

RESEARCH ARTICLE

Open Access



Predictive value of circulating NMR metabolic biomarkers for type 2 diabetes risk in the UK Biobank study

Fiona Bragg^{1,2*}, Eirini Trichia^{2†}, Diego Aguilar-Ramirez², Jelena Bešević², Sarah Lewington^{1,2,3} and Jonathan Emberson^{1,2}

Abstract

Background: Effective targeted prevention of type 2 diabetes (T2D) depends on accurate prediction of disease risk. We assessed the role of metabolomic profiling in improving T2D risk prediction beyond conventional risk factors.

Methods: Nuclear magnetic resonance (NMR) metabolomic profiling was undertaken on baseline plasma samples in 65,684 UK Biobank participants without diabetes and not taking lipid-lowering medication. Among a subset of 50,519 participants with data available on all relevant co-variates (sociodemographic characteristics, parental history of diabetes, lifestyle—including dietary—factors, anthropometric measures and fasting time), Cox regression yielded adjusted hazard ratios for the associations of 143 individual metabolic biomarkers (including lipids, lipoproteins, fatty acids, amino acids, ketone bodies and other low molecular weight metabolic biomarkers) and 11 metabolic biomarker principal components (PCs) (accounting for 90% of the total variance in individual biomarkers) with incident T2D. These 11 PCs were added to established models for T2D risk prediction among the full study population, and measures of risk discrimination (*c*-statistic) and reclassification (continuous net reclassification improvement [NRI], integrated discrimination index [IDI]) were assessed.

Results: During median 11.9 (IQR 11.1–12.6) years' follow-up, after accounting for multiple testing, 90 metabolic biomarkers showed independent associations with T2D risk among 50,519 participants (1211 incident T2D cases) and 76 showed associations after additional adjustment for HbA1c (false discovery rate controlled $p < 0.01$). Overall, 8 metabolic biomarker PCs were independently associated with T2D. Among the full study population of 65,684 participants, of whom 1719 developed T2D, addition of PCs to an established risk prediction model, including age, sex, parental history of diabetes, body mass index and HbA1c, improved T2D risk prediction as assessed by the *c*-statistic (increased from 0.802 [95% CI 0.791–0.812] to 0.830 [0.822–0.841]), continuous NRI (0.44 [0.38–0.49]) and relative (15.0% [10.5–20.4%]) and absolute (1.5 [1.0–1.9]) IDI. More modest improvements were observed when metabolic biomarker PCs were added to a more comprehensive established T2D risk prediction model additionally including waist circumference, blood pressure and plasma lipid concentrations (*c*-statistic, 0.829 [0.819–0.838] to 0.837 [0.831–0.848]; continuous NRI, 0.22 [0.17–0.28]; relative IDI, 6.3% [4.1–9.8%]; absolute IDI, 0.7 [0.4–1.1]).

*Correspondence: fiona.bragg@ndph.ox.ac.uk

†Fiona Bragg and Eirini Trichia contributed equally to this work.

¹ MRC Population Health Research Unit, Nuffield Department of Population Health, University of Oxford, Old Road Campus, Oxford OX3 7LF, UK

Full list of author information is available at the end of the article



© The Author(s) 2022. **Open Access** This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>. The Creative Commons Public Domain Dedication waiver (<http://creativecommons.org/publicdomain/zero/1.0/>) applies to the data made available in this article, unless otherwise stated in a credit line to the data.

Conclusions: When added to conventional risk factors, circulating NMR-based metabolic biomarkers modestly enhanced T2D risk prediction.

Keywords: Biomarkers, Diabetes, Metabolomics, Risk prediction

Background

Both population-level and individual high-risk prevention approaches are essential for addressing the major and rising global public health challenge of type 2 diabetes (T2D). Fundamental to the latter is the ability to accurately predict future T2D risk. This enables targeted or *precision* prevention of the disease among individuals at high risk [1] through evidence-based lifestyle or pharmacologic interventions capable of preventing or delaying the onset of T2D [1] and ultimately its complications [2]. Existing T2D risk prediction models perform well in their ability to discriminate between individuals at low or high future risk of T2D [3–5]. However, these models are imperfect, frequently over-estimating T2D risk [6] and often lacking sufficient specificity to be of use clinically [7]. Moreover, they characteristically rely on distal risk factors and consider, at best, only limited molecular pathways. This contrasts with the classical T2D prodrome, comprising dysregulation of multiple molecular pathways over a period of many years [8].

Through metabolomic profiling, large numbers of biomarkers across multiple biological pathways—proximal and distal—can be quantified in a single measurement, capturing the consequences of genetic variation, environmental influences, and their interactions. Prospective studies have established associations of diverse circulating metabolic biomarkers (e.g. amino acids, fatty acids, hexoses, lipids) with T2D [9]. As well as providing aetiological insights, these data might feasibly contribute valuable risk prediction information. Previous studies investigating the ability of metabolomics to improve T2D risk prediction over established risk factors have, with the exception of a small number of studies [10, 11], based their findings on limited T2D cases [12–17], frequently investigating only small numbers or single subclasses of metabolic biomarkers [11–16], or have used untargeted metabolomic profiling, including unknown biomarkers [16, 17] which are less easily and expeditiously translated into clinical use. The resulting inconsistent findings leave an ongoing uncertainty regarding the value of metabolomic profiling for T2D risk prediction.

Using recently available data from the UK Biobank study, we characterise the prospective associations of circulating metabolic biomarkers, quantified using a high-throughput targeted nuclear magnetic resonance (NMR) metabolomics platform, with the risk of incident T2D,

and examine whether the addition of these biomarkers to established models improves prediction of T2D risk.

Methods

Study population

Details of the UK Biobank (UKB) (<https://www.ukbiobank.ac.uk/>) study design and population have been described previously [18]. Briefly, postal invitations to participate were sent to 9.2 million adults aged 40–69 years, living in England, Wales or Scotland and registered with the UK National Health Service. A response rate of 5.5% was achieved, and 502,493 participants were enrolled.

Data collection

The baseline survey took place between 2006 and 2010 in 22 assessment centres. Self-administered touchscreen questionnaires collected information on sociodemographic and lifestyle factors (including diet, physical activity, smoking and alcohol drinking) and personal (supplemented by verbal interview) and family medical history. Physical measurements, including blood pressure, height, weight, waist circumference (WC) and hip circumference, were undertaken using calibrated instruments with standard protocols. A non-fasting venous blood sample was collected, with the time since the last food or drink recorded. After minimal processing at assessment centres, samples were shipped to a central facility for processing and long-term storage at -80°C . Biochemical biomarkers were measured on stored baseline samples at a central UK Biobank laboratory between 2014 and 2017 [19]. These included HDL cholesterol and triglycerides (AU5400; Beckman Coulter) and HbA1c (VARIANT II TURBO Hemoglobin Testing System; BioRad). Repeat surveys collected the same information as at baseline in addition to certain enhancements; they comprised a resurvey of $\sim 20,000$ participants in 2012–2013 and an ongoing survey of $\sim 100,000$ participants which commenced in 2014 [20, 21].

