



Universiteit  
Leiden  
The Netherlands

## **The association between body mass index and metabolite response to a liquid mixed meal challenge: a Mendelian randomization study**

Hughes, D.A.; Li-Gao, R.; Bull, C.J.; Mutsert, R. de; Rosendaal, F.R.; Mook-Kanamori, D.; ... ; Timpson, N.J.

### **Citation**

Hughes, D. A., Li-Gao, R., Bull, C. J., Mutsert, R. de, Rosendaal, F. R., Mook-Kanamori, D., ... Timpson, N. J. (2024). The association between body mass index and metabolite response to a liquid mixed meal challenge: a Mendelian randomization study. *The American Journal Of Clinical Nutrition*, 119(5), 1354-1370.  
doi:10.1016/j.ajcnut.2024.03.009

Version: Publisher's Version

License: [Creative Commons CC BY 4.0 license](#)

Downloaded from: <https://hdl.handle.net/1887/4198939>

**Note:** To cite this publication please use the final published version (if applicable).



## Original Research Article

# The association between body mass index and metabolite response to a liquid mixed meal challenge: a Mendelian randomization study

David A Hughes<sup>1,2,\*</sup>, Ruifang Li-Gao<sup>3</sup>, Caroline J Bull<sup>1,2</sup>, Renée de Mutsert<sup>3</sup>, Frits R Rosendaal<sup>3</sup>, Dennis O Mook-Kanamori<sup>3,4</sup>, Ko Willems van Dijk<sup>5,6</sup>, Nicholas J Timpson<sup>1,2</sup>

<sup>1</sup> MRC Integrative Epidemiology Unit, University of Bristol, Bristol, United Kingdom; <sup>2</sup> Population Health Science, Bristol Medical School, University of Bristol, Bristol, United Kingdom; <sup>3</sup> Department of Clinical Epidemiology, Leiden University Medical Center, Leiden, the Netherlands; <sup>4</sup> Department of Public Health and Primary Care, Leiden University Medical Center, Leiden, the Netherlands; <sup>5</sup> Department of Human Genetics, Leiden University Medical Center, Leiden, the Netherlands; <sup>6</sup> Division of Endocrinology, Department of Internal Medicine, Leiden University Medical Center, Leiden, the Netherlands

## ABSTRACT

**Background:** Metabolite abundance is a dynamic trait that varies in response to environmental stimuli and phenotypic traits, such as food consumption and body mass index (BMI, kg/m<sup>2</sup>).

**Objectives:** In this study, we used the Netherlands Epidemiology of Obesity (NEO) study data to identify observational and causal associations between BMI and metabolite response to a liquid meal.

**Methods:** A liquid meal challenge was performed, and Nightingale Health metabolite profiles were collected in 5744 NEO participants. Observational and one-sample Mendelian randomization (MR) analysis were conducted to estimate the effect of BMI on metabolites ( $n = 229$ ) in the fasting, postprandial, and response (or change in abundance) states.

**Results:** We observed 473 associations with BMI (175 fasting, 188 postprandial, and 110 response) in observational analyses. In MR analyses, we observed 20 metabolite traits (5 fasting, 12 postprandial, and 3 response) to be associated with BMI. MR associations included the glucogenic amino acid alanine, which was inversely associated with BMI in the response state ( $\beta: -0.081$ ; SE: 0.023;  $P = 5.91 \times 10^{-4}$ ), suggesting that as alanine increased in postprandial abundance, that increase was attenuated with increasing BMI.

**Conclusions:** Overall, this study showed that MR estimates were strongly correlated with observational effect estimates, suggesting that the broad associations seen between BMI and metabolite variation has a causal underpinning. Specific effects in previously unassessed postprandial and response states are detected, and these may likely mark novel life course risk exposures driven by regular nutrition.

**Keywords:** metabolomics, lipidomics, obesity, instrumental variable analysis, causal inference

## Introduction

The excess accumulation of body fat and obesity are an established global health burden, the prevalence of which is increasing. BMI, or the ratio of one's weight and height squared, is a common metric for measuring excess weight or fat and BMI thresholds  $\geq 25$  and 30 are used to designate the conditions of overweight and obesity, respectively [1–6]. Increases in BMI have been associated with decreased life

expectancy [7–10], some cancers [11–13] and cardiometabolic diseases including cardiovascular disease, and type 2 diabetes [14].

Despite the scale of the BMI-related health burden, there remains a need to elucidate how and by what intermediate physiological traits excess body fat increases disease risk. Possible intermediates are metabolites and lipoproteins, some of which—such as low density lipoprotein (LDL) or its' crucial protein apolipoprotein B—have been previously associated with both BMI and disease risk [15–23]. Previous

*Abbreviations:* AAA, aromatic amino acid; BCAA, branch chain amino acid; BMI-PGS, polygenic score for body mass index; HDL, high density lipoprotein; IDL, intermediate density lipoprotein; IV, instrumental variable; LDL, low density lipoprotein; LUMC, Leiden University Medical Center; MR, Mendelian randomization; NA, not applicable; NEO, Netherlands Epidemiology of Obesity Study; PGS, polygenic score; QC, quality control; VLDL, very low density lipoprotein; faw6fa, ratio of  $\omega$ -6 fatty acids to total fatty acids; lafa, ratio of 18:2 linoleic acid to total fatty acids; ornl, orthogonal nonlinear least square; pufafa, ratio of polyunsaturated fatty acids to total fatty acids; SE, standard error; SNP, single nucleotide polymorphism; wNEO, weighted Netherlands Epidemiology of Obesity framework.

\* Corresponding author. DAH at Pennington Biomedical Research Center, 6400 Perkins Rd, Baton Rouge, LA 70808.

E-mail address: [david.hughes@pbrc.edu](mailto:david.hughes@pbrc.edu) (D.A. Hughes).

<https://doi.org/10.1016/j.ajcnut.2024.03.009>

Received 9 October 2023; Received in revised form 30 January 2024; Accepted 12 March 2024; Available online 16 March 2024

0002-9165/© 2024 The Authors. Published by Elsevier Inc. on behalf of American Society for Nutrition. This is an open access article under the CC BY license (<http://creativecommons.org/licenses/by/4.0/>).

cross-sectional, Mendelian randomization (MR), recall-by-genotype, and randomized controlled trials have established both observational and causal associations between BMI and fasted metabolite and protein abundances [15,24–26]. Such studies have informed and motivated downstream studies aimed at elucidating the causal effects of BMI-driven metabolite variation on disease outcomes, such as colorectal cancer [27,28].

However, with most individuals spending much of their waking hours in a nonfasted or postprandial state, it is important to understand both the variation in postprandial metabolite abundance and the association and influence BMI may have on dynamic ranges in metabolomic response. Response, defined as the change in metabolite abundance between fasted and postprandial states, may prove to be an emergent trait in disease risk research. Given that response is dependent on the starting, fasted, abundance level—which itself may be influenced by BMI—it becomes an instructive trait to evaluate interplays among BMI, metabolite variation, and disease risk. For example, we predict that some BMI-driven metabolite traits may have neared or reached a physiologic plateau in the fasted state, such that any response due to a meal has a limited range of change. Resultantly, response may prove to be a more powerful phenotype to identify molecular traits influenced by BMI than single time points. Previous work has provided evidence that this dynamic state is different and potentially informative versus classically measured fasting levels. For example, a recent genome-wide association analysis identified genetic variants uniquely associated with metabolite response traits that did not share associations with fasting or postprandial abundances [29]. Moreover, estimates of genotype heritability for some response traits, such as the (branched-chain) amino acids, glucose and extremely large very low density lipoprotein (VLDL), are as large or larger than those observed for fasting and postprandial estimates [29]. Taken together, results suggest that such state-specific traits are likely to be qualitatively different and may prove informative in explaining the etiology of disease.

We combined data measuring the metabolic response to a liquid mixed meal and measures of the exposure, BMI, in observational and MR frameworks. MR is a framework using instrumental variable (IV) analysis to deploy genetic variants as exposure proxies in efforts to obtain evidence of causal relationships (Figure 1) [30,31]. These analyses, undertaken in the Netherlands Epidemiology of Obesity (NEO) study, aimed to estimate the relationship between variation in BMI and

metabolite abundance in both fasted and postprandial states and to assess metabolite response—or change in abundance, between fasted and postprandial states.

## Methods

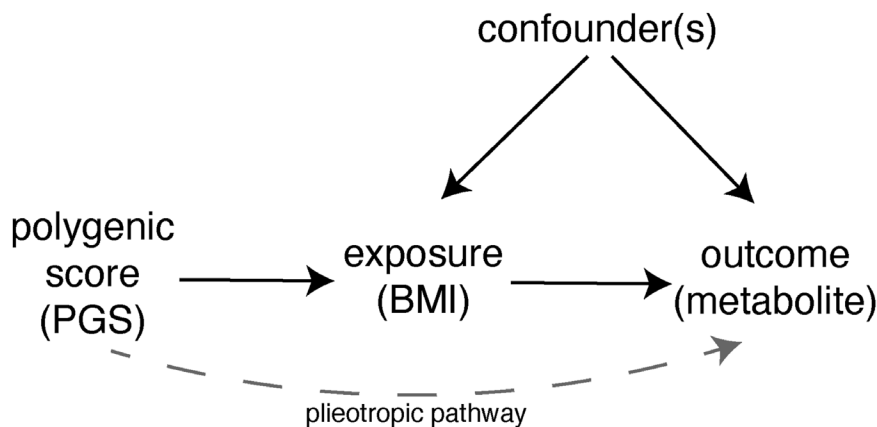
### Study sample

The sample population was 6671 participants of the NEO study, a prospective cohort study [32]. The NEO study was conducted at the Leiden University Medical Center (LUMC) and approved by the LUMC Medical Ethical Committee. All participants gave written informed consent. Recruitment was conducted between September 2008 and September 2012 by invitation, and sample size is a product of population response.

The NEO sample population is derived from 2 populations. First, all individuals between the ages of 45 and 65 y, living in the greater Leiden area, with a self-reported BMI of 27 or higher were invited to participate. Second, all individuals in the municipality of Leiderdorp of the same age range, regardless of BMI (Supplemental Figure 1) were invited to participate. We will refer to these 2 sample populations as “Leiden” and “Leiderdorp,” respectively, and the entire study sample as “NEO.” In this study, our study sample population had a size of 5744 defined by those participants with genotype data, metabolite data, and other phenotype or covariable data available. Further details of the NEO study and study design are previously published [32].

### Meal challenge

An overnight fasting was requested of all participants who subsequently visited the LUMC NEO study center and given a food challenge [29]. During the visit, first, an initial baseline, fasted blood sample was drawn. Second, within 5 min of the initial blood draw, participants were given a liquid mixed meal. The meal was 400 mL in volume with 600 kcal of energy, of which 16% of energy was derived from protein, 50% from carbohydrates, and 34% was from fat. Third, after completion of the food challenge, a blood sample was drawn at 30 min and again at 150 min. The NEO meal challenge occurred between the months of September 2008 and September 2012, with participants from Leiden undergoing sampling throughout this period. The participants from Leiderdorp were underwent sampling nonexclusively during the months of June 2011 and September 2012.



**FIGURE 1.** A schematic or directed acyclic graph of the Mendelian randomization (MR) framework. The MR framework in this study assumed that 1) the instrument is robustly associated with the outcome, 2) the instrument is not associated with variables that are associated to the outcome, and 3) the instrument acts on the outcome only through the exposure and not through a pleiotropic pathway.

The meal challenge sampling of Leiderdorp and Leiden population participants was structured by sampling date, and as a product, BMI was structured by sampling date. Specifically, BMI was associated with the date of measurement (univariable linear model  $F$  test  $P$  value =  $4.7 \times 10^{-229}$ ), explaining 20.3% of the total variation in BMI (Supplemental Figure 2). Despite this, the BMI polygenic score (BMI-PGS) had a smaller proportion (2.0%) of its variation explained by sampling date (univariate linear model  $F$  test  $P$  value =  $3.9 \times 10^{-7}$ ). As a product of this structure, sampling date was included as a covariate in all observational and MR analyses.

### Metabolite data

NEO participants had plasma-derived metabolite profiles measured using the Nightingale Health (Helsinki, Finland)  $^1\text{H}$  nuclear magnetic resonance platform [33]. For each participant, both their fasting and 150-min postprandial samples were assayed. We note here that metabolites are commonly defined as biological molecules of  $<1.5$  kDa in size and that many of the molecules (lipids) assayed in this study are larger than this threshold. For simplicity, we will refer to them all as metabolites. At the time of sampling, this metabolomics platform provided 229 measurements for 149 metabolites, including 80 derived ratio measurements (Supplemental Table 1) from 14 substance classes: amino acids ( $n = 8$ ), apolipoproteins ( $n = 3$ ), cholesterol ( $n = 9$ ), fatty acids ( $n = 11$ ), fatty acids ratios ( $n = 8$ ), fluid balance ( $n = 2$ ), glycerides and phospholipids ( $n = 9$ ), glycerides and phospholipid ratios ( $n = 2$ ), glycolysis-related metabolites ( $n = 3$ ), inflammation ( $n = 1$ ), ketone bodies ( $n = 2$ ), lipoprotein particle size ( $n = 3$ ), lipoprotein subclasses ( $n = 98$ ), and lipoprotein subclass ratios ( $n = 70$ ).

### Data quality control

Prior to data analysis, metabolites and covariable quality control (QC) and data filtering steps were implemented. First, a single individual who self-described themselves as “other” rather than “White” was excluded from the analysis. Second, any nonmetabolite covariable with  $<1000$  observations ( $n = 1$ ), with no variation ( $n = 1$ ), or binary covariables with fewer than 10 observations in either of the 2 binary classes ( $n = 1$ ) were removed. A complete list of evaluated nonmetabolite covariables is available in Supplemental Table 2.

