

Circulating metabolites and the risk of type 2 diabetes: a prospective study of 11,896 young adults from four Finnish cohorts

Ari V. Ahola-Olli^{1,2,3}, Linda Mustelin^{3,4,5}, Maria Kalimeri⁴, Johannes Kettunen^{6,7,8}, Jari Jokelainen^{6,9}, Juha Auvinen^{6,9,10}, Katri Puukka¹¹, Aki S. Havulinna^{3,12}, Terho Lehtimäki¹³, Mika Kähönen¹⁴, Markus Juonala¹⁵, Sirkka Keinänen-Kiukaanniemi^{6,9,16,17}, Veikko Salomaa¹², Markus Perola^{3,12,18}, Marjo-Riitta Järvelin^{6,9,19,20,21}, Mika Ala-Korpela^{7,8,22,23,24,25}, Olli Raitakari^{1,26}, Peter Würtz^{4,27}

1. Research Centre of Applied and Preventive Cardiovascular Medicine, University of Turku, Turku, Finland.
2. Department of Internal Medicine, Satakunta Central Hospital, Pori, Finland.
3. Institute for Molecular Medicine (FIMM), University of Helsinki, Helsinki, Finland.
4. Nightingale Health Ltd, Helsinki, Finland.
5. Department of Public Health, University of Helsinki, Helsinki, Finland.
6. Centre for Life Course Health Research, Faculty of Medicine, University of Oulu, Oulu, Finland.
7. Computational Medicine, Faculty of Medicine, University of Oulu and Biocenter Oulu, Finland
8. NMR Metabolomics Laboratory, School of Pharmacy, University of Eastern Finland, Kuopio, Finland
9. Unit of Primary Health Care and Medical Research Center, Oulu University Hospital, Oulu, Finland.
10. Oulunkaari Primary Health Care Unit, Ii, Finland
11. University of Oulu and Medical Research Center Oulu, Oulu, Finland; Nordlab Oulu, Oulu University Hospital and Department of Clinical Chemistry, University of Oulu, Oulu, Finland
12. National Institute for Health and Welfare, Helsinki, Finland.
13. Department of Clinical Chemistry, Fimlab Laboratories, and Finnish Cardiovascular Research Center - Tampere, Faculty of Medicine and Health Technology, Tampere University, Finland.
14. Department of Clinical Physiology, University of Tampere and Tampere University Hospital, Tampere, Finland
15. Department of Medicine, University of Turku and Division of Medicine, Turku University Hospital, Turku, Finland.
16. Healthcare and Social Services of Selanne, Pyhasalmi, Finland
17. Healthcare Services of City of Oulu, Diabetes Unit, Finland
18. Estonian Genome Center, University of Tartu, Tartu, Estonia.
19. Department of Epidemiology and Biostatistics, MRC-PHE Centre for Environment and Health, Imperial College London, London, UK.
20. Biocenter Oulu, University of Oulu, Finland
21. Department of Life Sciences, College of Health and Life Sciences, Brunel University London, UK
22. Medical Research Council Integrative Epidemiology Unit at the University of Bristol, Bristol, UK
23. Population Health Science, Bristol Medical School, University of Bristol, Bristol, UK
24. Systems Epidemiology, Baker Heart and Diabetes Institute, Melbourne, VIC, Australia
25. Department of Epidemiology and Preventive Medicine, School of Public Health and Preventive Medicine, Faculty of Medicine, Nursing and Health Sciences, The Alfred Hospital, Monash University, Melbourne, VIC, Australia
26. Department of Clinical Physiology and Nuclear Medicine, Turku University Hospital, Turku, Finland.
27. Research Programs Unit, Diabetes and Obesity, University of Helsinki, Helsinki, Finland.

Corresponding authors:

Dr Ari Ahola-Olli; avahol@utu.fi

University of Turku, Kiinamylynkatu 10, 20520 Turku, Finland.

Dr Peter Würtz; peter.wurtz@nightingalehealth.com

Nightingale Health Ltd, Mannerheimintie 164a, 00300 Helsinki, Finland; Phone: +358 504670900

ABSTRACT

Objective: Advances in metabolomics now allow high-throughput biomarker profiling of large population studies. We aimed to identify circulating metabolic biomarkers predictive of type 2 diabetes in young adults.

Methods: Nuclear magnetic resonance metabolomics was used to quantify 229 metabolic measures in 11,896 individuals from four Finnish cohorts (mean age 33 years, range 24–45). Associations between baseline metabolites and risk of type 2 diabetes onset during 8–15 years of follow-up (392 incident cases) were assessed by logistic regression adjusted for sex, age, body mass index, and fasting glucose.

Results: Out of 229 metabolic measures, 113 were associated with incident diabetes in meta-analysis of the four cohorts ($P < 0.0009$; odds ratios per 1-SD: 0.59–1.50). Among the strongest predictors of diabetes risk were branched-chained and aromatic amino acids (odds ratios 1.31–1.33), triglyceride fractions within the largest very-low-density lipoprotein particles (VLDL; odds ratios 1.33–1.50), as well as linoleic omega-6 fatty acids (odds ratio 0.75) and free cholesterol in large high-density lipoprotein particles (HDL; odds ratio 0.59). A biomarker score comprised of phenylalanine, free cholesterol in large HDL, and the ratio of cholesteryl esters to total lipids in large VLDL was predictive of the risk for future diabetes in an independent validation cohort (odds ratio 10.1 [95% confidence intervals 4.2–24.1] comparing individuals in upper vs lower fifth of biomarker score). Adjustment for routine lipids and insulin attenuated the odds ratio to 5.8 [2.2–15.1].

Conclusions: Metabolic aberrations across multiple molecular pathways are predictive of the long-term risk of type 2 diabetes in young adults. Comprehensive metabolic profiling may potentially help targeting preventive interventions for young asymptomatic individuals at increased risk for type 2 diabetes.

INTRODUCTION

The global prevalence of type 2 diabetes is increasing rapidly, particularly in low-income and middle-income countries (1). Type 2 diabetes is associated with increased mortality from vascular and numerous other causes, and reduced quality of life, causing an immense societal cost burden (2, 3). Given the availability of lifestyle interventions that are effective at preventing or delaying the onset of type 2 diabetes (4, 5), early identification of individuals at high risk is important. The risk for developing type 2 diabetes is to some extent reflected in current laboratory measures, including hyperglycaemia and dyslipidemia. Such metabolic aberrations are often present years before the diagnosis can be set. This has led to widespread implementation of prediabetes screening programs for identifying those at high risk for progression to diabetes. However, current screening approaches are ineffective in identifying high risk individuals, leading to missed opportunities for targeted prevention (6).

