

## Investigating gene-diet interactions impacting the association between macronutrient intake and glycemic traits

Kenneth E. Westerman<sup>1,2,3</sup>, Maura E. Walker<sup>4,5</sup>, Sheila M. Gaynor<sup>6</sup>, Jennifer Wessel<sup>7,8,9</sup>, Daniel DiCorpo<sup>10</sup>, Jiantao Ma<sup>11</sup>, Alvaro Alonso<sup>12</sup>, Stella Aslibekyan<sup>13</sup>, Abigail S. Baldrige<sup>14</sup>, Alain G. Bertoni<sup>15</sup>, Mary L. Biggs<sup>16,17</sup>, Jennifer A. Brody<sup>17,18</sup>, Yii-Der Ida Chen<sup>19</sup>, José Dupuis<sup>10</sup>, Mark O. Goodarzi<sup>20</sup>, Xiuqing Guo<sup>19</sup>, Natalie R. Hasbani<sup>21</sup>, Adam Heath<sup>21</sup>, Bertha Hidalgo<sup>22</sup>, Marguerite R. Irvin<sup>23</sup>, W. Craig Johnson<sup>16</sup>, Rita R. Kalyani<sup>24</sup>, Leslie Lange<sup>25</sup>, Rozenn N. Lemaitre<sup>17,26</sup>, Ching-Ti Liu<sup>10,27,28,29</sup>, Simin Liu<sup>30</sup>, Jee-Young Moon<sup>31</sup>, Rami Nassir<sup>32</sup>, James S. Pankow<sup>33</sup>, Mary Pettinger<sup>34</sup>, Laura Raffield<sup>35</sup>, Laura J. Rasmussen-Torvik<sup>14</sup>, Elizabeth Selvin<sup>36</sup>, Mackenzie K. Senn<sup>37</sup>, Aladdin H. Shadyab<sup>38</sup>, Albert V. Smith<sup>39</sup>, Nicholas L. Smith<sup>40,41,42</sup>, Lyn Steffen<sup>33</sup>, Sameera Talegakwar<sup>43</sup>, Kent D. Taylor<sup>19</sup>, Paul S. de Vries<sup>21</sup>, James G. Wilson<sup>44</sup>, Alexis C. Wood<sup>37</sup>, Lisa R. Yanek<sup>24</sup>, Jie Yao<sup>19</sup>, Yinan Zheng<sup>14</sup>, Eric Boerwinkle<sup>21,45</sup>, Alanna C. Morrison<sup>21</sup>, Miriam Fornage<sup>21</sup>, Tracy P. Russell<sup>46</sup>, Bruce M. Psaty<sup>17,18,40,47</sup>, Daniel Levy<sup>27,48</sup>, Nancy L. Head-Costa<sup>27,49</sup>, Vasani S. Ramachandran<sup>27,28,29</sup>, Rasika A. Mathias<sup>24</sup>, Donna K. Arnett<sup>50</sup>, Robert Kaplan<sup>51</sup>, Kari E. North<sup>52</sup>, Adolfo Correa<sup>53</sup>, April Carson<sup>54</sup>, Jerome Rotter<sup>19</sup>, Stephen S. Rich<sup>55</sup>, JoAnn E. Manson<sup>2</sup>, Alexander P. Reiner<sup>40</sup>, Charles Kooperberg<sup>34</sup>, Jose C. Florez<sup>2,3,56,57</sup>, James B. Meigs<sup>2,3,58</sup>, Jordi Merino<sup>2,3,59</sup>, Deirdre K. Tobias<sup>60,61</sup>, Han Chen<sup>21,62</sup>, Alisa K. Manning<sup>1,2,3</sup>

^ = co-corresponding authors

### AFFILIATIONS

1. Clinical and Translational Epidemiology Unit, Mongan Institute, Massachusetts General Hospital, Boston, MA.
2. Department of Medicine, Harvard Medical School, Boston, MA.
3. Programs in Metabolism and Medical and Population Genetics, Broad Institute of MIT and Harvard, Boston, MA.
4. Department of Medicine, Section of Preventative Medicine, Boston University School of Medicine, Boston, MA.
5. Department of Health Sciences, Sargent College of Health and Rehabilitation Sciences, Boston University, Boston, MA.
6. Department of Biostatistics, Harvard T.H. Chan School of Public Health, Boston, MA.
7. Department of Epidemiology, Richard M. Fairbanks School of Public Health, Indianapolis, IN.
8. Department of Medicine, Indiana University School of Medicine, Indianapolis, IN.
9. Department of Translation Research Center, Indiana University School of Medicine, Indianapolis, IN.
10. Department of Biostatistics, Boston University School of Public Health, Boston, MA.
11. Nutrition Epidemiology and Data Science, Friedman School of Nutrition Science and Policy, Tufts University, Boston, MA, USA
12. Department of Epidemiology, Rollins School of Public Health, Emory University, Atlanta, GA.
13. 23andme, Sunnyvale, CA.
14. Department of Preventive Medicine, Northwestern University Feinberg School of Medicine, Chicago, IL.
15. Department of Epidemiology & Prevention, Wake Forest School of Medicine, Winston-Salem, NC.
16. Department of Biostatistics, University of Washington, Seattle, WA.
17. Cardiovascular Health Research Unit, University of Washington, Seattle, WA.

18. Department of Medicine, University of Washington, Seattle, WA.
19. The Institute for Translational Genomics and Population Sciences, Department of Pediatrics, The Lundquist Institute for Biomedical Innovation at Harbor-UCLA Medical Center, Torrance, CA.
20. Department of Medicine, Division of Endocrinology Diabetes and Metabolism, Cedars-Sinai Medical Center, Los Angeles, CA.
21. Department of Epidemiology, Human Genetics, and Environmental Sciences, School of Public Health, The University of Texas Health Science Center at Houston, Houston, TX.
22. Ryals School of Public Health, University of Alabama at Birmingham, Birmingham, AL.
23. Department of Epidemiology, University of Alabama at Birmingham, Birmingham, AL.
24. GeneSTAR Research Program, Johns Hopkins University School of Medicine, Baltimore, MD.
25. Department of Medicine, Anschutz Medical Campus, University of Colorado Denver, Aurora, CO.
26. Department of Internal Medicine, University of Washington, Seattle, WA.
27. National Heart Lung and Blood Institute and Boston University's Framingham Heart Study, Framingham, MA.
28. Evans Department of Medicine, Section of Preventive Medicine and Epidemiology, Boston University School of Medicine, Boston, MA.
29. Evans Department of Medicine, Whitaker Cardiovascular Institute and Cardiology Section, Boston University School of Medicine, Boston, MA.
30. Center for Global Cardiometabolic Health (CGCH), Boston, MA.
31. Department of Epidemiology and Population Health, Albert Einstein College of Medicine, Bronx, NY.
32. Department of Pathology, School of Medicine, Umm Al-Qura'a University, Mecca, Saudi Arabia.
33. Division of Epidemiology and Community Health, School of Public Health, University of Minnesota, Minneapolis, MN.
34. Division of Public Health Sciences, Fred Hutchinson Cancer Research Center, Seattle, WA.
35. Department of Genetics, University of North Carolina, Chapel Hill, NC.
36. Department of Epidemiology, Johns Hopkins Bloomberg School of Public Health, Baltimore, MD.
37. USDA/ARS Children's Nutrition Research Center, Baylor College of Medicine, Houston, TX.
38. Herbert Wertheim School of Public Health and Human Longevity Science, University of California, San Diego, La Jolla, CA.
39. Department of Biostatistics, University of Michigan, Ann Arbor, MI.
40. Department of Epidemiology, University of Washington, Seattle, WA.
41. Kaiser Permanente Washington Health Research Institute, Seattle, WA.
42. Department of Veterans Affairs Office of Research and Development, Seattle Epidemiologic Research and Information Center, Seattle, WA.
43. Department of Exercise and Nutrition Sciences, Milken Institute School of Public Health, The George Washington University, Washington, District of Columbia.
44. Division of Cardiovascular Medicine, Beth Israel Deaconess Medical Center, Boston, MA.

45. Human Genome Sequencing Center, Baylor College of Medicine, Houston, TX.
46. Department of Pathology and Laboratory Medicine, University of Vermont Larner College of Medicine, Burlington, VT.
47. Department of Health Systems and Population Health, University of Washington, Seattle, WA.
48. The Population Science Branch, Division of Intramural Research, National Heart, Lung, and Blood Institute, Bethesda, MD.
49. Department of Neurology, Boston University School of Medicine, Boston, MA.
50. College of Public Health, University of Kentucky, Lexington, KY.
51. Clinical Excellence Research Center, School of Medicine, Stanford University, Stanford, CA.
52. Department of Epidemiology, University of North Carolina, Chapel Hill, NC.
53. Department of Population Health Science, University of Mississippi Medical Center, Jackson, MS
54. Department of Medicine, University of Mississippi Medical Center, Jackson, MS.
55. Center for Public Health Genomics, Department of Public Health Sciences, University of Virginia, Charlottesville, VA.
56. Diabetes Unit, Massachusetts General Hospital, Boston, MA.
57. Center for Genomic Medicine, Massachusetts General Hospital, Boston, MA.
58. Division of General Internal Medicine, Massachusetts General Hospital, Boston, MA.
59. Diabetes Unit and Center for Genomic Medicine, Massachusetts General Hospital, Boston, MA.
60. Department of Medicine, Division of Preventive Medicine, Brigham and Women's Hospital and Harvard Medical School, Boston, MA.
61. Department of Nutrition, Harvard T.H. Chan School of Public Health, Boston, MA.
62. Center for Precision Medicine, School of Public Health, The University of Texas Health Science Center, Houston, TX.

## **Conflicts of Interest**

KEW has provided consulting services for FOXO BioScience. LMR is a consultant for the TOPMed Administrative Coordinating Center (through Westat).

## **Funding**

KEW, HC, and AKM were supported by NIH R01 HL145025. LMR was supported by the National Center for Advancing Translational Sciences, National Institutes of Health, through Grant KL2TR002490 (LMR). VSR is supported in part by the Evans Medical Foundation and the Jay and Louis Coffman Endowment from the Department of Medicine, Boston University School of Medicine.

