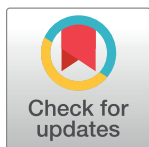RESEARCH ARTICLE

Biomarkers for predicting type 2 diabetes development—Can metabolomics improve on existing biomarkers?

Otto Savolainen[1]*, Björn Fagerberg[2], Mads Vendelbo Lind[1,3], Ann-Sofie Sandberg[1], Alastair B. Ross[1], Göran Bergström[4]

1 Division of Food and Nutrition Science, Department of Biology and Biological Engineering, Chalmers University of Technology, Gothenburg, Sweden, 2 Wallenberg Laboratory for Cardiovascular Research at the Center for Cardiovascular and Metabolic Research, Institute of Medicine, Sahlgrenska Academy at Gothenburg University, Gothenburg, Sweden, 3 Department of Nutrition, Exercise and Sports, Faculty of Science, University of Copenhagen, Frederiksberg, Denmark, 4 Department of Molecular and Clinical Medicine, Sahlgrenska Academy, University of Gothenburg, and Sahlgrenska University Hospital, Gothenburg, Sweden

* otto.savolainen@chalmers.se

Abstract

Aim

The aim was to determine if metabolomics could be used to build a predictive model for type 2 diabetes (T2D) risk that would improve prediction of T2D over current risk markers.

Methods

Gas chromatography-tandem mass spectrometry metabolomics was used in a nested case-control study based on a screening sample of 64-year-old Caucasian women (n = 629). Candidate metabolic markers of T2D were identified in plasma obtained at baseline and the power to predict diabetes was tested in 69 incident cases occurring during 5.5 years follow-up. The metabolomics results were used as a standalone prediction model and in combination with established T2D predictive biomarkers for building eight T2D prediction models that were compared with each other based on their sensitivity and selectivity for predicting T2D.

Results

Established markers of T2D (impaired fasting glucose, impaired glucose tolerance, insulin resistance (HOMA), smoking, serum adiponectin) alone, and in combination with metabolomics had the largest areas under the curve (AUC) (0.794 (95% confidence interval [0.738–0.850]) and 0.808 [0.749–0.867] respectively), with the standalone metabolomics model based on nine fasting plasma markers having a lower predictive power (0.657 [0.577–0.736]). Prediction based on non-blood based measures was 0.638 [0.565–0.711]).

Conclusions

Established measures of T2D risk remain the best predictor of T2D risk in this population. Additional markers detected using metabolomics are likely related to these measures as they did not enhance the overall prediction in a combined model.

Data Availability Statement: The data underlying the findings in our study are not freely and directly available in the manuscript, supplemental files, or in a public repository. The reason is that the original approval by the regional ethic´s board and the informed consent from the subjects participating in the studies do not include such a direct, free access. These restrictions are imposed by the Regional ethical review board in Gothenburg (Guldhedsgatan 5, 413 20 Gothenburg). If a reader of PLOS ONE wants access to the data underlying

Introduction

The global pandemic of type 2 diabetes (T2D) is a major challenge for health care systems globally and new tools are needed to reduce the burden of T2D. Prevention before the onset of T2D, or recognising and treating the early stages is the most cost effective way to treat T2D, before major systemic damage such as microvascular complications, retinopathy and nephropathy occur [1]. Population-based screening for T2D could be a useful tool to aid in efforts for prevention and early treatment, and number of diagnostic tools are available for early diagnosis of T2D risk, including impaired fasting glucose (IFG) test, impaired glucose tolerance (IGT) test, combined glucose tolerance (CGT) test, insulin sensitivity indexes, and anthropometric measurements. Many of these tools require a substantial investment of both patient and clinical time. New tools are needed to complement current diagnostic measures for predicting likely T2D incidence if population-based screening for T2D risk is to be carried out. Along with the well-established measures of T2D, such as fasting glucose and glucose tolerance tests, also non-glucose or insulin-related biomarkers for T2D exist, including low plasma adiponectin concentration that has been found to be a strong predictor for future T2D development [2,3]. This suggests that there is scope for establishing complementary biomarkers of T2D risk.

Emerging technologies, such as metabolomics, are increasingly popular in epidemiology [4] to capture broad based information about large sample sets. The possibility to find novel associations between disease risk and biochemical markers makes it an attractive tool not only for categorising disease states but also for finding new predictive biomarkers. Metabolomics has also proved useful in T2D research and has revealed new associations between metabolites and T2D [5], most notably that elevated branched chain amino acids (BCAA) are associated with hyperglycaemia and may be an early biomarker of T2D risk [6]. The involvement of BCAA metabolism in development of hyperglycaemia has been confirmed in several follow up studies [7,8], showing that it precedes diabetes development. Other potential metabolite biomarkers of diabetes risk include dimethylglycine (DMG) [9], 2-amino adipic acid and glycine [10,11], opening up not only new biomarker possibilities, but giving a broader view of T2D beyond those mechanisms immediately involved in glucose metabolism. Overall there is now a considerable body of research supporting metabolomics as a powerful tool for unravelling early metabolome differences related to later T2D risk [12].