All participants consented to be followed up through linkage to health-related records. These included prior and prospective data on dates and causes of hospital admissions (Hospital Episode Statistics in England, Patient Episode Database for Wales, and Scottish Morbidity Record) and primary care clinical events and prescribing (available for $\sim 45\%$ of participants), as well as

date and cause of death obtained from national death registries.

Metabolic biomarker quantification

A high-throughput NMR metabolomics platform [22, 23] was used to undertake metabolomic profiling in baseline plasma samples from a randomly selected subset of ~ 120,000 UKB participants [24]. This simultaneously quantified 249 metabolic biomarkers (168 directly measured and 81 ratios of these), including lipids, fatty acids, amino acids, ketone bodies and other low-molecular-weight metabolic biomarkers (e.g. gluconeogenesis-related metabolites), as well as lipoprotein subclass distribution, particle size and composition. A subset of 143 (Additional file 1: Table S1) was selected for inclusion in the presented analyses, focussing on those which were directly measured and could not be inferred from other biomarkers.

Assessment of incident type 2 diabetes status

Incident T2D status was ascertained through (i) self-report of T2D diagnosis or glucose-lowering medication use at repeat surveys; (ii) coded T2D diagnoses recorded in primary care, hospital admission or death registry data; or (iii) glucose-lowering medication prescribing in primary care data (Additional file 1: Table S2). Only those participants without diagnostic codes for other specified diabetes types (type 1/malnutrition-related/other specified diabetes) were considered to have T2D.

Statistical analysis

The analyses excluded those with previously diagnosed diabetes of any type (based on self-report, primary care or inpatient hospital data), taking regular glucose-lowering medication (based on self-report or primary care data) or with HbA1c $\geq 6.5\%$ (corresponding to 48 mmol/mol and consistent with undiagnosed diabetes) at the baseline survey. Those with missing or extreme NMR biomarker or covariate data (see below), or who were taking lipid-lowering medications at recruitment, were also excluded from the main analyses. This generated a 'risk prediction population' comprising 65,684 participants and, following the exclusion of a further 15,165 participants with missing data for additional covariates included only in association analyses, an 'association analyses population' of 50,519 participants (Additional file 1: Fig. S1).

All NMR biomarkers were log-transformed and standardised. Principal component analysis was then employed to reduce a large number of correlated NMR biomarkers (Additional file 1: Fig. S2) to a much smaller number of uncorrelated principal components (PCs) which retained most (> 90%) of the variance in the individual biomarkers. Cox regression among 50,519 participants in the

association analyses population was used to assess the individual relevance of each NMR biomarker (and each PC) to the risk of incident T2D. First, to examine the shape of the associations, participants were grouped into baseline categories defined by quartiles of their distributions and a test of trend performed across quartiles. Subsequently, continuous analyses of each NMR biomarker (and each PC) were done to estimate the HR per 1-SD higher baseline level. Cox models were stratified by age-at-risk (5-year age groups) and sex and adjusted for assessment centre (22 centres), Townsend Deprivation Index (numeric), ethnicity (6 categories), parental history of diabetes (2 categories), smoking (4 categories), alcohol drinking (4 categories), physical activity (numeric), dietary factors (whole and refined grains, fruit, vegetables, cheese, unprocessed red meat, processed meat, non-oily and oily fish, type of spread, tea [all 4 categories], coffee [caffeinated 4 categories; decaffeinated 3 categories], dietary supplements [4 categories]), body mass index (BMI) (numeric), waist-to-hip ratio (WHR) (numeric), fasting time (numeric) and spectrometer (6 spectrometers). Participants who did not develop incident T2D were censored at the earliest of death, loss to follow-up or 31 December 2020. For significance testing, the Benjamini-Hochberg method was used to control the false discovery rate (FDR) [25]. Statistical significance was defined as FDR controlled $p < 0.01$. Sensitivity analyses examined the associations separately by age (< 55 vs ≥ 55 years) and sex and after additional adjustment for HbA1c. In addition, the impact of excluding the first 3 years of follow-up was assessed and, for the analysis of each PC, mutual adjustment for all preceding PCs.

Then, to assess whether circulating NMR biomarkers could improve the prediction of T2D risk, the selected PCs were added to the 'traditional' T2D risk prediction models based on Framingham risk scores for T2D [3] among 65,684 participants in the risk prediction population. Two models were assessed: a 'concise' model, including age (< 50, 50-64, ≥ 65 years), sex, parental history of diabetes, BMI (< 25.0, 25.0-29.9, ≥ 30.0 kg/m²) and HbA1c (< 6.0% vs $\geq 6.0\%$), and a 'full' model, which additionally included blood pressure ($\leq 130/85$ mmHg and not taking anti-hypertensive medication vs > 130/85 mmHg or taking anti-hypertensive medication), HDL cholesterol (< 1.0 vs ≥ 1.0 mmol/L in men; < 1.3 vs ≥ 1.3 mmol/L in women), triglycerides (< 1.7 vs ≥ 1.7 mmol/L), and WC (≤ 102 vs > 102 cm in men; ≤ 88 vs > 88 cm in women) [3]. The discriminatory ability of each model before and after including the PCs was assessed using Harrell's *c*-statistic [26], and the likelihood ratio test was used to compare the fits of nested models (i.e. those including versus excluding the PCs). Relative and absolute integrated discrimination improvement (IDI)

[27] and continuous net reclassification improvement (NRI) [28] were estimated to assess risk reclassification. To avoid model optimism, bootstrapping was used to create bias-corrected estimates and CIs for the c-statistics, IDI and NRI. To test model calibration, observed T2D event rates for absolute predicted risk deciles were plotted against their predicted event rates, and calibration slopes were estimated using a Cox regression analysis of predicted risk on observed risk. Calibration slopes and their confidence intervals were estimated from 10-fold cross-validation (pooled using inverse variance weighting). Subsequent analyses assessed the performance of the four risk prediction models solely among 13,695 participants taking lipid-lowering medications at baseline. Sensitivity analyses separately assessed their performance after replacing WC with WHR and, where appropriate, including model covariates as continuous variables.

Analyses were conducted using SAS (version 9.4) and R (version 3.6.2).

Results

Of the original 502,493 UKB participants, a random subset of 118,036 (23%) had NMR biomarker data (Additional file 1: Fig. S1, Additional file 1: Table S3). Of these, 65,684 (56%) had no prior diabetes, were not taking lipid-lowering medication and had complete NMR biomarker (and other) data for inclusion in subsequent risk prediction analyses. The mean (SD) age of participants in this risk prediction population was 55.2 (8.0) years, and 58% ($n = 37,849$) were women (Table 1). During 0.8 million person-years of follow-up (median 11.9 [IQR 11.1–12.6]), 1719 cases of incident T2D were identified. Participants who developed T2D were more likely to be male and, at the time of recruitment, tended to be older and of lower socioeconomic status than those who did not develop T2D. They also had higher levels of adiposity, were more likely to be current regular smokers, but less likely to be current regular alcohol drinkers, and more frequently had a parental history of diabetes.

