



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 Puukka^{11,12,13} · Aki S. Havulinna^{3,14} · Terho Lehtimäki^{15,16} · Mika Kähönen^{15,17} · Markus Juonala¹⁸ · Sirkka Keinänen-Kiukaanniemi^{6,9,19,20} · Veikko Salomaa¹⁴ · Markus Perola^{3,14,21} · Marjo-Riitta Järvelin^{6,9,22,23,24} · Mika Ala-Korpela^{7,8,25,26,27,28} · Olli Raitakari^{1,29} · Peter Würtz^{4,30}

Received: 20 February 2019 / Accepted: 22 July 2019
© The Author(s) 2019

Abstract

Aims/hypothesis Metabolomics technologies have identified numerous blood biomarkers for type 2 diabetes risk in case–control studies of middle-aged and older individuals. We aimed to validate existing and identify novel metabolic biomarkers predictive of future diabetes in large cohorts of young adults.

Methods NMR metabolomics was used to quantify 229 circulating metabolic measures in 11,896 individuals from four Finnish observational cohorts (baseline age 24–45 years). Associations between baseline metabolites and risk of developing diabetes during 8–15 years of follow-up (392 incident cases) were adjusted for sex, age, BMI and fasting glucose. Prospective metabolite associations were also tested with fasting glucose, 2 h glucose and HOMA-IR at follow-up.

Results Out of 229 metabolic measures, 113 were associated with incident type 2 diabetes in meta-analysis of the four cohorts (ORs per 1 SD: 0.59–1.50; $p < 0.0009$). Among the strongest biomarkers of diabetes risk were branched-chain and aromatic amino acids (OR 1.31–1.33) and triacylglycerol within VLDL particles (OR 1.33–1.50), as well as linoleic *n*-6 fatty acid (OR 0.75) and non-esterified cholesterol in large HDL particles (OR 0.59). The metabolic biomarkers were more strongly associated with deterioration in post-load glucose and insulin resistance than with future fasting hyperglycaemia. A multi-metabolite score comprised of phenylalanine, non-esterified cholesterol in large HDL and the ratio of cholesteryl ester to total lipid in large VLDL was associated with future diabetes risk (OR 10.1 comparing individuals in upper vs lower fifth of the multi-metabolite score) in one of the cohorts (mean age 31 years).

Conclusions/interpretation Metabolic biomarkers across multiple molecular pathways are already predictive of the long-term risk of diabetes in young adults. Comprehensive metabolic profiling may help to target preventive interventions for young asymptomatic individuals at increased risk.

Keywords Branched-chain amino acid · Isoleucine · Leucine · Metabolomics · Type 2 diabetes

Abbreviations

DILGOM Dietary, Lifestyle, and Genetic Determinants of Obesity and Metabolic Syndrome Study

GlycA Glycoprotein acetyls
IDL Intermediate-density lipoprotein
MUFA Monounsaturated fatty acids
NFBC Northern Finland Birth Cohort
YFS Young Finns Study

Electronic supplementary material The online version of this article (<https://doi.org/10.1007/s00125-019-05001-w>) contains peer-reviewed but unedited supplementary material, which is available to authorised users.

✉ Ari V. Ahola-Olli
avahol@utu.fi

✉ Peter Würtz
peter.wurtz@nightingalehealth.com

Extended author information available on the last page of the article

Introduction

The global prevalence of type 2 diabetes is increasing rapidly, particularly in low- and middle-income countries [1]. Type 2 diabetes is associated with increased mortality risk from vascular and numerous other causes, and reduced quality of life,

Research in context

What is already known about this subject?

- Metabolomics studies have uncovered many associations of blood lipids and metabolites with increased risk for type 2 diabetes in middle-aged and older individuals
- Certain amino acids and fatty acids may reside on causal pathways contributing to the development of type 2 diabetes

What is the key question?

- Are blood metabolite biomarkers predictive of future type 2 diabetes in four cohorts of young men and women from the general population?

What are the new findings?

- Aberrations in blood concentrations of amino acids, fatty acids and lipoprotein lipids are associated with future diabetes risk in young individuals, with similar associations regardless of sex, duration of follow-up and fasting status; branched-chain amino acids are among the strongest biomarkers
- The metabolites are more strongly predictive of deterioration in glucose tolerance and insulin resistance than of worsening fasting hyperglycaemia
- A metabolite score comprising three metabolites was able to identify a subgroup with a tenfold increase in risk for incident diabetes in one of the studied cohorts (mean age 31 years)

How might this impact on clinical practice in the foreseeable future?

- These findings may eventually help target interventions to young individuals at high risk for type 2 diabetes

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 measures of hyperglycaemia and dyslipidaemia; however, these markers are ineffective for identifying high-risk individuals [6]. This has spurred interest in metabolite profiling technologies, also known as metabolomics, to identify biochemical changes occurring before the onset of diabetes to elucidate the pathophysiology and potentially aid risk prediction for better targeted prevention [7, 8].

Metabolomics is increasingly used in diabetes epidemiology [7, 8]. Multiple case–control studies have identified circulating lipids and metabolites associated with the risk for type 2 diabetes using a range of technological assays, based on MS or NMR [7, 9, 10]. Branched-chain and aromatic amino acids have been observed to be the most consistent metabolite biomarkers for type 2 diabetes [8]. Genetic evidence and experimental studies suggest that impaired metabolism of these amino acids may be causally implicated in the development of insulin resistance and type 2 diabetes [11, 12]. Also, *n*-6 and other fatty acids have emerged as robust biomarkers for future diabetes risk [8, 13, 14]. However, previous metabolomics studies have commonly involved a modest number of

participants in nested case–control settings and have almost exclusively been conducted in middle-aged and older individuals.

In this study, we aimed to assess if the metabolic biomarkers are already associated with future onset of type 2 diabetes in young adults, with blood sampling 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 years at blood draw. The high-throughput NMR platform allows us to validate many known metabolite biomarkers for diabetes and explore novel associations with detailed measures of lipoprotein metabolism. We also assessed of which hyperglycaemia measures the metabolite biomarkers were most strongly reflective, and if a multi-metabolite score would display a stronger association with early risk of type 2 diabetes than any individual metabolite biomarker.

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 electronic supplementary material (ESM) Fig. 1. Details of the individual cohorts are provided in ESM [Methods](#). All participants gave written informed consent and the studies were approved by local ethics committees. 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. The characteristics of each cohort are described in brief below.

Cardiovascular Risk in Young Finns Study In the Cardiovascular Risk in Young Finns Study (YFS), serum metabolites were quantified from 2248 individuals in the 2001 survey. The final sample consisted of 2141 individuals in the age range 24–39 years. The follow-up time was 10 years. Type 2 diabetes diagnoses at 10 year follow-up were based either on HbA_{1c} or fasting glucose assessed in the 2011 re-survey or nationwide registers of reimbursement for diabetes medication or inpatient hospital ICD-10 diagnosis of diabetes (<http://apps.who.int/classifications/icd10/browse/2016/en>; see ESM [Methods](#)) [15].