The initial metabolite data set consisted of 5744 individuals, 229 fasting metabolites, 229 postprandial metabolites, and 148 previously derived [29] orthogonal nonlinear least squares (ornls) response metabolite traits. Data QC of this data set was performed with the R package *metaboprep* [34]. A complete description of the procedure and parameters used with *metaboprep* can be found in Supplemental Methods. After data QC, 226 samples and 3 metabolites (all ornls response traits) were filtered from the data set. The log file and report generated by *metaboprep* are provided in Supplemental Materials. After running *metaboprep*, we also performed the following steps on the fasted and postprandial data. First, all zero values were converted into not applicable (NA) values. Second, for each metabolite (in the fasting and postprandial state, individually) all values of 10 interquartile distances from the median were also turned into NA values. Third, we used the expected correlated nature of the fasting and postprandial data to identify outliers of that relationship and turned them into NA values, in both dietary states. Further details on this procedure can be found in Supplemental Methods along with an illustration in Supplemental Figure 3.

### Response trait

A response trait, or a measure of change between the postprandial and fasting dietary state, was derived for each of the 229 metabolites traits by univariable Deming regression. A Deming regression was used in this study because it allows for error in both independent (fasting) and dependent (postprandial) variables of the model and, as such, is a type of total least squares regression. Following the above-described QC steps, we fit a Deming regression of postprandial on fasting data, for each metabolite. After model fitting, the residuals were extracted and used as the response trait. The function `deming()` from the `deming` (v1.4) R package was used in the analysis. The intercept and slope of Deming regression were recorded (Supplemental Table 1), along with trait annotations and population summary statistics for all metabolite traits.

The residuals of the Deming regressions derived in this study were used as metabolite traits of response. These response traits effectively mirror those of a simple delta, in that positive values indicate an increase in metabolite abundance in the postprandial state and negative values would indicate a decrease. Effect estimates from an association analysis with BMI, described further, therefore suggests whether response is associated with BMI. Positive effect estimates suggest that as BMI increases, so too does the change in metabolite abundance. Negative effect estimates suggest that as BMI increases the change in metabolite abundance decreases.

Deming regressions residuals were taken to be the most appropriate way to summarize response for second stage analyses—observational and MR—because it accounts for the measurement error in metabolite variation at both time points. However, in addition to the Deming regression, we also used 2 alternative methods to derive response. The first was a simple  $\Delta$  estimate (postprandial minus fasting abundance) and the second was an ornls regression. We note that simple  $\Delta$  values and Deming residuals correlate with a mean Pearson  $r$  of 0.94 (95% CI: 0.65, 0.99). The ornls response traits were derived and used previously in a genome-wide association study of metabolite response and was used in our metabolite QC steps described earlier [29]. Our result and discussion will focus exclusively on the Deming response traits, but all effect estimates for the  $\Delta$  and ornls response traits are provided in Supplemental Table 3.

### Effective number of tested metabolites

Although there were 687 metabolite traits (fasting = 229, postprandial = 229, response = 229), these were not independent (Supplemental Figure 4). We used the R package “identification of principal variable” (iPV; <https://github.com/hughesevoanth/iPVs>) to estimate the effective number of metabolites in the data set [35–37]. This allowed us to ensure that we were not overcorrecting for the number tests we performed in the study. We estimated 43 representative variables in the NEO metabolite data set. The data-reduced, study-wide Bonferroni-corrected  $P$  value was set to  $0.05/43$  or  $1.163 \times 10^{-3}$ . A full description of the iPV procedure and parameters can be found in Supplemental Methods.

### Metabolite data description

Two analyses were carried out to describe this Nightingale Health metabolomics data set. First, given the abundance of lipoproteins and their lipids in this data set, we estimated the mean abundance of each lipid for each lipoprotein in both the fasting and postprandial dietary states. Second, a paired Student  $t$  test was performed to determine whether mean abundances differed between the fasting and

postprandial states. Our threshold for declaring a change in mean abundance was 0.05/229, where 229 is the number of metabolites tested. We also estimated the change as a  $\Delta$  value between the postprandial and fasting states and then derived estimates for mean change and the 95% CI of the change distribution.

### Rank normal transformation of metabolite traits

Each metabolite trait distribution was tested for normality with a Shapiro-Wilk test. All W-statistics (untransformed distributions and model residuals) are reported in [supplemental tables](#). In total, 55.56% of all metabolite traits had a W-statistics <0.95, a threshold used in this study to define an inconsistency with normality. A total of 43.94% of all log-transformed metabolite traits had a W-statistics of <0.95. Consequently, for the purpose of signal discovery and to allow for parametric analysis, each metabolite was rank normal transformed (tied values randomly ranked), prior to linear modeling. These steps resulted in the residuals from all observational and MR models to have a W-statistic >0.9 and >0.99, respectively.

### Covariables

Variables from 14 categories were compiled and available to identify possible confounders by evaluating associations with BMI, BMI-PGS, and metabolite traits ([Supplemental Table 2](#)). These 14 categories are as follows: anthropomorphic ( $n = 8$ ); education ( $n = 5$ ); income ( $n = 2$ ); smoking ( $n = 3$ ); diet including alcohol intake in grams ( $n = 24$ ); medication including glucose and lipid lowering ( $n = 10$ ); systolic and diastolic blood pressure ( $n = 2$ ); health including diabetes, glucose, hypertension, and cancer status ( $n = 4$ ); physical exercise ( $n = 18$ ); basal metabolic rate ( $n = 1$ ); indirect calorimetry ( $n = 5$ ); and genotype principal components ( $n = 4$ ). In addition, 7 metrics of biological sample quality ( $n = 7$ ) and the sampling information (visit date and subpopulation,  $n = 2$ ) were included in the study to assess influence on metabolites. During QC, described earlier, 3 variables were filtered [Nightingale Health low protein flag, education type, medication C07B ( $\beta$ -blockers)]. As performed with aforementioned metabolites, we estimated the effect number of covariables in the data set. After running an iPV analysis with a tree cut height of 0.5, we estimated that there are 54 representative or independent covariables. As a product any test performed across all covariables was corrected for multiple testing at a  $P$  value of 0.05/54 or  $9.29 \times 10^{-4}$ .

### Genotype data

The NEO participants were genotyped on the Illumina HumanCoreExome-24 BeadChip (Illumina), at the Centre National de Génomique (Evry Cedex, France). Following genotype QC steps, as previously described [38], sample were imputed to the Haplotype Reference Consortium release 1.1 [39], with further details in the study by Li-Gao et al. [29].

### Construction of the BMI-PGS

BMI-PGS was derived for each individual using variants previously and independently associated with BMI ( $n = 656$ ) at a  $P$  value of  $<1 \times 10^{-8}$  from the study by Yengo et al. [40]. To ensure the integrity of the PGS, only those genetic variants, in the NEO genotype dosage data set, where the effect allele, the alternative allele, and the minor allele frequency ( $\pm 0.1$ ) matched the data from Yengo et al. [40] were retained. If a match could not be made, then that single nucleotide polymorphism (SNP) was omitted. All effect estimates were aligned to be positive,

such that the effect allele was always that which increased an individual's genotype predicted BMI. Finally, a weighted PGS was constructed by weighing the number of BMI-increasing alleles, at each SNP ( $n = 646$ , after data harmonization), by the effect estimate for that SNP and then summing across all values [40]. The R script "01\_generate\_bmi\_grs.R" for deriving the PGS can be found in the github repository [https://github.com/hughesevoanth/NEO\\_BMI\\_Metabolite\\_MR](https://github.com/hughesevoanth/NEO_BMI_Metabolite_MR).

### Observational and MR analysis

Two observational association analyses were undertaken in cross-section. The first was an association analysis between individual metabolite traits and BMI. The second uses a PGS as a proxy instrument for the exposure in a MR framework ([Figure 1](#)) to provide an estimate of the causal effect of the exposure on outcome (in this study, dietary state-specific metabolite values). We will refer to the first as an observational analysis and the later as an MR analysis.

Observational analyses were performed by a multivariable generalized linear regression with sampling date, subpopulation, age, and sex as additional covariables—taking the form of  $\text{glm}(\text{metabolite} \sim \text{visit date} + \text{subpopulation} + \text{age} + \text{sex} + \text{BMI})$ . For each generalized linear model, we also performed a Breusch-Pagan test of homoskedasticity using the `bptest()` function from the `lmtest` R package [41]. In addition, for each association analysis, we performed a type I ANOVA, producing a table of deviances and estimated an  $\eta^2$  statistic, providing an estimate of the variance explained, for the model and for the primary exposure, BMI. These estimates are reported in [supplemental tables](#).

One-sample MR analyses were performed using the `ivreg` R package (<https://cran.r-project.org/web/packages/ivreg/>), implementing a 2-stage least squares IV analysis. In all instances and given the relative viability of BMI instruments versus reciprocal metabolite instrumentation, BMI was defined as our exposure; the PGS described earlier was our instrument for BMI, and each metabolite trait was iteratively defined as the outcome. The same model and covariables as described earlier in the association analysis was used here. Along with MR effect estimates, we also report in [supplemental tables](#) the Breusch-Pagan test of homoskedasticity, the  $F$  statistic and  $F$  test  $P$  value testing for weak instruments, and summary statistics for the Durbin-Wu-Hausman endogeneity test.

Unless stated otherwise—because our outcomes are rank normal transformed (zero centered with a standard deviation of 1)—the effect estimates ( $\beta$ ) and standard errors (SEs) are reported as rank-normalized standard deviation unit change per unit cross-sectional increase in BMI. We will refer to increases in BMI or increases in metabolite abundances throughout the article to infer cross-sectional increases in BMI, not individual-level units of change.

### Study structure

In all observational and MR analyses, 4 (sub)population analyses were performed. The primary analysis, in all instances, was a weighted Netherlands Epidemiology of Obesity (wNEO) regression analysis using data from all available samples. wNEO observational and MR summary statistics are reported in [Supplemental Table 4](#). The wNEO analyses included weights for each sample that were previously derived to have the Leiden subpopulation BMI distribution emulate that of the randomly sampled Leiderdorp subpopulation [42]. Sensitivity analyses included the following: 1) a wNEO analysis with additional covariables to account for possible confounders

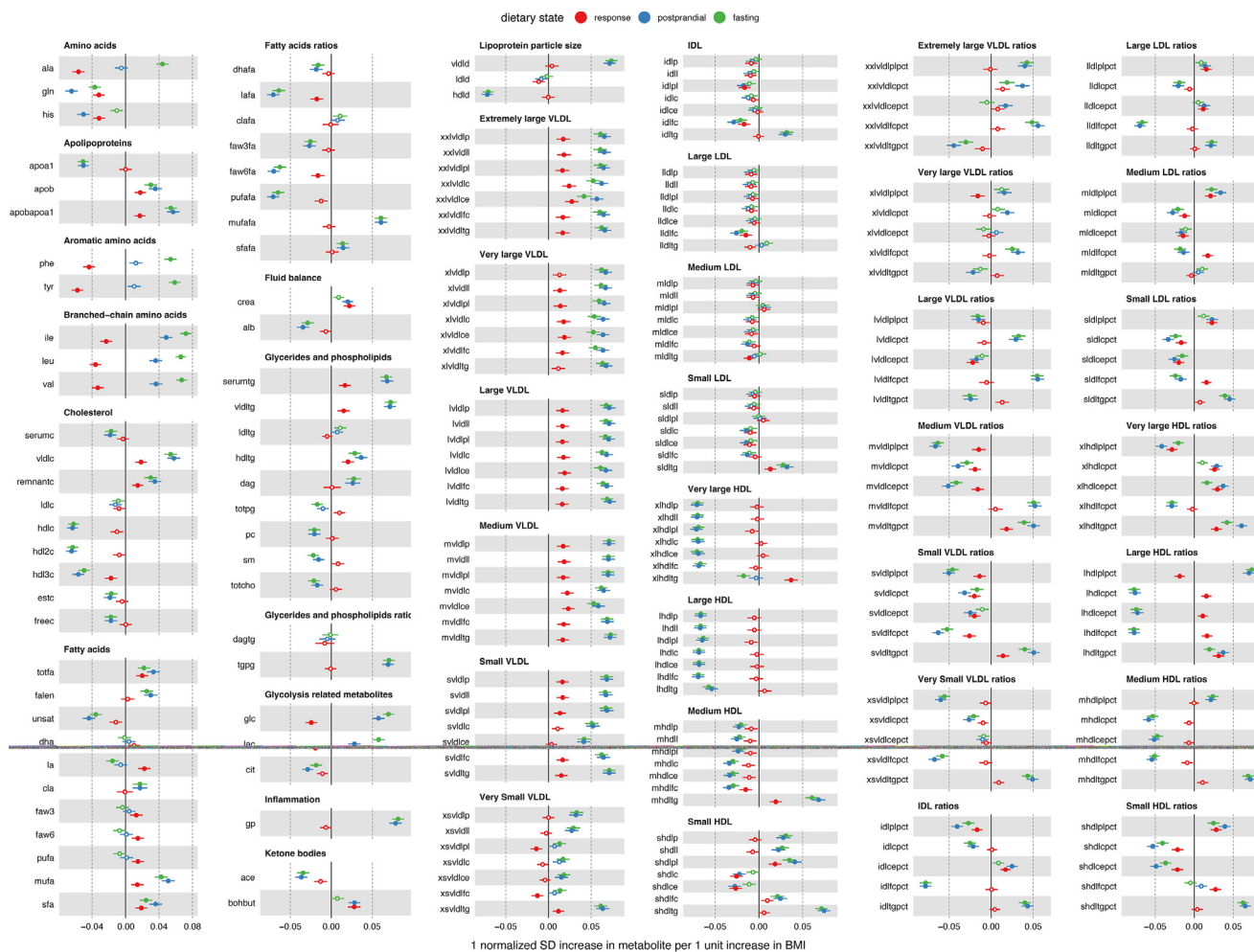
**TABLE 1**  
Population Summary Statistics

