UC Irvine
UC Irvine Previously Published Works

Title

A two-sample Mendelian randomization study explores metabolic profiling of different glycemic traits.

Permalink

https://escholarship.org/uc/item/45w1g1v6

Journal

Communications Biology, 7(1)

Authors

Wong, Tommy

Mo, Jacky

Zhou, Mingqi

et al.

Publication Date

2024-03-08

DOI

10.1038/s42003-024-05977-1

Article

A two-sample Mendelian randomization study explores metabolic profiling of different glycemic traits

Check for updates

Tommy H. T. Wong [1,5], Jacky M. Y. Mo[1,5], Mingqi Zhou [2,3], Jie V. Zhao [1], C. Mary Schooling [1,4], Baoting He[1], Shan Luo [1] ✉ & Shiu Lun Au Yeung[1]

We assessed the causal relation of four glycemic traits and type 2 diabetes liability with 167 metabolites using Mendelian randomization with various sensitivity analyses and a reverse Mendelian randomization analysis. We extracted instruments for fasting glucose, 2-h glucose, fasting insulin, and glycated hemoglobin from the Meta-Analyses of Glucose and Insulin-related traits Consortium ($n = 200,622$), and those for type 2 diabetes liability from a meta-analysis of multiple cohorts (148,726 cases, 965,732 controls) in Europeans. Outcome data were from summary statistics of 167 metabolites from the UK Biobank ($n = 115,078$). Fasting glucose and 2-h glucose were not associated with any metabolite. Higher glycated hemoglobin was associated with higher free cholesterol in small low-density lipoprotein. Type 2 diabetes liability and fasting insulin were inversely associated with apolipoprotein A1, total cholines, lipoprotein subfractions in high-density-lipoprotein and intermediate-density lipoproteins, and positively associated with aromatic amino acids. These findings indicate hyperglycemia-independent patterns and highlight the role of insulin in type 2 diabetes development. Further studies should evaluate these glycemic traits in type 2 diabetes diagnosis and clinical management.

Hyperglycemia, increased insulin, and type 2 diabetes likely increase the risk of cardiovascular disease[1]. A recent study suggested there were different roles of four glycemic traits in atherosclerotic and thrombotic conditions[2]. Based on the American Diabetes Association "Standards of Medical Care in Diabetes", type 2 diabetes may be diagnosed based on different glycemic traits. Diagnosis can be based on the glucose criteria (including fasting glucose, 2-h glucose after a 75-g oral glucose tolerance test) or the A1c criteria (glycated hemoglobin (HbA_{1c}))[3]. However, these could lead to heterogeneity in the profiling of type 2 diabetes given these glycemic traits do not always underlie the same pathophysiology relevant to glucose homeostasis[4]. Clarifying the similarities and differences in the association with downstream metabolites (e.g., lipid subfractions and amino acids) of these glycemic traits and liability to type 2 diabetes may help clarify the pathophysiology of these inter-related glycemic traits in various diseases, such as cardiovascular diseases shown in a previous Mendelian randomization study[2].

Previous observational studies have suggested glycemic traits are associated with particular metabolites. Studies of Finnish adults ($n = 7098$) found that branched-chain amino acids (BCAAs) and ketone body levels were associated with lower insulin sensitivity and higher risk of type 2 diabetes[5,6]. Another study of Finnish men ($n = 9399$) found that higher concentrations of very low-density lipoproteins (VLDL) subclass particles were associated with glucose intolerance and newly diagnosed type 2 diabetes[7]. A small observational study ($n = 733$) showed HbA_{1c} positively associated with some BCAAs, such as isoleucine and alanine, and lower apolipoprotein A1 (ApoA1)[8], whilst another small study ($n = 155$) suggested possible differences in amino acid signatures for prediabetes defined by different glycemic traits[9]. Nonetheless, these associations may be susceptible to residual confounding by obesity and physical inactivity. A Mendelian randomization study, which makes use of genetic endowment randomly allocated at conception, can overcome these limitations[10]. We used a Mendelian randomization study to infer the role of each glycemic

[1]School of Public Health, Li Ka Shing Faculty of Medicine, The University of Hong Kong, Hong Kong SAR, China. [2]Department of Biological Chemistry, School of Medicine, University of California Irvine, Irvine, CA, USA. [3]Center for Epigenetics and Metabolism, University of California Irvine, Irvine, CA, USA. [4]School of Public Health and Health Policy, City University of New York, New York, NY, USA. [5]These authors contributed equally: Tommy H. T. Wong, Jacky M. Y. Mo. ✉e-mail: luoshan@connect.hku.hk

trait (fasting glucose, 2-h glucose, HbA$_{1c}$ and fasting insulin) and liability to type 2 diabetes in metabolomic signatures with various sensitivity analyses and a reverse Mendelian randomization analysis.

Methods

This is a two-sample Mendelian randomization study using summary statistics from genome-wide association studies (GWAS), which relies on the three instrumental variable assumptions[11]. First, the instruments should be strongly associated with the exposure of interest. Second, there should be no unmeasured confounding of instruments on outcome. Third, the instruments should be independent of the outcome given the exposure and the confounders.