## **Corresponding Authors:**

Kenneth E. Westerman

[kewesterman@mgh.harvard.edu](mailto:kewesterman@mgh.harvard.edu)

Alisa K. Manning

[akmanning@mgh.harvard.edu](mailto:akmanning@mgh.harvard.edu)

1 **ABSTRACT**

2  
3 **Background:** Heterogeneity in the long-term metabolic response to dietary macronutrient  
4 composition can be partially explained by genetic factors. However, few studies have  
5 demonstrated reproducible gene-diet interactions (GDIs), likely due in part to measurement error  
6 in dietary intake estimation as well as insufficient capture of rare genetic variation. Discovery  
7 analyses in ancestry-diverse cohorts that include rare genetic variants from whole-genome  
8 sequencing (WGS) could help identify genetic variants modifying the effects of dietary  
9 macronutrient composition on glycemic phenotypes.

10  
11 **Objective:** We aimed to identify macronutrient GDIs across the genetic frequency spectrum  
12 associated with continuous glycemic traits in genetically and culturally diverse cohorts.

13  
14 **Methods:** We analyzed N=33,187 diabetes-free participants from 10 cohorts in the NHLBI  
15 Trans-Omics for Precision Medicine (TOPMed) program with WGS, self-reported diet, and  
16 glycemic traits (fasting glucose [FG], insulin [FI], and hemoglobin A1c [HbA1c]). We fit  
17 multivariable-adjusted linear mixed models for the main effect of diet, modeled as an isocaloric  
18 substitution of carbohydrate for fat, and for its interactions with genetic variants genome-wide.  
19 Tests were performed for both common variants and gene-based rare variant sets in each cohort  
20 followed by a combined cohort meta-analysis.

21  
22 **Results:** In main effect models, participants consuming more calories from carbohydrate at the  
23 expense of fat had modestly lower glycemic trait values ( $\beta$  per 250 kcal substitution for FG: -  
24 0.030 mmol/L,  $p=2.7\times 10^{-6}$ ; lnFI: -0.008 log(pmol/L),  $p=0.17$ ; HbA1c: -0.013 %,  $p=0.025$ ). In  
25 GDI analyses, a common African ancestry-enriched variant (rs79762542; 78 kb upstream of the  
26 *FRAS1* gene) reached study-wide significance ( $p = 1.14\times 10^{-8}$ ) indicating a higher HbA1c with  
27 greater proportion of calories from carbohydrate vs. fat among minor allele carriers only. This  
28 interaction was replicated in the UK Biobank cohort. Simulations revealed that there is (1) a  
29 substantial impact of measurement error on statistical power for GDI discovery at these sample  
30 sizes, especially for rare genetic variants, and (2) over 150,000 samples may be necessary to  
31 identify similar macronutrient GDIs under realistic assumptions about effect size and  
32 measurement error.

33  
34 **Conclusions:** Our analysis identified a potential genetic interaction modifying the dietary  
35 macronutrient-HbA1c association while highlighting the importance of precise exposure  
36 measurement and significantly increased sample size to identify additional similar effects.

## 39 INTRODUCTION

40  
41 Diet is an established modifiable factor associated with risk of type 2 diabetes (T2D) and related  
42 cardiometabolic diseases(1). However, evidence is mixed regarding the ideal dietary  
43 macronutrient composition for risk reduction. Dietary interventions with differing proportions of  
44 energy from carbohydrates versus fat have shown varied efficacy for T2D risk reduction and  
45 substantial between-person heterogeneity in effects on cardiometabolic risk factors(2–4).  
46 Further, acute glycemic responses to meals with specific macronutrient composition are  
47 reproducible within individuals(5,6). Genetically different mouse strains have varying sensitivity  
48 of glycemic biomarkers to a high-fat diet(7) and to human-relevant dietary patterns(8).  
49 Retrospective analysis of human trials manipulating macronutrient intake have found genetic  
50 modifiers of glycemic response(9). Taken together, such studies suggest that genetics could be a  
51 key contributor to variability in the association between dietary macronutrient composition and  
52 glycemic health.

53  
54 Gene-diet interaction (GDI) studies aim to identify genetic variants that modify the association  
55 between dietary behaviors and health. Furthermore, GDI studies support differential associations  
56 of dietary factors with glycemic traits according to genotypes, using both hypothesis-driven(10)  
57 and hypothesis-free(11,12) strategies. However, in general, discovery and replication of gene-  
58 diet interactions with T2D risk has been poor, possibly due to measurement error in assessing  
59 habitual diet, low statistical power for interaction analysis, and biological and behavioral  
60 heterogeneity across populations(13). Additionally, to date there has been little exploration of  
61 GDIs involving rare genetic variants, which affect a smaller proportion of the population but  
62 may have larger effect sizes(14).

63  
64 Our primary aim was to discover novel putative genetic modifiers for the association between  
65 dietary macronutrient composition and glycemic traits. To this end, we performed a GDI analysis  
66 using common and rare genetic variants in over 30,000 individuals with diverse ancestral  
67 backgrounds from the NHLBI Trans-Omics for Precision Medicine (TOPMed) program. We  
68 focused on modeling a dietary carbohydrate-fat exchange, which can be reasonably assessed via  
69 self-reported diet questionnaires and can be straightforwardly modified in the context of a  
70 healthful diet. Furthermore, the use of whole-genome sequencing (WGS) permitted the analysis  
71 of rare variants using set-based association tests. As a secondary aim, we sought to inform  
72 subsequent GDI research by exploring the impact of dietary exposure measurement error on  
73 statistical power in the context of realistic effect size estimates.

## 74 75 RESULTS

76  
77 We analyzed data from 33,178 individuals without diabetes (based on FG, HbA1c, or medication  
78 use) from 10 TOPMed program cohorts. Participants had diverse cohort-reported race/ethnicities  
79 including: African American (N = 6,158), American Indian (N = 35), Asian (N = 124), white (N  
80 = 19,721), and Hispanic/Latino (N = 7,114). Dietary carbohydrate and fat as a percentage of total  
81 energy intake on average were 50.5% (standard deviation = 8.5%) and 32.2% (6.9%),  
82 respectively, in the full pooled sample, estimated using validated food frequency questionnaires  
83 (FFQ) or 24-hour dietary recalls (24HR). Cohort-specific carbohydrate intake estimates (as  
84 percent of total energy [% kcal]), glycemic trait values (fasting glucose [FG], fasting insulin [FI;

85 or log-transformed, lnFI], and hemoglobin A1c [HbA1c]), and additional population  
86 characteristics are presented in Supp. Table S1 and Supp. Fig. S1.

87  
88 We first modeled the main association of macronutrient compositions with each of the glycemic  
89 traits. By adjusting for total energy and energy from protein, resulting regression estimates for  
90 carbohydrate represented a macronutrient exchange (increased 250kcal from carbohydrate  
91 replacing an equivalent 250kcal from fat; see Methods). Meta-analysis of the individual cohorts  
92 indicated that a higher proportion of kcal from carbohydrate at the expense of fat was associated  
93 with lower FG ( $-0.030$  mmol/L / 250kcal;  $p = 2.2 \times 10^{-6}$ ), lnFI ( $-0.008$  log(pmol/L) / 250kcal;  $p$   
94 = 0.15), and HbA1c ( $-0.012$  %HbA1c / 250kcal;  $p = 0.029$ ). Forest plots of these results are  
95 shown in Supp. Fig. S2.

### 96 97 **Common variant gene-diet interactions**

98  
99 We sought to identify macronutrient GDIs with the maximal sample available in TOPMed  
100 program cohorts in order to provide a baseline for discovery and evaluate our assumptions about  
101 expected effect sizes. Common variants (MAF > 1%) were analyzed in a primary, single-variant  
102 analysis of gene-carbohydrate interactions, with the same regression adjustments as above. This  
103 GDI analysis produces interaction estimates for the difference in the macronutrient-glycemic  
104 trait association per alternate allele at the variant of interest. After genome-wide, cohort-specific  
105 analysis and cross-cohort meta-analysis, one variant reached a study-wide significance threshold  
106 of  $1.67 \times 10^{-8}$  ( $5 \times 10^{-8}$  / 3 glycemic traits). Two additional variants passed a standard genome-wide  
107 threshold of  $5 \times 10^{-8}$  (Table 1). We note that this threshold is liberal given the greater testing  
108 burden involved in the analysis of multiple ancestry groups(15). Of these three, none had  
109 evidence of a genetic main effect on the associated trait. Results are visualized in Supp. Fig. S3  
110 for all variants and shown in Supp. Table S3 for variants with interaction  $p < 10^{-5}$ .

111  
112 As the only variant reaching study-wide significance in the primary analysis, we looked deeper  
113 into the biological function of rs79762542 and the functional form of its interaction. Variant  
114 rs79762542 is observed on African-ancestry haplotypes and was discovered with respect to  
115 HbA1c. The variant does not have known regulatory activity based on epigenomic assays in  
116 RegulomeDB, but there is evidence for a role in regulating expression of the nearby gene  
117 *FRAS1*, especially in thyroid, where this gene is most strongly expressed (GTEx project).  
118 Colocalization analysis did not support a shared causal signal between our interaction results and  
119 thyroid-specific eQTL signal (posterior probability of shared causal variant = 0.003%).

120  
121 In genotype-stratified meta-analysis, HbA1c showed a modest negative association with  
122 increasing carbohydrate relative to fat intake in major allele homozygotes ( $-0.033$   
123 %HbA1c/250kcal;  $p = 0.004$ ) versus a positive association in minor allele carriers ( $0.10$   
124 %HbA1c/250kcal;  $p = 0.42$ ) that may not have reached significance due to the much lower  
125 sample size in this group (N = 1,055 across all studies). This genetic effect modification was  
126 moderately consistent across the cohorts in which HbA1c levels were studied and the minor  
127 allele of rs79762542 was observed (Fig. 1a,b; identical visualization in the African American  
128 race/ethnicity subset in Supp. Fig. S4a,b). Finally, with respect to the other glycemic traits in our  
129 analysis, interaction effects were directionally consistent but did not reach nominal significance

130 (0.02 mmol/L/allele/250kcal;  $p = 0.07$  for FG and 0.003 log(pmol/L)/allele/250kcal;  $p = 0.74$   
131 for lnFI).