Trait	NEO	Males	Females	Leiderdorp	Leiden	Subpopulation $R^2$ (%)	$P$
Sample size	5571	2671	2846	1406	4111	NA	NA
Females ( $n$ )	2846	0	2846	783	2063	0.17	$3.65 \times 10^{-4}$
Age (y)	55.99 (45.00–65.00)	56.13 (45.00–65.00)	55.86 (45.00–65.00)	56.13 (46.00–65.00)	55.95 (45.00–65.00)	0.02	$3.26 \times 10^{-1}$
Height (m)	1.74 (1.57–1.92)	1.81 (1.68–1.95)	1.67 (1.55–1.79)	1.73 (1.57–1.92)	1.74 (1.57–1.92)	0.03	$2.03 \times 10^{-1}$
Weight (kg)	90.58 (60.00–126.80)	97.45 (72.95–131.60)	84.14 (57.20–121.52)	79.16 (54.00–116.15)	94.49 (69.60–128.85)	16.12	$9.63 \times 10^{-213}$
Waist circumference (cm)	102.01 (74.00–129.00)	106.32 (85.79–131.00)	97.96 (71.00–126.00)	91.22 (70.00–119.00)	105.70 (86.00–131.00)	22.87	$2.85 \times 10^{313}$
Hip circumference (cm)	110.23 (94.00–133.00)	108.57 (96.00–125.00)	111.79 (92.00–138.00)	103.40 (90.00–124.17)	112.57 (99.00–135.00)	16.22	$3.33 \times 10^{-214}$
Pack-y	11.21 (0.00–52.38)	13.71 (0.00–61.69)	8.88 (0.00–43.83)	8.69 (0.00–43.32)	12.09 (0.00–55.20)	0.83	$1.01 \times 10^{-10}$
Alcohol	15.52 (0.00–60.52)	21.40 (0.00–73.24)	10.00 (0.00–41.72)	14.58 (0.00–54.06)	15.84 (0.00–62.57)	0.10	$1.95 \times 10^{-2}$
Smoking (never/former/current)	34.11/49.85/16.03	30.32/50.49/19.19	37.67/49.26/13.07	39.29/46.76/13.95	32.34/50.91/16.74	0.38	$4.11 \times 10^{-6}$
Higher education (%)	38.58	42.27	35.12	50.5	34.48	1.52	$6.24 \times 10^{-26}$
BMI ( $\text{kg}/\text{m}^2$ )	29.98 (21.40–41.35)	29.74 (22.72–38.99)	30.20 (20.64–42.90)	26.24 (19.94–37.20)	31.26 (26.02–42.19)	20.70	$3.66 \times 10^{-280}$
BMI-PGS	10.20 (9.64–10.78)	10.20 (9.63–10.78)	10.20 (9.65–10.77)	10.14 (9.56–10.71)	10.22 (9.67–10.80)	1.63	$1.53 \times 10^{-21}$
$R^2$	0.0431, 0.0473 <sup>1</sup>	0.0439	0.0453	0.0427	0.0243	—	—
$\beta$ (SE)	3.38 (0.21), 3.26 (0.20) <sup>1</sup>	2.77 (0.25)	4.00 (0.34)	3.14 (0.40)	2.25 (0.22)	—	—
$P$	$8.10 \times 10^{-55}$ ; $3.61 \times 10^{-60.1}$	$7.23 \times 10^{-28}$	$1.57 \times 10^{-30}$	$5.10 \times 10^{-15}$	$8.22 \times 10^{-24}$	—	—

Population summary statistics for 1) the complete NEO cohort, 2) males, and 3) females of the NEO cohort and the 2 NEO subpopulations; 4) Leiderdorp; and 5) Leiden. For each variable—age, height, weight, waist circumference, hip circumference, smoking in pack-years, alcohol consumption in grams per day, smoking (never/former/current), the percentage of the population with higher education, BMI, and BMI-PGS—the mean are provided along with the (95% quantile intervals) of all distributions. In addition, the number of females in each sample (sub)population is provided. The last 3 rows of data provide summary statistics describing the univariable linear relationship between the instrumental variables (BMI-PGS) on the exposure (BMI). Specifically, the variance explained ( $R^2$ ), the effect estimate ( $\beta$ ) and standard error (SE), and the  $P$  value as derived by univariable linear regression are provided. The last 2 columns of data provide an estimate of the proportion of variation (in NEO) of the trait (in the row) explained by variation between the subpopulation (Leiderdorp vs. Leiden) samples (subpopulation  $R^2$ ) and the model  $P$  value, as estimated by a univariable linear regression.

Abbreviations: NEO, Netherlands Epidemiology of Obesity Study; PGS, polygenic score.

<sup>1</sup> Estimates are derived from linear models that include sample weights (wNEO; see Methods and Study structure).



**FIGURE 2.** Observational effect estimate forest plot. Point estimates ( $\beta$ ) and 95% CIs for BMI on fasting metabolite abundance (green), postprandial metabolite abundance (blue), and metabolite response (red). Estimates are derived from the weighted Netherlands Epidemiology of Obesity (wNEO) multi-variable generalized linear regression that include sample date, subpopulation, age, and sex as covariables. Metabolites are classified and organized by Nightingale Health subclass assignments. Metabolites that are associated with BMI, after multiple test correction ( $P < 0.05/43$ ), are indicated as solid color point estimates. Effect estimates are reported as rank-normalized standard deviation unit change per unit increase in BMI ( $\text{kg}/\text{m}^2$ ). All results can be found in Supplemental Table 4. Metabolite annotations and abbreviations can be found in Supplemental Table 1.

(Supplemental Table 5); 2) a wNEO analysis with no transformation of the outcome traits (Supplemental Table 6); 3) an unweighted NEO analysis (Supplemental Table 7); 4) a Leiderdorp subpopulation analysis (Supplemental Table 8); and 5) a Leiden subpopulation analysis (Supplemental Table 9). In addition, sex-specific analyses were performed using the primary wNEO framework (Supplemental Tables 10 and 11) and the Leiderdorp subpopulation (Supplemental Tables 12 and 13), as neither have a sampling bias in their BMI distributions.

**Data and code availability**

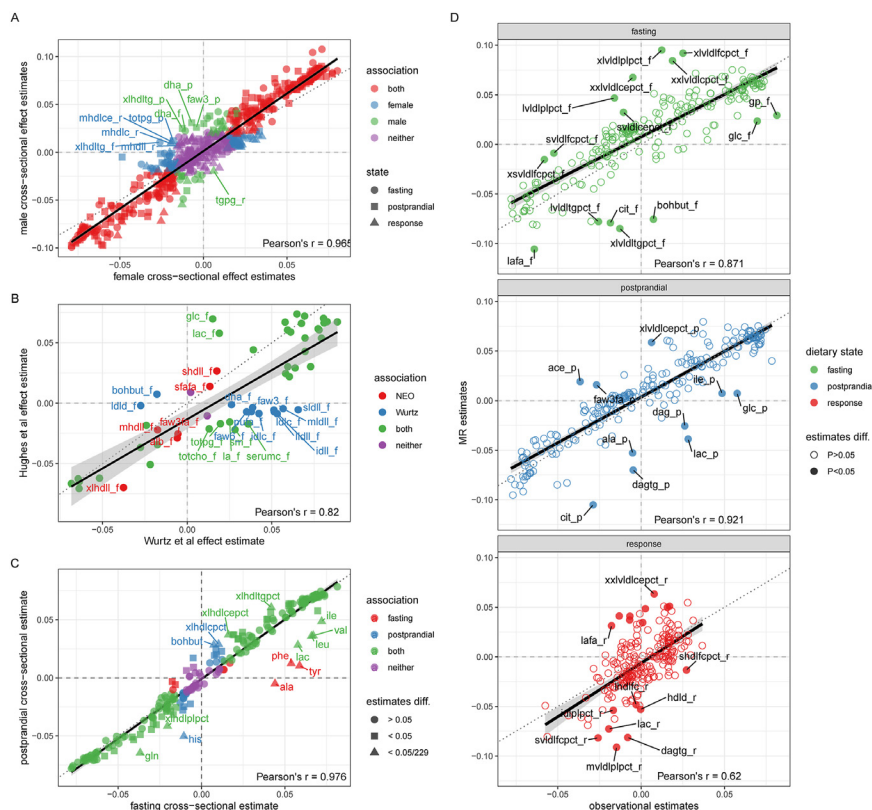
Individual-level NEO data are available to researchers according to the NEO research procedure, which can be obtained by contacting the scientific director of the NEO study for access [32]. All statistical analysis were conducted in the R language (v 4.0.2, Taking Off Again), and all bespoke functions and analytical code for the study can be found in the github repository [https://github.com/hughesevoanth/NEO\\_BMI\\_Metabolite\\_MR](https://github.com/hughesevoanth/NEO_BMI_Metabolite_MR).

**Results**

**NEO cohort description**

Following QC and derivation of the response trait, the data set consisted of 5517 individuals (51.6% female), 687 metabolite traits (229 fasting, 229 postprandial, and 229 response), and 85 covariables. The mean age of participants was 56 y, with an mean BMI of 29.98 (Table 1).

The NEO cohort is made up of 2 distinct subsamples, (1: “Leiderdorp”) a randomly sampled population of 1406 individuals from the municipality of Leiderdorp (mean BMI = 26.24, 95% quantile interval: 19.94, 37.2), and (2: “Leiden”) a sample population of 4111 participant from the city of Leiden who were oversampled for individuals with a BMI >27 (mean BMI = 31.26, 95% quantile interval: 26.02, 42.19). BMI distributions differed between the 2 subpopulations (generalized linear model Student *t* test  $P$  value =  $3.66 \times 10^{-280}$ ) (Supplemental Figure 1). In addition, the 2 subpopulations also differ in sex ratio, alcohol intake (in grams per day), smoking (pack-years), educational



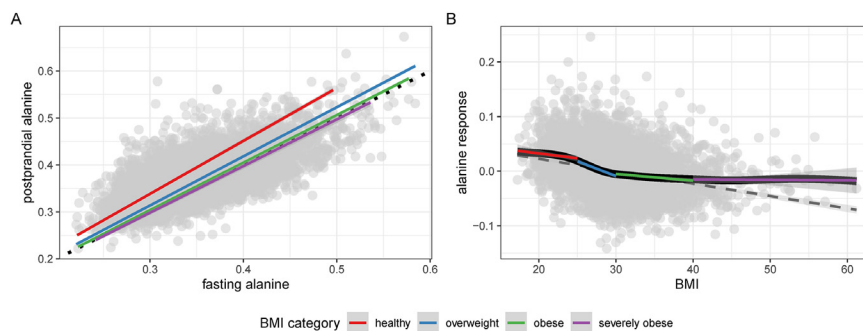
**FIGURE 3.** Scatter plots. In all plots, the solid black line is a best-fit regression line, the dotted black line is an equivalency line (intercept = 0, slope = 1), the dashed vertical and horizontal line indicate the zero values, and a Pearson correlation coefficient is in the bottom right corner. (A) Scatter plot of BMI–metabolite cross-sectional effect estimates for females (x-axis) and males (y-axis). Point estimates are colored red if there is an association between the metabolite and BMI in both sexes, blue if an association was only observed in females, green if an association was only observed in males, and purple if no association was observed in either sex. Point estimates are shaped as circles, squares, and triangles to represent the 3 dietary states fasting, postprandial, and response, respectively. All results for plot A can be found in [Supplemental Tables 10 and 11](#). (B) Scatter plot of BMI–metabolite cross-sectional effect estimates from Wurtz et al. [15] (x-axis) and Netherlands Epidemiology of Obesity (NEO; this study, y-axis). Point estimates are colored red if there is an association between the metabolite and BMI in the NEO study, blue if an association was only observed in the study by Wurtz et al., green if an association was observed in both studies, and purple if no association was observed in either study. (C) Scatter plot of BMI–metabolite cross-sectional effect estimates for fasting (x-axis) and postprandial (y-axis) dietary states. Point estimates are colored red if there is an association between the metabolite and BMI in the fasting state, blue if an association was only observed in the postprandial state, green an association was observed in both dietary states, and purple if no association was observed in either. Point estimates are shaped as circles if they are not different between the 2 dietary states—as determined by a *z* test, shaped as squares if they are nominally different ( $P < 0.05$ ), and triangles and labeled if they are different after correcting for multiple tests ( $P < 0.05/229$ ). All results for plot C can be found in [Supplemental Table 4](#). (D) Scatter plot of BMI–metabolite cross-sectional (x-axis) and Mendelian randomization (MR) (y-axis) effect in the fasting (top and green), postprandial (middle and blue), and response (bottom and red) dietary states. Metabolites with effect estimates that differ between the cross-sectional and MR analyses are solid circles and labeled with the metabolite name. All results for plot D can be found in [Supplemental Table 4](#). Metabolite annotations and abbreviations can be found in [Supplemental Table 1](#).

attainment, and BMI-PGS (generalized linear model Student *t* test *P* value of  $< 1.95 \times 10^{-2}$ ) but not in age or height ([Table 1](#)). Although the variance explained by subpopulations is large for BMI (20.7%) and closely related traits (weight, hip and waist circumference), it is an order smaller for other variables such as BMI-PGS, pack-years, and educational attainment (variance explained by subpopulation  $< 2\%$ ) ([Table 1](#)).