Genetic predictors of glycemic traits

We selected genetic instruments that were genome-wide significant ($p < 5 \times 10^{-8}$) and uncorrelated ($r^2 < 0.001$) for fasting glucose (mmol/L), 2-h glucose (mmol/L), fasting insulin (natural log transformed pmol/L), and HbA$_{1c}$ (%) from the Meta-Analyses of Glucose and Insulin-related traits Consortium (MAGIC), using data only from participants of European ancestry ($n = 200,622$)[12]. In this GWAS, participants with either type 1 or type 2 diabetes, who reported taking diabetes-related medications, had a fasting glucose ≥7 mmol/L, 2-h glucose ≥11.1 mmol/L, or HbA$_{1c}$ ≥ 6.5% were excluded[12]. Genetic associations of glycemic traits were obtained using multivariable linear regression adjusted for age, sex, body-mass index (BMI) (except for HbA$_{1c}$), study-specific covariates, and genomic control. The impact of collider bias due to BMI adjustment was minimal according to the original GWAS[12,13].

Genetic predictors of type 2 diabetes

We selected genetic instruments that were genome-wide significant ($p < 5 \times 10^{-8}$) and uncorrelated ($r^2 < 0.001$) related to liability to type 2 diabetes from the largest GWAS to-date, using data only from participants of European ancestry (148,726 cases, 965,732 controls)[14]. Type 2 diabetes cases were ascertained using study-specific criteria, including diagnosis codes, hospital admission records, biochemical results (fasting glucose ≥ 7.0 mmol/L, or 2-h glucose ≥11.1 mmol/L, or HbA$_{1c}$ ≥ 6.5%), and use of diabetes-related medications[14]. The genetic associations with type 2 diabetes were obtained using multivariable logistic regression adjusted for age, sex, and the top 10 principal components for genetic ancestry. Given type 2 diabetes is a binary variable, the interpretation of the corresponding Mendelian randomization study estimates using these instruments should be in terms of liability to type 2 diabetes[15], consistent with earlier Mendelian randomization studies using diseases as the exposure[16–18].

Genetic associations with metabolomic markers

We obtained genetic associations with circulating metabolomic markers ($n = 115,078$) from UK Biobank summary statistics, accessed via MR-Base[19]. In brief, a range of circulating metabolomic markers were quantified in non-fasting EDTA-plasma samples collected from a random subset of the UK Biobank participants using a high-throughput nuclear magnetic resonance (NMR) spectroscopy platform from Nightingale Health, the technical details of which have been published[20]. We considered all 167 metabolomic biomarkers measured as outcomes, including amino acids, lipids, apolipoproteins, and lipoprotein subclass distribution and excluding glucose (Supplementary Data 1). Measurements of all metabolomic markers were inverse rank-normalized. Genetic associations with each biomarker (in standard deviations) were obtained using a linear mixed model with a random effect accounting for potential confounding due to population stratification and genetic relatedness, with adjustment for age, sex, fasting status, and genotyping array[21,22]. Whenever genetic instruments for glycemic traits and liability to type 2 diabetes were not available, proxy instruments (pairwise $r^2 \geq 0.8$) were identified based on the 1000 Genomes Phase 3 dataset (version 5, CEU reference population)[23].

Statistical analyses

We aligned the effect estimates for both exposure and outcome studies so that they corresponded to the same effect allele. Given that palindromic instruments (G > C and A > T) have the same allele notation on both the forward and reverse strand, we additionally used effect allele frequency (EAF) to ensure the alleles in both studies referred to the same strand direction. However, variants with ambiguous EAF (0.42 < EAF < 0.58) were excluded from the analyses. For each genetic instrument, we calculated the instrument strength for the Cragg-Donald F statistic using $F = \frac{N-K-1}{K}\frac{R^2}{1-R^2}$, with an F statistic >10 indicating weak instrument bias is unlikely[24,25]. We also calculated the variance of each glycemic trait explained by the selected instruments (R^2) based on F statistics, effect estimates, standard errors of the instruments, and the corresponding GWAS sample size[26,27]. We assessed the association of each glycemic trait and liability to type 2 diabetes with each circulating metabolite using inverse variance weighted with multiplicative random effects, which assumes no unbalanced horizontal pleiotropy[28].

Sensitivity analyses

Sensitivity analyses were conducted to assess the robustness of the results, which involved estimators relying on different assumptions. These included the weighted median, which requires at least 50% of the weighted estimates to be derived from valid instruments[29], and MR-Egger, which allows all instruments to be pleiotropic as long as the instrument strength is independent of the direct effect, at the expense of a lower statistical power[30]. We also used the MR-Egger intercept term to explore the presence of unbalanced horizontal pleiotropy, where a statistically significant ($p < 0.05$) intercept implies its presence[30]. As shown in a previous GWAS, genetic instruments related to HbA$_{1c}$ displayed possible horizontal pleiotropic effects via hemoglobin[12]. Therefore, we conducted multivariable Mendelian randomization (MVMR) for HbA$_{1c}$-related analyses to control for horizontal pleiotropy by adjusting for hemoglobin concentration[31]. Genetic associations with hemoglobin concentration were obtained from the largest GWAS to date, which pooled data from the Blood Cell Consortium and UK Biobank and included 563,946 participants of European ancestry[32]. The genetic associations with hemoglobin concentration were obtained using multivariable linear regression adjusted for age, sex, principal components for genetic ancestry, and cohort specific-covariates[32]. The effect allele of each genetic variant was aligned to that for HbA$_{1c}$[19]. We reported estimates from both MVMR-inverse variance weighted and MVMR-MR-Egger because the latter is robust to directional pleiotropy[33]. Horizontal pleiotropy was assessed using the MVMR-MR-Egger intercept.