We hypothesised that by using metabolomics we would find biomarkers for T2D that would improve current tools used to predict and diagnose T2D. For this we used gas chromatography-tandem mass spectrometry (GC-MS/MS) metabolomics and built a new predictive model of T2D development in a cohort of 64 year old women, and compared this with models based on existing markers of T2D risk.

Materials and methods

Study participants

Both studies have been carried out in accordance with the Declaration of Helsinki and were approved by the Regional Ethical Review Board in Gothenburg and all participants gave written informed consent to participate.

A population-based cohort was initiated by inviting all 64-year-old women from the Gothenburg region to participate in a screening examination for T2D [13]. The screening examination included fasting capillary whole blood glucose measurements in women with overt diabetes and repeated oral glucose tolerance tests (OGTTs) [2,13] in women without overt diabetes. The definition of IGT and diabetes were based on WHO criteria [14]. The means of the two oral glucose tolerance tests were used in the classification of IGT at baseline. From this

Table 1. Baseline characteristics of the 64-year-old women.

	No diabetes at follow-up		T2D (n = 202)	New T2D at follow-up (n = 69)
	NGT(n = 188)	IGT(n = 203)		
Family history of T2D, n (%)	38(19)	81(40)	67(33)	29(42)
Smoking, n (%)				
Never	79(42)	89(44)	91(45)	24(35)
Previous	73(39)	77(38)	64(32)	30(43)
Current	36(19)	37(18)	47(23)	15(22)
Alcohol consumption, g/day	8.8(8.0)	7.2(7.1)	8.3(8.8)	7.7(8.0)
Waist circumference, cm	88(9)	92(12)	97(11)	95(13)
Systolic blood pressure, mm Hg	136(16)	149(17)	151(18)	148(17)
Serum adiponectin, μg/mL	17.5(7.1)	14.1(6.7)	12.5(6.2)	12.4(6.8)
HOMA IR (pmol/L)· (mmol/L)	1.2(0.6)	1.8(1.3)	2.7(1.5)	2.2(1.4)
IGT, n (%)				48(70)
IFG, n (%)				19(28)

All values are mean(+/-standard deviation) if not otherwise stated

initial screened cohort, cases with diabetes (9.5%), and impaired glucose tolerance (IGT) (14.4%) and a random sample of healthy subjects with normal glucose tolerance (NGT) were included in a nested case–control study, which included a baseline and a follow-up examination after 5.5 years. The baseline examination included anthropometry, assessment of lifestyle factors, fasting blood glucose, oral glucose tolerance test, plasma insulin and serum adiponectin [2]. With the exception of adiponectin, these examinations were also carried out at follow-up. The characteristics of the study subjects are shown in Table 1. In total, 629 women participated in the first examination and samples from 614 subjects were analysed by GC-MS/MS metabolomics (T1D n = 14, T2D n = 202; IGT, n = 203; NGT, n = 188) [2]. T1D was defined as serum GAD antibody level ≥4.6.

The re-examination was performed after 5.5 years in 500 women with diabetes (n = 159), IGT (n = 174) or NGT (n = 167) according to the classification at baseline. The remaining 129 women from the original study did not participate in the re-examination because of death (n = 23), severe disease (n = 3), because they were no longer living in the area (n = 12) or were unwilling to participate (n = 91) [2].

GC-MS/MS metabolomics analysis

GC-MS/MS metabolomics was performed as previously described [15] using plasma extracted with methanol:water, derivitised using methoxymation followed by silylation. Samples were injected onto a Shimadzu GCMS TQ-8030 GC-MS/MS (Shimadzu Europa GmbH, Duisberg, Germany) and both scan and multiple reaction monitoring data was collected for further data analysis.

Metabolomics data processing

A MATLAB (Mathworks, Natick, MA, USA) script and database developed at the Swedish Metabolomics Centre (Umeå, Sweden) was used for targeted analysis of the full scan data [16]. Internal standard normalisation of raw data was performed using the method of Jonsson et al [16]. Internal standard variables from each sample were scaled to unit variance and modelled together using principle components analysis (SIMCA+, Umetrics AB, Sweden) to describe overall variation in multivariate space. Vectors from the first component were then used to

correct for analytical variation between samples. The multiple reaction monitoring (MRM) data was processed and normalised as described earlier [15]. The two data sets, full scan and MRM, were then merged for statistical analyses.

Biochemical analysis

Blood glucose, insulin and adiponectin measurements were performed using standard clinical chemistry techniques at the clinical laboratory of the Sahlgrenska University Hospital, Gothenburg, as previously described [2].