Among 50,519 participants in the association analyses population, of whom 1211 developed incident T2D (Additional file 1: Table S4), after adjustment for potential confounding factors and accounting for multiple testing, 90 of the 143 metabolic biomarkers showed statistically significant associations with the risk of incident T2D at FDR controlled $p < 0.01$ (Fig. 1, Additional file 1: Table S1, Additional file 1: Fig. S3). Among the strongest positive associations were those of VLDL particle concentrations, particularly larger VLDL particles, and the lipid concentrations within them. Triglyceride concentrations in all 14 lipoprotein subclasses were also very strongly positively associated with incident T2D. Conversely, concentrations of larger HDL particles, and the cholesterol

Table 1 Baseline characteristics of 65,684 participants in the risk prediction population by incident type 2 diabetes status

Baseline characteristics ^a	Incident type 2 diabetes		Total
	Yes	No	
No. of participants	1719	63,965	65,684
Age, sex and socioeconomic factors			
Mean age (SD), years	57.1 (7.7)	55.1 (8.0)	55.2 (8.0)
Women, %	47	58	58
Townsend Deprivation Index (SD) ^b	0.3 (1.1)	0.0 (1.0)	0.0 (1.0)
Lifestyle factors			
Smoking, %			
Never or occasional	54	61	61
Previous	33	32	32
Current regular	14	7	7
Alcohol drinking, %			
Never or occasional	42	25	26
Previous	5	3	3
Current regular	53	71	71
Physical and blood-based measurements, mean (SD)			
BMI, kg/m ²	31.5 (5.2)	26.8 (4.5)	26.9 (4.4)
WC, cm	100 (13)	88 (12)	88 (13)
HC, cm	110 (10)	103 (9)	103 (9)
WHR	0.91 (0.08)	0.86 (0.09)	0.86 (0.09)
SBP, mmHg	142 (19)	136 (18)	136 (18)
DBP, mmHg	86 (11)	82 (10)	82 (10)
Total cholesterol, mmol/L ^c	6.0 (1.1)	5.9 (1.1)	5.9 (1.1)
LDL cholesterol, mmol/L ^c	3.8 (0.8)	3.7 (0.8)	3.7 (0.8)
HDL cholesterol, mmol/L ^c	1.3 (0.3)	1.5 (0.4)	1.5 (0.4)
Triglycerides, mmol/L ^c	2.4 (1.3)	1.7 (1.0)	1.7 (1.0)
HbA1c, %	5.8 (0.4)	5.3 (0.3)	5.3 (0.3)
Previously diagnosed coronary heart disease, %	1.8	0.6	0.7
Parental history of diabetes, %	38	18	19
Mean fasting time (SD), hours	4.0 (2.6)	3.7 (2.4)	3.7 (2.4)

Participants with missing data: total cholesterol, $n = 7$; LDL cholesterol $n = 83$

BMI body mass index, DBP diastolic blood pressure, HC hip circumference, HDL high-density lipoprotein, LDL low-density lipoprotein, SBP systolic blood pressure, WC waist circumference, WHR waist-to-hip ratio

^a Standardised to age and sex structure of the study population

^b Standardised Townsend Deprivation Index; higher scores represent higher levels of deprivation

^c Clinical chemistry derived concentrations

and phospholipids within those particles, were inversely associated with T2D. Higher branched-chain amino acid (BCAA)—leucine, isoleucine and valine—concentrations were associated with a higher risk of T2D, as were higher concentrations of alanine, phenylalanine and tyrosine. Glutamine and glycine were inversely associated with T2D. Relative to total fatty acids, higher concentrations of polyunsaturated, omega-3 and omega-6 fatty acids and