Reverse Mendelian randomization

To assess possible reverse causation, we conducted a reverse Mendelian randomization study to assess the association of NMR measured metabolomics ($n = 157$) on type 2 diabetes risk and glycemic traits. However, we did not include the panel on amino acids ($n = 10$) given this overlaps with ongoing and published work[34–36].

Statistics and reproducibility

All statistical analyses were conducted using R version 4.0.5, with the "TwoSampleMR"[19] package used for data harmonization, extraction and alignment, univariable Mendelian randomization analyses, and the "MendelianRandomization"[37] package used for MVMR. We calculated the number of principal components that explained 99% of the variance of the 167 metabolic measures using individual-level data from the UK Biobank (Application 14684), which gave 25 principal components. Given that 5 exposures were considered in this study, the threshold for statistical significance was set at $p < 0.0004$ (0.05/25/5).

Ethics approval

This study only used publicly available summary statistics from relevant GWAS and UK Biobank, thus no ethics approval is required. Respective

ethics approvals have been obtained by the GWAS investigator from all participating studies and the UK Biobank investigators from the North West Multi-center Research Ethic Committee.

Reporting summary

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

Results

Genetic predictors for glycemic traits and liability to type 2 diabetes

We selected 67 genetic instruments for fasting glucose (R^2: 4.7%, F statistics: 25 to 1662), 14 genetic instruments for 2-h glucose (R^2: 1.3%, F statistics: 26 to 111), 74 genetic instruments for HbA$_{1c}$ (R^2: 5.6%, F statistics: 25 to 1392), 38 genetic instruments for fasting insulin (R^2: 1.4%, F statistics: 22 to 173), and 228 instruments for liability to type 2 diabetes (F statistics: 29 to 3136) (Supplementary Data 2). Based on the F statistics, there was little evidence of weak instrument bias. Supplementary Fig. 1 shows the process on instrument selection.

The association of fasting glucose and 2-h glucose with circulating metabolites

Associations of all exposures with circulating metabolites are presented in Figs. 1–4. Results for fasting glucose and 2-h glucose are shown in Supplementary Data 3 and 4. None of the associations of these two glycemic traits with any of the metabolites achieved statistical significance.

The association of HbA$_{1c}$ with circulating metabolites

Higher HbA$_{1c}$ was associated with higher free cholesterol in small LDL (Fig. 4). Estimates from both MR Egger and weighted median were directionally consistent. No evidence of horizontal pleiotropy was observed (Supplementary Data 5). After adjusting for hemoglobin using MVMR, the positive association of HbA$_{1c}$ with free cholesterol in small LDL was slightly attenuated (Supplementary Data 6).

The association of fasting insulin with circulating metabolites

There were 42 associations of higher fasting insulin with circulating metabolites that achieved statistical significance, while only two had a positive effect estimate (Supplementary Data 7). Notably, higher fasting insulin was associated with higher aromatic amino acids (phenylalanine and tyrosine) (Fig. 1). Higher fasting insulin was associated with lower apolipoprotein A1, total cholesterol, HDL-cholesterol and other lipid components in HDL, and total cholines (Fig. 2). There was also a consistent inverse association of fasting insulin with lipoprotein subfractions in medium HDL, large HDL, and very large HDL, as well as IDL (except for triglycerides) (Fig. 4). No evidence of horizontal pleiotropy was observed for any of these associations (Supplementary Data 7).

The association of liability to type 2 diabetes with circulating metabolites

There were 88 associations of type 2 diabetes with circulating metabolites, comprising 34 positive and 54 inverse associations (Supplementary Data 8). The metabolomic signature of type 2 diabetes included most fasting insulin-associated signals (38 out of 42), except for degree of unsaturation in fatty acids, acetone (Fig. 1), phospholipids in medium HDL, and total lipids in medium HDL (Fig. 4). Besides aromatic amino acids, type 2 diabetes was also associated with higher BCAAs (isoleucine, leucine, valine and total BCAA), amino acid (alanine) and glycoprotein acetyls (Fig. 1). Moreover, type 2 diabetes showed positive associations with total triglycerides, triglycerides in HDL, and triglycerides in VLDL (Fig. 2). Consistently, positive associations of type 2 diabetes with lipoprotein subfractions in large VLDL, very large VLDL, and chylomicrons and extremely large VLDL were found (Fig. 3). There was a consistent inverse association of type 2 diabetes with lipoprotein subfractions in large LDL, medium HDL, large HDL, very large HDL and IDL (except triglycerides) (Fig. 4).

### The association of circulating metabolites with glycemic traits and T2D risk

In the reverse Mendelian randomization analyses (Supplementary Data 9–13), there were no associations of circulating metabolites with fasting glucose and 2-h glucose (Supplementary Figs. 2–4). Linoleic acid was associated with lower HbA$_{1c}$ (Supplementary Fig. 2). However, HDL-cholesterol and cholesteryl ester in HDL were inversely associated with fasting insulin, suggesting potential reverse causation (Supplementary Fig. 3). There were 19 circulating metabolites associated with lower type 2 diabetes risk, of which 10 overlapped with the associations of type 2 diabetes with metabolites, such as LDL-cholesterol and subfractions (cholesterol, cholesteryl ester, free cholesterol, phospholipids, total lipids) in large LDL (Supplementary Data 13).