Data analysis strategy

We used baseline data in order to explore potential metabolomics markers of T2D after exclusion of medicated T2D cases and the 69 participants who developed T2D during follow-up. These subjects were used for testing of the prediction models. The candidate markers for detection of incident T2D were tested on those subjects who developed T2D during 5.5 years of follow-up. Sixty of the 69 women who developed T2D during the follow up period had IGT at baseline. The subjects with T1D at baseline were not included in the analysis.

In the screening analyses the NGT and IGT groups, excluding the 69 women who developed T2D during follow up, were compared with the T2D group in two pairwise comparisons (NGT vs. T2D and IGT vs. T2D). Logistic regression adjusted for analytical batch was used to identify potential metabolic markers. Metabolites found to be significant in both of the comparisons were considered as potential predictive metabolites. To avoid statistical overfitting of the prediction models the number of metabolites used for prediction of incident T2D was restricted to the nine that separated T2D at baseline with the highest statistical difference.

In total eight different prediction models (Table 2) were tested for detection of the subjects that developed T2D (n = 69). The rationale for constructing these models was an attempt to delineate the importance of the metabolomics markers in the present study in relation to established predictors of future T2D (see Table 2): Model 1 –based on established non-invasive data indicating increased T2D risk [17]; Model 2 –same as Model 1 and the candidate metabolomic markers in the present study (Table 3); Model 3—serum adiponectin, HOMA IR, smoking, IGT, IFG, which are well established invasive risk markers of T2D [2]; Model 4 –Model 3 and candidate metabolic markers; Model 5 –the present candidate metabolic markers; Model 6 – was developed to determine if the 9 metabolites used in the metabolomics model could be

Table 2. Models used for prediction of incidence of type 2 diabetes with abbreviations and variables included in the models.

Model [ref]	Abbreviation	Variables
1: Non-invasive [2,17]	NI	waist circumference, alcohol consumption, smoking, systolic blood pressure, family history of T2D
2: Non-invasive+metabolomics	NIMet	AM + sorbitol, galacticol, mannose, galactose, uric acid, oxalic acid, glucaric acid-1,4-lactone, 3-methyl-2-oxopentanoic acid, 2-hydroxybutyric acid
3: Adiponectin [2]	AdM	serum adiponectin, HOMA IR, smoking, IGT, IFG
4: Adiponectin + metabolomics	AdMMet	AdM + sorbitol, galacticol, mannose, galactose, uric acid, oxalic acid, glucaric acid-1,4-lactone, 3-methyl-2-oxopentanoic acid, 2-hydroxybutyric acid
5: Metabolomics model	Met	sorbitol, galacticol, mannose, galactose, uric acid, oxalic acid, glucaric acid-1,4-lactone, 3-methyl-2-oxopentanoic acid, 2-hydroxybutyric acid
6: Optimized metabolomics	OMet	galacticol, mannose, galactose and 2-hydroxybutyric acid
7: Adiponectin without glucose measures	AdM2	serum adiponectin, HOMA IR, smoking
8: Optimized metabolomics with glucose measures	OMet2	IFG, IGT, galacticol, mannose, galactose and 2-hydroxybutyric acid

https://doi.org/10.1371/journal.pone.0177738.t002

Table 3. Metabolites used in prediction of incident T2D with average P-value, false discovery rate (FDR), odds ratio (OR) and 95% confidence interval (95% CI) from the pairwise comparisons of the NGT and IGT groups to the T2D group.

Metabolite	P-value	FDR	OR	95% CI
Sorbitol	7.54E-13	1.86E-11	11.15	6.06–22.65
Galacticol	4.85E-10	7.37E-11	6.56	3.90–11.95
Mannose	3.72E-08	2.66E-08	3.89	2.52–6.35
Galactose	9.95E-06	4.56E-05	2.44	1.68–3.69
Uric acid	1.10E-08	5.20E-08	3.22	2.21–4.89
Oxalic acid	2.19E-05	2.70E-07	2.56	1.79–3.79
Glucaric acid-1,4-lactone	8.68E-07	3.29E-09	3.80	2.47–6.16
3-Methyl-2-oxopentanoic acid	4.18E-06	1.74E-05	2.96	1.90–4.98
2-Hydroxybutyric acid	3.32E-06	5.61E-05	3.40	2.17–5.75

reduced with minimal loss of prediction to make an optimised metabolomics model; Model 7 —serum adiponectin, HOMA IR, smoking, without prediabetic blood glucose concentrations; Model 8 –Model 6 and IFG, IGT, which are strong predictors of T2D, in order to test the suggestion that these are the strongest predictors of T2D risk in this cohort.