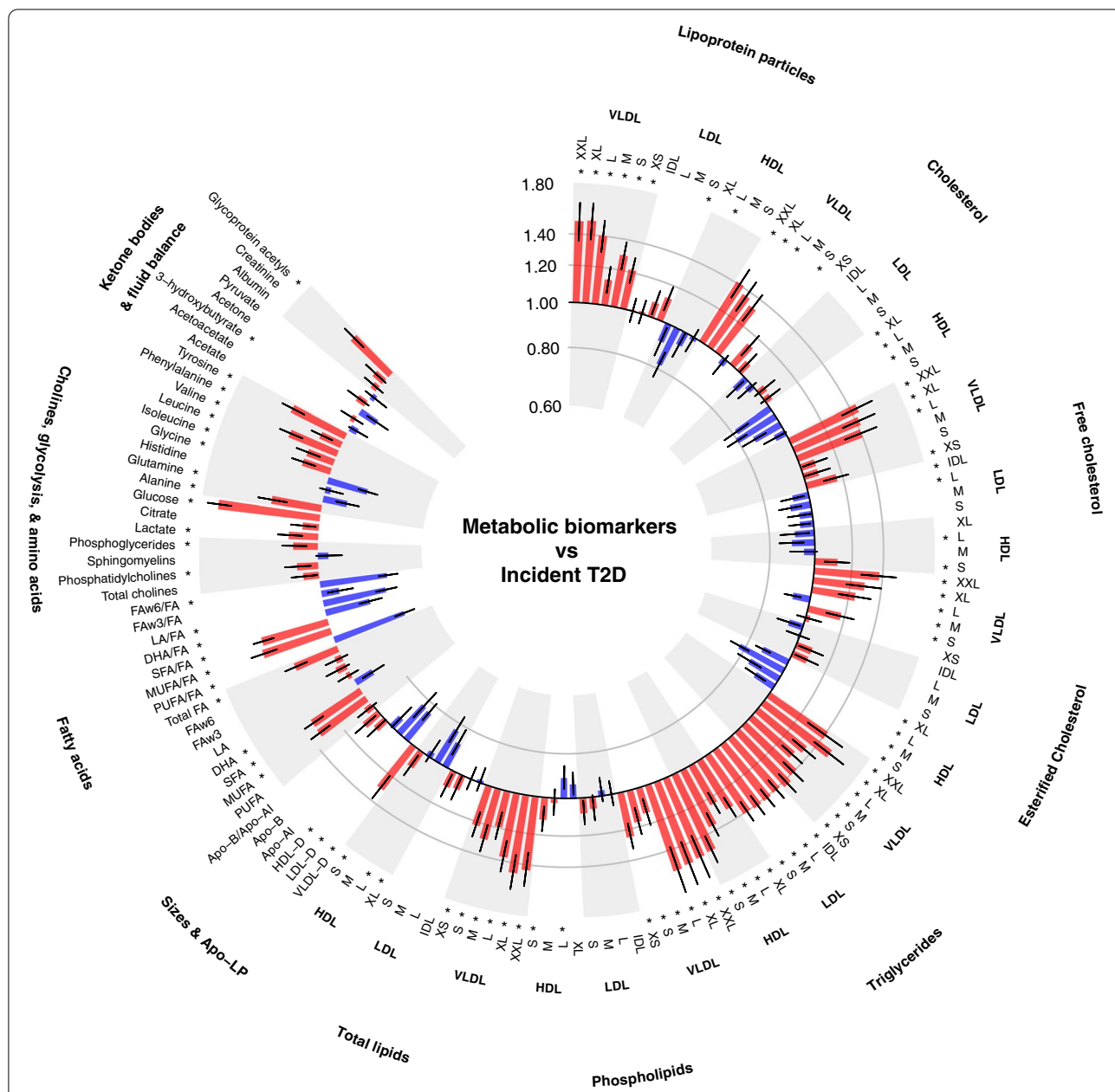


Fig. 1 Associations of metabolic biomarkers with risk of incident type 2 diabetes among 50,519 participants in the association analyses population. Hazard ratios (with 95% confidence intervals) are presented per 1-SD higher metabolic biomarker on the natural log scale, stratified by age-at-risk and sex and adjusted for assessment centre, Townsend Deprivation Index, ethnicity, parental history of diabetes, smoking, alcohol drinking, physical activity, dietary factors (whole and refined grains, fruit, vegetables, cheese, unprocessed red meat, processed meat, non-oily and oily fish, type of spread, caffeinated and decaffeinated coffee, tea and dietary supplements), body mass index, waist-to-hip ratio, fasting duration and spectrometer. *False discovery rate controlled $p < 0.01$. Apo-A1, apolipoprotein A1; Apo-B, apolipoprotein B; DHA, docosahexaenoic acid; FA, fatty acids; Faw3, omega-3 fatty acids; Faw6, omega-6 fatty acids; HDL, high-density lipoproteins; HDL-D, high-density lipoprotein particle diameter; IDL, intermediate-density lipoproteins; L, large; LA, linoleic acid; LDL, low-density lipoproteins; LDL-D, low-density lipoprotein particle diameter; LP, lipoprotein; M, medium; MUFA, monounsaturated fatty acids; PUFA, polyunsaturated fatty acids; S, small; SFA, saturated fatty acids; T2D, type 2 diabetes; VLDL, very low-density lipoproteins; VLDL-D, very low-density lipoprotein particle diameter; XL, very large; XS, very small; XXL, extremely large

docosahexaenoic and linoleic acids were associated with lower T2D risk, whereas higher concentrations of saturated and monounsaturated fatty acids were associated with higher T2D risk. Higher plasma glycoprotein acetyls, a marker of inflammation, were also associated with higher T2D risk.

After additional adjustment for HbA1c, many associations were moderately attenuated but statistically significant associations of most biomarkers ($n = 76$ at FDR controlled $p < 0.01$) with T2D remained (Additional file 1: Table S1). There were no marked differences in the relationships between men and women (Additional file 1: Fig. S4), by age at baseline (Additional file 1: Fig. S5), or after exclusion of the first 3 years of follow-up (Additional file 1: Fig. S6).

The first 11 PCs of the NMR biomarkers explained 90% of the total variance present in the 143 individual biomarkers (Additional file 1: Fig. S7). The PC loadings from these 11 PCs are shown in Fig. S8 (Additional file 1: Fig. S8) (the larger a biomarker's loading, positive or negative, the more it contributes to that PC), and the associations of these PCs with incident T2D are shown in Table S5 (Additional file 1: Table S5). The major contributors to PC1 were VLDL and LDL particle concentrations and the lipid concentrations within those particles, while for PC2, they included large HDL particles and lipid concentrations within them. PC1 and PC2 showed opposing associations with T2D (adjusted HR 1.25 [95% CI 1.17–1.34] and 0.81 [0.76–0.87], respectively). Biomarkers across multiple molecular pathways, including lipid concentrations in LDL and HDL particles and apolipoprotein concentrations, were prominent contributors to PC3 (HR 1.23 [95% CI 1.17–1.30]). Within PC4 (HR 1.13 [95% CI 1.06–1.20]), loadings were high for small and very large HDL particles and their lipid concentrations, and amino acids were the major contributors to PC5 (1.07 [1.00–1.14]). Fatty acids were dominant in PC6 (0.95 [95% CI 0.89–1.01]) and also PC7 (0.72 [0.67–0.76]), in which ketone bodies also had large factor loadings. Overall, 8 of the 11 PCs were associated with incident T2D independent of sociodemographic characteristics, parental history of diabetes, lifestyle factors, anthropometric measures and fasting time and largely remained so after sequential adjustment for preceding PCs.

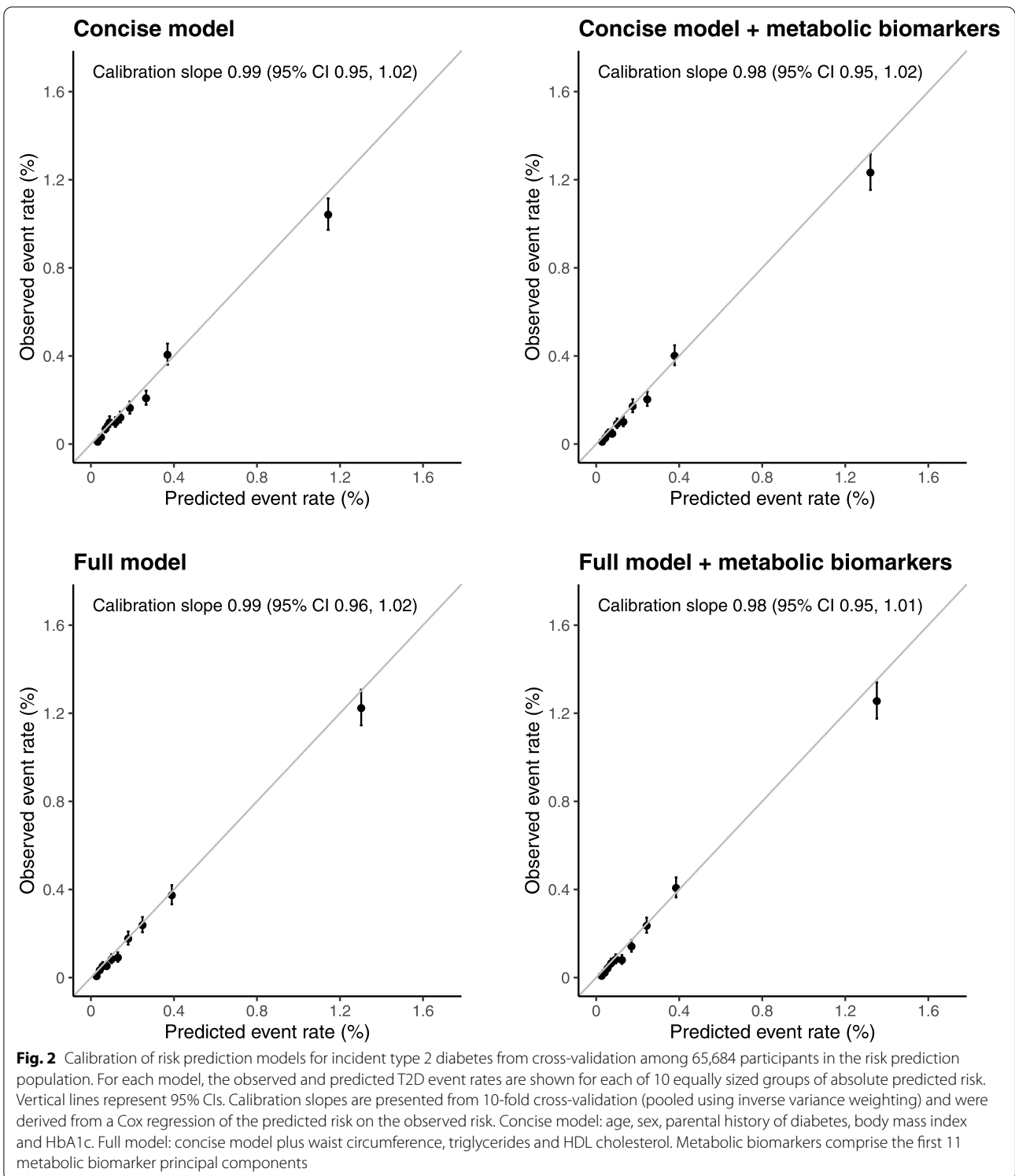
In the two traditional risk prediction models, all risk factors were strongly and independently associated with T2D risk among 65,684 participants in the risk prediction population (Additional file 1: Table S6). Older age, male sex, parental history of diabetes, higher levels of adiposity, blood pressure, HbA1c and triglycerides and lower HDL cholesterol concentration were all associated with higher risk. These relationships largely persisted, although with modest attenuation of some, when

metabolic biomarker PCs were added. For both models, 8 of the 11 PCs were significantly associated with T2D risk independently of all other risk factors.