132

133 Lookups for the other two variants passing  $p < 5 \times 10^{-8}$  revealed potential functional roles for  
134 these variants. Variant rs1288694 (common in multiple ancestries) impacted FG in our analysis.  
135 The variant is intronic to the *FOXPI* gene and may regulate splicing of the same gene (GTEx  
136 project). *FOXPI* has a demonstrated role in hepatic gluconeogenesis(16). Variant rs782681704 is  
137 observed on African-ancestry haplotypes and was discovered with respect to FI in our analysis.  
138 The variant is intronic to *BRCC3* and has likely regulatory activity (RegulomeDB score of 0.59)  
139 but does not have clear evidence as an eQTL for *BRCC3*.

140

141 We explored these three prioritized single-variant loci through sensitivity analysis (Supp. Figs.  
142 S5 [rs79762542] and S6 [all three variants]). Interaction effects were robust in population  
143 subsets: only males, only females, and individuals without obesity. Exclusion of individuals  
144 either with or without prediabetes (beyond the predefined exclusion of individuals with diabetes)  
145 partially attenuated the interaction signal; this might be expected due to the removal of a  
146 substantial portion of the glycemic trait spectrum. Further, adjustment for either a diet quality  
147 score (AHEI-2010) or smoking status (along with their genotype interactions) did not  
148 meaningfully impact estimates. Interaction estimates were also generally consistent in the  
149 African American race/ethnicity subset, indicating that the interactions for African ancestry-  
150 specific variants do not solely reflect population stratification.

151

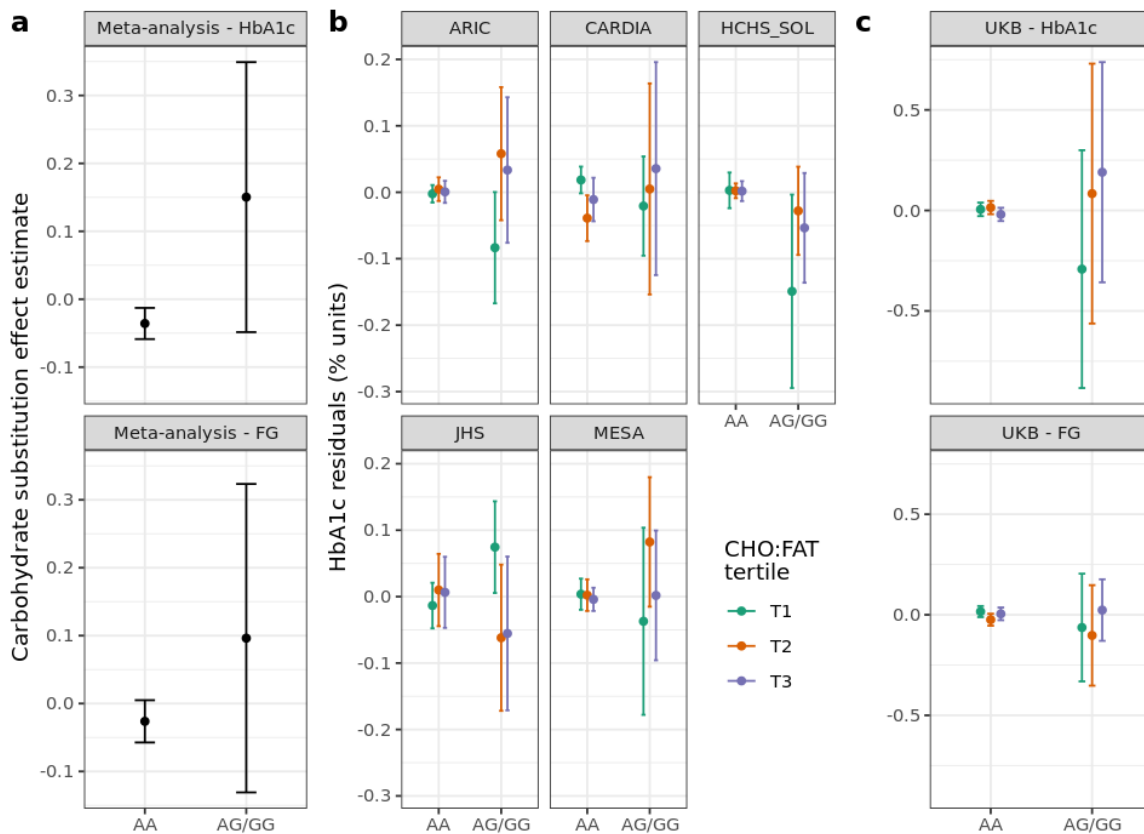
### 152 **Common variant replication**

153

154 For the three prioritized single-variant loci, we tested for replication of these signals in the UK  
155 Biobank (N=178,352 with 24-hour dietary assessment data(17) and glycemic biomarkers; see  
156 Methods). Of these, 5,183 individuals were included in fasting glucose analyses (based on  
157 fasting for at least eight hours prior to the associated blood draw). In the full, multi-ancestry  
158 group (Supp. Table S5), we saw nominal replication of the interaction at rs79762542 with  
159 respect to both HbA1c (the discovery trait;  $p = 0.025$ ) and fasting glucose ( $p = 0.013$ ; Fig. 1c).  
160 Because most of the prioritized variants were specific to African-ancestry individuals, we  
161 conducted a similar replication in just this subgroup of the UK Biobank (Supp. Table S6). This  
162 analysis supported the rs79762542 interaction influencing FG ( $p = 0.046$ ).

163





164  
 165 **Figure 1:** Exploration of the rs79762542 interaction and replication. (a) Genotype-stratified  
 166 dietary main effect estimates. (b) Stratified plots (one for each cohort with HbA1c available)  
 167 display residualized HbA1c within strata defined by both genotype at rs79762542 (none vs. any  
 168 minor alleles) and cohort-specific tertile of carbohydrate:fat ratio. (c) Similar stratified plots for  
 169 the UKB replication cohort. For (b) and (c), the y-axis displays residuals after regressing the  
 170 relevant trait (HbA1c or FG) on the set of covariates used in the replication analysis.

171  
 172 **Rare variant interactions**

173  
 174 Rare variants (MAF < 1%) were analyzed in gene-centric, set-based tests, which help to  
 175 overcome power limitations for low-MAF variants by aggregating signal across multiple variants  
 176 annotated to the same gene. We used three variant aggregation strategies to define sets: selecting  
 177 missense variants, loss-of-function variants, or a broader coding + non-coding variant set  
 178 annotated to each gene (see Methods). No rare variant interaction signals showed genome-wide  
 179 significance ( $p < 0.05 / 28,111$  total genes =  $1.78 \times 10^{-6}$ ; Supp. Table S4; Supp. Fig. S7).

180  
 181 Since the set of rare variants used does not overlap with those from the common-variant tests,  
 182 these gene-based tests can provide orthogonal evidence supporting common-variant signals  
 183 while further clarifying potential effector genes. Each of the three prioritized single-variant  
 184 findings were annotated to one or more genes based on proximity and/or expression-quantitative  
 185 trait locus (eQTL) data. None of these pairings showed supporting gene-based signal for the  
 186 corresponding glycemic trait, though the single study-wide significant variant (rs79762542,

187 discovered in relation to HbA1c) showed a nominal corresponding signal from the gene-based  
188 test of *FRAS1* impacting FG ( $p = 0.028$ ).