Detailed metabolic profiling technologies, also known as metabolomics, are increasingly used in epidemiological studies (7, 8). This has proven a powerful approach to identify biochemical changes occurring before disease onset and elucidate the pathophysiology of type 2 diabetes. Multiple nested case-control studies have identified circulating metabolites associated with the risk for type 2 diabetes using a range of technological assays, based on mass spectrometry or nuclear magnetic resonance (NMR) (7, 9, 10). The most consistent biomarker associations with type 2 diabetes have been observed for branched-chain and aromatic amino acids (8). Genetic evidence and experimental data suggest that these amino acids may be causally implicated in the development of insulin resistance and type 2 diabetes (11, 12). However, previous metabolomics studies have commonly involved a modest number of participants, and have almost exclusively been conducted for middle-aged and older individuals.

In this study, we characterized a comprehensive metabolic signature of increased risk for future type 2 diabetes in young adults, up to 15 years before disease onset. We used NMR metabolomics to quantify 229 metabolic measures in 11,896 individuals from four population-based cohorts with individuals aged 24–45 at blood draw. We aimed to describe the metabolic aberrations reflective of the risk for future type 2 diabetes in young adults, including their consistency between men and women and across cohorts. In addition, we explored the predictive value of a multi-biomarker score for early risk of type 2 diabetes.

METHODS

Study populations

The study involved 11,896 individuals from four prospective population-based cohorts in Finland. An overview of the study cohorts and participants included in the present analyses is shown in **Supplementary Figure 1**. Details of the individual cohorts are provided in **Supplementary Methods**. In all cohorts, we excluded individuals with diabetes at baseline, pregnant women, study participants aged over 45 years at the blood draw, and those lacking follow-up information on diabetes diagnosis. Briefly, the characteristics of each cohort were as follows:

1) In the Cardiovascular Risk in Young Finns Study (YFS), serum metabolites were quantified from 2,248 individuals in the 2001 survey. The final sample consisted of 2,141 individuals of age range 24–39 years. Follow-up time was 10 years. Type 2 diabetes diagnoses at 10-year follow-up were based either on HbA1c or fasting glucose assessed in the 2011 re-survey, or nationwide registers on reimbursement for diabetes medication or in-patient hospital ICD-10 diagnosis of diabetes (**Supplementary Methods**) (13).

2) In FINRISK-1997, serum metabolites were quantified from 7,603 individuals. The final sample consisted of 3,063 individuals when limiting analyses to participants aged 24–45. Follow-up time was 15 years. Type 2 diabetes diagnoses at follow-up were based on nationwide register data (14).

3) In the Dietary, Lifestyle, and Genetic Determinants of Obesity and Metabolic Syndrome Study (DILGOM), serum metabolites were quantified from 4,816 individuals in 2007. The final sample consisted of 1,421 individuals when limiting analyses to participants in the age range 25–45. Follow-up time was 7.8 years. Type 2 diabetes diagnoses at follow-up were based on fasting glucose at the re-survey conducted in 2014 or nationwide register data.

4) In the Northern Finland Birth Cohort (NFBC) of 1966, serum metabolites were quantified from 5,680 individuals in the 1997 survey. The final sample consisted of 5,275 individuals aged 30–32 years. Follow-up time was 15 years. Type 2 diabetes diagnoses were based on either fasting or 2-hour glucose at the 46-year follow-up conducted in 2012, or nationwide register data.

Metabolite quantification

A high-throughput NMR metabolomics platform (Nightingale Health Ltd, Helsinki, Finland) was used to quantify 229 metabolic measures from baseline serum samples (15). This metabolite panel captures multiple metabolic pathways, including amino acids, glycolysis-related metabolites, fatty acids and detailed lipoprotein lipid profiles, covering triglycerides, total cholesterol, free cholesterol, esterified cholesterol, and phospholipids within 14 subclasses. Details of the NMR

metabolomics experimentation have been described previously (15) and epidemiological applications have recently been reviewed (7).

Statistical analyses

Due to skewness of the metabolite distributions, all metabolite concentrations were $\log_e(\text{metabolite}+1)$ transformed prior to analyses. Although 229 metabolic measures in total were analyzed, the number of independent tests performed is lower due to the correlated nature of the measures (7). We calculated that 54 principal components explained 99% of the variation in the metabolic measures. Hence, we inferred statistical significance at meta-analysis P-value < 0.0009 ($0.05/54$). Odds ratios of 229 circulating metabolic measures with incidence of type 2 diabetes were assessed using logistic regression. Each metabolite was analyzed as predictor of diabetes in a separate model, adjusted for sex, baseline age, fasting glucose, and body mass index (BMI). We also assessed the influence of additional adjustment for fasting insulin. To facilitate comparison of the biomarker association magnitudes for measures with units and different concentration range, the odds ratios are scaled to 1-standard deviation (SD) increment in log-transformed metabolite concentration. Results from individual cohorts were combined using inverse variance-weighted fixed-effect meta-analysis. Analyses were also conducted separately for men and women.

To compare the pattern of metabolite associations for future diabetes risk with those of established risk factors, we subsequently assessed the cross-sectional associations of all metabolites with BMI, homeostatic model assessment insulin resistance (HOMA-IR) index, and fasting glucose in linear regression models adjusted for age and sex.