Statistical analysis

All statistical tests were performed using the R statistical environment (version 3.0.2). Data was checked for normal distribution and skewed variables were log transformed. Multivariate logistic regression models were used for pair-wise comparisons between the different groups (see Data Analysis Strategy). To put equal statistical weight for all metabolites, each metabolite was scaled to have a mean of zero and a standard deviation of one. Correction for multiple testing was done using the Benjamini-Hochberg method [18]. Metabolomics data results were reported as odds ratio (OR) for 1 standard deviation change in concentration of metabolite for significant metabolites. Area under the receiver operating characteristics curve (AUROC) was used to assess the discriminating power of the prediction models. AUROC was calculated with the trapezoidal rule using the R-package pROC (version 1.8). Using the same R-package the different ROC curves were tested for significant differences using a non-parametric comparison as described earlier [19]. Model reduction for the 9 significant metabolites was done by finding the optimal model using R package glmulti [20]. Akaikes information criterion (AIC) was used to evaluate goodness of fit of the models [21]. P-values corrected for false discovery rate (P_{corr}) and differences between ROC curves were considered significant at P<0.05.

Results

The characteristics of the patients included in the analysis are shown in Table 1.

Metabolomics

The pairwise comparisons between the glucose tolerance groups found 48 metabolites that differed significantly (P_{corr}<0.05) between NGT and T2D subjects and 17 when comparing subjects with IGT to subjects with T2D. The two lists were compared to find those that were significant in both comparisons, and the nine metabolites with the lowest average corrected P-values were selected for the overall metabolomics model. These metabolites were sorbitol, galacticol, mannose, galactose, uric acid, oxalic acid, glucaric acid-1,4-lactone, 3-methyl-2-oxopentanoic acid and 2-hydroxybutyric acid (Table 3), which were all positively associated with IGT and T2D.

Prediction of type 2 diabetes incidence

To investigate if the baseline concentrations of the selected metabolites could be used for prediction of incident T2D after 5 years and/or add value to the more traditional predictors, different ROC curves were constructed and compared with each other. AUC values for the different models with 95% confidence intervals and the change to the AUC by addition of the metabolomics model (%) are presented in Table 4 and p-values for the model comparisons in Table 5. From the eight models the combination of the adiponectin model and metabolomics provided the best sensitivity and selectivity (AUC = 0.8078[0.7491–0.8666]) for detecting the cases developing T2D after 5 years. However, the increase of 1.7% in AUC when compared with the adiponectin alone model was not significant (P = 0.74) suggesting that the addition of the metabolomics set does not improve sensitivity and selectivity of the detection of the subjects at risk of developing T2D compared to the adiponectin model alone. The combination of the metabolomics model with the non-invasive model increased AUC by 9.7% (AUC = 0.6377 [0.5645–0.7108] and 0.6997 [0.6261–0.7733], respectively), though this was not significant (P = 0.24). The combination of the metabolomics model to the non-invasive model numerically increased the AUC and, thus, also sensitivity and selectivity of the prediction, though not significantly. The metabolomics model alone provides similar prediction to the non-invasive model (P = 0.73).

To further evaluate if the variables in the metabolomics model can be further reduced to simplify data acquisition we tested if optimization of the model against AIC criteria can provide a model with fewer variables and similar prediction. The optimization further reduced the number of metabolites in the model from nine to four with an AUC of 0.6562 [0.5789–0.7335] for the optimized metabolomics model, a difference of only -0.0004 (P = 0.98) compared with the original metabolomics model. This model included galacticol, mannose, galactose and 2-hydroxybutyric acid as predictors of T2D.

We further investigated if prediction of the adiponectin model was driven by the non-fasting glucose measurements and found that when IGT and IFG classifications were removed from the model its AUC decreased by 21.2% (P = 0.003). There was no difference between the adiponectin model without the glucose measures IGT and IFG when compared to the metabolomics and non-invasive models (P = 0.97 and 0.75 respectively). This led us to conclude that the glucose measures IFG and IGT were driving the prediction model. This was further explored by testing if addition of IFG and IGT in the optimized metabolomics model (OM2)

Table 4. AUC values with 95% CI for each model and the % change in AUC where applicable.

Model[2]	AUC	Change in AUC (%)[1]
1: NI	0.6377 (0.5645–0.7108)	
2: NIMet	0.6997 (0.6261–0.7733)	+9.7
3: AdM	0.7941 (0.7382–0.8500)	
4: AdMMet	0.8078 (0.7491–0.8666)	+1.7
5: Met	0.6566 (0.5771–0.7361)	
6: OMet	0.6562 (0.5789–0.7335)	-0.1
7: AdM2	0.6550 (0.5800–0.7290)	
8: OMet2	0.7760 (0.7130–0.8390)	

[1]Change in AUC after addition of metabolomics into the prediction model

[2]Model 1: Non-invasive, Model 2: Noninvasive + metabolomics, Model 3: Adiponectin, Model 4: Adiponectin + metabolomics, Model 5: Metabolomics, Model 6: Optimized metabolomics, Model 7: Adiponectin without glucose measures and Model 8: Optimized metabolomics with glucose measures

Table 5. Comparison of the different ROC-prediction models based on known markers of type 2 diabetes risk and those derived using metabolomics. Numbers are P-values and the ones highlighted in grey are significantly different at P<0.05.