FINRISK-1997 Serum metabolites were quantified from 7603 individuals. The final sample consisted of 3063 individuals when limiting analyses to participants aged 24–45 years. The follow-up time was 15 years. Type 2 diabetes diagnoses at follow-up were based on nationwide register data [16].

Dietary, Lifestyle, and Genetic Determinants of Obesity and Metabolic Syndrome Study In the Dietary, Lifestyle, and Genetic Determinants of Obesity and Metabolic Syndrome Study (DILGOM), serum metabolites were quantified from 4816 individuals in 2007. The final sample consisted of 1421 individuals when limiting analyses to participants in the age range 25–45 years. The 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.

Northern Finland Birth Cohort In the Northern Finland Birth Cohort (NFBC) of 1966, serum metabolites were quantified from 5680 individuals in the 1997 survey. The final sample consisted of 5275 individuals aged 30–32 years. The follow-up time was 15 years. Type 2 diabetes diagnoses were based on either fasting or 2 h glucose at the 46 year follow-up conducted in 2012 or nationwide register data.

Metabolite quantification

A high-throughput NMR metabolomics platform (Nightingale Health, Helsinki, Finland) was used to quantify 229 metabolic measures from baseline serum samples [17]. This metabolite panel captures a range of established and emerging biomarkers from multiple metabolic pathways, including amino acids, glycolysis-related metabolites, fatty acids and detailed lipoprotein

lipid profiles, covering triacylglycerol, total cholesterol, non-esterified cholesterol, esterified cholesterol and phospholipids within 14 subclasses. The same experimental NMR setup and software library was used for metabolite quantification for all four cohorts. The mean levels and distributions of metabolite concentrations were coherent across the cohorts [18]. Details of the NMR metabolomics experimentation have been described previously [17] and epidemiological applications have recently been reviewed [7].

Statistical analyses

Owing to the skewness of the metabolite distributions, all metabolite concentrations were $\log_e(\text{metabolite}+1)$ transformed prior to analyses and scaled to SD concentrations separately for each cohort. Although 229 metabolic measures in total were analysed, the number of independent tests performed is lower because of the correlated nature of the measures [7]. We calculated that 54 principal components explained 99% of the variation in the metabolic measures. Alternative methods have yielded a similar number of independent tests in the NMR metabolite data [19, 20]. Hence, we inferred statistical significance at meta-analysis p value <0.0009 ($0.05/54$). The ORs of 229 circulating metabolic measures with incidence of type 2 diabetes were assessed using logistic regression. Each metabolite was analysed for association with incident diabetes in a separate model, adjusted for sex, baseline age, fasting glucose and BMI. To facilitate comparison of the magnitudes of biomarker association for measures with units and different concentration ranges, the ORs are scaled to 1 SD increments in \log_e -transformed metabolite concentration. Results from individual cohorts were combined using inverse variance-weighted fixed-effect meta-analysis. We also assessed the influence of additional adjustment for HOMA-IR index, tested results separately for men and women and compared the pattern of metabolite associations with incident type 2 diabetes with that of impaired fasting glucose (≥ 6.0 mmol/l) at follow-up.

Metabolite associations were also assessed cross-sectionally with BMI, HOMA-IR and fasting glucose using linear regression models adjusted for age and sex, and prospectively with fasting glucose, 2 h glucose, HbA_{1c} and HOMA-IR at follow-up, adjusting for sex, baseline age, fasting glucose and BMI.

Last, we examined the association with future diabetes risk using a multi-metabolite score, composed as the weighted sum of metabolite concentrations. The metabolite selection and weights in the multi-metabolite score were derived by meta-analysis of three of the cohorts (YFS, FINRISK-1997 and DILGOM, constituting approximately half of the incident cases) using forward stepwise logistic model testing of all metabolites. Age, sex, baseline fasting glucose and BMI were always included as covariates in the models for metabolite selection. 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 $p < 0.0009$ in meta-analysis of the three derivation cohorts. The multi-metabolite score was defined as the sum of concentrations of the three selected metabolites weighted by β -coefficients in the final stepwise model. This multi-metabolite score was then evaluated for association with diabetes risk in NFBC, as this cohort had the highest number of cases and most reliably ascertained diagnoses. ORs of the multi-metabolite score were assessed both as a continuous marker and by quintile, with adjustment for sex, baseline age, fasting glucose and BMI. The influence of further adjustment for HOMA-IR, triacylglycerol and HDL-cholesterol was also assessed. The risk discrimination when adding the multi-metabolite score to models containing these two sets of clinical variables were compared in terms of C-statistic, integrated discrimination improvement and continuous reclassification [21]. Statistical analyses were performed in R version 3.1.3 (R Foundation for Statistical Computing, Vienna, Austria; <https://www.R-project.org/>).

Results

The study included 11,896 individuals from four Finnish cohorts. The 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 years). 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 ESM Table 1. The ORs of 104 selected metabolic measures with incident type 2 diabetes are shown in Figs 1 and 2; results for the remaining 125 metabolic measures assayed are found in ESM Fig. 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 ORs ranging from 1.18 to 1.50 for direct associations and from 0.59 to 0.86 for inverse associations per 1 SD metabolite concentration.

Amino acids, glycolysis and inflammation

The branched-chain amino acids isoleucine, leucine and valine (ORs 1.20–1.33) and the aromatic amino acids phenylalanine and tyrosine (ORs 1.31 [95% CI 1.18, 1.46] and 1.18 [95% CI 1.06, 1.32], respectively) were associated with the risk of type 2 diabetes (Fig. 1). Glycerol was also associated with increased risk (OR 1.22 [95% CI 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 (OR 1.37 [95% CI 1.24, 1.51]).

Fatty acids

The total concentration of circulating fatty acids (OR 1.23 [95% CI 1.11, 1.36]) and the relative amount of monounsaturated fatty acids ([MUFA] ratio to total fatty acids) were directly associated with increased risk for type 2 diabetes (OR 1.32 [95% CI 1.18–1.48]). In contrast, higher relative concentrations of *n*-6 fatty acids were associated with decreased risk for type 2 diabetes (OR 0.75 [95% CI 0.69, 0.83]). This inverse association was primarily driven by linoleic acid, whereas the association for arachidonic acid was weaker.