189

### 190 **Power calculations incorporating measurement error**

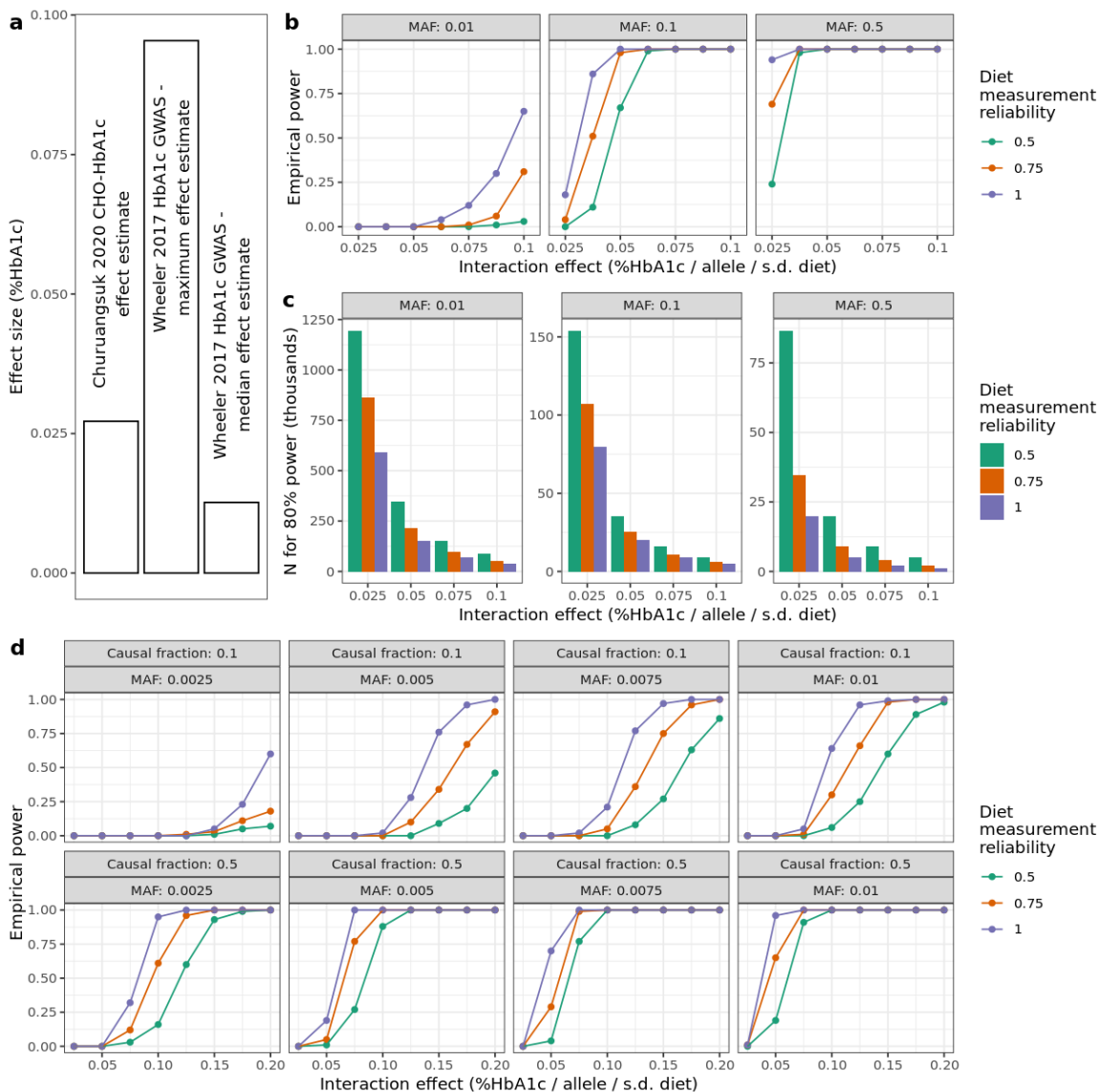
191

192 Given the modest discovery of GDIs despite the use of the maximal sample available within  
193 TOPMed cohorts and substantial harmonization effort, we sought to better understand the  
194 necessary power to detect expected GDI effects. Simulation-based power calculations for single-  
195 variant tests were conducted with added noise to account for the known random measurement  
196 error in dietary data. Using literature-based anchors for expected effect sizes (Fig. 2a) and  
197 assuming a conservative but realistic dietary measurement reliability of 0.5 (see Methods for  
198 details), we established that a sample size of over 150,000 would be required to detect a GDI  
199 effect of 0.025 %HbA1c / allele / s.d. carbohydrate at genome-wide significance for a variant  
200 with MAF = 0.1 (Fig. 2b,c). As previously explored in the literature(18,19), power scaled  
201 approximately linearly with exposure measurement fidelity. If we alternatively assume perfect  
202 dietary exposure measurement, the associated sample size to detect the same effect was reduced  
203 to about 80,000 indicating the importance of accounting for this measurement error. The  
204 necessary sample size given realistic measurement reliabilities increased even further for lower-  
205 frequency variants (e.g., almost 1.2 million for MAF = 1%).

206

207 We extended this simulation-based power calculation approach to test multiple variants jointly,  
208 mimicking the variant set-based test implemented for rare variants. Assuming similar  
209 measurement fidelity and effect sizes as for single variants and fixing the sample size to match  
210 the full sample used here (~35,000), an aggregate test of 20 rare variants with a causal fraction of  
211 50% and MAF = 0.25% had negligible power (Fig. 2d). Power increased somewhat but remained  
212 low when incorporating larger effect sizes (as are known to be present for rare variant main  
213 effects on cardiometabolic traits)(20). For example, using an effect size of 0.1, approximately  
214 equal to the largest genetic main effect on HbA1c reported by Wheeler and colleagues(21),  
215 power increased to 0.16.

216



217  
 218 **Figure 2:** Power calculations for GEI incorporating exposure measurement error. For all plots  
 219 above, HbA1c is used as a basis for parameter choices. (a) Genetic and dietary effect sizes on  
 220 HbA1c for reference for potential interaction effects. (b) Simulation-based empirical power  
 221 estimates are shown as a function of the interaction effect ( $x$ -axis), minor allele frequency (MAF;  
 222 panels left-to-right), and diet measurement reliability (colors). (c) Bar plots show the estimated  
 223 sample size needed to achieve 80% statistical power. Panels and colors are as in (b). (d) As in (a)  
 224 but modeling empirical power for simulated aggregate tests of 20 rare variants with a causal  
 225 fraction of 0.1 or 0.5 (indicated in panel labels). Additional assumptions for these simulations  
 226 (full details in Methods):  $N = 35,000$ , phenotype mean of 5.5, phenotype standard deviation of  
 227 0.5, exposure mean of 0, exposure standard deviation of 1, genetic main effect of 0.01, and  
 228 environmental main effect of 0.2.

229  
 230  
 231

## 232 DISCUSSION

233

234 Our goal was to investigate genotype-related variability in the association of dietary  
235 macronutrient composition with glycemic traits. Importantly, this was based on a regression  
236 strategy modeling an isocaloric increase in dietary carbohydrate at the expense of fat(22). We  
237 conducted our comprehensive analyses in cohorts with racial/ethnic diversity with data for both  
238 common and rare variants from WGS. We examined multiple single variants with potential  
239 modifying roles for the relationship of carbohydrate intake with glycemic traits, but did not find  
240 substantial evidence from gene-based tests for a role of rare variants in modifying this diet-  
241 glycemia relationship. Furthermore, our simulation-based power analysis highlighted the impact  
242 of dietary measurement error on statistical power for the gene-diet interaction tests.

243

244 Dietary main effect models indicated that an increase in carbohydrate at the expense of dietary  
245 fat was associated with lower fasting glucose and HbA1c. The impact of this macronutrient  
246 exchange on glucose homeostasis and diabetes risk is complex and likely depends on the  
247 respective macronutrient quality. Prior studies suggest null associations of total carbohydrate to  
248 total fat exchange on diabetes risk(23,24). However, an exchange of animal-sourced fat for  
249 carbohydrate or vegetable fat appears to have favorable associations with HbA1c(25,26).

250

251 Our primary genome-wide common variant interaction analysis yielded an interaction between a  
252 250kcal carbohydrate-fat substitution and HbA1c with the African-ancestry rs79762542 variant,  
253 which was validated in the UK Biobank. Genotype stratified analyses suggested that minor allele  
254 carriers generally had a small negative association between carbohydrate and HbA1c (-0.033  
255 %HbA1c/250kcal;  $p = 0.004$ ) versus a larger but non-significant association in minor allele  
256 carriers where the sample size was much lower (-0.10 %HbA1c/250kcal;  $p = 0.42$ ). However,  
257 this precise pattern was not observed in all cohorts. These results warrant further exploration in  
258 additional cohorts with African ancestry individuals and dietary intervention studies to examine  
259 whether glycemic traits among minor allele carriers may benefit from higher fat and lower  
260 carbohydrate diet composition. This primary discovery was made with HbA1c as an outcome,  
261 but our results from set-based rare variant analysis and the UK Biobank replication suggest  
262 similar patterns with respect to other traits such as FG. Beyond GDI discovery, the GWIS results  
263 provided an opportunity to evaluate the effect size assumptions used in the power calculations.  
264 For example, the rs79762542 interaction had an effect size of 0.048 %HbA1c/allele/250 kcal  
265 carbohydrate, or 0.068 %HbA1c/allele/std. dev. carbohydrate. This effect size is somewhat larger  
266 than, but of a similar magnitude to, the relevant anchor for the power analysis (the referenced  
267 main effect association of carbohydrate with HbA1c(27)).

268

269 This analysis leveraged WGS data along with multi-variant set-based tests to better incorporate  
270 rare variants (minor allele frequency [MAF] < 0.01). While these variants don't contribute  
271 substantially to the overall population variance of glycemic or other traits, they tend to have  
272 larger effect sizes and thus may be important for the specific individuals carrying them(14). For  
273 example, phenylketonuria, a well-known inborn error of metabolism, acts through a rare-variant  
274 GDI in which severe adverse effects of phenylalanine intake are seen only in individuals with a  
275 particular genotype(28). In our study, despite helping to reinforce common variant signals, the  
276 rare variant analysis did not contribute additional findings after aggregation at the gene level.

277 Substantially larger sample sizes will likely be necessary to uncover macronutrient GDIs  
278 involving rare variants.

279  
280 We explored the statistical power for these interaction tests through simulations incorporating  
281 random dietary measurement error using available simulation-based power calculation software  
282 (ESPRESSO.GxE(29), for single variants) with additional extensions to allow for aggregate rare  
283 variant tests. We estimated that substantially higher sample sizes (almost five times that used in  
284 this study) are required for sufficient power to detect macronutrient-gene interactions at expected  
285 effect sizes obtained from genetics and nutrition literature. This prompts two directions of further  
286 inquiry. First, it suggests the importance of complementary approaches that assess where there is  
287 any whole-genome contribution to the diet-glycemia association, at least in observational  
288 datasets. These whole-genome analyses trade resolution for statistical power(30) and have a  
289 precedent for GDIs in smaller study samples(31). Second, it reinforces the importance of  
290 collecting dietary intake data in the growing group of large-scale biobanks and cohorts.  
291 Improvements in study design, data collection methods, and analysis that can improve quality of  
292 dietary assessments are also warranted. For example, conducting rigorous validation studies of  
293 the data collection tools and approaches and ascertaining repeated dietary data can greatly  
294 improve the precision of these measurements on a population level. Advancements in  
295 objectively quantifying habitual diet from biospecimen samples are also underway and have  
296 potential to improve discovery for genetic analyses.