### Description of metabolite data

Of the 229 assayed metabolites 98 are lipoproteins and their particle concentrations and an additional 70 are lipoprotein ratios. The expected lipid concentrations were observed across lipoprotein classes ([Supplemental Figure 5](#)). Namely, triglycerides dominate VLDLs, cholesterol dominates intermediate density lipoprotein (IDL) and LDLs, and phospholipids and cholesterol dominate high density lipoproteins

(HDLs), on average. This remains true in both the fasting and postprandial data. All except 14 metabolites differed in mean concentration or ratio between the fasted and postprandial states (paired *t* test,  $P < 0.05/229$ ) ([Supplemental Table 14](#)). These 14 include the concentration of particles: total lipids, total cholesterol, cholesterol esters, and free cholesterol in medium LDL. It also includes total cholesterol in LDL, large LDL, and small LDL, and cholesterol esters in small LDL. Overall, 126 metabolites increased in abundance, and 89 decreased postprandially. Those with the largest centered and scaled, mean increase were the amino acids tyrosine, leucine, valine, isoleucine, and phenylalanine ([Supplemental Figure 6](#)). Those with the largest mean decrease were the ratio of saturated fatty acids to total fatty acids, the ratio of triglycerides to total lipids in very large VLDL, the ratio of phospholipids to total lipids in small VLDL and the ketone body 3-hydroxybutyrate. For all metabolites, 150-min postprandial change



**FIGURE 4.** Alanine response. (A) Scatter plot illustrating the correlation between fasting (x-axis) and postprandial (y-axis) alanine abundance, with a best-fit line for 4 BMI ( $\text{kg}/\text{m}^2$ ) classes of individuals: healthy ( $\text{BMI} < 25$ ), overweight ( $25 \leq \text{BMI} < 30$ ), obese ( $30 \leq \text{BMI} < 40$ ), and severely obese ( $\text{BMI} > 40$ ). All best-fit lines were forced through an intercept of zero. Note that the slope of the best-fit line is the largest for individuals who have a healthy BMI and the smallest for those who have severe obesity. (B) Scatter plot illustrating the inverse relationship between BMI (x-axis) and alanine response (y-axis). The blue line is a best-fit linear regression line; the black line is a best-fit generalized additive model smooth (nonlinear regression); and the red, blue, green, and purple line intervals are best-fit regression for each clinically defined portion of the BMI distribution, which mirror the nonlinear regression.

was not in the same direction for all participants (Supplemental Figure 6).

### Estimating the observational effect of BMI on metabolite trait variation

wNEO observational analyses suggested broad association between BMI and metabolite traits, with 473 metabolite traits (across fasting, postprandial, and response states) showing evidence of association with BMI ( $P < 1.163 \times 10^{-3}$ ) (Figure 2, Supplemental Table 4, Supplemental Figure 7). Effect estimates correlated between sexes (Pearson  $r = 0.965$ ), but more associations were observed in females (426 total: 162 fasting, 175 postprandial, 89 response) than males (369 total: 161 fasting, 168 postprandial, 40 response) (Supplemental Figure 4). A total of 333 associations were shared between females and males, with 93 specific to females, 36 specific to males, and 225 metabolite traits having no association with BMI in either sex (Supplemental Tables 10 and 11). Among the associated traits all but 10 are directionally consistent among females and males (Figure 3A). Larger effect sizes were observed in males (mean absolute  $\beta$ : 0.0356) than those in females (mean  $\beta$ : 0.0291), across all traits on average (Figure 3A), but the variance explained by BMI, across all traits on average was similar among males and females ( $\eta^2$ : males = 1.737%, females = 1.734%).

### Observational results—fasting

A total of 175 or 76% of fasting traits showed evidence of association with BMI. The strongest association observed was with the inflammation marker glycoprotein acetyls ( $\beta$ : 0.081; SE: 0.0037;  $P = 2.89 \times 10^{-101}$ ). This was followed closely by the branched-chain amino acids (BCAA) isoleucine ( $\beta$ : 0.072; SE: 0.0033;  $P = 4.18 \times 10^{-101}$ ), leucine ( $\beta$ : 0.066; SE: 0.0031;  $P = 1.16 \times 10^{-94}$ ), and valine ( $\beta$ : 0.067; SE: 0.0032;  $P = 8.70 \times 10^{-92}$ ). The strongest inverse associations were observed for mean diameter of HDL particles ( $\beta$ : -0.071; SE: 0.0035;  $P = 6.74 \times 10^{-95}$ ), the ratio of free cholesterol to total lipids in IDL ( $\beta$ : -0.078; SE: 0.0036;  $P = 1.04 \times 10^{-100}$ ), and large HDL ( $\beta$ : -0.077; SE: 0.0037;  $P = 8.09 \times 10^{-95}$ ). If data are summarized by metabolite annotation class (Supplemental Table 1), we find that the classes with the largest mean absolute BMI effect are inflammation (mean absolute  $\beta$ : 0.081), amino acids (0.051), lipoprotein particle size (0.049), and glycolysis-related metabolites (0.049).

Those with the smallest mean absolute effect are ketone bodies (0.021), fluid balance (0.019) and fatty acids (0.018) (Supplemental Figure 7).

### Observational results—postprandial

A total of 188 or 82% of postprandial traits showed evidence of association with BMI. Fasting and postprandial point estimates correlated strongly (Pearson  $r = 0.976$ ) with each other, with no mean difference in their distributions ( $t$  test,  $P = 0.83$ ) and 166 shared associations. However, 31 associations are unique to 1 of the 2 dietary states, with 9 associations specific to fasting and 22 specific to postprandial data (Figure 3C). We tested for a difference in effect estimates between the dietary states and observed 14 estimates to differ ( $z$  test,  $P < 0.05/229$ ) (Figure 3C). Eight of those 14 are amino acids (tyrosine, alanine, phenylalanine, histidine, valine, leucine, glutamine, and isoleucine). The other 6 that differ between dietary states include 4 very large HDL ratios, the ketone  $\beta$ -hydroxybutyrate, and the glycolysis-related metabolite lactate.

### Observational results—response

A total of 70 response traits showed evidence of a positive association with BMI and 40 traits had an inverse association. Response and fasting (Pearson  $r = 0.34$ ;  $P = 1.39 \times 10^{-7}$ ) and response and postprandial (Pearson  $r = 0.514$ ;  $P = 7.96 \times 10^{-17}$ ) point estimates modestly correlated with each other, indicative of the relative independence of the response trait. The strongest associations were all amino acids, which (ordered by association  $P$  value: tyrosine, alanine, phenylalanine, leucine, valine, glutamine, histidine, and isoleucine) were all inversely associated with BMI (Supplemental Figure 7). For example, although alanine levels (fasting mean = 0.359 mmol/L) were elevated after a liquid mixed meal (postprandial mean = 0.376 mmol/L; paired  $t$  test,  $P = 2.53 \times 10^{-181}$ ), the effect attenuated as BMI increased ( $\beta$ : -0.056; SE: 0.004;  $P = 5.72 \times 10^{-51}$ ) (Figure 4A) such that the effect of BMI on alanine response was negative (Figure 4B). This relationship was observed for all amino acids in the data set (Figure 2).

The strongest positive response association was observed for triglycerides in very large HDL ( $\beta$ : 0.037; SE: 0.004;  $P = 5.88 \times 10^{-21}$ ). This was followed closely by the ratio of 1) cholesterol esters, 2) triglycerides, and 3) total cholesterol to total lipids in very large HDL. Conversely, the response trait ratio of phospholipids to total lipids in very large HDL decreased as BMI increased. This suggests that after a



phospholipids in large VLDL to total lipids in large VLDL ratio; lvldltg, triglycerides in large VLDL; lvldltgpcpt, triglycerides in large VLDL to total lipids in large VLDL ratio; mhdhc, total cholesterol in medium HDL; mhdhce, cholesterol esters in medium HDL; mhdhcepcpt, cholesterol esters in medium HDL to total lipids in medium HDL ratio; mhdhcepcpt, total cholesterol in medium HDL to total lipids in medium HDL ratio; mhdhfc, free cholesterol in medium HDL; mhdhfcpcpt, free cholesterol in medium HDL to total lipids in medium HDL ratio; mhdhll, total lipids in medium HDL; mhdhlp, concentration of medium HDL particles; mhdhlp, phospholipids in medium HDL; mhdhlp, phospholipids in medium HDL to total lipids in medium HDL ratio; mhdhltg, triglycerides in medium HDL; mhdhltgpcpt, triglycerides in medium HDL to total lipids in medium HDL ratio; mldlc, total cholesterol in medium LDL; mldlce, cholesterol esters in medium LDL; mldlcepcpt, cholesterol esters in medium LDL to total lipids in medium LDL ratio; mldlcepcpt, total cholesterol in medium LDL to total lipids in medium LDL ratio; mldlfc, free cholesterol in medium LDL; mldlfcpcpt, free cholesterol in medium LDL to total lipids in medium LDL ratio; mldlll, total lipids in medium LDL; mldlpl, concentration of medium LDL particles; mldlpl, phospholipids in medium LDL; mldlplpct, phospholipids in medium LDL to total lipids in medium LDL ratio; mldltg, triglycerides in medium LDL; mldltgpcpt, triglycerides in medium LDL to total lipids in medium LDL ratio; mufa, monounsaturated fatty acids; 16:1, 18:1; mufafa, Ratio of monounsaturated fatty acids to total fatty acids; mvldlc, total cholesterol in medium VLDL; mvldlce, cholesterol esters in medium VLDL; mvldlcepcpt, cholesterol esters in medium VLDL to total lipids in medium VLDL ratio; mvldlcepcpt, total cholesterol in medium VLDL to total lipids in medium VLDL ratio; mvldlfc, free cholesterol in medium VLDL; mvldlfcpcpt, free cholesterol in medium VLDL to total lipids in medium VLDL ratio; mvldlll, total lipids in medium VLDL; mvldlp, concentration of medium VLDL particles; mvldlp, phospholipids in medium VLDL; mvldlp, phospholipids in medium VLDL to total lipids in medium VLDL ratio; mvldltg, triglycerides in medium VLDL; mvldltgpcpt, triglycerides in medium VLDL to total lipids in medium VLDL ratio; ormls, orthogonal nonlinear least squares; pc, phosphatidylcholine and other cholines; phe, Phenylalanine; pufa, polyunsaturated fatty acids; pufafa, ratio of polyunsaturated fatty acids to total fatty acids; remnantc, remnant cholesterol; se, standard error; serumc, serum total cholesterol; serumtg, serum total triglycerides; sfa, saturated fatty acids; sfafa, ratio of saturated fatty acids to total fatty acids; shdlc, total cholesterol in small HDL; shdlce, cholesterol esters in small HDL; shdlcepcpt, cholesterol esters in small HDL to total lipids in small HDL ratio; shdlcepcpt, total cholesterol in small HDL to total lipids in small HDL ratio; shdlfc, free cholesterol in small HDL; shdlfcpcpt, free cholesterol in small HDL to total lipids in small HDL ratio; shdll, total lipids in small HDL; shdhp, concentration of small HDL particles; shdhp, phospholipids in small HDL; shdhp, phospholipids in small HDL to total lipids in small HDL ratio; shdltg, triglycerides in small HDL; shdltg\_f, fasting small HDL triglycerides; shdltgpcpt, triglycerides in small HDL to total lipids in small HDL ratio; sldlc, total cholesterol in small LDL; sldlce, cholesterol esters in small LDL; sldlcepcpt, cholesterol esters in small LDL to total lipids in small LDL ratio; sldlcepcpt, total cholesterol in small LDL to total lipids in small LDL ratio; sldlfc, free cholesterol in small LDL; sldlfcpcpt, free cholesterol in small LDL to total lipids in small LDL ratio; sldll, total lipids in small LDL; sldlp, concentration of small LDL particles; sldlp, phospholipids in small LDL; sldlp, phospholipids in small LDL to total lipids in small LDL ratio; sldlplpct, phospholipids in small LDL to total lipids in small LDL ratio; sldltg, triglycerides in small LDL; sldltgpcpt, triglycerides in small LDL to total lipids in small LDL ratio; sm, sphingomyelins; svldlc, total cholesterol in small VLDL; svldlce, cholesterol esters in small VLDL; svldlcepcpt, cholesterol esters in small VLDL to total lipids in small VLDL ratio; svldlcepcpt, total cholesterol in small VLDL to total lipids in small VLDL ratio; svldlfc, free cholesterol in small VLDL; svldlfcpcpt, free cholesterol in small VLDL to total lipids in small VLDL ratio; svdlll, total lipids in small VLDL; svldlp, concentration of small VLDL particles; svldlp, phospholipids in small VLDL; svldlp, phospholipids in small VLDL to total lipids in small VLDL ratio; svldltg, triglycerides in small VLDL; svldltgpcpt, triglycerides in small VLDL to total lipids in small VLDL ratio; tggp, ratio of triglycerides to phosphoglycerides ratio; tocho, total cholines; totfa, total fatty acids; totpg, total phosphoglycerides; tyr, tyrosine; unsat, estimated degree of unsaturation; val, valine; vldlc, total cholesterol in VLDL; vldld, mean diameter for VLDL particles; vldltg, triglycerides in VLDL; wNEO, weighted NEO regression analysis, the primary analysis framework; xlhdhc, total cholesterol in very large HDL; xlhdhce, cholesterol esters in very large HDL; xlhdhcepcpt, cholesterol esters in very large HDL to total lipids in very large HDL ratio; xlhdhcepcpt, total cholesterol in very large HDL to total lipids in very large HDL ratio; xlhdhfc, free cholesterol in very large HDL; xlhdhfcpcpt, free cholesterol in very large HDL to total lipids in very large HDL ratio; xlhdll, total lipids in very large HDL; xlhdhp, concentration of very large HDL particles; xlhdhp, phospholipids in very large HDL; xlhdhp, phospholipids in very large HDL to total lipids in very large HDL ratio; xlhdltg, triglycerides in very large HDL; xlhdltgpcpt, triglycerides in very large HDL to total lipids in very large HDL ratio; xlvdlc, total cholesterol in very large VLDL; xlvdlce, cholesterol esters in very large VLDL; xlvdlcepcpt, cholesterol esters in very large VLDL to total lipids in very large VLDL ratio; xlvdlcepcpt, total cholesterol in very large VLDL to total lipids in very large VLDL ratio; xlvdlfc, free cholesterol in very large VLDL; xlvdlfcpcpt, free cholesterol in very large VLDL to total lipids in very large VLDL ratio; xlvdlll, total lipids in very large VLDL; xlvdlp, concentration of very large VLDL particles; xlvdlp, phospholipids in very large VLDL; xlvdlp, phospholipids in very large VLDL to total lipids in very large VLDL ratio; xlvdltg, triglycerides in very large VLDL; xlvdltgpcpt, triglycerides in very large VLDL to total lipids in very large VLDL ratio; xsvldc, total cholesterol in very small VLDL; xsvldlce, cholesterol esters in very small VLDL; xsvldlcepcpt, cholesterol esters in very small VLDL to total lipids in very small VLDL ratio; xsvldlcepcpt, total cholesterol in very small VLDL to total lipids in very small VLDL ratio; xsvldlfc, free cholesterol in very small VLDL; xsvldlfcpcpt, free cholesterol in very small VLDL to total lipids in very small VLDL ratio; xsvdlll, total lipids in very small VLDL; xsvdhp, concentration of very small VLDL particles; xsvdhp, phospholipids in very small VLDL; xsvdhp, phospholipids in very small VLDL to total lipids in very small VLDL ratio; xsvdltg, triglycerides in very small VLDL; xsvdltg\_f, fasting very small VLDL triglycerides; xsvdltgpcpt, triglycerides in very small VLDL to total lipids in very small VLDL ratio; xxlvldc, total cholesterol in chylomicrons and extremely large VLDL; xxlvldlce, cholesterol esters in chylomicrons and extremely large VLDL ratio; xxlvldlcepcpt, cholesterol esters in chylomicrons and extremely large VLDL to total lipids in chylomicrons and extremely large VLDL ratio; xxlvldlcepcpt, total cholesterol in chylomicrons and extremely large VLDL to total lipids in chylomicrons and extremely large VLDL ratio; xxlvldlfc, free cholesterol in chylomicrons and extremely large VLDL; xxlvldlfcpcpt, free cholesterol in chylomicrons and extremely large VLDL to total lipids in chylomicrons and extremely large VLDL ratio; xxlvdlll, total lipids in chylomicrons and extremely large VLDL; xxlvdhp, concentration of chylomicrons and extremely large VLDL particles; xxlvdhp, phospholipids in chylomicrons and extremely large VLDL; xxlvdhp, phospholipids in chylomicrons and extremely large VLDL to total lipids in chylomicrons and extremely large VLDL ratio; xxlvdltg, triglycerides in chylomicrons and extremely large VLDL; xxlvdltgpcpt, triglycerides in chylomicrons and extremely large VLDL to total lipids in chylomicrons and extremely large VLDL ratio.