Table 1 Baseline characteristics of participants in the four prospective cohorts

Characteristic	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 (% women)	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
BMI (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
Triacylglycerol (mmol/l)	1.3 ± 0.8	1.3 ± 1.0	1.3 ± 0.9	1.2 ± 0.7
Plasma insulin (pmol/l)	52.8 ± 36.1	39.6 ± 38.2	38.9 ± 27.1	57.6 ± 27.1
Lipid-lowering medication (%)	0.3	0.3	1.3	0.1

Values are mean ± SD

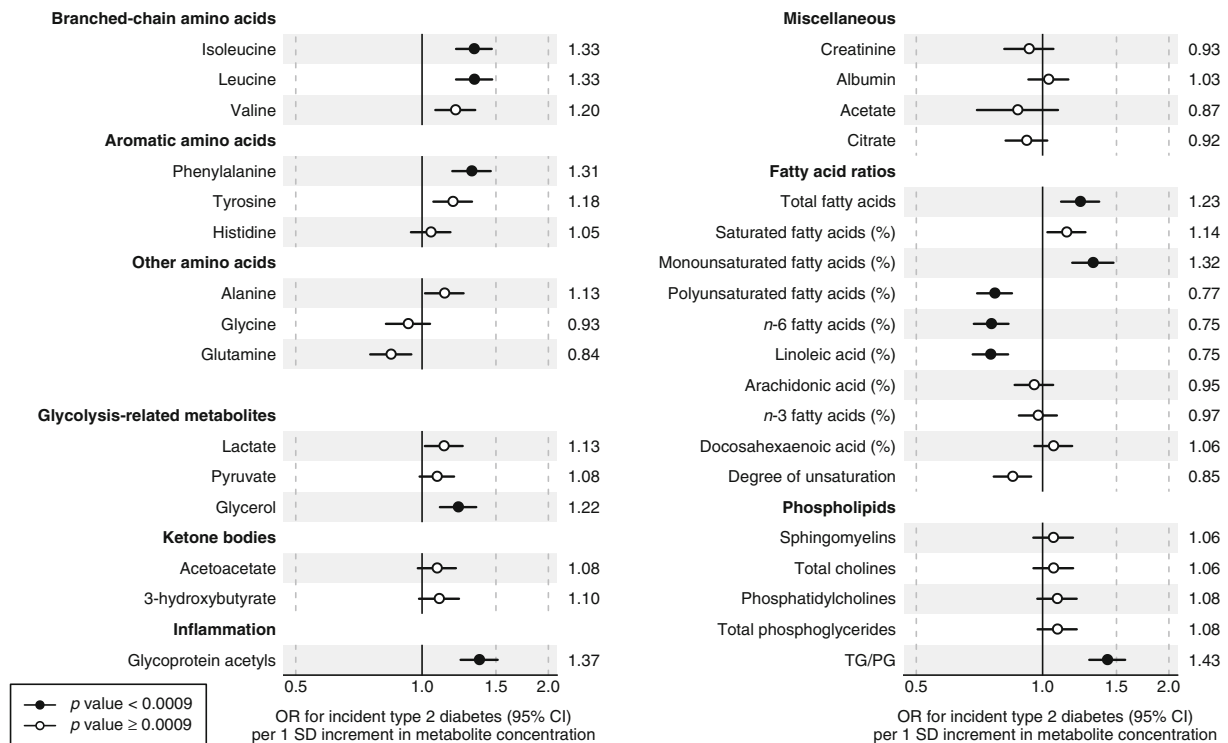


Fig. 1 Relationship between baseline circulating metabolite concentrations and risk of future type 2 diabetes. Values are ORs (95% CIs) per 1 SD \log_{10} -transformed metabolite concentration. ORs were adjusted for

Lipoprotein measures

Both lipid measures used in routine clinical settings and more fine-grained lipoprotein subclass measures were quantified by the NMR metabolomics platform. The associations of routine lipids, as well as cholesterol and triacylglycerol concentrations in 14 lipoprotein subclasses, with type 2 diabetes risk are shown in Fig. 2. Additional lipoprotein subclass measures are shown in ESM Fig. 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. Cholesterol in very large and large HDL particles was particularly strongly associated with decreased diabetes risk. The association patterns were similar for non-esterified cholesterol and cholesteryl esters; the strongest biomarker for decreased diabetes risk was non-esterified cholesterol in large HDL (OR 0.59 [95% CI 0.50, 0.68]; ESM Fig. 2). However, this pattern of lipoprotein lipid association was different for triacylglycerols: increased triacylglycerol concentrations in all VLDL, intermediate-density lipoprotein (IDL) and LDL as well as medium-sized and small HDL subclasses were strongly associated with increased type 2 diabetes risk. The prominent importance of triacylglycerols was also evident when examining the associations for the relative fraction of triacylglycerol in each lipoprotein subclass, i.e. the percentage of triacylglycerol per total lipid concentration in a given size

sex, baseline age, BMI and fasting glucose. The results were meta-analysed for 11,896 young adults from four prospective cohorts. PG, phosphoglyceride; TG, triacylglycerol

of lipoprotein particle: a higher relative abundance of triacylglycerols within lipoprotein particles was strongly associated with higher diabetes risk (Fig. 2). Because a higher relative triacylglycerol content in lipoprotein particles generally reflects a lower cholesterol content, then the relative fraction of cholesterol in most lipoprotein subclasses was inversely associated with future diabetes risk.

Concentration of apolipoproteins, the structural proteins of lipoprotein particles, was also associated with increased risk for type 2 diabetes. In particular, the ratio of apolipoprotein B to apolipoprotein A1 was among the strongest predictors (OR 1.40 [95% CI 1.25, 1.58]). Further, larger VLDL particle size was associated with increased diabetes risk (OR 1.32 [95% CI 1.19, 1.47]), whereas larger HDL particle size displayed an inverse association (OR 0.62 [95% CI 0.54, 0.72]).

Consistency across cohorts and influence of adjustment for insulin resistance

The patterns of association 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 (ESM Fig. 3). The metabolite associations were highly similar for men and women (ESM Fig. 4). Most associations between metabolites and future risk of type 2 diabetes were moderately attenuated when including HOMA-IR as covariate, but the overall