Discussion

This Mendelian randomization studies explored the metabolomic signatures of different glycemic traits and liability to type 2 diabetes. Specifically, fasting glucose and 2-h glucose did not impact lipid profile, consistent with a previous Mendelian randomization study but not other observations studies[9,38,39], although these studies did not investigate 2-h glucose. HbA$_{1c}$ was positively associated with LDL-cholesterol, as well as with free cholesterol and phospholipids in LDL, which were consistent with a previous Mendelian randomization study conducted in Chinese ($n = 11,935$)[40] and other observational studies[41]. Although these glycemic traits were broadly used in defining type 2 diabetes, we found that liability to type 2 diabetes had a strikingly different metabolomic signature compared to fasting glucose and 2-h glucose, and HbA$_{1c}$, such as higher BCAAs, aromatic amino acids, alanine, and lower lipoprotein subfractions, which has been reported in a previous Mendelian randomization study[38]. Our study adds by showing that metabolomic signatures associated with liability to type 2 diabetes resemble the signatures for fasting insulin, which implies that signals of liability to type 2 diabetes cannot be solely explained by hyperglycemia but is likely more related to the consequence of elevated insulin.

As with all study designs, the validity of Mendelian randomization studies depends on assumptions[42]. Whilst Mendelian randomization studies are less susceptible to confounding due to the use of genetics randomly allocated at conception and weak instrument bias was unlikely given the high F statistics, there could be issues with violation of exclusion restriction assumptions where there were signs of horizontal pleiotropy for some analyses. However, the results of sensitivity analyses, which utilized estimators based on different sets of assumptions for validity, gave similar conclusions and hence the associations were unlikely driven completely by the violation of assumptions. We acknowledge that 37.5% of the participants in the GWAS of type 2 diabetes also provided data for the outcome GWAS but biases arising from using two-sample Mendelian randomization methods were likely little given the large F statistics for the instruments. For MR-Egger, the high instrument variability (I^2_{GX} for all included instruments of type 2 diabetes was 0.99) indicated that biases in MR-Egger estimates due to sample overlap are likely minimal[43,44]. Furthermore, we were not able to explore the effect of type 2 diabetes on metabolic signatures since we only instrumented on liability to type 2 diabetes. As such, results from liability to type 2 diabetes should be interpreted with caution[15], where the results could be a mixture of both causes, consequences and merely biomarkers related to type 2 diabetes in the general population with some diagnosed with type 2 diabetes[45]. Despite these limitations, Mendelian randomization studies in general give findings more consistent with randomized controlled trials than conventional observational studies[46].

Previous observational studies showed positive associations of fasting glucose with BCAAs[9], cholesterols in VLDLs, or saturated and unsaturated fatty acids, as well as with lower phospholipids and sphingomyelins[39]. However, confounding cannot be ruled out completely, especially as these studies had limited sample sizes which impacted the ability to adjust for confounders. Furthermore, reverse causation could be an issue, where prospective cohort studies showed these metabolomic markers associated with higher risk of impaired fasting glucose[47]. The lack of association in our
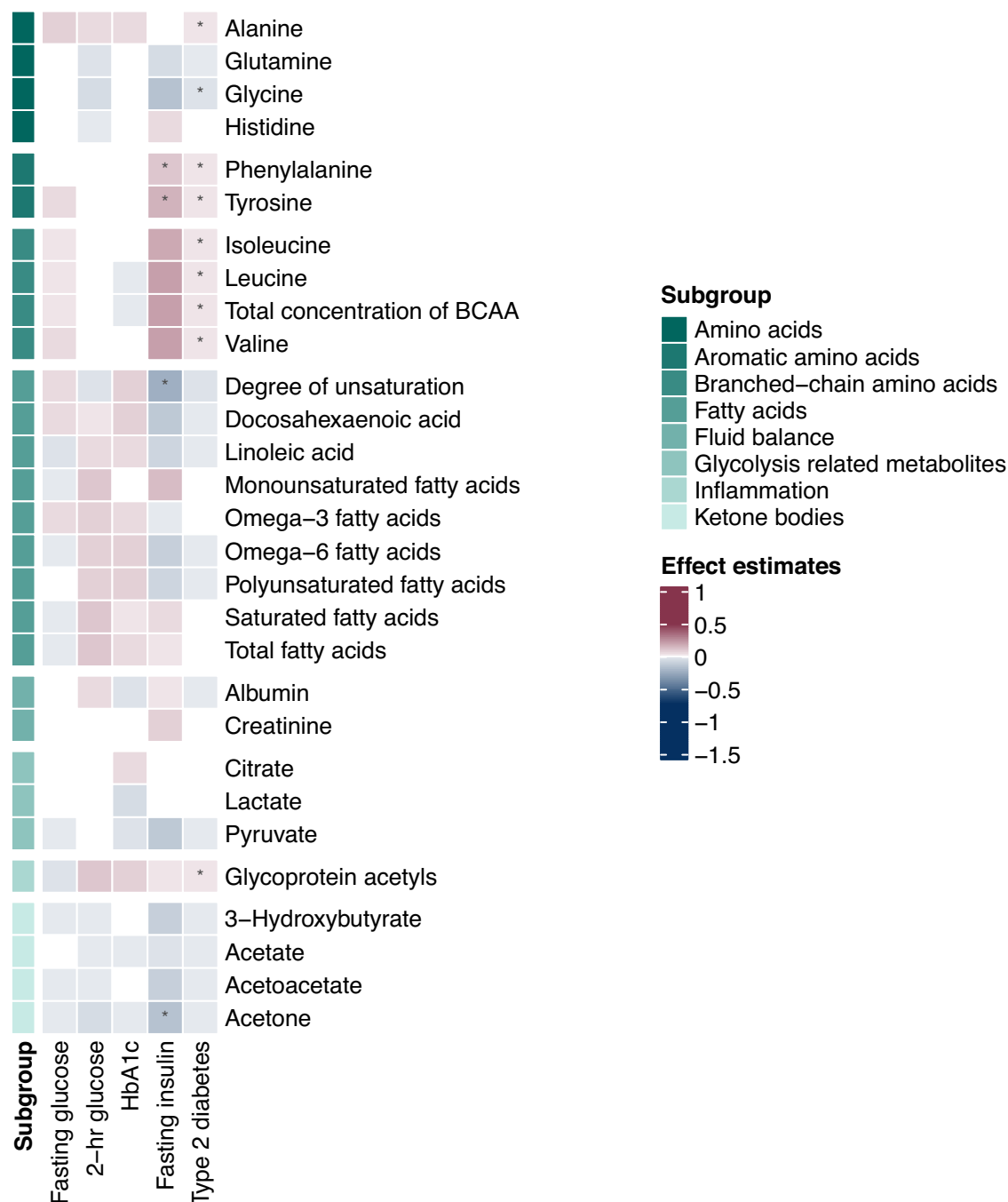
Fig. 1 | Heatmap of associations of genetically predicted glycemic traits (fasting glucose, 2-h glucose, HbA1c and fasting insulin) and of genetic liability to type 2 diabetes with amino acids, fatty acids, and various low-molecular weight metabolites. Circulating metabolites include: amino acids, aromatic amino acids, branched-chain amino acids, fatty acids, biomarkers of fluid balance, glycolysis related metabolites, inflammation, and ketone bodies. The estimates were obtained from Mendelian randomization analyses using the inverse variance weighted method. Asterisks depict statistical significance ($p < 0.0004$). BCAA branched-chain amino acid.