Lastly, we examined the predictive value of a combination of metabolites in the NFBC validation cohort based on a biomarker summary score derived in the three other cohorts (YFS, FINRISK-1997, and DILGOM; constituting approximately half of the incident cases). Here, a parsimonious biomarker summary score was derived based on forward step-wise logistic models testing of all metabolites, using meta-analysis of the three derivation cohorts. Age, sex, baseline fasting glucose, and BMI were always included as covariates in the models. In each step, the metabolite with the lowest P-value was added as a covariate, and associations of all remaining metabolites with diabetes risk were assessed. This process was repeated until no further metabolites were significant at $P < 0.0009$. The biomarker summary score was then defined as the sum of concentrations of the three selected metabolites weighted by beta-coefficients in the final stepwise model. The predictive ability of this biomarker summary score was then evaluated in NFBC, since this validation cohort

had the highest number of cases and most reliably ascertained diagnoses. Odds ratios of the biomarker summary score were assessed both as a continuous marker and by quintiles, with adjustment for sex, baseline age, fasting glucose, and BMI. The influence of further adjusting for fasting insulin, triglycerides and HDL-cholesterol was also assessed. The potential improvement in risk discrimination when adding the biomarker summary score to models containing these two sets of clinical variables were compared in terms of C-statistic, integrated discrimination improvement and continuous reclassification (16). Statistical analyses were performed in R version 3.1.3.

RESULTS

The study included 11,896 individuals from four Finnish cohorts. Characteristics of the study participants at the time of blood sampling are shown in **Table 1**. The mean age was 32.9 years (range 24–45). The follow-up time ranged from 8 to 15 years, during which a total of 392 incident cases of type 2 diabetes occurred. Mean concentrations and SDs of all metabolic measures are listed in **Supplementary Table 1**. The odds ratios of 104 selected metabolic measures with incident type 2 diabetes are shown in **Figure 1** and **Figure 2**; results for the remaining 125 metabolic measures assayed are found in **Supplementary Figure 2**. In meta-analysis of all four cohorts, 113 out of the 229 metabolic measures were robustly associated with incident type 2 diabetes ($P < 0.0009$) when adjusting for sex, baseline age, BMI, and fasting glucose. The biomarkers associated with risk of future type 2 diabetes risk spanned multiple metabolic pathways of polar metabolites, fatty acids and detailed lipoprotein lipid measures, with significant odds ratios ranging from 1.18–1.50 for direct associations and from 0.59–0.86 for inverse associations per 1-SD metabolite concentration.

Amino acids, glycolysis, and inflammation

Branched-chain amino acids isoleucine, leucine, and valine (odds ratios 1.20–1.33) and the aromatic amino acids phenylalanine and tyrosine (1.33 and 1.18, respectively) were associated with the risk of type 2 diabetes (**Figure 1**). Glycerol was also associated with increased risk (1.22 [95% confidence interval 1.10–1.35]) while other glycolysis related metabolites had weaker associations. The inflammatory biomarker glycoprotein acetyls (GlycA) displayed one of the strongest associations for type 2 diabetes risk (1.37 [1.24–1.51]).

Fatty acids

The total concentration of circulating fatty acids as well as the relative amount of monounsaturated fatty acids (MUFA; ratio to total fatty acids) were directly associated with increased risk for type 2

diabetes (1.32 [1.18–1.48]). In contrast, higher relative concentrations of omega-6 fatty acids were associated with decreased risk for type 2 diabetes (0.75 [0.69-0.83]). This inverse association was primarily driven by linoleic acid, whereas the association for arachidonic acid was weaker.

Lipoprotein measures

Both lipid measures used in routine clinical settings and more fine-grained lipoprotein subclass measures are quantified by the NMR metabolomics platform. The associations of routine lipids, as well as cholesterol and triglyceride concentrations in 14 lipoprotein subclasses, with type 2 diabetes risk are shown in **Figure 2**. Additional lipoprotein subclass measures are shown in **Supplementary Figure 2**.

Overall, the cholesterol concentration within VLDL particles was associated with increased risk for type 2 diabetes, whereas the cholesterol in HDL particles was associated with decreased risk. Triglycerides in both VLDL, LDL, and HDL particles were strongly associated with increased diabetes risk. The analysis of 14 lipoprotein subclasses and their lipid content provides a more nuanced picture of these associations. Cholesterol in large HDL, in particular, was associated with decreased diabetes risk. The association patterns were similar for free cholesterol and cholesteryl esters; the strongest biomarker for decreased diabetes risk was free cholesterol in large HDL (0.59 [0.50-0.68]; **Supplementary Figure 2**). The absolute triglyceride concentrations within most of the lipoprotein subclasses were strongly associated with increased diabetes risk. This was also the case for the relative fraction of triglyceride within each subclass, i.e. the percentage of triglycerides to total lipids in a given lipoprotein particle size. Higher relative triglyceride content in lipoprotein particles generally reflects a lower cholesterol content, which is why inverse associations with future diabetes risk are seen for the percentage fractions of cholesterol in most of the lipoprotein subclasses.

Concentrations of apolipoproteins, the structural proteins of lipoprotein particles, were also associated with increased risk for type 2 diabetes. In particular, the ratio of apolipoprotein B to apolipoprotein A1 was among the strongest predictors (1.40 [1.25-1.58]). Further, larger VLDL particle size was associated with increased diabetes risk (1.32 [1.19-1.47]) whereas larger HDL particle size displayed an inverse association (0.62 [0.54-0.72]).

Consistency across cohorts and influence of insulin adjustment

The association patterns between metabolites and incident type 2 diabetes were highly consistent in all four cohorts despite between-cohort differences in fasting status and ascertainment of diabetes

diagnoses at follow-up (**Supplementary Figure 3**). The metabolite associations were also similar for men and women (**Supplementary Figure 4**).

Since the metabolite associations with type 2 diabetes risk in our study are reminiscent of those previously reported with glycemic traits and adiposity (17, 18), we examined the cross-sectional associations of metabolites with BMI, HOMA-IR index, and fasting glucose (**Supplementary Figure 5**). We found a high consistency in the overall pattern of metabolite associations for diabetes risk with those observed for BMI and HOMA-IR index, whereas the associations with fasting glucose were substantially weaker in these young adults.

To evaluate whether the metabolite biomarkers provide information beyond insulin resistance, we further adjusted analyses for baseline insulin levels. Most associations between metabolites and future risk of type 2 diabetes were moderately attenuated when including insulin as covariate, but the overall pattern persisted and 71 of the metabolic measures remained significant at $P < 0.0009$ (**Supplementary Figure 6**).