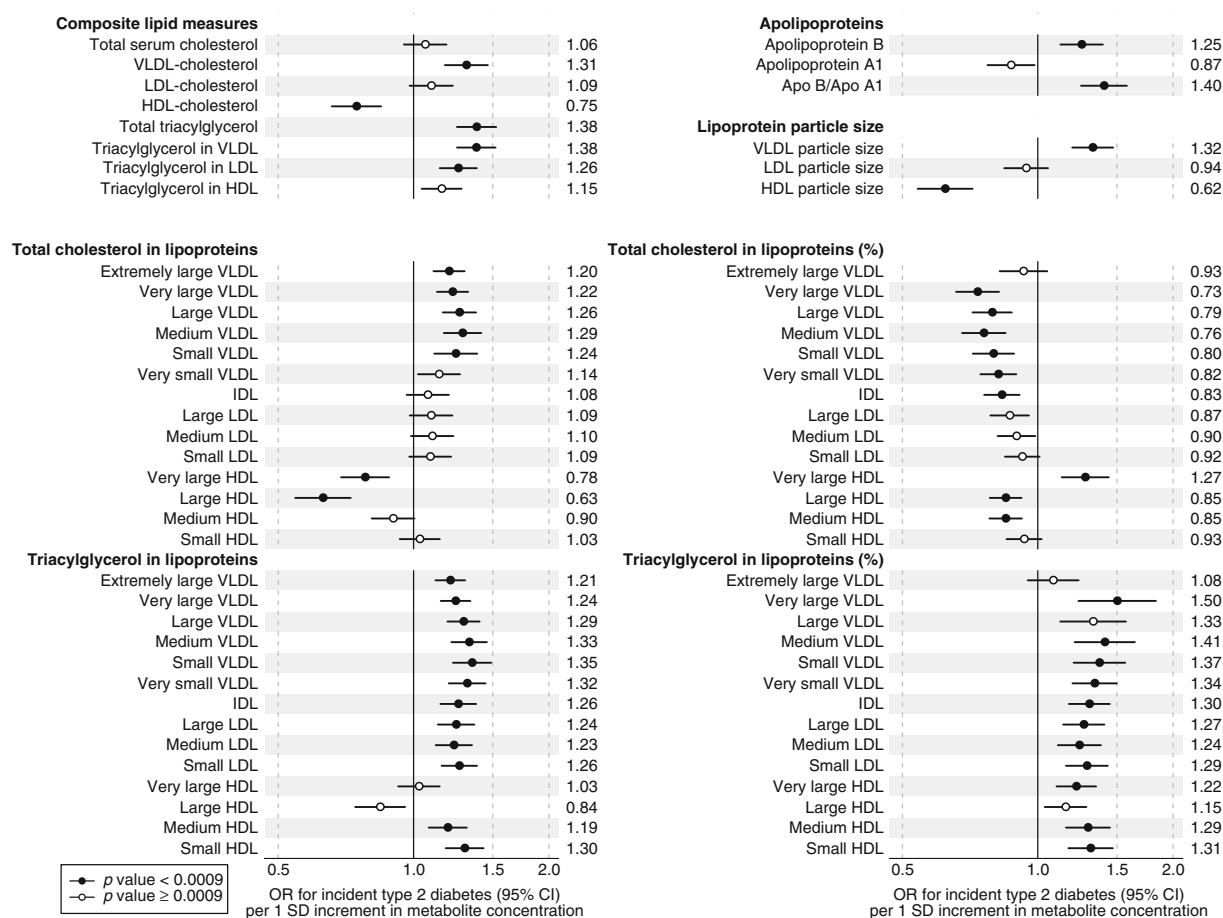


Fig. 2 Relationship between baseline circulating lipoprotein measures and risk of future type 2 diabetes. Values are ORs (95% CIs) per 1 SD log_e-transformed metabolite concentration. ORs were adjusted for sex, baseline age, BMI and fasting glucose. The results were meta-analysed

pattern persisted and 71 of the metabolic measures remained significant at $p < 0.0009$ (ESM Fig. 5). Results were almost identical if random-effects rather than fixed-effects were used in meta-analyses and if time-to-event Cox models were used instead of logistic regression (ESM Table 2).

Prospective metabolite associations with measures of hyperglycaemia

To clarify the aspects of hyperglycaemia reflected most closely by the observed metabolic aberrations, we examined the metabolite associations with fasting glucose, 2 h glucose and HOMA-IR measured in the follow-up surveys 8–15 years after the baseline (Fig. 3). The overall pattern of metabolite associations was similar for the three continuous measures of blood glucose; however, the magnitudes of associations were, on average, 2.2-fold stronger for HOMA-IR and 1.7-fold stronger for 2 h glucose compared with association magnitudes for fasting glucose (ESM Fig. 6). Consistently, the ORs were almost twice as strong for metabolite associations with incident type 2 diabetes compared with incident impaired

for 11,896 young adults from four prospective cohorts. ORs for the remaining 125 metabolic measures assayed are shown in ESM Fig. 2. ApoA1, apolipoprotein A1; ApoB, apolipoprotein B

fasting glucose (≥ 6.0 mmol/l at follow-up; ESM Fig. 7). In line with these prospective analyses, we found that the metabolite associations were strongly associated with HOMA-IR and BMI as assessed cross-sectionally, whereas the associations with fasting glucose at baseline were substantially weaker in these young adults (ESM Fig. 8).

Multi-metabolite score strongly associates with future diabetes

To examine if a combination of metabolites would be more strongly associated with diabetes risk than any individual metabolite biomarker, we derived a multi-metabolite score. The weights for adding up the metabolite concentrations in the multi-metabolite score were derived using a stepwise modelling approach based on three of the cohorts. In this manner, three metabolic measures were selected as independent predictors of diabetes: phenylalanine, non-esterified cholesterol in large HDL and cholesteryl ester to total lipid ratio within large VLDL. The association of this multi-metabolite score was then evaluated separately in the NFBC study: the multi-