	Model 1	Model 2	Model 3	Model 4	Model 5	Model 6	Model 7	Model 8
Model 1	x	0.2419	0.0009	0.0004	0.7313	0.7330	0.7486	0.0050
Model 2	x	x	0.0456	0.0247	0.4363	0.4248	0.4007	0.1212
Model 3	x	x	x	0.7406	0.0057	0.0047	0.0035	0.6780
Model 4	x	x	x	x	0.0028	0.0023	0.0016	0.4729
Model 5	x	x	x	x	x	0.9764	0.9733	0.2094
Model 6	x	x	x	x	x	x	0.9789	0.0184
Model 7	x	x	x	x	x	x	x	0.0147
Model 8	x	x	x	x	x	x	x	x

Model 1: Non-invasive, Model 2: Noninvasive + metabolomics, Model 3: Adiponectin, Model 4: Adiponectin + metabolomics, Model 5: Metabolomics, Model 6: Optimized metabolomics, Model 7: Adiponectin without glucose measures and Model 8: Optimized metabolomics with glucose measures

https://doi.org/10.1371/journal.pone.0177738.t005

can provide similar prediction to that of the original adiponectin model (AdM). There were no significant differences observed between OM2 and AdM (P = 0.68) confirming that non-fasting measures of glucose metabolism (IFG and IGT) were responsible for better prediction of T2D rather than adiponectin alone.

Discussion

In this study we evaluated if using metabolomics can add value to the prediction of the onset of T2D over the already established T2D predictors in older Swedish women. The results found that the adiponectin model previously described for this cohort [2] was the strongest predictor of development of T2D, while the metabolomics model performed similarly to the non-invasive model, with 17% lower AUC value than the adiponectin model. The association between circulating adiponectin levels, insulin resistance and T2D is strong and has been previously found to be a good predictor of T2D [22,23]. However, when removing IFG and IGT measures from the adiponectin model, all models performed similarly for prediction of T2D, suggesting that these direct measures remain the strongest predictors of T2D risk in this cohort. This is not surprising given that diabetes stems from elevated glucose concentrations and is diagnosed based on these measures. A further observation is that there is no support for the concept that genetically low circulating adiponectin levels might be causally associated with increased risk of diabetes [22].

Metabolomics

The objective of this work was to find predictive markers of T2D rather than look for mechanisms behind development of T2D, though these metabolites may provide information on potential mechanisms. The high number of sugars that were predictive is perhaps not so surprising given the overall effects of T2D on glucose metabolism and activation of the polyol pathway [24,25] which in the present work was seen as increased circulating levels of sorbitol, a reduction product of glucose, and galacticol, a reduction product of galactose. An advantage of applying predictive modelling to the metabolomics data is that these predictive metabolites are selected in an unbiased manner, bringing attention to potential links between some metabolites that hitherto have not been linked to the development of T2D. Such methods could later be used to build targeted methods measuring the wide range of biomarkers found to be predictive of T2D in earlier studies [26]. This could lead to a method for predicting diabetes development

that accounts for a far broader range of factors beyond the well-defined symptoms of elevated fasting and post-prandial glucose concentrations.

Comparability to earlier studies

A number of studies have used mass spectrometry based metabolomics for both discovering metabolites associated with glucose tolerance and for prediction of the disease when prospective data has been available [5,6,10,27–30]. Mass spectrometry based metabolomics have also been applied in prediction of other diseases, such as breast cancer [31], and the overall use of the method in epidemiology has been reviewed recently [4]. In the current literature studies utilizing prospective data and metabolomics AROC values for prediction of incident T2D range between 0.52 and 0.912 depending on the use of different variables and metabolomics data [10,11,28,32]. The present findings are comparable to the earlier studies and the sensitivity and selectivity of the prediction models are within the range of reported AROC values. Much of the variation in the predictions in these studies can be due to the different cohorts used for the studies as more diverse cohorts will give poorer prediction while more homogenous cohorts will have better prediction. Although previous studies have highlighted aromatic and branched chain amino acids (BCAA) to be associated with T2D development in other cohorts [6,28,33], these were not found among the nine most significant biomarkers used for prediction of T2D in this study, though the BCAA were still found to be associated with glucose tolerance.