Biomarker score identifies individuals with increased risk of diabetes

To examine potential predictive utility of a combination of metabolic measures, we derived a biomarker summary score for the risk of future type 2 diabetes. Using a stepwise modeling approach based on three of the cohorts, three metabolic measures were selected as independent predictors of diabetes: phenylalanine, free cholesterol in large HDL, and cholesteryl esters (CE) to total lipids ratio within large VLDL. The biomarker summary score was then evaluated in the NFBC validation cohort. Here, the odds ratio of incident type 2 diabetes was 10.1 [4.2-24.1] among individuals in the upper fifth of the biomarker summary score compared to those in the lower fifth, when adjusting for age, sex, and baseline glucose and BMI (**Table 2**). If further adjusting for baseline insulin, triglycerides, and HDL-cholesterol, the odds ratio for individuals in the highest vs. lowest fifth of the biomarker summary score was attenuated to 5.79 [2.22-15.1]. The discrimination in absolute risk for future type 2 diabetes is presented in **Supplementary Tables 3 and 4** and **Supplementary Figure 7**; the prediction model enhanced with the biomarker summary score displayed improved risk discrimination above the basic clinical model (age, sex, BMI and fasting glucose), but only modest improvements above a model extended with triglycerides, HDL cholesterol, and fasting insulin.

DISCUSSION

This large, multi-cohort study describes a metabolic signature of increased risk for type 2 diabetes in young adults up to 15 years prior to disease onset. Metabolic aberrations related to incident type 2 diabetes risk spanned amino acids, fatty acid balance, inflammation, and detailed lipoprotein particle composition, with consistent results across the four cohorts. Among the strongest biomarkers were higher concentrations of branched-chained and aromatic amino acids, VLDL particle measures, and the enrichment of triglycerides in all lipoprotein subclasses. Moreover, higher circulating levels of GlycA, glycerol, and MUFA also predicted increased risk for type 2 diabetes, whereas glutamine, linoleic acid, as well as HDL particle size and certain lipid measures within large HDL were associated with lower risk. A biomarker score consisting of three metabolic measures was predictive of a 10-fold elevation in the long-term risk for type 2 diabetes among 31-year old men and women.

The metabolic signature for type 2 diabetes risk described here included biomarkers across multiple molecular pathways. Branched-chain and aromatic amino acids were among the first biomarkers for type 2 diabetes risk identified by metabolomics (10). Their association with diabetes risk has since been replicated in several epidemiological studies (8, 9, 19) and extended to insulin resistance and glycemia (9, 18, 20). The odds ratios of all amino acids assayed in this study were consistent with a recent meta-analysis of prospective studies (8). We extend these prior findings by showing that branched-chain and aromatic amino acid levels are predictive of the long-term risk of type 2 diabetes already in young adults.

The causal relation between amino acid levels and type 2 diabetes risk is not yet fully clear. Mendelian randomization studies have indicated that adiposity and insulin resistance leads to increased branched-chain amino acid levels (12, 17), but also that the metabolism of these amino acids play a causal role in the development of type 2 diabetes (11). In addition, physiological studies have suggested mechanisms by which alterations in BCAA metabolism might cause insulin resistance and impairment of insulin secretion (21, 22). Altered amino acid metabolism may also represent a link between diabetes and cardiovascular diseases, common complications of type 2 diabetes (23, 24). Our results in young adults support the notion that amino acid profiling may prove helpful in monitoring of cardiometabolic health in asymptomatic individuals, with potential to facilitate targeted interventions and tracking of their effects (25).

Pervasive alterations in the lipoprotein profile were found to be predictive of future diabetes risk. These included both established lipid measures and novel associations of increased particle size and lipid content of VLDL, and enrichment of triglycerides in all lipoprotein subclasses. These lipid modulations are similar to those previously reported in cross-sectional settings for individuals with impaired glucose tolerance (26, 27). They may reflect early stages of the aberrations in lipoprotein metabolism characteristic of insulin resistance: increased production of large VLDLs, increased catabolism of HDLs, and increased transfer of triglycerides to HDL and LDL particles (28). Our findings indicate that such distortions of lipoprotein metabolism may already be present in normoglycemic young adults and reflect an increased risk for type 2 diabetes.

Increasing evidence suggests that certain fatty acid levels are predictive of type 2 diabetes. Our finding that a higher relative concentration of omega-6 fatty acids was associated with decreased diabetes risk, whereas higher MUFA levels were associated with increased diabetes risk is consistent with previous investigations (29, 30). Consistent with our results in young adults, a recent large-scale study pooling data from 20 prospective cohorts reported that higher levels of linoleic acid, the predominant omega-6 fatty acid, has long-term benefits for the prevention of type 2 diabetes (30). These fatty acids are likely reflective of both dietary composition and endogenous metabolism (31). Dietary counseling aiming to replace saturated fat with unsaturated fat in the diet, in accordance with Nordic dietary recommendations, has been shown to decrease circulating MUFA and increase circulating omega-3 and omega-6 levels (32). If these fatty acids play a causal role in the development of type 2 diabetes, then our results suggest that interventions modifying dietary fatty acid composition could be effective in prevention.

In addition to modulations in lipoprotein metabolism, metabolic measures related to lipolysis (glycerol) and inflammation (GlycA; a marker of chronic inflammation (33, 34)) were predictive biomarkers, illustrating how many different pathways are perturbed long before onset of type 2 diabetes. The overall metabolic signature of increased diabetes risk was reminiscent of the patterns of metabolite associations for adiposity and insulin resistance index. This is keeping with previous large-scale metabolic profiling studies (17, 18, 20) and consistent with the pathophysiology of type 2 diabetes, where insulin sensitivity gradually declines years before clinical disease onset (35). The similarity of the metabolic signatures for type 2 diabetes risk to those for higher BMI and insulin levels could suggest that the metabolic biomarkers for diabetes risk are manifestations of developing insulin resistance. Nonetheless, the overall pattern of biomarker associations remained

predictive after controlling for baseline BMI and insulin levels. This indicates that comprehensive metabolic profiling is sensitive to detect subtle metabolic changes that precede insulin resistance and hyperglycemia in apparently healthy young adults.

Whereas the comprehensive signature of single-biomarkers for type 2 diabetes risk provide a picture of the numerous metabolic pathways reflective of the disease development, the measurement of multiple biomarkers in one go may prove beneficial for disease prediction. Supporting this notion, we found that a simple biomarker summary score comprised of phenylalanine and two detailed lipoprotein measures was a stronger predictor of diabetes risk than any of the individual biomarkers. The 10-fold elevation in diabetes risk observed for those in the highest compared to lowest fifth of the biomarker summary score indicates that multi-metabolite scores could potentially aid identification of high-risk individuals at young age. In this age group, the absolute risk for type 2 diabetes is low and a focus on relative risk compared to peers may be most effective for targeted prevention efforts. Future studies are needed to evaluate the potential of such scores for risk identification and health tracking in clinical settings.