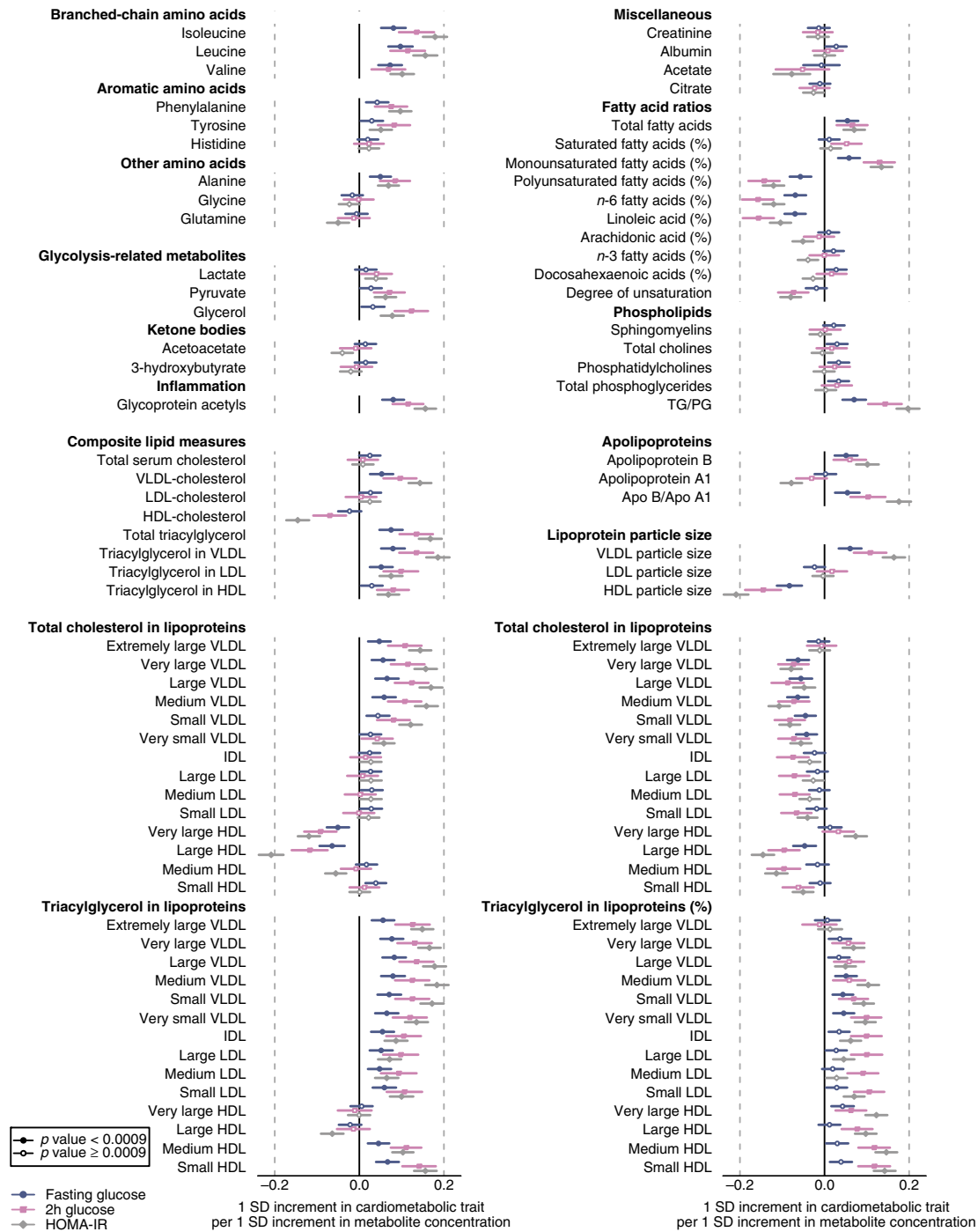


Fig. 3 Relationship between baseline circulating metabolites and lipids to blood glucose measures at follow-up. The prospective associations were assessed for fasting glucose ($n = 5017$), 2 h glucose ($n = 3028$) and HOMA-IR ($n = 5010$). Values are β -coefficients (95% CIs) scaled to 1

SD in each of the measures of blood glucose per 1 SD \log_e -transformed metabolite concentration. Associations were adjusted for sex, baseline age, BMI and fasting glucose. ApoA1, apolipoprotein A1; ApoB, apolipoprotein B; PG, phosphoglyceride; TG, triacylglycerol

metabolite score was more strongly associated with incident type 2 diabetes than any individual metabolite measure (OR 1.76 [95% CI 1.48, 2.09] per SD)). When dividing individuals based on quintiles of their multi-metabolite score, the OR was 10.1 (95% CI 4.2, 24.1) among individuals in the upper fifth compared with those in the lower fifth when adjusting for age,

sex and baseline glucose and BMI (Table 2). If further adjusting for baseline HOMA-IR, triacylglycerol and HDL-cholesterol, the OR for the highest vs lowest fifth of the multi-metabolite score was attenuated to 5.79 (95% CI 2.22, 15.1). The discrimination in absolute risk for future type 2 diabetes is presented in ESM Tables 3–5 and ESM Fig. 9.

Table 2 Multi-metabolite score for the risk of type 2 diabetes during the 15 year follow-up, assessed for 5271 individuals aged 31 years at blood sampling

	Incident type 2 diabetes cases ^a , <i>n</i> (%)	OR (95% CI)	
		Model 1 ^b	Model 2 ^c
Score quintile			
Lower fifth	6 (0.6)	Reference	Reference
20–40%	14 (1.3)	2.17 (0.83, 5.67)	1.95 (0.74, 5.16)
40–60%	36 (3.4)	4.09 (2.08, 12.0)	3.93 (1.59, 9.70)
60–80%	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) ^d	1.42 (1.14, 1.76) ^e

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

^a The lower fifth quantile contains 1055 individuals and the other quintiles 1054 individuals

^b With age, sex, BMI and fasting glucose as covariates

^c Model 1 + triacylglycerol, HDL-cholesterol and HOMA-IR as additional covariates

^d $p = 2 \times 10^{-10}$

^e $p = 0.002$

Discussion

This large multi-cohort study describes the metabolic signature of increased type 2 diabetes risk in young adults up to 15 years prior to disease onset. Metabolic aberrations related to incident type 2 diabetes spanned amino acids, fatty acid balance, inflammation and detailed lipoprotein particle composition, with consistent results across the four cohorts. Many of these metabolic measures have previously been associated with future diabetes in middle-aged and older individuals. Among the strongest biomarkers were higher concentrations of branched-chained and aromatic amino acids, VLDL particle measures and the enrichment of triacylglycerol in all lipoprotein subclasses. Moreover, higher circulating levels of GlycA, glycerol and MUFA were also associated with increased risk for type 2 diabetes, whereas glutamine, linoleic acid, HDL particle size and certain lipid measures within large HDL were associated with lower risk. These metabolic aberrations were more strongly predictive of deterioration of insulin sensitivity and impaired post-load glucose levels over long-term follow-up than worsening of fasting hyperglycaemia. A multi-metabolite score consisting of three metabolic measures was associated with a tenfold elevation in the long-term risk for type 2 diabetes in one of the cohorts, comprising 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 future diabetes has since been replicated in several epidemiological studies [8, 9, 22]

and extended to insulin resistance and blood glucose [9, 23, 24]. The ORs 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 already associate with the long-term risk of type 2 diabetes in young adults. Our results also show that the perturbed amino acid levels are more strongly indicative of future impaired glucose tolerance and insulin resistance than of worsening in fasting glucose levels.

The mechanistic underpinnings and causal relation between amino acid levels and type 2 diabetes risk are not yet fully clear [25]. Mendelian randomisation studies have indicated that adiposity and insulin resistance lead to increased branched-chain amino acid levels [12, 26]; other Mendelian randomisation studies suggest that the metabolism of these amino acids may play a causal role in the development of type 2 diabetes [11]. In addition, physiological studies have suggested mechanisms by which alterations in branched-chain amino acid metabolism might cause insulin resistance and impairment of insulin secretion [27, 28]. Altered amino acid metabolism may also represent a link between diabetes and cardiovascular diseases [29, 30]. Our results in young adults support the notion that amino acid profiling may prove helpful for monitoring cardiometabolic health in asymptomatic individuals, with the potential to facilitate targeted interventions [31].

Increasing evidence suggests that levels of certain fatty acids are associated with type 2 diabetes risk. Our finding that a higher relative concentration of *n*-6 fatty acids was associated with decreased diabetes risk, whereas higher MUFA levels were associated with increased diabetes risk is

consistent with previous investigations [13, 14]. Consistent with our results in young adults, a recent study from 20 prospective cohorts reported that higher levels of linoleic acid in serum and different lipid compartments is associated with lower risk of type 2 diabetes [14]. The circulating fatty acid biomarkers are reflective of both dietary composition and endogenous metabolism [32]. Dietary counselling 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 *n*-3 and *n*-6 levels [33]. If these fatty acids play a causal role in the development of type 2 diabetes, then our results suggest that interventions modifying the circulating fatty acid composition could be effective in prevention.