Insulin and glucose, in combination with anthropometric and lifestyle related variables are the main diagnostic tools for T2D, and have been also used to predict the onset of T2D [5], though their use remains primarily diagnostic. Recently studies utilising metabolomics have reported additional selectivity and/or sensitivity in the prediction models of T2D after inclusion of metabolomics data in their prediction models [10,11,28], in line with our findings of metabolomics being able to add value to some of the existing prediction models. However, more advanced prediction models, such as those including measurements of diabetes status (e.g. IFG and IGT) and related compounds such as insulin and adiponectin were not significantly improved by the addition of metabolomics data. There was also no significant difference when comparing the optimized metabolomics model with the addition of IFG and IGT classification to the adiponectin model. As the diagnosis of T2D is based on fasting and 2-hour glucose, early measurements of these are by nature explanatory for later T2D. This makes prediction models complicated to build because patients could be close to the threshold of diagnosis and a simple measurement imprecision (199 mg/dl vs. 201 mg/dl) could change their status from IGT to T2D. Such small differences might have large impact on diagnosis but little impact on possible predictive biomarkers. Thus, finding the perfect prediction biomarker or perfect set of biomarkers based on metabolomics alone might be more complex than apparent at first glance and the standard classification tools, IFG and IGT, remain useful tools for T2D prediction.

Applicability of the results

With the increasing burden of non-communicable diseases in the healthcare systems worldwide, using population-wide prediction of the onset of diseases, such as T2D, is an intriguing solution to counter the rise in T2D incidence. In this study the traditional measures of T2D, insulin and OGTT are among the strongest predictors for future T2D, which is expected as these form the basis of T2D diagnosis. We found that the use of metabolite markers determined by metabolomics could achieve similar prediction to these measures, as well as other established risk factors including adiponectin and anthropometry. In order to have useful

biomarker-based prediction of T2D it is necessary to have tests that can be quickly and easily carried out to minimise cost and patient time. Prediction based on a single sample is less of a burden for both patients and clinical staff, provided it has adequate precision, and can also be applied in population based cohorts for mapping of diabetes risk in larger groups. Our finding that four metabolites detected using metabolomics can predict T2D risk at a similar rate to established measures suggests that there is a future scope for adding new measures of T2D risk. Although the method used to discover these markers is advanced, mass spectrometers are now common in clinical laboratories and methodology for measurement of the best biomarkers identified using metabolomics should not be a barrier to implementation after further validation in other cohorts. Ultimately a combination of different markers that reflect different risk factors and mechanisms behind T2D development could be used to get not only a measure of T2D risk, but also an indication of what is inducing dysregulation of glucose metabolism, which will be more instructive for clinicians treating people at risk of T2D.

The limitations of the present study are that we only studied women at the age of 64 and that the sample size of around 600 subjects, while adequate for detecting differences in glucose tolerance, there may not have been enough new T2D cases over five years for building a robust general prediction model based on metabolomics data only. Women at this age have a high and rising incidence of diabetes [34], making this an important target group for prediction of T2D. In this population diabetes is often undiagnosed for many years unless an oral glucose tolerance test is performed. In the present study the majority of cases with T2D at baseline were newly diagnosed and had no treatment that might have interfered with the metabolomics measurements. Selection of this specific population, while limiting the generalisability of the results, also removes variation due to gender and age, making the cohort suitable for exploratory work on T2D.

Conclusions

We have established a set of metabolites that can predict T2D risk in 64-year-old women with similar precision to established predictors of T2D risk. These metabolites do not improve on current prediction models but can be used as an alternative method for screening purposes that could be more applicable for larger population based studies. The results suggest that in this population, commonly used clinical tools provide adequate prediction of future T2D. Thus, finding suitable tools to predict T2D is lesser of an issue than implementing widespread screening programmes to identify people at risk of T2D.

Acknowledgments

Author Contributions

Conceptualization: OS ABR BF GB A-SS.

Data curation: BF GB OS MVL.

Formal analysis: OS MVL ABR.

Funding acquisition: BF GB ABR A-SS.

Investigation: BF OS.

Methodology: OS ABR MVL GB BF.

Project administration: OS ABR.

Resources: OS ABR BF GB A-SS.

Writing – original draft: OS.

Writing – review & editing: ABR MVL AS-S BF BG OS.

References

1. Bailes BK. Diabetes Mellitus and its Chronic Complications. AORN Journal. 2002; 76: 265–282. https://doi.org/10.1016/S0001-2092(06)61065-X

2. Fagerberg B, Kellis D, Bergstrom G, Behre CJ. Adiponectin in relation to insulin sensitivity and insulin secretion in the development of type 2 diabetes: a prospective study in 64-year-old women. Journal of internal medicine. Blackwell Publishing Ltd; 2011; 269: 636–643. https://doi.org/10.1111/j.1365-2796.2010.02336.x PMID: 21198995

3. Li S, Shin HJ, Ding EL, van Dam RM. Adiponectin Levels and Risk of Type 2 Diabetes: A Systematic Review and Meta-analysis. JAMA. American Medical Association; 2009; 302: 179–188. https://doi.org/10.1001/jama.2009.976 PMID: 19584347