Our study has both strengths and limitations. Strengths include the large sample size, and the profiling of multiple prospective cohorts. Our results were consistent across cohorts despite differences in age distribution, fasting status, and diagnostic ascertainment. The multi-cohort design allowed derivation of the biomarker summary score and validation in an independent cohort. Some limitations also need to be considered. First, because type 2 diabetes is relatively rare among young adults, the number of cases was modest despite the large sample size. The power for evaluating the predictive value of the biomarker scores was therefore limited. Second, since all cohorts were Finnish, our results cannot necessarily be generalized beyond white Europeans. However, previous research shows that amino acid measures may be even stronger predictors of type 2 diabetes in South Asians as compared to Europeans (36). Third, the NMR metabolomics platform is not able to quantify metabolites present in blood in very low concentrations, and we could therefore not replicate several previously reported metabolomic biomarkers for diabetes (8, 9, 37). Nonetheless, the NMR metabolomics method is high-throughput and consistent over time, and therefore particularly suited for large cohorts.

In conclusion, we have described a metabolic signature of increased risk for future type 2 diabetes in population-based cohorts of young adults with long follow-up. Metabolic aberrations were observed across multiple biological pathways, including inflammation, fatty acid balance, and

aspects of lipoprotein metabolism. Our results extend the evidence of amino acid alterations as strong predictors of type 2 diabetes to young adults. If branched-chain amino acids, MUFAs or omega-6 fatty acids are proven to be causal in the pathogenesis of type 2 diabetes, then interventions aimed at altering the circulating levels may be beneficial already in early adulthood. Finally, we derived and validated a biomarker summary score that stratified the risk for future type 2 diabetes beyond risk factors used in primary care settings. These results support the possibility that early screening by detailed metabolic profiling could help targeting early interventions for type 2 diabetes prevention.

References

1. NCD Risk Factor Collaboration. Worldwide trends in diabetes since 1980: a pooled analysis of 751 population-based studies with 4.4 million participants. *Lancet*. 2016;387(10027):1513-30.
2. Rao Kondapally Seshasai S, Kaptoge S, Thompson A, Di Angelantonio E, Gao P, Sarwar N, et al. Diabetes mellitus, fasting glucose, and risk of cause-specific death. *N Engl J Med*. 2011;364(9):829-41.
3. Di Angelantonio E, Kaptoge S, Wormser D, Willeit P, Butterworth AS, Bansal N, et al. Association of Cardiometabolic Multimorbidity With Mortality. *JAMA*. 2015;314(1):52-60.
4. Lindstrom J, Ilanne-Parikka P, Peltonen M, Aunola S, Eriksson JG, Hemio K, et al. Sustained reduction in the incidence of type 2 diabetes by lifestyle intervention: follow-up of the Finnish Diabetes Prevention Study. *Lancet*. 2006;368(9548):1673-9.
5. Lindstrom J, Peltonen M, Eriksson JG, Ilanne-Parikka P, Aunola S, Keinanen-Kiukaanniemi S, et al. Improved lifestyle and decreased diabetes risk over 13 years: long-term follow-up of the randomised Finnish Diabetes Prevention Study (DPS). *Diabetologia*. 2013;56(2):284-93.
6. Barry E, Roberts S, Oke J, Vijayaraghavan S, Normansell R, Greenhalgh T. Efficacy and effectiveness of screen and treat policies in prevention of type 2 diabetes: systematic review and meta-analysis of screening tests and interventions. *BMJ*. 2017;356:i6538.
7. Würtz P, Kangas AJ, Soininen P, Lawlor DA, Davey Smith G, Ala-Korpela M. Quantitative Serum Nuclear Magnetic Resonance Metabolomics in Large-Scale Epidemiology: A Primer on -Omic Technologies. *Am J Epidemiol*. 2017;186(9):1084-96.
8. Guasch-Ferre M, Hruby A, Toledo E, Clish CB, Martinez-Gonzalez MA, Salas-Salvado J, et al. Metabolomics in Prediabetes and Diabetes: A Systematic Review and Meta-analysis. *Diabetes Care*. 2016;39(5):833-46.
9. Floegel A, Stefan N, Yu Z, Muhlenbruch K, Drogan D, Joost HG, et al. Identification of serum metabolites associated with risk of type 2 diabetes using a targeted metabolomic approach. *Diabetes*. 2013;62(2):639-48.
10. Wang TJ, Larson MG, Vasan RS, Cheng S, Rhee EP, McCabe E, et al. Metabolite profiles and the risk of developing diabetes. *Nat Med*. 2011;17(4):448-53.
11. Lotta LA, Scott RA, Sharp SJ, Burgess S, Luan J, Tillin T, 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*. 2016;13(11):e1002179.
12. Wang Q, Holmes MV, Davey Smith G, Ala-Korpela M. Genetic Support for a Causal Role of Insulin Resistance on Circulating Branched-Chain Amino Acids and Inflammation. *Diabetes Care*. 2017;40(12):1779-86.
13. Pitkänen N, Juonala M, Ronnema T, Sabin MA, Hutri-Kahonen N, Kahonen M, et al. Role of Conventional Childhood Risk Factors Versus Genetic Risk in the Development of Type 2 Diabetes and Impaired Fasting Glucose in Adulthood: The Cardiovascular Risk in Young Finns Study. *Diabetes Care*. 2016;39(8):1393-9.
14. Salomaa V, Havulinna A, Saarela O, Zeller T, Jousilahti P, Jula A, et al. Thirty-one novel biomarkers as predictors for clinically incident diabetes. *PLoS One*. 2010;5(4):e10100.
15. Soininen P, Kangas AJ, Wurtz P, Suna T, Ala-Korpela M. Quantitative serum nuclear magnetic resonance metabolomics in cardiovascular epidemiology and genetics. *Circ Cardiovasc Genet*. 2015;8(1):192-206.
16. Pencina MJ, D'Agostino RB, Sr., Steyerberg EW. Extensions of net reclassification improvement calculations to measure usefulness of new biomarkers. *Stat Med*. 2011;30(1):11-21.
17. Würtz P, Wang Q, Kangas AJ, Richmond RC, Skarp J, Tiainen M, et al. Metabolic signatures of adiposity in young adults: Mendelian randomization analysis and effects of weight change. *PLoS Med*. 2014;11(12):e1001765.
18. Würtz P, Makinen VP, Soininen P, Kangas AJ, Tukiainen T, Kettunen J, et al. Metabolic signatures of insulin resistance in 7,098 young adults. *Diabetes*. 2012;61(6):1372-80.