Pervasive alterations in the lipoprotein profile were also found to be associated with future diabetes risk. These included both established lipids and novel findings based on detailed lipoprotein subclass measures. The lipid modulations shown here to reflect diabetes risk in young adults are similar to those previously reported in cross-sectional settings for older individuals with impaired glucose tolerance [34, 35]. The results for VLDL and HDL particle size are consistent with a large study of American women [36]. In addition, we report novel associations of lipoprotein composition, showing increased risk associated with a higher relative fraction of triacylglycerol in VLDL, LDL as well as HDL. Higher percentage triacylglycerol in VLDL subclasses was associated with the strongest increase in diabetes risk among all metabolic measures assayed. These results 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 triacylglycerol to HDL and LDL particles [37]. Consistent with this, we showed that the lipoprotein lipid perturbations were strongly reflective of future insulin resistance and impaired glucose tolerance. Our findings indicate that such distortions of lipoprotein metabolism may already be present in normoglycaemic young adults and reflect an increased risk for insulin resistance and type 2 diabetes.

In addition to modulations in lipoprotein metabolism, metabolic measures related to lipolysis (glycerol) and inflammation (GlycA, a marker of chronic inflammation [38, 39]) were predictive biomarkers, illustrating that many different pathways are perturbed long before the 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, cross-sectionally and prospectively. This is keeping with previous large-scale metabolic profiling studies [23, 24, 26] and consistent with the pathophysiology of type 2 diabetes, where insulin sensitivity gradually declines years before clinical disease onset [40]. It suggests that the metabolic biomarkers for type 2 diabetes are predominantly manifestations of developing insulin

resistance. Nonetheless, the overall pattern of biomarker associations remained predictive after controlling for baseline BMI and HOMA-IR. These results indicate that metabolomic profiling is sensitive to subtle metabolic changes that precede insulin resistance and hyperglycaemia in apparently healthy young adults.

Whereas the comprehensive signature of single biomarkers for type 2 diabetes risk provides 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. We found that a simple multi-metabolite score comprised of phenylalanine and two detailed lipoprotein measures was a stronger predictor of diabetes risk than any of the individual biomarkers. The tenfold elevation in diabetes risk observed here for those in the highest fifth compared with the lowest fifth of the multi-metabolite score indicates that multi-metabolite scores hold potential to aid identification of high-risk individuals at a young age. Future studies with a larger number of incident diabetes cases 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. Its 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 study design allowed derivation and validation of the multi-metabolite score in independent cohorts. 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 multi-metabolite score was therefore limited. Second, as all cohorts were Finnish, our results cannot necessarily be generalised beyond white Europeans. However, previous research shows that amino acid measures may be even stronger predictors of type 2 diabetes in South Asians compared with Europeans [41]. Third, the NMR metabolomics platform is not able to quantify metabolites present in blood in very low concentrations, and therefore we could not replicate several previously reported metabolomic biomarkers for diabetes [8, 9, 42]. Nonetheless, the NMR metabolomics method is high-throughput and consistent over time, and therefore it is particularly suited for large cohorts. We acknowledge the lack of coherent dietary information across the cohorts and that a large fraction of the samples were non-fasting; however, we observed highly consistent biomarker associations between cohorts with fasting samples and the FINRISK 1997 cohort with non-fasting samples.

In conclusion, we have described a metabolic signature of increased risk for future type 2 diabetes in large 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 *n*-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 in early adulthood. The detailed metabolic profiling was shown to capture aspects of the development of insulin resistance and post-load hyperglycaemia, which are missed by fasting glucose and other risk markers used in primary care settings. These results support the possibility that screening aided by detailed metabolic profiling could help targeting interventions for type 2 diabetes prevention in young adults.

Acknowledgements Open access funding provided by University of Turku (UTU) including Turku University Central Hospital.

Data availability The individual-level data from DILGOM and FINRISK are available from THL Biobank (<https://thl.fi/en/web/thl-biobank>). Individual-level data for NFBC 1966 can be applied for from the University of Oulu Infrastructure for Population Studies (www.oulu.fi/nfbc/) and for YFS by request to the University of Turku Research Centre of Applied and Preventive Cardiovascular Medicine (youngfinnsstudy.utu.fi).

Funding This study was supported by the Academy of Finland (grant numbers 297338, 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 (MRC) 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 YFS 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; Sigrid Juselius Foundation; Tampere Tuberculosis Foundation; Emil Aaltonen Foundation; Yrjö Jahnsson Foundation; Signe and Ane Gyllenberg Foundation; Diabetes Research Foundation of Finnish Diabetes Association; EU Horizon 2020 (755320); and European Research Council (742927). NFBC 1966 received financial support from: Academy of Finland (104781, 120315, 129269, 1114194, 24300796, 85547 and 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.

Duality of interest LM, MKa and PW are shareholders and employees of Nightingale Health, 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. All other authors declare that there is no duality of interest associated with their contribution to this manuscript.

Contribution statement AVA-O and PW designed the study. AVA-O, MKa and PW conducted statistical analyses. AVA-O, LM and PW drafted the manuscript and contributed to interpretation of data. JK, JJ, JA, KP, ASH, TL, MKä, MJ, SK-K, VS, MP, M-RJ, MA-K and OR collected and contributed data and analysis tools, contributed to the discussion and reviewed/edited the manuscript. PW is the guarantor of this work and, as such, had full access to all of the data in the study and takes responsibility for the integrity of the data and the accuracy of the data analysis. All authors approved the final content of the manuscript.



Open Access This article is distributed under the terms of the Creative Commons Attribution 4.0 International License (<http://creativecommons.org/licenses/by/4.0/>), which permits unrestricted use, distribution, and reproduction in any medium, provided you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license, and indicate if changes were made.