297  
298 An important strength of this study is the breadth of ethnic and cultural diversity of the sample  
299 (increasing the likelihood that findings are robust) and of genetic variation (with WGS data  
300 enabling exploration of ancestry-specific genetic variation across the frequency spectrum). We  
301 also conducted a systematic investigation into the available statistical power while incorporating  
302 both realistic degrees of measurement error and evidence-based estimations of realistic effect  
303 sizes for gene-macronutrient interactions. However, the diversity of the included study sample  
304 also introduces heterogeneity that may be problematic. For example, the cohorts used different  
305 dietary assessment tools to capture habitual intake, leading to differences in the degree and  
306 direction of random and systematic measurement error. This is compounded by general,  
307 culturally-driven differences in food intake across race/ethnicity groups. Further heterogeneity  
308 arises from the time of data collection; the perceptions of carbohydrate intake have trended as  
309 more and less healthful in recent decades, potential resulting in differential confounding between  
310 diet and other health-related behaviors depending on the time of data collection(32). Future work  
311 can step beyond broad macronutrient categories by harmonizing intakes of specific foods or  
312 dietary patterns and analyzing macronutrient subtypes (e.g., added sugar and specific fatty acids).  
313 These approaches, combined with improved methods for detecting rare variant gene-  
314 environment interactions, will help utilize the increasing volume of WGS data to discover new  
315 GDIs relevant for metabolic disease risk.

## 316 317 **METHODS**

### 318 **Whole-genome sequencing**

319 All genetic and phenotypic data used for this study were obtained from the NCBI Database of  
320 Genotypes and Phenotypes (dbGaP) and the research was approved under the Mass General  
321 Brigham IRB (protocol 2017P000531). WGS was conducted through the NHLBI TOPMed

322 program (Freeze 8 data release). Sequencing and alignment to the GRCh38 reference genome  
323 was performed across six centers across the US: Broad Institute of MIT and Harvard, Northwest  
324 Genomics Center, New York Genome Center, Illumina Genomic Services, PSOMAGEN  
325 (formerly Macrogen), Baylor College of Medicine Human Genome Sequencing Center,  
326 McDonnell Genome Institute (MGI) at Washington University. Data harmonization and joint  
327 variant discovery and genotype calling were performed within the TOPMed Informatics  
328 Research Center at the University of Michigan. Sequence quality control filters were as follows:  
329 estimated DNA sample contamination below 10% and at least 95% of the genome having  
330 coverage of at least 10x. After genotyping, variants were further filtered for Mendelian  
331 inconsistency (based on a support vector machine classifier) and excess heterozygosity.  
332 Additional sample quality control was performed within the Data Coordinating Center at the  
333 University of Washington, including: matching sex as annotated and inferred from WGS,  
334 concordance of WGS genotypes with prior array-based “fingerprints”, and agreement of inferred  
335 relatedness with expectations based on pedigrees. Additional details on the processing steps are  
336 available at: <https://www.nhlbiwgs.org/topmed-whole-genome-sequencing-methods-freeze-8>.

337 Global measures of ancestry and relatedness were calculated on the entire TOPMed Freeze 8  
338 sample by the TOPMed Data Coordinating Center. Genetic principal components reflecting  
339 ancestry were calculated using the PC-AiR method (allowing for related individuals)(33), and  
340 kinship matrices were calculated using the PC-Relate method (accounting for principal  
341 components)(34), both from the GENESIS R package. A sparse genetic relationship matrix  
342 containing only relationships of degree four or closer was extracted for analysis. Samples were  
343 grouped into race/ethnicity categories based on cohort-reported values.

#### 344 **Harmonization of glycemic traits**

345 Phenotypes were harmonized across the 10 studies based on a protocol developed within the  
346 TOPMed Diabetes Working Group. Glycemic traits, including fasting glucose (FG; mmol/L),  
347 fasting insulin (FI; pmol/L), and glycated hemoglobin (HbA1c; %), were collected where  
348 available. Fasting (for FG and FI) was defined as at least 8 hours without food or drink. FG  
349 measurements made in blood rather than plasma were adjusted by multiplying by a correction  
350 factor of 1.13. When multiple values were available for a given participant, blood draws were  
351 chosen to favor measurements made at study baselines and in order to maximize overlap with  
352 time points in which dietary data were collected. Participants were excluded if their glycemic  
353 trait blood draw was more than one year before or after diet measurement or if they had diabetes  
354 (defined as any of: taking anti-diabetic medication,  $FG \geq 7$  mmol/L, or  $HbA1c \geq 6.5\%$ ). Further  
355 study-specific details are available in the Supplementary Methods. Phenotype data  
356 harmonization and all other post-genome-wide analyses and visualizations were conducted using  
357 R version 4.1.1(35). Unless otherwise noted, all analyses including harmonization were  
358 performed on the NHLBI BioData Catalyst cloud computing platform(36).

#### 359 **Harmonization of dietary data**

360 Estimates of dietary intake were derived from self-reported diet questionnaires, either food  
361 frequency questionnaires (FFQ), diet history, or 24-hour recalls (24HR). Reported quantities of  
362 food and beverage consumption were converted into daily nutrient intake estimates via standard  
363 nutrient databases (see Supplementary Methods for study-specific details), with energy and  
364 macronutrients (carbohydrate, protein and total fat) expressed in kilocalories/day (kcal/d).

365 Participants were excluded if responses were deemed implausible, based on having total caloric  
366 intake <600 kcals/d or >4800 kcals/d. Nutrient intake values were analyzed in units of kcals/d  
367 and winsorized at three standard deviations from the mean. Dietary fiber was represented in  
368 grams/day (g/d) and alcohol intake was reported as number of drinks per day.

### 369 **Genome-wide gene-diet interaction scans**

370 For each cohort and glycemic trait, four genome-wide gene-diet interaction scans were  
371 performed to identify diet-interacting loci: one for common variants, and three gene-based  
372 aggregate tests for rare variants using different variant masks (described below). Mixed linear  
373 models were used to allow for random effects of kinship capturing close family relationships  
374 (degree four relatives or closer). The linear model setup was as follows:

$$FG/\ln FI/HbA1c \sim g + CHO + g * CHO + \gamma + covariates$$

375 Where  $g$  is the genotype at the variant of interest,  $CHO$  is dietary carbohydrate intake (kcals/d),  
376 and  $\gamma$  is a random effect governed by a sparse kinship matrix. General covariates included sex,  
377 age, age<sup>2</sup>, five genetic principal components to capture genetic ancestry, cohort-reported  
378 race/ethnicity to capture potential confounding by ethnicity-related dietary behavior, and  
379 additional study-specific covariates (Supp. Table S2). Dietary protein intake and total energy  
380 (also expressed in kcals/d) were included as covariates to set up an isocaloric substitution model  
381 in which increases in CHO were implicitly exchanged for decreases in dietary fat. Dietary fiber  
382 (g/d), alcohol intake (standard drinks/d), and body mass index (BMI; kg/m<sup>2</sup>) were included as  
383 additional covariates to account for further lifestyle-related confounding. During null model  
384 fitting, heterogeneous variances were allowed within each cohort-reported race/ethnicity group  
385 (equivalent to including a random effect for this grouping variable).

386 Genome-wide interaction analysis was performed using the *MAGEE* package(37). Single-  
387 variant analysis (*glmm.gci* function) was conducted for variants with MAF > 1%. *METAL*(38)  
388 was used to perform fixed effects meta-analysis across cohorts. Specifically, the 2-degree of  
389 freedom joint meta-analysis patch was used(39), with genetic main effect and interaction p-  
390 values derived downstream based on the resulting effect and standard error estimates.

391 Gene-centric, set-based rare-variant analysis (*MAGEE* function) was conducted for variants with  
392 MAF < 1%. Variant annotations derived from the WGS v0.8 and WGSAParsr v6.3.8 were  
393 retrieved from the NCBI dbGaP database. A genome-wide interaction meta-analysis was  
394 conducted for each of three variant masks: loss of function variants  
395 (VEP\_ensembl\_Consequence has terms *transcript\_ablation*, *splice\_acceptor\_variant*,  
396 *splice\_donor\_variant*, *stop\_gained*, *frameshift\_variant*, *stop\_lost*, *start\_lost* or  
397 *transcript\_amplification*), missense variants (VEP\_ensembl\_Consequence has the term  
398 *missense\_variant*), and a broad coding and noncoding filter (containing high-confidence loss-of-  
399 function variants, missense variants, protein-altering variants, synonymous variants, variants  
400 overlapping enhancers, and variants overlapping promoters). *MAGEE* calculates three interaction  
401 p-values: an adjusted variance component-like test, a burden test (assuming a consistent direction  
402 of effect for all variants), and a hybrid test (which combines the first two p-values using Fisher's  
403 method). P-values from the hybrid test were used here to balance the increased power of the

404 burden test with the possibility that its assumption of homogeneous effect directions is violated.  
405 Meta-analysis was then performed using a fixed-effects strategy.

406 Linear mixed models without genotype terms, meant to understand the marginal dietary effects  
407 prior to considering genetic effects, were fit in R using analogous models to those with diet-  
408 genotype interaction terms. Diet-genetic principal component interaction terms were excluded  
409 from these models and individuals in cohort-reported race/ethnicity groups with less than five  
410 members were excluded. Fixed effect meta-analysis of the CHO association (implicitly modeling  
411 an exchange with FAT due to the additional dietary covariates) was conducted using the *meta*  
412 package.

### 413 **Variant follow-up**

414 Sensitivity analysis was conducted to understand the impact of modeling choices on the  
415 interaction effect estimates derived in the genome-wide analysis. These linear mixed models  
416 were fit in R, with G×CHO interaction terms subject to fixed-effects meta-analysis using the  
417 *meta* package as with the models without genotype effects. Some of these involved subsets of the  
418 population: male and female subsets were tested separately, as well as subsets without obesity  
419 (BMI < 30 kg/m<sup>2</sup>) and with and without prediabetes (defined as FG > 5.6 mmol/L or HbA1c >  
420 5.7%). Additional models included adjustment for either smoking status (never/former/current,  
421 coded as 0/1/2 and analyzed as a continuous variable), the Alternative Healthy Eating Index  
422 2010(40) (a diet quality score), or a categorical coding of alcohol intake (none, modest [ $<1$   
423 drink/day for females or  $<2$  drinks/day for males], or high), where available. These models with  
424 additional covariate adjustments also included adjustment for their interactions with genotype.  
425 Finally, a model including genotype interaction terms for other main dietary components and  
426 lifestyle confounders (total energy, protein, fiber, and alcohol) was included. This type of  
427 residual confounding by genotype-covariate interaction terms has been previously  
428 documented(41), but would have decreased statistical power if included in the genome-wide  
429 analysis, especially for lower-frequency variants.