study supported that these previous observations were likely non-causal. Our study is also explored the association of 2-h glucose with metabolomic signature, which showed null associations with included metabolomic signatures.

HbA$_{1c}$ is often used to proxy average glucose over 2–3 months. As these associations were not observed for fasting glucose, these may reflect differences in hyperglycemia being proxied by these two glycemic traits or glycemia-independent effects of HbA$_{1c}$. One possibility is hemoglobin[48], which HbA$_{1c}$ is strongly linked to and may also impact lipid levels, although these associations remained after adjusting for hemoglobin using MVMR[49]. Similarly, other diseases such as glucose-6-phosphate dehydrogenase

deficiency and changes in iron homeostasis markers may also reduce HbA$_{1c}$[50]. Better understanding of the underlying mechanisms has substantial implications for understanding the impact of hyperglycemia, as proxied by fasting glucose, 2-h glucose or HbA$_{1c}$.

The similarities in metabolic signature between liability to type 2 diabetes and fasting insulin, but not other glycemic traits, imply signals associated with type 2 diabetes liability are likely the result of elevated insulin in response to insulin resistance instead of overall hyperglycemia. The lipid signatures, such as inverse associations with IDL and HDL and potentially positive association with VLDL, were similar to a previous smaller Mendelian randomization although that study only focused on liability to type 2

Fig. 2 | Heatmap of associations of genetically predicted glycemic traits (fasting glucose, 2-h glucose, HbA₁c and fasting insulin) and of genetic liability to type 2 diabetes with cholesterol metabolites. Measures of cholesterol metabolites include: apolipoproteins, cholesterol, cholesteryl esters, free cholesterol, lipoprotein particle concentrations, lipoprotein particle sizes, other lipids, phospholipids, total lipids, and triglycerides. The estimates were obtained from Mendelian randomization analyses using the inverse variance weighted method. Asterisks depict statistical significance ($p < 0.0004$). LDL low-density lipoprotein, HDL high-density lipoprotein, VLDL very low-density lipoprotein.

diabetes[51]. Our finding concerning insulin is also consistent with a previous study showing insulin resistance associated with higher BCAAs[52], possibly via decreased BCAA metabolism due to impaired insulin action[53]. Previous studies have debated whether BCAA is a cause of insulin resistance[53], although previous Mendelian randomization studies suggested a non-causal role of BCAAs in insulin resistance[54]. Additional mechanistic studies would be useful to identify the different roles of BCAAs, insulin, and type 2 diabetes and hence targets of intervention.

Fig. 3 | Heatmap of associations of genetically predicted glycemic traits (fasting glucose, 2-h glucose, HbA1c and fasting insulin) and of genetic liability to type 2 diabetes with lipoprotein subclasses in very small, small, medium, large, very large VLDL, and chylomicrons and extremely large VLDL. The estimates were obtained from Mendelian randomization analyses using the inverse variance weighted method. Asterisks depict statistical significance ($p < 0.0004$). VLDL very low-density lipoprotein.