References

1. NCD Risk Factor Collaboration (NCD-RisC) (2016) Worldwide trends in diabetes since 1980: a pooled analysis of 751 population-based studies with 4.4 million participants. *Lancet* 387(10027):1513–1530. [https://doi.org/10.1016/S0140-6736\(16\)00618-8](https://doi.org/10.1016/S0140-6736(16)00618-8)
2. Emerging Risk Factors Collaboration (2011) Diabetes mellitus, fasting glucose, and risk of cause-specific death. *N Engl J Med* 364(9):829–841. <https://doi.org/10.1056/NEJMoa1008862>
3. Emerging Risk Factors Collaboration (2015) Association of cardio-metabolic multimorbidity with mortality. *JAMA* 314(1):52–60. <https://doi.org/10.1001/jama.2015.7008>
4. Lindström J, Ilanne-Parikka P, Peltonen M et al (2006) Sustained reduction in the incidence of type 2 diabetes by lifestyle intervention: follow-up of the Finnish Diabetes Prevention Study. *Lancet* 368(9548):1673–1679. [https://doi.org/10.1016/S0140-6736\(06\)69701-8](https://doi.org/10.1016/S0140-6736(06)69701-8)
5. Lindström J, Peltonen M, Eriksson JG et al (2013) Improved lifestyle and decreased diabetes risk over 13 years: long-term follow-up of the randomised Finnish Diabetes Prevention Study (DPS). *Diabetologia* 56(2):284–293. <https://doi.org/10.1007/s00125-012-2752-5>
6. Barry E, Roberts S, Oke J, Vijayaraghavan S, Normansell R, Greenhalgh T (2017) 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* 356: i6538. <https://doi.org/10.1136/bmj.i6538>
7. Würtz P, Kangas AJ, Soininen P, Lawlor DA, Davey Smith G, Ala-Korpela M (2017) Quantitative serum nuclear magnetic resonance metabolomics in large-scale epidemiology: a primer on -omic technologies. *Am J Epidemiol* 186(9):1084–1096. <https://doi.org/10.1093/aje/kwx016>
8. Guasch-Ferré M, Hruby A, Toledo E et al (2016) Metabolomics in prediabetes and diabetes: a systematic review and meta-analysis. *Diabetes Care* 39(5):833–846. <https://doi.org/10.2337/dc15-2251>
9. Floegel A, Stefan N, Yu Z et al (2013) Identification of serum metabolites associated with risk of type 2 diabetes using a targeted metabolomic approach. *Diabetes* 62(2):639–648. <https://doi.org/10.2337/db12-0495>
10. Wang TJ, Larson MG, Vasan RS et al (2011) Metabolite profiles and the risk of developing diabetes. *Nat Med* 17(4):448–453. <https://doi.org/10.1038/nm.2307>
11. Lotta LA, Scott RA, Sharp SJ et al (2016) 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):e1002179. <https://doi.org/10.1371/journal.pmed.1002179>
12. Wang Q, Holmes MV, Davey Smith G, Ala-Korpela M (2017) Genetic support for a causal role of insulin resistance on circulating branched-chain amino acids and inflammation. *Diabetes Care* 40(12):1779–1786. <https://doi.org/10.2337/dc17-1642>
 13. Mahendran Y, Cederberg H, Vangipurapu J et al (2013) Glycerol and fatty acids in serum predict the development of hyperglycemia and type 2 diabetes in Finnish men. *Diabetes Care* 36(11):3732–3738. <https://doi.org/10.2337/dc13-0800>
 14. Wu JHY, Marklund M, Imamura F et al (2017) Omega-6 fatty acid biomarkers and incident type 2 diabetes: pooled analysis of individual-level data for 39740 adults from 20 prospective cohort studies. *Lancet Diabetes Endocrinol* 5(12):965–974. [https://doi.org/10.1016/S2213-8587\(17\)30307-8](https://doi.org/10.1016/S2213-8587(17)30307-8)
 15. Pitkänen N, Juonala M, Rönnemaa T et al (2016) 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* 39(8):1393–1399. <https://doi.org/10.2337/dc16-0167>
 16. Salomaa V, Havulinna A, Saarela O et al (2010) Thirty-one novel biomarkers as predictors for clinically incident diabetes. *PLoS One* 5(4):e10100. <https://doi.org/10.1371/journal.pone.0010100>
 17. Soininen P, Kangas AJ, Würtz P, Suna T, Ala-Korpela M (2015) Quantitative serum nuclear magnetic resonance metabolomics in cardiovascular epidemiology and genetics. *Circ Cardiovasc Genet* 8(1):192–206. <https://doi.org/10.1161/CIRCGENETICS.114.000216>
 18. Wang Q, Würtz P, Auro K et al (2016) Metabolic profiling of pregnancy: cross-sectional and longitudinal evidence. *BMC Med* 14(1):205. <https://doi.org/10.1186/s12916-016-0733-0>
 19. Barrios C, Zierer J, Würtz P et al (2018) Circulating metabolic biomarkers of renal function in diabetic and non-diabetic populations. *Sci Rep* 8(1):15249. <https://doi.org/10.1038/s41598-018-33507-7>
 20. Zheng J, Richardson TG, Millard LAC et al (2018) PhenoSpD: an integrated toolkit for phenotypic correlation estimation and multiple testing correction using GWAS summary statistics. *Gigascience* 7(8). <https://doi.org/10.1093/gigascience/giy090>
 21. Pencina MJ, D'Agostino RB, Steyerberg EW (2011) Extensions of net reclassification improvement calculations to measure usefulness of new biomarkers. *Stat Med* 30(1):11–21. <https://doi.org/10.1002/sim.4085>
 22. Stancakova A, Civelek M, Saleem NK et al (2012) Hyperglycemia and a common variant of GCKR are associated with the levels of eight amino acids in 9,369 Finnish men. *Diabetes* 61(7):1895–1902. <https://doi.org/10.2337/db11-1378>
 23. Würtz P, Makinen V-P, Soininen P et al (2012) Metabolic signatures of insulin resistance in 7,098 young adults. *Diabetes* 61(6):1372–1380. <https://doi.org/10.2337/db11-1355>
 24. Ho JE, Larson MG, Ghorbani A et al (2016) Metabolomic profiles of body mass index in the Framingham Heart Study reveal distinct cardiometabolic phenotypes. *PLoS One* 11(2):e0148361. <https://doi.org/10.1371/journal.pone.0148361>
 25. White PJ, Newgard CB (2019) Branched-chain amino acids in disease. *Science* 363(6427):582–583. <https://doi.org/10.1126/science.aav0558>
 26. Würtz P, Wang Q, Kangas AJ et al (2014) Metabolic signatures of adiposity in young adults: Mendelian randomization analysis and effects of weight change. *PLoS Med* 11(12):e1001765. <https://doi.org/10.1371/journal.pmed.1001765>
 27. Jang C, Oh SF, Wada S et al (2016) A branched-chain amino acid metabolite drives vascular fatty acid transport and causes insulin resistance. *Nat Med* 22(4):421–426. <https://doi.org/10.1038/nm.4057>
 28. Newgard CB (2012) Interplay between lipids and branched-chain amino acids in development of insulin resistance. *Cell Metab* 15(5):606–614. <https://doi.org/10.1016/j.cmet.2012.01.024>
 29. Magnusson M, Lewis GD, Ericson U et al (2013) A diabetes-predictive amino acid score and future cardiovascular disease. *Eur Heart J* 34(26):1982–1989. <https://doi.org/10.1093/eurheartj/ehs424>
 30. Wurtz P, Havulinna AS, Soininen P et al (2015) Metabolite profiling and cardiovascular event risk: a prospective study of 3 population-based cohorts. *Circulation* 131(9):774–785. <https://doi.org/10.1161/CIRCULATIONAHA.114.013116>
 31. Batch BC, Shah SH, Newgard CB et al (2013) Branched chain amino acids are novel biomarkers for discrimination of metabolic wellness. *Metabolism* 62(7):961–969. <https://doi.org/10.1016/j.metabol.2013.01.007>
 32. Baylin A, Campos H (2006) The use of fatty acid biomarkers to reflect dietary intake. *Curr Opin Lipidol* 17(1):22–27. <https://doi.org/10.1097/01.mol.0000199814.46720.83>
 33. Lehtovirta M, Pakkala K, Niinikoski H et al (2018) Effect of dietary counseling on a comprehensive metabolic profile from childhood to adulthood. *J Pediatr* 195:190–198. <https://doi.org/10.1016/j.jpeds.2017.11.057>
 34. Wang J, Stančáková A, Soininen P et al (2012) Lipoprotein subclass profiles in individuals with varying degrees of glucose tolerance: a population-based study of 9399 Finnish men. *J Intern Med* 272(6):562–572. <https://doi.org/10.1111/j.1365-2796.2012.02562.x>
 35. Fizelova M, Miilunpohja M, Kangas AJ et al (2015) Associations of multiple lipoprotein and apolipoprotein measures with worsening of glycemia and incident type 2 diabetes in 6607 non-diabetic Finnish men. *Atherosclerosis* 240(1):272–277. <https://doi.org/10.1016/j.atherosclerosis.2015.03.034>
 36. Mora S, Otvos JD, Rosenson RS, Pradhan A, Buring JE, Ridker PM (2010) Lipoprotein particle size and concentration by nuclear magnetic resonance and incident type 2 diabetes in women. *Diabetes* 59(5):1153–1160. <https://doi.org/10.2337/db09-1114>
 37. Vergès B (2015) Pathophysiology of diabetic dyslipidaemia: where are we? *Diabetologia* 58(5):886–899. <https://doi.org/10.1007/s00125-015-3525-8>
 38. Otvos JD, Shalaurava I, Wolak-Dinsmore J et al (2015) GlycA: a composite nuclear magnetic resonance biomarker of systemic inflammation. *Clin Chem* 61(5):714–723. <https://doi.org/10.1373/clinchem.2014.232918>
 39. Ritchie SC, Würtz P, Nath AP et al (2015) The biomarker GlycA is associated with chronic inflammation and predicts long-term risk of severe infection. *Cell Syst* 1(4):293–301. <https://doi.org/10.1016/j.cels.2015.09.007>
 40. Tabák AG, Jokela M, Akbaraly TN, Brunner EJ, Kivimäki M, Witte DR (2009) Trajectories of glycaemia, insulin sensitivity, and insulin secretion before diagnosis of type 2 diabetes: an analysis from the Whitehall II study. *Lancet* 373(9682):2215–2221. [https://doi.org/10.1016/S0140-6736\(09\)60619-X](https://doi.org/10.1016/S0140-6736(09)60619-X)
 41. Tillin T, Hughes AD, Wang Q et al (2015) 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* 58(5):968–979. <https://doi.org/10.1007/s00125-015-3517-8>
 42. Menni C, Fauman E, Erte I et al (2013) Biomarkers for type 2 diabetes and impaired fasting glucose using a nontargeted metabolomics approach. *Diabetes* 62(12):4270–4276. <https://doi.org/10.2337/db13-0570>