The concise T2D risk prediction model (based on the Framingham 'personal' model for T2D risk prediction and incorporating age, sex, parental history of diabetes, BMI and HbA1c) demonstrated good calibration of observed versus predicted T2D rates across deciles of predicted risk (calibration slope, 0.99 [95% CI 0.95–1.02]) (Fig. 2). This did not meaningfully change after the addition of metabolic biomarker PCs (0.98 [95% CI 0.95–1.02]). Table 2 summarises the measures of model fit and performance. The addition of the PCs to the concise model resulted in a 17% increase in the chi-square statistic and yielded an increase in the c-statistic from 0.802 (95% CI 0.791–0.812) to 0.830 (0.822–0.841). Improved T2D risk prediction on addition of the PCs was also evidenced by estimates of the overall continuous NRI (0.44 [95% CI 0.38–0.49]), with an improvement of 0.15 (0.12–0.20) in events and 0.28 (0.26–0.31) in non-events and both absolute (1.5 [1.0–1.9]) and relative (15.0% [10.5–20.4%]) IDI. The full model (concise model plus blood pressure, WC, HDL cholesterol and triglycerides, based on the Framingham 'clinical' model for T2D risk prediction) achieved a c-statistic of 0.829 (95% CI 0.819–0.838). Modest improvements in model fit and performance were observed following the addition of metabolic biomarker PCs to this model, with a 6% increase in the chi-square statistic, a c-statistic of 0.837 (95% CI 0.831–0.848), an overall continuous NRI of 0.22 (0.17–0.28), an absolute IDI of 0.7 (0.4–1.1) and a relative IDI of 6.3% (4.1–9.8%). The full model was well-calibrated, both with and without the inclusion of metabolic biomarker PCs (0.99 [95% CI 0.96–1.02] and 0.98 [0.95–1.01], respectively). When analyses were repeated among participants taking lipid-lowering medications at baseline, c-statistics for the four individual T2D risk prediction models were lower than in the main study population, but estimates of the relative performance of the nested models were broadly comparable (Additional file 1: Table S7). Sensitivity analyses replacing WC in the full model with WHR did not materially affect the performance of the model (Additional file 1: Table S8). Including covariates as continuous variables improved the discriminatory ability of both the concise and the full model; although moderately diminished, the ability of metabolic biomarker PCs to improve T2D risk prediction remained (Additional file 1: Table S9).

Discussion

This prospective population-based cohort study of over 65,000 middle-aged adults with 1719 cases of new-onset T2D is, to our knowledge, the largest study



to date to examine the predictive value of circulating metabolic biomarkers for T2D risk. Strong independent associations of diverse biomarkers, quantified using targeted NMR-based metabolomic profiling, including

lipoprotein particle size and composition, amino acids and fatty acids, with risk of incident T2D were observed. When added to established risk prediction models, PCs derived from 143 circulating

Table 2 Performance of risk prediction models for incident type 2 diabetes among 65,684 participants in the risk prediction population

Performance metric	Concise model ^a	Concise model ^a plus metabolic biomarkers ^b	Full model ^c	Full model ^c plus metabolic biomarkers ^b
C-statistic (CI)^d	0.802 (0.791, 0.812)	0.830 (0.822, 0.841)	0.829 (0.819, 0.838)	0.837 (0.831, 0.848)
Metrics of relative performance				
χ^2 ^{e,f}		453 ($p < 0.0001$)		177 ($p < 0.0001$)
%increase χ^2		17		6
Absolute IDI ^{f,g}		1.5 (1.0, 1.9)		0.7 (0.4, 1.1)
Relative IDI (%) (CI) ^{f,g}		15.0 (10.5, 20.4)		6.3 (4.1, 9.8)
Continuous NRI (CI) ^{f,h}				
Events		0.15 (0.12, 0.20)		0.10 (0.06, 0.14)
Non-events		0.28 (0.26, 0.31)		0.12 (0.09, 0.14)
Overall		0.44 (0.38, 0.49)		0.22 (0.17, 0.28)

DF degrees of freedom, IDI integrated discrimination improvement, NRI net reclassification improvement; T2D type 2 diabetes

^a Concise model: age, sex, parental history of diabetes, body mass index and HbA1c

^b Metabolic biomarkers comprise the first 11 metabolic biomarker principal components

^c Full model: concise model plus waist circumference, blood pressure, triglycerides and HDL cholesterol

^d The c-statistic measures the ability of a model to rank participants from low to high risk. Given two randomly selected individuals, one who develops T2D and one who does not, the c-statistic is the probability that the model will give a higher predicted risk for the individual who develops T2D. An uninformative model will have a c-statistic of 0.5 and a model that discriminates perfectly will have a c-statistic of 1.0

^e 11 DF

^f Bias-corrected estimates and confidence intervals were derived using 200 bootstrap samples

^g The IDI quantifies the difference between two models in their ability to predict risk. It is calculated as the difference between the two models in the mean predicted T2D risk among those who did develop T2D minus the mean predicted risk of T2D in those who did not develop T2D (i.e. it is the difference between two differences). When metabolic biomarkers were added to the concise model, the separation in the mean predicted T2D risk between those who did develop T2D, compared with those who did not develop T2D, increased in relative terms by 15.0%. Positive IDI values indicate improved T2D risk classification following the addition of metabolic biomarkers to the risk prediction model

^h The continuous NRI quantifies the appropriateness of the change in predicted probabilities of T2D between two models. The 'events' NRI is calculated among those who developed T2D, and the 'non-events' NRI is calculated among those who did not develop T2D. Both statistics are calculated as the probability of an 'appropriate' change in predicted risk (after the addition of metabolic biomarkers to the model) minus the probability of an 'inappropriate' change in predicted risk. For those who developed T2D, an appropriate change would be a higher predicted T2D risk after the addition of metabolic biomarkers to the model. An inappropriate change would be a lower predicted T2D risk after the addition of metabolic biomarkers to the model. When metabolic biomarkers were added to the concise model, among those who developed T2D, 15% more were assigned a higher predicted T2D risk than were assigned a lower predicted risk. The overall NRI is the sum of the 'events' and 'non-events' NRI statistics. Positive NRI values indicate that the addition of metabolic biomarkers results in a superior model

biomarkers achieved modest improvements in T2D risk prediction.