Although this study used Mendelian randomization, which is less susceptible to confounding, we caution readers with several limitations when interpreting the findings. First, valid causal inference depends on satisfying the instrumental variable assumptions when fully assessing pleiotropy is challenging[55]. Nonetheless, several sensitivity analyses with different assumptions about pleiotropy yielded consistent conclusions. Second, since this study utilized data from European populations only, our findings may not generalize to other populations. Assessing metabolic signatures across ethnicities would be helpful, when suitable GWAS becomes available. Third, there were signs of reverse causation for lipids related traits

Fig. 4 | Heatmap of associations of genetically predicted glycemic traits (fasting glucose, 2-h glucose, HbA$_{1c}$ and fasting insulin) and of genetic liability to type 2 diabetes with lipoprotein subfractions in small, medium, and large LDL, in small, medium, large, and very large HDL, and those in IDL. The estimates were obtained from Mendelian randomization analyses using the inverse variance weighted method. Asterisks depict statistical significance ($p < 0.0004$). LDL low-density lipoprotein, HDL high-density lipoprotein, IDL intermediate-density lipoprotein, VLDL very low-density lipoprotein.

after correcting for multiple comparisons, which were consistent with previous Mendelian randomization studies[56,57]. However, these results should be interpreted with caution, in particular the instruments used for NMR measured traits could be pleiotropic, where correction of horizontal

pleiotropy may not be adequate using standard sensitivity analyses. Based on the findings from this study, future studies with more specific hypotheses (e.g., focusing on specific lipid phenotype), coupled with the use of MVMR which could better account for horizontal pleiotropy via other lipids, and the use of larger lipid GWAS[58], will help ascertain the findings from our reverse Mendelian randomization study[2,59,60]. Fourth, genetic instruments for glycemic traits were obtained from MAGIC, which excluded participants with type 2 diabetes. Although such an approach should reduce the likelihood of reverse causation, this may inevitably introduce possible selection bias[42]. Lastly, the metabolomic markers used in this study were mainly related to lipids although other markers relevant to type 2 diabetes, such as BCAAs, were also included. Whether there are shared signatures in other metabolomic markers across glycemic traits require further investigation.

In conclusion, fasting glucose, 2-h glucose, and HbA$_{1c}$ had little evidence of a metabolomic signature for the metabolites considered. Fasting insulin and liability to type 2 diabetes had similar metabolic signatures encompassing a wide range of lipids and amino acids. As such, glycemic traits likely reflect symptoms of type 2 diabetes while insulin also has a role in the pathophysiology of type 2 diabetes independent of hyperglycemia. The distinct characterization of these inter-related glycemic traits may help us better understand the mechanisms underpinning the relation of these traits with downstream clinical outcomes such as cardiovascular diseases, as well as diagnosis and clinical management of type 2 diabetes using these traits.

Data availability

Summary statistics of glycemic traits are available from the MAGIC, with details provided in Supplementary Data 2. Summary statistics of NMR metabolites are available from the IEU GWAS database (https://gwas.mrcieu.ac.uk/), with the IDs for each metabolite listed in Supplementary Data 1.

References

1. Emerging Risk Factors, C. et al. Diabetes mellitus, fasting blood glucose concentration, and risk of vascular disease: a collaborative meta-analysis of 102 prospective studies. *Lancet* **375**, 2215–2222 (2010).
2. Yuan, S. et al. Differentiating associations of glycemic traits with atherosclerotic and thrombotic outcomes: Mendelian randomization investigation. *Diabetes* **71**, 2222–2232 (2022).
3. American Diabetes Association Professional Practice C. 2. Classification and diagnosis of diabetes: standards of medical care in diabetes—2022. *Diabetes Care* **45**, S17–S38 (2021).
4. Beck, R. W. et al. The fallacy of average: how using HbA1c alone to assess glycemic control can be misleading. *Diabetes Care* **40**, 994 (2017).
5. Würtz, P. et al. Metabolic signatures of insulin resistance in 7,098 young adults. *Diabetes* **61**, 1372–1380 (2012).
6. Mahendran, Y. et al. Association of ketone body levels with hyperglycemia and type 2 diabetes in 9,398 Finnish men. *Diabetes* **62**, 3618–3626 (2013).
7. Wang, J. et al. Lipoprotein subclass profiles in individuals with varying degrees of glucose tolerance: a population-based study of 9399 Finnish men. *J. Intern. Med.* **272**, 562–572 (2012).
8. Zhang, X. et al. Association of hemoglobin A1C with circulating metabolites in Dutch with European, African Surinamese and Ghanaian background. *Nutr. Diabetes* **9**, 15 (2019).
9. Jun, G. et al. Metabolomic profiles associated with subtypes of prediabetes among Mexican Americans in Starr County, Texas, USA. *Diabetologia* **63**, 287–295 (2020).
10. Lawlor, D. A. et al. Mendelian randomization: using genes as instruments for making causal inferences in epidemiology. *Stat. Med.* **27**, 1133–1163 (2008).

11. Davies, N. M., Holmes, M. V. & Davey Smith, G. Reading Mendelian randomisation studies: a guide, glossary, and checklist for clinicians. *BMJ* **362**, k601 (2018).