430 Variant rs79762542 was investigated in greater depth as the only variant reaching study-wide  
431 significance. Based on its eQTL relationship impacting *FRAS1* gene expression in thyroid from  
432 the GTEx v8 dataset (<https://gtexportal.org/>), we tested for colocalization of this signal with the  
433 carbohydrate interaction signal impacting HbA1c. Interaction summary statistics were retrieved  
434 in a window of 1Mb around the index variant rs79762542, and all thyroid-specific *cis*-eQTL  
435 summary statistics related to *FRAS1* were retrieved from GTEx. Colocalization was tested using  
436 the *coloc* package for R, assuming a single causal variant (*coloc.abf* function).

### 437 **Replication analysis in the UK Biobank**

438 UKB is a large prospective cohort with both deep phenotyping and molecular data, including  
439 genome-wide genotyping, on over 500,000 individuals ages 40-69 living throughout the UK  
440 between 2006-2010(42). Genotyping, imputation, and initial quality control on the genetic  
441 dataset have been described previously(43). Analyses were conducted on genetic data release  
442 version 3, with imputation to a joint reference panel including the Haplotype Reference  
443 Consortium and the 1000 Genomes Project (1KGP), under UK Biobank application 27892. This



444 work was conducted under a Not Human Subjects Research determination (NHSR-4298 at the  
445 Broad Institute of MIT and Harvard).

446  
447 Ancestry group labels, genetic principal components, and labels defining an unrelated subset of  
448 individuals were retrieved from the Pan-UKBB project (<https://pan.ukbb.broadinstitute.org/>; data  
449 retrieved from UKB return of results number 2442). Only unrelated individuals were used for  
450 analysis, with additional removal of individuals who were pregnant or had diabetes at the study  
451 center visit. Two glycemic traits were available for testing in UKB: HbA1c and glucose  
452 (collected as a random glucose measurement and later to remove non-fasting individuals).  
453 Outliers for both traits (defined as more than 5 standard deviations from the mean) were  
454 removed. Dietary data came from one or more Oxford WebQ 24-hour dietary assessments(17)  
455 completed at the study center or during online follow-up over the course of approximately two  
456 years. Daily nutrient intake estimates (calculated centrally by the UKB) were averaged across all  
457 questionnaires for each individual and winsorized at three standard deviations from the mean.  
458 After all exclusions, 178,352 individuals without diabetes had available genotype, biomarker,  
459 and dietary data.

460  
461 Regression analysis in the UKB mirrored that of the primary analysis, replacing cohort-reported  
462 race/ethnicity with genetically-defined ancestry groups as defined by the Pan-UKBB project.  
463 Given the larger sample size available, gene-covariate interactions were included for dietary  
464 covariates (total energy, protein, fiber, and alcohol). When analyzing glucose, only the subset of  
465 individuals with reported fasting times of at least eight hours were included, reducing the sample  
466 size to 5,183. Due to the African ancestry-specificity of some of the top variants, a second  
467 replication analysis was performed in the African-ancestry subset of UKB.

## 468 **Power Calculations**

469 Interaction test power calculations were performed using the *ESPRESSO.GxE* R package, which  
470 uses a simulation-based approach to calculate empirical power estimates (given some sample  
471 size) and sample size requirements (to achieve 80% power). The following parameters were  
472 fixed for this analysis, chosen to mimic an analysis of HbA1c: random seed = 1, significance  
473 threshold =  $5 \times 10^{-8}$ , phenotype mean = 5.5, phenotype standard deviation = 0.5, phenotype  
474 reliability = 1, genetic main effect = 0.1, exposure mean = 0, exposure standard deviation = 1,  
475 exposure main effect = 0.2. The following parameters were varied: interaction effect {0.025,  
476 0.0375, 0.05, 0.0625, 0.075, 0.0875, 0.1}, MAF {0.01, 0.05, 0.1, 0.5}, and exposure reliability  
477 {0.25, 0.5, 0.75, 1}. Here, reliability is used to quantify the simulated measurement error of the  
478 phenotype and exposure and is equivalent to an intraclass correlation coefficient (ratio of  
479 between-subject variance to total [between-subject plus measurement error] variance).

480 To enable simulation-based power calculations for aggregate tests of rare variants while  
481 accounting for exposure measurement error, we developed an extension of the *ESPRESSO.GxE*  
482 package, called *ESPRESSO.GxE.RV*. In this extension, the basic structure of the simulations  
483 remains the same, but an additional parameter allows the user to specify a number of variants  
484 (*M*) to test in aggregate. Within each simulation run, *M* variants are simulated, with some portion  
485 having equal interaction and main effects on the outcome (according to a user-specified causal  
486 variant fraction) and the rest generated randomly. The final p-value from that simulation is

487 calculated using Fisher's method on the full set of M p-values. The following parameters were  
488 given different values for this set of simulations: interaction effect {0.025, 0.05, 0.075, 0.1,  
489 0.125, 0.15, 0.175, 0.2}, MAF {0.0025, 0.005, 0.0075, 0.01}. Other parameters were specific to  
490 rare variant tests: number of variants per aggregate test {1, 5, 10, 20} and causal fraction {0.1,  
491 0.25, 0.5, 1}. Code for this extension of the package can be found on GitHub:  
492 <https://github.com/kwesterman/ESPRESSO.GxE.RV>.

493 To provide context for realistic gene-diet interaction effect sizes despite few well-replicated  
494 examples of such interactions for glycemic traits in the literature, we retrieved results from  
495 variants reaching significance in a recent trans-ancestry GWAS for HbA1c(21) and an estimated  
496 effect for the carbohydrate-HbA1c relationship from a recent nutritional epidemiological  
497 analysis(27).

## 498 499 **AUTHOR CONTRIBUTIONS**

500 KEW, MEW, JM, and AKM designed the research. KEW, SMG, JW, DKT, and AKM  
501 contributed to the harmonization of glycemic trait variables across studies; KEW conducted the  
502 research and performed the primary data analysis; JCF, JBM, DKT, HC, and AKM supervised  
503 the research; KEW and MEW wrote the manuscript; all additional authors contributed to the  
504 collection and curation of the study-specific or TOPMed-wide datasets; KEW and AKM had  
505 primary responsibility for the final content. All authors read and approved the final manuscript:  
506 KEW, MEW, SMG, JW, DD, JM, AA, SA, ASB, AGB, MLB, JAB, YIC, JD, MOG, XG, NRH,  
507 AH, BH, MRI, WCJ, RRK, LL, RNL, CL, SL, JM, RN, JSP, MP, LR, LJR, ES, MKS, AHS,  
508 AVS, NLS, LS, ST, KDT, PSD, JGW, ACW, LRY, JY, YZ, EB, ACM, MF, TPR, BMP, DL,  
509 NLH, VSR, RAM, DKA, RK, KEN, AC, AC, JR, SSR, JEM, APR, CK, JCF, JBM, JM, DKT,  
510 HC, AKM.

## 512 513 **ACKNOWLEDGMENTS**

514  
515 The Genotype-Tissue Expression (GTEx) Project was supported by the Common Fund of the  
516 Office of the Director of the National Institutes of Health, and by NCI, NHGRI, NHLBI, NIDA,  
517 NIMH, and NINDS. The data used for the analyses described in this manuscript were obtained  
518 from the GTEx Portal on 03/28/2022.

519  
520 Support for this work was provided by the National Institutes of Health, National Heart, Lung,  
521 and Blood Institute, through the BioData Catalyst program (award 1OT3HL142479-01,  
522 1OT3HL142478-01, 1OT3HL142481-01, 1OT3HL142480-01, 1OT3HL147154-01). Any  
523 opinions expressed in this document are those of the author(s) and do not necessarily reflect the  
524 views of NHLBI, individual BioData Catalyst team members, or affiliated organizations and  
525 institutions. Further, the authors wish to acknowledge the contributions of the consortium  
526 working on the development of the NHLBI BioData Catalyst ecosystem.

527  
528 The authors would also like to acknowledge the contributions of L. Adrienne Cupples to the  
529 Framingham Heart Study and the TOPMed program.

530

531 **TOPMed Acknowledgments**

532 Molecular data for the Trans-Omics in Precision Medicine (TOPMed) program was supported by  
 533 the National Heart, Lung and Blood Institute (NHLBI). See the TOPMed Omics Support Table  
 534 below for study specific omics support information. Core support including centralized genomic  
 535 read mapping and genotype calling, along with variant quality metrics and filtering were  
 536 provided by the TOPMed Informatics Research Center (3R01HL-117626-02S1; contract  
 537 HHSN268201800002I). Core support including phenotype harmonization, data management,  
 538 sample-identity QC, and general program coordination were provided by the TOPMed Data  
 539 Coordinating Center (R01HL-120393; U01HL-120393; contract HHSN268201800001I). We  
 540 gratefully acknowledge the studies and participants who provided biological samples and data  
 541 for TOPMed.