19. Stancakova A, Civelek M, Saleem NK, Soininen P, Kangas AJ, Cederberg H, et al. Hyperglycemia and a common variant of GCKR are associated with the levels of eight amino acids in 9,369 Finnish men. *Diabetes*. 2012;61(7):1895-902.
20. Ho JE, Larson MG, Ghorbani A, Cheng S, Chen MH, Keyes M, et al. Metabolomic profiles of body mass index in the Framingham Heart Study reveal distinct cardiometabolic phenotypes. *PLoS One*. 2016;11(2):e0148361.
21. Jang C, Oh SF, Wada S, Rowe GC, Liu L, Chan MC, et al. A branched-chain amino acid metabolite drives vascular fatty acid transport and causes insulin resistance. *Nat Med*. 2016;22(4):421-6.
22. Newgard CB. Interplay between lipids and branched-chain amino acids in development of insulin resistance. *Cell Metab*. 2012;15(5):606-14.
23. Magnusson M, Lewis GD, Ericson U, Orho-Melander M, Hedblad B, Engstrom G, et al. A diabetes-predictive amino acid score and future cardiovascular disease. *Eur Heart J*. 2013;34(26):1982-9.
24. Würtz P, Havulinna AS, Soininen P, Tynkkynen T, Prieto-Merino D, Tillin T, et al. Metabolite profiling and cardiovascular event risk: a prospective study of 3 population-based cohorts. *Circulation*. 2015;131(9):774-85.
25. Batch BC, Shah SH, Newgard CB, Turer CB, Haynes C, Bain JR, et al. Branched chain amino acids are novel biomarkers for discrimination of metabolic wellness. *Metabolism*. 2013;62(7):961-9.
26. Wang J, Stancakova A, Soininen P, Kangas AJ, Paananen J, Kuusisto 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*. 2012;272(6):562-72.
27. Fizeleva M, Miilunpohja M, Kangas AJ, Soininen P, Kuusisto J, Ala-Korpela M, et al. Associations of multiple lipoprotein and apolipoprotein measures with worsening of glycemia and incident type 2 diabetes in 6607 non-diabetic Finnish men. *Atherosclerosis*. 2015;240(1):272-7.
28. Verges B. Pathophysiology of diabetic dyslipidaemia: where are we? *Diabetologia*. 2015;58(5):886-99.
29. Mahendran Y, Cederberg H, Vangipurapu J, Kangas AJ, Soininen P, Kuusisto J, et al. Glycerol and fatty acids in serum predict the development of hyperglycemia and type 2 diabetes in Finnish men. *Diabetes Care*. 2013;36(11):3732-8.
30. Wu JHY, Marklund M, Imamura F, Tintle N, Ardisson Korat AV, de Goede J, et al. Omega-6 fatty acid biomarkers and incident type 2 diabetes: pooled analysis of individual-level data for 39 740 adults from 20 prospective cohort studies. *Lancet Diabetes Endocrinol*. 2017;5(12):965-74.
31. Baylin A, Campos H. The use of fatty acid biomarkers to reflect dietary intake. *Curr Opin Lipidol*. 2006;17(1):22-7.
32. Lehtovirta M, Pahkala K, Niinikoski H, Kangas AJ, Soininen P, Lagstrom H, et al. Effect of Dietary Counseling on a Comprehensive Metabolic Profile from Childhood to Adulthood. *J Pediatr*. 2018.
33. Otvos JD, Shalaurova I, Wolak-Dinsmore J, Connelly MA, Mackey RH, Stein JH, et al. GlycA: A Composite Nuclear Magnetic Resonance Biomarker of Systemic Inflammation. *Clinical chemistry*. 2015;61(5):714-23.
34. Ritchie SC, Wurtz P, Nath AP, Abraham G, Havulinna AS, Fearnley LG, et al. The biomarker GlycA is associated with chronic inflammation and predicts long-term risk of severe infection. *Cell Systems*. 2015;1(4):293-301.
35. Tabak AG, Jokela M, Akbaraly TN, Brunner EJ, Kivimaki M, Witte DR. Trajectories of glycaemia, insulin sensitivity, and insulin secretion before diagnosis of type 2 diabetes: an analysis from the Whitehall II study. *Lancet*. 2009;373(9682):2215-21.
36. Tillin T, Hughes AD, Wang Q, Wurtz P, Ala-Korpela M, Sattar N, et al. Diabetes risk and amino acid profiles: cross-sectional and prospective analyses of ethnicity, amino acids and diabetes

in a South Asian and European cohort from the SABRE (Southall And Brent REvisited) Study. *Diabetologia*. 2015;58(5):968-79.

37. Menni C, Fauman E, Erte I, Perry JR, Kastenmuller G, Shin SY, et al. Biomarkers for type 2 diabetes and impaired fasting glucose using a nontargeted metabolomics approach. *Diabetes*. 2013;62(12):4270-6.

DISCLOSURES

LM, MK and PW are shareholders and employees of Nightingale Health Ltd, a company offering NMR based metabolic profiling. JK reports owning stock options for Nightingale Health. VS has participated in a conference trip sponsored by Novo Nordisk. No other authors reported disclosures.

SOURCES OF FUNDING

This study was supported by the Academy of Finland (grant numbers 297338 and 307247, 312476 and 312477), the Novo Nordisk Foundation (NNF17OC0026062 and 15998), Strategic Research Funding from the University of Oulu, Finland, the Sigrid Juselius Foundation, the Finnish Foundation for Cardiovascular Research, the UK Medical Research Council via the MRC University of Bristol Integrative Epidemiology Unit (MC_UU_12013/1 and MC_UU_12013/5), and the National Health and Medical Research Council of Australia (APP1158958).