12. Chen, J. et al. The trans-ancestral genomic architecture of glycemic traits. *Nat. Genet.* **53**, 840–860 (2021).

13. Scott, R. A. et al. Large-scale association analyses identify new loci influencing glycemic traits and provide insight into the underlying biological pathways. *Nat. Genet.* **44**, 991–1005 (2012).

14. Vujkovic, M. et al. Discovery of 318 new risk loci for type 2 diabetes and related vascular outcomes among 1.4 million participants in a multi-ancestry meta-analysis. *Nat. Genet.* **52**, 680–691 (2020).

15. Burgess, S. & Labrecque, J. A. Mendelian randomization with a binary exposure variable: interpretation and presentation of causal estimates. *Eur. J. Epidemiol.* **33**, 947–952 (2018).

16. Au Yeung, S. L., Zhao, J. V. & Schooling, C. M. Evaluation of glycemic traits in susceptibility to COVID-19 risk: a Mendelian randomization study. *BMC Med.* **19**, 72 (2021).

17. Dixon, P. et al. Estimating the causal effect of liability to disease on healthcare costs using Mendelian Randomization. *Econ. Hum. Biol.* **46**, 101154 (2022).

18. Yuan, S. et al. Genetic liability to insomnia in relation to cardiovascular diseases: a Mendelian randomisation study. *Eur. J. Epidemiol.* **36**, 393–400 (2021).

19. Hemani, G. et al. The MR-Base platform supports systematic causal inference across the human phenome. *eLife* **7**, e34408 (2018).

20. Würtz, P. et al. Quantitative serum nuclear magnetic resonance metabolomics in large-scale epidemiology: a primer on -omic technologies. *Am. J. Epidemiol.* **186**, 1084–1096 (2017).

21. Bycroft, C. et al. The UK Biobank resource with deep phenotyping and genomic data. *Nature* **562**, 203–209 (2018).

22. Loh, P.-R. et al. Efficient Bayesian mixed-model analysis increases association power in large cohorts. *Nat. Genet.* **47**, 284–290 (2015).

23. Auton, A. et al. A global reference for human genetic variation. *Nature* **526**, 68–74 (2015).

24. Pierce, B. L., Ahsan, H. & VanderWeele, T. J. Power and instrument strength requirements for Mendelian randomization studies using multiple genetic variants. *Int. J. Epidemiol.* **40**, 740–752 (2011).

25. Lee, S. H. et al. A better coefficient of determination for genetic profile analysis. *Genet. Epidemiol.* **36**, 214–224 (2012).

26. Au Yeung, S. L. et al. Impact of lung function on cardiovascular diseases and cardiovascular risk factors: a two sample bidirectional Mendelian randomisation study. *Thorax* **77**, 164 (2022).

27. Yarmolinsky, J. et al. Circulating selenium and prostate cancer risk: a Mendelian randomization analysis. *J. Natl Cancer Inst.* **110**, 1035–1038 (2018).

28. Bowden, J. et al. A framework for the investigation of pleiotropy in two-sample summary data Mendelian randomization. *Stat. Med.* **36**, 1783–1802 (2017).

29. Bowden, J. et al. Consistent estimation in Mendelian randomization with some invalid instruments using a weighted median estimator. *Genet Epidemiol.* **40**, 304–314 (2016).

30. Bowden, J., Davey Smith, G. & Burgess, S. Mendelian randomization with invalid instruments: effect estimation and bias detection through Egger regression. *Int. J. Epidemiol.* **44**, 512–525 (2015).

31. Luo, S., Au Yeung, S. L. & Schooling, C. M. Assessing the linear and non-linear association of HbA1c with cardiovascular disease: a Mendelian randomisation study. *Diabetologia* **64**, 2502–2510 (2021).

32. Vuckovic, D. et al. The polygenic and monogenic basis of blood traits and diseases. *Cell* **182**, 1214–31.e11 (2020).

33. Rees, J. M. B., Wood, A. M. & Burgess, S. Extending the MR-Egger method for multivariable Mendelian randomization to correct for both measured and unmeasured pleiotropy. *Stat. Med.* **36**, 4705–4718 (2017).

34. Mo, J. et al. Evaluating the role of amino acids in type 2 diabetes risk: a Mendelian randomization study. *medRxiv* 2023:2023.08.27.23294702. https://doi.org/10.1101/2023.08.27.23294702.

35. Zhao, J. V., Fan, B., Burgess, S. Using genetics to examine the overall and sex-specific associations of branch-chain amino acids and the valine metabolite, 3-hydroxyisobutyrate, with ischemic heart disease and diabetes: a two-sample mendelian randomization study. *Atherosclerosis* https://doi.org/10.1016/j.atherosclerosis.2023.117246.

36. Huang, X., Zhao, J. V. The associations of genetically predicted plasma alanine with coronary artery disease (CAD) and CAD risk factors: a Mendelian randomization study. *Am. J. Clin. Nutr.* https://doi.org/10.1016/j.ajcnut.2023.08.011.

37. Yavorska, O. O. & Burgess, S. MendelianRandomization: an R package for performing Mendelian randomization analyses using summarized data. *Int. J. Epidemiol.* **46**, 1734–1739 (2017).

38. Liu, J. et al. A Mendelian randomization study of metabolite profiles, fasting glucose, and type 2 diabetes. *Diabetes* **66**, 2915–2926 (2017).