liquid mixed meal, the increase of phospholipids relative to total lipids in very large HDL is attenuated by BMI and, conversely, the ratio of cholesterol and triglycerides to total lipids in very large HDL increased as BMI increased (Figure 2).

Outside of amino acids and lipoproteins, other metabolite response traits that had a positive association with BMI included 3-hydroxybutyrate (ketone body), creatinine (fluid balance), linoleic acid, and total fatty acids (fatty acids). Others with an inverse association with BMI

included glucose and lactate (glycolysis), the ratio of linoleic acid to total fatty acids, and the ratio of  $\omega$ -6 (n-6) fatty acids to total fatty acids (fatty acid ratios) (Supplemental Table 4).

### Possible confounders

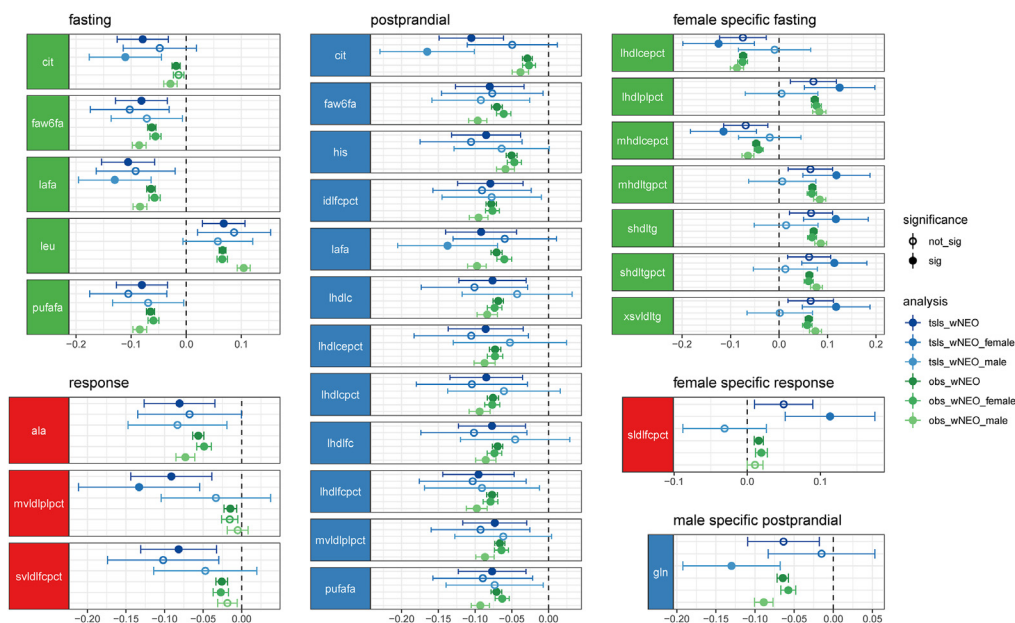
An exhaustive assessment of the association between BMI and BMI-PGS with 91 study covariables was performed, by univariable linear modeling and analysis of variance, to identify possible

**TABLE 2**  
Metabolite Traits Causally Associated With BMI

Trait	Metabolite	Class	Dietary state	$\beta$	SE	<i>P</i>
lafa_f	Ratio of linoleic acid (18:2) to total fatty acids (%)	Fatty acids ratios	Fasting	−0.106	0.025	$1.80 \times 10^{-5}$
faw6fa_f	Ratio of $\omega$ -6 fatty acids to total fatty acids (%)	Fatty acids ratios	Fasting	−0.081	0.024	$6.97 \times 10^{-4}$
pufafa_f	Ratio of polyunsaturated fatty acids to total fatty acids (%)	Fatty acids ratios	Fasting	−0.081	0.024	$6.80 \times 10^{-4}$
cit_f	Citrate (mmol/L)	Glycolysis-related metabolites	Fasting	−0.079	0.024	$8.63 \times 10^{-4}$
leu_f	Leucine (mmol/L)	Amino acids	Fasting	0.068	0.020	$5.62 \times 10^{-4}$
lhdlc_p	Total cholesterol in large HDL (mmol/L)	Lipoprotein subclasses	Postprandial	−0.076	0.023	$1.12 \times 10^{-3}$
lhdlfc_p	Free cholesterol in large HDL (mmol/L)	Lipoprotein subclasses	Postprandial	−0.077	0.023	$9.73 \times 10^{-4}$
mvldlplpct_p	Phospholipids in medium VLDL to total lipids in medium VLDL ratio (%)	Lipoprotein subclass ratios	Postprandial	−0.073	0.022	$9.98 \times 10^{-4}$
idlfcpct_p	Free cholesterol in IDL to total lipids in IDL ratio (%)	Lipoprotein subclass ratios	Postprandial	−0.079	0.023	$5.08 \times 10^{-4}$
lhdlcpcpct_p	Total cholesterol in large HDL to total lipids in large HDL ratio (%)	Lipoprotein subclass ratios	Postprandial	−0.085	0.025	$7.90 \times 10^{-4}$
lhdlcpcpct_p	Cholesterol esters in large HDL to total lipids in large HDL ratio (%)	Lipoprotein subclass ratios	Postprandial	−0.086	0.026	$1.03 \times 10^{-3}$
lhdlfcpcpct_p	Free cholesterol in large HDL to total lipids in large HDL ratio (%)	Lipoprotein subclass ratios	Postprandial	−0.095	0.025	$1.20 \times 10^{-4}$
lafa_p	Ratio of linoleic acid (18:2) to total fatty acids (%)	Fatty acids ratios	Postprandial	−0.092	0.025	$2.00 \times 10^{-4}$
faw6fa_p	Ratio of $\omega$ -6 fatty acids to total fatty acids (%)	Fatty acids ratios	Postprandial	−0.080	0.024	$7.83 \times 10^{-4}$
pufafa_p	Ratio of polyunsaturated fatty acids to total fatty acids (%)	Fatty acids ratios	Postprandial	−0.077	0.024	$1.10 \times 10^{-3}$
cit_p	Citrate (mmol/L)	Glycolysis-related metabolites	Postprandial	−0.105	0.022	$2.83 \times 10^{-6}$
his_p	Histidine (mmol/L)	Amino acids	Postprandial	−0.085	0.024	$3.96 \times 10^{-4}$
mvldlplpct_r	Phospholipids in medium VLDL to total lipids in medium VLDL ratio (%)	Lipoprotein subclass ratios	Response	−0.091	0.027	$6.75 \times 10^{-4}$
svldlfcpct_r	Free cholesterol in small VLDL to total lipids in small VLDL ratio (%)	Lipoprotein subclass ratios	Response	−0.082	0.025	$1.11 \times 10^{-3}$
ala_r	Alanine (mmol/L)	Amino acids	Response	−0.081	0.023	$5.91 \times 10^{-4}$

MR summary statistics for the 20 metabolite traits causally associated with BMI. Provided are the trait name, metabolite description and units, metabolite class as defined by Nightingale Heath, the dietary state, and the MR effect estimate, standard error, and *P* value. All summary statistics were derived from the wNEO 2-stage-least squares 1-sample MR framework, using the NEO data set with a sample size of 5744. However, each metabolite has a varying amount of missingness, and exact sample sizes for each trait can be found in [Supplemental Table 4](#). The effect estimates are in normalized standard deviation units of change per 1 unit increase of BMI ( $\text{kg/m}^2$ ).

Abbreviations: ala\_r, alanine response; cit\_f, fasting citrate; cit\_p, postprandial citrate; faw6fa\_f, fasting ratio of omega-6 fatty acids to total fatty acids; faw6fa\_p, postprandial omega-6 fatty acids to total fatty acids; HDL, high density lipoprotein; his\_p postprandial histidine; IDL, intermediate density lipoprotein; idlfcpct\_p, postprandial free cholesterol in IDL to total lipids in IDL; lafa\_f, fasting ratio of linoleic acid to total fatty acids; lafa\_p, postprandial ratio of linoleic acid to total fatty acids; leu\_f, fasting leucine; lhdlc\_p, postprandial total cholesterol in large HDL; lhdlfc\_p, postprandial free cholesterol in large HDL; lhdlcpcpct\_p, postprandial cholesterol esters in large HDL to total lipids in large HDL; lhdlcpcpct\_p, postprandial total cholesterol in large HDL to total lipids in large HDL; lhdlfcpcpct\_p, postprandial free cholesterol in large HDL to total lipids in large HDL; MR, Mendelian randomization; mvldlplpct\_p, postprandial phospholipids in medium VLDL to total lipids in medium VLDL; mvldlplpct\_r, response phospholipids in medium VLDL to total lipids in medium VLDL ratio; NEO, Netherlands Epidemiology of Obesity Study; pufafa\_f, fasting ratio of polyunsaturated fatty acids to total fatty acids; pufafa\_p, postprandial ratio of ployunsaturated fatty acids to total fatty acids; svldlfcpct\_r, response free cholesterol in small VLDL to total lipids in small VLDL ratio; VLDL, very low density lipoprotein.



**FIGURE 6.** Metabolites associated with BMI in Mendelian randomization (MR) analyses. A forest plot of effect estimates (points) and 95% CIs (whiskers) for MR (blue, “tsts” prefix) and observational (green, “obs” prefix) effect estimates for the weighted Netherlands Epidemiology of Obesity (wNEO), weighted NEO female (wNEO\_female), and weighted NEO male (wNEO\_male) (sub)populations. Fasting, postprandial, and response association observed in the general population and those association seen in only females and males. All results can be found in [Supplemental Tables 4, 10, and 11](#), respectively. Metabolite annotations and abbreviations can be found in [Supplementary Table 1](#). ala, alanine; cit, citrate; faw6fa, ratio of ω-6 fatty acids to total fatty acids; gln, Glutamine; his, histidine; idlfcpt, free cholesterol in IDL to total lipids in IDL ratio; lafa, ratio of 18:2 linoleic acid to total fatty acids; leu, Leucine; lhdic, total cholesterol in large HDL; lhdicept, cholesterol esters in large HDL to total lipids in large HDL ratio; lhdplpct, total cholesterol in large HDL to total lipids in large HDL ratio; lhdhlc, total cholesterol in large HDL to total lipids in large HDL ratio; mhdicept, cholesterol esters in medium HDL to total lipids in medium HDL ratio; mhdltgpcpt, triglycerides in medium HDL to total lipids in medium HDL ratio; mvldlplpct, phospholipids in medium VLDL to total lipids in medium VLDL ratio; pufafa, ratio of polyunsaturated fatty acids to total fatty acids; shdltg, triglycerides in small HDL; shdltgpcpt, triglycerides in small HDL to total lipids in small HDL ratio; svldlfcpt, free cholesterol in small VLDL to total lipids in small VLDL ratio; xsvldltg, triglycerides in very small VLDL.