The Cardiovascular Risk in Young Finns Study has been financially supported by the Academy of Finland (286284, 134309, 126925, 121584, 124282, 129378, 117787, and 41071); the Social Insurance Institution of Finland; the Social Insurance Institution of Finland; Competitive State Research Financing of the Expert Responsibility area of Kuopio, Tampere and Turku University Hospitals (X51001); Tampere University Hospital Supporting Foundation, Juho Vainio Foundation; Paavo Nurmi Foundation; Finnish Foundation for Cardiovascular Research; Finnish Cultural Foundation; The Sigrid Juselius Foundation; Tampere Tuberculosis Foundation; Emil Aaltonen Foundation; Yrjö Jahnsson Foundation; Signe and Ane Gyllenberg Foundation; Diabetes Research Foundation of Finnish Diabetes Association; and EU Horizon 2020 (755320); and European Research Council (742927). NFBC1966 received financial support from the Academy of Finland (104781, 120315, 129269, 1114194, 24300796, 85547, 285547), Biocenter Oulu (75617), University of Oulu Grant (65354), Oulu University Hospital (2/97, 8/97), Ministry of Health and Social Affairs (23/251/97, 160/97, 190/97), National Institute for Health and Welfare, Helsinki (54121), Regional Institute of Occupational Health, Oulu (50621, 54231), ERDF European Regional Development Fund Grant (539/2010 A31592), the EU H2020-PHC-2014 DynaHEALTH action (633595), and EU H2020-HCO-2004 iHEALTH Action.

Table 1. Baseline characteristics of participants in the four prospective cohorts.

	YFS	FINRISK-1997	DILGOM	NFBC
Number of individuals	2141	3063	1421	5271
Number of incident type 2 diabetes cases	65	110	18	199
Follow-up time (years)	10	15	8	15
Sex (% females)	53.8	52.7	55.7	50.2
Baseline Age (years)	31.7 ± 4.7	35.3 ± 6.0	35.7 ± 6.2	31.2±0.4
Body-mass index (kg/m ²)	25.0 ± 4.4	25.1 ± 4.2	25.6 ± 4.4	24.6 ± 4.1
Glucose (mmol/l)	5.0 ± 0.4	4.7 ± 0.6	5.6 ± 0.4	5.0 ± 0.4
Total cholesterol (mmol/l)	5.1 ± 1.0	5.1 ± 1.0	5.0 ± 0.9	5.0 ± 1.0
HDL cholesterol (mmol/l)	1.3 ± 0.3	1.4 ± 0.3	1.4 ± 0.4	1.5 ± 0.4
Triglycerides (mmol/l)	1.3 ± 0.8	1.3 ± 1.0	1.3 ± 0.9	1.2 ± 0.7
Plasma Insulin (mU/l)	7.6 ± 5.2	5.7 ± 5.5	5.6 ± 3.9	8.3 ± 3.9
Lipid-lowering medication (%)	0.3	0.3	1.3	0.1

Values are mean ± SD.

YFS, The Cardiovascular risk in Young Finns Study; DILGOM, Dietary, Lifestyle, and Genetic determinants of Obesity and Metabolic syndrome study; NFBC, Northern Finland Birth Cohort of 1966.

Table 2. Biomarker summary score for the risk of type 2 diabetes during 15-year follow-up, assessed for 5,271 individuals aged 31 at blood sampling.

Individuals divided into fifths of biomarker summary score	Number of incident type 2 diabetes cases	Model 1: Age, sex, BMI and fasting glucose	Model 2: Model 1 + triglycerides, HDL-cholesterol and insulin as additional covariates
	N (%)	Odds ratio (95% CI)	Odds ratio (95% CI)
Lower fifth	6 (0.6)	Reference	Reference
20% to 40%	14 (1.3)	2.17 (0.83-5.67)	1.95 (0.74-5.16)
40% to 60%	36 (3.4)	4.09 (2.08-12.0)	3.93 (1.59-9.70)
60% to 80%	47 (4.5)	5.92 (2.47-14.2)	4.11 (1.63-10.3)
Upper fifth	96 (9.1)	10.1 (4.20-24.1)	5.80 (2.22-15.1)
Per 1-SD increment		1.76 (1.48-2.09) P=2×10 ⁻¹⁰	1.42 (1.14-1.76) P=0.002

The biomarker summary score was calculated as the weighted sum of concentrations of three circulating metabolites: phenylalanine (weight = 0.320), free cholesterol in large HDL (weight = -0.474), and ratio of cholesteryl esters to total lipids in large VLDL (weight = -0.321). The beta-coefficients used as weights for the biomarkers score were derived by meta-analysis of three derivation cohorts.