Affiliations

Ari V. Ahola-Olli^{1,2,3}  · **Linda Mustelin**^{3,4,5} · **Maria Kalimeri**⁴ · **Johannes Kettunen**^{6,7,8} · **Jari Jokelainen**^{6,9} · **Juha Auvinen**^{6,9,10} · **Katri Puukka**^{11,12,13} · **Aki S. Havulinna**^{3,14} · **Terho Lehtimäki**^{15,16} · **Mika Kähönen**^{15,17} · **Markus Juonala**¹⁸ · **Sirkka Keinänen-Kiukaanniemi**^{6,9,19,20} · **Veikko Salomaa**¹⁴ · **Markus Perola**^{3,14,21} · **Marjo-Riitta Järvelin**^{6,9,22,23,24} · **Mika Ala-Korpela**^{7,8,25,26,27,28} · **Olli Raitakari**^{1,29} · **Peter Würtz**^{4,30} 

¹ Research Centre of Applied and Preventive Cardiovascular Medicine, University of Turku, Kiinamylynkatu 10, 20520 Turku, Finland

² Department of Internal Medicine, Satakunta Central Hospital, Sairaalanatie 3, 28500 Pori, Finland

³ Institute for Molecular Medicine (FIMM), University of Helsinki, Tukholmankatu 8, 00014 Helsinki, Finland

⁴ Nightingale Health Ltd, Mannerheimintie 164a, 00300 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, 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 Medical Research Center Oulu, Oulu University Hospital and University of Oulu, Oulu, Finland

¹² Nordlab Oulu, Oulu University Hospital, Oulu, Finland

¹³ Department of Clinical Chemistry, University of Oulu, Oulu, Finland

¹⁴ National Institute for Health and Welfare, Helsinki, Finland

¹⁵ Department of Clinical Chemistry, Fimlab Laboratories, Tampere, Finland

¹⁶ Finnish Cardiovascular Research Center–Tampere, Faculty of Medicine and Health Technology, Tampere University, Tampere, Finland

¹⁷ Department of Clinical Physiology, Tampere University Hospital, Tampere, Finland

¹⁸ Department of Medicine, University of Turku and Division of Medicine, Turku University Hospital, Turku, Finland

¹⁹ Healthcare and Social Services of Selanne, Pyhasalmi, Finland

²⁰ Diabetes Unit, Healthcare Services of City of Oulu, Oulu, Finland

²¹ Estonian Genome Center, University of Tartu, Tartu, Estonia

²² Department of Epidemiology and Biostatistics, Medical Research Council–Public Health England Centre for Environment and Health, Imperial College London, London, UK

²³ Biocenter Oulu, University of Oulu, Oulu, Finland

²⁴ Department of Life Sciences, College of Health and Life Sciences, Brunel University, London, UK

²⁵ Medical Research Council Integrative Epidemiology Unit, University of Bristol, Bristol, UK

²⁶ Population Health Science, Bristol Medical School, University of Bristol, Bristol, UK

²⁷ Systems Epidemiology, Baker Heart and Diabetes Institute, Melbourne, Victoria, Australia

²⁸ 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, Victoria, Australia

²⁹ Department of Clinical Physiology and Nuclear Medicine, Turku University Hospital, Turku, Finland

³⁰ Research Programs Unit, Diabetes and Obesity, University of Helsinki, Helsinki, Finland