Study	Acknowledgements
<p>Atherosclerosis Risk in Communities study</p>	<p>ARIC: The Atherosclerosis Risk in Communities study has been funded in whole or in part with Federal funds from the National Heart, Lung, and Blood Institute, National Institutes of Health, Department of Health and Human Services, under Contract nos. (75N92022D00001, 75N92022D00002, 75N92022D00003, 75N92022D00004, 75N92022D00005). The authors thank the staff and participants of the ARIC study for their important contributions.</p> <p>WGS for “NHLBI TOPMed: Atherosclerosis Risk in Communities (ARIC)” (phs001211) was performed at the Baylor College of Medicine Human Genome Sequencing Center (HHSN268201500015C and 3U54HG003273-12S2) and the Broad Institute for MIT and Harvard (3R01HL092577-06S1).</p> <p>The Genome Sequencing Program (GSP) was funded by the National Human Genome Research Institute (NHGRI), the National Heart, Lung, and Blood Institute (NHLBI), and the National Eye Institute (NEI). The GSP Coordinating Center (U24 HG008956) contributed to cross program scientific initiatives and provided logistical and general study coordination. The Centers for Common Disease Genomics (CCDG) program was supported by NHGRI and NHLBI, and whole genome sequencing was performed at the Baylor College of Medicine Human Genome Sequencing Center (UM1 HG008898).</p>
<p>CARDIA</p>	<p>The Coronary Artery Risk Development in Young Adults Study (CARDIA) is conducted and supported by the National Heart, Lung, and Blood Institute (NHLBI) in collaboration with the University of Alabama at Birmingham (HHSN268201800005I &amp; HHSN268201800007I), Northwestern University (HHSN268201800003I), University of Minnesota (HHSN268201800006I), and Kaiser Foundation Research Institute (HHSN268201800004I). CARDIA was also partially supported by the Intramural Research Program of the National Institute on Aging (NIA) and an intra-agency agreement between NIA and NHLBI (AG0005).</p>

CHS	<p>This research was supported by contracts 75N92021D00006, HHSN268201200036C, HHSN268200800007C, HHSN268201800001C, N01HC55222, N01HC85079, N01HC85080, N01HC85081, N01HC85082, N01HC85083, N01HC85086, and grants R01HL120393, U01HL080295 and U01HL130114 from the National Heart, Lung, and Blood Institute (NHLBI), with additional contribution from the National Institute of Neurological Disorders and Stroke (NINDS). Additional support was provided by R01AG023629 from the National Institute on Aging (NIA). A full list of principal CHS investigators and institutions can be found at <a href="http://CHS-NHLBI.org">CHS-NHLBI.org</a>. The content is solely the responsibility of the authors and does not necessarily represent the official views of the National Institutes of Health.</p>
Framingham Heart Study	<p>WGS for "NHLBI TOPMed: Whole Genome Sequencing and Related Phenotypes in the Framingham Heart Study (phs000974.v1.p1) was performed at the Broad Institute of MIT and Harvard (3R01HL092577-06S1 (AFGen)). This research was conducted in part using data and resources from the Framingham Heart Study of the National Heart Lung and Blood Institute of the National Institutes of Health and Boston University School of Medicine. The Framingham Heart Study (FHS) acknowledges the support of contracts NO1-HC-25195, HHSN268201500001I and 75N92019D00031 from the National Heart, Lung and Blood Institute and grant supplement R01 HL092577-06S1 for this research. We also acknowledge the dedication of the FHS study participants without whom this research would not be possible.</p>
GeneSTAR	<p>GeneSTAR was supported by grants from the National Institutes of Health/National Heart, Lung, and Blood Institute (U01 HL72518, HL087698, HL49762, HL59684, HL58625, HL071025, HL092165, HL099747, K23HL105897, HL112064), by grants from the National Institutes of Health/National Institute of Nursing Research (NR0224103, NR008153), by a grant from the National Institutes of Health/National Institute of Neurological Disorders and Stroke (NS062059), by a grant from the National Institutes of Health/National Center for Research Resources (M01-RR000052) to the Johns Hopkins General Clinical Research Center, and by a grant from the National Center for Advancing Translational Sciences, National Institutes of Health (UL1TR001079) to the Johns Hopkins Institute for Clinical &amp; Translational Research.</p>
GOLDN	<p>GOLDN biospecimens, baseline phenotype data, and intervention phenotype data were collected with funding from National Heart, Lung and Blood Institute (NHLBI) grant U01 HL072524. Whole-genome sequencing in GOLDN was funded by NHLBI grant R01 HL104135-04S1. We thank GOLDN participants and investigators for their significant contributions.</p>
HCHS/SOL	<p>The Hispanic Community Health Study/Study of Latinos is a collaborative study supported by contracts from the National Heart, Lung, and Blood Institute (NHLBI) to the University of North Carolina (HHSN268201300001I / N01-HC-65233), University of Miami (HHSN268201300004I / N01-HC-65234), Albert Einstein College of Medicine (HHSN268201300002I / N01-HC-65235), University of Illinois at Chicago – HHSN268201300003I / N01-HC-65236 Northwestern Univ), and San Diego State University</p>

	<p>(HHSN268201300005I / N01-HC-65237). The following Institutes/Centers/Offices have contributed to the HCHS/SOL through a transfer of funds to the NHLBI: National Institute on Minority Health and Health Disparities, National Institute on Deafness and Other Communication Disorders, National Institute of Dental and Craniofacial Research, National Institute of Diabetes and Digestive and Kidney Diseases, National Institute of Neurological Disorders and Stroke, NIH Institution-Office of Dietary Supplements.</p>
<p>JHS</p>	<p>The Jackson Heart Study is supported and conducted in collaboration with Jackson State University (HHSN268201800013I), Tougaloo College (HHSN268201800014I), the Mississippi State Department of Health (HHSN268201800015I) and the University of Mississippi Medical Center (HHSN268201800010I, HHSN268201800011I and HHSN268201800012I) contracts from the National Heart, Lung, and Blood Institute (NHLBI) and the National Institute on Minority Health and Health Disparities (NIMHD). The authors thank the participants and data collection staff of the Jackson Heart Study. The views expressed in this manuscript are those of the authors and do not necessarily represent the views of the National Heart, Lung, and Blood Institute; the National Institutes of Health; or the U.S. Department of Health and Human Services.</p>
<p>MESA</p>	<p>Whole genome sequencing (WGS) for the Trans-Omics in Precision Medicine (TOPMed) program was supported by the National Heart, Lung and Blood Institute (NHLBI). WGS for “NHLBI TOPMed: Multi-Ethnic Study of Atherosclerosis (MESA)” (phs001416.v1.p1) was performed at the Broad Institute of MIT and Harvard (3U54HG003067-13S1). Centralized read mapping and genotype calling, along with variant quality metrics and filtering were provided by the TOPMed Informatics Research Center (3R01HL-117626-02S1). Phenotype harmonization, data management, sample-identity QC, and general study coordination, were provided by the TOPMed Data Coordinating Center (3R01HL-120393-02S1), and TOPMed MESA Multi-Omics (HHSN2682015000031/HSN26800004). The MESA projects are conducted and supported by the National Heart, Lung, and Blood Institute (NHLBI) in collaboration with MESA investigators. Support for the Multi-Ethnic Study of Atherosclerosis (MESA) projects are conducted and supported by the National Heart, Lung, and Blood Institute (NHLBI) in collaboration with MESA investigators. Support for MESA is provided by contracts 75N92020D00001, HHSN268201500003I, N01-HC-95159, 75N92020D00005, N01-HC-95160, 75N92020D00002, N01-HC-95161, 75N92020D00003, N01-HC-95162, 75N92020D00006, N01-HC-95163, 75N92020D00004, N01-HC-95164, 75N92020D00007, N01-HC-95165, N01-HC-95166, N01-HC-95167, N01-HC-95168, N01-HC-95169, UL1-TR-000040, UL1-TR-001079, UL1-TR-001420, UL1TR001881, DK063491, and R01HL105756. The authors thank the other investigators, the staff, and the participants of the MESA study for their valuable contributions. A full list of participating MESA investigators and institutes can be found</p>

	at <a href="http://www.mesa-nhlbi.org">http://www.mesa-nhlbi.org</a> . Also supported in part by the NHLBI contracts R01HL151855 and R01HL146860.
WHI	The WHI program is funded by the National Heart, Lung, and Blood Institute, National Institutes of Health, U.S. Department of Health and Human Services through contracts 75N92021D00001, 75N92021D00002, 75N92021D00003, 75N92021D00004, 75N92021D00005.

542

543

544

Data described in the manuscript will be made available upon request pending application and approval from NHLBI Trans-Omics for Precision Medicine (TOPMed) program.

## REFERENCES

1. Ajala O, English P, Pinkney J. Systematic review and meta-analysis of different dietary approaches to the management of type 2 diabetes. *Am J Clin Nutr* 2013;97:505–16.
2. Salas-Salvadó J, Bulló M, Babio N, Martínez-González MÁ, Ibarrola-Jurado N, Basora J, Estruch R, Covas MI, Corella D, Arós F, et al. Reduction in the Incidence of Type 2 Diabetes With the Mediterranean Diet. *Diabetes Care* 2011;34:14–9.
3. The Diabetes Prevention Program (DPP) Research Group. The Diabetes Prevention Program (DPP). *Diabetes Care* 2002;25:2165–71.
4. Davis NJ, Tomuta N, Schechter C, Isasi CR, Segal-Isaacson CJ, Stein D, Zonszein J, Wylie-Rosett J. Comparative Study of the Effects of a 1-Year Dietary Intervention of a Low-Carbohydrate Diet Versus a Low-Fat Diet on Weight and Glycemic Control in Type 2 Diabetes. *Diabetes Care* 2009;32:1147–52.
5. Korem T, Zeevi D, Zmora N, Weissbrod O, Bar N, Lotan-Pompan M, Avnit-Sagi T, Kosower N, Malka G, Rein M, et al. Bread Affects Clinical Parameters and Induces Gut Microbiome-Associated Personal Glycemic Responses. *Cell Metab* 2017;
6. Berry SE, Valdes AM, Drew DA, Asnicar F, Mazidi M, Wolf J, Capdevila J, Hadjigeorgiou G, Davies R, Al Khatib H, et al. Human postprandial responses to food and potential for precision nutrition. *Nat Med* 2020;26:964–73.
7. Bachmann AM, Morel J-D, El Alam G, Rodríguez-López S, Imamura de lima T, Goeminne LJE, Benegiamo G, Loric S, Conti M, Sleiman MB, et al. Genetic background and sex control the outcome of high-fat diet feeding in mice. *iScience* 2022;25:104468.
8. Barrington WT, Wulfridge P, Wells AE, Rojas CM, Howe SYF, Perry A, Hua K, Pellizzon MA, Hansen KD, Voy BH, et al. Improving Metabolic Health Through Precision Dietetics in Mice. *Genetics* 2018;208:399–417.
9. Qi B, Bray GA, Hu FB, Sacks FM, Qi L. Weight-loss diets modify glucose-dependent insulinotropic polypeptide receptor rs2287019 genotype effects on changes in body weight, fasting glucose, and insulin resistance: The preventing overweight using novel dietary strategies trial. *Am J Clin Nutr* 2012;95.
10. Corella D, Carrasco P, Sorli J V., Estruch R, Rico-Sanz J, Martinez-Gonzalez MA, Salas-Salvado J, Covas MI, Coltell O, Aros F, et al. Mediterranean Diet Reduces the Adverse Effect of the TCF7L2-rs7903146 Polymorphism on Cardiovascular Risk Factors and Stroke Incidence: A randomized controlled trial in a high-cardiovascular-risk population. *Diabetes Care* 2013;36:3803–11.
11. Westerman KE, Miao J, Chasman DI, Florez JC, Chen H, Manning AK, Cole JB. Genome-wide gene–diet interaction analysis in the UK Biobank identifies novel effects on hemoglobin A1c. *Hum Mol Genet* 2021;
12. Francis M, Li C, Sun Y, Zhou J, Li X, Brenna JT, Ye K. Genome-wide association study of fish oil supplementation on lipid traits in 81,246 individuals reveals new gene-diet interaction loci. Cordell HJ, editor. *PLOS Genet* 2021;17:e1009431.
13. Franks PW, Merino J. Gene-lifestyle interplay in type 2 diabetes. *Curr Opin Genet Dev* 2018;50:35–40.
14. Mahajan A, Taliun D, Thurner M, Robertson NR, Torres JM, Rayner NW, Payne AJ, Steinthorsdottir V, Scott RA, Grarup N, et al. Fine-mapping type 2 diabetes loci to single-variant resolution using high-density imputation and islet-specific epigenome maps. *Nat Genet* 2018;50:1505–13.