39. Xu, F. et al. Metabolic signature shift in type 2 diabetes mellitus revealed by mass spectrometry-based metabolomics. *J. Clin. Endocrinol. Metab.* **98**, E1060–E1065 (2013).

40. Jia, X. et al. Mendelian randomization analysis support causal associations of HbA1c with circulating triglyceride, total and low-density lipoprotein cholesterol in a Chinese population. *Sci. Rep.* **9**, 5525 (2019).

41. Du, T. et al. Sex differences in the effect of HbA1c-defined diabetes on a wide range of cardiovascular disease risk factors. *Ann. Med.* **48**, 34–41 (2016).

42. Sanderson, E. et al. Mendelian randomization. *Nat. Rev. Methods Prim.* **2**, 6 (2022).

43. Minelli, C. et al. The use of two-sample methods for Mendelian randomization analyses on single large datasets. *Int. J. Epidemiol.* **50**, 1651–1659 (2021).

44. Bowden, J. et al. Assessing the suitability of summary data for two-sample Mendelian randomization analyses using MR-Egger regression: the role of the I2 statistic. *Int. J. Epidemiol.* **45**, 1961–1974 (2016).

45. Holmes, M. V. & Davey Smith, G. Can Mendelian randomization shift into reverse gear? *Clin. Chem.* **65**, 363–366 (2019).

46. Holmes, M. V., Ala-Korpela, M. & Davey Smith, G. Mendelian randomization in cardiometabolic disease: challenges in evaluating causality. *Nat. Rev. Cardiol.* **14**, 577–590 (2017).

47. Ahola-Olli, A. V. et al. Circulating metabolites and the risk of type 2 diabetes: a prospective study of 11,896 young adults from four Finnish cohorts. *Diabetologia* **62**, 2298–2309 (2019).

48. Bae, J. C. et al. Hemoglobin A1c values are affected by hemoglobin level and gender in non-anemic Koreans. *J. Diabetes Investig.* **5**, 60–65 (2014).

49. Hämäläinen, P. et al. Hemoglobin level and lipoprotein particle size. *Lipids Health Dis.* **17**, 10 (2018).

50. Liang, Y. et al. Association of iron homeostasis biomarkers in type 2 diabetes and glycaemic traits: a bidirectional two-sample Mendelian randomization study. *Int. J. Epidemiol.* https://doi.org/10.1093/ije/dyad093 (2023).

51. Bell, J. A. et al. Early metabolic features of genetic liability to type 2 diabetes: cohort study with repeated metabolomics across early life. *Diabetes Care* **43**, 1537–1545 (2020).

52. Wang, Q. et al. Genetic support for a causal role of insulin resistance on circulating branched-chain amino acids and inflammation. *Diabetes Care* **40**, 1779–1786 (2017).

53. Lynch, C. J. & Adams, S. H. Branched-chain amino acids in metabolic signalling and insulin resistance. *Nat. Rev. Endocrinol.* **10**, 723–736 (2014).

54. Mahendran, Y. et al. Genetic evidence of a causal effect of insulin resistance on branched-chain amino acid levels. *Diabetologia* **60**, 873–878 (2017).

55. Bowden, J., Burgess, S. & Smith, G. D. Difficulties in testing the instrument strength independent of direct effect assumption in Mendelian randomization. *JAMA Cardiol.* **2**, 929–930 (2017).

56. Fall, T. et al. Using genetic variants to assess the relationship between circulating lipids and type 2 diabetes. *Diabetes* **64**, 2676–2684 (2015).

57. White, J. et al. Association of lipid fractions with risks for coronary artery disease and diabetes. *JAMA Cardiol.* **1**, 692–699 (2016).

58. Graham, S. E. et al. The power of genetic diversity in genome-wide association studies of lipids. *Nature* **600**, 675–679 (2021).

59. Karjalainen, M. K. et al. Genome-wide characterization of circulating metabolic biomarkers reveals substantial pleiotropy and novel disease pathways. medRxiv https://doi.org/10.1101/2022.10.20.22281089 (2022).

60. Burgess, S., Dudbridge, F. & Thompson, S. G. Re: "Multivariable Mendelian randomization: the use of pleiotropic genetic variants to estimate causal effects. *Am. J. Epidemiol.* **181**, 290–291 (2015).

Acknowledgements

Author contributions

S.L.A.Y. conceptualized the study, T.W.H.T., J.M.Y.M., and M.Z. conducted the analysis with feedback from S.L.A.Y. and S.L.; T.W.H.T., J.M.Y.M., M.Z., S.L., and S.L.A.Y. interpreted the results. C.M.S. and J.V.Z. provided additional interpretation concerning insulin related analyses. T.W.H.T., J.M.Y.M., and S.L. wrote the first draft of the manuscript with critical feedback and revisions from J.V.Z., C.M.S., B.H., and S.L.A.Y. All authors gave final approval of the version to be published. T.W.H.T. and J.M.Y.M. are the guarantors.

Competing interests

The authors declare no competing interests.

Additional information

Supplementary information The online version contains supplementary material available at https://doi.org/10.1038/s42003-024-05977-1.

Correspondence and requests for materials should be addressed to Shan Luo.

Peer review information *Communications Biology* thanks Segun Fatumo, and the other, anonymous, reviewer(s) for their contribution to the peer review of this work. Primary Handling Editor: George Inglis.

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

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