4. Lind MV, Savolainen OI, Ross AB. The use of mass spectrometry for analysing metabolite biomarkers in epidemiology: methodological and statistical considerations for application to large numbers of biological samples. Eur J Epidemiol. Springer Netherlands; 2016;: 1–17. https://doi.org/10.1007/s10654-016-0166-2 PMID: 27230258

5. Herder C, Kowall B, Tabak AG, Rathmann W. The potential of novel biomarkers to improve risk prediction of type 2 diabetes. Diabetologia. 2014; 57: 16–29. https://doi.org/10.1007/s00125-013-3061-3 PMID: 24078135

6. Menni C, Fauman E, Erte I, Perry JRB, Kastenmüller G, Shin S-Y, et al. Biomarkers for type 2 diabetes and impaired fasting glucose using a nontargeted metabolomics approach. Diabetes. American Diabetes Association; 2013; 62: 4270–4276. https://doi.org/10.2337/db13-0570 PMID: 23884885

7. Palmer ND, Stevens RD, Antinozzi PA, Anderson A, Bergman RN, Wagenknecht LE, et al. Metabolomic Profile Associated With Insulin Resistance and Conversion to Diabetes in the Insulin Resistance Atherosclerosis Study. The Journal of Clinical Endocrinology & Metabolism. Endocrine Society Chevy Chase, MD; 2014; 100: E463–E468. https://doi.org/10.1210/jc.2014-2357 PMID: 25423564

8. Wang TJ, Larson MG, Vasan RS, Cheng S, Rhee EP, McCabe E, et al. Metabolite profiles and the risk of developing diabetes. Nature medicine. Nature Publishing Group; 2011; 17: 448–453. https://doi.org/10.1038/nm.2307 PMID: 21423183

9. Magnusson M, Wang TJ, Clish C, Engström G, Nilsson P, Gerszten RE, et al. Dimethylglycine Deficiency and the Development of Diabetes. Diabetes. American Diabetes Association; 2015; 64: 3010–3016. https://doi.org/10.2337/db14-1863 PMID: 25795213

10. 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. https://doi.org/10.1038/msb.2012.43 PMID: 23010998

11. Floegel A, Stefan N, Yu Z, Mühlenbruch K, Drogan D, Joost H-G, et al. Identification of serum metabolites associated with risk of type 2 diabetes using a targeted metabolomic approach. Diabetes. American Diabetes Association; 2013; 62: 639–648. https://doi.org/10.2337/db12-0495 PMID: 23043162

12. Brunius C, Shi L, Landberg R. Metabolomics for Improved Understanding and Prediction of Cardiometabolic Diseases—Recent Findings from Human Studies. Curr Nutr Rep. Springer US; 2015; 4: 348–364. https://doi.org/10.1007/s13668-015-0144-4

13. Brohall G, Behre CJ, Hulthe J, Wikstrand J, Fagerberg B. Prevalence of diabetes and impaired glucose tolerance in 64-year-old Swedish women: experiences of using repeated oral glucose tolerance tests. Diabetes care. 2006 ed. 2006; 29: 363–367. PMID: 16443888

14. Alberti KGMM Zimmet PZ, Consultation WHO. Definition, diagnosis and classification of diabetes mellitus and its complications. Part 1: diagnosis and classification of diabetes mellitus. Provisional report of a WHO Consultation. Diabet Med. John Wiley & Sons, Ltd; 1998; 15: 539–553. https://doi.org/10.1002/(SICI)1096-9136(199807)15:7<539::AID-DIA668>3.0.CO;2-S PMID: 9686693

15. Savolainen OI, Sandberg A-S, Ross AB. A Simultaneous Metabolic Profiling and Quantitative Multimetabolite Metabolomic Method for Human Plasma Using Gas-Chromatography Tandem Mass

Spectrometry. J Proteome Res. American Chemical Society; 2016; 15: 259–265. https://doi.org/10.1021/acs.jproteome.5b00790 PMID: 26615962

16. Jonsson P, Johansson AI, Gullberg J, Trygg J, A J, Grung B, et al. High-throughput data analysis for detecting and identifying differences between samples in GC/MS-based metabolomic analyses. Analytical chemistry. 2005 ed. 2005; 77: 5635–5642. https://doi.org/10.1021/ac050601e PMID: 16131076

17. Noble D, Mathur R, Dent T, Meads C, Greenhalgh T. Risk models and scores for type 2 diabetes: systematic review. BMJ. BMJ Group; 2011; 343: d7163–d7163. https://doi.org/10.1136/bmj.d7163 PMID: 22123912

18. Benjamini Y, Hochberg Y. Controlling the false discovery rate: a practical and powerful approach to multiple testing. Journal of the Royal Statistical Society Series B 1995. https://doi.org/10.2307/2346101

19. DeLong ER, DeLong DM, Clarke-Pearson DL. Comparing the Areas under Two or More Correlated Receiver Operating Characteristic Curves: A Nonparametric Approach. Biometrics. 1988; 44: 837. https://doi.org/10.2307/2531595 PMID: 3203132

20. Calcagno V, de Mazancourt C. glmulti: An R Package for Easy Automated Model Selection with (Generalized) Linear Models. Journal of Statistical Software. 2010; 34: 1–29.