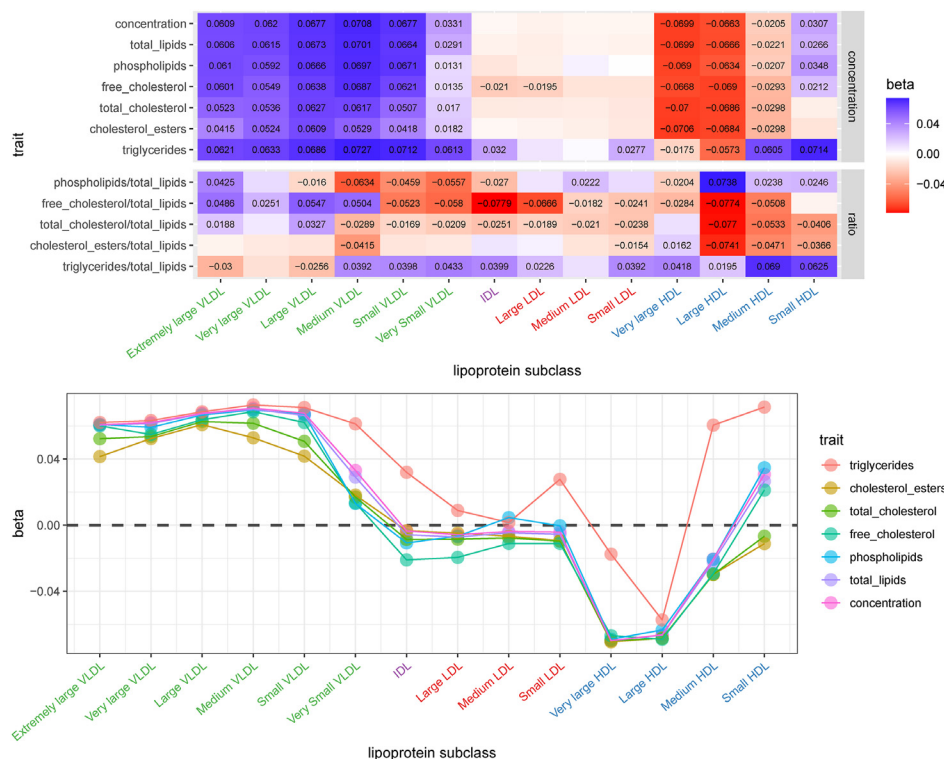
confounders or those that may violate the second MR assumption (exchangeability). In brief, 60 covariables were associated with BMI (wNEO;  $P < 9.26 \times 10^{-4}$ ), and 17 were associated with BMI-PGS ([Supplemental Figures 8 and 9](#), [Supplemental Table 15](#)). The 17 covariables that associated with the BMI-PGS were tested for an association with each metabolite trait (the outcomes), by univariable linear modeling, to identify possible confounders that violate MR assumption 2 ([Figure 1](#)). Results illustrated broad association across all traits and covariables, defining each covariable as a confounder in MR analysis leading to sensitivity analyses discussed further ([Supplemental Figure 10](#)). A more comprehensive description of these analyses is in [Supplemental Results](#).

### Estimating the causal effect of BMI on metabolite trait variation

MR analyses (wNEO) support broad associations between BMI and metabolites as suggested by the overall strong correlation between observational and MR estimates (Pearson  $r = 0.852$ ;  $P = 1.02 \times 10^{-194}$ ; intercept =  $1.54 \times 10^{-3}$ ; slope = 0.911). The correlation was the strongest for postprandial traits (Pearson  $r = 0.921$ ;  $P = 4.44 \times 10^{-95}$ ; intercept =  $3.59 \times 10^{-3}$ ; slope = 0.922) and weaker for fasting (Pearson  $r = 0.871$ ;  $P = 7.71 \times 10^{-72}$ ; intercept =  $7.62 \times 10^{-3}$ ; slope = 0.855) and response traits (Pearson  $r = 0.62$ ;  $P = 1.11 \times 10^{-25}$ ; intercept =  $-6.36 \times 10^{-3}$ ; slope = 1.08) ([Figure 3D](#)). Forty-two observational and MR effect estimates showed evidence of being different from each other

( $z$  test,  $P < 0.05$ ), yet none remain so after correcting for multiple tests ( $P < 0.05/687$ ). Overall, these observations indicate that cross-sectional estimates have reasonable power at predicting MR estimates for these exposure-outcome analyses, consistent with previous work [15].

A total of 201 metabolite traits (across fasting, postprandial, and response states) ([Supplemental Table 4](#), [Figure 5](#)) showed nominal evidence of association ( $P < 0.05$ ) with BMI and 20 ([Table 2](#), [Figure 6](#)) showed evidence of association with BMI ( $P < 1.16 \times 10^{-3}$ ) ([Supplemental Figure 11](#)). MR effect estimates varied more between sexes (Pearson  $r = 0.368$ ;  $P = 1.93 \times 10^{-23}$ ) than observed in observational results, with 86 showing evidence of being different from each other ( $z$  test,  $P < 0.05$ ), yet none remaining so after correcting for multiple tests ( $P < 0.05/687$ ). Nine metabolite traits (7 fasting and 2 response) showed evidence of association with BMI in females, 8 of which were unique to females ([Figure 6](#)). This included fasting small HDL triglycerides and fasting very small VLDL triglycerides, both of which increased with increases in BMI. By contrast, 5 metabolites were associated with BMI in males, one of which was unique to males—postprandial glutamine, which had an inverse relationship with BMI ([Figure 6](#)). The other 4 metabolites shared an association with the general population. They are fasting and postprandial citrate and the ratio of linoleic acid to total fatty acids in both the fasting and postprandial states. Moreover, each of them were also inversely associated with BMI. Mean MR effect estimates were larger in females (absolute mean  $\beta$ : 0.047) than those in males (absolute mean  $\beta$ : 0.026; paired  $t$  test,  $P = 1.43 \times 10^{-45}$ ).



**FIGURE 7.** Fasting lipoprotein cross-sectional profile. (Upper) A tile plot of cross-sectional effect estimates for lipoproteins and lipoprotein ratios in the fasting state. Tiles with an effect estimate provided in text are those associated with BMI ( $P < 1.16 \times 10^{-3}$ ). The lipoproteins are organized by size or density along the x-axis, and the component or ratio being measured is along the y-axis. (Lower) A dot plot or profile of cross-sectional effect estimates for lipoproteins (x-axis) ordered by lipoprotein size or density with effect estimates (y-axis). The component or measurement of each lipoprotein is defined by the color as described in the key. All results can be found in [Supplemental Table 4](#). HDL, high density lipoprotein; IDL, intermediate density lipoprotein; LDL, low density lipoprotein; VLDL, very low density lipoprotein.

**MR results—fasting**

A total of 87 fasting traits exhibited nominal evidence of an association with BMI in the general population, 5 remained after correction for multiple testing ([Table 2](#), [Supplemental Table 4](#), [Figure 6](#)). The strongest association observed was with the ratio of linoleic acid to total fatty acids (lafa;  $\beta$ :  $-0.106$ ; SE:  $0.025$ ;  $P = 1.80 \times 10^{-5}$ ). This was followed closely by 4 other inverse associations: the ratio polyunsaturated fatty acids (pufafa;  $\beta$ :  $-0.081$ ; SE:  $0.024$ ;  $P = 6.80 \times 10^{-4}$ ) to total fatty acids, the ratio  $\omega$ -6 to total fatty acids (faw6fa;  $\beta$ :  $-0.081$ ; SE:  $0.024$ ;  $P = 6.97 \times 10^{-4}$ ), and the abundance of citrate ( $\beta$ :  $-0.079$ ; SE:  $0.024$ ;  $P = 8.63 \times 10^{-4}$ ). In addition, 1 fasting metabolite showed evidence of a positive association with BMI, the BCAA leucine ( $\beta$ :  $0.068$ ; SE:  $0.020$ ;  $P = 5.62 \times 10^{-4}$ ). If data are summarized by metabolite annotation class ([Supplemental Table 1](#)), we find that the classes with the largest mean absolute BMI effect are ketone bodies (mean  $\beta$ :  $-0.047$ ,  $n = 2$ ), amino acids ( $0.036$ ,  $n = 8$ ), inflammation ( $0.029$ ,  $n = 1$ ), glycerides and phospholipids ( $0.027$ ,  $n = 9$ ), and lipoprotein subclasses ( $0.023$ ,  $n = 98$ ). Those with the smallest mean absolute effect are lipoprotein particle size ( $0.009$ ,  $n = 3$ ), lipoprotein subclass ratios ( $0.003$ ,  $n = 70$ ), and cholesterol ( $0.002$ ,  $n = 9$ ) ([Supplemental Figure 11](#)).

**MR results—postprandial**

A total of 96 postprandial metabolites exhibited nominal evidence of an association with BMI, 12 remained after correcting for multiple testing ([Table 2](#), [Supplemental Table 4](#), [Figure 6](#)). Like the fasting

metabolite traits, the fatty acid ratios linoleic acid (lafa;  $\beta$ :  $-0.092$ ; SE:  $0.025$ ;  $P = 2.0 \times 10^{-4}$ ),  $\omega$ -6 (faw6fa;  $\beta$ :  $-0.080$ ; SE:  $0.024$ ;  $P = 7.83 \times 10^{-4}$ ), polyunsaturated fatty acids (pufafa;  $\beta$ :  $-0.077$ ; SE:  $0.024$ ;  $P = 1.10 \times 10^{-3}$ ), and citrate ( $\beta$ :  $-0.105$ ; SE:  $0.022$ ;  $P = 2.83 \times 10^{-6}$ ), which had the largest effect, each also had an inverse relationship with BMI ([Figures 5 and 6](#)). In addition, the amino acid histidine, large HDL cholesterol, and free cholesterol decreased with increases in BMI. Furthermore, the ratio of cholesterol, cholesterol esters, and free cholesterol to total lipids in HDL decreased with increases in BMI. Finally, the ratio of free cholesterol to total lipids in IDL and the ratio of phospholipids to total lipids in medium VLDL decreased with increases in BMI. If data are summarized by metabolite annotation class ([Supplemental Table 1](#)), we find that the classes with the largest mean absolute BMI effect are glycolysis-related metabolites (mean  $\beta$ :  $-0.046$ ,  $n = 3$ ), inflammation ( $0.40$ ,  $n = 1$ ), and lipoprotein subclasses ( $0.02$ ,  $n = 98$ ). The mean effect among amino acids is not negative ( $-0.019$ ,  $n = 8$ ) in opposition to that seen in fasting data ([Supplemental Figure 11](#)).

**MR results—response**

A total of 18 response metabolites exhibited nominal evidence of an association with BMI, 3 remained after correcting for multiple testing ([Table 2](#), [Supplemental Table 4](#), [Figure 6](#)). The largest effect was observed for the amino acid alanine ( $\beta$ :  $-0.081$ ; SE:  $0.023$ ;  $P = 5.91 \times 10^{-4}$ ). Data suggested that increases in BMI attenuates the increase in abundance of the amino acid alanine after a liquid mixed meal

(Figures 5 and 6)—consistent with earlier observational analyses (Figure 4). The other 2 response traits that associated with BMI were the ratio of phospholipids in medium VLDL to total lipids in medium VLDL ( $\beta$ : 0.091; SE: 0.027;  $P = 6.75 \times 10^{-4}$ ) and the ratio of free cholesterol in small VLDL to total lipids in small VLDL ( $\beta$ : 0.082; SE: 0.025;  $P = 1.11 \times 10^{-3}$ ). As a class, amino acids (mean  $\beta$ :  $-0.059$ ,  $n = 8$ ) had the largest mean BMI response effect estimate (Supplemental Figure 11). This was followed by the ketone bodies acetate and  $\beta$ -hydroxybutyrate (mean  $\beta$ : 0.048,  $n = 2$ ) and the glycolysis-related metabolites citrate, glucose, and lactate (mean  $\beta$ :  $-0.042$ ,  $n = 3$ ).

### Subpopulation and sensitivity analyses

All observational and MR analyses were repeated in each of the 2 subpopulations and in the NEO cohort without the inclusion of weights. These analyses allowed us to evaluate the variability in effect estimates measured in a random (Leiderdorp) and biased (Leiden) population sample as and the effectiveness of the weights in our primary analyses (wNEO). We found that effect estimates from the Leiderdorp and the wNEO frameworks are strongly correlated (observational Pearson  $r = 0.983$ ; MR = 0.855) and correlation coefficients weakened when compared with the unweighted NEO and the Leiden frameworks (Supplemental Figure 12). Decreases in agreement between analyses as the sample population mean BMI shifts would indicate that there are either unaccounted confounders influencing the results or that the relationship between BMI and metabolite trait variation is not always linear.

To evaluate the influence of confounders on MR effect estimates, we reran the association analysis in the wNEO data set. Three additional covariates were included in the model (smoking, on a weight loss diet, and principal component 3) and 101 sample with a peroxidation sample quality flag, a sample quality metric identified by Metabolon, were removed. Overall primary (wNEO) observational (Pearson  $r = 0.998$ ) and MR (Pearson  $r = 0.963$ ) effect estimates correlated strongly with those from the sensitivity analysis (Supplemental Table 5). A more comprehensive description of these analyses is in Supplemental Results along with illustrations in Supplemental Figures 13–15.

