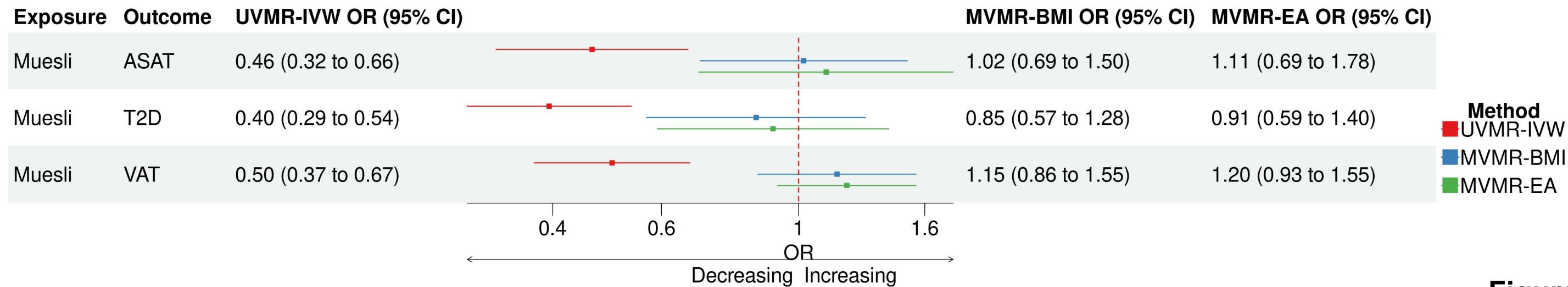


Figure 5

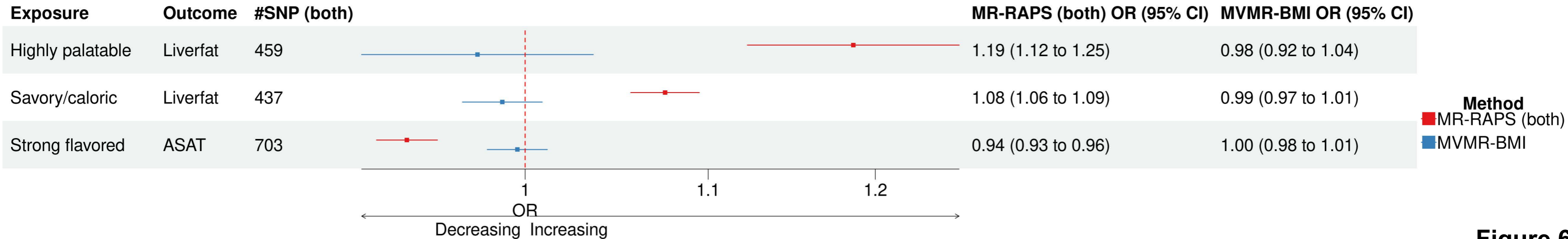


Figure 6

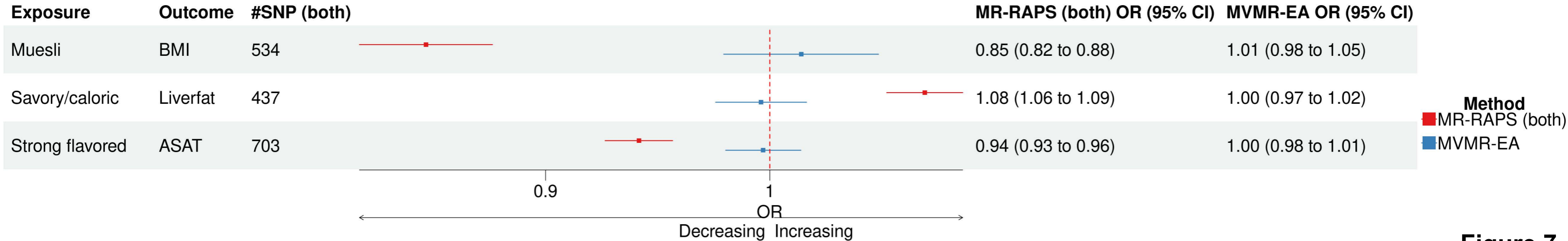


Figure 7

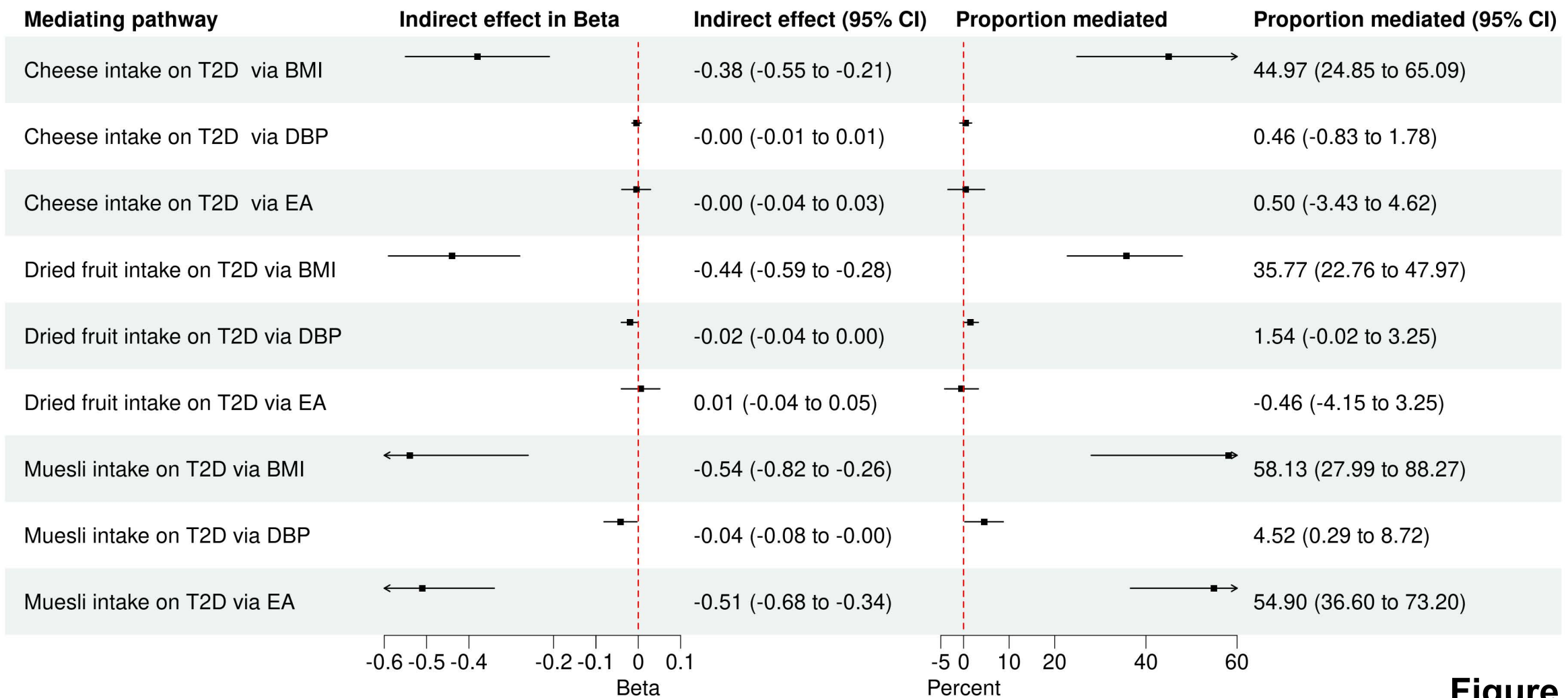


Figure 8