21. Akaike H. Information Theory and an Extension of the Maximum Likelihood Principle. Breakthroughs in Statistics. New York, NY: Springer New York; 1992. pp. 610–624. https://doi.org/10.1007/978-1-4612-0919-5_38

22. Yaghootkar H, Lamina C, Scott RA, Dastani Z, Hivert M-F, Warren LL, et al. Mendelian Randomisation Studies Do Not Support a Causal Role for Reduced Circulating Adiponectin Levels in Insulin Resistance and Type 2 Diabetes. Diabetes. American Diabetes Association; 2013; 62: DB_130128–3598. https://doi.org/10.2337/db13-0128 PMID: 23835345

23. Li S, Shin HJ, Ding EL, van Dam RM. Adiponectin levels and risk of type 2 diabetes: a systematic review and meta-analysis. JAMA. American Medical Association; 2009; 302: 179–188. https://doi.org/10.1001/jama.2009.976 PMID: 19584347

24. Lorenzi M. The Polyol Pathway as a Mechanism for Diabetic Retinopathy: Attractive, Elusive, and Resilient. Journal of Diabetes Research. Hindawi Publishing Corporation; 2007; 2007: 1–10. https://doi.org/10.1155/2007/61038 PMID: 18224243

25. Oates PJ. Polyol pathway and diabetic peripheral neuropathy. Neurobiology of Diabetic Neuropathy. Elsevier; 2002. pp. 325–392. https://doi.org/10.1016/S0074-7742(02)50082-9

26. Meiss E, Werner P, John C, Scheja L, Herbach N, Heeren J, et al. Metabolite targeting: development of a comprehensive targeted metabolomics platform for the assessment of diabetes and its complications. Metabolomics: Official journal of the Metabolomic Society. Springer US; 2016; 12: –15. https://doi.org/10.1007/s11306-016-0958-0

27. Wang TJ, Ngo D, Psychogios N, Dejam A, Larson MG, Vasan RS, et al. 2-Aminoadipic acid is a biomarker for diabetes risk. The Journal of clinical investigation. 2013 ed. 2013; 123: 4309–4317. https://doi.org/10.1172/JCI64801 PMID: 24091325

28. Wang TJ, Larson MG, Vasan RS, Cheng S, Rhee EP, McCabe E, et al. Metabolite profiles and the risk of developing diabetes. Nature medicine. 2011 ed. 2011; 17: 448–453. https://doi.org/10.1038/nm.2307 PMID: 21423183

29. Roberts LD, Koulman A, Griffin JL. Towards metabolic biomarkers of insulin resistance and type 2 diabetes: progress from the metabolome. Lancet Diabetes Endocrinol. 2014; 2: 65–75. https://doi.org/10.1016/S2213-8587(13)70143-8 PMID: 24622670

30. Fiehn O, Garvey WT, Newman JW, Lok KH, Hoppel CL, Adams SH. Plasma metabolomic profiles reflective of glucose homeostasis in non-diabetic and type 2 diabetic obese African-American women. Gimble JM, editor. PloS one. 2010; 5: e15234. https://doi.org/10.1371/journal.pone.0015234 PMID: 21170321

31. Bro R, Kamstrup-Nielsen MH, Engelsen SB, Savorani F, Rasmussen MA, Hansen L, et al. Forecasting individual breast cancer risk using plasma metabolomics and biocontours. Metabolomics: Official journal of the Metabolomic Society. Springer US; 2015; 11: 1376–1380. https://doi.org/10.1007/s11306-015-0793-8 PMID: 26366139

32. Ferrannini E, Natali A, Camastra S, Nannipieri M, Mari A, Adam K-P, et al. Early Metabolic Markers of the Development of Dysglycemia and Type 2 Diabetes and Their Physiological Significance. Diabetes. American Diabetes Association; 2013; 62: 1730–1737. https://doi.org/10.2337/db12-0707 PMID: 23160532

33. Newgard CB, An J, Bain JR, Muehlbauer MJ, Stevens RD, Lien LF, et al. A branched-chain amino acid-related metabolic signature that differentiates obese and lean humans and contributes to insulin resistance. Cell metabolism. 2009; 9: 311–326. https://doi.org/10.1016/j.cmet.2009.02.002 PMID: 19356713

34. The DECODE Study Group. Age- and sex-specific prevalences of diabetes and impaired glucose regulation in 13 European cohorts. Diabetes care. 2002nd ed. 2003; 26: 61–69. PMID: 12502659