### Comparison with previous work

We compared our observational effect estimates with 57 matching fasting metabolites measured using the same platform, in a young adult cohort [15]. Overall effect estimates correlated well (Pearson  $r = 0.82$ ) with the caveat that estimates cannot be directly compared given differences in data transformations between the studies. However, there are 37 shared associations, 6 unique to this study, 12 unique to the other, and 2 yielding no association in either study (Figure 3B). Five of the shared associations were different across studies (total cholesterol, sphingomyelins, phosphatidylcholines, phosphoglycerides, and linoleic acid). Among the 12 metabolites unique to existing work before that presented in this study, are 10 metabolites that previously exhibited a strong positive effect but showed no evidence for an association or negative point estimates (total lipids in IDL, large LDL, medium LDL and small LDL, total cholesterol in IDL and LDL, and 4 fatty acid metabolites  $\omega$ -3 and  $\omega$ -6, polyunsaturated fatty acids, and docosahexaenoic acid). These differences may be the product of nonlinear associations between these metabolites and BMI in combination with age effects and differences in the distribution of BMI between these studies. Despite these difference, 57 fasting MR estimates that did overlap with previous work showed strong correlation between the 2 studies (Pearson  $r = 0.862$ ;  $P = 7.99 \times 10^{-18}$ ) (Supplemental Figure 16), with 1 shared fasting MR association—the amino acid leucine.

### Atherogenic lipoprotein profile

Given the prevalence of lipoproteins and their lipids on this platform, we looked for the expected atherogenic lipoprotein profile—that is decreases in HDL and increases in non-HDL lipoproteins with increase in BMI [43–46]. Both observational and MR associations between BMI and fasting and postprandial dietary states do suggest a worsening atherogenic lipoprotein profile with BMI (Figure 7, Supplemental Figures 17 and 18). As illustrated with the fasting observational effect estimates in Figure 7, the atherogenic VLDL lipoproteins are higher with increases in BMI whereas nonatherogenic lipoproteins—or those involved in the reverse cholesterol transport system (HDL), were lower. In contrast to expectations, other atherogenic lipoproteins IDL and LDL, largely do not associate with BMI. In addition, triglycerides in nonatherogenic medium HDL and small HDL increase with increases in BMI. Furthermore, as HDL lipoprotein density increases, the inverse relationship with BMI weakens, and as seen with small HDL, the association becomes positive (Figure 7). This is consistent with the aforementioned observation that the strongest inverse association with BMI is that with mean diameter of HDL particles, suggesting that as BMI increases there is an associated shift toward smaller HDL particle size. Along with these changes, other observations were consistent with obesity influencing lipid profiles that are associated with cardiometabolic disease [14,22,47]. These included increases in apolipoprotein B, decreases in apolipoprotein A-1, and increases in the ratio of B to A-1 with increases in BMI, in both the fasting and postprandial states [45,48].

### Discussion

This study provided effect estimates for observational and MR associations between BMI and metabolites in a 1) fasted, 2) postprandial, and 3) response state, using individual-level data from a middle-aged cohort of 5517 individuals of Northern European ancestry including a liquid mixed meal challenge. Each dietary state provides a unique examination at metabolite variation, the value of which are actively being explored. Broad observational associations were observed between BMI and metabolites (69% of tested traits). In addition, MR estimates were largely concordant with the observational estimates (Figure 3D). This is consistent with previous research and suggests that, at least for this lipidomic platform, observational estimates are reasonably well aligned to MR effect estimates [15]. This generalization holds less true for response traits than for postprandial and fasting trait (Figure 3D). After correcting for multiple testing, 20 metabolite traits maintain evidence of a causal BMI effect (Figure 6). This includes the inverse association with fasting and postprandial ratios of linoleic acid (lafa),  $\omega$ -6 (faw6fa), and polyunsaturated fatty acids (pufafa) to total fatty acids and fasting and postprandial citrate abundance. In addition, the amino acid leucine has a positive MR association with BMI in the fasting state, histidine has an inverse MR association with BMI in the postprandial state, and alanine has an inverse MR association with BMI in the response state.

The MR association of BMI across BCAAs and aromatic amino acids (AAAs) were similar with positive effects in the fasting state, no effect in the postprandial state, and an inverse association in the response state (Figures 2 and 5). Meanwhile, the dietary profile of (glucogenic) amino acids suggests no effect in the fasting state and inverse effects in both the postprandial and response states. These observations may suggest that BMI may have a causal influence on the synthesis or metabolism of amino acids in general, particularly

regarding their inverse association with BMI in the response state (Figures 2 and 5). BCAAs and AAAs have long been associated with obesity, glucose, insulin (resistance), and type 2 diabetes [49–55]. In support of early observations, numerous recent studies that have shown associations between amino acids and BMI, visceral adipose tissue and weight change [15,56–59], between BCAAs and AAAs with insulin resistance and type 2 diabetes [60–63], used MR to show that insulin resistance increase BCAAs concentration [64] and used a randomized control trial to show that restricting BCAAs can improve glucose tolerance and reduce fat accumulation [65]. Indeed evidence now suggests a potential causal pathway from BMI to type 2 diabetes through the intermediate traits of insulin resistance and BCAAs, respectively [64]. Observations in this study reinforce the causal association between BMI and amino acids, with the additional observations that postprandial abundance of the glucogenic amino acids decreases with BMI and that the relative change in amino acid abundance in response to a liquid mixed meal decreases with increases in BMI.

As the study was centered in the Netherlands, study samples were limited to individuals of Northern European ancestry, and as such, inferences made from results should be limited to populations of similar ancestry and environment. Sampling was limited to individuals of middle age (range 44 to 66 y and a mean of 56 y) (Table 1). Although this provides specificity to a middle age population it also limits the results to this population age range as well. The NEO population sampling was performed in 2 distinct batches, one focused on oversampling individuals of high BMI (Leiden) and the second, a random, and representative sampling of the population (Leiderdorp). The product of this led to not only the use of weighted linear regression analyses to maintain sample size and power but also providing estimates that are representative of the study population at large. The validity of this assumption rests in the credibility of the weights, which were specifically designed to make the BMI distribution of the Leiden sample mirror that of the Leiderdorp sample [42]. We have illustrated through comparison with the Leiderdorp sample that effect estimates are consistent across the 2 analyses, imparting support to the weighted analysis (Supplemental Figure 12). With these factors, there was just a single postprandial time point evaluated in this study limiting the inferences that can be drawn. It is reasonable to speculate that an earlier time point or a complete time series would provide additional information about how BMI might influence postprandial variation. Finally, although a single complex meal standardizes the complexities in evaluating metabolite response and postprandial abundances alternative meals could provide novel insights in the interplay among BMI, nutrition, and metabolite variation.

In conclusion, using a middle-aged cohort of 5517 individuals, we derived effect estimates for BMI on metabolite traits in the fasting, postprandial, and response dietary states. We were able to show that results are broadly correlated between observational and one-sample MR analyses and in the context of the numerous documented associations among BMI, metabolites, and disease, this gives support to a conclusion that metabolites may act as intermediates—or are at least biomarkers of underlying driver physiology—between adiposity accumulation and disease. Furthermore, this work suggested that the dynamic metabolome may potentially flag common biological events, which will systematically vary by BMI, which may be linked to the etiology of disease, and which are linked to life course events such as feeding and metabolic response.

## Acknowledgments

We thank Pat van Beelen for data collection, Petra Noordijk for laboratory management, and Ingeborg de Jonge for data management of the NEO study. We thank Nutricia Research, Utrecht, The Netherlands, for providing the liquid mixed meal.

## Author contributions

The authors' responsibilities were as follows – DAH, RL-G, CJB, DOM-K, KWvD, NJT: designed the study; FRR, RdM: designed and conducted the NEO study and managed the data; DAH, NJT: analyzed the data; DAH, RL-G, KWvD, NJT: wrote the manuscript; DAH, NJT: had primary responsibility for final content; and all authors: read and approved the final manuscript.

## Conflict of interest

Ruifang Li-Gao is a part-time contractor of Metabolon Inc.

## Funding

This research was supported by the Netherlands Cardiovascular Research Initiative: an initiative with support from the Dutch Heart Foundation (CVON2014-02 ENERGISE). NJT is a Wellcome Trust Investigator (202802/Z/16/Z); is the PI of the Avon Longitudinal Study of Parents and Children (MRC & WT 102215/2/13/2) is supported by the University of Bristol NIHR Biomedical Research Centre (BRC-1215-20011), the MRC Integrative Epidemiology Unit (MC\_UU\_12013/3); and works within the CRUK Integrative Cancer Epidemiology Programme (C18281/A19169). DAH was supported by NJT's Wellcome Investigator Award (202802/Z/16/Z) and along with CJB was also supported by the University of Bristol and UK Medical Research Council (MC\_UU\_00011/1 and MC\_UU\_00011/6). RL-G was supported by JPI HDHL-DIYUFOOD project. DOM-K is supported by the Dutch Science Organization (ZonMW-VENI Grant 916.14.023).

## Data availability

Data described in the manuscript are available to researchers according to the NEO research procedure. Contact the scientific director of the NEO study for access. All analytic codes are publicly and freely available without restriction at the github repository ([https://github.com/hughesevoanth/NEO\\_BMI\\_Metabolite\\_MR](https://github.com/hughesevoanth/NEO_BMI_Metabolite_MR)).

## Appendix A. Supplementary data

Supplementary data to this article can be found online at <https://doi.org/10.1016/j.ajcnut.2024.03.009>.

## References

- [1] Clinical guidelines on the identification, evaluation, and treatment of overweight and obesity in adults—the evidence report, National Institutes of Health, *Obes. Res.* 6 (Suppl 2) (1998) 51S–209S.
- [2] M.D. Jensen, D.H. Ryan, C.M. Apovian, J.D. Ard, A.G. Comuzzie, K.A. Donato, et al., 2013 AHA/ACC/TOS guideline for the management of overweight and obesity in adults: a report of the American College of Cardiology/American Heart Association Task Force on Practice Guidelines and The Obesity Society, *J. Am. Coll. Cardiol.* 63 (25 Pt B) (2014) 2985–3023.
- [3] D.O. Okorodudu, M.F. Jumean, V.M. Montori, A. Romero-Corral, V.K. Somers, P.J. Erwin, et al., Diagnostic performance of body mass index to identify obesity as defined by body adiposity: a systematic review and meta-analysis, *Int. J. Obes (Lond).* 34 (5) (2010) 791–799.