Our study found strong positive associations of VLDL particle measures and triglyceride concentrations with incident T2D risk and inverse associations of HDL particle size and lipids within larger HDL particles. These findings are qualitatively, and broadly quantitatively, consistent with previous studies [29, 30], and are characteristic of lipoprotein profiles associated with insulin resistance [31]. This is also thought to underlie the strong positive associations of BCAAs—leucine, isoleucine and valine—with the risk of T2D observed in UKB and in previous studies among diverse populations [9, 11, 29]. More specifically, genetic association studies have shown increased BCAA levels as a consequence of insulin resistance [32], which, in turn, appear to be causally related to T2D [33]. We replicated the findings of studies showing higher levels of phenylalanine, tyrosine and alanine [9, 11, 29], and lower concentrations of glutamate [9, 29] and glycine [9] several

years prior to T2D diagnosis, and the observed T2D-associated fatty acid profiles are broadly consistent with previous investigations [29, 34]. Insulin resistance and inflammation are postulated to underlie some or all of these associations [9, 11, 34], but the nature of the relationships of these and other metabolic biomarkers with T2D, including their causal significance, remains uncertain. Despite this, these findings provide clear evidence of the relevance of diverse metabolic biomarkers to T2D risk.

The 'traditional' T2D risk prediction models examined in the present study, based on the Framingham 'personal' and 'clinical' diabetes risk scores, demonstrated good discriminatory ability in the UKB population (concise model: c-statistic 0.80; full model: c-statistic 0.83). This is consistent with the performance of similar models across varied populations [35], highlighting one of the major challenges of identifying novel predictive biomarkers for T2D. That is, that established clinical risk factors perform so well in predicting T2D risk that achieving clinically

meaningful improvements above and beyond these is difficult.

The addition of metabolic biomarkers to the concise model in the present study improved, albeit modestly, model fit and risk discrimination (*c*-statistic 0.83). Although some previous studies have observed no improvement in risk discrimination with the addition of metabolic biomarkers to similar traditional risk prediction models [12, 15, 36], several have investigated the impact of only limited biomarkers [12, 15]. The inclusion of more diverse biomarkers has tended to achieve greater gains in model discrimination [10, 17, 29]. For example, in a case-cohort study in Germany, comprising 800 T2D cases and a randomly selected subcohort of 2282 adults (mean follow-up of 7 years), the addition of 14 metabolic biomarkers (including hexoses, amino acids and fatty acids) to an established T2D risk score, comprising clinical risk factors and glycaemia, resulted in moderate, but statistically significant, improvement in risk discrimination (increase in *c*-statistic from 0.901 to 0.912; $p < 0.0001$) [10]. However, even these studies have tended to investigate highly selected subsets of biomarkers. In contrast, the use of principal component analysis in the present study facilitated the inclusion of information from all 143 metabolic biomarkers, despite their highly correlated nature. Many individual biomarkers most strongly associated with T2D were prominent contributors to the PCs selected for inclusion in risk prediction models, most of which were associated with incident T2D in fully adjusted regression models. The more modest, and non-significant, gains in risk discrimination when metabolic biomarkers were added to the full model may reflect the insensitivity of the *c*-statistic to improvements in predictive performance with the addition of new, even strong, risk predictors to established models [37]. It likely also reflects an overlap between measured metabolic biomarkers and blood-based risk factors included in this model. These findings suggest that there may be limited value for T2D risk discrimination of adding metabolic biomarkers to a risk prediction model which already includes routine clinical chemistry lipid measures. However, the inclusion of metabolic biomarkers instead of these routine lipid measures would be expected to enhance T2D risk prediction model performance. Moreover, metabolomic profiling data may be of wider clinical relevance (e.g. for diagnosis and risk assessment of other cardiometabolic diseases) [23].

More global measures of model performance provided supportive evidence of the value of metabolic biomarkers for T2D risk prediction. Their addition to both traditional risk prediction models in the present study was associated with improvement in the prediction of T2D using measures of risk reclassification, specifically the

IDI and continuous NRI. Of note, the NRI appeared to be driven more by reductions in predicted risk among participants who did not develop T2D, suggesting metabolomic profiling may be particularly valuable for reducing unnecessary prevention interventions among individuals at low risk of T2D. Increasing availability of standardised, quantitative, high-throughput metabolomics platforms, such as that used in the current study, underscores the potential translational relevance of these findings.

In addition to a large number of incident T2D events, our study has several strengths. An established targeted NMR metabolomics platform, with existing clinical regulatory approvals [24], was used; as well as enabling quantification of diverse biomarkers, this facilitates comparisons between study populations and enhances the potential clinical relevance. Moreover, high levels of correlation between NMR- and standard clinical chemistry-derived concentrations of a subset of biomarkers (Additional file 1: Fig. S9) support the validity of the approach [38]. The exclusion of participants taking lipid-lowering medication avoided treatment-associated biases, although the broadly comparable performance of the nested risk prediction models in this subpopulation (with a higher frequency of incident T2D) demonstrates the wider generalisability of our findings. Finally, the cohort study design avoided potential biases and loss of precision which may affect more frequently used nested case-control and case-cohort designs. However, the study also has limitations. Incident T2D was limited to diagnosed cases; although resulting misclassification would likely underestimate associations of metabolic biomarkers with T2D, the relative improvements in model performance (between models with versus without metabolic biomarkers) should be largely unaffected by misclassification in outcome assessment. Blood samples were taken in the non-fasting state, and so would be subject to greater variability in metabolic biomarker concentrations than fasting samples (although fasting duration has previously been found to account for only a small proportion of variation in plasma metabolic biomarker concentrations [39]). However, our analyses were adjusted for fasting time, as well as for extensive dietary factors, which should have limited any material impact of the use of non-fasting samples on the findings. Moreover, although risk prediction would ideally incorporate repeat biomarker levels measured longitudinally, the use of single measurements more closely reflects the practical implementation of risk prediction models in the clinical setting. Independent validation of the risk prediction findings was not performed. However, the observed associations of metabolic biomarkers with T2D replicate previous study findings [9, 11, 29, 30, 34], with no novel associations identified. Finally, the more favourable lifestyle and health-related

characteristics of the UKB population when compared with the general UK population (e.g. participants were less likely to be obese or to smoke and had fewer chronic diseases at recruitment) [40] would not be expected to impact on the generalisability of observed associations of metabolic biomarkers with T2D risk [40, 41]. However, the risk prediction findings may not necessarily be generalisable to other populations at higher risk of future T2D.

Conclusions

In summary, this study provides large-scale evidence of the incremental predictive value of metabolomic profiling for the prediction of T2D risk. Addition of data on 143 circulating metabolic biomarkers, with replicated prospective associations with T2D, to an established risk prediction model comprising basic clinical risk factors and HbA1c improved T2D risk discrimination and classification. More modest improvements were observed when metabolic biomarkers were added to a model additionally incorporating WC, blood pressure, and plasma lipid measures. The study serves to illustrate the utility of large-scale biobanks for the assessment of the clinical relevance and value of emerging biomarkers. Moreover, given increasing availability, including in clinical settings, of high-throughput, comprehensive, targeted metabolomic profiling, these findings may have translational relevance for T2D risk stratification and precision prevention.