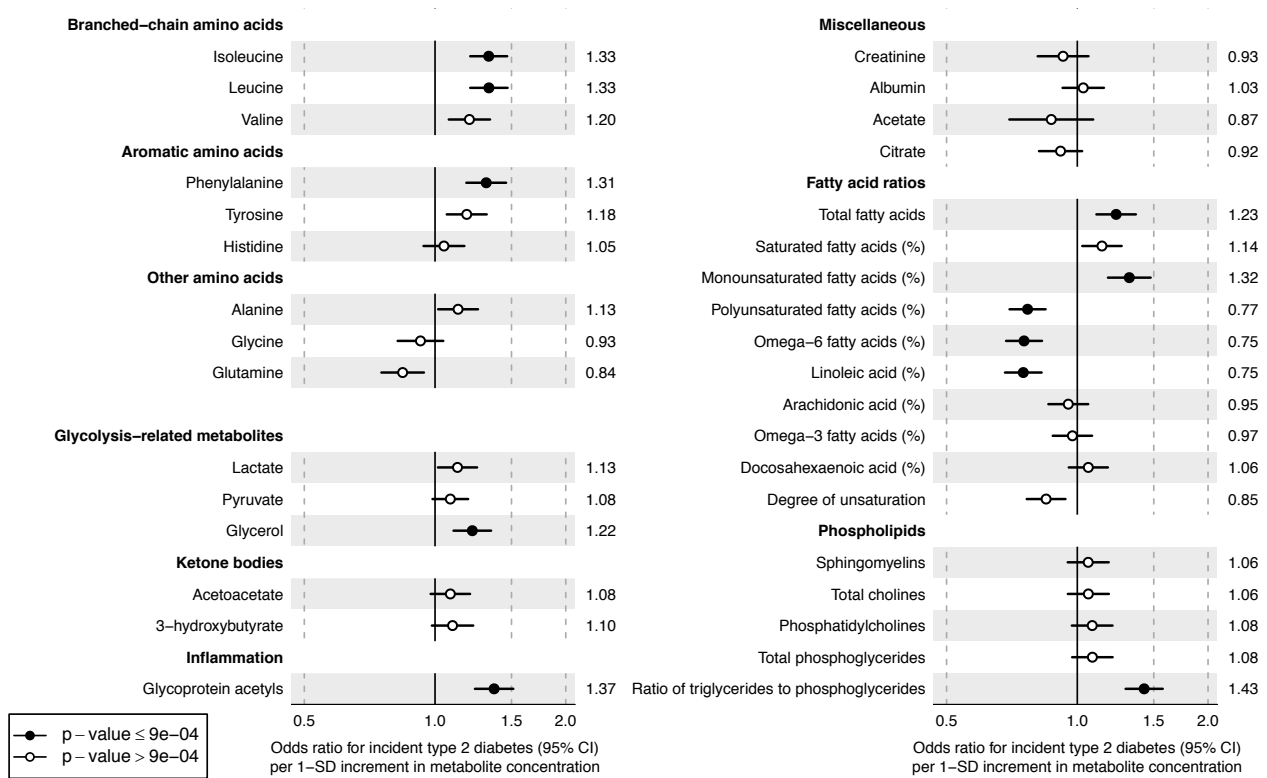


Figure 1. Relation of baseline circulating metabolite concentrations to risk of future type 2 diabetes. Values are odds ratios (95% confidence intervals) per 1-SD log-transformed metabolite concentration. Odds ratios were adjusted for sex, baseline age, body mass index and fasting glucose. The results were meta-analyzed for 11,896 young adults from four prospective cohorts.

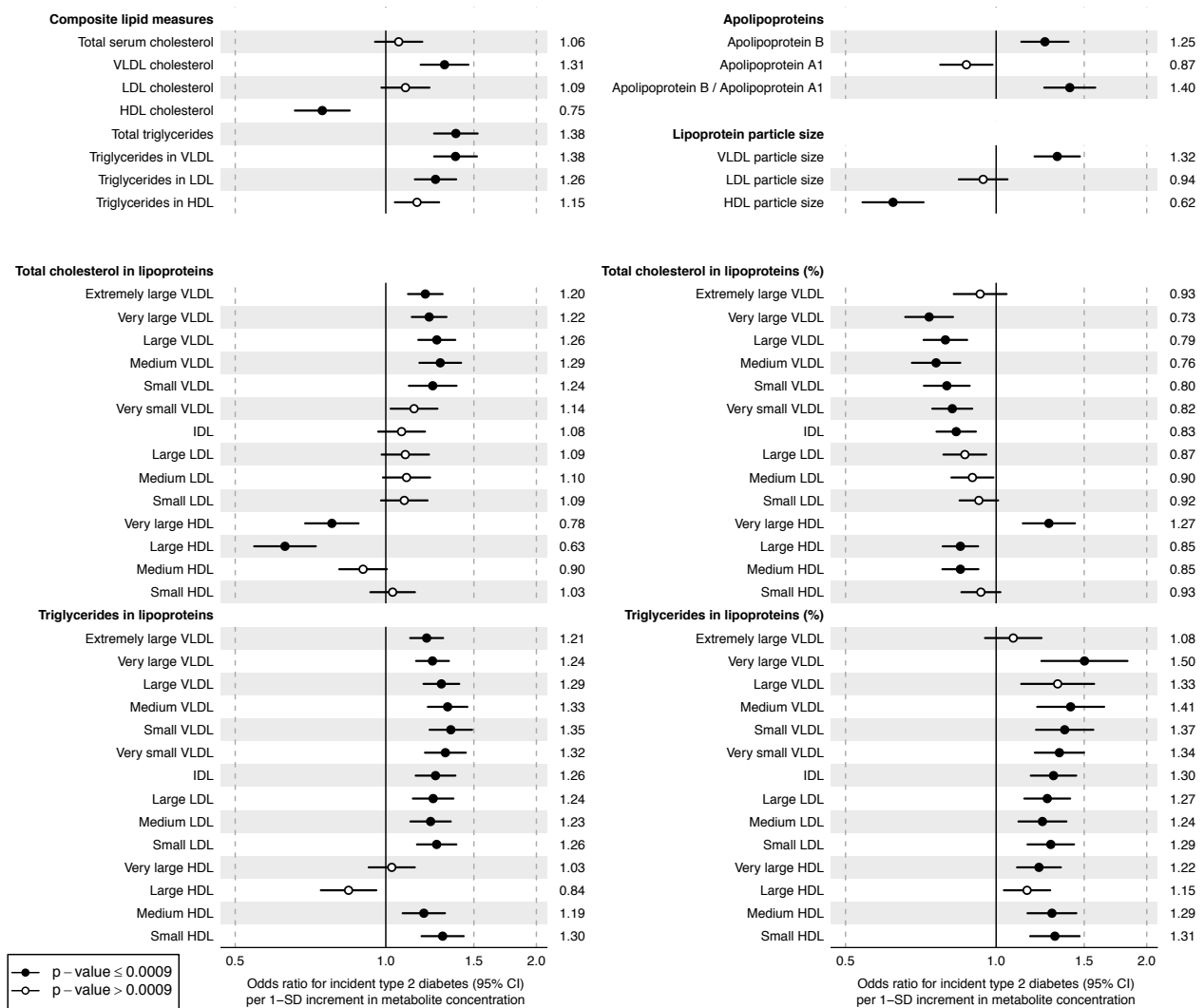


Figure 2. Relation of baseline circulating lipoprotein measures to risk of future type 2 diabetes. Values are odds ratios (95% confidence intervals) per 1-SD log-transformed metabolite concentration. Odds ratios were adjusted for sex, baseline age, body mass index and fasting glucose. The results were meta-analyzed for 11,896 young adults from four prospective cohorts. Odds ratios for the remaining 125 metabolic measures assayed are shown in Supplementary Figure 2.