- [4] J.J. Reilly, J. Kelly, D.C. Wilson, Accuracy of simple clinical and epidemiological definitions of childhood obesity: systematic review and evidence appraisal, *Obes. Rev.* 11 (9) (2010) 645–655.
- [5] A. Javed, M. Jumean, M.H. Murad, D. Okorodudu, S. Kumar, V.K. Somers, et al., Diagnostic performance of body mass index to identify obesity as defined by body adiposity in children and adolescents: a systematic review and meta-analysis, *Pediatr. Obes.* 10 (3) (2015) 234–244.
- [6] J.J. Reilly, A. El-Hamdouchi, A. Diouf, A. Monyeke, S.A. Somda, Determining the worldwide prevalence of obesity, *Lancet.* 391 (10132) (2018) 1773–1774.
- [7] K.H. Wade, D. Carslake, P. Tynelius, G.D. Smith, R.M. Martin, Variation of all-cause and cause-specific mortality with body mass index in one million Swedish parent-son pairs: an instrumental variable analysis, *PLOS Med* 16 (8) (2019) 1–15.
- [8] D.A. Jenkins, K.H. Wade, D. Carslake, J. Bowden, N. Sattar, R.J.F. Loos, et al., Estimating the causal effect of BMI on mortality risk in people with heart disease, diabetes and cancer using Mendelian randomization, *Int. J. Cardiol.* 330 (2021) 214–220.
- [9] C.G. Solomon, J.E. Manson, Obesity and mortality: a review of the epidemiologic data, *Am. J. Clin. Nutr.* 66 (4 Suppl) (1997) 1044S–1050S.
- [10] J.P. Reis, C.A. Macera, M.R. Araneta, S.P. Lindsay, S.J. Marshall, D.L. Wingard, Comparison of overall obesity and body fat distribution in predicting risk of mortality, *Obesity, Silver Spring* 17 (6) (2009) 1232–1239.
- [11] J. Pearson-Stuttard, B. Zhou, V. Kontis, J. Bentham, M.J. Gunter, M. Ezzi, Worldwide burden of cancer attributable to diabetes and high body-mass index: a comparative risk assessment, *Lancet Diabetes Endocrinol* 6 (6) (2018) e6–e15.
- [12] B. Lauby-Secretan, C. Scoccianti, D. Loomis, Y. Grosse, F. Bianchini, K. Straif, Body fatness and cancer—viewpoint of the IARC Working Group, *N. Engl. J. Med.* 375 (8) (2016) 794–798.
- [13] H. Sung, R.L. Siegel, L.A. Torre, J. Pearson-Stuttard, F. Islami, S.A. Fedewa, et al., Global patterns in excess body weight and the associated cancer burden, *CA Cancer J. Clin.* 69 (2018) 88–112.
- [14] T.M. Powell-Wiley, P. Poirier, L.E. Burke, J.-P.P. Després, P. Gordon-Larsen, C.J. Lavie, et al., Obesity and cardiovascular disease: a scientific statement from the American Heart Association, *Circulation* 143 (21) (2021) E984–E1010.
- [15] P. Würtz, Q. Wang, A.J. Kangas, R.C. Richmond, J. Skarp, M. Tiainen, et al., Metabolic signatures of adiposity in young adults: Mendelian randomization analysis and effects of weight change, *PLoS Med* 11 (12) (2014) e1001765.
- [16] P. Würtz, A.S. Havulinna, P. Soininen, T. Tynkynen, D. Prieto-Merino, T. Tillin, et al., Metabolite profiling and cardiovascular event risk: a prospective study of 3 population-based cohorts, *Physiol. Behav.* 176 (1) (2016) 139–148.
- [17] R. von Krüchten, R. Lorbeer, K. Müller-Peltzer, S. Rospleszcz, C. Storz, E. Askani, et al., Association between adipose tissue depots and dyslipidemia: the KORA-MRI population-based study, *Nutrients* 14 (4) (2022) 797.
- [18] B.A. Ference, W. Yoo, I. Alesh, N. Mahajan, K.K. Mirowska, A. Mewada, et al., Effect of long-term exposure to lower low-density lipoprotein cholesterol beginning early in life on the risk of coronary heart disease: a Mendelian randomization analysis, *Ration Pharmacother Cardiol* 9 (1) (2013) 90–98.
- [19] C. Krittanawong, M. Khawaja, R.S. Rosenson, C.I. Amos, V. Nambi, C.J. Lavie, et al., Association of PCSK9 variants with the risk of atherosclerotic cardiovascular disease and variable responses to PCSK9 inhibitor therapy, *Curr. Probl. Cardiol.* 47 (2022) 101043.
- [20] T.J. Wilt, H.E. Bloomfield, R. MacDonald, D. Nelson, I. Rutks, M. Ho, et al., Effectiveness of statin therapy in adults with coronary heart disease, *Arch. Intern. Med.* 164 (13) (2004) 1427–1436.
- [21] J.A. Berliner, M. Navab, A.M. Fogelman, J.S. Frank, L.L. Demer, P.A. Edwards, et al., Atherosclerosis: basic mechanisms. Oxidation, inflammation, and genetics, *Circulation* 91 (9) (1995) 2488–2496.
- [22] T.G. Richardson, E. Sanderson, T.M. Palmer, M.A. Korpelaid, B.A. Ference, G.D. Smith, et al., Evaluating the relationship between circulating lipoprotein lipids and apolipoproteins with risk of coronary heart disease: a multivariable Mendelian randomisation analysis, *PLoS Med* 17 (3) (2020) 1–22.
- [23] M. Ala-Korpela, The culprit is the carrier, not the loads: cholesterol, triglycerides and apolipoprotein B in atherosclerosis and coronary heart disease, *Int. J. Epidemiol.* 48 (5) (2019) 1389–1392.
- [24] L.J. Corbin, V.Y. Tan, D.A. Hughes, K.H. Wade, D.S. Paul, K.E. Tansey, et al., Formalising recall by genotype as an efficient approach to detailed phenotyping and causal inference, *Nat. Commun.* 9 (1) (2018) 711.
- [25] S. Fang, K.H. Wade, D.A. Hughes, S. Fitzgibbon, V. Yip, N.J. Timpson, et al., A multivariate recall-by-genotype study of the metabolomic signature of BMI, *Obesity* 30 (6) (2022) 1298–1310.
- [26] L.J. Goudswaard, J.A. Bell, D.A. Hughes, L.J. Corbin, K. Walter, G. Davey Smith, et al., Effects of adiposity on the human plasma proteome: observational and Mendelian randomisation estimates, *Int. J. Obes.* 45 (10) (2021) 2221–2229.
- [27] C. Bull, E. Hazelwood, J.A. Bell, V. Tan, A.E. Constantinescu, C. Borges, et al., Identifying metabolic features of colorectal cancer liability using Mendelian randomization, *Elife* 12 (2023) RP87894.
- [28] C.J. Bull, J.A. Bell, N. Murphy, E. Sanderson, G. Davey Smith, N.J. Timpson, et al., Adiposity, metabolites, and colorectal cancer risk: Mendelian randomization study, *BMC Med* 18 (1) (2020) 396. Available from: <https://pubmed.ncbi.nlm.nih.gov/33327948/>.
- [29] R. Li-Gao, D.A. Hughes, J.B. van Klinken, R. de Mutsert, F.R. Rosendaal, D.O. Mook-Kanamori, et al., Genetic studies of metabolomics change after a liquid meal illuminate novel pathways for glucose and lipid metabolism, *Diabetes* 70 (12) (2021) 2932–2946.
- [30] K.H. Wade, D. Lawlor, M.C. Borges, T. Palmer, F.P. Hartwig, G. Hemani, et al., [Internet], MR Dictionary (2021) [cited 2022 Mar 1]. Available from: <https://mr-dictionary.mrcieu.ac.uk>.
- [31] G. Davey Smith, G. Hemani, Mendelian randomization: genetic anchors for causal inference in epidemiological studies, *Hum. Mol. Genet.* 23 (R1) (2014) R89–R98.
- [32] R. de Mutsert, M. den Heijer, T.J. Rabelink, J.W. Smit, J.A. Romijn, J.W. Jukema, et al., The Netherlands Epidemiology of Obesity (NEO) study: study design and data collection, *Eur. J. Epidemiol.* 28 (6) (2013) 513–523.
- [33] P. Soininen, A.J. Kangas, P. Würtz, T. Suna, M. Ala-Korpela, Quantitative serum nuclear magnetic resonance metabolomics in cardiovascular epidemiology and genetics, *Circ. Cardiovasc. Genet.* 8 (1) (2015) 192–206.
- [34] D.A. Hughes, K. Taylor, N. McBride, M.A. Lee, D. Mason, D.A. Lawlor, et al., Metaboprep: an R package for preanalysis data description and processing, *Bioinformatics* 38 (7) (2022) 1980–1987.
- [35] J.M. Cheverud, A simple correction for multiple comparisons in interval mapping genome scans, *Heredity (Edinb.)* 87 (1) (2001) 52–58.
- [36] J. Li, L. Ji, Adjusting multiple testing in multilocus analyses using the eigenvalues of a correlation matrix, *Heredity (Edinb.)* 95 (3) (2005) 221–227.
- [37] X. Gao, J. Starmer, E.R. Martin, A multiple testing correction method for genetic association studies using correlated single nucleotide polymorphisms, *Genet. Epidemiol.* 32 (4) (2008) 361–369.
- [38] L.L. Blauw, R. Li-Gao, R. Noordam, R. de Mutsert, S. Trompet, J.F.P. Berbee, et al., CETP (cholesteryl ester transfer protein) concentration: a genome-wide association study followed by Mendelian randomization on coronary artery disease, *Circ. Genom. Precis. Med.* 11 (5) (2018) e002034.
- [39] S. McCarthy, S. Das, W. Kretzschmar, O. Delaneau, A.R. Wood, et al., A reference panel of 64,976 haplotypes for genotype imputation, *Nat. Genet.* 48 (10) (2016) 1279–1283.
- [40] L. Yengo, J. Sidorenko, K.E. Kemper, Z. Zheng, A.R. Wood, M.N. Weedon, et al., Meta-analysis of genome-wide association studies for height and body mass index in ~700,000 individuals of European ancestry, *Hum. Mol. Genet.* 27 (20) (2018) 3641–3649.
- [41] A. Zeileis, T. Hothorn, Diagnostic checking in regression relationships, *R News [Internet]* 2 (3) (2002) 7–10. Available from: <https://cran.r-project.org/doc/Rnews/>.
- [42] I.A. Dekkers, R. de Mutsert, A.P.J. de Vries, F.R. Rosendaal, S.C. Cannegieter, J.W. Jukema, et al., Determinants of impaired renal and vascular function are associated with elevated levels of procoagulant factors in the general population, *J. Thromb. Haemost.* 16 (3) (2018) 519–528.
- [43] B.G. Nordestgaard, A test in context: lipid profile, fasting versus nonfasting, *J. Am. Coll. Cardiol.* 70 (2017) 1637–1646.
- [44] B.J. Arsenaault, J.S. Rana, E.S.G. Stroes, J.-P. Després, P.K. Shah, J.J.P. Kastelein, et al., Beyond low-density lipoprotein cholesterol: respective contributions of non-high-density lipoprotein cholesterol levels, triglycerides, and the total cholesterol/high-density lipoprotein cholesterol ratio to coronary heart disease risk in apparently healthy men and women, *J. Am. Coll. Cardiol.* 55 (1) (2009) 35–41.
- [45] S.S. Carr, A.J. Hooper, D.R. Sullivan, J.R. Burnett, Non-HDL-cholesterol and apolipoprotein B compared with LDL-cholesterol in atherosclerotic cardiovascular disease risk assessment, *Pathology* 51 (2) (2019) 148–154.
- [46] A.D. Sniderman, S. Islam, M. McQueen, M. Pencina, C.D. Furberg, G. Thanassoulis, et al., Age and cardiovascular risk attributable to apolipoprotein B, low-density lipoprotein cholesterol or non-high-density lipoprotein cholesterol, *J. Am. Heart. Assoc.* 5 (10) (2016) e003665.
- [47] M.V. Holmes, L.A. Lange, T. Palmer, M.B. Lanktree, K.E. North, B. Almqvister, et al., Causal effects of body mass index on cardiometabolic traits and events: a Mendelian randomization analysis, *Am. J. Hum. Genet.* 94 (2) (2014) 198–208.
- [48] B.J. Arsenaault, J.-P. Després, E.S.G. Stroes, N.J. Wareham, J.J.P. Kastelein, K.-T. Khaw, et al., Lipid assessment, metabolic syndrome and coronary heart disease risk, *Eur. J. Clin. Invest.* 40 (12) (2010) 1081–1093.
- [49] P. Felig, E. Marliss, G.F. Cahill, Plasma amino acid levels and insulin secretion in obesity, *N. Engl. J. Med.* 281 (15) (1969) 811–816.
- [50] S.A. Adibi, Influence of dietary deprivations on plasma concentration of free amino acids of man, *J. Appl. Physiol.* 25 (1) (1968) 52–57.

- [51] J.M. Luck, S.W. Morse, The effects of insulin and adrenaline on the amino-acid content of blood, *Biochem. J.* 27 (5) (1933) 1648–1654.
- [52] J.C.J. Floyd, S.S. Fajans, J.W. Conn, R.F. Knopf, J. Rull, Stimulation of insulin secretion by amino acids, *J. Clin. Invest.* 45 (9) (1966) 1487–1502.
- [53] J.H. Ivy, M. Svec, S. Freeman, Free plasma levels and urinary excretion of eighteen amino acids in normal and diabetic dogs, *Am. J. Physiol.* 167 (1) (1951) 182–192.
- [54] A. Carlsten, B. Hallgren, R. Jagenburg, A. Svanborg, L. Werkö, Amino acids and free fatty acids in plasma in diabetes. I. The effect of insulin on the arterial levels, *Acta Med. Scand.* 179 (3) (1966) 361–370.
- [55] P.J. White, C.B. Newgard, Branched-chain amino acids in disease, *Science*. 363 (6427) (2019) 582–583.
- [56] I.J. Neeland, S.C. Boone, D.O. Mook-Kanamori, C. Ayers, R.A.J. Smit, I. Tzoulaki, et al., Metabolomics profiling of visceral adipose tissue: results from MESA and the NEO study, *J. Am. Heart Assoc.* 8 (9) (2019) e010810.
- [57] N. Geidenstam, M. Al-Majdoub, M. Ekman, P. Spéjel, M. Ridderstråle, Metabolite profiling of obese individuals before and after a one year weight loss program, *Int. J. Obes.* 41 (9) (2017) 1369–1378.
- [58] B.A.M. Schutte, E.B. Van Den Akker, J. Deelen, O. Van De Rest, D. Van Heemst, E.J.M. Feskens, et al., The effect of standardized food intake on the association between BMI and <sup>1</sup>H-NMR metabolites, *Sci. Rep.* 6 (2016) 1–6.
- [59] N. Geidenstam, A.P.H. Danielsson, P. Spéjel, M. Ridderstråle, Changes in glucose-elicited blood metabolite responses following weight loss and long term weight maintenance in obese individuals with impaired glucose tolerance, *Diabetes Res. Clin. Pract.* 113 (2016) 187–197.
- [60] L. Weng, E. Quinlivan, Y. Gong, A.L. Beitelshes, M.H. Shahin, S.T. Turner, et al., Association of branched and aromatic amino acids levels with metabolic syndrome and impaired fasting glucose in hypertensive patients, *Metab. Syndr. Relat. Disord.* 13 (5) (2015) 195–202.
- [61] T.J. Wang, M.G. Larson, R.S. Vasan, S. Cheng, E.P. Rhee, E. McCabe, et al., Metabolite profiles and the risk of developing diabetes, *Nat. Med.* 17 (4) (2011) 448–453.
- [62] E. Porcu, F. Gilardi, L. Darrous, L. Yengo, N. Bararpour, M. Gasser, et al., Triangulating evidence from longitudinal and Mendelian randomization studies of metabolomic biomarkers for type 2 diabetes, *Sci. Rep.* 11 (1) (2021) 6197.
- [63] L.A. Lotta, R.A. Scott, S.J. Sharp, S. Burgess, J. Luan, T. Tillin, et al., Genetic predisposition to an impaired metabolism of the branched-chain amino acids and risk of type 2 diabetes: a Mendelian randomisation analysis, *PLOS Med* 13 (11) (2016) e1002179.
- [64] Q. Wang, M.V. Holmes, G. Davey Smith, M. Ala-Korpela, Genetic support for a causal role of insulin resistance on circulating branched-chain amino acids and inflammation, *Diabetes Care* 40 (12) (2017) 1779–1786.
- [65] L. Fontana, N.E. Cummings, S.I. Arriola Apelo, J.C. Neuman, I. Kasza, B.A. Schmidt, et al., Decreased consumption of branched-chain amino acids improves metabolic health, *Cell. Rep* 16 (2) (2016) 520–530.