Abbreviations

BCAA: Branched-chain amino acid; BMI: Body mass index; CI: Confidence interval; DBP: Diastolic blood pressure; FDR: False discovery rate; HC: Hip circumference; HDL: High-density lipoprotein; IDI: Integrated discrimination index; IQR: Interquartile range; LDL: Low-density lipoprotein; NMR: Nuclear magnetic resonance; NRI: Net reclassification improvement; PC: Principal component; SBP: Systolic blood pressure; SD: Standard deviation; T2D: Type 2 diabetes; UKB: UK Biobank; VLDL: Very low-density lipoprotein; WC: Waist circumference; WHR: Waist-to-hip ratio.

Supplementary Information

The online version contains supplementary material available at <https://doi.org/10.1186/s12916-022-02354-9>.

Additional file 1: Predictive value of circulating NMR metabolic biomarkers for type 2 diabetes risk in the UK Biobank study. **Table S1.** Distribution of metabolic biomarkers and their associations with incident type 2 diabetes among 50,519 participants in the association analyses population. **Table S2.** Diagnosis and medication codes for assessment of type 2 diabetes status in primary and secondary healthcare and death registry records and UK Biobank verbal interview. **Table S3.** Baseline characteristics of participants with and without NMR-metabolomics profiling. **Table S4.** Baseline characteristics of 50,519 participants in the association analyses population by incident type 2 diabetes status. **Table S5.** Associations of the first 11 metabolic biomarker principal components with risk of incident type 2 diabetes among 50,519 participants in the association analyses population. **Table S6.** Regression models for risk of incident type 2 diabetes among 65,684 participants in the risk prediction population. **Table S7.** Performance of risk prediction models for incident type 2

diabetes among 13,695 participants taking lipid-lowering medication at recruitment. **Table S8.** Performance of risk prediction models including waist-to-hip ratio for incident type 2 diabetes among 65,684 participants in the risk prediction population. **Table S9.** Performance of risk prediction models for incident type 2 diabetes incorporating co-variables, where relevant, as continuous variables among 65,684 participants in the risk prediction population. **Fig. S1.** Participant exclusions to derive risk prediction and association analyses populations. **Fig. S2.** Cross-correlations of metabolic biomarkers. **Fig. S3.** Associations of metabolic biomarkers with risk of incident type 2 diabetes among 50,519 participants in the association analyses population. **Fig. S4.** Associations of metabolic biomarkers with risk of incident type 2 diabetes by sex among 50,519 participants in the association analyses population. **Fig. S5.** Associations of metabolic biomarkers with risk of incident type 2 diabetes by age among 50,519 participants in the association analyses population. **Fig. S6.** Associations of metabolic biomarkers with risk of incident type 2 diabetes excluding the first three years of follow-up in the association analyses population. **Fig. S7.** Importance of the first 20 metabolic biomarker principal components. **Fig. S8.** Characterisation of the first 11 metabolic biomarker principal components among 65,684 participants in the risk prediction population. **Fig. S9.** Comparison of biomarkers measured by NMR and routine clinical chemistry assays among 65,684 participants in the risk prediction population.

Acknowledgements

This research used the UK Biobank resource (application number 30418). We thank the participants of UK Biobank for their contribution to the resource. The authors are grateful to Nightingale Health Ltd. for providing early access to UK Biobank NMR-metabolomics data during Nightingale Health's exclusivity period, prior to the resource being made openly available to the scientific community.

Authors' contributions

All authors were involved in the study design, analysis of data, interpretation or writing of the report. The authors read and approved the final manuscript.

Funding

The British Heart Foundation, Medical Research Council and Cancer Research UK provide core funding to the Oxford CTSU. DAR acknowledges the support from the BHF Centre of Research Excellence, Oxford (Grant code RE/13/1/30181). SL reports grants from the Medical Research Council (MRC) and research funding from the US Centers for Disease Control and Prevention Foundation (with support from Amgen) during the conduct of the study. JE reports grants from Boehringer Ingelheim, Regeneron and Astra Zeneca outside the submitted work.

Availability of data and materials

The underlying data are open access through application to the UK Biobank, and materials and methods will be made freely available through the UK Biobank as part of this project.

Declarations

Ethics approval and consent to participate

Ethics approval for the UK Biobank was obtained from the North West Multi-centre Research Ethics Committee (Ref: 11/NW/0382). All participants provided informed written consent.

Consent for publication

Not applicable

Competing interests

The authors declare that they have no competing interests.

Author details

¹MRC Population Health Research Unit, Nuffield Department of Population Health, University of Oxford, Old Road Campus, Oxford OX3 7LF, UK. ²Clinical Trial Service Unit & Epidemiological Studies Unit, Nuffield Department

of Population Health, University of Oxford, Old Road Campus, Oxford OX3 7LF, UK. ³UKM Medical Molecular Biology Institute (UMBI), Universiti Kebangsaan Malaysia, Kuala Lumpur, Malaysia.