15. Fadista J, Manning AK, Florez JC, Groop L. The (in)famous GWAS P-value threshold revisited and updated for low-frequency variants. *Eur J Hum Genet* 2016;24:1202–5.
16. Zou Y, Gong N, Cui Y, Wang X, Cui A, Chen Q, Jiao T, Dong X, Yang H, Zhang S, et al. Forkhead Box P1 (FOXP1) Transcription Factor Regulates Hepatic Glucose Homeostasis. *J Biol Chem* 2015;290:30607–15.
17. Liu B, Young H, Crowe FL, Benson VS, Spencer EA, Key TJ, Appleby PN, Beral V. Development and evaluation of the Oxford WebQ, a low-cost, web-based method for assessment of previous 24 h dietary intakes in large-scale prospective studies. *Public Health Nutr* 2011;14:1998–2005.
18. Wong M, Day N, Luan J, Chan K, Wareham N. The detection of gene–environment interaction for continuous traits: should we deal with measurement error by bigger studies or better measurement? *Int J Epidemiol* 2003;32:51–7.
19. Osazuwa-Peters OL, Schwander K, Waken RJ, de las Fuentes L, Kilpeläinen TO, Loos RJJ, Racette SB, Sung YJ, Rao DC. The Promise of Selecting Individuals from the Extremes of Exposure in the Analysis of Gene-Physical Activity Interactions. *Hum Hered* 2018;83:315–32.
20. Jurgens SJ, Choi SH, Morrill VN, Chaffin M, Pirruccello JP, Halford JL, Weng L-C, Nauffal V, Roselli C, Hall AW, et al. Analysis of rare genetic variation underlying cardiometabolic diseases and traits among 200,000 individuals in the UK Biobank. *Nat Genet* 2022;
21. Wheeler E, Leong A, Liu C-T, Hivert M-F, Strawbridge RJ, Podmore C, Li M, Yao J, Sim X, Hong J, et al. Impact of common genetic determinants of Hemoglobin A1c on type 2 diabetes risk and diagnosis in ancestrally diverse populations: A transethnic genome-wide meta-analysis. Gregg E, editor. *PLOS Med* 2017;14:e1002383.
22. Arnold KF, Berrie L, Tennant PWG, Gilthorpe MS. A causal inference perspective on the analysis of compositional data. *Int J Epidemiol* 2020;49:1307–13.
23. Salmerón J, Hu FB, Manson JE, Stampfer MJ, Colditz GA, Rimm EB, Willett WC. Dietary fat intake and risk of type 2 diabetes in women. *Am J Clin Nutr* 2001;73:1019–26.
24. Merino J, Guasch-Ferré M, Ellervik C, Dashti HS, Sharp SJ, Wu P, Overvad K, Sarnowski C, Kuokkanen M, Lemaitre RN, et al. Quality of dietary fat and genetic risk of type 2 diabetes: individual participant data meta-analysis. *BMJ* 2019;14292.
25. Harding A-H, Sargeant LA, Welch A, Oakes S, Luben RN, Bingham S, Day NE, Khaw K-T, Wareham NJ. Fat Consumption and HbA1c Levels. *Diabetes Care* 2001;24:1911–6.
26. Imamura F, Micha R, Wu JHY, de Oliveira Otto MC, Otite FO, Abioye AI, Mozaffarian D. Effects of Saturated Fat, Polyunsaturated Fat, Monounsaturated Fat, and Carbohydrate on Glucose-Insulin Homeostasis: A Systematic Review and Meta-analysis of Randomised Controlled Feeding Trials. Ma RCW, editor. *PLOS Med* 2016;13:e1002087.
27. Churuangsu C, Lean MEJ, Combet E. Lower carbohydrate and higher fat intakes are associated with higher hemoglobin A1c: findings from the UK National Diet and Nutrition Survey 2008–2016. *Eur J Nutr* 2020;59:2771–82.
28. Pérusse L, Bouchard C. Gene-diet interactions in obesity. *Am J Clin Nutr* 2000;72:1285s-1290s.
29. Gaye A, Burton TWY, Burton PR. ESPRESSO: Taking into account assessment errors on outcome and exposures in power analysis for association studies. *Bioinformatics* 2015;31.
30. Dahl A, Nguyen K, Cai N, Gandal MJ, Flint J, Zaitlen N. A Robust Method Uncovers Significant Context-Specific Heritability in Diverse Complex Traits. *Am J Hum Genet*



- 2020;106:71–91.
31. Zheng J-S, Arnett DK, Lee Y-C, Shen J, Parnell LD, Smith CE, Richardson K, Li D, Borecki IB, Ordovás JM, et al. Genome-Wide Contribution of Genotype by Environment Interaction to Variation of Diabetes-Related Traits. *PLoS One* 2013;8:e77442.
  32. Oster E. Health Recommendations and Selection in Health Behaviors. *Am Econ Rev Insights* 2020;2:143–60.
  33. Conomos MP, Miller MB, Thornton TA. Robust Inference of Population Structure for Ancestry Prediction and Correction of Stratification in the Presence of Relatedness. *Genet Epidemiol* 2015;39:276–93.
  34. Conomos MP, Reiner AP, Weir BS, Thornton TA. Model-free Estimation of Recent Genetic Relatedness. *Am J Hum Genet* 2016;98:127–48.
  35. R Core Team. R: A Language and Environment for Statistical Computing. Vienna, Austria: R Foundation for Statistical Computing; 2021.
  36. The NHLBI BioData Catalyst. National Heart, Lung, and Blood Institute, National Institutes of Health, U.S. Department of Health and Human Services; 2020.
  37. Wang X, Lim E, Liu CT, Sung YJ, Rao DC, Morrison AC, Boerwinkle E, Manning AK, Chen H. Efficient gene–environment interaction tests for large biobank-scale sequencing studies. *Genet Epidemiol* 2020;44.
  38. Willer CJ, Li Y, Abecasis GR. METAL: fast and efficient meta-analysis of genomewide association scans. *Bioinformatics* 2010;26:2190–1.
  39. Manning AK, LaValley M, Liu C-T, Rice K, An P, Liu Y, Miljkovic I, Rasmussen-Torvik L, Harris TB, Province MA, et al. Meta-analysis of gene-environment interaction: joint estimation of SNP and SNP  $\times$  environment regression coefficients. *Genet Epidemiol* 2011;35:11–8.
  40. Chiuvè SE, Fung TT, Rimm EB, Hu FB, McCullough ML, Wang M, Stampfer MJ, Willett WC. Alternative Dietary Indices Both Strongly Predict Risk of Chronic Disease. *J Nutr* 2012;142:1009–18.
  41. Keller MC. Gene  $\times$  Environment Interaction Studies Have Not Properly Controlled for Potential Confounders: The Problem and the (Simple) Solution. *Biol Psychiatry* 2014;75:18–24.
  42. Sudlow C, Gallacher J, Allen N, Beral V, Burton P, Danesh J, Downey P, Elliott P, Green J, Landray M, et al. UK Biobank: An Open Access Resource for Identifying the Causes of a Wide Range of Complex Diseases of Middle and Old Age. *PLOS Med* 2015;12:e1001779.
  43. Bycroft C, Freeman C, Petkova D, Band G, Elliott LT, Sharp K, Motyer A, Vukcevic D, Delaneau O, O’Connell J, et al. The UK Biobank resource with deep phenotyping and genomic data. *Nature* 2018;562:203–9.

## TABLES

**Table 1:** Top variants interacting with carbohydrate intake from the common-variant GWIS.

Trait	rsID	Chromosome	Effect Allele	Avg. EAF	Main effect estimate	Interaction estimate	$P_{interaction}$
HbA1c	rs79762542	4	G	0.03	-0.013 [-0.038-0.012]	0.048 [0.031-0.064]	$1.1 \times 10^{-8}$
FG	rs1288694	3	C	0.61	-0.003 [-0.011-0.004]	0.016 [0.01-0.022]	$1.9 \times 10^{-8}$
lnFI	rs782681704	X	G	0.01	-0.049 [-0.209-0.11]	0.284 [0.182-0.385]	$4.6 \times 10^{-8}$

Interaction estimates with 95% CIs are given in units of [trait units / allele / 250 kcal carbohydrate]. All variants passed a significance threshold of  $P_{interaction} < 5 \times 10^{-8}$ .