Received: 1 December 2021 Accepted: 28 March 2022

Published online: 03 May 2022

References

- Gillies CL, Abrams KR, Lambert PC, Cooper NJ, Sutton AJ, Hsu RT, et al. Pharmacological and lifestyle interventions to prevent or delay type 2 diabetes in people with impaired glucose tolerance: systematic review and meta-analysis. *BMJ*. 2007;334(7588):299.
- Gong Q, Zhang P, Wang J, Ma J, An Y, Chen Y, et al. Morbidity and mortality after lifestyle intervention for people with impaired glucose tolerance: 30-year results of the Da Qing diabetes prevention outcome study. *Lancet Diabetes Endocrinol*. 2019;7(6):452–61.
- Wilson PW, Meigs JB, Sullivan L, Fox CS, Nathan DM, D'Agostino RB Sr. Prediction of incident diabetes mellitus in middle-aged adults: the Framingham offspring study. *Arch Intern Med*. 2007;167(10):1068–74.
- Lindström J, Tuomilehto J. The diabetes risk score: a practical tool to predict type 2 diabetes risk. *Diabetes Care*. 2003;26(3):725–31.
- Hippisley-Cox J, Coupland C, Robson J, Sheikh A, Brindle P. Predicting risk of type 2 diabetes in England and Wales: prospective derivation and validation of QDScore. *BMJ*. 2009;338:b880.
- Abbasi A, Peelen LM, Corpeleijn E, van der Schouw YT, Stolk RP, Spijkerman AMW, et al. Prediction models for risk of developing type 2 diabetes: systematic literature search and independent external validation study. *BMJ*. 2012;345:e5900.
- Noble D, Mathur R, Dent T, Meads C, Greenhalgh T. Risk models and scores for type 2 diabetes: systematic review. *BMJ*. 2011;343:d7163.
- Færch K, Witte DR, Tabák AG, Perreault L, Herder C, Brunner EJ, et al. Trajectories of cardiometabolic risk factors before diagnosis of three subtypes of type 2 diabetes: a post-hoc analysis of the longitudinal Whitehall II cohort study. *Lancet Diabetes Endocrinol*. 2013;1(1):43–51.
- Guasch-Ferré M, Hruby A, Toledo E, Clish CB, Martínez-González MA, Salas-Salvadó J, et al. Metabolomics in prediabetes and diabetes: a systematic review and meta-analysis. *Diabetes Care*. 2016;39(5):833–46.
- Floegel A, Stefan N, Yu Z, Mühlenbruch 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.
- Qiu G, Zheng Y, Wang H, Sun J, Ma H, Xiao Y, et al. Plasma metabolomics identified novel metabolites associated with risk of type 2 diabetes in two prospective cohorts of Chinese adults. *Int J Epidemiol*. 2016;45(5):1507–16.
- 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.
- Wang-Sattler R, Yu Z, Herder C, Messias AC, Floegel A, He Y, et al. Novel biomarkers for pre-diabetes identified by metabolomics. *Mol Syst Biol*. 2012;8:615.
- Ferrannini E, Natali A, Camastra S, Nannipieri M, Mari A, Adam KP, et al. Early metabolic markers of the development of dysglycemia and type 2 diabetes and their physiological significance. *Diabetes*. 2013;62(5):1730–7.
- Tillin T, Hughes AD, Wang Q, Würzt 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.
- Zhao J, Zhu Y, Hyun N, Zeng D, Uppal K, Tran VT, et al. Novel metabolic markers for the risk of diabetes development in American Indians. *Diabetes Care*. 2015;38(2):220–7.
- Peddinti G, Cobb J, Yengo L, Froguel P, Kravić J, Balkau B, et al. Early metabolic markers identify potential targets for the prevention of type 2 diabetes. *Diabetologia*. 2017;60(9):1740–50.
- Sudlow C, Gallacher J, Allen N, Beral V, Burton P, Danesh J, et al. UK Biobank: an open access resource for identifying the causes of a wide range of complex diseases of middle and old age. *PLoS Med*. 2015;12(3):e1001779.
- Allen N, Arnold M, Parish S, Hill M, Sheard S, Callen H, et al. Approaches to minimising the epidemiological impact of sources of systematic and random variation that may affect biochemistry assay data in UK Biobank. *Wellcome Open Res*. 2021;5:222.
- Littlejohns TJ, Holliday J, Gibson LM, Garratt S, Oesingmann N, Alfaro-Almagro F, et al. The UK Biobank imaging enhancement of 100,000 participants: rationale, data collection, management and future directions. *Nat Commun*. 2020;11(1):2624.
- UK Biobank. Repeat assessment data, version 1.0: UK Biobank; 2013.
- Würzt 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.
- Soininen P, Kangas AJ, Würzt 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.
- Julkunen H, Cichońska A, Slagboom PE, Würzt P, Nightingale Health UKBl. Metabolic biomarker profiling for identification of susceptibility to severe pneumonia and COVID-19 in the general population. *Elife*. 2021;10:e63033.
- Benjamini Y, Hochberg Y. Controlling the false discovery rate: a practical and powerful approach to multiple testing. *J R Stat Soc Ser B Methodol*. 1995;57(1):289–300.
- Harrell FE Jr, Lee KL, Mark DB. Multivariable prognostic models: issues in developing models, evaluating assumptions and adequacy, and measuring and reducing errors. *Stat Med*. 1996;15(4):361–87.
- Pencina MJ, D'Agostino RB Sr, D'Agostino RB Jr, Vasan RS. Evaluating the added predictive ability of a new marker: from area under the ROC curve to reclassification and beyond. *Stat Med*. 2008;27(2):157–72.
- 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.
- Ahola-Olli AV, Mustelin L, Kalimeri M, Kettunen J, Jokelainen J, Auvinen J, et al. Circulating metabolites and the risk of type 2 diabetes: a prospective study of 11,896 young adults from four Finnish cohorts. *Diabetologia*. 2019;62(12):2298–309.
- Mackey RH, Mora S, Bertoni AG, Wassel CL, Carnethon MR, Sibley CT, et al. Lipoprotein particles and incident type 2 diabetes in the multi-ethnic study of atherosclerosis. *Diabetes Care*. 2015;38(4):628–36.
- Garvey WT, Kwon S, Zheng D, Shaughnessy S, Wallace P, Hutto A, et al. Effects of insulin resistance and type 2 diabetes on lipoprotein subclass particle size and concentration determined by nuclear magnetic resonance. *Diabetes*. 2003;52(2):453–62.
- Mahendran Y, Jonsson A, Have CT, Allin KH, Witte DR, Jørgensen ME, et al. Genetic evidence of a causal effect of insulin resistance on branched-chain amino acid levels. *Diabetologia*. 2017;60(5):873–8.
- 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.
- Qian F, Ardisson Korat AV, Imamura F, Marklund M, Tintle N, Virtanen JK, et al. N-3 fatty acid biomarkers and incident type 2 diabetes: an individual participant-level pooling project of 20 prospective cohort studies. *Diabetes Care*. 2021;44:1133–42.
- Pearson E, Adamski J. The search for predictive metabolic biomarkers for incident T2DM. *Nat Rev Endocrinol*. 2018;14(8):444–6.
- Fall T, Salihovic S, Brandmaier S, Nowak C, Ganna A, Gustafsson S, et al. Non-targeted metabolomics combined with genetic analyses identifies bile acid synthesis and phospholipid metabolism as being associated with incident type 2 diabetes. *Diabetologia*. 2016;59(10):2114–24.
- Herder C, Kowall B, Tabak AG, Rathmann W. The potential of novel biomarkers to improve risk prediction of type 2 diabetes. *Diabetologia*. 2014;57(1):16–29.
- Tikkanen E, Jägerroos V, Rodosthenous R, Holmes MV, Sattar N, Ala-Korpela M, Jousilahti P, Lundqvist A, Perola M, Salomaa V et al. Metabolic Biomarkers for Peripheral Artery Disease Compared with Coronary Artery Disease: Lipoprotein and metabolite profiling of 31,657 individuals from five prospective cohorts. *medRxiv*. 2020.07.24.20158675. <https://doi.org/10.1101/2020.07.24.20158675>.
- Li-Gao R, Hughes DA, le Cessie S, de Mutsert R, den Heijer M, Rosendaal FR, Willems van Dijk K, Timpson NJ, Mook-Kanamori DO: assessment

of reproducibility and biological variability of fasting and postprandial plasma metabolite concentrations using ^1H NMR spectroscopy. *PLoS One*. 2019;14(6):e0218549.

40. Fry A, Littlejohns TJ, Sudlow C, Doherty N, Adamska L, Sprosen T, et al. Comparison of sociodemographic and health-related characteristics of UK Biobank participants with those of the general population. *Am J Epidemiol*. 2017;186(9):1026–34.
41. Batty GD, Gale CR, Kivimäki M, Deary IJ, Bell S. Comparison of risk factor associations in UK Biobank against representative, general population based studies with conventional response rates: prospective cohort study and individual participant meta-analysis. *BMJ*. 2020;368:m131.

Publisher's Note

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

Ready to submit your research? Choose BMC and benefit from:

- fast, convenient online submission
- thorough peer review by experienced researchers in your field
- rapid publication on acceptance
- support for research data, including large and complex data types
- gold Open Access which fosters wider collaboration and increased citations
- maximum visibility for your research: over 100M website views per year

At BMC, research is always in progress.

Learn more biomedcentral.com/submissions

